



**UNIVERSIDADE FEDERAL DA PARAÍBA**  
**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE**

**JULIANA BARBOSA MEDEIROS**

**ANÁLISE DA QUALIDADE DOS DADOS POPULACIONAIS E**  
**REGISTROS DE ÓBITOS DOS CENTENÁRIOS DO SEMIÁRIDO**  
**BRASILEIRO NO PERÍODO DE 2000 A 2020**

João Pessoa-PB

2023

**JULIANA BARBOSA MEDEIROS**

**ANÁLISE DA QUALIDADE DOS DADOS POPULACIONAIS E REGISTROS DE  
ÓBITOS DOS CENTENÁRIOS DO SEMIÁRIDO BRASILEIRO NO PERÍODO DE  
2000 A 2020.**

Trabalho de Tese apresentado ao Programa de Pós-Graduação em Modelos de Decisão e Saúde - Nível Doutorado - do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba.

**Área de Concentração:** Modelos de Decisão

**Orientadores:**

Prof. Dr. Neir Antunes Paes

Prof. Dr. Rodrigo Pinheiro de Toledo Vianna

João Pessoa-PB

2023

UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE

**ANÁLISE DA QUALIDADE DOS DADOS POPULACIONAIS E REGISTROS DE  
ÓBITOS DOS CENTENÁRIOS DO SEMIÁRIDO BRASILEIRO NO PERÍODO DE  
2000 A 2020.**

Trabalho de Tese de investigação científica submetido à Coordenação do Programa de Pós-Graduação em Modelos de Decisão e Saúde do Centro de Ciências Exatas e da Natureza (CCEN), da Universidade Federal da Paraíba (UFPB) como requisito final para Conclusão de Doutorado, apresentada pela doutoranda Juliana Barbosa Medeiros, tendo obtido o conceito APROVADA, conforme apreciação da Banca Examinadora constituída pelos avaliadores:

**BANCA EXAMINADORA**

---

**Prof. Dr. Neir Antunes Paes – UFPB**  
Orientador - Ciências Exatas

---

**Prof. Dr. Rodrigo Pinheiro de Toledo Vianna – UFPB**  
Orientador – Ciências da Saúde

---

**Profª Drª. Everlane Suane de Araújo da Silva– UFPB**  
Examinadora – Interno

---

**Profª Drª. Caliandra Maria Bezerra Luna Lima – UFPB**  
Examinadora – Interno

---

**Profª Drª. Carlos Sérgio Araújo dos Santos - UFCG**  
Examinador – Externo

---

**Profª Drª. Lara Barbosa de Melo Andrade - UFRN**  
Examinadora – Externo

Aprovado em: 29 de março de 2023.

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

M488a Medeiros, Juliana Barbosa.

Análise da qualidade dos dados populacionais e registros de óbitos dos centenários do semiárido brasileiro no período de 2000 a 2020 / Juliana Barbosa Medeiros. - João Pessoa, 2023.  
169 f. : il.

Orientação: Neir Antunes Paes, Rodrigo Pinheiro de Toledo Vianna.

Tese (Doutorado) - UFPB/CCEN.

1. Saúde - Modelos de decisão - Demografia. 2. Idosos centenários. 3. Censo demográfico. 4. Registros de mortalidade. I. Paes, Neir Antunes. II. Vianna, Rodrigo Pinheiro de Toledo. III. Título.

UFPB/BC

CDU 614:314(043)

*“A vida é difícil meu caro,  
E quando vejo um idoso, olho admirado e sempre me pergunto:  
Como ele conseguiu chegar até lá?  
(Matheus Honorato)*

*Dedico este trabalho aos meus três filhos, Ana Júlia, Kauã e Ana Luísa, por causa deles muitos disseram que eu não conseguiria, mas na verdade tê-los em minha vida é ter dia a dia uma fonte inesgotável de força e motivação, por eles, com eles e para eles eu vou além e a frente, sempre!*

## AGRADECIMENTOS

Quero fazer um agradecimento especial ao meu orientador, professor Neir Paes, que com a sua experiência, conhecimento científico e pensamento crítico me apresentou o campo da demografia. E assim foi construída essa jornada, com atenção, zelo e sabedoria, um orientador que acreditou que eu seria capaz, que insistiu e insistiu mais uma vez, encarando a minha resistência com sensibilidade e paciência, sempre mostrando o melhor caminho a seguir. Obrigada acima de tudo pela confiança, obrigada por ter proporcionado em mim uma postura aberta a novos desafios, hoje sei que sou capaz de muito basta ter dedicação e disciplina. Foi uma honra poder contar com a orientação de alguém tão competente e comprometido com o que faz.

Agradecer ao meu esposo Ítalo Kleber, sem ele nada seria possível, seu apoio foi fundamental para eu conseguir chegar até aqui, sempre acreditando em mim, não me deixou desistir, vibrando a cada conquista e avanço. Que presente é caminhar ao seu lado e compartilhar os momentos com você, sejam eles bons ou ruins, foram dias de luta, mas conseguimos. Que parceria a nossa!

À minha família, em especial minha mãe Socorro Barbosa e minhas irmãs Kaline e Liana, pelo carinho, afeto e cuidado, sendo a base de amparo, confiança e cumplicidade, sendo sempre elas meu porto seguro.

À Universidade Federal da Paraíba, lugar especial que faz parte da minha história desde a graduação, que guardo tantas memórias afetivas e lembranças de momentos felizes. Ao Programa de Pós-Graduação em Modelos de Decisão e Saúde e aos professores pela dedicação e conhecimentos compartilhados durante as disciplinas, vivenciando o cotidiano de intensa aprendizagem com foco na interdisciplinaridade e transcendência. Aos funcionários, que sempre foram solícitos e que realizam seu trabalho com seriedade e compromisso.

Aos colegas do PPGMDS, que dádiva e experiência maravilhosa conviver com vocês, quantos momentos de extremo estresse e superação vivenciamos, mas vencemos cada um deles.

À professora Everlane Suane pelo apoio, sensibilidade e compreensão.

A todos do Hospital Universitário Onofre Lopes (HUOL), especialmente à Bárbara Dantas e à minha supervisora Viviany Medeiros, pela compreensão e apoio constante durante a minha jornada de doutoranda e enfermeira, por todo incentivo e atenção aos pedidos de escala e folgas para cursar as disciplinas e concluir essa Tese, sem colocar obstáculos até a conclusão do curso.

## RESUMO

Há um histórico de superestimação da contagem populacional censitária de centenários brasileiros e os registros de óbitos desses indivíduos apresentam problemas de qualidade que podem comprometer seriamente a fidedignidade dos indicadores deles derivados, cujas questões se tornam ainda mais problemáticas quando se trata de regiões como a do Semiárido brasileiro, uma das mais iníquas da América Latina. Nessa perspectiva, o presente estudo teve por objetivo avaliar a qualidade e ajustar os dados populacionais e dos registros de óbitos dos centenários do Semiárido brasileiro no período de 2000 a 2020. Trata-se de um estudo ecológico, transversal, com abordagem quantitativa, referente aos registros de óbitos e dados populacionais de indivíduos com 100 anos ou mais dos 1.262 municípios, regionalizados nas categorias urbano, intermediário e rural do Semiárido brasileiro. Na primeira etapa do estudo foi desenvolvida uma proposta para a avaliação da consistência da contagem dos dados censitários dos centenários. A aplicação desse método revelou uma superestimação dos centenários do Semiárido brasileiro, cujo tamanho da população cresceu no decorrer dos anos. A qualidade da contagem censitária melhorou, passando da classificação *baixa* para *regular*, entre os censos de 2000 e 2010. Porém, do ponto de vista da comparabilidade internacional os dados censitários ainda são considerados de preocupante qualidade, principalmente naqueles municípios caracterizados como rurais. Na análise da qualidade dos registros de óbitos houve um importante avanço no período, porém a cobertura de óbitos como um todo não atingiu níveis acima de 90%. O percentual das causas mal definidas foi classificado como de qualidade “pouco confiável”, de 16% a 30%, para a maioria dos espaços regionalizados. Na Declaração de Óbitos, a variável sociodemográfica “Escaridade” apresentou um percentual de incompletude de 44,6%, enquanto “Assistência médica” ficou acima de 50% em todas as situações analisadas. Após a redistribuição das Causas Mal Definidas, o grupo das doenças do aparelho circulatório foi a causa definida mais frequente entre os centenários, seguida pelas doenças do aparelho respiratório e as doenças endócrinas, nutricionais e metabólicas; juntas responderam por cerca de 78% dos óbitos dos centenários, após a redistribuição. A análise dos fatores associados ao registro das Causas Mal Definidas evidenciou que os solteiros, separados ou divorciados, residentes em municípios rurais/intermediários, assim como aqueles centenários que faleceram fora de estabelecimentos de saúde e sem assistência médica tiveram maiores chances de apresentarem registros de óbitos mal definidos. Os resultados deste estudo propiciaram uma compreensão, em primeira mão, da situação da contagem dos dados populacionais e dos registros de óbitos dos centenários, os quais podem auxiliar nas avaliações e nas tomadas de decisão das políticas de atenção à saúde, programas e serviços geriátricos e de assistência a essa população tão pouco estudada à despeito do acelerado processo de envelhecimento em que se encontra a população brasileira e em particular a do Semiárido.

**Palavras-chave:** Centenários. Censos. Registros de Mortalidade. Confiabilidade dos Dados.

## ABSTRACT

There is an overrated historic in census population counting of Brazilian centenarians and the death records of these individuals have been presenting quality problems that can seriously compromise the trust of the indicators that come from them, whose issues become even more problematic when it comes to regions such as the Brazilian Semiarid, one of the most unjust in Latin America. From this perspective, the present study aimed to analyze the quality of population data and death records of centenarians in the Brazilian Semiarid in the period from 2000 to 2020. This is an ecological, cross-sectional study, with a quantitative approach, referring to death records and population data of individuals aged 100 years or more from 1,226 municipalities, regionalized in the urban, intermediate and rural categories of the Brazilian Semiarid region. The application of this proposal revealed an overestimation of centenarians in the Brazilian Semi-arid region, whose population size has grown over the years. The quality improved from the poor to regular classification between the 2000 and 2010 censuses. However, from the point of view of international comparability, the data are still considered of "low" quality, especially in those areas characterized as rural. In the analysis of the quality of death records there was also an important advance in the period, but the coverage of deaths as a whole did not reach levels above 90%. The percentage of poorly defined causes was classified as having "poor" quality, 16% - 30%, for most of the regionalized areas. In the study of the completeness of the variables contained in Death Certificates, the variable "Education" showed the highest percentages of incompleteness 44.6% among the others studied. The variable "Medical care" presented percentages of missings above 50% in all analyzed situations. After the redistribution of the Ill-Defined Causes, the group of circulatory system diseases was the most frequent defined cause among the centenarians, followed by respiratory system diseases and endocrine, nutritional and metabolic diseases, together accounting for about 78% of deaths of the centenarians after the redistribution. The analysis of the factors associated with the registration of ill-defined causes showed that single, separated or divorced people, residents of rural/intermediate municipalities, as well as those centenarians who died outside of health facilities and without medical assistance have a greater chance of presenting their ill-defined death records. It is expected that this study may bring more assertive lights for a better understanding of the situation of population counting data and death records, as well as assist in evaluations and decision-making in health care policies, programs and geriatric services and assistance to this population so little studied despite the accelerated aging process in which the Brazilian population is found, particularly in the semi-arid region.

**Keywords:** Centenarians. Censuses. Mortality Records. Data Reliability.

## LISTA DE ILUSTRAÇÕES

|                                                                                                                                                                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 2.1- Pirâmide etária do Brasil em 1970.....                                                                                                                                                                                                                            | 24  |
| Figura 2.2 - Pirâmide etária do Brasil em 2010.....                                                                                                                                                                                                                           | 24  |
| Figura 2.3 – Mapa do Semiárido brasileiro .....                                                                                                                                                                                                                               | 55  |
| Figura 3.1 – Exemplo de dendograma.....                                                                                                                                                                                                                                       | 64  |
| Figura 4.1 - Representação dos indicadores de qualidade para avaliação dos dados<br>populacionais dos centenários, para o Semiárido brasileiro, 2000, 2010 e 2019.....                                                                                                        | 99  |
| Figura 4.2.1 - Distribuição do percentual dos óbitos dos centenários por causa mal definida<br>da Declaração de Óbito, entre os homens, nos espaços regionalizados do Semiárido<br>brasileiro, 1999-2001, 2009-2011, 2018-2020.....                                           | 119 |
| Figura 4.2.2: Distribuição do percentual dos óbitos dos centenários por causa mal definida da<br>Declaração de Óbito, entre as mulheres, nos espaços regionalizados do Semiárido brasileiro,<br>1999-2001, 2009-2011, 2018-2020.....                                          | 120 |
| Figura 4.3.1 – Número e Proporção de dados faltantes em diferentes aspectos do banco de<br>dados.....                                                                                                                                                                         | 128 |
| Figura 4.3.2 - Padrões de não resposta das variáveis do estudo, entre homens e mulheres,<br>gráficos de barras com percentagem de casos em cada padrão, Teste MCAR de Little.....                                                                                             | 129 |
| Gráfico 4.3.3 – Representação da dispersão dos pontos, linha de tendência e a equação da reta,<br>método de Ledermann, para sexo masculino, grupo de causa básica do óbito: Doenças do<br>aparelho circulatório, entre os centenários, Semiárido brasileiro, 2010 – 2020..... | 136 |
| Gráfico 4.3.4 – Representação da dispersão dos pontos, linha de tendência e a equação da reta,<br>método de Lederman, para o sexo feminino, grupo de causa básica do óbito: Doenças do<br>aparelho circulatório, entre os centenários, Semiárido brasileiro, 2010 – 2020..... | 136 |
| Figura 4.4.1 - Curva ROC para o modelo final (área sob a curva = 0,696).....                                                                                                                                                                                                  | 145 |
| Figura 4.4.2 – Gráfico dos resíduos padronizados do modelo final.....                                                                                                                                                                                                         | 146 |
| Figura 4.4.3 – Gráfico da Distância de Cook do modelo final.....                                                                                                                                                                                                              | 147 |
| Figura 4.4.4 – Gráfico de Independência dos erros do modelo final.....                                                                                                                                                                                                        | 147 |

## LISTA DE QUADROS

|                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Quadro 2.1 - Critérios utilizados para nova tipologia de classificação municipal.....                                                        | 53  |
| Quadro 3.1 - Descrição do código da variável, variável e categorias das variáveis<br>sociodemográficas contidas na DO e causas de óbito..... | 59  |
| Quadro 3.2 - Descrição dos Capítulos CID-10ª Revisão, correspondentes à variável Causa<br>Básica do óbito da Declaração de Óbito.....        | 60  |
| Quadro 3.3 - Classificação dos escores para avaliar a qualidade dos dados censitários do<br>contingente de idosos centenários. ....          | 67  |
| Quadro 4.4.1 – Variáveis selecionadas para o ajuste do modelo de regressão logística segundo<br>categorias.....                              | 139 |
| Quadro 4.4.2 – Recategorização das variáveis inseridas no ajuste do modelo de regressão<br>logística proposto.....                           | 140 |

## LISTA DE TABELAS

|                                                                                                                                                                                                                                                      |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabela 1 - Distribuição de centenários recenseados no Brasil e nas cinco grandes regiões, segundo o sexo, 2000 e 2010.....                                                                                                                           | 32  |
| Tabela 4.1.1 - Descrição da população total e centenária, entre os espaços regionalizados do Semiárido brasileiro, Nordeste e Brasil.....                                                                                                            | 88  |
| Tabela 4.1.2 - Projeção da população total e centenários, de acordo com os métodos aritmético, geométrico e exponencial, para o Semiárido brasileiro, Nordeste e Brasil.....                                                                         | 91  |
| Tabela 4.1.3 - Média, desvio-padrão, valor mínimo e valor máximo dos indicadores selecionados. Índice de Desenvolvimento Humano (IDH) e expectativa de vida ao nascer, dos 85 países selecionados.....                                               | 92  |
| Tabela 4.1.4 - Coeficiente de Determinação e indicadores da regressão linear simples entre o Índice de Desenvolvimento Humano (IDH) e os indicadores construídos a partir dos dados censitários dos centenários dos 85 países selecionados.....      | 93  |
| Tabela 4.1.5 - Agrupamento dos 85 países -segundo os três indicadores selecionados, IDH e expectativa de vida ao nascer nos quatro clusters estimados pela análise de cluster multivariada.....                                                      | 94  |
| Tabela 4.1.6 - Apresentação dos intervalos de referência para cada indicador selecionado, construídos a partir da segregação dos grupos, para os 85 países selecionados.....                                                                         | 96  |
| Tabela 4.1.7 - Classificação da qualidade dos dados censitários dos idosos centenários, a partir dos três indicadores propostos, de acordo com os pesos dos três intervalos padrão e os cinco escores definidos, para os 85 países selecionados..... | 96  |
| Tabela 4.1.8 - Classificação da qualidade dos dados censitários dos centenários no Brasil, 2010.....                                                                                                                                                 | 99  |
| Tabela 4.1.9 - Descrição do Padrão de qualidade de cada indicador, soma dos pesos e classificação final da qualidade dos dados censitários dos centenários no Semiárido brasileiro, para os anos 2000, 2010.....                                     | 101 |
| Tabela 4.1.10 - Distribuição do número de centenários com base em informações censitárias e estimativas indiretas com base nos óbitos corrigidos, segundo sexo, entre os espaços regionalizados do Semiárido brasileiro, 2000, 2010 e 2019.....      | 103 |
| Tabela 4.1.11 - Razão de sexo dos centenários do Semiárido brasileiro, segundo o Censo 2000 e 2010, Projeção de 2019, estimativas indiretas, 2000, 2010 e 2019.....                                                                                  | 106 |

|                                                                                                                                                                                                                                                      |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabela 4.2.1: Cobertura de óbitos adultos dos espaços regionalizados do Semiárido brasileiro, segundo o sexo, 2000, 2010 e 2019.....                                                                                                                 | 107 |
| Tabela 4.2.2 - Número de óbitos de idosos centenários notificados ao SIM, corrigidos e fatores de correção para os espaços regionalizados e sexo do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.....                                       | 109 |
| Tabela 4.2.3 - Características sociodemográficas e operacionais dos registros de óbitos dos centenários ocorridos no Semiárido brasileiro, 2000 - 2020.....                                                                                          | 111 |
| Tabela 4.2.4 - Descrição das principais causas básicas de óbitos dos centenários, ocorridos no Semiárido brasileiro, 2000 - 2020.....                                                                                                                | 115 |
| Tabela 4.2.5 - Distribuição do percentual de óbitos dos centenários cuja causa básica é desconhecida, segundo sexo, nos espaços regionalizados do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.....                                         | 122 |
| Tabela 4.3.1 - Número e percentual de informações ignoradas ou não preenchidas segundo variáveis selecionadas da Declaração de Óbito dos centenários, por sexo, nos espaços regionalizados do Semiárido brasileiro, nos triênios de 2000 a 2020..... | 125 |
| Tabela 4.3.2 - Eficiência Relativa segundo variáveis sociodemográficas selecionadas de acordo com a escolha de cinco imputações, espaços regionalizados do Semiárido brasileiro, 2000 a 2020.....                                                    | 130 |
| Tabela 4.3.3 - Distribuição do percentual das categorias das variáveis sociodemográficas, antes e depois da imputação de dados, entre os homens, segundo espaços regionalizados, nos triênios de 2000 a 2020.....                                    | 132 |
| Tabela 4.3.4 - Distribuição do percentual das categorias das variáveis sociodemográficas, antes e depois da imputação de dados, entre as mulheres, segundo espaços regionalizados, nos triênios de 2000 a 2020.....                                  | 133 |
| Tabela 4.3.5 - Coeficientes de redistribuição e óbitos para os principais grupos de causas de óbitos por sexo, Semiárido brasileiro, 2010-2020.....                                                                                                  | 137 |
| Tabela 4.4.1- Modelo ajustado, variáveis sociodemográficas associadas a CMD, Centenários, Semiárido brasileiro, 2018-2020.....                                                                                                                       | 141 |
| Tabela 4.4.2 - Tabela de Contingência para o Modelo de Regressão.....                                                                                                                                                                                | 146 |

## LISTA DE ABREVIATURAS

|       |                                                 |
|-------|-------------------------------------------------|
| CID   | Classificação Internacional de Doenças          |
| CMD   | Causa mal definida                              |
| DC    | Distância de Cook                               |
| DN    | Declaração de nascido vivo                      |
| DO    | Declaração de óbito                             |
| EM    | Expextation-Maximization                        |
| ER    | Eficiência relativa                             |
| IBGE  | Instituto Brasileiro de Geografia e Estatística |
| EUA   | Estados Unidos da América                       |
| IC    | Intervalo de confiança                          |
| IDH   | Índice de Desenvolvimento Humano                |
| IM    | Imputação múltipla                              |
| LED   | Laboratório de estudos demográficos             |
| MAR   | Aleatório                                       |
| MCAR  | Completamente aleatório                         |
| MCMC  | Cadeias de markov                               |
| MS    | Ministério da saúde                             |
| NMAR  | Não aleatório                                   |
| OMS   | Organização mundial de saúde                    |
| ONU   | Organização das nações unidas                   |
| OR    | Odds ratio                                      |
| PIB   | Produto interno bruto                           |
| PMM   | Método da média preditiva                       |
| RELAP | Revista latinoamericana de poblacion            |
| RIPSA | Rede interagencial de informações para a saúde  |
| RS    | Razão de sexo                                   |
| SAMU  | Serviço de atendimento móvel de urgência        |
| SEADE | Fundação sistema estadual de análise de dados   |
| SES   | Secretaria estadual de saúde                    |
| SIDRA | Sistema de recuperação automática               |
| SIH   | Sistema de informações hospitalares             |

|        |                                                   |
|--------|---------------------------------------------------|
| SIM    | Sistema de informações sobre mortalidade          |
| SINAN  | Sistemas de informação de agravos de notificações |
| SINASC | Sistema de informações de nascidos vivos          |
| SIS    | Sistema de informações em saúde                   |
| SUDENE | Superintendência de desenvolvimento do Nordeste   |
| UF     | Unidade de federação                              |
| UNDP   | Programa das nações unidas para o desenvolvimento |

## SUMÁRIO

|                                                                                                      |    |
|------------------------------------------------------------------------------------------------------|----|
| 1 INTRODUÇÃO.....                                                                                    | 15 |
| 1.1 Problematização e justificativas .....                                                           | 19 |
| 1.2 Objetivos.....                                                                                   | 21 |
| 1.2.1 Objetivo Geral .....                                                                           | 21 |
| 1.2.2 Objetivos Específicos .....                                                                    | 21 |
| 2 REFERENCIAL TEÓRICO.....                                                                           | 22 |
| 2.1 Envelhecimento Populacional .....                                                                | 22 |
| 2.1.1 Transição Demográfica.....                                                                     | 22 |
| 2.1.2 Transição Epidemiológica .....                                                                 | 26 |
| 2.2 Estudo dos Centenários: o mundo e o regionalismo brasileiro .....                                | 29 |
| 2.3 Fontes de Dados Demográficos.....                                                                | 33 |
| 2.4 Os Sistemas Nacionais de Informações das Estatísticas Vitais.....                                | 35 |
| 2.4.1 Sistemas de Informações do IBGE .....                                                          | 36 |
| 2.4.2 Sistemas de Informações do Ministério da Saúde .....                                           | 37 |
| 2.4.2.1 Sistema de Informações sobre Mortalidade – SIM .....                                         | 38 |
| 2.5 Dados Populacionais e Registros de Óbitos dos Centenários .....                                  | 39 |
| 2.5.1 A qualidade dos dados censitários dos centenários .....                                        | 39 |
| 2.5.2 Os registros de óbitos dos centenários.....                                                    | 43 |
| 2.6 O Problema da Incompletude das Informações.....                                                  | 44 |
| 2.6.1 Imputação dos Dados .....                                                                      | 45 |
| 2.6.1.1 Imputação única.....                                                                         | 47 |
| 2.6.1.2 Imputação Múltipla.....                                                                      | 48 |
| 2.7 Caracterização dos Espaços Regionais.....                                                        | 50 |
| 2.8 Caracterização do Semiárido brasileiro .....                                                     | 54 |
| 3 PROCEDIMENTOS METODOLÓGICOS .....                                                                  | 57 |
| 3.1 Tipologia e Desenvolvimento do Estudo .....                                                      | 57 |
| 3.2 Fontes dos Dados e Variáveis do Estudo .....                                                     | 57 |
| 3.3 Análise da Qualidade dos Dados Populacionais .....                                               | 61 |
| 3.3.1 Proposta de classificação para avaliar a qualidade dos dados censitários dos centenários ..... | 61 |
| 3.3.2 Método indireto para estimar a população de centenários .....                                  | 68 |
| 3.4 Análise da Qualidade dos Registros de Óbitos.....                                                | 70 |
| 3.4.1 Correção dos sub-registros dos óbitos .....                                                    | 71 |

|                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.4.2 Correção dos problemas de incompletude das variáveis .....                                                                                    | 75  |
| 3.4.3 Redistribuição de óbitos por causas mal definidas .....                                                                                       | 79  |
| 3.5 Associação das causas mal definidas com as variáveis da DO .....                                                                                | 80  |
| 3.5.1 O Modelo de Regressão Logística.....                                                                                                          | 81  |
| 3.5.2 Razão de Chances (Odds Ratio) .....                                                                                                           | 84  |
| 3.5.3 Análise ROC.....                                                                                                                              | 85  |
| 3.5.4 Análise de resíduos.....                                                                                                                      | 86  |
| 4 RESULTADOS E DISCUSSÕES.....                                                                                                                      | 88  |
| 4.1 Análise da qualidade dos dados populacionais dos centenários do Semiárido brasileiro ..                                                         | 88  |
| 4.1.1 Panorama populacional dos centenários residentes no Semiárido brasileiro, Nordeste e Brasil.....                                              | 88  |
| 4.1.2 Proposta de classificação para avaliar a qualidade dos dados censitários dos centenários. ....                                                | 91  |
| 4.1.3 Implementação de métodos indiretos para estimar a população centenária para os anos de 2000, 2010 e 2019.....                                 | 102 |
| 4.2 Avaliação da qualidade dos registros de óbitos dos centenários do Semiárido brasileiro                                                          | 107 |
| 4.2.1 Cobertura dos óbitos.....                                                                                                                     | 107 |
| 4.2.2 Principais características sociodemográficas e operacionais dos óbitos dos centenários .....                                                  | 110 |
| 4.2.3 O estudo dos “missing values” e causas mal definidas .....                                                                                    | 113 |
| 4.3 Análise da completude das variáveis contidas nos registros de óbitos dos centenários do Semiárido brasileiro, nos triênios de 2000 a 2020. .... | 124 |
| 4.3.1 Completude das variáveis dos registros de óbitos.....                                                                                         | 124 |
| 4.3.2 Implementação do Método de Imputação Múltipla.....                                                                                            | 127 |
| 4.3.3 Redistribuição de óbitos por causas mal definidas (CMD) .....                                                                                 | 135 |
| 4.4 Associação das causas mal definidas com as variáveis da DO .....                                                                                | 139 |
| 4.4.1 Avaliação do ajuste do modelo.....                                                                                                            | 140 |
| 4.4.2 Estimação dos parâmetros do modelo .....                                                                                                      | 141 |
| 4.4.3 Curva Roc .....                                                                                                                               | 144 |
| 4.4.4 Análise De Resíduos.....                                                                                                                      | 146 |
| 5 CONSIDERAÇÕES FINAIS .....                                                                                                                        | 149 |
| REFERÊNCIAS .....                                                                                                                                   | 154 |

## 1 INTRODUÇÃO

O envelhecimento populacional é um fenômeno mundial decorrente do processo da transição demográfica, principalmente devido ao declínio dos níveis de fecundidade. Como consequência, tem-se uma redução na proporção da população jovem e um consequente aumento da participação de idosos no total da população. Esse processo é muito mais amplo do que uma simples modificação de proporções de determinada faixa etária, pois altera a vida dos indivíduos, as estruturas familiares, a demanda por políticas públicas e a distribuição de recursos na sociedade (CAMARANO, KANSO, 2016).

Mesmo sendo um acontecimento evidenciado mundialmente, o envelhecimento populacional apresenta características distintas de acordo com o nível econômico do país. Nos países desenvolvidos, este processo foi bastante lento e ocorreu concomitantemente a melhorias nas condições sociais, econômicas e de saúde da população. Nos países em desenvolvimento, estas mudanças ocorreram em curto espaço de tempo, sem o devido preparo para a nova demanda e sem melhorias de suas características econômicas e sociais (PATRÍCIO, 2013).

A situação do Brasil reflete esse panorama, pois está entre as sociedades que envelhece mais rapidamente no mundo. Nos últimos 70 anos, o número absoluto de pessoas com 60 anos ou mais aumentou doze vezes. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), em 1940 foram recenseados 1,7 milhão de idosos no Brasil e, em 2010, 20,5 milhões. A participação relativa dessa população que era de 5,1%, em 1970, passou para 8,6%, em 2000 e, segundo o último Censo em 2010, representou 10,8% da população total. Espera-se que em 2025 o Brasil tenha a sexta maior população idosa do mundo, com aproximadamente 31,8 milhões de pessoas com 60 anos ou mais, correspondentes a 15% da população esperada. Com o progressivo e contínuo envelhecimento vivenciado pela população brasileira, estima-se que em 2050 existirão 64 milhões de pessoas com 60 anos ou mais no Brasil, o que representará uma população, aproximadamente, três vezes maior que aquela recenseada em 2010 (CHAIMOVZC, 2016).

Uma característica do envelhecimento populacional é o aumento significativo da população de centenários em algumas regiões do mundo. Centenários, por definição, são sobreviventes que chegaram aos cem anos, ou seja, viveram cerca de 20 anos a mais do que a expectativa de vida média dos países desenvolvidos. Desde 1960, o número de centenários vem dobrando a cada dez anos, reflexo do crescente número de indivíduos com mais de 80 anos de idade na população mundial (MAZO, et al., 2019). Em diversos países este segmento populacional cresce rapidamente. O número estimado de centenários em países desenvolvidos

duplicou a cada década a partir de 1950 (UNITED NATIONS, 2005). Em 2013, havia cerca de 441.000 centenários em todo mundo. Em 2050, as projeções indicam que haverá cerca de 3,7 milhões, um aumento de cerca de oito vezes (UNITED NATIONS, 2015).

Os dados dos censos demográficos brasileiros evidenciam que a população de centenários tem aumentado ao longo dos anos. Segundo o IBGE, esse contingente passou de 9.689 em 1950 para 11.990 centenários em 1980. Em 1991, foram recenseados 13.865 indivíduos com 100 anos ou mais no Brasil e, em 2000, 24.576, representando um aumento de 77% em nove anos. Como esperado, em função de diferenciais de mortalidade por sexo, cerca de 60% dos centenários, em 2000, eram mulheres. No último censo, em 2010, foram recenseados um total de 24.236 centenários, destes 10.408 residiam na região Nordeste e 5.259 no Semiárido Brasileiro (IBGE, 2000; IBGE, 2010).

O envelhecimento da população traz também uma mudança nos padrões de mortalidade e morbidade, ocasionando o fenômeno conhecido como transição epidemiológica. Segundo Onram (1971), essa transição está centralizada nas complexas mudanças nos padrões de saúde e doença e nas interações desses padrões com seus determinantes demográfico, econômico e social, assumindo características diferentes em cada sociedade. Essa mudança do perfil de morbi-mortalidade faz com que o cenário caracterizado por uma população jovem, com maior incidência de doenças infecciosas, transforme-se em outro, típico de uma população mais envelhecida, onde predominam agravos crônicos e o consequente aumento nos custos assistenciais, gerados pelas alterações na utilização dos serviços de saúde, tais como: aumento das hospitalizações e necessidade de incrementação de serviços de reabilitação (MAZO et al., 2019).

Neste contexto, o Brasil está estabelecido em um processo de transição epidemiológica incompleta, na qual coexistem importantes níveis de morbi-mortalidade por doenças infecciosas e parasitárias, com elevado número de idosos com significativa prevalência de doenças crônicas. Esse fato implica no crescimento da incapacidade funcional, dependência, necessidade de cuidados de longa duração e comprometimento da qualidade de vida entre esses indivíduos. Deste modo, a transição epidemiológica remete à necessidade de se estabelecerem políticas diferenciadas para a população que irá envelhecer e a que já envelheceu, com atuação nos diversos âmbitos na área de Saúde Pública.

Conhecer o perfil epidemiológico e as causas de mortalidade da população são fundamentais para o planejamento de políticas públicas eficazes, sendo possível identificar as melhores estratégias para melhorar as condições de saúde dos idosos. Através da análise das

causas de mortalidade entre idosos pode-se compreender melhor suas condições e hábitos de vida. De acordo com a análise dos fatores demográficos, epidemiológicos e socioeconômicos é possível explorar questões relacionadas aos óbitos e inferir as razões pelos quais eles ocorreram (CHAIMOVZC, 2016). Portanto, a curiosidade de saber como um indivíduo chega a viver mais de 100 anos de idade pode ser explicada através da análise do perfil de mortalidade desses indivíduos.

A mortalidade nos idosos aumenta proporcionalmente em relação às outras faixas etárias, principalmente nos chamados longevos, aqueles acima de 80 anos, em comparação com os idosos “mais jovens”, principalmente aqueles com idade entre 60 a 69 anos. Estudo realizado por Chaimovzki (2016), evidenciou que para os idosos longevos brasileiros, incluindo os centenários, as causas mais frequentes de óbitos são as doenças do aparelho circulatório, seguidas das doenças respiratórias e das causas mal definidas. Porém, há alternância nas posições que ocupam proporcionalmente e na magnitude que atinge cada região do país, de acordo com as características pessoais e contextos socioeconômicos e ambientais a que estes indivíduos estão expostos (OLIVEIRA et al., 2019).

Portanto, no Brasil, também há diferenciais marcantes entre os estados ou regiões e, possivelmente, a mortalidade atinge de forma desigual os idosos dentro do país, o que contribui, junto à fecundidade e as migrações, para uma distribuição do envelhecimento populacional diversa entre as espacialidades. Do mesmo modo, os fatores usualmente relacionados à mortalidade diferem em exposição de acordo com os contextos e faixas etárias, mesmo entre os idosos. Tais variações ainda são percebidas nas tendências em que as causas básicas assumem e como a qualidade da informação interfere nos registros desses óbitos.

Sendo o Brasil um país de contrastes, através de um olhar social, ou pela ótica geográfica, os distintos “Brasis” se expressam em variadas formas: na desigualdade de renda, de acesso à educação, aos serviços de saúde, ao poder político, nível de instrução, problemas no registro de eventos vitais e qualidade dos dados dos censos demográficos. As diferentes regiões refletem diferentes graus de desigualdade provocados pelo modelo de desenvolvimento por elas adotado, que conduzem a sociedade brasileira a díspares condições e formas de viver (CASTRO; RODRIGUES JÚNIOR, 2012).

Ao analisar esses contrastes regionais, desponta a região semiárida brasileira. Por sua vez, tem sido marcante a deficiência histórica das estatísticas vitais, registros e qualidade dos dados do Semiárido brasileiro, sendo as mais problemáticas do país seja na cobertura, regularidade ou qualidade das informações. Daí ser flagrante a problemática que existe quanto

à qualidade dos indicadores baseados nas informações provenientes dos dados demográficos dessa região, que estão relacionados com baixos níveis de desenvolvimento social e econômico (PAES; SILVA; MACIEL, 2020). Tratando-se esses indicadores com relação à população idosa e centenária as dificuldades sobre a cobertura e fidedignidade deles se exacerbam.

### **1.1 Problematização e justificativas**

Para que seja possível produzir estimativas demográficas confiáveis é necessário que os dados possuam alta qualidade ou que, mesmo na presença de erros nos dados, possa-se detectá-los e corrigi-los. Boa parte dos problemas enfrentados pelos estudiosos da área de demografia é devotada à qualidade dos dados populacionais e registros de óbitos (PAES, 2018).

Apesar do crescimento da população centenária, a contagem dessas pessoas está sujeita a erros, até mesmo em países desenvolvidos (SPENCER, 1987; ANDRADE et al. 2018). Os tipos de erros e magnitude do seu efeito sobre os dados dos centenários variam de um censo para outro. Os principais problemas na contagem da população centenária ocorrem pelos seguintes motivos: viés oriundo da imputação das idades omitidas na entrevista, erros na declaração de idade dos entrevistados, questões de forma do questionário, captura de dados, erros de codificação e processamento (TURRA, 2012). Os erros na declaração de idade ocorrem quando o requerido não responde à pergunta com precisão. Particularmente, em relação à população de centenários, os erros podem resultar de baixos níveis de educação formal entre os mais velhos, das falhas do registro de nascimentos no passado, de dificuldades cognitivas, de respostas errôneas dadas por outra pessoa no domicílio, ou simplesmente pelo valor cultural ou econômico de se alcançar o status de centenário (PRESTON et al. 1999).

Já existem alguns estudos internacionais (CHUANWAN et. al. 2012; TERBLANCHE; WILSON, 2015) e nacionais (GOMES; TURRA, 2009) que estimam de forma indireta o número de pessoas em grupos de idades mais avançadas, incluindo o grupo de pessoas com 100 anos ou mais, com intuito de avaliar a consistência do número de centenários recenseados pelo censo demográfico. Após estimativas através de métodos indiretos distintos foi possível evidenciar que, em alguns países com histórico de problemas com a qualidade dos dados e registros, incluindo o Brasil, a população centenária está sendo superestimada.

Quando se trata de averiguar a qualidade dos registros de óbitos, dos indivíduos com 100 anos ou mais, levando em consideração não somente o problema do registro, mas também o das causas de mortes, fica evidente a escassez de estudos científicos que analisem a mortalidade desse grupo de idosos especificamente. De maneira geral, o que já se tem

conhecimento é que países em desenvolvimento, como o Brasil, os sistemas de registros de óbitos não são considerados confiáveis, pois nestes, a cobertura é incompleta, a completude das informações não alcança níveis satisfatórios e a regularidade na captação dos registros não é estável (PAES, 2018). Essas limitações impõem restrições aos usuários quando estes se propõem a medir a mortalidade com confiança, particularmente para alguns grupos etários, como é o caso dos centenários.

O estudo dos centenários fornece oportunidade ímpar de examinar como e porque alguns indivíduos vivem até idades próximas ao limite máximo de sobrevida do ser humano. Como atingir uma vida mais longa, com condições fisiológicas para manter-se vivo, rompendo os desafios enfrentados e superando os limites das estatísticas vitais, é um segredo que só algumas pessoas alcançam. Para tanto, diante das inúmeras dificuldades para se mensurar a saúde de um determinado grupo populacional, a maior disponibilidade em conseguir documentar estatisticamente as informações de saúde é pelo seu avesso, ou seja, a perspectiva do olhar científico é pelo complemento da saúde, sua ausência, quantificando e descrevendo a ocorrência de determinados agravos à saúde, que tem como consequência final a morte.

Nesse sentido, é notória a necessidade da ampliação dos conhecimentos quanto ao perfil de mortalidade dos centenários, considerando as características sociodemográficas e as principais causas de óbitos que podem estar relacionadas à sua diversidade especificamente. Como já foi ressaltado, estudos sobre a mortalidade da população com 100 anos ou mais são raros. Os que existem estão restritos a populações específicas com 80 anos ou mais, e aqueles que buscam analisar tendências temporais e avaliar a qualidade da informação são ainda mais escassos. Considerando que com o aumento da longevidade haverá uma ampliação, para diante, das curvas de mortalidade, o que tornará o estudo ainda mais dispare, principalmente, pela importância de se conhecer os indivíduos que foram a óbito nos extremos da velhice nos últimos anos, amplia-se a necessidade de compreensão dos elementos e fatores importantes que determinam a mortalidade para dar luz as gerações que ainda virão.

Nesse contexto, diante de tantas adversidades socioeconômicas e demográficas enfrentadas o Semiárido brasileiro tem experimentado o fenômeno da longevidade e o estudo de grupos etários longevos remete a necessidade de uma análise peculiar em determinados espaços geográficos e de tempo, que tenha a capacidade de identificar e compreender como se apresenta a mortalidade nesse estrato limítrofe na região. Portanto, construir indicadores de mortalidade que sejam mais confiáveis e fidedignos trará uma relevante importância e contribuirá sobremaneira para o estudo do envelhecimento nas regiões menos desenvolvidas do

país, propiciando bases e informações que sejam utilizadas para a formulação de políticas públicas que realmente impactem nesse estrato populacional tão desejado e almejado, que é o status de centenários. Nessa perspectiva, foram traçados os seguintes objetivos.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

Avaliar a qualidade e ajustar os dados populacionais e dos registros de óbitos dos centenários do Semiárido brasileiro.

### **1.2.2 Objetivos Específicos**

- Propor uma metodologia de avaliação da qualidade dos dados censitários.
- Qualificar os dados populacionais no período de 1999 a 2020.
- Qualificar os óbitos segundo características presentes nos certificados de registro e as principais causas dos óbitos no período de 2000 a 2020.
- Analisar os fatores sociodemográficos associados aos óbitos por causas mal definidas no período de 2018 a 2020.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Envelhecimento Populacional**

O envelhecimento populacional é uma realidade mundial, resultante de um processo de reestruturação demográfica no qual o Brasil também se encontra inserido. Decorre de mudanças expressivas ao longo do século XX, ditadas, fundamentalmente, pelo comportamento das taxas de fecundidade e de mortalidade, mas é o resultado, também, de melhorias na vida dos indivíduos, com maior acesso aos serviços de saúde, cobertura previdenciária, medicamentos e alimentação, bem como de melhores condições sanitárias (CLOSS; SCHWANKE, 2017).

Antes de iniciar a discussão sobre o processo de envelhecimento populacional, convém ressaltar as diferenças entre envelhecimento individual e populacional. Um indivíduo envelhece à medida que sua idade aumenta. Este é um processo irreversível, natural e individual. É acompanhada por perdas progressivas de função e papéis sociais, um processo único que depende de capacidades básicas, adquiridas e do meio ambiente, que pode ser medida pela esperança de vida ao nascer (CAMARANO; KANSO, 2016),

Já o envelhecimento populacional ocorre quando aumenta a participação de idosos no total da população. É acompanhado pelo aumento da idade média da população, um processo que pode ser revertido se a fecundidade aumentar. Sob o ponto de vista demográfico, o envelhecimento populacional é o resultado da manutenção, por um período de tempo prolongado, de taxas de crescimento da população idosa superiores às da população mais jovem. Isso implica uma mudança nas proporções dos diversos grupos etários no total da população (CAMARANO; KANSO, 2016).

A proporção de pessoas com 60 anos ou mais no mundo está crescendo mais rapidamente que a de qualquer outra faixa etária (LIMA; KONRAD, 2020). Entre 1970 e 2025, espera-se um crescimento de 223% ou em torno de 694 milhões, no número de pessoas mais velhas. Em 2025, existirá um total de aproximadamente 1,2 bilhão de pessoas com 60 anos ou mais. Até 2050 haverá dois bilhões, sendo 80% nos países em desenvolvimento. Esse processo de envelhecimento populacional, olhado pela perspectiva da distribuição etária, tem encontrado suporte na teoria da transição demográfica.

#### **2.1.1 Transição Demográfica**

A Transição Demográfica pode ser considerada como o pano de fundo para as principais modificações e processos relacionados à população, dentre os quais está o

Envelhecimento Populacional (OLIVEIRA, 2019). A transição demográfica é um processo de mudança da população de um estágio inicial caracterizado por alta fecundidade, alta mortalidade e preponderância de jovens para um estágio diferente, caracterizado por baixa fecundidade, baixa mortalidade e preponderância de idosos, modificando a estrutura etária das populações. As causas mais frequentemente atribuídas à queda da mortalidade são os avanços na área da medicina, saneamento, higiene e condições de vida. À queda da fecundidade, creditam-se fatores como urbanização, industrialização, inserção da mulher no mercado de trabalho, expansão da escolaridade etc.

Segundo a Organização Mundial de Saúde (2015), uma redução nas taxas de fecundidade total (TFT) e um aumento da longevidade irão assegurar o contínuo envelhecimento da população mundial. Observa-se quedas abruptas nas taxas de fecundidade em todo o mundo e estima-se que, até 2025, 120 países terão alcançado TFT abaixo do nível de reposição (média de fecundidade de 2,1 crianças por mulher). Isso representa um aumento substancial se comparado a dados de 1975, quando apenas 22 países possuíam uma TFT menor ou igual ao nível de reposição (LIMA; KONRAD, 2020).

Um fato importante é que, de modo geral, essa transição demográfica em países desenvolvidos ocorreu bem antes do envelhecimento significativo de suas populações. Esses países adquiriram uma base sólida dentro de aspectos socioeconômicos e de saúde, além de estratégias institucionais visando atender essa população e equilibrar os efeitos das desigualdades residuais. Isso pode ser explicado, tendo em vista que países na América do Norte e Europa já possuíam um grau de desenvolvimento social e econômico à medida que sua população envelhecia, ou seja, os países desenvolvidos ficaram estruturados antes de envelhecer e o Brasil está envelhecendo antes de se estruturar economicamente (CORTEZ et. al., 2019).

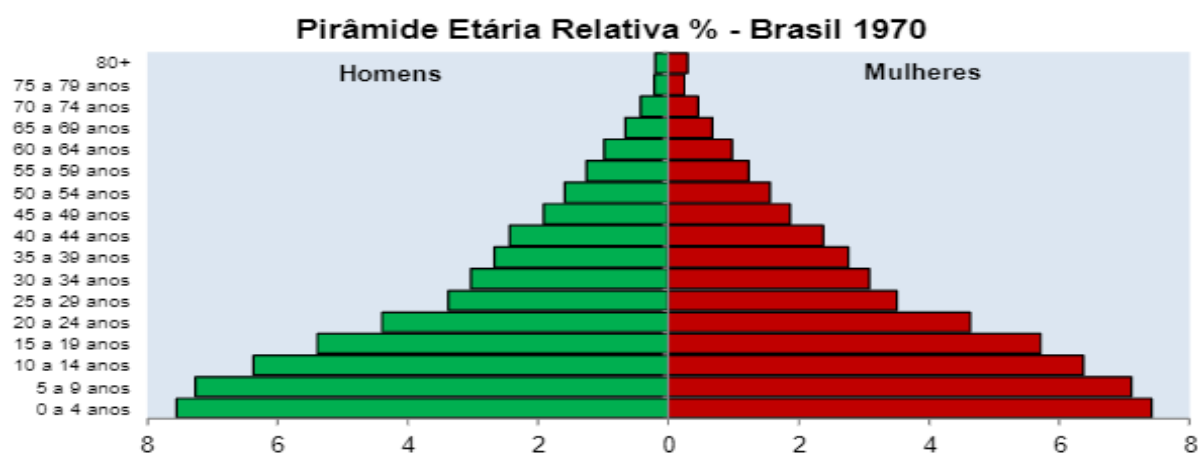
Os países da América Latina estão envelhecendo muito rapidamente, mas nem sempre acompanhados de um desenvolvimento socioeconômico. A velocidade do processo traz novos desafios que se instalam de maneira muito rápida e acabam impactando na estrutura da sociedade e de como a mesma se comporta (BRASIL, 2010).

Em 1970, o Censo Demográfico já indicava o início da transição demográfica – o grupo etário de 0 a 4 anos baixou de 16% da população, observados entre 1940 e 1960, para 14,9%, em 1970. Desde então, o envelhecimento da população brasileira tem sido contínuo e progressivo. Em cinco décadas, o número de brasileiros com 60 anos ou mais aumentou mais de seis vezes - em 1950, eram 2,3 milhões, ou 3,5% da população; em 2000, eram mais de 14,5

milhões, ou 8,5% dos brasileiros estavam nessa faixa etária. Na última década o salto populacional foi grande, e em 2010 a representação passou para 10,8% da população, foram recenseados 20,5 milhões de idosos (IBGE, 2010)

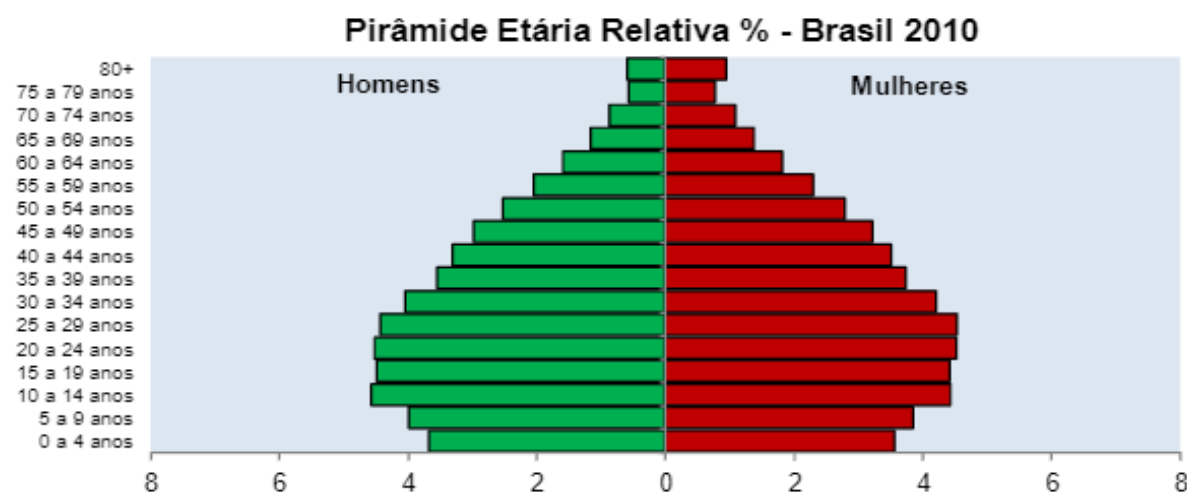
A comparação das pirâmides etárias permite a observação das principais mudanças na distribuição dos grupos etários no país. A Figura 2.1 mostra a distribuição dos grupos etários por sexo no país em 1970 e a Figura 2.2 a mesma distribuição no país, porém no ano de 2010.

Figura 2.1- Pirâmide etária do Brasil em 1970.



Fonte: Oliveira (2019).

Figura 1.2 - Pirâmide etária do Brasil em 2010.



Fonte: Oliveira (2019)

Em 1970, a pirâmide etária do país refletia o comportamento demográfico da época em que predominavam altos níveis de fecundidade. A pirâmide se configurava com a forma típica que é encontrada em países pobres, os quais apresentam baixos níveis de desenvolvimento econômico e social. A situação socioeconômica do país refletia no comportamento demográfico

da população. Ou seja, as famílias eram bem numerosas, resultado do elevado número de nascimentos. Desse modo, a pirâmide etária apresentava uma base bem larga, com uma grande participação de crianças e jovens na população total. Nessa configuração a base e a parte mediana são alargadas e o ápice é bem estreito, devido ao reduzido número de idosos. A fecundidade alta impactava no processo de envelhecimento, que era pouco significativo, sendo a expectativa de vida não muito elevada.

Em 2010 as transformações na distribuição etária da população do Brasil seguiram uma trajetória de redução das barras que representam as crianças e alargamento das barras referentes aos jovens, adultos e idosos. A Figura 2.2 apresenta diferenças em relação à Figura 2.1 e o formato da pirâmide etária do país mostra que está em um processo de transição. A pirâmide do Brasil que em 1970, possuía o formato tradicional, típico de nações mais pobres, vem ganhando uma nova forma, a qual se aproxima dos países mais desenvolvidos.

Houve uma redução na base, reflexo da diminuição do número de nascimentos e o ingresso do menor número de crianças na população, ocasionado pela redução da fecundidade. Houve alargamento do meio da pirâmide decorrente da elevação do grupo dos adultos e aumento do ápice devido à maior participação de idosos. Em 2010, a pirâmide etária seguiu refletindo mudanças do comportamento demográfico do país com a melhoria das condições sociais e econômicas, além de mudanças culturais que impactaram de modo significativo na distribuição dos grupos de idade.

A pesquisa nacional por amostra de domicílios contínua (PNAD) realizada em 2017 e divulgada pelo IBGE, descreve que a população brasileira manteve a tendência de envelhecimento dos últimos anos e ganhou 4,8 milhões de idosos desde 2012, superando a marca dos 30,2 milhões em 2017. Em 2012, a população com 60 anos ou mais era de 25,4 milhões. Os 4,8 milhões de novos idosos em cinco anos correspondem a um crescimento de 18% desse grupo etário, que tem se tornado cada ano mais representativo no Brasil (IBGE, 2017).

É necessário evidenciar que grande parte desses idosos são mulheres, ou seja, há uma feminização da velhice. Essa tendência é consequência da sobremortalidade masculina em todas as faixas etárias e para a maioria das causas de morte. Possivelmente, isso se deve à maior exposição dos homens a fatores de risco como álcool e fumo, aos acidentes de trabalho e homicídios e ao aumento da mortalidade por câncer de pulmão e doenças cardiovasculares (CHAIMOVICZ, 2016). Além disso, o século XX presenciou um importante declínio na mortalidade materna perinatal e por câncer do útero; fatores hormonais são ainda determinantes

de uma proteção cardiovascular que se estende até alguns anos após a menopausa. A aparente vantagem das mulheres é parcialmente atenuada pela maior prevalência de demências, depressão e dependência funcional neste grupo, reduzindo seu tempo de vida livre de incapacidades (CLOSS; SCHWANKE, 2017).

Confirmada em números, esta feminização do envelhecimento pode ser evidenciada por meio da razão de sexos, ou “número de homens para cada 100 mulheres” que constitui um índice demográfico que retrata a proporção dos sexos em uma população. No Brasil, em 1950 existiam aproximadamente 100 homens para cada 100 mulheres na faixa etária de 65 a 69 anos, mas em 2000 eram 81 e em 2025 deverão ser 78 homens para cada 100 mulheres (IBGE, 2013).

### **2.1.2 Transição Epidemiológica**

Aliando-se à transição demográfica, as alterações no perfil de morbi-mortalidade das populações, influenciadas pelo envelhecimento dos indivíduos, ocasiona o fenômeno conhecido como transição epidemiológica (TE). Segundo um dos primeiros formuladores dessa teoria, Onram (1971), essa transição está centralizada nas complexas mudanças nos padrões de saúde e doença e nas interações desses padrões com seus determinantes demográfico, econômico e social; assumindo características diferentes em cada sociedade.

A Teoria da transição epidemiológica se caracteriza como um processo fundamentado em algumas proposições: 1) tem início com a premissa principal de que a mortalidade é um fator fundamental na dinâmica populacional. Diante disso, tem-se como exemplo os países em desenvolvimento, onde a taxa de mortalidade começou a declinar rapidamente a partir de 1945 e a taxa de natalidade sofreu pequenas alterações. Dessa maneira, esse hiato demográfico provocou um grande crescimento populacional; 2) durante a transição, ocorre uma mudança nos padrões de mortalidade e da doença, na qual as epidemias e pandemias existentes são substituídas de forma gradual pelo predomínio de doenças crônicas não transmissíveis e provocadas pelo homem, de modo que elas se tornam as principais formas de morbidades e de causas de óbitos; 3) durante a transição epidemiológica, as mudanças mais profundas nos padrões de saúde e doença ocorrem entre crianças e mulheres jovens; 4) as mudanças nos processos de saúde e de doença fazem parte do complexo da modernização e estão altamente relacionadas às mudanças demográficas e socioeconômicas vivenciadas pelas sociedades (OMRAN, 1971).

Os padrões de mortalidade ocorrem em três estágios: a) a "era da fome e das pestilências", representadas pela incidência de grandes epidemias e guerras, nas quais as taxas

de mortalidade por doenças transmissíveis eram altas e altas taxas de natalidade eram observadas; b) a "era do declínio das pandemias", na qual as pandemias diminuíram mas as doenças transmissíveis ainda respondiam pela maior parte das causas de óbitos e; c) a "era das doenças degenerativas e causadas pelo homem", onde houve melhora nas condições de vida da população, que passou a desfrutar de melhores condições de habitação, saneamento, alimentação e educação e, além disso, houve significativa queda na ocorrência de doenças transmissíveis. Dessa maneira, as doenças cardiovasculares e neoplasias malignas passaram a ocupar o topo das principais causas de óbito (OMRAN, 1971).

No entanto, a transição epidemiológica ocorreu e ocorre de maneira distinta entre alguns países e, por isso, três modelos diferentes de transição foram inicialmente propostos: a) o Modelo Clássico ou Ocidental, o qual consiste em uma transição gradual e progressiva de altos níveis de mortalidade e de natalidade para níveis reduzidos de ambos os indicadores e é representado pelos países da Europa e EUA; b) o Modelo Acelerado, no qual ocorre uma queda acelerada nas taxas de mortalidade a partir de 1920, como é o caso do Japão. A grande distinção desse modelo para o modelo clássico é a queda na taxa de mortalidade ocorrer de forma muito mais rápida e; c) o Modelo Atrasado ou Contemporâneo, que descreve a transição epidemiológica mais recente e que ainda precisa ser concluída nos países em desenvolvimento (OMRAN, 1971).

Logo, a teoria da transição epidemiológica formulada por Omran (1971) identifica a mortalidade como uma força fundamental na mudança do tamanho e estrutura da população, particularmente na fase inicial da transição, com a fecundidade sendo um fator crucial que promoverá mudanças futuras na população, quando a mortalidade estiver caminhando para a estabilização em níveis baixos (OMRAN, 1983).

Apesar das quedas ocorridas nos indicadores de natalidade, fecundidade e mortalidade no Brasil, o país apresentava significativas diferenças acerca do processo de Transição Epidemiológica, uma vez que, apesar de apresentar um aumento da expectativa de vida ao nascer, juntamente com o aumento da proporção de óbitos atribuídos às características do envelhecimento, é ainda observado um número significativo de óbitos atribuídos às doenças transmissíveis e causas externas (VASCONCELOS; GOMES, 2012). Em grande parte, isso pode ser atribuído ao fato do Brasil pertencer à América Latina, uma região bastante heterogênea que apresenta desigualdades, ocasionadas pelas distintas realidades econômicas, sociais e culturais. Sendo assim, apesar do país se apresentar em processo mais avançado de transição demográfica, as taxas de mortalidade por doenças transmissíveis ainda são

relativamente altas.

O conceito de transição epidemiológica tem merecido críticas, uma vez que, a transformação dos padrões de saúde não obedece aos mesmos parâmetros de sequência, intensidade e velocidade, em diferentes regiões (SCHRAMM et al., 2004). Segundo esses autores, no Brasil a diferenciação regional coloca barreiras ao avanço do processo de transição epidemiológica fazendo com que as regiões Sul e Sudeste encontrem-se em patamares avançados, a região centro-oeste em situação intermediária e as regiões Norte e Nordeste atrasadas em relação às outras. Destacam que há uma sobreposição entre as etapas nas quais predominam as doenças transmissíveis e crônico-degenerativas, assim como a reintrodução de doenças como dengue e cólera ou o recrudescimento de outras como malária, hanseníase e leishmaniose (SCHRAMM et al., 2004).

Desse modo, a emergência para lidar com as doenças transmissíveis, como: AIDS, zika, dengue, cólera, malária, esquistossomose, hanseníase e leishmaniose tornam o cenário da Transição Epidemiológica ainda mais complexo no Brasil, colocando o país em uma situação muito distinta da vivenciada pelos países desenvolvidos, onde as doenças crônicas não transmissíveis assumiram papel principal. Portanto, esse cenário exige maior atenção das políticas públicas a serem adotadas (BORGES, 2017), o país precisa enfrentar o aumento da morbidade e da mortalidade atribuídas às doenças crônicas não transmissíveis, mas também, precisa lidar com o agravamento e a permanência de doenças transmissíveis (ARAÚJO, 2012).

Além dessa sobreposição nas etapas do processo de transição epidemiológica, é necessário observar o incremento das causas externas de morte que destoam das causas de morte esperadas em estágios avançados da transição epidemiológica. Os estudos de carga de doenças demonstram que o Brasil vive uma situação de tripla carga de doenças em que se manifestam, coetaneamente, uma agenda não concluída de infecções agudas e parasitárias, o acometimento de causas externas e o enfrentamento das doenças crônicas (MENDES, 2018).

O processo de transição epidemiológica no Brasil sofreu transformações que ocorrem no modelo epidemiológico do país, com o avanço das causas externas de mortalidade. Essas mudanças são, segundo os autores, decorrentes:

Não apenas da reemergência ou presença constante dos casos de doenças infecciosas e parasitárias – se não com casos elevados de mortalidade, mas ainda com casos relevantes de morbidade, detectada pelo sistema de vigilância epidemiológica e pelos registros ambulatoriais e hospitalares – como também pelo importante crescimento das causas externas, expressão da violência social em suas diferentes formas (PEREIRA et al., 2015, p. 101-102).

Em resumo pode-se observar que o processo de TE é um fenômeno de difícil análise, sobretudo quando o objeto de estudo são países em desenvolvimento, como o caso do Brasil. Aspectos como velocidade do processo, polarização epidemiológica, possibilidade de reversão e sobreposição de etapas são questões que adicionam maior especificidade a transição, especialmente em análises intrarregionais.

## **2.2 Estudo dos Centenários: o mundo e o regionalismo brasileiro**

Segundo dados demográficos, as pessoas com mais de 80 anos correspondem ao estrato populacional que mais cresce nos países desenvolvidos. Um grupo especial dos mais idosos, e que merece destaque, dentro do estrato dos longevos, são os centenários - pessoas que atingiram o limite extremo do tempo de vida humano (PATRÍCIO, 2013; TEIXEIRA et. al., 2017). Apesar de, atualmente, ainda representar uma pequena proporção da população mundial, o relatório da United Nations World Population Aging prevê que o número de centenários será de 20,1 milhões em 2100, ou seja, essa faixa etária limítrofe tem perspectivas consideradas de crescimento ao longo dos anos (BATISTA, 2019).

definem os centenários são “sobreviventes que chegaram aos cem anos, ou seja, viveram cerca de 20 anos a mais do que a expectativa de vida média dos países desenvolvidos” (KUMON et al., 2009). A partir dessa definição, entende-se que, devido a determinados fatores biopsicossociais, alguns indivíduos conseguem viver muito mais tempo do que o restante da população. Portanto, o estudo dos centenários fornece oportunidade ímpar para examinar porquê e como alguns indivíduos vivem até idades próximas ao limite máximo de sobrevida do ser humano (ROWE; KAHN, 1987).

O conceito de uma vida longa, fundamentado no termo longevidade, traz em si a compreensão de indivíduos que anseiam por melhores condições de vida em sua totalidade, permitindo viver de maneira ativa e consciente, como compromisso da integralidade no cotidiano da vida (ARAÚJO, 2018). Como atingir uma vida mais longa e saudável, estendendo a qualidade de vida até o final da existência humana, é um segredo que só algumas pessoas alcançam. Sabe-se, no entanto, que a longevidade humana pode ser considerada uma característica multifatorial, altamente dependente de componentes genéticos e ambientais (PEDRO; BÓS, 2016).

Atualmente várias são as teorias que tentam explicar o porquê de alguns indivíduos atingirem essas idades e outros não, no entanto em todas elas existe um fator comum atribuído à interação do indivíduo com o meio que o rodeia, relacionado com hábitos e estilos de vida

que podem ser catalogados como mais ou menos saudáveis (CAMPS CALZADILLA et. al., 2013).

Embora, a longevidade seja determinada por fatores genéticos, o tempo de vida também é afetado por fatores modificáveis, tais como ambientais, psicológicos, socioculturais, estilos de vida, atividade física, fatores alimentares e médicos (PIETRUSZKA et al., 2006; KOLLAJTIS-DOLOWY et al., 2007; KWAK et al., 2012). Porém, segundo Kumon et. al. (2009) não existe um perfil único desses indivíduos longevos ou uma “receita única para alcançar a longevidade”, já que cada centenário é singular e possui vivências únicas que permitiram alcançar a atual idade.

A heterogeneidade existente entre os centenários tem dificultado as pesquisas científicas que procuram explicar, de maneira única, a complexidade que envolve a longevidade humana (AFONSO et. al., 2018). Heterogeneidade essa vivenciada no estado de saúde, seja físico ou cognitivo, provavelmente resultante de coortes diferentes no espaço (geografia) e tempo (nascimento), e da consequente adaptação a diferentes ambientes socioeconômicos e culturais (FRANCESCHI et. al., 2017). Por um lado, existem centenários frágeis, com múltiplas patologias, déficit cognitivo e/ou dependência funcional. Por outro lado, existem indivíduos autônomos, sem condições particulares de doença ou distúrbios cognitivos (AFONSO et. al., 2018; VACANTE et al., 2012). Estes últimos têm sido estudados como fenômenos de “*successful aging*” (PIETRUSZKA et. al., 2006), contrariando o mito “*the older you get, the sicker you get*”, vivendo a maior parte das suas vidas em excelente estado de saúde, podendo surgir algumas doenças apenas nos últimos anos de vida (PERLS et. al., 2002). Ou seja, envelhecer de forma relativamente lenta e atrasar e/ou escapar das doenças associadas à idade parece ser a solução e o desafio para atingir idades iguais ou superiores a 100 anos (PERLS et. al., 2002).

Cada vez mais pessoas têm alcançado o status centenário. Segundo informações do Programa das Nações Unidas para o desenvolvimento (PNUD, 2015), em 2015, existiam por volta de 434 mil pessoas com 100 anos ou mais de idade no mundo e supõe-se que esse número de pessoas aumente consideravelmente nas próximas décadas, a projeção é que em 2050 existam 3,8 milhões de centenários em todo o mundo.

Os dados disponíveis revelam que os EUA possuem a maior população de centenários do mundo, seguido pelo Japão, China, Índia e Itália. China e Índia estão na lista por apresentarem enormes estoques populacionais (REIS; TURRA, 2016). De fato, os centenários ainda representam uma parcela muito pequena da população total dos países relativamente

jovens, como China e Índia. Por outro lado, os centenários compõem uma fração maior de populações onde o processo de envelhecimento está mais adiantado, como são os casos de Japão e Itália. Em 2015, existiam 4,8 centenários para cada 10.000 pessoas no Japão e 4,1 na Itália. Em comparação, nos EUA, que está envelhecendo a um ritmo mais lento, em 2015, existiam 2,2 centenários para cada 10.000 pessoas (PNUD, 2015).

Em 2008, o *National Institute on Aging*, identificaram comunidades em todo o mundo onde as pessoas estão a viver mais anos e mensuravelmente melhor. Essas zonas foram designadas de *Blue Zones*. Uma dessas comunidades é Okinawa, uma das prefeituras do Japão, país conhecido por apresentar a maior longevidade do mundo (WILLCOX et. al., 2009; MAHAN et. al., 2012). Em 2010, contava com 34 centenários por 100 000 indivíduos (KWAK et al., 2012). Igualmente, a península de Nicoya na Costa Rica, a ilha Ikaria na Grécia e a Sardenha na Itália, foram consideradas *Blue Zones* (JOPP et al., 2016; MAHAN et al., 2012).

Os centenários no Brasil compõem uma parcela muito pequena da população total, mas que tem aumentado ao longo dos anos. Segundo o IBGE, em 1950, havia apenas 9.689 centenários no Brasil, ou 1,86 centenários por 10.000 pessoas e, em 1980, esse número aumentou para 11.390, ou 0,95 centenário por 10.000 pessoas. Em 1991, foram recenseados 13.865 indivíduos com 100 anos ou mais, o que correspondia a 0,94 centenários por 10.000 pessoas. Em 2000, comparado ao censo anterior, o número de pessoas com mais de cem anos quase duplicou, havia 24.576 centenários no país, ou 1,44 centenários por 10.000 pessoas, sendo 14.153 mulheres e 10.423 homens (IBGE, 2000).

No último censo brasileiro, realizado em 2010, foram contabilizados mais de 24.236 centenários no Brasil (77% do sexo feminino), distribuídos principalmente nos Estados de São Paulo (4.457), Bahia (2.808) e Minas Gerais (2.765). Houve uma redução no número total de centenários recenseados em relação ao ano de 2000, consequentemente, a proporção de centenários na população total também diminuiu para 1,27 centenários por 10.000 pessoas (IBGE, 2010). A Tabela 1 apresenta o panorama da distribuição dos centenários recenseados nas cinco regiões brasileiras nos anos de 2000 e 2010.

Tabela 1 - Distribuição de centenários recenseados no Brasil e nas cinco grandes regiões, segundo o sexo, 2000 e 2010.

| Área                | Ano x Sexo |        |        |        |       |        |
|---------------------|------------|--------|--------|--------|-------|--------|
|                     | 2000       |        |        | 2010   |       |        |
|                     | Total      | Homem  | Mulher | Total  | Homem | Mulher |
| <b>Brasil</b>       | 24.576     | 10.423 | 14.153 | 24.236 | 7.247 | 16.989 |
| <b>Norte</b>        | 1.877      | 827    | 1.050  | 1.887  | 640   | 1.247  |
| <b>Nordeste</b>     | 8.038      | 3.385  | 4.653  | 10.408 | 3.200 | 7.208  |
| <b>Sudeste</b>      | 9.702      | 4.083  | 5.619  | 8.128  | 2.168 | 5.960  |
| <b>Sul</b>          | 3.223      | 1.338  | 1.885  | 2.377  | 687   | 1.690  |
| <b>Centro-Oeste</b> | 1.736      | 790    | 946    | 1.436  | 552   | 884    |

Fonte: IBGE, 2000 e 2010.

Observando a Tabela 1, é possível identificar diferenças na distribuição dos centenários no Brasil ao longo de 10 anos. Houve diminuição do número total de centenários, dentre esses os homens, porém no período em questão o número de mulheres centenárias aumentou. No tocante as regiões brasileiras, em 2000 a região com maior número de centenários era a região Sudeste, contudo em 2010 o Nordeste foi a região com o maior quantitativo desse grupo, concentrando quase metade dos centenários recenseados no Brasil no referido ano. A maioria das regiões brasileiras diminuíram a quantidade de centenários recenseados no decorrer do período, exceto Norte e o Nordeste. Outra informação peculiar da região nordestina é que quase duplicou o número de mulheres centenárias ao longo da década.

Como já foi mencionado, o estudo dos centenários e a heterogeneidade desse grupo etário específico remete a necessidade de uma análise peculiar em determinados espaços geográficos e de tempo. Caso esses territórios de análise sejam marcados por baixos níveis socioeconômicos, desigualdades históricas, dificuldade de acesso a água e alimentos, torna-se ainda mais interessante compreender como existem centenários nessas regiões pobres, indivíduos que rompem as barreiras e dificuldades locais, conseguindo viver mais 25 anos além da expectativa de vida brasileira.

Um dos espaços geográficos que enfrenta maiores desafios e dificuldades econômicas no Brasil é a região semiárida (ARAÚJO, 2013). Atualmente, o semiárido brasileiro engloba 1.262 municípios nos nove estados do Nordeste (Alagoas, Bahia, Ceará, Maranhão, Pernambuco, Paraíba, Piauí, Rio Grande do Norte e Sergipe), além de Minas Gerais. A área é marcada pelo conflito entre meio ambiente e sociedade, que se sustenta na implacabilidade das secas e no aumento da desertificação como produtores das desigualdades, onde a população residente tem que lutar contra a seca e conviver constantemente com ela (ASA, 2020). Políticas

públicas brasileiras como Programa Fome Zero, Programa de Cisternas, Plano Nacional Combate à Fome e de Promoção da Segurança Alimentar e Nutricional no Brasil, a partir de 2003, que tinham como áreas prioritárias o semiárido (Diniz & Piraux, 2011). Dados do IBGE (2010) compilados pelo INSA (2017) mostram o semiárido como uma área de baixo dinamismo econômico, sendo que em torno de 61% dos municípios possuem IDH baixo e muito baixo.

Contudo, mesmo diante de tantas adversidades enfrentadas, o semiárido brasileiro tem experimentado o fenômeno da longevidade, segundo dados do IBGE o número de centenários vem aumentando no semiárido brasileiro ao longo dos anos. Em 2000, foram recenseados 4367 centenários, sendo 1.831 homens e 2.536 mulheres, em 2010 a população residente com 100 anos ou mais totalizava 5.259, destes 1.769 eram homens e 3.490 eram mulheres. Portanto, devido a heterogeneidade do processo de envelhecimento, a demanda exige estudos científicos para compreender de maneira qualificada onde e como vivem esses centenários, assim como qual o perfil de morbimortalidade deles. Porém essas questões só serão possíveis de serem respondidas com a análise minuciosa da qualidade dos dados e registros tanto da população quando dos óbitos da região, e das fontes primordiais que geram essas informações.

### **2.3 Fontes de Dados Demográficos**

A demografia é uma ciência primordialmente empírica e dedica cuidadosa atenção a questões sobre medição, disponibilidade e qualidade dos dados, a fim de estudar processos no nível populacional agregado (PRESTON, 1993). Segundo Borges e Nepomuceno (2020), uma das principais características dos estudos demográficos, desde os primórdios da disciplina, é o seu foco em dados, com grande atenção dedicada à sua qualidade. Boa parte do trabalho do demógrafo é destinada a testes que avaliam a qualidade dos dados na tentativa de aprimorá-los e melhor interpretá-los.

As fontes de dados contemporâneas mais relevantes são os censos demográficos e os registros civis. Outras fontes administrativas incluem os controles dos movimentos migratórios, as campanhas de saúde pública, os dados das representações ministeriais e as pesquisas especiais e etc. (PAES, 2018).

Apesar das desvantagens de alto custo, divulgação demorada e frequência reduzida, e a despeito de existirem hoje várias alternativas de coleta de informação, o censo demográfico ainda é o principal instrumento para obter dados sobre a população, principalmente nos países em desenvolvimento, onde existem relativamente poucas alternativas. Segundo a definição das Nações Unidas (1980), um censo é “o processo total de coleta, processamento, avaliação,

análise e divulgação de dados demográficos, econômicos e sociais referentes a todas as pessoas dentro de um país ou de uma parte bem definida de um país num momento específico".

Os censos nacionais são realizados praticamente em todos os países do mundo e produzem o registro das pessoas em um determinado momento, ou seja, produz uma informação “fotográfica” transversal da população existente, o seu tamanho, sua estrutura, e como ela é observada no momento do censo (PAES, 2018).

No Brasil, o órgão responsável pela realização dos censos é o Instituto Brasileiro de Geografia e Estatística (IBGE). Os recenseamentos brasileiros foram realizados em 1872, 1920, e anos decenais a partir de 1940 (exceto o de 1990 que somente foi realizado em 1991). O censo mais recente foi realizado em 2010, na sequência dos anos decenais deveria ter acontecido em 2020, porém por diversos motivos, incluindo cortes de verbas e pandemia da COVID-19, o censo brasileiro foi realizado em 2022.

Os erros e vieses no recenseamento estão presentes em todas as etapas e são praticamente inevitáveis, pois envolvem uma complexidade logística e técnica. Segundo Paes, (2018), os principais erros dos censos são:

- a) Cobertura incompleta da população – a contagem de pessoas mais de uma vez ou a não contagem de alguns segmentos populacionais;
- b) Erros de declaração – aparecem quando o respondente, por algum motivo deliberado ou não, comete erros ao responder perguntas feitas.

Enquanto o censo consiste no levantamento de dados sobre todos os indivíduos de uma população em momentos preestabelecidos, o registro civil visa acompanhar as ocorrências de eventos vitais (nascimentos e óbitos), os quais modificam o tamanho ou a composição da população ao longo do tempo. A unidade de enumeração do registro civil, portanto, é o evento demográfico, enquanto a unidade de enumeração do censo é o indivíduo. Além das suas finalidades estatísticas, o registro civil cumpre uma função legal, uma vez que os eventos registrados modificam a situação das pessoas perante leis que variam de país para país e têm uma maior especificidade e durabilidade do que a legislação referente aos censos (HAKKERT, 1996).

As estatísticas vitais são efetivas em somente poucos países (BORGES, 2017). Nos países em desenvolvimento, esses sistemas de coleta de informações são, na maioria das vezes, incompletos, e por vezes não existem, ou não são explorados com fins demográficos. Mesmo em países desenvolvidos, existem muitos problemas com a qualidade dos dados coletados (PAES, 2018).

Ainda segundo o autor supracitado, os métodos para testar os dados incompletos e incorretos dos registros vitais são inúmeros, sendo os mais importantes divididos em quatro categorias: cobertura, completude, regularidade e fidedignidade.

A cobertura (ou notificação) significa magnitude ou nível, medido geralmente em percentual, com que os registros vitais são efetivamente enumerados nos sistemas. A completude representa a magnitude ou nível de declaração de uma determinada variável informada quando o registro foi realizado no sistema. Os dois correspondem aos indicadores mais preocupantes de qualidade dos registros vitais, particularmente a cobertura, já que a contabilidade estatística depende do número de óbitos para ser implementado em diversos âmbitos, inclusive na implementação de políticas públicas, emergências epidemiológicas e repasse financeiro.

A regularidade diz respeito com que frequência um sistema de registro mantém a captação dos dados em um determinado nível durante o tempo.

A fidedignidade significa o grau de confiança ou de informações corretas declaradas nas variáveis observadas nos registros vitais.

O Brasil possui duas fontes oficiais responsáveis pela produção contínua de estatísticas vitais. A primeira, fica sob responsabilidade da Fundação IBGE, cujos resultados são publicados em “Estatísticas do Registro Civil” (IBGE, 2018). A segunda, fica sob responsabilidade do Ministério da Saúde – MS, denominados Sistema de Informações sobre Mortalidade – SIM e o Sistema de Informações sobre Nascidos Vivos – SINASC.

## **2.4 Os Sistemas Nacionais de Informações das Estatísticas Vitais**

O Brasil tem uma vasta tradição na produção de Estatísticas Vitais, haja vista o Cartório de Registro Civil de Pessoas Naturais ser uma instituição secular. Todavia, os diferentes estágios de desenvolvimento econômico e social e as assimetrias regionais, associados às largas distâncias de um país continental, fizeram com que ao longo de longo período a cobertura e a qualidade de dados sobre os eventos de nascimentos e óbitos ficassem muito aquém das desejadas. Essas deficiências, somadas às necessidades de dados epidemiológicos, por parte do Ministério da Saúde, fizeram com que fossem criados o Sistema de Informações sobre Nascidos Vivos (SINASC) e o Sistema de Informações sobre Mortalidade (SIM) em paralelo às Estatísticas do Registro Civil, produzidas pela Fundação Instituto Brasileiro de Geografia e Estatística - IBGE (Fundação IBGE, 2018), ambas as fontes possuem disponibilidade anual para todos os municípios brasileiros.

### **2.4.1 Sistemas de Informações do IBGE**

As grandes fontes de dados históricos sobre condições de vida, saúde e morbidade referidas dos brasileiros são as Estatísticas de Registro Civil, os Censos Demográficos decenais, as Pesquisas Nacionais por Amostra de Domicílios (PNAD) e os respectivos suplementos especiais de temas relacionados à saúde. Mais recentemente, esses suplementos desdobraram-se em uma nova pesquisa especial, a Pesquisa Nacional de Saúde (PNS) (IBGE, 2015).

O IBGE, a partir da aprovação da Lei n. 6.015, de 31/12/1973, passou a ser o responsável, inicialmente, pela centralização do recebimento dos formulários e, posteriormente, por toda a sistematização e produção de Estatísticas do Registro Civil. Dessa forma, o IBGE passou a coletar, em todo território nacional, os dados sobre os eventos vitais registrados nos cartórios (IBGE, 2018).

As Estatísticas de Registro Civil foram sistematizadas a partir de 1974, incluindo os nascidos vivos, os óbitos, incluindo os fetais e os casamentos. A partir de 1984, passou a considerar também as separações judiciais e os divórcios. Quando se analisa a primeira publicação destes registros é curioso notar que desde sua origem, os mesmos já ocorriam e eram divulgados por município. A pesquisa, publicada anualmente, permite análises locais e regionais sobre fecundidade, nupcialidade e mortalidade e ainda viabiliza estudos para o aprimoramento de programas governamentais na área da educação, previdenciária, econômica e social (PINTO, FREITAS e FIGUEIREDO, 2018).

Os três últimos Censos Demográficos Populacionais foram realizados em 1991, 2000 e 2010. A menor unidade de análise para fins de coleta de dados utilizada no trabalho de campo é o setor censitário, que corresponde a menor unidade territorial, criada para fins de controle cadastral da coleta de dados. Quando se realizam pesquisas na área da atenção primária à saúde, muitas vezes ainda é possível combinar setores censitários com as chamadas “microáreas” das Equipes de Saúde da Família, que correspondem a um conjunto de logradouros, quarteirões, vielas de moradia da população cadastrada pelas equipes (IBGE, 2015).

Para atender à necessidade de estudos nos períodos intercensitários surgiu de forma pioneira a Pesquisa Nacional por Amostra de Domicílios (PNAD) a partir de 1967, hoje, denominada PNAD Contínua. Ao longo dos últimos 50 anos, as PNAD pesquisaram dezenas de áreas temáticas em pesquisas complementares ao questionário básico, abordando desde aspectos demográficos, trabalho e rendimento, trabalho infantil, aspectos sociais e a saúde propriamente dita. Nesse caso, as dimensões investigadas referem-se ao acesso a serviços de saúde, características de saúde dos moradores, serviços preventivos de saúde da mulher,

mobilidade física, fatores de risco e proteção à saúde, tabagismo, prática de esportes e atividade física e no campo da Política Nacional de Atenção Básica e características da Estratégia Saúde da Família (IBGE, 2015).

#### **2.4.2 Sistemas de Informações do Ministério da Saúde**

Os Sistemas de Informações a respeito dos eventos vitais no âmbito do Ministério da Saúde – SIM e SINASC – foram desenvolvidos com a finalidade de proporcionar dados para o acompanhamento e análise epidemiológica, orientar a gestão da saúde e proporcionar a construção de políticas nessa área. Isto sinaliza objetivos distintos daqueles das Estatísticas do Registro Civil, como apresentado anteriormente, o que justifica a necessária coexistência dos Sistemas (IBGE, 2018).

A gestão do Sistema Único de Saúde (SUS) é um processo pelo qual equipes gestoras (sujeitos) tomam decisões na implementação da Política de Saúde (TAMAKI et al., 2012). Os bancos de dados e Sistemas de Informação em Saúde (SIS) constituem ferramentas importantes para o planejamento e a avaliação das políticas de saúde, assim como dos serviços, redes e sistemas de saúde (FERLA; CECIM; ALBA, 2012).

No Brasil, diversos SIS federais foram instituídos antes mesmo da implantação do SUS e os seus objetivos estavam voltados para os governos federal e/ou estadual. Os municípios assumiam o papel de coletores de dados e, frequentemente, ocorria a subutilização das informações. Assim, eles permaneciam alijados do processo de elaboração dos planejamentos, dispondo de pouca ou nenhuma experiência/autonomia para a formulação de políticas e para a tomada de decisão (VIDOR, FISHER e BORDIN, 2011).

Os Sistemas de Informação em Saúde podem ser definidos como um conjunto de componentes inter-relacionados que coletam, processam, armazenam e distribuem a informação para apoiar o processo de tomada de decisão e auxiliar no controle das organizações de saúde. Assim, os sistemas de informação em saúde congregam um conjunto de dados, informações e conhecimento utilizados na área de saúde para sustentar o planejamento, o aperfeiçoamento e o processo decisório dos múltiplos profissionais da área da saúde envolvidos no atendimento aos pacientes e usuários do sistema de saúde (MARIN, 2010).

Os SIS são influenciados pela organização do SUS, e como tal, integram suas estruturas organizacionais e contribuem para sua missão. São constituídos por vários subsistemas e tem como objetivo geral facilitar a formulação e avaliação das políticas, planos e programas de saúde, subsidiando o processo de tomada de decisões. Para tanto, deve contar com os requisitos

técnicos e profissionais necessários ao planejamento, coordenação e supervisão das atividades relativas à coleta, registro, processamento, análise, apresentação e difusão de dados e geração de informações (BRASIL, 2015).

Um de seus objetivos específicos, na concepção do SUS, é possibilitar a análise da situação de saúde no nível local, e considerando, necessariamente, as condições de vida da população na determinação do processo saúde-doença. O nível local tem, então, responsabilidade não apenas com a alimentação do sistema de informação em saúde, mas também com sua organização e gestão (BRASIL, 2005).

No Brasil, existem cinco principais SIS ligados ao Ministério da Saúde: Sistema de Informações sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC), Sistema de Informação de Agravos de Notificação (SINAN), Sistema de Informações Hospitalares do SUS (SIH-SUS) e Sistema de Informações Ambulatoriais do SUS (SIA-SUS) (BRASIL, 2019).

Existem ainda muitos outros sistemas que são operados pela rede de serviços do SUS, servindo como fontes complementares de dados, como: Estratégia de Informatização da Atenção Básica (e-SUS AB), Sistema de Informações de Vigilância Alimentar e Nutricional (SISVAN), Sistema de Informações do Programa Nacional de Imunização (SI-PNI), Sistema de Informações do Câncer da Mulher (SISCAM), Sistema de Informação sobre Orçamento Público em Saúde (SIOPS), Sistema de Informações da Anvisa, entre outros (BRASIL, 2019). Porém, devido à importância no que diz respeito aos objetivos dessa Tese, o enfoque será dado ao SIM.

#### **2.4.2.1 Sistema de Informações sobre Mortalidade – SIM**

O primeiro sistema de informação sobre estatísticas vitais implantado pelo MS foi o Sistema de Informação sobre Mortalidade (SIM), na década de 70, que surgiu para suprir as falhas detectadas no conjunto de dados coletados pelo registro civil, possibilitando a ampliação do conhecimento do perfil epidemiológico dos óbitos ocorridos no Brasil (FRIAS *et al.*, 2008).

A Declaração de Óbito (DO) é o instrumento que contém as informações que alimentam a base de dados do SIM, com um formulário pré-enumerado em três vias, onde a primeira fica no estabelecimento de saúde, outra vai para o cartório de registro civil e a terceira fica em posse do serviço funerário (BRASIL, 2011). Sendo composto por nove blocos que contém informações sobre identificação e residência; causas do óbito básica e associadas (segundo a Décima Revisão da Classificação Internacional de Doenças-CID-10); informações básicas

sobre o médico, e no caso de mortes de menores de um ano de idade há campos com informações relacionados a mãe e ao parto (PAES, 2018). Uma atualização da Classificação Internacional de Doenças (CID-11) foi lançada em 2018 e ficou de ser implantada em 2022 (OMS, 2019).

Atualmente o SIM tem como menor nível territorial de análise o município e periodicidade de divulgação de dados anual, sendo alimentado pelas secretarias municipais e estaduais de saúde e gerenciado pelo departamento de Análise de Situação de Saúde da SVS.

Dentre os principais benefícios e funcionalidade do SIM o Ministério da Saúde destaca: *a) Declaração de óbito informatizada; b) Agilidade na transmissão e divulgação dos dados via internet; c) Produção de estatísticas de mortalidade; d) Construção dos principais indicadores; e) Análises estatística epidemiológica e demográfica; f) Controle na distribuição das DN (municipal, regional, estadual e federal).*

Os dados do SIM permitem calcular importantes indicadores para a vigilância epidemiológica, como: coeficiente de mortalidade e mortalidade proporcional por grandes grupos de causas, por causas específicas, faixa etária, sexo, escolaridade, ocupação, o delineamento do perfil de mortalidade de uma área, etc. (BRASIL, 2010).

A captação de eventos de mortalidade no Brasil tem apresentado significativa melhora nas últimas décadas, porém o País tem sido classificado em perfis intermediários em pesquisas sobre cobertura e completude de eventos vitais (MAHAPATRA et al., 2007, MATHERS et al., 2005), apresenta volume significativo de registros com causas não definidas (LIMA; QUEIROZ, 2011, FRANÇA et al., 2013) e problemas na notificação e no fluxo de informação de óbitos (SIVIERO et al., 2013, FIGUEIROA et al., 2013).

## **2.5 Dados Popacionais e Registros de Óbitos dos Centenários**

### **2.5.1 A qualidade dos dados censitários dos centenários**

Apesar do evidente crescimento da população centenária, Coale e Caselli (1990) desde os anos 90 do século passado, já ressaltavam que o incremento acentuado da população em idades avançadas nem sempre é verdadeiro, já que a contagem desses indivíduos está sujeita a erros, mesmo em países desenvolvidos. Nos censos demográficos, problemas na contagem da população centenária ocorrem em função, principalmente de dois tipos de erros: (i) problemas na imputação das idades omitidas na entrevista e, (ii) erros na declaração de idade dos entrevistados que, normalmente, geram uma transferência líquida de pessoas de idades mais jovens para idades mais velhas (PRESTON et al, 1999).

Os tipos de erros e magnitude do seu efeito sobre os dados dos centenários variam de um censo para outro. Os principais problemas na contagem da população centenária ocorrem pelos seguintes motivos: viés oriundo da imputação das idades omitidas na entrevista, erros na declaração de idade dos entrevistados, questões de forma do questionário, captura de dados, erros de codificação e processamento (SPENCER, 1987; KRACH; VELKOFF, 1999; HUMES; VELKOFF, 2007; TURRA; 2012).

Os erros na declaração de idade ocorrem quando o requerido não responde à pergunta com precisão. Particularmente, em relação à população de centenários, os erros podem resultar de baixos níveis de educação formal entre os mais velhos, das falhas do registro de nascimentos no passado, de dificuldades cognitivas, de respostas errôneas dadas por outra pessoa no domicílio, ou simplesmente pelo valor cultural ou econômico de se alcançar o status de centenário (PRESTON et al., 1999).

Erros na contagem da população por idade e seus efeitos sobre as estimativas de mortalidade são uma das principais razões pelas quais as tábuas de vida são finalizadas em 80 ou 85 anos e mais de idade em várias populações, incluindo a população brasileira (GONZAGA, 2008). No caso do Brasil, a solução de encerrar as tabelas de vida aos oitenta anos tem efeito ainda pouco conhecido sobre as estimativas de esperança de vida, com possíveis implicações para o estudo dos determinantes da longevidade no país (CAMARANO et al, 2004; JASSEN et al, 2005) e suas consequências para as políticas públicas.

Existem alguns estudos internacionais (CHUANWAN et. al., 2012; TERBLANCHE; WILSON 2015) e nacionais (GOMES; TURRA, 2009) que estimam de forma indireta o número de pessoas em grupos de idades mais avançadas, incluindo o grupo de pessoas com 100 anos e mais, com intuito de avaliar a consistência do número de centenários recenseados pelo censo demográfico. Esses estudos evidenciaram que para estimar o número de pessoas com 100 anos e mais com base no número de óbitos dos centenários, os métodos indiretos mais apropriados seriam três: o método de Gerações Extintas e os métodos propostos por Rosenwaike (1968) e Coale e Caselli (1990).

O primeiro deles é conhecido como método das Gerações Extintas. Introduzido em 1951 pelo francês Paul Vicente, esse método tem sido muito utilizado nos estudos sobre mortalidade. Embora seja um método inconveniente pelo fato de depender de dados de coorte, coletados em um período relativamente longo de tempo, sua grande vantagem é o fato de não depender de nenhum pressuposto importante para sua aplicação.

No método das Gerações Extintas, o número de pessoas sobreviventes à idade exata  $x$ ,

no ano  $t$ , é igual ao número de óbitos com  $x$  anos ou mais de idade ocorridos no ano  $t$  e nos anos seguintes, até que o último indivíduo da geração – ou coorte – venha a falecer. Assume-se que o número de óbitos é distribuído uniformemente ao longo dos anos e que, portanto, os óbitos de cada coorte, entre as idades  $x$  e  $x+1$ , é aproximadamente igual à metade do número total de óbitos observados entre as idades  $[x, x+1)$  e  $[x+1, x+2)$ , no ano  $t$ . Considera-se ainda que a população é fechada (ROSENWAIKE, 1968; COALE & CASELLI, 1990; BOURBEAU; LEBEL, 2000).

Nos casos em que não há dados de óbitos suficientes, principalmente para aquelas coortes que ainda não foram extintas, é possível aplicar métodos alternativos. Um desses métodos foi proposto por Rosenwaike (1968) e substitui os dados de coorte faltantes por dados de período, supondo que a população em estudo é fechada e estacionária. Sendo assim, o autor propõe que o número de sobreviventes a cada idade  $x$  é aproximadamente igual ao número de óbitos com  $x+1$  anos e mais de idade observados no último ano calendário considerado. Embora de fácil aplicação, a metodologia sugerida por Rosenwaike (1968) é deficiente ao não considerar que coortes distintas podem ter diferentes tamanhos em função de variações nas taxas de crescimento decorrentes de diferentes histórias de mortalidade e fecundidade.

Os possíveis problemas gerados pelo pressuposto de estacionariedade no método proposto por Rosenwaike (1968) podem ser minimizados com o uso de métodos mais robustos que consideram os diferentes ritmos de crescimento experimentados por cada coorte. Desta forma pode-se utilizar estimativas realizadas com base na metodologia proposta por Coale & Caselli (1990). De acordo com este método, em uma população fechada com taxa de crescimento entre as idades  $x$  e  $x+1$  igual a  $1r_x$ , o número de indivíduos que atinge a idade  $x$  difere do número de indivíduos que atinge a idade  $x+1$  em virtude de variações na mortalidade corrente, observada entre as idades  $x$  e  $x+1$  anos e na história demográfica de cada coorte expressa através das diferentes taxas de crescimento de cada grupo etário.

Para o Brasil, Gomes e Turra (2009) avaliaram a consistência do número de centenários recenseados no Brasil e suas grandes regiões, comparando-o com estimativas indiretas calculadas com base no método de Gerações Extintas e nos métodos propostos por Rosenwaike (1968) e Coale & Caselli (1990). Segundo os autores, em 1991 e 2000, observaram-se diferenças significativas entre os resultados estimados indiretamente e a população recenseada pelo IBGE. Em 1991, os métodos indiretos mostram um número de centenários no país quase quatro vezes menor do que o recenseado pelo IBGE (GOMES; TURRA, 2009). Com base no método gerações extintas, os autores estimaram um total de 4.539 centenários de ambos os

sexos, contra 13.865 pessoas recenseadas no censo de 1991. Em 2000, a estimativa indireta baseada no método de Coale e Caselli (1990) indica um número de centenários também bem inferior, aproximadamente de 6.177 pessoas para ambos os sexos, contra 24.576 pessoas recenseadas pelo censo demográfico daquele ano.

Reis e Turra (2016) realizaram uma análise exploratória da distribuição espacial dos centenários no Brasil por microrregião em 2000 e 2010. A pesquisa evidenciou variação regional na qualidade dos dados censitários dos centenários e dois fatores merecem ser destacados a partir das análises.

Em primeiro lugar, notou-se que há uma alta prevalência de centenários, especialmente do sexo masculino, em regiões geográficas que são reconhecidas por terem baixa qualidade de registros vitais. As regiões Norte e Nordeste do Brasil, além do norte de Minas Gerais são reconhecidamente áreas de sub-registro de nascimentos e óbitos (LIMA; QUEIROZ, 2011; SOUZA; FREIRE, 2012) o que está atrelado aos seus piores indicadores socioeconômicos e a um sistema de coleta e registro de dados mais frágil.

Em segundo lugar, notou-se uma mudança importante nos indicadores de qualidade na contagem da população centenária em 2010, especialmente nestas áreas geográficas, reforçando a ideia de que os dados mais antigos são de pior qualidade e que na medida em que coortes mais escolarizadas, cujos membros tem melhor conhecimento sobre seus próprios eventos vitais substituem as coortes anteriores, há uma melhoria nos resultados, convergindo para os de países desenvolvidos.

Estudo realizado no Semiárido brasileiro por Medeiros e Paes (2020), corroboram com essas evidências. A pesquisa evidenciou que os dados censitários referentes aos centenários melhoraram com o tempo. Porém, mesmo com o avanço observado, os indicadores do Semiárido brasileiro ainda estão distantes dos encontrados nos países mais desenvolvidos. Com valores bem destoantes e elevados destacaram-se alguns estados do semiárido, dentre eles: Bahia e Minas Gerais. Para fins comparativos, os indicadores de centenários desses estados foram encontrados apenas no Quirquístão, Nicarágua e Andorra, países esses com baixa qualidade dos registros dos centenários. Portanto, valores referentes aos dados populacionais elevados como esses não foram observados em nenhum país que tem histórico de boa qualidade em seus dados e registros, fato esse preocupante, que merece atenção e cuidados na captação desse contingente populacional no próximo censo demográfico de 2022, principalmente em determinadas regiões do país.

### 2.5.2 Os registros de óbitos dos centenários

No Brasil, segundo dados do SIM, em 1996 a principal causa de mortalidade entre os centenários era devido a Sinais, Sintomas e Achados Anormais (Capítulo XVII do CID-10), com cerca de 35% das mortes por essa causa nessa faixa etária. As doenças circulatórias apareciam em segundo lugar com 30%, seguidas das respiratórias (15%) e as neoplasias (pouco menos que 5%). Entre os anos de 1996 a 2007 observou-se uma transformação nesse quadro: queda das causas de mortes por capítulo XVIII com aumento substancial da participação das doenças circulatórias e respiratórias, sendo as circulatórias a causa mais frequente entre os centenários (BÓS, 2010).

As mortes por Sinais, Sintomas e Achados Anormais, a partir de 2005, apresentaram queda contínua, caindo a proporções muito próxima a terceira causa em 2009, ocupada anteriormente pelas doenças respiratórias (BÓS, 2010; PEDRO; BÓS, 2017).

Analisando-se dados sobre mortalidade de centenários brasileiros em 2010, observa-se que 21,5% dos óbitos referem-se a causas mal definidas. A proporção de mortes por essas causas reflete a falta de assistência médica e a dificuldade de se estabelecer uma causa básica de óbitos nos muito idosos. No Brasil, dentre os óbitos de centenários classificados como Sinais, Sintomas e Achados Anormais clínicos e laboratoriais no ano de 2010, 34,4% ocorreram sem assistência médica e 28,4% devido à senilidade (PEDRO; BÓS, 2017).

A causa mal definida (CMD), tem sua especificação definida no capítulo XVIII da 10 revisão da CID, como “Sinais, Sintomas e Achados Anormais de exames clínicos e laboratoriais, não classificados em outra parte”, a não especificação da causa básica do óbito ocorre pelo mau preenchimento da declaração de óbito (PAES, 2018). A proporção dos óbitos por CMD é considerada como um importante indicador da qualidade dos registros de óbitos, no tocante a completude das informações sobre as causas básicas de morte. Portanto, se a proporção das CMD está elevada, isto sinaliza que a mortalidade pelas outras causas está subestimada, o que pode comprometer seriamente as avaliações e análises da mortalidade por causas dos centenários.

A mortalidade proporcional por CMD aumenta com a idade, tendo sido menor na faixa etária de 60-69 anos e maior entre centenários (JORGE, 2008). O autor relata que estudos internacionais confirmam essa dificuldade, comparando causas registradas na DO com resultados de autópsias. Os fatores mais importantes associados ao erro diagnóstico são a idade avançada e a incerteza do diagnóstico prévio. Existe uma relação considerada entre aumento de CMD e problemas de acesso aos serviços básicos de saúde. Estados brasileiros com menor

gasto per capita em saúde apresenta maiores taxas de mortalidade por CMD em idosos longevos (PEDRO; BÓS, 2017).

## 2.6 O Problema da Incompletude das Informações

A evolução na disponibilidade das bases de dados dos eventos vitais, principalmente através do SIS, torna o uso das análises quantitativas cada vez mais inerentes ao meio dos estudos epidemiológicos, que utilizam a estatística como principal ferramenta para trabalhar com as informações contidas nas bases (OLIVEIRA, M. M. DE et al., 2018).

Nesta busca pelo conhecimento empírico, a qualidade da informação é imprescindível, uma vez que, várias técnicas e aplicações ficam limitadas a completude das informações. A completude é um dos pilares de avaliação da qualidade de uma base de dados, uma vez que, a incompletude refere-se as informações em branco contidas nas variáveis o que compromete a consistência dos dados (ROMERO; CUNHA, 2006).

A perda de informações dentro de um formulário, é conhecido como problema de dados faltantes ou missing-values, que em alguns casos provocam a exclusão total do sujeito. Essa exclusão pode provocar um viés nos resultados, uma vez que, os sujeitos excluídos, podem pertencer a determinadas categorias que não terão representatividade na amostra final. Em outros casos, os valores não declarados inviabilizam o uso de determinada variável, pois a quantidade de dados faltantes na mesma é muito alta (NUNES, L. N.; KLÜCK; FACHEL, 2010).

Como critérios de avaliação da completude pode ser classificada em: excelente (menos de 5%); boa (5 a 9%); regular (10 a 19%); ruim (20 a 49%); e muito ruim (50% ou mais) (PAES, 2010). Na tentativa de obter o maior número de informações, técnicas são empregadas para resgatar esses valores por meio de re-verificação, autópsia verbal, *linkage* ou imputação dos valores, resultando esta última alternativa em uma completude total das informações finais.

O *linkage* e a imputação de dados têm sido cada vez mais utilizados por pesquisadores e órgãos governamentais por possuírem baixo custo operacional e eficiência no resgate dos dados. Contudo, a maioria das aplicações utilizam apenas uma dessas técnicas, cuja escolha depende das variáveis estudadas. Utiliza-se o *linkage* para realizar o pareamento com outras bases de dados, ou seja, o *linkage* é uma técnica de relacionamento de bases de dados que busca identificar de forma precisa se o registro de um indivíduo também está contido em outra base de dados. Contudo, as variáveis de estudo dessa tese, contidas na DO, não estão contidas em outras bases de dados de interesse, como no caso do SIH, portanto para corrigir os problemas

de incompletude das variáveis o foco traçado neste trabalho é concentrado na imputação de dados.

### **2.6.1 Imputação dos Dados**

Um problema comum em investigações científicas é a ocorrência de dados faltantes (*missing data*), especialmente na área da Saúde e das Ciências Sociais. Com frequência a ocorrência de dados faltantes dificulta o uso de métodos estatísticos tradicionais para análise dos dados e, quando os mesmos são adotados, existe o risco dos estimadores se tornarem tendenciosos por conta das diferenças existentes entre os valores da variável de interesse para os respondentes e para os não respondentes. Tal viés dificilmente é eliminado, pois na prática, não se conhece o real motivo que gerou a não resposta (RUBIN, 1987). O desenvolvimento de métodos estatísticos direcionados a solucionar problemas de dados faltantes tem sido uma área de pesquisa bastante ativa nas últimas décadas (RUBIN, 1987; RUBIN, 1996; SCHAFER, 1999).

Uma das formas usadas para lidar com dados faltantes é a exclusão daqueles indivíduos que não possuem seus dados completos para se fazer a análise estatística. Mas esta abordagem, por muitas vezes, restringe a amostra a uma proporção menor da amostra original, e possivelmente o preço a se pagar por isso é a perda de poder e precisão (STERNE et al., 2009). Ignorar o problema da não resposta não é a melhor solução, uma vez que as estimativas obtidas em tais análises podem estar viesadas e, conseqüentemente, levar a conclusões errôneas. Para contornar este problema uma alternativa comumente recorrida é a imputação de dados.

A imputação é o preenchimento dos dados ausentes com valores plausíveis para uma posterior análise dos dados completos. Diversos são os métodos ou procedimentos que se ocupam do preenchimento de valores ausentes. As primeiras técnicas desenvolvidas possuíam relativa simplicidade, denominadas de imputação única: imputação por média, imputação por substituição pelo vizinho mais próximo e por regressão linear. Contudo, a partir da imputação múltipla, onde para cada dado faltante são imputados mais de um dado, associando assim a variabilidade dos dados aos resultados, os métodos ficaram mais precisos e sofisticados. Os métodos mais atuais, como regressão estocástica, modelagem Markoviana, interpolação, redes neurais, entre outras, viabilizam a imputação múltipla de dados (FERREIRA, SANTOS e LUCIO, 2017).

A fim de se dar um tratamento adequado aos dados, é importante que o pesquisador defina algumas características dos dados faltantes, como os mecanismos que os geraram, se

existe algum padrão entre eles e a proporção de dados faltantes na base (VERONEZE, 2011).

### *Mecanismos de ausência de dados*

Segundo Little e Rubin (2002) os valores ausentes podem ser divididos em três classes de acordo com seu valor de aleatoriedade:

- Ausentes de forma completamente aleatória: Os valores ausentes distribuídos de forma completamente aleatória (*Missing completely at random* - MCAR) são os de maior grau de aleatoriedade. Eles ocorrem quando o valor da variável ausente não está relacionado com essa variável ou qualquer outra no conjunto de dados.
- Ausentes de forma aleatória: Os valores ausentes de forma aleatória (*Missing at random* - MAR) ocorrem quando o valor da variável não está relacionado com essa variável, mas sim com uma ou outras variáveis do conjunto de dados.
- Ausentes de forma não aleatória: Os valores ausentes de forma não aleatória (*Not missing at random* - NMAR) ocorrem quando o valor da variável ausente está relacionado à própria variável e a um ou mais atributos do conjunto de dados.

### *Padrões de dados ausentes*

Muitos conjuntos de dados podem ser arranjados na forma de matriz, onde as linhas são os indivíduos e as colunas são as variáveis. Os dados ausentes podem também ser classificados de acordo com o padrão de não resposta. Com a presença de unidades ausentes em uma matriz de dados multivariados, é de suma importância observar a forma com que ocorre essa ausência, o que pode ser verificado por meio dos padrões de comportamento dos dados ausentes, os quais descrevem a localização dos valores em falta (LITTLE e RUBIN, 2002).

Os tipos de padrões segundo Little e Rubin (2002) podem ser especificados como segue:

- Padrão univariado: apresenta uma falta de dados isoladamente em uma variável, o que é comum em estudos experimentais;
- Padrão monótono: geralmente ocorre em pesquisas clínicas, onde os indivíduos participantes da pesquisa em algum momento não podem continuar no estudo devido a alguns fatores. Este tipo de padrão de dados em falta é característico de experimentos longitudinais, sendo as variáveis medidas ao longo do tempo;
- Padrão geral (não-monótono): padrão conhecido como arbitrário que consiste numa dispersão de unidades ausentes por toda matriz de dados. Aparentemente é aleatório,

porém pode existir uma relação entre a falta de valores de uma variável e a tendência da falta de dados referente à outra variável medida.

Dentre os padrões descritos, os principais são os monótonos e gerais, sendo que o padrão monótono é utilizado pela maioria dos algoritmos de imputação, o que na prática não ocorre, pois normalmente as respostas estão associadas às características do indivíduo (LITTLE e RUBIN, 2002).

#### **2.6.1.1 Imputação única**

Métodos de imputação simples ou única são métodos de substituição de dados faltantes quando os dados perdidos são substituídos uma única vez por algum dos métodos citados a seguir (ENGELS; DIEHR, 2003):

##### *Substituição por um valor de tendência central*

Os dados faltantes das variáveis quantitativas são substituídos pela média dos valores conhecidos da variável. Pode ser a média geral, ou seja, a média dos valores observados, ou a média de um grupo mais similar ao do caso com dado faltante, identificado por uma ou mais variáveis categóricas presentes no banco de dados. Também podem substituir os dados faltantes pela mediana dos valores conhecidos da variável ou pela mediana de um grupo de casos mais similares. Sempre que existirem valores extremos (outliers) na amostra, é recomendado utilizar o valor da mediana ao invés do valor da média. Se a variável com dados faltantes é categórica ordinal, utiliza-se a mediana ou pode-se também utilizar o valor modal para substituição do dado faltante. Se a variável é categórica não ordinal, é recomendado utilizar a moda e se não houver moda, sorteia-se uma categoria daquelas com maior frequência.

##### *“Hot deck”*

Os valores de respondentes, que são similares em relação a variáveis auxiliares, são selecionados para a imputação. São os chamados “doadores”. Ou seja, localiza-se o indivíduo com dado observado mais parecido com o indivíduo com dado faltante em relação às variáveis auxiliares e substitui-se o dado faltante pelo valor do respondente pareado. Se houver mais de um respondente pareado, é usado o método de imputação do “vizinho mais próximo”, onde algum critério de classificação é desenvolvido para determinar o registro mais semelhante àquele com o dado faltante e esse registro se torna doador desses dados.

### *Regressão (média predita)*

Os valores imputados são preditos através de regressão simples ou múltipla, que pode ser usada simplesmente utilizando uma ou mais variáveis existentes para prever os valores faltantes de outra variável altamente correlacionada com as anteriores. Dois tipos de regressão para imputação podem ser utilizados: regressão simples e regressão com termo de adição da variância do erro. A imputação que usa a regressão faz com que indivíduos que têm os mesmos valores nas mesmas covariáveis fiquem com o mesmo valor imputado, pois o valor predito é o mesmo.

### *Estimativa de Máxima Verossimilhança*

Esse método faz referência ao algoritmo EM (Expectation-Maximization) e é atualmente um método bastante comum de imputação. O algoritmo EM é utilizado quando se deseja estimar parâmetros a partir de um conjunto de dados incompletos. É um processo iterativo em que se repetem dois passos até que haja convergência: E (Estimação) e M (Maximização). No passo E se estima os dados faltantes para completar a matriz de dados. No passo M, com os dados completados, há uma aprendizagem das probabilidades e então essas probabilidades são usadas para se fazer a inferência no passo E, assim, sucessivamente, o algoritmo é processado até que haja convergência.

#### **2.6.1.2 Imputação Múltipla**

Com a tentativa de construir um método que reflita a incerteza sobre as previsões dos dados em falta, Rubin (1987) descreveu o método de imputação múltipla, o qual substitui cada valor ausente por um conjunto de valores plausíveis que representa esta incerteza sobre o valor a ser imputado. Essa técnica possibilita, além da estimativa pontual dos parâmetros, a inclusão da incerteza da imputação dos dados na variância dos resultados estimados, corrigindo o maior problema associado à imputação única (RUBIN, 1987; SCHAFER, 1999).

A ideia por trás da Imputação Múltipla (IM) é que para cada dado faltante são imputados vários valores, por exemplo,  $m$ , ao invés de um. Com isso, são obtidos  $m$  bancos de dados completos e cada conjunto de dados é analisado usando-se procedimentos para dados completos. Após, obtém-se a estimativa pontual do parâmetro que é obtida através da média das múltiplas imputações e o seu erro padrão obtido através da variância das múltiplas imputações.

Existem três vantagens importantes da IM sobre a imputação única. Primeiro, quando

imputações são realizadas aleatoriamente numa tentativa de representar a distribuição dos dados, a IM aumenta a eficiência da estimação. A segunda vantagem da IM é que quando são feitas as  $m$  imputações sob um mesmo modelo para não-resposta, inferências válidas - isto é, que reflitam a variabilidade adicional devido aos dados faltantes sob este modelo - são obtidas simplesmente combinando inferências de dados completos de maneira simples. A terceira vantagem é que gerando imputações múltiplas sob diferentes modelos é possível um estudo da sensibilidade das inferências para vários modelos de não-resposta (RUBIN, 1987).

Existem três desvantagens claras da IM sobre a imputação única. Primeiro: é necessário mais trabalho para produzir os valores a serem imputados. Segundo: mais espaço é necessário para armazenar os dados e resultados obtidos com a IM. Terceiro: é necessário mais trabalho para analisar os bancos de dados completos pela imputação múltipla do que o banco completo pela imputação única.

Essas desvantagens não são sérias quando os  $s$  e  $m$  são modestos,  $m$  modestos são adequados quando a fração de dados faltantes é pequena. Quando a fração de dados faltantes é grande, imputações múltiplas com  $m$  modesto podem não ser completamente satisfatórias, entretanto a imputação única seria muito mais desastrosa (RUBIN, 1987).

Para realização da inferência por meio da imputação múltipla é necessário à realização de três passos, descritas a seguir.

1. Os dados ausentes são preenchidos em  $m$  ( $m > 1$ ) tempos gerando  $m$  conjuntos de dados completos por meio de técnicas adequadas de imputação. De acordo com a literatura  $m$  fica entre 3 e 10 imputações.
2. Os  $m$  conjuntos de dados completos são analisados usando técnicas estatísticas padrões.
3. Os resultados dos  $m$  conjuntos de dados completos são combinados para produzir inferência dos resultados a serem imputados.

O primeiro passo é a parte fundamental da IM, pois as técnicas de imputação utilizadas têm que preservar a relação das observações faltantes e observadas, levar em conta o mecanismo de ausência (MCAR, MAR ou NMAR) e o padrão dos dados faltantes (monotônicos ou não-monotônicos).

As Regras de Rubin (*Rubin rules*) referem-se à combinação dos resultados obtidos nas diferentes análises e podem ser usadas, independentemente do método de IM. As regras de Rubin estão amplamente divulgadas na literatura que trata de imputação múltipla, pois se trata de regras simples que resolvem o passo três da IM, isto é, a combinação dos resultados obtidos em diferentes análises. Essas regras podem ser usadas independentemente do método que se

usou para se fazer a IM (SCHAFER; GRAHAM, 2002).

A partir das  $m$  imputações realizadas, o passo dois da IM pode ser realizado, ou seja, os  $m$  bancos de dados são analisados por métodos tradicionais de análise. Finalmente, os  $m$  resultados obtidos podem ser combinados de um modo simples e apropriado como proposto por Rubin (1987).

A imputação é uma poderosa ferramenta e tem a grande vantagem de flexibilidade em manusear os dados faltantes. Entretanto, é importante que se tenha cautela, assim como em qualquer estatística que se faça uso. É claro que se o modelo de imputação não consegue capturar o mecanismo não resposta, as análises de imputação estarão comprometidas.

## **2.7 Caracterização dos Espaços Regionais**

Os conceitos de rural e urbano servem, antes de tudo, para designar dois tipos diferentes de produção e organização do espaço, embora esta dimensão não esteja dissociada do seu conteúdo relacional e social. A cada um destes tipos de espaço correspondem formas de uso social definidas, geralmente, por fatores demográficos como o número de habitantes, a densidade populacional ou por fatores econômicos como o tipo de atividade econômica, ou ainda por indicadores sociais, como a presença de determinados bens de uso coletivo. Assim, pode argumentar-se que o conceito de urbano, geralmente associado à ideia de cidade, remete para um ambiente produzido e modificado. Esta definição não parece ser suficiente, pois o espaço rural é também modificado, ainda que nele as características do ambiente natural sejam mais visíveis (RODRIGUES, 2014).

A definição de rural está sempre subsidiada pela definição de urbano, sendo na maioria das vezes, o rural classificado como o que está fora do urbano (MIRANDA; SILVA, 2013). Para Mormont (1996), “o rural é uma categoria historicamente situada, que emerge com o processo que vê as forças conjugadas da industrialização e da urbanização (a mobilidade mais que a expansão das cidades) integrarem progressivamente os campos dentro de um sistema econômico e sociopolítico unificado”.

Na década de 1980, o surgimento da abordagem territorialista, que se consolidaria nas décadas seguintes, representa uma quebra de paradigma no estudo do rural. A mudança de abordagem não se resume à perspectiva econômica. Nota-se também que o processo de modernização da agricultura levou aos campos produtos e serviços antes considerados tipicamente urbanos. Isto não significa que o rural deixou de ser espaço da agricultura, mas que

e torna um espaço multifuncional, marcado, sobretudo pela pluriatividade (KAYSER, 1990; JOLLIVET, 1997; MATHIEU, 1998).

Esta mudança de abordagem indica que a relação entre rural e urbano deixa de ser analisada de forma dicotômica, passando estas duas realidades a ser vistas como partes complementares de um mesmo território. Do ponto de vista da diferenciação quanto ao urbano, Kayser (1990) sintetiza o rural a partir de quatro dimensões:

- a) uma baixa densidade de habitantes e de construções, com a cobertura vegetal como paisagem predominante;
- b) um uso econômico do solo predominante por atividades agro-silvo- pastoril;
- c) um modo de vida dos seus habitantes caracterizado pela pertença a uma coletividade de tamanho limitado e por uma relação com a natureza.
- d) uma identidade e uma representação específicas fortemente marcadas pela cultura camponesa.

Em resumo, podem-se sistematizar as definições conceituais de rural fornecidas pela bibliografia internacional em duas abordagens paradigmáticas: a abordagem tradicional de rural, na qual este é visto como sinônimo de espaço agrícola e simétrico do urbano; e a abordagem territorialista de rural, uma abordagem em que o rural adquire identidade própria, é economicamente multifuncional e com modalidades de articulação com os espaços urbanos que ditam a sua especificidade. Em ambas as abordagens, o rural é sempre definido tendo o urbano como referência.

O debate sobre o rural brasileiro até então era centrado em dois aspetos: o seu *tamanho* e a sua *natureza*. Quando se fala em *tamanho* do rural pretende-se referir o tamanho da população rural. Ou seja, o número de pessoas que vivem nas áreas consideradas rurais. A menção à *natureza* do rural corresponde, por sua vez, às características socioeconômicas e culturais dos espaços rurais e dos seus habitantes. Deste modo são identificadas duas formas de perceber o rural e o urbano no Brasil. A primeira, através do conceito de população rural medida pelos censos realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A segunda corresponde a aspetos socioeconômicos como a presença de atividades agrícolas e a ausência de bens e serviços modernos. Os conceitos de população rural e população urbana utilizados no Brasil derivam do que o IBGE chama situação de domicílio. Isto é, a localização do domicílio quanto à área urbana ou rural de cada município (RODRIGUES, 2014).

Em relação ao meio rural vale destacar elementos como o aumento das atividades não agrícolas, a mecanização, a intensificação da pluriatividade, a valorização da biodiversidade, a

expansão do setor terciário e a intensificação de fluxos materiais e imateriais na caracterização e maior compreensão de suas dinâmicas. Por outro lado, a intensa urbanização vivenciada no país deve levar em conta hoje não apenas os processos migratórios como também o fenômeno da peri-urbanização tanto pela difusão do modo de vida urbano quanto pela construção de novas zonas residenciais.

Tendo em vista o presente cenário, o IBGE publicou a nova proposta de *Classificação e caracterização dos espaços rurais e urbanos do Brasil* (IBGE, 2017) que propõe a discussão sobre os critérios utilizados na delimitação do território nacional. O objetivo do estudo é aprimorar a divulgação do próximo Censo Demográfico do Brasil, que seria realizado em 2021 e adiado para 2022, e oferecer à sociedade avanços na diferenciação de áreas rurais e urbanas que possam servir de base para a otimização de políticas públicas e do planejamento privado. De acordo com a nova proposta, 76% da população brasileira era “urbana” em 2010, enquanto a classificação usada atualmente, via legislação municipal, indica 84,4%.

A metodologia propõe-se a definir uma tipologia rural–urbano para o recorte territorial municipal, mesmo reconhecendo a generalização necessária nessa escala de análise, uma vez que dentro de praticamente todos os municípios brasileiros encontram-se uma variedade de situações que vão desde os espaços eminentemente rurais às grandes densidades urbanas.

O critério fundamental escolhido para essa metodologia é a densidade demográfica, alinhada com tipologias bem aceitas internacionalmente como a da OCDE (Organização para a Cooperação e Desenvolvimento Econômico) e da União Europeia. Essa escolha facilita a comparabilidade dos resultados brasileiros com um número significativo de países. Além disso, optou-se por considerar o critério de acessibilidade a centros com alto nível hierárquico em relação a rede urbana. Desse modo, buscou-se critérios alternativos e complementares àqueles mais frequentes em tipologias oficiais: a patamares demográficos de localidades e a de critérios legais – que vigora atualmente no país (IBGE, 2017).

Adicionalmente, para ampliar o entendimento das áreas urbanas e rurais, qualificou-se a tipologia com um critério que leva em conta o acesso dos municípios a bens e serviços mais complexos. A localização ou a acessibilidade aos centros urbanos mais estruturados também se apresenta como um elemento importante, já que a relação das cidades menores com os centros urbanos de maior hierarquia reflete diretamente no modo de vida e na configuração do espaço. Esse aspecto contribui com o objetivo de se construir uma tipologia que rompa com a abordagem dicotômica que separa os espaços rurais dos espaços urbanos (IBGE, 2017).

Esquemáticamente, os critérios utilizados para esta nova tipologia foram assim classificados segundo IBGE (2017) e apresentados no Quadro 2.1.

Assim, a definição da tipologia efetuou-se segundo um processo de classificações e cruzamentos matriciais sucessivos com base nos seguintes critérios: população em áreas de ocupação densa, proporção da população em áreas de ocupação densa em relação à população total e localização, cujo processo metodológico encontra-se descrito detalhadamente em IBGE (2017), resultou em cinco tipologias rural/urbana: Urbano, Intermediário Adjacente, Intermediário Remoto, Rural Adjacente e Rural Remoto.

Segundo o IBGE (2017), 84,4% da população brasileira se encontra em municípios considerados predominantemente urbanos, correspondendo somente a 26,0% do total de municípios. A maior parte dos municípios brasileiros foram classificados como predominantemente rurais (60,4%), sendo 54,6% como rurais adjacentes e 5,8% como rurais remotos.

Quadro 2.1 - Critérios utilizados para nova tipologia de classificação municipal segundo o IBGE (2017).

| <b>Classificação</b>                                       | <b>Critérios para os municípios em Unidade Populacionais</b>                                                                                                                                                                        |
|------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Município predominantemente urbano</b>                  | mais de 50.000 habitantes em área de ocupação densa                                                                                                                                                                                 |
|                                                            | entre 25.000 e 50.000 habitantes em área de ocupação densa com grau de urbanização superior a 50%                                                                                                                                   |
|                                                            | entre 10.000 e 25.000 habitantes em área de ocupação densa com grau de urbanização superior a 75%                                                                                                                                   |
| <b>Município intermediário</b>                             | entre 25.000 e 50.000 habitantes em área de ocupação densa com grau de urbanização entre 25 e 50%                                                                                                                                   |
|                                                            | entre 10.000 e 25.000 habitantes em área de ocupação densa com grau de urbanização entre 50 e 75%                                                                                                                                   |
|                                                            | entre 3.000 e 10.000 habitantes em área de ocupação densa com grau de urbanização superior a 75%                                                                                                                                    |
| <b>Município predominantemente rural</b>                   | entre 25.000 e 50.000 habitantes em área de ocupação densa com grau de urbanização inferior a 25%                                                                                                                                   |
|                                                            | entre 10.000 e 25.000 habitantes em área de ocupação densa com grau de urbanização inferior a 50%                                                                                                                                   |
|                                                            | entre 3.000 e 10.000 habitantes em área de ocupação densa com grau de urbanização inferior a 75%                                                                                                                                    |
| <b>Município rural adjacente e intermediário adjacente</b> | (...) a dimensão da localização de forma que se possibilite distinguir, dentre os municípios classificados como intermediários e rurais, aqueles adjacentes a centros urbanos de maior hierarquia daqueles que se encontram remotos |

Fonte: Classificação e caracterização dos espaços rurais e urbanos do Brasil: uma primeira aproximação / IBGE (2017).

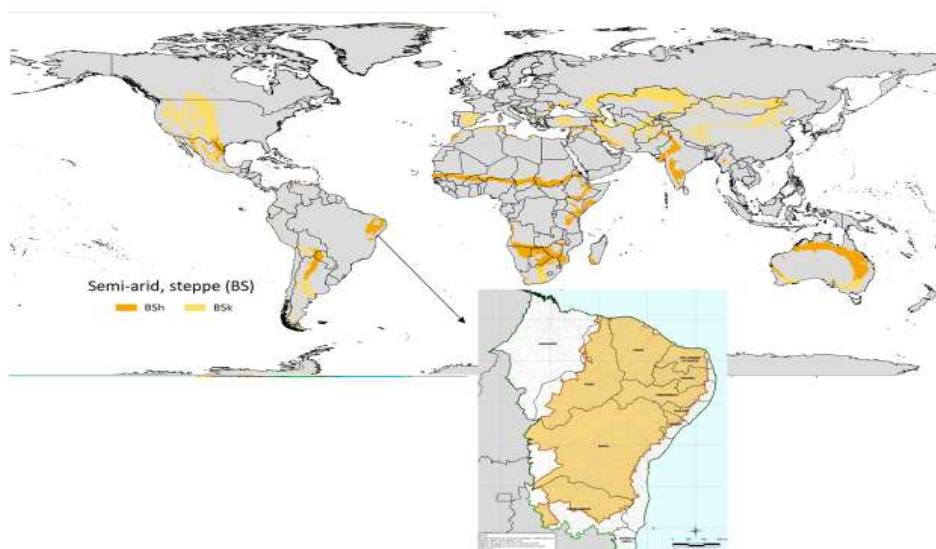
## **2.8 Caracterização do Semiárido brasileiro**

O Semiárido Brasileiro compreende uma área total de 1.128.697 km<sup>2</sup>, cobrindo quase 12% do território nacional. Esta região é composta, atualmente, por 1.262 municípios, pertencentes aos estados do Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia e Minas Gerais (SUDENE, 2017). Em 2017, viviam aproximadamente 27,9 milhões de pessoas, que representavam 48% da população do Nordeste e cerca de 13% do total da população brasileira (IBGE, 2017), sendo considerado o semiárido mais povoado do mundo, com 30 habitantes por km<sup>2</sup> (SUDENE, 2017).

De acordo com a nova tipologia de caracterização municipal proposta pelo IBGE (2017), que estabelece cinco tipos de regiões, os municípios do Semiárido brasileiro foram caracterizados em três denominações: urbano, intermediário (adjacente e remoto) e rural (adjacente e remoto). Para fins de simplificação, nesta tese a tipologia dos 1.262 municípios que pertencem ao Semiárido brasileiro foi classificada da seguinte forma e denominada de Espaços regionalizados: Urbano (total de 134 municípios), Intermediário (total de 167 municípios), Rural (total de 961 municípios). Estes tipos são considerados tanto para o estudo da população quanto para o de mortalidade.

A Resolução 1115/17 ampliou a primeira delimitação, instituída pela lei nº 7827/89, integrando municípios do Maranhão e ampliando a região de Minas Gerais. Nessa perspectiva, é considerada região semiárida aquela com precipitação pluviométrica média anual igual ou inferior a 800 mm; índice de Aridez de Thorntwaite igual ou inferior a 0,50; e percentual diário de déficit hídrico igual ou superior a 60%, considerando todos os dias do ano (INSA, 2017). O semiárido brasileiro está representado na Figura 2.3.

Figura 2.3 – Mapa do mundo e Semiárido brasileiro



Fonte: BAPTISTA; CAMPOS (2013); SUDENE (2017).

O Semiárido enfrenta vulnerabilidades climáticas e grande parte dos solos encontra-se degradada. Os recursos hídricos da região são insuficientes ou apresentam níveis elevados de poluição. As ações humanas vêm prejudicando a fauna e a flora. E a falta de proteção dos frágeis ecossistemas regionais ameaçam a sobrevivência de muitas espécies vegetais e animais, além disso, gerando riscos à ocupação humana, inclusive associados a processos, em curso, de desertificação (BRASIL, 2015b).

Dados do IBGE (2010) compilados pelo INSA (2017) mostram o Semiárido como uma área de baixo dinamismo econômico, sendo que em torno de 61% dos municípios possuem IDH baixo e muito baixo. A realidade social do semiárido é marcada por diferentes cenários. Com regiões latifundiárias, dos idos do coronelismo, (CARONE, 1971; CARVALHO, 2010; BUAINAIN; GARCIA, 2013), o “semiárido é um espaço com grande concentração de terra, da água e dos meios de comunicação, que historicamente sempre estiveram nas mãos de uma pequena elite” (ASA, 2018). Ainda segundo ASA (2018), a realidade gera distorcidos níveis de exclusão social e de degradação ambiental, fatores determinantes da crise socioambiental e econômica vivenciada pela região. A partir do INSA (2014) e do censo IBGE (2010) é possível reforçar as magnitudes dos indicadores socioeconômicos das cidades circunscritas nesta região como:

- Trata-se de uma região eminentemente rural e desértica;
- 93% dos municípios do Semiárido possui população idosa;

- O Índice de Gini acima de 0,60 para aproximadamente 1/3 de seus municípios, sinalizando uma elevada concentração da renda na região;

- 50% da população semiárida não possui renda monetária ou possui como única fonte de rendimento os benefícios governamentais. Destes, aproximadamente 60% são mulheres.

A hidrografia do Semiárido é pobre, sendo um reflexo das condições climáticas dominantes de semiaridez. O rio São Francisco tem uma significância especial para as populações ribeirinhas e da zona do Sertão que serve como sustento hídrico em longos períodos de ausência de precipitações (BRASIL, 2015).

Alguns fatores agravam as condições de sobrevivência nessa região, dentre eles é possível citar: questões relacionadas a processos produtivos exóticos à região; titularidade da terra; dimensões do latifúndio; baixa oferta de alimentos; rede de serviços insuficiente e baixa oferta de trabalho e renda. Esses fatores, contraditoriamente, têm ocupado um lugar limitado nos temas discutidos pelos gestores públicos por ocasião das discussões, formulação e implementação de políticas para promoção do bem comum nessa região (UNICEF, 2012).

Para o ano 2010, o nível da Taxa de Fecundidade Total (TFT) na região Semiárida era de 2,2 filhos em média por mulher, quase atingindo o nível de reposição populacional, estabelecido em 2,1 (BRASIL, 2012). No entanto, esse nível era de 4,55 em 1991 e de 3,15 em 2000, revelando o rápido descenso da fecundidade na região, o que implica decisivamente no processo de envelhecimento populacional. Em 2010, apresentou uma Taxa Bruta de Natalidade de 14,98 nascidos vivos para cada grupo de mil habitantes, um valor superior aos observados para as regiões Sul e Sudeste e, inferior à das regiões Norte, Centro-Oeste e Nordeste. Nos últimos censos brasileiros foram recenseados na região semiárida: em 2000 – 23.847.537 indivíduos, sendo 159.422 de idosos com 85 anos ou mais e 4.367 centenários (com 100 anos ou mais), destes 1.831 eram homens e 2.536 eram mulheres; em 2010 – 25.863.199 indivíduos sendo 237.403 de idosos com 85 anos ou mais sendo 5.259 centenários (com 100 anos ou mais), destes 1.769 eram homens e 3.490 eram mulheres.

### **3 PROCEDIMENTOS METODOLÓGICOS**

#### **3.1 Tipologia e Desenvolvimento do Estudo**

O presente estudo foi do tipo ecológico, transversal, referente aos óbitos e a população centenária do Semiárido brasileiro, com abordagem quantitativa e que fez uso de dados censitários e informações sociodemográficas e das causas dos óbitos presentes nas Declarações de Óbitos, no período de 2000 a 2020.

A população alvo para o estudo foi a população de centenários (indivíduos com 100 anos ou mais) captados e registrados pelos censos demográficos brasileiros de 2000 e 2010, os dados populacionais de 2019, foram obtidos através de técnicas matemáticas de projeção populacional.

Para o estudo da mortalidade teve-se como foco todos os casos de óbitos de indivíduos com 100 anos ou mais registrados na base de microdados do Sistema de Informações de Mortalidade (SIM). Visando minimizar as flutuações aleatórias no número de óbitos anuais, onde pequenos números geram altas variabilidades nas taxas, optou-se por trabalhar com os triênios: 2000, 2001 e 2002; 2009, 2010 e 2011; 2018, 2019 e 2020.

A unidade de análise observacional do estudo foram os centenários dos municípios pertencentes ao Semiárido brasileiro, de acordo com Resolução 1115/17 que determina a atual delimitação da região (SUDENE, 2017).

Três etapas foram traçadas para o desenvolvimento deste estudo. Na primeira, procedeu-se a análise exploratória da qualidade da contagem censitária dos centenários no Semiárido brasileiro em 2000 e 2010, com a proposta de uma metodologia de avaliação da consistência desses dados censitários; após avaliar a qualidade dos dados populacionais foram implementados métodos indiretos para estimar a população centenária, para os anos de 2000, 2010 e 2019. Na segunda, tratou-se da análise e correção dos registros de óbitos dos centenários para quantificar a cobertura, correção dos problemas na incompletude das variáveis sociodemográficas presentes na DO e na redistribuição de causas mal definidas, no período estudado. Na terceira etapa, analisou-se os fatores sociodemográficos associados aos registros dos óbitos por causas mal definidas, no último triênio - 2018, 2019, 2020.

#### **3.2 Fontes dos Dados e Variáveis do Estudo**

Os dados populacionais de centenários do Semiárido brasileiro, considerados como eventos raros, foram extraídos dos resultados do universo divulgados pelo IBGE, referentes aos

censos de 2000 e 2010, gerados com base no levantamento minucioso de todos os domicílios do país. A contagem dos indivíduos com 100 anos ou mais, de acordo com o sexo e idade, foram acessados através do site do IBGE, especificamente, no Sistema IBGE de Recuperação Automática (SIDRA). Ademais, foi realizada uma projeção dos centenários para 2019 com base nos dados dos censos de 2000 e 2010 através da média das projeções obtidas pela aplicação dos métodos de projeção populacional: Aritmético, Geométrico e Exponencial. Os dados populacionais foram, por sua vez, agregados em espaços regionalizados, denominados de acordo com a nova tipologia de caracterização municipal (Urbano, Intermediário e Rural), com o propósito de se obter melhor representatividade quantitativa.

Os registros de óbitos dos centenários foram obtidos através da base de dados do Sistema de Informação de Mortalidade, através dos microdados, por Declaração de Óbito (DO), disponibilizados no sítio do DATASUS pelo TABNET e processados com o aplicativo de tabulação do Ministério da Saúde para Windows, chamado de Tabwin.

As variáveis da DO estão divididas em blocos no formulário, conforme especificações: Bloco I – Identificação; Bloco II – Residência; Bloco III – Ocorrência; Bloco IV – Óbito fetal ou menor de um ano; Bloco V – Condições e causas do óbito; Bloco VI – Médico (informações básicas sobre o médico que assina a DO); Bloco VII – Causas externas; Bloco VIII- Cartório; IX – Localidade sem Médico. O modelo da declaração de óbito está demonstrado no Anexo 1.

Para o presente estudo foram utilizadas as variáveis dos Blocos I, II, III e V. As 13 (treze) variáveis do Bloco I referem-se à identificação e características do indivíduo declarado. As 4 (quatro) variáveis do bloco II referem-se aos dados de residência do indivíduo declarado. O Bloco III contém 11 (onze) variáveis que descrevem a ocorrência do óbito e o Bloco V contém 8 (oito) variáveis acerca das condições e as causas do óbito. Para a seleção das variáveis do estudo procedeu-se anteriormente a uma análise preliminar das variáveis, de cada bloco, sendo escolhidas aquelas que representavam as características principais dos centenários falecidos no período em questão e as informações acerca do óbito, contudo foram excluídas do estudo variáveis que apresentavam alta proporção de incompletude, ou seja, igual ou superior a 50%.

Por fim, foram selecionadas 9 (nove) variáveis: 5 (cinco) variáveis sociodemográficas do Bloco I; 1 (uma) variável do Bloco II referente ao município de residência do indivíduo declarado, essa variável foi classificada segundo critérios de ruralidade e urbanização utilizado pelo IBGE (2017), ou seja, os municípios foram classificados em: Urbano, Intermediário (adjacente e remoto), Rural (adjacente e remoto) e, logo, a nova variável foi renomeada para

caracterização municipal, 1 (uma) variável do Bloco III, referente ao local de ocorrência do óbito e 2 (duas) variáveis referentes a Assistência Médica e Causa Básica do óbito pertencentes ao Bloco V, mostradas no Quadro 2. A variável Assistência Médica corresponde a seguinte pergunta na DO: Recebeu Assistência Médica durante a Doença que ocasionou a Morte? A resposta Sim significa que o falecido, durante a doença que ocasionou a morte, teve uma assistência médica todo o tempo; a resposta Não significa que o falecido não teve assistência médica continuada, talvez tendo apenas por ocasião do óbito.

Quadro 3.1 - Descrição do código da variável, variável e categorias das variáveis sociodemográficas contidas na Declaração de Óbito e causas de óbito.

(Continua)

| <b>Código da variável</b> | <b>Variável</b>                                  | <b>Categorias</b>                  |
|---------------------------|--------------------------------------------------|------------------------------------|
| <b>V1</b>                 | <b>Sexo</b>                                      | Masculino                          |
|                           |                                                  | Feminino                           |
|                           |                                                  | Ignorado                           |
| <b>V2</b>                 | <b>Idade</b>                                     | 100 – 104 anos                     |
|                           |                                                  | 105 – 109 anos                     |
|                           |                                                  | 110 anos ou mais                   |
| <b>V3</b>                 | <b>Raça/Cor</b>                                  | Branca                             |
|                           |                                                  | Preta                              |
|                           |                                                  | Amarela                            |
|                           |                                                  | Parda                              |
|                           |                                                  | Indígena                           |
| <b>V4</b>                 | <b>Situação Conjugal</b>                         | Solteiro                           |
|                           |                                                  | Casado                             |
|                           |                                                  | Viúvo                              |
|                           |                                                  | Separado Judicialmente/Divorciado  |
|                           |                                                  | União Estável                      |
|                           |                                                  | Ignorado                           |
| <b>V5</b>                 | <b>Escolaridade (anos de estudos concluídos)</b> | Sem escolaridade                   |
|                           |                                                  | 1 a 3 anos                         |
|                           |                                                  | 4 a 7 anos                         |
|                           |                                                  | 8 a 11 anos                        |
|                           |                                                  | 12 anos ou mais                    |
|                           |                                                  | Ignorado                           |
| <b>V6</b>                 | <b>Caracterização municipal</b>                  | Urbano                             |
|                           |                                                  | Intermediário (Adjacente e Remoto) |
|                           |                                                  | Rural (Adjacente e Remoto)         |
| <b>V7</b>                 | <b>Local de ocorrência do óbito</b>              | Hospital                           |
|                           |                                                  | Outros estabelecimentos de saúde   |
|                           |                                                  | Domicílio                          |
|                           |                                                  | Via pública                        |
|                           |                                                  | Outros                             |
|                           |                                                  | Ignorado                           |

(Continuação)

| <b>Código da variável</b> | <b>Variável</b>           | <b>Categorias</b>                                                                 |
|---------------------------|---------------------------|-----------------------------------------------------------------------------------|
| <b>V8</b>                 | <b>Assistência Médica</b> | Sim                                                                               |
|                           |                           | Não                                                                               |
|                           |                           | Ignorado                                                                          |
| <b>V9</b>                 | <b>Causa Básica</b>       | Causa básica que levou ao óbito - causa terminal -, classificada segundo a CID-10 |

Fonte: Elaborado pelo autor (2021).

Para o estudo os registros do óbitos dos idosos com 100 anos ou mais foram selecionados a partir do preenchimento do item 15 da Declaração de Óbito (DO), que define quantos anos possuía o falecido no dia de seu óbito. Contudo a idade foi confirmada pela data de nascimento também contida na DO. Para análise das Causas Básicas de Óbito foram consideradas os agrupamentos de acordo com os Capítulos CID- 10ª Revisão, descritos no Quadro 3.2.

Quadro 2.2 - Descrição dos Capítulos CID-10ª Revisão, correspondentes à variável Causa Básica do óbito da Declaração de Óbito.

(Continua)

| <b>Capítulo</b> | <b>Descrição</b>                                                               |
|-----------------|--------------------------------------------------------------------------------|
| I               | Algumas doenças infecciosas e parasitárias                                     |
| II              | Neoplasmas [tumores]                                                           |
| III             | Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários |
| IV              | Doenças endócrinas, nutricionais e metabólicas                                 |
| V               | Transtornos mentais e comportamentais                                          |
| VI              | Doenças do sistema nervoso                                                     |
| VII             | Doenças do olho e anexos                                                       |
| VIII            | Doenças do ouvido e da apófise mastoide                                        |
| IX              | Doenças do aparelho circulatório                                               |
| X               | Doenças do aparelho respiratório                                               |
| XI              | Doenças do aparelho digestivo                                                  |
| XII             | Doenças da pele e do tecido subcutâneo                                         |
| XIII            | Doenças do sistema osteomuscular e do tecido conjuntivo                        |
| XIV             | Doenças do aparelho geniturinário                                              |
| XV              | Gravidez, parto e puerpério                                                    |
| XVI             | Algumas afecções originadas no período perinatal                               |
| XVII            | Malformações congênitas, deformidades e anomalias cromossômicas                |

(Continuação)

| Capítulo | Descrição                                                                                                 |
|----------|-----------------------------------------------------------------------------------------------------------|
| XVIII    | Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte |
| XIX      | Lesões, envenenamentos e algumas outras consequências de causas externas                                  |
| XX       | Causas externas de morbidade e de mortalidade                                                             |
| XXI      | Fatores que influenciam o estado de saúde e o contato com os serviços de saúde                            |

Fonte: CID-10.

Os capítulos XIX (Lesões, envenenamentos e algumas outras consequências de causas externas) e XXI (Fatores que influenciam o estado de saúde e o contato com os serviços de saúde) são utilizados exclusivamente para a classificação de dados de morbidade (BRASIL, 2013). Sendo assim, tais capítulos não foram abordados no presente estudo.

O banco de dados foi elaborado utilizando-se o aplicativo Microsoft Office Excel. Para a análise dos dados, foi utilizado o software SPSS 20.0 (*Statistical Package for the Social Sciences*, Chicago, EUA). Por se tratar de um estudo que emprega apenas dados secundários provenientes de bancos de dados de domínio público, disponibilizados online, justificou-se o não encaminhamento deste estudo para aprovação por comitê de ética em pesquisa, segundo o que estabelece o inciso III, artigo primeiro da Resolução nº 510/16 do Conselho Nacional de Saúde.

### 3.3 Análise da Qualidade dos Dados Populacionais

#### 3.3.1 Método de classificação para avaliar a qualidade dos dados censitários dos centenários

Foram coletados dados populacionais censitários dos 193 países que fazem parte da Organização das Nações Unidas (ONU) utilizando o sistema estatístico das Nações Unidas denominado “UNdata” (UNSD, 2020). Durante o processo de coleta de dados foram considerados os dados provenientes do último censo de cada país (em sua maioria realizados em 2010 e 2011) e definidos os seguintes critérios para inclusão dos países:

- 1) Ter a informação específica da quantidade de centenários recenseados, bem como da população de 85 anos ou mais, tanto para homens como para mulheres;
- 2) Ter no último censo pelo menos 1 (um) centenário recenseado do sexo masculino e do feminino;
- 3) A população do país ter no mínimo 100 (cem) mil habitantes.

De acordo com esses critérios, ao final, foram coletados dados referentes a 85 países.

Após a coleta de dados, procedeu-se a construção de três indicadores referentes aos centenários, os quais foram utilizados para a classificação proposta neste estudo: 1) Razão de Sexo (RS) dos centenários, que representa o quociente entre o total de centenários homens pelo total de mulheres; 2) Razão da população masculina com 100 anos ou mais com relação aos idosos homens com 85 anos ou mais, multiplicada por 100; 3) Razão da população feminina com 100 anos ou mais com relação às idosas com 85 anos ou mais, multiplicada por 100. Esses indicadores foram calculados para os 85 países selecionados.

Para investigar a pertinência dos indicadores selecionados como proxy para avaliar a qualidade dos dados censitários regionais das populações centenárias, utilizou-se como medida relacional o Índice de Desenvolvimento Humano (IDH) conforme definido pela Organização das Nações Unidas (UNDP, 2018). Considerou-se como hipótese que quanto maior o grau de desenvolvimento de um país melhor é a qualidade dos seus registros censitários concernentes aos centenários e população de 85 ou mais. Para tanto, foi implementado o método de regressão linear simples, com um Intervalo de Confiança de 95%.

Verificada a relação entre os indicadores selecionados e o IDH, em seguida, procedeu-se a aplicação da técnica multivariada de classificação “*cluster analysis*”. Segundo Hair et al. (2009) a análise de cluster é um conjunto de técnicas estatísticas cujo objetivo é agrupar objetos segundo suas características, formando grupos ou conglomerados homogêneos. Roses e Leis (2002) chamam a atenção no sentido de que os conglomerados obtidos devem apresentar tanto uma homogeneidade interna (dentro de cada conglomerado), como uma grande heterogeneidade externa (entre conglomerados).

O conceito de *cluster analysis*, utilizado primeiramente por Tryon (apud STATSOFT, s.d.), engloba uma série de diferentes métodos e algoritmos para agrupar objetos do mesmo tipo nas respectivas categorias. Em outras palavras, *cluster analysis* é uma ferramenta de análise que visa à triagem de diferentes objetos em grupos, de modo que o grau de associação entre dois objetos é máximo, se eles pertencem ao mesmo grupo, e mínimo em caso contrário.

Genericamente, a análise de agrupamento compreende cinco etapas (AAKER et al., 2001): 1. A seleção de elementos ou de uma amostra de elementos a serem agrupados; 2. A definição de um conjunto de variáveis a partir das quais serão obtidas informações necessárias ao agrupamento dos elementos; 3. A definição de uma medida de semelhança ou distância entre os elementos; 4. A escolha de um algoritmo estatístico de partição/classificação; 5. Por último, a validação dos resultados encontrados.

Como o objetivo da análise de cluster é agrupar objetos semelhantes, é necessária uma medida de similaridade entre os mesmos, normalmente expressa como uma função distância ou métrica. Os objetos com menor distância entre si são mais semelhantes, logo são aglomerados em um mesmo conglomerado. Já os mais distantes participam de conglomerados distintos. Segundo Cormack (1971) as distâncias mais utilizadas em análise de agrupamento são:

1. Distância Euclidiana: a distância entre dois casos (i e j), é dada pela raiz quadrada do somatório dos quadrados das diferenças entre valores de i e j para todas as variáveis ( $v = 1, 2, \dots, p$ ).

$$d_{ij} = \sqrt{\sum_{v=1}^p (X_{iv} - X_{jv})^2}$$

sendo:

$X_{iv}$  representa o valor da variável  $V$  do elemento  $i$ ,

$X_{jv}$  representa o valor da variável  $V$  do elemento  $j$ ,

$p$  é o número variáveis

2. Distância Euclidiana ao Quadrado: a distância entre dois casos (i e j), é definida como o somatório dos quadrados das diferenças entre os valores de i e j para todas as variáveis ( $v = 1, 2, \dots, p$ ).

$$d_{ij}^2 = \sum_{v=1}^p (X_{iv} - X_{jv})^2$$

sendo:

$X_{iv}$  representa o valor da variável  $V$  do elemento  $i$ ,

$X_{jv}$  representa o valor da variável  $V$  do elemento  $j$ ,

$p$  é o número variáveis

Define-se o nível de similaridade como:

$$S_{il} = \left(1 - \frac{d_{il}}{\max\{d_{il}, j, k = 1, 2, \dots, n\}}\right) \times 100$$

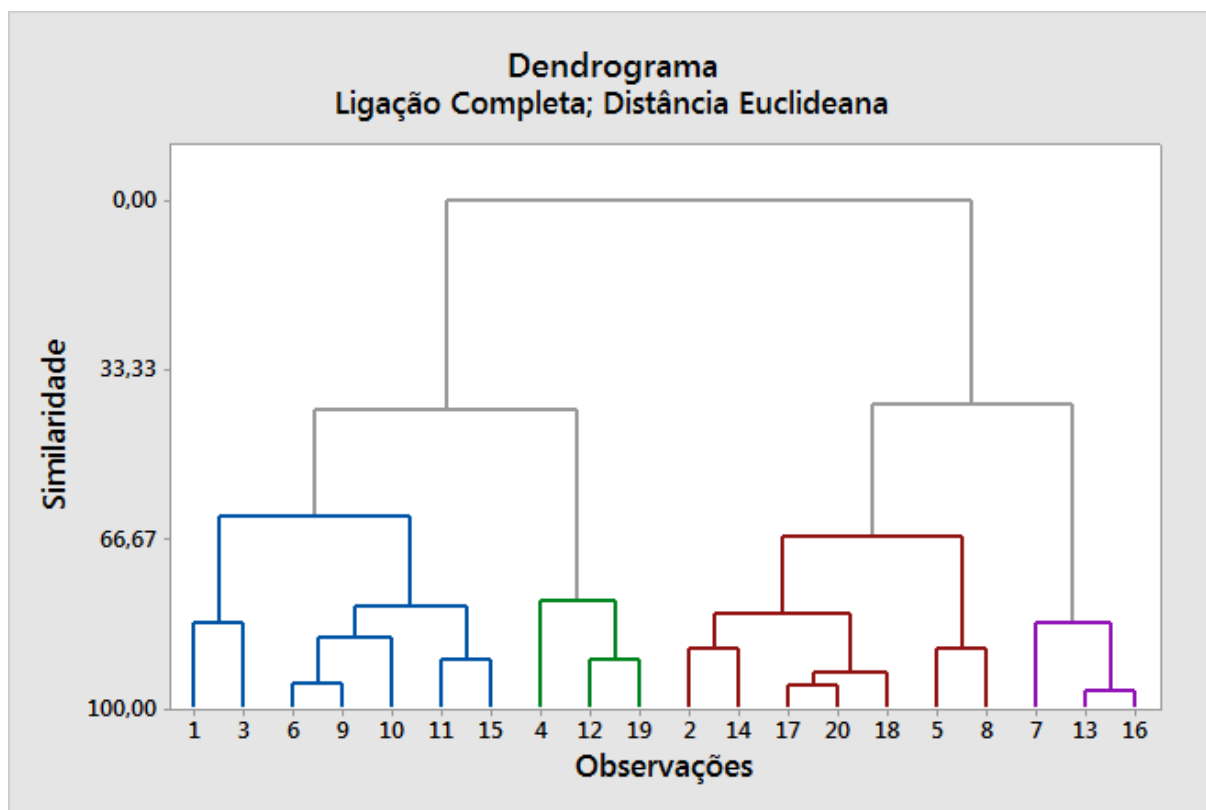
Onde,  $\max\{d_{jk}, j, k = 1, 2, \dots, n\}$  é a maior distância entre os  $n$  elementos amostrais na matriz de distância  $D_{n \times m}$  do primeiro estágio do processo de agrupamento (MINGOTI,

2005).

Segundo Malhotra (2001), existem dois tipos de processos de aglomeração que podem ser hierárquicos e não-hierárquicos. A aglomeração hierárquica se caracteriza pelo estabelecimento de uma hierarquia ou estrutura em forma de árvore, podendo ser aglomerativos ou divisivos. Enquanto a aglomeração não-hierárquica, também chamada de *K-means clustering*, inicialmente, determina ou assume um centro de conglomerado e em seguida grupa todos os objetos que estão a menos de um valor pré-estabelecido do centro.

O método hierárquico de Cluster, segundo Bussab (1990), consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, onde os elementos são agregados ou desagregados. Os grupos, nos métodos hierárquicos, são geralmente representados por um diagrama bi-dimensional chamado de dendograma ou diagrama de árvore. Neste diagrama, cada ramo representa um elemento, enquanto a raiz representa o agrupamento de todos os elementos. A Figura 3.1 ilustra um exemplo o dendograma.

Figura 3.1 – Exemplo de dendograma



Fonte: Silva, 2013.

De acordo com Mingoti (2005), no método aglomerativo, cada elemento inicia-se representando um grupo, e a cada passo, um grupo ou elemento é ligado a outro de acordo com

sua similaridade, até o último passo, onde é formado um grupo único com todos os elementos. Existe uma variedade de métodos aglomerativos, que são caracterizados de acordo com o critério utilizado para definir as distâncias entre grupos. Entretanto, a maioria dos métodos são basicamente formulações alternativas de três grandes conceitos de agrupamento aglomerativo (ANDERBERG, 1973): 1) Métodos de ligação single linkage (ligação simples), complete linkage (ligação completa), average linkage (ligação das médias), median linkage (ligação de medianas); 2) Método de centróide; 3) Métodos de minimização da soma de erros quadráticos ou variâncias (método de Ward).

Neste estudo foi aplicado o método de agrupamento de Ward. Ward (1963), o qual propõe um processo geral de classificação em que  $n$  elementos são progressivamente reunidos dentro de grupos através da minimização de uma função objetivo para cada  $(n - 2)$  passos de fusão. Inicialmente, este algoritmo admite que cada um dos elementos se constituía em um único agrupamento. Em cada passo do agrupamento dois grupos são unidos com base no valor da soma de quadrados dentro dos grupos.

Basicamente essa medida quantifica a variabilidade dos elementos alocados em um mesmo grupo em relação ao vetor de médias do grupo. Cada grupo da partição tem a sua soma de quadrados e a adição das somas de quadrados de todos os grupos dá origem a soma de quadrados total dentro dos grupos da partição. Para cada possibilidade de agrupamento é calculado a soma de quadrados total dentro dos grupos, sendo unidos os grupos que resultam no menor valor numérico dessa soma de quadrados. Dessa forma, o método de agrupamento busca a partição que minimiza a variabilidade dentro dos grupos formados, daí ser conhecido como método de mínima variância. Pode ser mostrado (Ward, 1963), que esse critério de agrupamento é equivalente ao agrupamento formado utilizando-se a distância definida para comparação dos conglomerados em cada passo do agrupamento, sendo unidos sempre os dois grupos que geram o menor valor numérico, de acordo com a equação a seguir:

$$d(C_l, C_i) = \left[ \frac{n_l n_i}{n_l + n_i} \right] (\bar{X}_l - \bar{X}_i)' (\bar{X}_l - \bar{X}_i)$$

Sendo:

$n_i$  o número de elementos no conglomerado  $C_i$ ,

$n_l$  o número de elementos no conglomerado  $C_l$ ,

$\bar{X}_l$  o centróide do conglomerado  $C_l$

$\bar{X}_i$  o centróide do conglomerado  $C_i$

$C_i$  e  $C_l$  os conglomerados que estão sendo comparados.

Ao contrário do método hierárquico, o método não-hierárquico de agrupamento não produz “árvore (ou dendograma)”, para demonstrar o resultado do agrupamento feito em cada passo. Não há propriedade de hierarquia, ou seja, grupos unidos num determinado passo podem se separar em passos posteriores. No entanto, para execução do algoritmo é necessário que a quantidade de grupos ( $k$ ) deve ser pré-estabelecida.

A partição dos dados se dá respeitando duas premissas: a coesão interna e o isolamento dos grupos é o que informa Mingoti (2005). Existem vários métodos não hierárquicos mas para o estudo deste presente trabalho o método K-médias (Everitt et. al, 2001), foi utilizado apenas como validação da partição escolhida pelo método hierárquico.

O método das K-médias é um dos mais conhecidos e utilizados em casos práticos, segundo Hartingan (1979) e citado por Mingoti (2007). É um método iterativo no qual em cada passo do algoritmo cada elemento da amostra é comparado com o vetor de médias do grupo (sementes do grupo), sendo alocado no grupo cuja distância é menor. Em cada passo os vetores de médias são re-calculados usando os elementos amostrais que foram alocados aos grupos no passo anterior. A execução do algoritmo é interrompida quando não há na re-alocação dos elementos nos vários grupos formados previamente. Para a inicialização do algoritmo é necessário definir as sementes (vetores) que definem o perfil inicial de cada grupo.

No presente estudo para implementar o *cluster analysis* foram utilizadas as seguintes variáveis para os 85 países: os três indicadores referidos acima, os quais foram construídos com os dados censitários; a expectativa de vida ao nascer e o IDH. A partir dessas variáveis os países foram discriminados em diferentes grupos (*clusters*) de acordo com o grau de similaridade entre eles.

Para a aplicação da *cluster analysis* foi utilizado neste estudo um algoritmo híbrido combinando-se os métodos hierárquico e k-médias. Uma vez que o k-médias é bastante sensível à definição dos centroides iniciais de cada *cluster*, o *cluster* híbrido utiliza os resultados encontrados pelo método hierárquico como referência para os argumentos iniciais do k-médias. O número de *clusters* foi definido por meio do dendograma. Foi feita uma inspeção visual da sobreposição entre os *clusters*, como meio de avaliar o poder de separação dos agrupamentos escolhidos. Posteriormente, foi possível identificar a existência de quatro grupos de países no dendograma. Para o uso da técnica de clusterização foi utilizado o método de Ward e a medida de distância euclidiana ao quadrado.

Após a definição dos grupos de países, foi possível identificar um cluster formado pelos países que tiveram os mais altos valores do IDH e as expectativas de vida ao nascer mais elevadas, e outro cluster contendo países em situação oposta, ou seja, baixos valores do IDH e baixas expectativas de vida ao nascer. Entre esses dois grupos díspares encontraram-se aqueles países com nível de desenvolvimento (medidos pelo IDH) e valores de expectativa de vida ao nascer intermediários.

A fim de construir intervalos de referências e, por conseguinte, as faixas classificatórias, para cada indicador selecionado foram definidos três intervalos de referência pelo valor mínimo e máximo:

- 1) intervalo de valores encontrados no cluster dos países com IDH e expectativas de vida mais elevados, denominado aqui como de “Elevada” qualidade dos dados censitários.
- 2) intervalo de valores encontrados no cluster dos países com IDH e expectativas de vida baixas, denominado como de “Reduzida” qualidade dos dados censitários.
- 3) Os valores dos indicadores encontrados entre os intervalos anteriores foram classificados como de qualidade “Intermediária”.

Destaca-se que entre os três intervalos de referência identificados não houve intersecção entre eles.

Para aqueles países categorizados como de qualidade “Intermediária” ocorreu que algum dos três indicadores foi enquadrado como de qualidade “Elevada” ou “Reduzida”. Devido a essas situações foram atribuídos pesos, para cada uma das três faixas de intervalo dos indicadores, aqui denominadas de “intervalo padrão” para avaliar de forma sistemática e coerente a qualidade dos dados censitários dos centenários:

- Intervalo padrão de “Elevada” qualidade- Peso 3
- Intervalo padrão de “Intermediária” qualidade- Peso 2
- Intervalo padrão de “Reduzida” qualidade- Peso 1

Finalmente, a partir da observância das faixas de intervalo padrão e da soma dos seus pesos atribuídos para os três indicadores verificados em cada país, foram designados cinco escores ou classificações para avaliar a qualidade dos dados censitários do contingente de idosos centenários dos países (Quadro 3.3).

Quadro 3.3 - Classificação dos escores para avaliar a qualidade dos dados censitários do contingente de idosos centenários.

| CLASSIFICAÇÃO | SOMA DOS PESOS |
|---------------|----------------|
| Muito Boa     | 9              |
| Boa           | 8 ou 7         |
| Regular       | 6              |
| Baixa         | 5 ou 4         |
| Muito Baixa   | 3              |

Fonte: Elaborado pelo autor, 2021.

Após a construção desta proposta de classificação, a metodologia classificatória foi implementada para o Brasil e o Semiárido brasileiro com o objetivo de avaliar a qualidade dos dados censitários dos centenários do país e da região semiárida, para os anos de 2000 e 2010, últimos censos realizados no Brasil, assim como para o ano de 2019, dados populacionais construídos através de projeções provenientes dos censos anteriores

### 3.3.2 Método indireto para estimar a população de centenários

Conforme explicitado na Seção 2.5.1, referente à qualidade dos dados censitários dos centenários, no Capítulo 2, os métodos mais usados para estimar a população de centenários são: o método de Gerações Extintas e os métodos propostos por Rosenwaike (1968) e Coale e Caselli (1990). Para este trabalho, o número de centenários estimados para os anos 2000, 2010 foi calculado através do método das Gerações Extintas, contudo para 2019, como não é possível implementar a análise da coorte, foi utilizado o método indireto de Rosenwaike. Não foi utilizado o método de Coale e Caselli, pois de acordo com estudos internacionais (BRAVO, 2007; TERBLANCHE; WILSON, 2015) o método de Gerações Extintas e suas generalizações (método de Rosenwaike) são mais adequadas para o estudo de centenários. As estimativas indiretas foram construídas para o Semiárido brasileiro e seus espaços regionalizados.

#### *Método das Gerações Extintas*

O método das Gerações Extintas faz uso de uma coorte de um número de sobreviventes em um tempo  $t$  que vai se extinguindo ao longo do tempo subsequente. Este método tem sido bastante aplicado para população com 80 anos ou mais e sua grande vantagem é o fato de não depender de nenhum pressuposto importante para sua aplicação, embora seja um método inconveniente pelo fato de depender de dados de coorte, coletados em um período longo de tempo. Trata-se de um método relativamente robusto na presença de erros na idade declarada no censo e no registro de mortes, e também quando o diferencial de cobertura entre dois censos

varia com a idade (COALE; CASELLI, 1990; PRESTON et al, 1998; BOURBEAU; LEBEL, 2000; ROSENWAIKESTONE, 2003).

A aplicação do método das Gerações Extintas é simples: considerando que a população é fechada, o número de pessoas sobreviventes à idade exata  $x$ , no ano  $t$ , é igual ao número de óbitos com  $x$  anos ou mais de idade ocorridos no ano  $t$  e nos anos seguintes, até que o último indivíduo dessa geração – ou coorte – venha a falecer. O número de óbitos dessa geração, entre as idades  $x$  e  $x+1$ , é aproximadamente igual ao total de óbitos observados no mesmo intervalo etário no ano  $t$  e no ano  $t+1$ . Por exemplo, todos os sobreviventes da geração nascidos em 1909 atingiriam 100 anos de idade em 1º de janeiro de 2010. Pressupõe-se que durante o ano calendário 2010 a metade dos óbitos observados de pessoas com 100 anos pertenciam à coorte nascida em 1909 e a outra metade à coorte nascida em 1910. O mesmo é aplicado para as demais idades, porém nos anos posteriores somasse todos os óbitos que ocorreram na referida idade. A soma de todos os óbitos ocorridos até o último ano calendário, que representa o momento que a referida coorte foi extinta, corresponde a população com 100 anos ou mais em 1º de janeiro de 2010 (Rosenwaike, 1968; Coale & Caselli, 1990; Bourbeau & Lebel, 2000).

Sendo assim, a população em idade  $x$  pode ser descrita pela expressão:

$$P_x^t = \left[ \sum_{i=0}^n (d_{x+i}^{t+i} + d_{x+i+1}^{t+i}) \right], \text{ onde}$$

$P_x^t$  : é a população na idade  $x$  no início do ano  $t$ ;

$n$ : é o número de anos necessários até que a coorte seja extinta;

$d$ : representa o número de óbitos em cada ano calendário considerado.

Para a aplicação do método de Gerações Extintas, assume-se que a coorte de centenários extingue-se aos 110 anos. Embora haja registro de pessoas em todo o mundo que sobreviveram às idades superiores a esse limite, o pressuposto de idade máxima de 110 anos é consistente com trabalhos internacionais. Wilmoth (1997), por exemplo, demonstra que é possível obter estimativas populacionais confiáveis assumindo um limite de vida de 109 anos para as mulheres e de 107 anos para os homens, portanto o limite de 110 anos para a idade de extinção da coorte torna-se viável.

#### *Método proposto por Rosenwaike*

Nos casos em que não há dados de óbitos suficientes, principalmente para aquelas coortes que ainda não foram extintas, é possível aplicar métodos alternativos. Um desses métodos foi proposto por Rosenwaike (1968) e substitui os dados de coorte faltantes por dados de período, supondo que a população em estudo é fechada e estacionária. Sendo assim, o autor propõe que o número de sobreviventes a cada idade  $x$  é aproximadamente igual ao número de óbitos com  $x+1$  anos e mais de idade observados no último ano calendário considerado. Por exemplo, a coorte com 100 anos de idade em 1º de janeiro de 2018 completará 103 anos até o final do ano calendário 2021. Como há sobreviventes dessa geração com mais de 103 anos que morrerão em períodos seguintes para os quais não se dispõe de informações, seu número pode ser aproximado pelo número de óbitos ocorridos em 2021 entre pessoas de 104 anos e mais de idade naquele ano. Embora de fácil aplicação, a metodologia sugerida por Rosenwaike (1968) é deficiente ao não considerar que coortes distintas podem ter diferentes tamanhos em função de variações nas taxas de crescimento decorrentes de diferentes histórias de mortalidade e fecundidade. No entanto, considera-se que essas diferenças históricas não são tão grandes em virtude da proximidade dos anos das coortes, os quais passaram por padrões de mortalidade e de fecundidade similares.

### **3.4 Análise da Qualidade dos Registros de Óbitos**

Os métodos para avaliar os registros de óbitos incompletos e incorretos podem ser divididos em quatro categorias: cobertura, completude, regularidade e fidedignidade (PAES, 2018). Os dois primeiros, a cobertura (enumeração correta dos óbitos) e a completude (informações completas) são considerados como os indicadores mais preocupantes nas avaliações da qualidade dos registros de óbitos, particularmente a cobertura, já que para a contabilidade estatística a quantidade de registros tem uma maior relevância político-administrativa e de cidadania.

O cálculo das reais magnitudes e a precisão das estimativas dos indicadores de mortalidade do Semiárido brasileiro têm enfrentado importantes entraves devido à subenumeração dos registros de óbitos e aos problemas na qualidade das informações contidas nas declarações, dentre elas as causas básicas de óbito. Assim, a utilização de métodos de correção reduz problemas de subestimação dos níveis de mortalidade, proporcionando uma melhor qualidade da informação que pode servir mais adequadamente à gestão pública dos sistemas de informação.

Neste estudo, os óbitos de centenários no Semiárido foram analisados para definição da sua qualidade e depois foram corrigidos para sub-registro, redistribuição de óbitos por causas mal definidas, e por fim, correção da incompletude das variáveis sociodemográficas disponíveis que fazem parte da DO. A seguir, é descrito cada uma das etapas da correção.

### **3.4.1 Correção dos sub-registros dos óbitos**

Para mensuração e estimação das cobertura dos óbitos foram desenvolvidos métodos estatísticos que podem ser classificados em dois tipos: diretos - que fazem uso de dados observados e buscam extrair conclusões após análise observacional e indiretos – baseados em informações censitárias ou de pesquisas e fazem uso de modelos, equações e funções matemáticas.

Pelo menos dez métodos na literatura mundial foram propostos nas últimas décadas, os quais permitem estimar o grau de cobertura dos óbitos. A aplicação deles sugere que mesmo dados deficientes podem se constituir em referências úteis para estimar os níveis e tendências de mortalidade. Contudo, esses métodos diferem quanto aos pressupostos, as exigências e o grau de precisão.

Segundo Paes (2005) em uma tentativa de síntese, os métodos propostos podem ser alocados em três categorias:

- 1) Métodos que exigem o suposto de estabilidade da população, entre eles, destaca-se a Equação do Balanço de Crescimento, proposta por Brass (1975);
- 2) Métodos que flexibilizam o suposto de estabilidade da população, por exemplo, o método da Equação Geral de Balanço (“*Generalized Growth Balance*” ou GGB, em inglês), formulado por Hill (1987); o método das Gerações Extintas Sintéticas (“*Synthetic Extinct Generations*” ou SEG, em inglês), desenvolvido por Bennett e Horiuchi (1981); o método que realiza combinação entre o método estimador bayesiano empírico (BE) e Equação Geral de Balanceamento (EGB), proposto por Freire e Justino (2016).
- 3) Métodos que dispensam o suposto de estabilidade da população, como o de Courbage e Fargues (1979).

Esses métodos são utilizados para estimar a cobertura de óbitos da população adulta, porém não há na literatura proposta de técnicas específicas para o contingente populacional de centenários. Diante deste desafio e com o objetivo de atendê-lo, ainda que parcialmente, para assim encontrar um caminho que possibilite qualificar os registros dos óbitos dos centenários, estabeleceu-se alguns direcionamentos:

1) usar outros sistemas de informação, no caso o Sistema de Informações Hospitalares (SIH) ou outra fonte de dados, a exemplo da Previdência Social, como comparativo de total de óbitos confrontado com os registros encontrados no SIM, durante a coleta de dados;

2) utilizar as estimativas da cobertura de óbitos da população adulta, que já foram construídas em estudos científicos anteriores, para o Semiárido brasileiro, como sendo o percentual de cobertura máxima para essa região, baseado na suposição de que a cobertura dos centenários é mais baixa.

Buscando atender ao direcionamento 1), após a coleta de dados e uma análise preliminar das variáveis do banco de dados do presente estudo, foi possível identificar que dos 28.985 registros de óbitos, ocorridos no Semiárido, no período de 2000 a 2020, 69,2% tiveram como local de ocorrência o domicílio, ou seja, fora do ambiente hospitalar, o que dificultaria utilizar o SIH para comparar a cobertura dos óbitos para essa faixa etária. Quanto ao acesso e investigação dos bancos de dados da Previdência Social, as investigações seguem como uma possibilidade de uso.

Dessa forma, o direcionamento 2) pareceu mais promissor. Em virtude da dificuldade de implementar os métodos indiretos em faixas etárias limítrofes, como é o caso dos centenários, assim como pela dificuldade de acesso e confiabilidade dos dados e registros desse contingente populacional, o caminho proposto como alternativa foi o de calcular as coberturas de óbitos da população adulta utilizando um método convencional e em seguida comparar com as estimativas da cobertura de óbitos já calculadas para o Semiárido brasileiro (PAES, 2013; PAES; SILVA; BRAGA, 2014), os quais foram desenvolvidos pelo Laboratório de Estudos Demográficos, da Universidade Federal da Paraíba.

Nos trabalhos já desenvolvidos, para calcular a cobertura de óbitos em adultos na região semiárida foi utilizado o método da equação de balanço de crescimento proposto por Brass (1975) e o método adaptado por Hill (1987), os quais se baseiam na teoria das populações estáveis e configuram-se como robustos, mesmo em situação de descumprimento de alguns dos supostos. A escolha desses métodos foi baseada na flexibilidade de uso, nas suposições e nas exigências quanto aos dados básicos. Desta maneira, para o cálculo das estimativas das coberturas geradas nesta tese seguiu-se os mesmos procedimentos adotados nas referências.

#### *Método da Equação do Balanceamento de Brass*

Para a aplicação do método é necessário o conhecimento da distribuição da população por faixa etária além da distribuição dos óbitos. Estimar a cobertura de óbitos de uma

determinada região por este procedimento equivale ao ajuste do seguinte modelo de regressão:

$$\frac{N(a)}{N(a+)} = r + k \cdot \frac{D'(a+)}{N(a+)}$$

Onde:

$N(a)$  = população na idade exata  $a$  ;

$N(a+)$  = somatório de pessoas que estão na idade exata até um limite de idade “w” qualquer;

$D'(a+)$  = óbitos registrados e afetados por erros na idade  $a +$ ;

$r$  = taxa de crescimento natural;

$k$  = fator de correção para a cobertura dos óbitos.

Para que o método de Brass possa fornecer bons resultados é necessário que alguns pressupostos sejam satisfeitos, além de que ele funciona melhor para as idades adultas. Sendo os seguintes pressupostos: 1) Estabilidade relativa da população; 2) A distribuição por idade dos óbitos deve apresentar um grau de cobertura constante; 3) A distribuição por idade da população não deve apresentar problemas de enumeração ou de declaração da idade.

No entanto, Paes (2007) recomenda algumas estratégias para a superação dessas dificuldades quando elas estão presentes nos dados. Para a seleção dos grupos etários mais estáveis a serem usados no modelo de regressão, os seguintes critérios foram estabelecidos para a estimativa das coberturas dos óbitos:

1. Considerar um limite de variação das faixas etárias entre 20 e 65 anos;
2. Calcular a mediana de pelo menos cinco combinações sequenciais de faixas etárias das coberturas calculadas que variaram entre o limite adotado: 20-65 anos. Por exemplo: 20-45, 20-50, 20-55, 20-60, 20-65. Em seguida, 25-45, 25-50,..., 25-65, e assim por diante;
3. Calcular a média das medianas das coberturas entre as várias combinações quando estas foram muito diferentes;
4. Coberturas acima de 100% foram descartadas.

#### *Método da equação do Balanço do Crescimento Geral de Hill*

O método do Balanço do Crescimento Geral (General Growth Balance – GGB) proposto por Hill (1987) é uma generalização da Equação do Balanço do Crescimento proposta por Brass (1975). O GGB utiliza a distribuição de crescimento da população por faixa etária em dois censos e o total de óbitos por grupo etário no período intercensitário para estimar as taxas de crescimento específicas por idade.

Com a relação obtida entre a diferença das taxas de entrada e de crescimento e a taxa de mortalidade em cada grupo etário, obtém-se um intercepto com o qual se estima uma inclinação que serve como um indicador do grau de cobertura do registro de óbitos

em relação de óbitos em relação à cobertura média de ambos os censos, além de capturar a variação no grau de cobertura no período intersensitário (HILL, 1987).

A equação proposta por Hill é dada por:

$$\frac{N(a)}{N(a+)} = r(a+) + \frac{1}{t} \cdot \ln\left(\frac{k_1}{k_2}\right) + \frac{(k_1 \cdot k_2)^{1/2}}{C} \cdot \frac{D(a+)}{N(a+)}$$

Em que:

$N(a)$  = número de pessoas-anos vividos em uma idade exata  $a$ ;

$N(a+)$  = número de pessoas-anos vividos em uma idade exata  $a$  em diante durante um período intercensitário  $t$ ;

$D(a+)$  = é o número de mortes de uma idade exata  $a$  em diante no período intercensitário  $t$ ;

$r(a+)$  = é a taxa de crescimento cumulativa por idade;

$k_1$  e  $k_2$  = as coberturas do primeiro e do segundo censo, respectivamente;

$C$  = fator que representa a cobertura dos registros de óbitos.

Estima-se  $C$  assumindo que o maior valor entre  $k_1$  e  $k_2$  seja igual a 1. Assim, se  $\frac{k_1}{k_2} < 1$ , assume-se  $k_2$  igual a 1 e  $k_1 = e^{a(t_2 - t_1)}$  e  $C = \frac{e^{a(t_2 - t_1)}}{b}$ ; se  $\frac{k_1}{k_2} > 1$ , assume-se  $k_1$  igual a 1 e  $k_2 = e^{a(t_2 - t_1)}$  e  $C = \frac{e^{-a(t_2 - t_1)}}{b}$ .

Onde,

$$a = \frac{\ln\left(\frac{k_1}{k_2}\right)}{t_2 - t_1} \text{ e } b = \frac{(k_1 \cdot k_2)^{1/2}}{C}$$

$t_2 - t_1$  se referem, respectivamente, a dois momentos, usualmente censitários.

Nos estudos realizados pelo LED/UFPB, foi possível estimar a cobertura de óbitos entre os estados que compõem o Semiárido brasileiro fazendo uso da técnica do Balanço de Crescimento proposto por Brass (1975) e de Hill (1987) para os anos censitários: 1991, 2000 e 2010. Para a série histórica proposta nesta pesquisa, utilizou-se cálculo da cobertura dos registros de óbitos para a população adulta pelo método da equação de balanço

de crescimento (Growth Balance Method - GBM) originalmente proposto por Brass (1975) para idades acima dos quatro anos.

Esse método, como as demais propostas clássicas de estimação de cobertura dos óbitos, não foi desenhado para contemplar populações de 100 anos ou mais. Como não está disponível na literatura um método específico para estimar a cobertura dos óbitos dos indivíduos com 100 anos ou mais, a estimativa da cobertura obtida para os adultos foi admitida como uma proxy para os centenários, embora possa ser considerado que a cobertura dos óbitos dos centenários seja menor que a dos adultos. As coberturas foram construídas para os anos de 2000, 2010 e 2019, para os espaços regionalizados do Semiárido brasileiro (urbano, intermediário e rural).

Por fim, devido aos erros inerentes a toda estimativa e para facilitar a interpretação das estimativas das coberturas dos óbitos, utilizou-se a classificação adaptada por Paes (2007) que estabeleceu quatro categorias para classificação das coberturas dos óbitos adultos para ambos os sexos: muito bom ( $>90\%$ ); bom ( $81\%–90\%$ ); regular ( $71\%–80\%$ ) e deficiente ( $\leq 70\%$ ).

### **3.4.2 Correção dos problemas de incompletude das variáveis**

A perda de informações dentro de um formulário, é conhecido como problema de dados faltantes ou *missing-values*, que em alguns casos provocam a exclusão total do sujeito. Essa exclusão pode provocar um viés nos resultados, uma vez que, os sujeitos excluídos, podem pertencer a determinadas categorias que não terão representatividade na amostra final. Em outros casos, os valores não declarados inviabilizam o uso de determinada variável, quando a quantidade de dados faltantes na mesma é muito alta (NUNES; KLÜCK; FACHEL, 2010).

Como critérios de avaliação da completude das variáveis neste estudo utilizou-se a classificação proposta por Romero e Cunha (2006), que define a completude das variáveis de acordo com o percentual de *missing-values* : excelente (menos de 5%); boa (5 a 9%); regular (10 a 19%); ruim (20 a 49%); e muito ruim (50% ou mais).

Na tentativa de obter o maior número de informações, técnicas são empregadas para resgatar esses valores por meio de re-verificação, autópsia verbal, linkage ou imputação dos valores, resultando esta última alternativa em uma completude total das informações finais. O linkage e a imputação de dados têm sido cada vez mais utilizados por pesquisadores e órgãos governamentais por possuírem baixo custo operacional e eficiência no resgate dos dados.

Como as variáveis de estudo dessa tese, contidas na DO, não estão contidas em outras bases de dados de interesse, como no caso do SIH, para corrigir os problemas de incompletude das variáveis utilizou-se a técnica de imputação de valores para completar as informações

ausentes das variáveis contidas na DO selecionadas neste estudo. Para aplicação desta técnica foi necessário avaliar o mecanismo e o padrão de dados ausentes dos dados a serem imputados, cujas descrições e significados foram dados na Sessão 2.6. Os três mecanismos de não-resposta segundo Little e Rubin (2019) são:

- i) ausentes de forma completamente aleatória;
- ii) ausentes de forma aleatória e
- iii) ausentes de forma não-aleatória.

Os padrões de não-resposta se referem à forma com que os valores ausentes estão distribuídos em uma base de dados e podem ser classificados em:

- i) padrão univariado,
- ii) padrão monotônico
- iii) padrão não-monotônico.

Após essas verificações foram imputadas as informações ausentes das variáveis sociodemográficas e local de ocorrência do óbito.

De acordo com os percentuais de dados ausentes nas variáveis foi possível determinar o método de imputação a ser utilizado. Segundo Harrell Jr (2001) é possível serem definidas linhas gerais para a escolha entre os métodos de imputação de acordo com a proporção de dados faltantes em qualquer uma das variáveis.

- Proporção  $\leq 0,05 \rightarrow$  Neste caso pode ser usada imputação única ou analisar somente os dados completos.
- Proporção entre 0,05 e 0,15  $\rightarrow$  Imputação única pode ser usada aqui provavelmente sem problemas, entretanto o uso da imputação múltipla é indicado.
- Proporção  $\geq 0,15 \rightarrow$  A imputação múltipla é indicada na maior parte dos casos.

Optou-se pela imputação múltipla, pois essa técnica produz resultados não viesados e com erros padrão apropriados.

Uma questão relevante na imputação múltipla é a escolha da quantidade de imputações  $m$  a serem realizadas. Alguns autores argumentam que um  $m$  entre 3 e 5 já é suficiente para gerar resultados satisfatórios (RUBIN, 1996; SCHFER, 1999). A decisão sobre a quantidade de imputações se baseia em um indicador denominado por Rubin (RUBIN, 1996) de Eficiência Relativa ( $ER$ ), expresso como função da quantidade de imputações ( $m$ ) e do percentual de dados ausentes da variável ( $\lambda$ ). O resultado do indicador aponta o percentual de eficiência dos valores imputados de cada variável. Assim, o cálculo para obtenção da Eficiência Relativa é dado por:

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1}$$

De acordo com os dados observados e o percentual de dados faltantes, foi calculado a eficiência relativa para cada variável e após esse passo foi optado o número de imputações para os dados faltantes das variáveis do estudo.

A técnica de imputação múltipla cria  $m$  cópias da base de dados onde os valores ausentes são substituídos por valores plausíveis imputados através de técnicas adequadas de estimação. Os valores imputados para o padrão monotônico são obtidos por meio de métodos estatísticos inferenciais como Método da Regressão Linear Bayesiana ou Método da Média Preditiva e para o padrão não-monotônico tem-se o método de Monte Carlo baseado em Cadeias de Markov. Um número  $m$  de bancos distintos e completos são gerados, e cada um deles deve ser analisado (RUBIN, 1996; SCHFER, 1999). Para a combinação entre todas as  $m$  estimativas individuais de todas as imputações realizadas, recorreu-se às Regras de Rubin (RUBIN, 1987) que se utiliza de estimativas da média e da variância entre as imputações.

Para cada análise das  $m$  bases de dados completas, obtém-se uma estimativa para um parâmetro escalar de interesse  $Q$ , ou seja,  $Q_j, j=1,2,\dots,m$ . Segundo Schafer (1999),  $Q$  pode ser qualquer medida escalar a ser estimada, tal como média, correlação, coeficiente de regressão ou razão de chances. Então a estimativa combinada  $\bar{Q}$  será a média das estimativas individuais ( $\widehat{Q}_j$ ):

$$\bar{Q} = \frac{1}{m} \sum_{j=1}^m \widehat{Q}_j$$

Para a variância combinada, primeiramente calcula-se a variância dentro das imputações:

$$\bar{U} = \frac{1}{m} \sum_{j=1}^m U_j$$

e a variância entre imputações:

$$B = \frac{1}{m-1} \sum_{j=1}^m (\widehat{Q}_j - \bar{Q})^2$$

Então, a variância total, que é a variância combinada, será:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right)B$$

Aprofundando o estudo da incompletude das variáveis, foi realizada separadamente a análise da completude da variável Causas Básicas de óbito, as quais foram consideradas os agrupamentos de acordo com os Capítulos CID- 10ª Revisão, descritos no Quadro 3.2 da Sessão 3.2 desta Tese.

A falta de especificação na causa básica de óbito pelo responsável do preenchimento da DO foi classificada pelas Nações Unidas como “Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte” sendo o capítulo XVIII da CID-10, portanto quando a Causa Básica de Óbito for preenchida pelo Capítulo XVIII trata-se de Causa Mal Definida (CMD).

Para quantificar a magnitude das causas mal definidas, calculou-se seu percentual em relação ao total de mortes de cada Espaço Regionalizado, nos triênios estudados. Utilizou-se a classificação proposta por Chackiel (1987) que estabeleceu quatro categorias para os percentuais de óbitos com causas mal definidas para as regiões, de acordo com o seguinte critério: adequado (<10%); pouco adequado (10%–15%); inadequado (16%–30%) e altamente inadequado (>30%). Para fins de compatibilização com a notação usada posteriormente, para as causas desconhecidas, foi adotada respectivamente a seguinte terminologia: boa, satisfatória, pouco confiável ou regular e deficiente.

Outro percentual calculado foi o de óbitos cuja causa básica é desconhecida. A mortalidade por “causas desconhecidas” representa o resultado da combinação de dois fatores: os óbitos não registrados, cujas causas de morte, conseqüentemente, não são conhecidas, e os óbitos registrados, cujas causas básicas foram mal definidas e que, também, não se conhecem (PAES, 2007).

A partir de informações sobre o percentual da cobertura dos óbitos e o percentual de causas mal definidas foi possível chegar ao percentual de óbitos com causas básicas desconhecidas, o qual pode ser obtido pela seguinte expressão:

$$O_d = (100 - C) + \left(\frac{C * M_d}{100}\right)$$

onde:

$O_d$  : Percentual de óbitos cuja causa básica é desconhecida.

$C$  : Percentual de cobertura dos óbitos.

$M_d$ : Percentual dos óbitos com causa básica mal definida.

Para entender a diversidade regional e qualidade dos registro de acordo com a análise das “causas desconhecidas” adotou-se uma variante de uma outra classificação proposta por Chackiel (1987) para países da América Latina (PAES, 2007). Regiões com informação: boa (<5%); satisfatória (5%–25%); pouco confiável ou regular (25%–40%); deficiente ( $\geq 40\%$ ).

### 3.4.3 Redistribuição de óbitos por causas mal definidas

Para combater o problema na declaração das causas mal definidas, poderia ser feita uma redistribuição proporcional dos óbitos entre as outras causas bem definidas, porém essa solução só deve ser aplicada quando o percentual de causas mal definidas é menor que 5% (PAES, 2018).

Uma outra maneira de recuperação de óbitos mal definidos é através do método conhecido como Autópsia Verbal (MATHERS, et. al, 2005), definida pela Organização Mundial de Saúde (WHO, 2007) como uma entrevista domiciliar com a família e/ou cuidador do indivíduo, utilizando um questionário estruturado para induzir sinais e sintomas ou outras informações que possam ser utilizadas para atribuir a provável causa subjacente de morte. Essa busca ativa faz parte de estratégias implantadas pelo MS desde 2006, através da criação da Rede Nacional de Serviços de Verificação de Óbito (SVO) e Esclarecimento da Causa Mortis.

Um outro método foi proposto por Ledermann (1955), o qual foi extensivamente explorado por Vallin (1987) para regiões da França.

No presente estudo, utilizou-se o método de Ledermann para redistribuir entre os Capítulos da Classificação Internacional das Doenças (CID-10) os óbitos de centenários no Semiárido brasileiro, desagregados entre os espaços regionalizados, classificados no grupo de causas mal definidas (Capítulo XVIII).

O método foi escolhido por ter um caráter mais teórico, pois utiliza a regressão linear simples na sua aplicação. Para tanto, foi considerada como variável resposta a proporção de óbitos dos centenários pelos principais grupos de causas: Doenças infecciosas e parasitárias (Cap. I), Neoplasias (Cap. II), Doenças endócrinas, nutricionais e metabólicas (Cap. IV), Doenças do aparelho circulatório (Cap. IX), Doenças do aparelho respiratório (Cap. X), Doenças do Aparelho digestivo (Cap. XI), Outras Causas e como covariável a proporção de óbitos por causas mal definidas, desagregadas por espaços regionais.

Em formalidade matemática, sejam  $Y_{ij}$  a variável que representa a proporção de óbitos

da  $i$ -ésima microrregião (ou município) atribuído ao  $j$ -ésimo grupo de causas da CID-10 e  $X_i$  a proporção de óbitos por causas mal definidas correspondente a  $i$ -ésima região. A  $j$ -ésima equação de regressão linear de  $X_i$  sobre  $Y_{ij}$  estimada é dada por:

$$Y_{ij} = \alpha_j + \beta_j X_i + e_{ij}$$

Em que  $\alpha_j$  e  $\beta_j$  são os coeficientes linear e angular, respectivamente, obtidos pelo método de mínimos quadrados ordinários (MORETTIN; BUSSAB, 2017). Pode ser demonstrado que  $\sum_{j=1}^r \alpha_j = 0$  e  $\sum_{j=1}^r \beta_j = -1$ , em que  $r$  é o número do grupo de causas. Em geral espera-se uma regressão linear inversa entre a proporção de óbitos pelas causas mal definidas e os grupos de causas bem definidas. Assim, é razoável que  $-1 < \beta_j < 0$ .

A equação de redistribuição dos óbitos mal definidos tem a seguinte expressão:

$$O_j = Y_j - 1 \cdot (X \cdot \beta_j)$$

Em que:

$O_j$  são os óbitos redistribuídos da causa  $j$ ;

$Y_j$  são os óbitos observados da causa  $j$ ;

$\beta_j$  é o fator de redistribuição da causa  $j$ ;

$X$  são os óbitos da causa mal definida.

Uma vez que as causas mal definidas foram redistribuídas pelos grupos de causas definidas, nos espaços regionalizados, os mesmos fatores de redistribuição  $\beta$  admitiu-se que eles foram mantidos constantes para todos os grupos etários entre os centenários. Para o uso desse procedimento, foram consideradas as Unidades de Federação – UF (que compõem o Semiárido brasileiro) como unidades geográficas e classificadas de acordo com os três espaços regionalizados (Urbano, Intermediário e Rural) para efeito da aplicação da regressão proposta por Ledermann. Foram utilizados os dados referentes aos anos de 2010 a 2020.

### 3.5 Associação das causas mal definidas com as variáveis da DO

Devido a importância da problemática dos óbitos mal definidos no contexto dos centenários, buscou-se explorar ainda mais essa questão, ao procurar compreender a existência

de uma associação das causas mal definidas com fatores sociodemográficos dos óbitos identificados na DO para o Semiárido brasileiro. Para tanto, optou-se por aplicar uma regressão logística binária para o triênio de 2018, 2019 e 2020.

Considerou-se uma modelagem dicotomizada tendo como variável dependente o registro de óbito por causa mal definida. Para tanto, foram divididos dois grupos: 0 (registros contendo a causa básica de óbito por outros capítulos, que não fossem o Capítulo XVIII) e 1 (registros contendo a causa básica de óbito pelo Capítulo XVIII, ou seja, pela causa mal definida). As variáveis independentes, categorizadas, são as descritas no Quadro 3.1, oriundas da Declaração de Óbito.

### 3.5.1 O Modelo de Regressão Logística

O Modelo de Regressão Logística (MRL) tem se apresentado enquanto alternativa para investigação de desfechos com respostas binárias por meio de um conjunto de variáveis independentes. Os avanços do seu uso são pertinentes devido aos avanços teóricos e científicos em relação à sua otimização e ao processamento computacional (BARRETO, 2011; CORDEIRO, LIMA NETO, 2004).

Para investigação de casos cuja variável resposta poderá assumir apenas dois valores, “0” ou “1”, supõe-se que o modelo tem a seguinte representação:

$$y_i = x_i^T \beta + \varepsilon_i$$

No qual,  $x_i^T = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$ ,  $\beta^T = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$  e a variável dependente  $y_i$  assume os valores 0 ou 1. Assume-se que a variável dependente  $y_i$  é uma variável aleatória de Bernoulli, sendo  $\pi_i$  a probabilidade do valor  $y_i = 1$ , seguindo a distribuição (AGRESTI, 2007):

| $y_i$ | Probabilidade            |
|-------|--------------------------|
| 1     | $P(y_i = 1) = \pi_i$     |
| 0     | $P(y_i = 0) = 1 - \pi_i$ |

Sendo  $E(x_i) = 0$ , o valor esperado de  $y_i$  é:

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

Logo:

$$E(y_i) = x_i^T \beta = \pi_i$$

Isso implica que o resultado esperado dado pela função resposta  $E(y_i) = x_i^T \beta$  corresponde à probabilidade da variável resposta assumir o valor 1 (MONTGOMERY; PECK; VINING, 2001). No entanto, identificam-se algumas particularidades com o modelo de regressão  $y_i = x_i^T \beta + \varepsilon_i$ . Primeiramente, observa-se que se a resposta é binária, o erro  $\varepsilon_i$  poderá assumir apenas dois valores:

$$\varepsilon_i = 1 - x_i^T \beta, \text{ quando } y_i = 1$$

$$\varepsilon_i = -x_i^T \beta, \text{ quando } y_i = 0$$

Assim, os erros neste modelo não apresentarão a possibilidade de seguir a distribuição normal de probabilidade e a variância destes não será constante, visto que:

$$\begin{aligned} \sigma_{y_i}^2 &= E\{y_i - E(y_i)\}^2 \\ &= (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i) \\ &= \pi_i(1 - \pi_i) \end{aligned}$$

Essa expressão pode ser apresentada por:

$$\sigma_{y_i}^2 = E(y_i)[1 - E(y_i)]$$

Se  $E(y_i) = x_i^T \beta = \pi_i$ , a variância das observações (que é a mesma variância dos erros, já que  $\varepsilon_i = y_i - \pi_i$  e  $\pi_i$  é uma constante) é função da média (HOSMER, LEMESHOW, 2000). Por fim, há uma restrição da função resposta, já que

$$0 \leq E(y_i) = \pi_i \leq 1$$

Esse pressuposto repercute na escolha da função linear, já que, neste caso, o modelo poderia ser ajustado com valores preditivos para função resposta fora do intervalo entre 0 e 1. Em geral, quando a variável resposta é binária, há evidências empíricas de que a forma da função resposta é não linear. Nessa situação, faz-se necessária uma transformação na expressão, que passa a ser denominada de função resposta logística (MONTGOMERY; PECK; VINING, 2001) sob a forma:

$$E(y) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}$$

Ou equivalente:

$$E(y) = \frac{1}{1 + \exp(-x^T \beta)}$$

Esta transformação se justifica pelo fato desta função apresentar muitas propriedades desejáveis de um modelo de regressão linear. A função logística é linear em seus parâmetros, podendo ser contínua e variar entre  $-\infty$  e  $+\infty$ , dependendo da propriedade de  $x$  (AGRESTI, 2007).

O modelo logístico viabiliza o ajuste dos dados de um problema a partir de um modelo, favorecendo a sua análise e interpretação. As variáveis explicativas podem ser tanto de natureza categórica quanto variáveis contínuas. A equação logística é definida por:

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i = \eta_i + \varepsilon_i$$

onde:

$\text{logit}(\pi_i)$  = logaritmo da razão de chances entre  $p(y=1)$  e  $p(y=0)$

$\pi_i = p(y_i = 1)$

$\beta_0$  e  $\beta_j$  = parâmetros a serem estimados

$x_{ij}$  = variáveis independentes,  $j = 1, \dots, k$ .

$\varepsilon_i$  = Erro aleatório

Modelos lineares generalizados incluem uma função de ligação que associa os valores esperados da resposta aos preditores lineares no modelo. A função *logit* utilizada neste trabalho,

tem funções de distribuições específicas para a realização do cálculo da probabilidade, que é a logística. Os parâmetros dessa função de ligação são estimados de forma iterativa pelo método da máxima verossimilhança, pois é uma transformação da distribuição acumulada (FREITAS et al, 2013).

### 3.5.2 Razão de Chances (Odds Ratio)

Cada coeficiente estimado da regressão logística fornece uma estimativa na qual o seu logaritmo natural ( $\ln$ ) resulta em uma medida de associação, que é a *Odds Ratio* (OR), ou razão de chances, para todas as variáveis do modelo. A OR é a razão entre as chances e a chance é a probabilidade de que um evento ocorra dividido pela probabilidade de que ele não ocorra (BLAND; ALTMAN, 2000). Ela é obtida pela comparação de indivíduos que diferem apenas na variável de interesse e que apresentam as outras características constantes (HOSMER; LEMESHOW, 2000).

Considerando inicialmente o caso onde o preditor linear tem apenas um único regressor, o valor do modelo ajustado para um particular valor de  $x_i$  é:

$$\eta(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

O valor preditivo para  $x_i + 1$  é:

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1 (x_i + 1)$$

E a diferença entre os dois valores preditivos é:

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Agora  $\hat{\eta}(x_i)$  é apenas o log-odds quando a variável regressora é igual a  $x_i$ , e  $\hat{\eta}(x_i + 1)$  é apenas a log-odds quando o regressor é igual a  $x_i + 1$ . Então, a diferença entre dois valores preditivos é:

$$\begin{aligned} \hat{\eta}(x_i + 1) - \hat{\eta}(x_i) &= \ln(odds_{x_{i+1}}) - \ln(odds_{x_i}) \\ &= \ln \left[ \frac{odds_{x_{i+1}}}{odds_{x_i}} \right] \\ &= \hat{\beta}_1 \end{aligned}$$

Se forem aplicados os antilogs, será obtido a *Odds Ratio*:

$$O_R = \frac{odds_{xi+1}}{odds_{xi}} = e^{\beta_j}$$

A interpretação da *Odds Ratio* é dada como o aumento/decrécimo esperado na probabilidade de sucesso devido a uma mudança de uma unidade no valor da variável preditora. No caso em que a variável resposta é nominal, a interpretação pode ser feita em termos de percentagem. A interpretação dos coeficientes de regressão para o caso de um modelo de regressão logística múltiplo é similar ao caso em que o modelo tem apenas um único regressor. Em tais casos, a quantidade  $\exp(\hat{\beta}_j)$  é a *Odds Ratio* por regressor  $x_j$ , assumindo que todas as outras variáveis preditivas são constantes (MONTCOMERY; PECK; VINING, 2001).

### 3.5.3 Análise ROC

A curva ROC é uma técnica gráfica utilizada para avaliar a capacidade de um teste ou modelo fazer a distinção entre duas classes. Permite fazer análises visuais entre sensibilidade e especificidade relativamente a diversos pontos de corte (MADEIRA et al, 2008). A curva é obtida através do cálculo da Sensibilidade e da Especificidade para cada ponto de corte e representam-se graficamente os pontos de coordenadas (1-Especificidade, Sensibilidade). Por convenção, 1-Especificidade é representada no eixo das abcissas e a Sensibilidade é representada no eixo das ordenadas, variando ambas de 0 a 1 (0-100%). A curva ROC no plano unitário resulta da representação gráfica dos índices de precisão 1-Especificidade versus Sensibilidade, oriundas da variação do ponto de corte. Assim, a curva ROC é uma descrição empírica da capacidade e, consequentemente, da qualidade de um teste de diagnóstico diferenciar duas classes num universo (POLO; MIOT, 2020).

Quanto maior for a dimensão da amostra mais pontos de corte existirão para a representação da curva e, consequentemente, mais precisa será a curva ROC representada. Os valores de x equivalem ao valor de 1-Especificidade e os valores de y ao valor da Sensibilidade. Conclui-se, assim, que cada ponto de corte dará origem a um ponto (1-E; S) para a construção da curva ROC. Por último, marcam-se todos os pontos (1-E; S) no gráfico, os quais unem-se por retas e obtém-se a curva ROC. Daí a importância de recolher o maior número de observações possíveis, para que a separação entre pontos seja menos notória (MADEIRA et al,

2008). Neste estudo procedeu-se à construção da curva ROC para definir a adequabilidade do Modelo de Regressão Logística (MRL) utilizado.

#### 3.5.4 Análise de resíduos

Uma vez que o modelo tenha sido ajustado, realiza-se a análise de resíduos. Estes resíduos consistem na diferença entre os valores previstos pelo modelo e os valores observados na amostra, e representam o erro amostral que está presente no modelo. “Se o modelo se ajusta bem aos dados da amostra, todos os resíduos devem ser pequenos, se o modelo aderir perfeitamente aos dados, todos os pontos estarão sobre a linha de regressão e todos os resíduos serão iguais a zero” (FIELD, 2009, p. 382).

Para esta pesquisa foi utilizada a análise gráfica dos resíduos padronizados, com o diagrama de dispersão entre o resíduo e o predito, esperando que eles estejam concentrados em pelo menos 95% entre -2 e 2 e pouquíssimos pontos acima de 3 ou abaixo de -3 (HAIR et al, 2017).

Ainda na análise de resíduos, foi realizado o teste de Durbin-Watson a fim de detectar a presença de autocorrelação (dependência) nos resíduos. Esse teste é baseado na suposição de que os erros no modelo de regressão são gerados por um processo autorregressivo de primeira ordem, de acordo com:

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

em que  $\varepsilon_i$  é o termo do erro do modelo na  $i$ -ésima observação. A partir disso, tem-se a estatística de teste de Durbin Watson, que é dada por:

$$dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Recomenda-se que o resultado desse teste seja entre  $0 \leq dw \leq 4$ , indicando que os resíduos do modelo são independentes (HAIR, 2017).

No intuito de corroborar todos esses resultados, foi calculada a Distância de Cook ( $DC_j$ ), que consiste na estimação de uma medida análoga de uma determinada observação  $j$  na estimação dos coeficientes do modelo. Essa medida indica a variação dos resíduos quando a observação  $j$  é eliminada do ajustamento do modelo. Usando a *leverage* e os resíduos de Pearson

estandardizados (*Standardized Pearson Residual*) , essa medida pode ser calculada como (PREGIBON, 1981):

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

Na qual  $r_j^2$  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. Recomenda-se que os valores que estejam acima ou próximo de 1 sejam eliminados da amostra (FÁVERO *et al.*, 2009).

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Análise da qualidade dos dados populacionais dos centenários do Semiárido brasileiro

#### 4.1.1 Panorama populacional dos centenários residentes no Semiárido brasileiro, Nordeste e Brasil.

A tabela 4.1.1 descreve o panorama da população recenseada em 2000, 2010 e a população projetada para 2019, total e dos centenários, residentes no Nordeste, Brasil e nos espaços regionalizados do Semiárido brasileiro, classificados de acordo com a nova tipologia municipal (IBGE, 2017), caracterizados em região Urbana, Intermediário e Rural, desagregadas por sexo.

Tabela 4.1.1 - Descrição da população total e centenária, entre os espaços regionalizados do Semiárido brasileiro, Nordeste e Brasil, 2000, 2010 e 2019.

| Semnando Brasília, Nordeste e Brasil, 2000, 2010 e 2019. |             |                  |       |             |                  |       |
|----------------------------------------------------------|-------------|------------------|-------|-------------|------------------|-------|
| Regiões                                                  | Homens      |                  |       | Mulheres    |                  |       |
|                                                          | Total       | 100 anos ou mais |       | Total       | 100 anos ou mais |       |
|                                                          |             | N                | %     |             | N                | %     |
| 2000                                                     |             |                  |       |             |                  |       |
| Brasil                                                   | 83.576.015  | 10.423           | -     | 86.223.155  | 14.153           | -     |
| Nordeste                                                 | 23.413.914  | 3.385            | -     | 24.327.797  | 4.653            | -     |
| Semiárido                                                | 11.888.738  | 1.841            | 100,0 | 12.123.397  | 2.550            | 100,0 |
| Urbano                                                   | 4.439.539   | 618              | 33,6  | 4.713.411   | 864              | 33,9  |
| Intermediário                                            | 1.826.916   | 275              | 14,9  | 1.856.208   | 411              | 16,1  |
| Rural                                                    | 5.622.283   | 948              | 51,5  | 5.553.778   | 1.275            | 50,0  |
| 2010                                                     |             |                  |       |             |                  |       |
| Brasil                                                   | 93.406.990  | 7.247            | -     | 97.348.809  | 16.989           | -     |
| Nordeste                                                 | 25.909.046  | 3.200            | -     | 27.172.904  | 7.208            | -     |
| Semiárido                                                | 12.865.008  | 1.781            | 100,0 | 13.196.156  | 3.521            | 100,0 |
| Urbano                                                   | 5.029.607   | 608              | 34,1  | 5.368.100   | 1.287            | 36,5  |
| Intermediário                                            | 1.933.171   | 288              | 16,2  | 1.980.268   | 569              | 16,2  |
| Rural                                                    | 5.902.230   | 885              | 49,7  | 5.847.788   | 1.665            | 47,3  |
| 2019*                                                    |             |                  |       |             |                  |       |
| Brasil                                                   | 103.258.439 | 5.237            | -     | 108.607.451 | 20.034           | -     |
| Nordeste                                                 | 28.385.472  | 3.042            | -     | 30.022.260  | 10.715           | -     |
| Semiárido                                                | 13.895.758  | 1.729            | 100,0 | 14.335.508  | 4.714            | 100,0 |
| Urbano                                                   | 5.419.346   | 588              | 34,0  | 5.877.558   | 1.744            | 37,0  |
| Intermediário                                            | 2.084.364   | 277              | 16,0  | 2.150.326   | 754              | 16,0  |
| Rural                                                    | 6.392.048   | 864              | 50,0  | 6.307.624   | 2.216            | 53,0  |

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE, 2000, 2010 e 2019 (projeção).

Nota: \* Dados populacionais projetados para 2019.

O IBGE contabilizou no Censo Demográfico 2010, último censo brasileiro, uma população de 190.755.799 habitantes no Brasil, sendo 53.081.950 localizados na Região Nordeste e 26.061.164 residentes na região do Semiárido brasileiro. Em 2010, foram recenseados 24.236 centenários no Brasil e quase metade desse contingente populacional residia na região Nordeste.

A projeção populacional para 2019, estimou um total de 211.865.890 habitantes para o Brasil, sendo 58.407.732 para a região Nordeste, que representou quase 30% da população brasileira (Tabela 4.1.1). O Nordeste é a segunda região mais populosa do país e a terceira em densidade demográfica, contando com cerca de 38 habitantes por quilômetro quadrado. Para o Semiárido brasileiro, em 2019, estimou-se uma população de 24.012.135 habitantes, que representou quase 15% da população brasileira, sendo destes a maioria mulheres e residentes em municípios caracterizados como Rurais.

Nos anos estudados, o conjunto de municípios caracterizado como Rural concentrou cerca de 50% da população total do Semiárido brasileiro, assim como no estrato dos centenários, entre os sexos. Esse panorama é influenciado pelo fato de quase 80% dos municípios do Semiárido brasileiro ser caracterizado como Rural. Não obstante, o contingente populacional da zona Urbana merece destaque, pois os municípios predominantemente urbanos equivalem apenas a 11% do Semiárido brasileiro. A Tabela 4.1.1 mostra que tanto na população total como entre os centenários, o número de pessoas recenseadas na região urbana era bastante elevado e quase se assemelha ao número absoluto de pessoas residentes na região Rural.

Já em termos de diferença entre os sexos, ao longo de 19 anos (2000-2019), a população centenária absoluta masculina diminuiu no Nordeste e no Semiárido, sendo que no Brasil essa diminuição foi quase pela metade, em contrapartida a população centenária feminina quase que dobrou no Semiárido e no Brasil, já na região Nordeste aumentou mais de 100%,

Seguindo uma tendência nacional, a análise da proporção de centenários entre os homens vem se mantendo estável ao longo dos anos. No entanto, entre as mulheres, a proporção de centenárias aumentou no Semiárido e no Nordeste, já no Brasil observou-se uma estabilidade na proporcionalidade. Assim, fica muito claro, a evolução do processo de envelhecimento da população semiárida e nordestina, principalmente no estrato feminino.

Embora não haja uma definição universal para o termo rural, conforme Steinberger (2006), ele pode ser entendido como uma fração do espaço, assim como o urbano, representando uma forma ou conteúdo particular. Para Neuman (2000), as áreas ou regiões

rurais constituem um sistema de cidades e vilas circundadas por áreas abertas, onde a infraestrutura de transporte, utilidades e comunicação são reduzidas. No Brasil, a discussão de rural e ruralidades desenvolvida por Galizoni (2019) revela que os lugares rurais têm assumido diferentes perfis, como reflexo de formas de ocupação de um vasto território, regidas por diversas peculiaridades ambientais e socioculturais. Apesar de grandes dificuldades enfrentadas nessas áreas, os números mostram que o processo de envelhecimento tem acontecido no âmbito rural, mais da metade dos centenários do Semiárido brasileiro estão vivenciando a ruralidade, o que aumenta a importância de estudos voltados para essas áreas específicas.

O fenômeno conhecido como feminização da velhice conforme mostrado na Tabela 4.1.1 e já descrito anteriormente nesse trabalho, também é evidenciado em diversas outras localidades mundiais, pois a maioria da população de centenários no mundo são mulheres (CEVENINI et al., 2008; VACANTE et al., 2012). Sugere-se que a relação centenária entre mulheres e homens provavelmente está relacionada com fatores ambientais e sociais, que têm um efeito diferente nos dois gêneros (FRANCESCHI et al., 2017). As mulheres têm uma esperança média de vida mais longa que os homens (cerca de 6 anos a mais em países desenvolvidos) e uma taxa de mortalidade mais baixa em todas as idades (TEIXEIRA et al., 2017). Segundo Rochon et al. (2014), atrelado ao evidente predomínio feminino entre os centenários, as mulheres apresentam maior probabilidade de serem portadoras de múltiplas síndromes geriátricas e maior prevalência de doenças musculoesqueléticas e déficits sensoriais. Os autores supracitados afirmam que a tendência esperada é a diminuição do número de homens e aumento das mulheres centenárias ao longo dos anos, o que corrobora com dados encontrados na Tabela 4.1.1.

Para a construção da projeção populacional do ano 2019, foram implementados os métodos de projeção Aritmético, Geométrico e Exponencial, com base nos censos de 2000 e 2010, conforme apresentado na Tabela 4.1.2. Após proceder o cálculo a partir de cada método, para cada região e para o Brasil, entre os sexos, foi calculada a média dos valores encontrados nos três métodos e, posteriormente, incluída nos dados projetados para 2019 apresentados na Tabela 4.1.1.

Tabela 4.1.2 - Projeção da população total e centenários, de acordo com os métodos aritmético, geométrico e exponencial, para o Semiárido brasileiro, Nordeste e Brasil, 2019.

| Regiões          | Método de projeção | Homem       |             | Mulheres    |             |
|------------------|--------------------|-------------|-------------|-------------|-------------|
|                  |                    | Total       | Centenários | Total       | Centenários |
| <b>Semiárido</b> | Aritmético         | 13.815.803  | 1.729       | 14.247.071  | 4.728       |
|                  | Geométrico         | 14.059.460  | 1.729       | 14.516.885  | 4.707       |
|                  | Exponencial        | 13.812.012  | 1.729       | 14.242.566  | 4.707       |
| <b>Nordeste</b>  | Aritmético         | 28.393.972  | 3.043       | 30.032.959  | 10.770      |
|                  | Geométrico         | 28.381.222  | 3.042       | 30.016.911  | 10.688      |
|                  | Exponencial        | 28.381.222  | 3.042       | 30.016.911  | 10.688      |
| <b>Brasil</b>    | Aritmético         | 103.295.636 | 5.260       | 108.653.919 | 20.053      |
|                  | Geométrico         | 103.239.840 | 5.225       | 108.584.216 | 20.024      |
|                  | Exponencial        | 103.239.840 | 5.225       | 108.584.216 | 20.024      |

Fonte: Elaboração própria

Observando os valores encontrados nos cálculos das projeções é possível identificar que não houve grande diferença nos valores entre os métodos implementados. Uma das principais razões para isso é o intervalo de tempo relativamente pequeno entre o período da projeção, o que leva as projeções dos métodos geométrico e exponencial ficarem com praticamente os mesmos valores. Segundo Paes (2018), embora existam diversas fórmulas de crescimento possíveis de serem usadas, ademais do método demográfico dos componentes, quatro modelos matemáticos são os mais efetivos e simples para projetar populações humanas: Linear (aritmético), geométrico, exponencial e logístico. Uma razão é que as curvas associadas com esses modelos são suaves, particularmente, os três primeiros, os quais são apresentados na Tabela 4.1.2, pois eles assumem que o crescimento populacional se dá de uma maneira mais regular e não de forma brusca ou com picos acentuados.

#### 4.1.2 Proposta de classificação para avaliar a qualidade dos dados censitários dos centenários.

Embora tenha-se como propósito principal a investigação da qualidade da contagem censitária dos centenários e como não há na literatura nenhum método de avaliação específico para este grupo etário, partiu-se do entendimento de que a partir de referências de países com qualidade reconhecida de contagem dos centenários seria possível propor algum método para avaliar a contagem das regiões menos desenvolvidas, como é o caso do Semiárido brasileiro. Desse modo, a proposta metodológica que se segue consistiu na elaboração de uma classificação que permita avaliar a qualidade da contagem censitária dos idosos centenários, a partir de determinados indicadores para contingentes populacionais regionais do mundo. A

metodologia completa da referida proposta já foi descrita na Sessão 3.3.1, que utilizou três indicadores referentes aos centenários: 1) Razão de Sexo (RS); 2) Razão da população masculina com 100 anos ou mais com relação aos idosos homens com 85 anos ou mais, multiplicada por 100; 3) Razão da população feminina com 100 anos ou mais com relação às idosas com 85 anos ou mais, multiplicada por 100. Esses indicadores foram construídos para 85 países selecionados de acordo com os critérios estabelecidos na Sessão 3.3.1.

Os resultados da análise descritiva dos dados referentes aos três indicadores dos centenários, além dos valores do IDH e da expectativa de vida ao nascer, segundo a UNDP (2018), dos 85 países incluídos neste estudo, estão apresentados na Tabela 4.1.3

Tabela 4.1.3 - Média, desvio-padrão, valor mínimo e valor máximo dos indicadores selecionados. Índice de Desenvolvimento Humano (IDH) e expectativa de vida ao nascer, dos 85 países selecionados.

| <b>Medidas</b> | <b>RS</b> | <b>Cent./85+( H)</b> | <b>Cent./85+(M)</b> | <b>IDH</b> | <b>Exp. de vida</b> |
|----------------|-----------|----------------------|---------------------|------------|---------------------|
| Média          | 0,372     | 1,371                | 1,905               | 0,808      | 76,5                |
| Desvio Padrão  | 0,235     | 1,867                | 2,186               | 0,110      | 5,5                 |
| Valor Mínimo   | 0,029     | 0,044                | 0,123               | 0,411      | 58,0                |
| Valor Máximo   | 1,103     | 7,607                | 9,514               | 0,954      | 86,0                |

Fonte: UNSD, 2020.

Nota: RS: Razão de Sexo H: Homens M:Mulheres

De acordo com os dados descritivos apresentados na Tabela 4.1.3, foi possível compreender a variação existente entre os valores mínimos e máximos para os três indicadores construídos, para o IDH e para a expectativa de vida, o que permite a possibilidade da segregação de valores em grupos e à construção de intervalos.

Neste estudo, para construir a proposta de classificação da qualidade dos dados populacionais dos centenários, utilizou-se como ferramenta de classificação os três primeiros indicadores apresentados na Tabela 4.1.3.

A Razão de Sexo é um indicador que expressa uma baixa qualidade na contagem da população centenária quando está elevada. A razão varia em função da idade, porém em idades avançadas, o decréscimo deste indicador é consequência da maior mortalidade masculina em relação à feminina (CAMARANO; KANSO, 2016). A feminização da velhice já é um fenômeno retratado na literatura (CHAIMOVZC, 2016; CAMARANO; KANSO, 2016). Devido às diferenças da mortalidade por sexo ao longo da vida, a proporção de mulheres na população aumenta com a idade. Isso é especialmente notável nas idades mais elevadas, em que a proporção de mulheres aumenta acentuadamente, o que foi também apresentado nos resultados deste estudo conforme descrito na Tabela 4.1.1.

Estudo realizado por Meyer (2012) evidenciou que em países desenvolvidos a razão de sexo entre centenários é sabidamente inferior 0,25. Por exemplo, nos EUA, em 2010, havia entre os centenários apenas 20,7 homens a cada 100 mulheres, desta forma o indicador da provável baixa qualidade na contagem da população centenária é a elevada razão de sexo, muito comum em países em desenvolvimento.

A razão da população com 100 anos ou mais e 85 anos ou mais pode indicar possíveis problemas de falta de confiabilidade nos dados recenseados para a população de centenários. Segundo Reis e Turra (2016), essa razão não ultrapassa o limite de 1% nos países com referência de boa qualidade dos dados. São poucas as situações que ultrapassam esse limite, nos países desenvolvidos, como no caso da proporção de mulheres do Canadá, EUA, França e Japão para o ano de 2010, devido ao fenômeno já descrito denominado de feminização da velhice. De acordo com a Tabela 4.1.3, a amplitude desse indicador para os homens é de 0,044 a 7,607 e de 0,123 a 9,514 para as mulheres, indicando uma grande variabilidade dos valores, apontando a existência de países com níveis distintos de qualidade, variando de uma baixa a alta qualidade dos dados, refletidos pelos respectivos desvios padrões.

Os contrastes regionais podem ser evidenciados pelas magnitudes diferenciadas do IDH baixo a muito alto (de 0,411 a 0,954) e expectativa de vida ao nascer com 28 anos de diferença (58 anos a 86 anos), como foi apresentado na Tabela 4.1.3.

Para compreender a relação desses indicadores selecionados com a qualidade dos dados e registros dos países, procedeu-se a análise de regressão linear simples de cada indicador com o IDH dos países estudados, cujos principais resultados são descritos na Tabela 4.1.4.

Tabela 4.1.4 - Coeficiente de Determinação e indicadores da regressão linear simples entre o Índice de Desenvolvimento Humano (IDH) e os indicadores construídos a partir dos dados censitários dos centenários dos 85 países selecionados.

| <b>Indicadores dos centenários</b>     | <b>R<sup>2</sup></b> | <b>β</b> | <b>IC (95%)</b>  | <b>Valor de p</b> |
|----------------------------------------|----------------------|----------|------------------|-------------------|
| Razão de Sexo                          | 0,310                | -0,557   | (- 1,58; - 0,80) | <0,001            |
| Centenários/85 anos ou mais (homens)   | 0,432                | -0,657   | (-13,98; - 8,39) | <0,001            |
| Centenários/85 anos ou mais (mulheres) | 0,502                | -0,709   | (-17,19; -11,05) | <0,001            |

Fonte: Elaboração própria.

Legenda: R<sup>2</sup>: Coeficiente de determinação; β: Coeficiente de correlação; IC95%:Intervalo de 95% de confiança.

Na Tabela 4.1.4 é possível observar, embora não tão elevada, que houve uma correlação significativa entre o IDH e os três indicadores selecionados para a proposta de classificação: RS (R<sup>2</sup> = 0,310; p < 0,001); razão da população com 100 anos ou mais com relação aos idosos com 85 anos ou mais de idade: homens (R<sup>2</sup> = 0,432; p < 0,001); e mulheres (R<sup>2</sup> = 0,502; p <

0,001). Portanto, quanto mais elevado for o IDH, menores foram os valores dos indicadores selecionados, conforme sinaliza o parâmetro  $\beta$ . Relações estas que também foram descritas em outros estudos (CHUANWAN et al, 2012; TERBLANCHE; WILSON, 2015; REIS; TURRA 2016).

Composto por três dimensões - uma vida longa e saudável mensurada pela expectativa de vida ao nascer; aptidão de lograr conhecimento medida pela média de anos de estudo; possibilidade de atingir um padrão de vida digno, medido pelo rendimento nacional bruto per capita - foi elaborado o Índice de Desenvolvimento Humano (IDH) (MARTINS; PONTES; HIGA, 2018). Diante disso, através do IDH permite-se uma abordagem mais global em relação a outras abordagens de caráter além da riqueza, principalmente no que se refere à qualidade dos dados e informações fornecidas pelos países e o aporte técnico/logístico para a realização dos censos demográficos. Estudos anteriores já utilizavam o IDH como variável para testar indicadores de qualidade de dados (PAES; MAIOR; MELO; SILVA, 2016; PEREIRA; CLEMENTE, 2014; THUM; BALDISSEROTTO; CELESTE, 2019).

A fim de proceder à proposta de classificação, a partir da construção de intervalos de valores dos três indicadores selecionados, realizou-se primeiro a análise de agrupamento entre os 85 países selecionados, através das cinco variáveis (os três indicadores dos centenários, IDH e expectativa de vida ao nascer), na qual foram encontrados quatro *clusters*, utilizando a distância euclidiana ao quadrado, conforme indicado na Tabela 4.1.5.

Tabela 4.1.5 - Agrupamento dos 85 países segundo os três indicadores selecionados, IDH e expectativa de vida ao nascer nos quatro *clusters* estimados pela análise de cluster multivariada.

(Continua)

| Cluster 1      | Cluster 2 e 3    | Cluster 4       |
|----------------|------------------|-----------------|
| Noruega        | República Tcheca | África do Sul   |
| Suíça          | Estônia          | Marrocos        |
| Irlanda        | Polônia          | Nicarágua       |
| Alemanha       | Lituânia         | Índia           |
| Islândia       | Eslováquia       | Butão           |
| Austrália      | Letônia          | Ruanda          |
| Suécia         | Chile            | Costa do Marfim |
| Cingapura      | Hungria          | Senegal         |
| Holanda        | Croácia          | Chade           |
| Dinamarca      | Omã              |                 |
| Finlândia      | Argentina        |                 |
| Canadá         | Federação Russa  |                 |
| Nova Zelândia  | Cazaquistão      |                 |
| Reino Unido    | Belarus          |                 |
| Estados Unidos | Romênia          |                 |

| Cluster 1     | Cluster 2 e 3        | Cluster 4 |
|---------------|----------------------|-----------|
| Bélgica       | Bulgária             |           |
| Japão         | Montenegro           |           |
| Áustria       | Uruguai              |           |
| Luxemburgo    | Turquia              |           |
| Coreia do Sul | Bahamas              |           |
| Israel        | Sérvia               |           |
| Eslovênia     | Irã                  |           |
| Espanha       | Maurício             |           |
| França        | Panamá               |           |
| Malta         | Costa Rica           |           |
| Itália        | Albânia              |           |
| Grécia        | Geórgia              |           |
| Portugal      | Cuba                 |           |
|               | México               |           |
|               | Tailândia            |           |
|               | Brasil               |           |
|               | Colômbia             |           |
|               | Argélia              |           |
|               | Macedônia            |           |
|               | Peru                 |           |
|               | Equador              |           |
|               | China                |           |
|               | Azerbaijão           |           |
|               | Ucrânia              |           |
|               | República Dominicana |           |
|               | Mongólia             |           |
|               | Venezuela            |           |
|               | Jamaica              |           |
|               | Filipinas            |           |
|               | Moldávia             |           |
|               | Samoa                |           |
|               | Quirguistão          |           |
|               | Tajiquistão          |           |

Fonte: Elaboração própria

Após proceder à análise de *clusters* foi possível identificar claramente a presença de dois grupos díspares, no tocante ao IDH e a expectativa de vida ao nascer dos países. O *cluster* 1 composto pelos países com IDH e expectativas de vida ao nascer elevados, definido como o grupo “padrão de elevada qualidade”; o *cluster* 4 composto pelos países com IDH e expectativas de vida ao nascer mais baixos do rol de países selecionados, definido como o grupo “padrão de reduzida qualidade” e os *clusters* 2 e 3, formados pelos grupos compostos por países com IDH e expectativas de vida ao nascer, definidos com valores situados entre os outros dois grupos polarizados, definidos como “padrão intermediário de qualidade”.

Após essa segregação dos grupos de países, selecionou-se os valores dos três indicadores, sendo possível a construção de intervalos, para cada indicador, considerados neste

estudo como “intervalo padrão”, os quais foram definidos a partir do valor mínimo e máximo dos indicadores encontrados em cada grupo de países da análise de *clusters* (Tabela 4.1.6).

Tabela 4.1.6 - Apresentação dos intervalos de referência para cada indicador selecionado, construídos a partir da segregação dos grupos, para os 85 países selecionados.

| Qualidade do Padrão           | RS                                     | Cent./85+ (H)                          | Cent./85+ (M)                          |
|-------------------------------|----------------------------------------|----------------------------------------|----------------------------------------|
| Elevado (Cluster 1)           | (0,138; 0,353)                         | (0,203; 0,762)                         | (0,536; 1,405)                         |
| Reduzido (Cluster 4)          | (0,588; 1,102)                         | (2,410; 7,607)                         | (2,952; 9,514)                         |
| Intermediário (Cluster 2 e 3) | Valores fora dos intervalos anteriores | Valores fora dos intervalos anteriores | Valores fora dos intervalos anteriores |

Fonte: Elaboração própria.

Para a construção dos intervalos de referência no *Cluster 1* foram excluídos os valores referentes aos seguintes países: Cingapura e Israel. Estes países apresentaram *outliers* com relação aos valores dos indicadores, dificultando a construção de intervalos. Portanto, os indicadores desses países foram excluídos dos intervalos, o que demandaria uma análise particularizada com relação aos dados censitários dos idosos centenários nesses países.

De acordo com a Tabela 4.1.6, para cada uma das três faixas de intervalo dos indicadores, aqui denominadas de “intervalo padrão” e para avaliar de forma sistemática e coerente a qualidade dos dados censitários dos centenários, foram atribuídos pesos:

- Intervalo padrão de “Elevada” qualidade- Peso 3
- Intervalo padrão de “Intermediária” qualidade- Peso 2
- Intervalo padrão de “Reduzida” qualidade- Peso 1

Assim, observando os intervalos de referência como padrão de avaliação da qualidade mostrados na Tabela 4.1.6 e com a definição dos pesos e a soma destes, apresentados no Quadro 3.3, da sessão 3.3.1, obteve-se a classificação final da qualidade dos dados censitários dos idosos centenários, para os 85 países selecionados, os quais são mostrados na Tabela 4.1.7.

Tabela 4.1.7 - Classificação da qualidade dos dados censitários dos idosos centenários, a partir dos três indicadores propostos, de acordo com os pesos dos três intervalos padrão e os cinco escores definidos, para os 85 países selecionados, 2021.

(Continua)

| Países    | RS (Peso) | Cent./85+ (H) (Peso) | Cent./85+ (M) (Peso) | Classif. Final* |
|-----------|-----------|----------------------|----------------------|-----------------|
| Noruega   | 3         | 3                    | 3                    | Muito boa       |
| Suíça     | 3         | 3                    | 3                    | Muito boa       |
| Irlanda   | 3         | 3                    | 3                    | Muito boa       |
| Alemanha  | 3         | 3                    | 3                    | Muito boa       |
| Islândia  | 3         | 3                    | 3                    | Muito boa       |
| Austrália | 3         | 3                    | 3                    | Muito boa       |

(Continuação)

| <b>Países</b>  | <b>RS (Peso)</b> | <b>Cent./85+ (H) (Peso)</b> | <b>Cent./85+ (M) (Peso)</b> | <b>Classif. Final*</b> |
|----------------|------------------|-----------------------------|-----------------------------|------------------------|
| Suécia         | 3                | 3                           | 3                           | Muito boa              |
| Cingapura      | 2                | 1                           | 2                           | Baixa                  |
| Holanda        | 3                | 3                           | 3                           | Muito boa              |
| Dinamarca      | 3                | 3                           | 3                           | Muito boa              |
| Finlândia      | 3                | 3                           | 3                           | Muito boa              |
| Canadá         | 3                | 3                           | 3                           | Muito boa              |
| Nova Zelândia  | 3                | 3                           | 3                           | Muito boa              |
| Reino Unido    | 3                | 3                           | 3                           | Muito boa              |
| Estados Unidos | 3                | 3                           | 3                           | Muito boa              |
| Bélgica        | 3                | 3                           | 3                           | Muito boa              |
| Japão          | 3                | 3                           | 3                           | Muito boa              |
| Áustria        | 3                | 3                           | 3                           | Muito boa              |
| Luxemburgo     | 3                | 3                           | 3                           | Muito boa              |
| Coreia do Sul  | 3                | 3                           | 3                           | Muito boa              |
| Israel         | 1                | 2                           | 2                           | Baixa                  |
| Eslovênia      | 3                | 3                           | 3                           | Muito boa              |
| Espanha        | 3                | 3                           | 3                           | Muito boa              |
| França         | 3                | 3                           | 3                           | Muito boa              |
| Rep. Tcheca    | 3                | 3                           | 2                           | Boa                    |
| Malta          | 3                | 3                           | 3                           | Muito boa              |
| Itália         | 3                | 3                           | 3                           | Muito boa              |
| Estônia        | 3                | 3                           | 3                           | Muito boa              |
| Grécia         | 3                | 3                           | 3                           | Muito boa              |
| Polônia        | 3                | 3                           | 3                           | Muito boa              |
| Lituânia       | 2                | 2                           | 3                           | Boa                    |
| Eslováquia     | 3                | 3                           | 2                           | Boa                    |
| Letônia        | 3                | 3                           | 2                           | Boa                    |
| Portugal       | 3                | 3                           | 3                           | Muito boa              |
| Chile          | 1                | 2                           | 3                           | Regular                |
| Hungria        | 2                | 2                           | 3                           | Boa                    |
| Croácia        | 3                | 3                           | 2                           | Boa                    |
| Omã            | 1                | 1                           | 1                           | Muito baixa            |
| Argentina      | 3                | 3                           | 3                           | Muito boa              |
| Fed. Russa     | 3                | 3                           | 3                           | Muito boa              |
| Cazaquistão    | 2                | 3                           | 2                           | Boa                    |
| Belarus        | 3                | 3                           | 3                           | Muito boa              |
| Romênia        | 2                | 3                           | 2                           | Boa                    |
| Bulgária       | 2                | 3                           | 2                           | Boa                    |
| Montenegro     | 2                | 2                           | 3                           | Boa                    |
| Uruguai        | 3                | 3                           | 3                           | Muito boa              |
| Turquia        | 3                | 2                           | 3                           | Boa                    |
| Bahamas        | 3                | 2                           | 2                           | Boa                    |
| Sérvia         | 2                | 2                           | 2                           | Regular                |
| Irã            | 1                | 2                           | 2                           | Baixa                  |
| Maurício       | 3                | 3                           | 3                           | Muito boa              |
| Panamá         | 2                | 2                           | 2                           | Regular                |
| Costa Rica     | 1                | 2                           | 2                           | Baixa                  |
| Albânia        | 3                | 2                           | 2                           | Boa                    |

(Continuação)

| <b>Países</b>   | <b>RS (Peso)</b> | <b>Cent./85+ (H) (Peso)</b> | <b>Cent./85+ (M) (Peso)</b> | <b>Classif. Final*</b> |
|-----------------|------------------|-----------------------------|-----------------------------|------------------------|
| Geórgia         | 2                | 2                           | 3                           | Boa                    |
| Cuba            | 2                | 2                           | 2                           | Regular                |
| México          | 1                | 1                           | 2                           | Baixa                  |
| Tailândia       | 2                | 2                           | 2                           | Regular                |
| Brasil          | 2                | 2                           | 2                           | Regular                |
| Colômbia        | 1                | 2                           | 2                           | Baixa                  |
| Argélia         | 1                | 2                           | 3                           | Regular                |
| Macedônia       | 3                | 2                           | 2                           | Boa                    |
| Peru            | 2                | 2                           | 2                           | Regular                |
| Equador         | 2                | 2                           | 2                           | Regular                |
| China           | 2                | 3                           | 3                           | Boa                    |
| Azerbaijão      | 2                | 3                           | 1                           | Regular                |
| Ucrânia         | 3                | 3                           | 2                           | Boa                    |
| Rep. Dominicana | 1                | 1                           | 1                           | Muito baixa            |
| Mongólia        | 3                | 3                           | 2                           | Boa                    |
| Venezuela       | 3                | 2                           | 1                           | Regular                |
| Jamaica         | 3                | 2                           | 1                           | Regular                |
| Filipinas       | 2                | 2                           | 2                           | Regular                |
| Moldávia        | 1                | 1                           | 3                           | Baixa                  |
| Samoa           | 1                | 2                           | 2                           | Baixa                  |
| África do Sul   | 1                | 1                           | 1                           | Muito baixa            |
| Marrocos        | 1                | 1                           | 1                           | Muito baixa            |
| Quirguistão     | 2                | 2                           | 1                           | Baixa                  |
| Tajiquistão     | 2                | 2                           | 1                           | Baixa                  |
| Nicarágua       | 1                | 1                           | 1                           | Muito baixa            |
| Índia           | 1                | 1                           | 1                           | Muito baixa            |
| Butão           | 1                | 1                           | 1                           | Muito baixa            |
| Ruanda          | 1                | 1                           | 1                           | Muito baixa            |
| Costa do Marfim | 1                | 1                           | 1                           | Muito baixa            |
| Senegal         | 1                | 1                           | 1                           | Muito baixa            |
| Chade           | 1                | 1                           | 1                           | Muito baixa            |

Fonte: Elaboração própria. \*Classificação Final definida de acordo com a soma dos pesos.

De acordo com a Tabela 4.1.7 pode-se identificar que os países estão dispostos de acordo com o nível de desenvolvimento, do mais desenvolvido para o menos desenvolvido. Ou seja, países com maiores níveis de desenvolvimento foram classificados no escore de qualidade “muito boa”, exceto Israel e Cingapura. Em contraponto, a maioria dos países com menores níveis de desenvolvimento, foram classificados no escore de qualidade “muito baixa”. Essa classificação sinaliza o quanto a qualidade dos dados e registros, sejam eles censitários ou não, são sensíveis e influenciados pelo nível de desenvolvimento dos países e regiões.

Sabe-se que regiões menos desenvolvidas, invariavelmente associadas a baixos níveis de escolaridade, são aquelas onde as pessoas deixam de se registrar com maior frequência, ou registram fora dos prazos legais, ou fornecem informações equivocadas (HORTA, 2016). Portanto, é evidente que a qualidade dos dados censitários dos idosos centenários está diretamente relacionada com as condições socioeconômicas e culturais da população. Desta

forma, os países nessa categoria de desenvolvimento foram todos classificados como de qualidade “muita baixa”.

Para exemplificar metodologicamente a proposta aqui apresentada e retratar de forma sistemática, o passo a passo até a classificação final, foram utilizados os dados censitários dos centenários do Brasil como exemplo prático. No último censo demográfico brasileiro, realizado em 2010, foram recenseados 24.236 idosos com 100 anos ou mais, destes 7.247 eram homens e 16.989 eram mulheres. No mesmo ano, foram recenseados também 464.499 homens e 804.113 mulheres com 85 anos ou mais. A Tabela 4.1.8 apresenta a construção dos três indicadores, os pesos e a classificação final da qualidade dos dados censitários para o Brasil.

Tabela 4.1.8 - Classificação da qualidade dos dados censitários dos centenários no Brasil, 2010.

| Indicadores dos centenários      | Cálculo                          | Padrão*       | Peso    |
|----------------------------------|----------------------------------|---------------|---------|
| Razão de Sexo                    | $7247/16989=0,43$                | Intermediário | 2       |
| Cent./85 anos ou mais (homens)   | $7.247/46.449 \times 100=1,56$   | Intermediário | 2       |
| Cent./85 anos ou mais (mulheres) | $16.989/80.4113 \times 100=2,11$ | Intermediário | 2       |
| Soma dos pesos                   |                                  |               | 6       |
| Classificação Final**            |                                  |               | Regular |

Fonte: Elaboração própria.

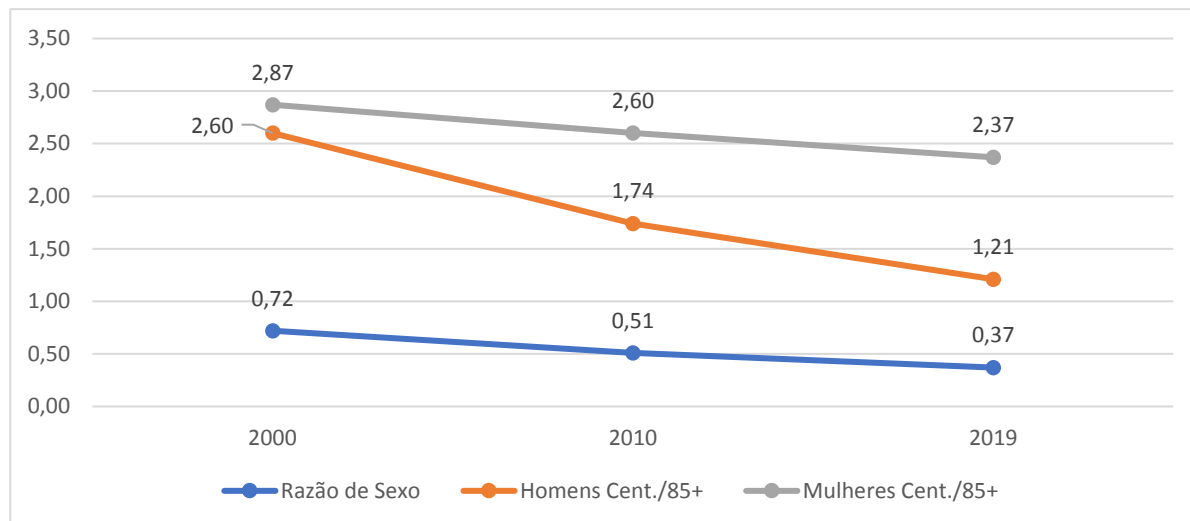
\*Padrão: intervalo de referência que o valor do cálculo corresponde, conforme classificação da Tabela 4.6.

\*\*Classificação final conforme apresentado no Quadro 3.3.

De acordo com os valores encontrados no cálculo dos indicadores, foi possível identificar o intervalo padrão correspondente para cada indicador conforme a Tabela 4.1.8. Como foi descrito na metodologia desta tese, o intervalo padrão de qualidade definido como “Intermediário” tem peso 2. Posteriormente, procedeu-se a soma dos pesos dos três indicadores ( $2 + 2 + 2=6$ ). Tomando como referência o Quadro 3.3, a soma dos pesos igual a 6 corresponde a classificação final da qualidade dos dados censitários dos centenários como “Regular”.

Como o foco deste estudo é o Semiárido brasileiro e ressaltando a importância (demonstrada estatisticamente) do uso dos indicadores utilizados na proposta apresentada, a Figura 4.1 representa o gráfico com a evolução de cada indicador selecionado na proposta de classificação ao longo dos anos estudados. Para a construção dos referidos indicadores foram elaboradas projeções para 2019, que já foram descritas na Tabela 4.1.2.

Figura 4.1 - Representação dos indicadores de qualidade para avaliação dos dados populacionais dos centenários: Razão de Sexo, Razão do número de Centenários e o número idosos com 85 anos ou mais (homens), Razão do número de Centenários e o número de idosos com 85 anos ou mais (mulheres), para o Semiárido brasileiro, 2000, 2010 e 2019.



Fonte: Elaboração própria

A classificação proposta neste estudo foi implementada a partir de dados censitários, porém a análise de comportamento de cada indicador, ao longo dos anos, não impede que sejam utilizados dados provenientes de projeções.

Como já mencionado, um indicador que expressa uma baixa qualidade na contagem da população centenária é uma elevada Razão de Sexo. A Figura 4.1 representa a RS dos centenários no Semiárido brasileiro nos anos 2000, 2010 e 2019. Em 2000, notou-se que o Semiárido possuía RS maior que 0,7 sendo bastante elevado e que se enquadra dentro do intervalo padrão de reduzida qualidade, apresentado na Tabela 4.1.6. Comparando os valores entre os anos, notou-se que a RS em 2010 diminuiu consideravelmente e, em 2019, caiu quase pela metade comparado ao ano 2000, o que expressa uma melhora significativa especificamente deste indicador analisando a série histórica.

Nos anos estudados, os valores da razão de sexo estavam todos fora do intervalo de referência mundial (0,138-0,353) de padrão de elevada qualidade, descrito na Tabela 4.1.6. Desse modo, os valores do Semiárido brasileiro só poderiam ser explicados por variações significativas nos erros de declaração de idade por sexo ou por algum efeito de concentração de população masculina nessas áreas em razão de migrações passadas (REIS; TURRA 2016), embora essa última explicação seja menos provável devido ao arrefecimento dos movimentos migratórios no Nordeste na última década. Como já mencionado, a redução da RS em 2010 e 2019 pode sinalizar uma melhora da qualidade dos dados dos centenários, principalmente em

2019, que o valor da RS foi de 0,37, muito próximo de ser incluído no intervalo padrão de elevada qualidade. Contudo, mesmo identificado esse avanço, os valores das RS do Semiárido ainda estão distantes dos países mais desenvolvidos considerados com alta qualidade para dados censitários de centenários, dentre eles: Suécia (0,18), Suíça (0,18), Noruega (0,19), Alemanha (0,14), Irlanda (0,15), Reino Unido (0,20), Estados Unidos (0,20).

Outro indicador para identificar possíveis erros de contagem censitária é a razão entre o número de centenários com 100 anos ou mais e a população com 85 anos ou mais, apresentado na Figura 4.1, com os dados do Semiárido, entre homens e mulheres. Observou-se que houve diminuição dos valores destes indicadores no decorrer dos anos, o que reforça a existência de erros na contagem censitária dos centenários em censos mais antigos.

De forma geral, observou-se que em 2000, 2010 e 2019, nenhum valor desses dois últimos indicadores esteve dentro do intervalo de parâmetro mundial referente a elevada qualidade - Homens (0,203-0,762), Mulheres (0,536-1,40) - de acordo com a Tabela 4.1.6, porém o que observou-se foi que, em 2000, o indicador para homens (2,60) esteve dentro do padrão de reduzida qualidade (2,41-7,61). Considerando alguns países que pertencem ao intervalo de qualidade elevada, percebeu-se o quanto o Semiárido ficou distante em termos de qualidade desse indicador, tendo como exemplo: Suécia, Suíça, Noruega, Alemanha, Irlanda, Reino Unido, Estados Unidos. Para esses países as razões não ultrapassam o limite de 1%, o que evidencia que no Semiárido houve melhora desses indicadores no decorrer das décadas, mas ainda com uma qualidade dos dados insatisfatória.

A Tabela 4.1.9 detalha a implementação da proposta de classificação da qualidade dos dados censitários dos centenários no Semiárido brasileiro, para os dois últimos censos demográficos realizados no Brasil, 2000 e 2010.

Tabela 4.1.9 - Descrição do Padrão de qualidade de cada indicador, soma dos pesos e classificação final da qualidade dos dados censitários dos centenários no Semiárido brasileiro, para os anos 2000, 2010.

| Anos | Razão de Sexo | Cent./85+ (H) | Cent./85+ (M) | Soma de Pesos | Classificação Final* |
|------|---------------|---------------|---------------|---------------|----------------------|
| 2000 | Reduzido      | Reduzido      | Intermediário | 4             | Baixa                |
| 2010 | Intermediário | Intermediário | Intermediário | 6             | Regular              |

Fonte: Elaboração própria.

Padrão: intervalo de referência que o valor do cálculo corresponde, conforme classificação da Tabela 4.6.

\*Classificação final conforme apresentado no Quadro 3.3.

De acordo com a Tabela 4.1.9 e analisando a classificação final de acordo com a proposta aqui apresentada, os dados censitários dos centenários no Semiárido brasileiro

melhoraram de um censo para o outro, passando de Baixa qualidade em 2000 para qualidade Regular em 2010. Essa classificação definida como “Regular”, no último censo, assemelha-se à classificação encontrada em países como Venezuela, Equador, Peru, Chile e o próprio Brasil, conforme descrito na Tabela 4.1.7.

Os resultados encontrados a partir da análise dos três indicadores e da implementação da proposta de classificação são consistentes com estudos anteriores sobre a qualidade dos dados da população centenária para a população brasileira, os quais fizeram uso dos mesmos indicadores (GOMES; TURRA, 2009; REIS; TURRA, 2016). De acordo com os resultados desses estudos e os encontrados nesta tese, dois fatores merecem ser destacados a partir dos resultados. Em primeiro lugar, nos estudos referidos, observou-se uma alta proporção de centenários, especialmente do sexo masculino, em regiões brasileiras que são reconhecidas por terem baixa qualidade de registros vitais, o que influenciava em elevados valores de RS. Neste trabalho, o mesmo foi observado para o Semiárido brasileiro. Nesta mesma linha, Paes (2020) chamou a atenção para o fato de que a região semiárida, na década passada, possuía níveis de sub-registros importantes, tanto dos nascimentos quanto dos óbitos, os quais estariam atrelados a condições de vida e de um sistema de coleta e registro de dados mais frágil.

Em segundo lugar, observou-se uma mudança importante nos indicadores de qualidade na contagem da população centenária em 2010, reforçando a ideia de que os dados mais antigos são de pior qualidade e que na medida em que coortes mais escolarizadas, cujos membros tem melhor conhecimento sobre seus próprios eventos vitais substituem as coortes anteriores, há uma melhoria nos resultados (KUMON et al, 2009). Porém mesmo com esse avanço ao longo das décadas, os resultados ainda estão distantes daqueles encontrados nos países desenvolvidos. O processo de envelhecimento no Brasil e no Semiárido ainda está atrás dos países desenvolvidos. Países como Estados Unidos, Japão e Itália já são reconhecidos há anos com seu grande incremento populacional de idosos, o que reflete nos indicadores estudados e evidencia-se no real e confiável aumento do número de centenários nesses países (KUMON et al, 2011).

#### **4.1.3 Implementação de métodos indiretos para estimar a população centenária para os anos de 2000, 2010 e 2019.**

A Tabela 4.1.10 apresenta os resultados das estimativas da população centenária no Semiárido brasileiro e seus espaços regionalizados, para os anos de 2000, 2010 e 2019, com base em métodos indiretos descritos na sessão 3.4.2. Os resultados são comparados aos dados

apresentados nos censos dos respectivos anos de 2000 e 2010. Em 2019, as estimativas indiretas foram comparadas às projeções populacionais descritas na Tabela 4.1.1.

Tabela 4.1.10 - Distribuição do número de centenários com base em informações censitárias e estimativas indiretas com base nos óbitos corrigidos, segundo sexo, entre os espaços regionalizados do Semiárido brasileiro, 2000, 2010 e 2019.

| Grandeza do Semáforo Branco, 2000, 2010 e 2019* |        |                                              |                   |          |                                              |                   |
|-------------------------------------------------|--------|----------------------------------------------|-------------------|----------|----------------------------------------------|-------------------|
| Regiões                                         | Homens |                                              |                   | Mulheres |                                              |                   |
|                                                 | Censo  | Estimativa com base nos<br>óbitos corrigidos |                   | Censo    | Estimativa com base nos<br>óbitos corrigidos |                   |
|                                                 |        | Estimado                                     | Diferença/Razão** |          | Estimado                                     | Diferença/Razão** |
| 2000                                            |        |                                              |                   |          |                                              |                   |
| Urbano                                          | 618    | 107                                          | 511 / 5,78        | 864      | 204                                          | 660 / 4,24        |
| Intermediário                                   | 275    | 40                                           | 235 / 6,87        | 411      | 101                                          | 310 / 4,07        |
| Rural                                           | 948    | 150                                          | 798 / 6,32        | 1275     | 292                                          | 983 / 4,37        |
| Semiárido                                       | 1.841  | 296                                          | 1.545 / 6,22      | 2.550    | 595                                          | 1.955 / 4,29      |
| 2010                                            |        |                                              |                   |          |                                              |                   |
| Urbano                                          | 608    | 240                                          | 368 / 2,53        | 1.287    | 441                                          | 846/ 2,92         |
| Intermediário                                   | 288    | 127                                          | 161 / 2,27        | 569      | 182                                          | 387 / 3,13        |
| Rural                                           | 885    | 341                                          | 544 / 2,60        | 1.665    | 544                                          | 1.121 / 3,06      |
| Semiárido                                       | 1.781  | 705                                          | 1.076 / 2,54      | 3.521    | 1.159                                        | 2.362 / 3,04      |
| 2019*                                           |        |                                              |                   |          |                                              |                   |
| Urbano                                          | 588    | 333                                          | 255 / 1,77        | 1.744    | 701                                          | 1.043 / 2,49      |
| Intermediário                                   | 277    | 180                                          | 97 / 1,54         | 754      | 335                                          | 419 / 2,25        |
| Rural                                           | 864    | 576                                          | 288 / 1,50        | 2.216    | 975                                          | 1.241 / 2,27      |
| Semiárido                                       | 1.729  | 1.082                                        | 647 / 1,60        | 4.714    | 2.006                                        | 2.708 / 2,34      |

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE, Censos 2000, 2010; 2019 (projeção) e SIM/MS.

\* Valor projetado para 2019, com base nos censos de 2000 e 2010.

\*\*Diferença Absoluta e a Razão entre Censo e Estimativas indiretas.

Na Tabela 4.1.10 o resultado da contagem do censo de 2000 e 2010 e da projeção para 2019 foi comparado com estimativas indiretas obtidas por meio da aplicação do método das Gerações Extintas (para os anos 2000 e 2010) e pelo método proposto por Rosenwaike para o ano de 2019. Como esses métodos fornecem estimativas indiretas da população baseadas no número de óbitos, foi necessário implementar a correção do sub-registro de óbitos, segundo sexo. A cobertura dos registros de óbitos e os respectivos fatores de correção, do Semiárido brasileiro e seus espaços regionalizados, para 2000, 2010 e 2019, também foi estimado nesse estudo e é apresentado na sessão 4.2.1. Sendo assim, é necessário ressaltar que as estimativas indiretas apresentadas na Tabela 4.1.10 já foram corrigidas para sub-registros de óbitos, de acordo com os fatores de correção apresentados na Tabela 4.2.1, da próxima sessão.

Parte-se de 2000, ano em que foi possível implementar o método das Gerações Extintas e verificar as maiores discrepâncias e variações significativas entre os resultados das estimativas indiretas e a população recenseada. Os resultados sugerem um número de centenários no Semiárido brasileiro, em 2000, 6 (seis) vezes menor para homens e 4 (quatro) vezes menor para mulheres, do que o enumerado pelo IBGE naquele ano. Entre os espaços regionalizados, as

maiores variações da razão entre censo e método indireto ocorreu entre os homens residentes em municípios caracterizados como Intermediário (6,87), já entre as mulheres a variação mais elevada foi na região Rural (4,37). Apesar de possíveis erros nos dados do censo, a consistência dos resultados entre o método indireto empregado, também levanta sérias dúvidas com relação à veracidade dos registros de óbitos dos centenários do Semiárido brasileiro em 2000. Além disso, vale ressaltar, que os erros foram maiores entre os homens, o que difere sobretudo dos anos posteriores, em que as diferenças foram mais altas entre mulheres.

Como forma de checagem dos registros de óbitos e com o objetivo de diminuir a presença de erros provenientes do banco de dados, foi realizada uma nova coleta do microdados do SIM e procedeu-se o processo de checagem dos dados. Como esperado, foram encontrados os mesmos valores anteriores dos registros de óbitos dos centenários para o Semiárido brasileiro, o que reforça que possa ter ocorrido sub-registro de óbitos e erro de captação desses registros especificamente em 2000, pois as diferenças foram bastante significativas.

Estudo realizado por Gomes e Turra (2009) também evidenciou discrepâncias maiores nos resultados gerados pelos métodos indiretos, no ano 2000. Os autores explicam que estas diferenças podem ser explicadas por mudanças nas taxas de crescimento dos óbitos ao longo dos anos, que provavelmente passaram a ter um peso maior na definição do tamanho das coortes.

Em 2010, ano em que também foi possível aplicar o método indireto das Gerações Extintas para estimar o número de centenários, verificou-se variações importantes entre os valores estimados e a população recenseada. No geral e para todos os espaços regionalizados, entre os sexos, os resultados indicaram um número de centenários estimados menor do que o enumerado pelo IBGE neste ano, cujas diferenças variaram com uma razão de 2,27 (Intermediário- homens) a 3,13 (Intermediário-mulheres). Portanto, o IBGE no último censo recenseou cerca de 2 a 3 vezes mais centenários que o número de centenários estimados pelos métodos indiretos no Semiárido brasileiro, segundo os resultados da Tabela 4.1.10.

Em função dos diferenciais relativos de mortalidade por sexo, tanto no censo quanto nas estimativas, nota-se que houve um maior volume populacional de mulheres, atingindo quase o dobro com relação aos homens. Essa predominância com vantagens para as mulheres também foi encontrada por Gomes e Turra (2009) ao analisar a consistência dos dados recenseados em 2000 para o Brasil e regiões, assim como os resultados de Teberlanche e Wilson (2015), usando os mesmos métodos aplicados aqui para as estimativas populacionais dos centenários.

Um olhar mais específico com relação ao sexo, identificou-se que em 2010 na região Rural e Intermediária houve uma maior variação entre censo e estimativa indireta entre as mulheres, para os homens essa diferença foi maior no Semiárido como um todo e na área Rural.

Observando as variações entre o censo de 2010 e a projeção de 2019, para todo o Semiárido brasileiro, o número de centenários diminuiu entre os homens e aumentou entre as mulheres o que pode equilibrar a superestimação do número de centenários recenseados entre os homens, ao longo dos anos, porém isso não ocorreu entre as mulheres, onde o número de centenárias recenseadas aumentou de forma robusta.

Além disso, em 2019, ao contrário do que foi observado em 2010, houve menores discrepâncias nos resultados comparados entre a projeção populacional com base no censo e a estimativa gerada pelo método indireto proposto por Rosenwaike. Em parte, estas diferenças podem ser explicadas pela melhoria nas bases de registros dos óbitos entre 2010 e 2019, ou ainda pelos erros embutidos nos cálculos da projeção.

No geral, as diferenças absolutas e razão entre número observado e estimado de centenários, por sexo, deram sinais de uma superestimação do número de centenários nos censos e que, evidentemente, afetou as projeções. Apesar de possíveis erros nos dados de óbitos e a consistência dos resultados entre os métodos indiretos empregados, levanta-se sérias dúvidas com relação à veracidade da contagem dos centenários nos anos estudados para ambos os sexos na região semiárida. Admitindo-se uma sobrecontagem dos centenários em 2010, bem como em 2000 e esse efeito contaminou as projeções que deram uma vantagem para os homens em 2019 (com menor diferencial).

Aprofundando o estudo da contagem da população centenária e a relação entre os sexos, utilizando a Razão de Sexo (RS) indicador base para a construção da proposta de classificação da qualidade da contagem dos centenários, apresentada nesta tese na sessão 4.1.2 e publicada na Revista Latinoamericana de Poblacion (RELAP) por Medeiros e Paes (2022), foi possível utilizar a RS para auxiliar na discussão sobre o diferencial nos anos estudados, entre o censo, projeção e os valores estimados. A Tabela 4.1.11 descreve a RS dos centenários do Semiárido brasileiro, construída com os dados do censo de 2000 e 2010, projeção de 2019 e valores estimados por métodos indiretos, nos referidos anos.

Tabela 4.1.11 - Razão de sexo dos centenários do Semiárido brasileiro, segundo o Censo 2000 e 2010, Projeção de 2019, estimativas indiretas, 2000, 2010 e 2019.

| <b>Método utilizado</b>    | <b>Razão de Sexo</b> |      |       |
|----------------------------|----------------------|------|-------|
|                            | 2000                 | 2010 | 2019* |
| <b>Censo IBGE</b>          | 0,72                 | 0,51 | 0,37  |
| <b>Métodos indiretos**</b> | 0,50                 | 0,61 | 0,54  |

Fonte: Elaboração própria

\*Projeção 2019 construída a partir dos censos 2000 e 2010

\*\* Métodos indiretos baseados nos óbitos: 2000 e 2010 - Gerações Extintas; 2019 – Rosenwaike.

De acordo com os dados oriundos de 85 países (MEDEIROS, PAES, 2022) e apresentados nessa tese na Tabela 4.1.6, para aqueles países considerados de elevada qualidade nos seus dados e registros censitários dos centenários, a razão de sexo quantificada variou entre (0,14-0,35). Analisando os valores de razão de sexo da Tabela 4.1.11, nenhum resultado da aplicação dos métodos indiretos ou da contagem censitária ficou dentro da faixa de referência de boa qualidade dos dados e registros populacionais na coleta das informações da declaração de idade, o que sugere que mesmo com possíveis avanços alcançados ao longo dos anos ainda são necessários melhorias principalmente quando se trata da faixa etária dos centenários.

Os resultados da RS dos dados obtidos através dos Censos diminuíram entre 2000 e 2019, chegando neste ano a uma RS de 0,37, valor perto de se enquadrar no intervalo de elevada qualidade. Porém, analisando a RS proveniente de métodos indiretos, em 2000 o resultado foi o menor (0,50), aumentando em 2010 (0,61) e diminuindo em 2019 (0,54), cujas flutuações parecem reforçar que os dados de óbitos de 2000 possa ter uma subnotificação ou algum tipo de problema, como já foi reforçado na discussão da Tabela 4.1.10.

Contudo, mesmo com a melhoria dos resultados da RS entre os anos, eles ainda não foram suficientes para produzirem uma RS compatível com o nível de qualidade observado para países desenvolvidos. Isto significa que muito provavelmente persistem problemas com a qualidade dos registros populacionais e de óbitos no sistema de captação de dados, seja nas declarações da idade, montante, formas de captação dos dados, erros no processamento, entre outros.

Por fim, diante das discrepâncias encontradas nos resultados produzidos neste estudo, as análises realizadas podem servir como alerta e sinalizadoras para minimizar equívocos específicos e evitar repetições de erros na avaliação da consistência da contagem dos dados, uma vez que o Censo Demográfico do Brasil de 2022 se encontra em fase de finalização até a apresentação desta tese. Superar as lacunas na captação de centenários e na declaração de idade é necessário para revelar com mais precisão o processo de envelhecimento no Brasil.

## 4.2 Avaliação da qualidade dos registros de óbitos dos centenários do Semiárido brasileiro

### 4.2.1 Cobertura dos óbitos

A Tabela 4.2.1 mostra a cobertura de óbitos dos adultos, por sexo, para os espaços regionalizados do Semiárido, nos anos 2000, 2010 e 2019 considerados como referência. No comparativo entre as três regiões, o espaço Rural revelou níveis mais baixos de cobertura que os espaços Intermediário e Urbano, cuja situação foi identificada em todos os anos estudados, o que sugere uma desvantagem no tocante à cobertura dos óbitos para aqueles municípios caracterizados como rurais. De certa forma, este fato poderia ser explicado por esta região apresentar condições de vida mais precárias que os demais espaços, pois já foi descrito na literatura (BAPTISTA; CAMPOS, 2013; PAES; SILVA; BRAGA, 2014; ASA, 2020) que os aglomerados urbanos têm historicamente mostrado vantagens na cobertura dos óbitos comparados ao interior e zona rural que são predominantes no Semiárido representando 76% dos seus municípios. Agrega-se o fato dos cartórios estarem sempre alocados em áreas urbanas e intermediárias o que pode dificultar o acesso dos residentes em áreas rurais. Sem embargo foi observado o avanço e melhoria da cobertura dos óbitos ao longo dos anos, em ambos os sexos, conforme apresentado, particularmente entre 2000 e 2010, com maiores diferenciais nas coberturas.

Tabela 4.2.1: Cobertura de óbitos adultos dos espaços regionalizados do Semiárido brasileiro, segundo o sexo, 2000, 2010 e 2019.

| Regiões       | Homens |       |       | Mulheres |       |       |
|---------------|--------|-------|-------|----------|-------|-------|
|               | 2000   | 2010  | 2019  | 2000     | 2010  | 2019  |
| Urbano        | 83,25  | 91,63 | 93,78 | 82,37    | 90,37 | 92,38 |
| Intermediário | 80,11  | 89,37 | 92,75 | 78,32    | 88,02 | 89,92 |
| Rural         | 78,88  | 88,12 | 89,73 | 76,77    | 86,28 | 88,21 |
| Semiárido     | 80,75  | 89,71 | 92,09 | 79,15    | 88,22 | 90,17 |

Fonte dos dados básicos: Ministério da Saúde/SIM; IBGE/ Censo Demográfico.

Observou-se que a cobertura dos óbitos dos adultos foi maior entre os homens que entre as mulheres, em todos os espaços regionalizados do Semiárido brasileiro, nos três anos estudados. Usualmente aponta-se que um dos fatores que concorre para tal diferencial deve-se a um maior registro dos óbitos por Causas Externas referentes ao sexo masculino. Aponta-se ainda a legislação brasileira que até décadas recentes privilegiava o usufruto de pensões e benefícios legais às mulheres que registravam os óbitos dos seus maridos ou companheiros para receber tais benefícios, além de uma cultura que favorece um maior controle social e de

segurança que implicam em um maior registro para os homens, fatos estes que poderiam ter influenciado as coberturas obtidas (PAES; SILVA; BRAGA, 2014).

No sexo masculino, as coberturas dos óbitos para os espaços Rural e Intermediário nos anos 2000, não ultrapassaram 81%, configurando-se como *regular*. Já o espaço Urbano, em 2000, Rural e Intermediário em 2010, Rural em 2019, foram classificados como *bom*. No espaço Urbano de 2010 e 2019 e no Intermediário de 2019 as coberturas variaram entre 91% a 93%, classificadas como  *muito bom*. Com relação à cobertura do Semiárido como um todo, para 2000, obteve-se uma cobertura menor que 81%, classificada como *regular*, já para os anos de 2010 configurou-se como *bom* e 2019 os valores da cobertura foram classificados como  *muito bom*.

No sexo feminino, as coberturas foram mais baixas comparadas aos homens. No ano 2000 todos os espaços regionalizados estudados apresentaram uma cobertura entre 76% e 82%, configurando-se como *regular* nas áreas Rural e Intermediário e *bom* na área Urbana. Em 2010, foi observado que em todos os espaços regionalizados os valores de cobertura encontravam-se dentro da classificação *bom*. Em 2019, os espaços Rural e Intermediário permaneceram com as coberturas menores que 90%, configurando-se ainda como *bom*, apenas o espaço Urbano configurou-se como  *muito bom*. Com relação à cobertura do Semiárido como um todo, em 2000 observou-se uma cobertura menor que 80% classificada como *regular*, contudo em 2010 e 2019 a cobertura configurou-se como *bom*.

O que se depreende como análise geral é que houve para a população adulta uma melhora nas magnitudes das coberturas tanto no tempo como na geografia partindo do rural para o urbano e em ambos os sexos, com uma tendência à homogeneização dos níveis. De fato, à medida que ocorre avanços na melhoria dos níveis de cobertura dos óbitos, eles tendem a ficarem mais próximos entre áreas e entre os sexos, principalmente quando atingem a barreira dos 90%, classificados como  *muito bom* (PAES, 2018)

Um estudo foi realizado pelo IBGE (2022) acerca da cobertura dos óbitos de indivíduos com 80 anos ou mais, através do método de Captura-Recaptura proposto por Laplace (1783). A estimação pelo Método de Captura-Recaptura consiste em se aproveitar da superposição dos registros presentes em fontes sabidamente incompletas, mas disponíveis, para medir a subestimação das fontes. Isso permite a correção das estatísticas, de maneira a produzir indicadores mais próximos da realidade.

O referido estudo utilizou as fontes de registros de óbitos provenientes do IBGE e do Ministério da Saúde, por município (IBGE, 2022). De acordo com esse estudo a cobertura de

óbitos, para o conjunto dos 1262 municípios do Semiárido brasileiro, em 2018, não segregado por sexo, foi a seguinte: Cobertura de óbitos IBGE – 91%; Ministério da Saúde – 94%. Ambas fontes fizeram uso dos dados de óbitos do SIM/MS. Essas magnitudes quando comparadas com as coberturas de óbitos obtidas neste trabalho que também fez uso da mesma base de dados do SIM/MS (cerca de 91% para o Semiárido como um todo em 2019) a diferença máxima recaiu em torno de 3%, o que está dentro da margem de tolerância embutido nas estimativas quando da aplicação dos respectivos métodos. Com isto, é possível especular que a utilização das coberturas obtidas para os adultos neste trabalho e que foram consideradas como “proxy” para as idades centenárias, não devem ser tão diferentes entre os dois grupos etários.

Assumindo que as coberturas estimadas para os adultos sejam admitidas como aproximações para a cobertura dos centenários, mostra-se na Tabela 4.2.2 os fatores de correção estimados para ambos os sexos, com os valores notificados e corrigidos dos óbitos para os espaços regionalizados do Semiárido, nos triênios estudados. A diferença nos óbitos notificados entre os sexos, evidencia um maior número de notificação de óbitos entre as mulheres com relação aos homens, que aumenta ao longo dos anos.

Tabela 4.2.2 - Número de óbitos de idosos centenários notificados ao SIM, corrigidos e fatores de correção para os espaços regionalizados e sexo do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.

| Regiões       | 1999-2001  |           |       |            |           |       |
|---------------|------------|-----------|-------|------------|-----------|-------|
|               | Homens     |           |       | Mulheres   |           |       |
|               | Notificado | Corrigido | Fator | Notificado | Corrigido | Fator |
| Urbano        | 83         | 100       | 1,20  | 144        | 174       | 1,21  |
| Intermediário | 47         | 59        | 1,25  | 75         | 96        | 1,28  |
| Rural         | 110        | 140       | 1,27  | 171        | 222       | 1,30  |
| Semiárido     | 240        | 299       | 1,24  | 390        | 492       | 1,26  |
| Regiões       | 2009-2011  |           |       |            |           |       |
|               | Notificado | Corrigido | Fator | Notificado | Corrigido | Fator |
|               | Notificado | Corrigido | Fator | Notificado | Corrigido | Fator |
| Urbano        | 499        | 544       | 1,09  | 1.024      | 1137      | 1,11  |
| Intermediário | 242        | 271       | 1,12  | 404        | 461       | 1,14  |
| Rural         | 652        | 737       | 1,13  | 1.096      | 1.271     | 1,16  |
| Semiárido     | 1.393      | 1.552     | 1,11  | 2.524      | 2.869     | 1,13  |
| Regiões       | 2018-2020  |           |       |            |           |       |
|               | Notificado | Corrigido | Fator | Notificado | Corrigido | Fator |
|               | Notificado | Corrigido | Fator | Notificado | Corrigido | Fator |
| Urbano        | 936        | 1.002     | 1,07  | 1.831      | 1.977     | 1,08  |
| Intermediário | 487        | 526       | 1,08  | 828        | 919       | 1,11  |
| Rural         | 1.513      | 1.679     | 1,11  | 2.398      | 2.710     | 1,13  |
| Semiárido     | 2.936      | 3.207     | 1,09  | 5.057      | 5.606     | 1,11  |

Fonte: DATASUS/SIM/MS.

O maior número de óbitos notificados entre as mulheres pode ser explicado pelo fenômeno conhecido como feminização da velhice, já discutido anteriormente nessa tese. As mulheres chegam em maior número a faixa etária de centenários, o que consequentemente aumenta o número de óbitos entre as mulheres nessa faixa etária, quando comparadas aos

homens, pois elas apresentam maior probabilidade de serem portadoras de múltiplas síndromes geriátricas e maior prevalência de doenças musculoesqueléticas e déficits sensoriais (ROCHON et al, 2014)

Inversamente ao que já foi evidenciado nos valores da cobertura, os fatores de correção das mulheres são maiores quando comparados aos dos homens e, no estudo dos triênios, o valor dos fatores de correção vai diminuindo, em ambos os sexos, evidenciando também uma melhoria dos registros no decorrer dos anos. Ou seja, os percentuais de ajustes para o Semiárido brasileiro, para os sexos conjugados, diminuíram de 25%, 12% e 10% respectivamente para 2000, 2010 e 2019.

#### **4.2.2 Principais características sociodemográficas e operacionais dos óbitos dos centenários**

Dando sequência à exploração do estudo da mortalidade dos centenários da região foi realizada uma análise descritiva das principais características sociodemográficas e operacionais dos óbitos ocorridos ao longo de 20 anos. O conhecimento do perfil epidemiológico e as causas de mortalidade da população centenária são fundamentais para o planejamento de políticas públicas eficazes, sendo possível identificar as melhores estratégias para melhorar as condições de saúde dos centenários, principalmente daqueles idosos que residem em regiões marcadas por baixos níveis de desenvolvimento social e econômico, como é o caso do Semiárido brasileiro.

A análise descritiva foi composta por 28.985 registros de óbitos de indivíduos com 100 anos ou mais ocorridos no Semiárido brasileiro de 2000 a 2020. A Tabela 4.2.3 descreve a caracterização de importantes variáveis sociodemográficas, operacionais, assim como os dados faltantes (*missing-values*) para cada variável e suas categorias dos registros de óbitos dos centenários ocorridos no período estudado.

Tabela 4.2.3 - Características sociodemográficas e operacionais dos registros de óbitos dos centenários ocorridos no Semiárido brasileiro, 2000 - 2020.

| <b>Variável</b>                                          | <b>Categorias</b>     | <b>n</b> | <b>%</b> |
|----------------------------------------------------------|-----------------------|----------|----------|
| <b>Sexo</b>                                              | Masculino             | 10.504   | 36,24    |
|                                                          | Feminino              | 18.474   | 63,74    |
|                                                          | <i>Missing-values</i> | 7        | 0,02     |
| <b>Idade</b>                                             | 100 – 104 anos        | 24.143   | 83,30    |
|                                                          | 105 – 109 anos        | 4.215    | 14,54    |
|                                                          | 110 anos ou mais      | 627      | 2,16     |
|                                                          | <i>Missing-values</i> | 0        | 0        |
| <b>Raça/Cor</b>                                          | Branca                | 9.357    | 32,28    |
|                                                          | Preta                 | 2.221    | 7,66     |
|                                                          | Amarela               | 117      | 0,40     |
|                                                          | Parda                 | 14.619   | 50,44    |
|                                                          | Indígena              | 68       | 0,23     |
|                                                          | <i>Missing-values</i> | 2.603    | 8,99     |
| <b>Situação Conjugal</b>                                 | Solteiro              | 7.309    | 25,22    |
|                                                          | Casado                | 3.120    | 10,76    |
|                                                          | Viúvo                 | 15.319   | 52,85    |
|                                                          | Separado/Divorciado   | 113      | 0,39     |
|                                                          | União Estável         | 288      | 1,00     |
|                                                          | <i>Missing-values</i> | 2.836    | 9,78     |
| <b>Escolaridade<br/>(anos concluídos)</b>                | Sem escolaridade      | 17.891   | 61,73    |
|                                                          | 1 – 3 anos            | 3.196    | 11,03    |
|                                                          | 4 – 7 anos            | 976      | 3,37     |
|                                                          | 8-11 anos             | 269      | 0,92     |
|                                                          | 12 anos ou mais       | 61       | 0,21     |
|                                                          | <i>Missing-values</i> | 6.592    | 22,74    |
| <b>Caracterização do<br/>Município de<br/>Residência</b> | Rural                 | 13.709   | 47,30    |
|                                                          | Intermediário         | 4.851    | 16,74    |
|                                                          | Urbano                | 10.425   | 35,96    |
|                                                          | <i>Missing-values</i> | 0        | 0,00     |
| <b>Recebeu assistência<br/>médica?</b>                   | Sim                   | 8.216    | 28,35    |
|                                                          | Não                   | 2.667    | 9,21     |
|                                                          | <i>Missing-values</i> | 18.102   | 62,44    |
| <b>Local de ocorrência<br/>do óbito</b>                  | Hospital              | 8.322    | 28,71    |
|                                                          | Outros est. de saúde  | 314      | 1,08     |
|                                                          | Domicílio             | 20.010   | 69,06    |
|                                                          | Via pública           | 80       | 0,26     |
|                                                          | Outros                | 205      | 0,71     |
|                                                          | <i>Missing-values</i> | 54       | 0,18     |
| <b>Total</b>                                             |                       | 28.985   | 100,0    |

Fonte dos dados básicos: Sistema de Informação de Mortalidade – SIM, 2020. *Missing-values*: Dados faltantes.

Com relação as características sociodemográficas, observou-se, uma predominância nas seguintes categorias das variáveis: mulheres (63,7%); faixa etária 100-104 anos (83,3%); raça/cor parda (50,4%); viúvos (52,84%); sem escolaridade (61,7%), residentes em municípios caracterizados como Rural (47,3%).

Assim como já foi explicado anteriormente, nos centenários há diferença marcante entre os sobreviventes com relação ao sexo e essa diferença torna-se bastante acentuada ao longo dos anos, pois como a mortalidade masculina é maior nas faixas etárias anteriores, há uma vantagem das mulheres ao atingirem idades longevas, o que explica a maior mortalidade feminina neste estudo. Contudo, a aparente vantagem das mulheres é parcialmente atenuada pela maior prevalência de demências, depressão e dependência funcional neste grupo, reduzindo seu tempo de vida livre de incapacidades (CLOSS; SCHWANKE, 2017).

Em relação à estratificação etária, observou-se a maioria dos óbitos entre 100 e 104 anos. Essa situação é esperada, pois os eventos fatais entre os idosos em velhice avançada são precedidos de longos períodos de doenças, múltiplas síndromes geriátricas e maior prevalência de doenças musculoesqueléticas, déficits sensoriais, incapacidades e sofrimentos (CAMPS CALZADILLA et. al., 2013; ARAÚJO, 2018). Portanto, apesar da heterogeneidade da longevidade, a maioria dos indivíduos chegam aos 100 anos já apresentando comorbidades acentuadas, provenientes de doenças crônicas, o que leva a eventos fatais nas primeiras idades centenárias (KWAK et al., 2012). Porém, existem aqueles indivíduos que são capazes de atrasar e/ou escapar das doenças associadas à idade, rompendo as estatísticas de mortalidade, atingindo idades iguais ou superiores a 105 anos (PERLS et. al., 2002), que no Semiárido correspondeu a 16,7% dos óbitos nessa faixa etária.

Neste estudo, verificou-se que, dos centenários que morreram, a maioria eram viúvos. Esse resultado também é reflexo da maior mortalidade masculina em idades anteriores, evidenciando que as mulheres chegam as idades mais avançadas, porém sem companheiros. Essa condição, poderia, também, por outro lado, indicar uma condição de maior fragilidade, visto que alguns deles, possivelmente, não teriam condições de morar sozinhos, quer pelas doenças apresentadas, quer pelas limitações funcionais que poderiam existir em decorrência delas, necessitando de cuidados de familiares e cuidadores, sem laços conjugais (CEVENINI et al., 2008; VACANTE et al., 2012).

Outra característica dos centenários que faleceram no Semiárido brasileiro é a maior frequência da raça/cor parda e sem nível de escolaridade. Este achado com relação à etnia pode estar relacionado à predisposição genética, às piores condições de vida, ao menor acesso aos serviços de saúde e ao estresse devido à discriminação racial (OSLEN, et al, 2012; MALTA et al, 2017). Portanto, as desigualdades sociais existentes neste grupo étnico no país, principalmente entre aqueles com baixos níveis de instrução, principalmente entre as regiões mais pobres, como é o caso da região semiárida, possibilitam que esses indivíduos tenham

dificuldades de acesso aos serviços de saúde, escassa informação de prevenção de agravos e promoção da saúde, baixa condição econômica e frágil rede de apoio (CASTRO; RODRIGUES JÚNIOR, 2012), contribuindo para a maior mortalidade entre os pardos e com baixa escolaridade.

Com relação às variáveis operacionais referentes aos óbitos dos centenários apresentadas na Tabela 4.2.3, apenas 28,3% dos indivíduos receberam assistência médica durante a doença que ocasionou a morte, e 69,1% dos óbitos ocorreram no domicílio do idoso. Esses resultados apontam para um fato preocupante e muito prevalente entre idosos, o aumento expressivo do número de mortes em residências e sem assistência médica, o que indica claramente falhas no sistema de saúde e da assistência social, com carência de atenção oportuna e preventiva (COSTA et al, 2016). Segundo Nyatanga (2015), a combinação de fatores - como o excesso de mortalidade entre as pessoas idosas que não tiveram assistência médica durante a doença que ocasionou a morte e os óbitos em domicílio - apontam para um quadro de desassistência aos indivíduos que enfrentam doenças no final da vida e que necessitam de acompanhamento contínuo devido ao predomínio de doenças crônicas e de cuidados especiais.

#### **4.2.3 O estudo dos “*missing values*” e das causas mal definidas**

Um fato preocupante com relação a qualidade dos registros de óbitos dos centenários é a análise do percentual de dados faltantes ou *missing values*, também apresentado na Tabela 4.2.3. Segundo Romero e Cunha (2006), a completude das variáveis é um dos pilares de avaliação da qualidade de uma base de dados, uma vez que, a incompletude refere-se as informações em branco contidas nas variáveis, que compromete a consistência, interpretação e utilização desses dados. Os referidos autores definiram os critérios de avaliação da completude das variáveis de acordo com o percentual de *missing-values*, da seguinte forma: excelente (menos de 5%); boa (5 a 9%); regular (10 a 19%); ruim (20 a 49%); e muito ruim (50% ou mais).

Para cada variável do estudo, foi apresentada na Tabela 4.2.3 a quantidade e percentual de *missing-values*: as variáveis Sexo, Idade, Local de Ocorrência do óbito e Caracterização Municipal, apresentaram menos que 0,2% de dados faltantes; Raça/Cor: 8,9%; Situação Conjugal: 9,8%; Escolaridade: 22,7%. A variável com maior percentual de dados faltantes foi Assistência médica com 62,4%. Utilizando os critérios propostos por Romero e Cunha (2006), foi possível analisar a completude de cada variável estudada dentro do formulário: Sexo, Idade,

Caracterização Municipal e Local de Ocorrência do óbito - *excelente*; Situação Conjugal e Raça/Cor – *boa*; Escolaridade – *ruim*; Assistência médica – *muito ruim*.

Uma revisão sistemática realizada por Correira, Padilha e Vasconcelos (2014) evidenciou nos resultados que 36,8% das pesquisas referentes a análise da completude das variáveis do SIM utilizaram o escore elaborado por Romero e Cunha (2006), sendo a classificação da completude mais utilizada nos estudos. Na análise da completude das variáveis que compõem a DO, um estudo realizado por Rios et al (2016) com idosos que faleceram no estado da Bahia, entre 1996 e 2010, que também utilizou o escore acima citado, evidenciou que dentre as variáveis sociodemográficas a variável sexo foi totalmente preenchida, a escolaridade apresentou o maior índice de não preenchimento, seguida da Raça/cor e Situação conjugal, as duas últimas apresentaram no início da análise escore *regular* passando para *bom* em 2010. Já a variável assistência médica foi classificada como *muito ruim* em todos os anos analisados e Local de ocorrência apresentou os melhores índices, com escore *excelente* em todo o período. Todos esses resultados corroboram com a análise da completude das variáveis entre os centenários do Semiárido apresentados na Tabela 4.2.3.

Outra análise da completude realizada, no Brasil e regiões, em 2016, na faixa etária de 30 a 69 anos, por Muzy, Castanheira e Romero (2021), evidenciou que na região Nordeste e no Brasil, as variáveis da DO apresentaram os seguintes percentuais de *missing values*, respectivamente: Raça/cor (6% e 4%), Situação conjugal (14% e 10,7%), Escolaridade (25% e 22,7%), Assistência médica (36% e 36,7), evidenciando que os dados faltantes dos centenários no Semiárido estão aproximados aos encontrados no Brasil e no Nordeste.

Nos estudos científicos, a perda de informações, em alguns casos, provoca a exclusão total do sujeito da pesquisa. Essa exclusão pode provocar um viés nos resultados, uma vez que, os sujeitos excluídos, podem pertencer a determinadas categorias que não terão representatividade na amostra final. Em outros casos, os valores não declarados inviabilizam o uso de determinada variável, pois a quantidade de dados faltantes na mesma é muito alta (NUNES; KLÜCK; FACHEL, 2010). Na tentativa de obter o maior número de informações, técnicas são empregadas para resgatar esses valores por meio de re-verificação, autópsia verbal, *linkage* ou imputação dos valores, resultando esta última alternativa em uma completude total das informações finais, com a composição de um banco completo de informações.

Buscando compreender as razões que levaram os centenários do Semiárido a morrer, as causas de óbito foram classificadas e distribuídas segundo a CID-10, conforme apresentado na Tabela 4.2.4.

Tabela 4.2.4: Descrição das principais causas básicas de óbitos dos centenários, ocorridos no Semiárido brasileiro, 2000 - 2020.

| Capítulos CID-10 | Causa Básica de Óbito (CID – 10)           | n             | %          |
|------------------|--------------------------------------------|---------------|------------|
| I                | Algumas doenças infecciosas e parasitárias | 1.512         | 5,22       |
| II               | Neoplasmas [tumores]                       | 1.042         | 3,59       |
| III              | Doenças endócrinas, nutr. e metabólicas    | 1.547         | 5,34       |
| V                | Transtornos mentais e comportamentais      | 184           | 0,63       |
| VI               | Doenças do sistema nervoso                 | 416           | 1,44       |
| IX               | Doenças do aparelho circulatório           | 8.967         | 30,94      |
| X                | Doenças do aparelho respiratório           | 4.850         | 16,73      |
| XI               | Doenças do aparelho digestivo              | 542           | 1,87       |
| XIV              | Doenças do aparelho geniturinário          | 657           | 2,27       |
| XVIII            | Causas Mal definidas                       | 8.188         | 28,25      |
| XX               | Causas externas                            | 497           | 1,71       |
| -                | Outras causas*                             | 583           | 2,01       |
| <b>Total</b>     |                                            | <b>28.985</b> | <b>100</b> |

Fonte dos dados básicos: Sistema de Informação de Mortalidade – SIM, 2021. \*Outras causas: Não inclui causas externas.

Verificou-se na Tabela 4.2.4 que os principais grupos de causas de óbito entre os centenários neste estudo coincidem com as estatísticas nacionais (PEDRO; BÓS, 2017). Na análise descritiva de todo o período (2000-2020), as doenças do aparelho circulatório representaram 30,9% do total dos óbitos. A Causa mal definida (CMD), que tem sua especificação no capítulo XVIII da CID-10, foi a segunda causa mais frequente com 28,2%, seguida das Doenças do aparelho respiratório (16,7%). Essas duas causas em conjunto com as mal definidas foram responsáveis por mais de dois terços (76%) do total de óbitos dos centenários no período de estudo. Em menor proporção, encontraram-se as Doenças endócrinas, nutricionais e metabólicas (5,3%) e doenças infecciosas e parasitárias (5,2%). As causas externas representaram apenas 1,7% do total de registros.

No Brasil, segundo dados do SIM, em 1996 a principal causa de mortalidade entre os centenários era devido a Sinais, Sintomas e Achados Anormais de exames clínicos e laboratoriais, não classificados em outra parte (Capítulo XVIII do CID-10), ou seja, CMD, com cerca de 35% de mortes mal definidas nessa faixa etária. As doenças circulatórias apareciam em segundo lugar com 30%, seguidas das respiratórias (15%) e as neoplasias (pouco menos que 5%). Entre os anos de 1996 e 2007 observou-se uma transformação nesse quadro: queda das causas de mortes por capítulo XVIII com aumento substancial da participação das doenças

circulatórias e respiratórias, sendo as circulatórias a causa mais frequente entre os centenários (PEDRO; BÓS, 2017).

A importância das doenças cardiovasculares no perfil epidemiológico é compatível com à realidade mundial, responsáveis por 30% de todas as mortes por causa conhecida (MARQUES; CONFORTIN, 2015). Apesar do decréscimo que essas doenças vêm apresentando nas taxas de mortalidade da população total no Brasil e no mundo em decorrência de medidas preventivas, ainda persiste uma tendência crescente quando se trata do segmento de idosos, com um aumento de 53,7% do número global de mortes entre 1990 e 2016 para os indivíduos com 70 anos ou mais de idade, segundo o Global Burden of Diseases, Injuries, and Risk Factors Study (NAGHAVI et al, 2017). Parte desse aumento no Brasil pode estar relacionada ao paralelo declínio das taxas de óbitos por causas mal definidas, dado que a maioria delas, que são mais frequentes em idosos, tem as doenças do aparelho circulatório como provável causa. Estudos mostraram que a tendência de óbitos por doenças cardiovasculares nas regiões Nordeste e Centro-Oeste brasileiras apresentaram aumento do risco de morte ao longo dos anos (MANSUR et al, 2006; OLIVEIRA et al, 2020; MUZY; CASTANHEIRA; ROMERO, 2021). O mesmo foi observado em pesquisa realizada nos Estados Unidos, entre 2000 e 2014, a qual também evidenciou que a principal causa de morte entre os centenários foram as doenças cardiovasculares de forma equivalente em ambos os sexos (XU J, 2016).

Aliadas às doenças cardiovasculares, as doenças respiratórias fazem parte da hierarquia das principais causas de mortalidade de idosos no Brasil (KANSO et al, 2011). No presente estudo, estas representaram a segunda maior causa de óbito entre os centenários, se excluídas as causas mal definidas. Uma alta magnitude das taxas de mortalidade por doenças respiratórias encontradas entre os idosos centenários também foi verificada em outros estudos brasileiros (OLIVEIRA; MEDEIROS; LIMA, 2015; SOUZA et al, 2021) e internacionais (ROBINE; CUBAYNES, 2017). Na cidade de Hong Kong as causas respiratórias foram a principal causa de mortalidade entre os centenários (39,8%) (YU R, et al., 2017). A prevalência de causas respiratórias pode estar associada à maior susceptibilidade fisiológica e imunológica com o avançar da idade. Além disso, o alto percentual de óbitos por doenças respiratórias entre os centenários pode estar relacionado à maior frequência de acamados. Os indivíduos nesta condição podem desenvolver complicações relacionadas à fraqueza muscular diafragmática e intercostal, dificuldade em eliminar secreção, tosse menos eficaz e diminuição da função ciliar, resultando em acúmulo de secreção, predispondo-os assim a atelectasia e infecções respiratórias (OLIVEIRA; MEDEIROS; LIMA, 2015).

Uma das principais estratégias adotadas para o controle das doenças respiratórias se refere à imunização com a vacina contra o vírus *Influenza*, que contribui consideravelmente para a redução da mortalidade e dos gastos com internações e tratamentos. No Brasil, a vacina contra a gripe é direcionada aos idosos e portadores de doenças crônicas. Todavia, são necessárias adaptações do calendário de vacinação de acordo com as variações regiões de sazonalidade, para que a campanha de vacinação obtenha resultados ainda mais positivos, além de se considerar a circulação de outros patógenos de grande importância clínica e epidemiológica para as doenças do aparelho respiratório (GOMES et al, 2013).

As doenças endócrinas, metabólicas e nutricionais apresentaram importância semelhante entre os centenários do Semiárido brasileiro como causa de mortalidade, com percentual de aproximadamente 6%. Entre essas doenças, a literatura aponta o diabetes *mellitus* como aquela que detém a maior carga de óbitos, apresentando tendência de crescimento exponencial de sua prevalência com a transição nutricional e adquirindo proporções epidêmicas em vários países (FERREIRA; SAMPAIO, 2022). Segundo esses autores, as alterações no estilo de vida, com ênfase na alimentação e prática de atividades físicas, associadas ao aumento da esperança de vida dos brasileiros, são atualmente apontados como os principais fatores responsáveis pelo aumento da prevalência do diabetes tipo 2 observado no país. No entanto, acredita-se que há uma subestimação do diabetes *mellitus* como causa básica, devido ao mal preenchimento da declaração de óbito, na qual esta doença é muitas vezes erroneamente classificada como causa contribuinte (CASCÃO; COSTA; KALE, 2012).

A proporção de óbitos por doenças infecciosas e parasitárias (pouco mais de 5%) entre os centenários do Semiárido brasileiro não acompanha o perfil de morbidade da população mais jovem, que ainda apresenta elevadas taxas de tuberculose, hanseníase, malária e leishmaniose. Porém, comparando esse percentual de mortes por doenças infecciosas no Semiárido com outros estudos de mortalidade entre centenários percebe-se um percentual mais elevado no Semiárido, em outras pesquisas o percentual de óbitos ficou em torno de 1% (XU J, 2016; YU R, et al., 2017 SOUZA et al, 2021).

As causas externas, as doenças do aparelho digestivo, as neoplasias e as doenças do aparelho geniturinário apresentaram menor impacto no padrão de mortalidade da população estudada, com proporção total dessas causas reunidas inferior a 9%. No que se refere às duas primeiras causas, estudos revelam que produzem maior impacto entre os dados de mortalidade dos idosos mais jovens e do sexo masculino (OLIVEIRA; MEDEIROS; LIMA, 2015). Segundo esses autores há acentuada sobre mortalidade masculina para essas duas causas, sendo

as doenças do aparelho digestivo destacadamente representadas pela cirrose hepática (informada ou não com associação ao alcoolismo) e causas externas por homicídios e acidentes de transporte. Ademais, acredita-se que os indivíduos longevos, dentre eles os centenários, estão menos expostos aos fatores de risco relacionados a tais causas específicas.

As neoplasias representam a segunda causa de morte em relação às causas conhecidas no Nordeste na população geral (CORRÊA; RIBEIRO, 2017). No entanto, sua proporção é menor entre os centenários do Semiárido, ocupando apenas a sexta posição, devido à maior taxa óbitos por causas mal definidas e doenças respiratórias nesse grupo. Tendo em vista a longevidade populacional e o fato de que o envelhecimento constitui o principal fator de risco para o desenvolvimento das neoplasias, estas têm apresentado uma tendência de crescimento na população, estimando-se que em breve ultrapassará as doenças do aparelho circulatório em países desenvolvidos. Apesar da relevância, ainda é limitado o acesso dos pacientes idosos aos serviços de saúde de qualidade, inclusive para prevenção, informação e rastreamento clínico (GOMES et al, 2013; BONFIM; CAMARGOS, 2021)

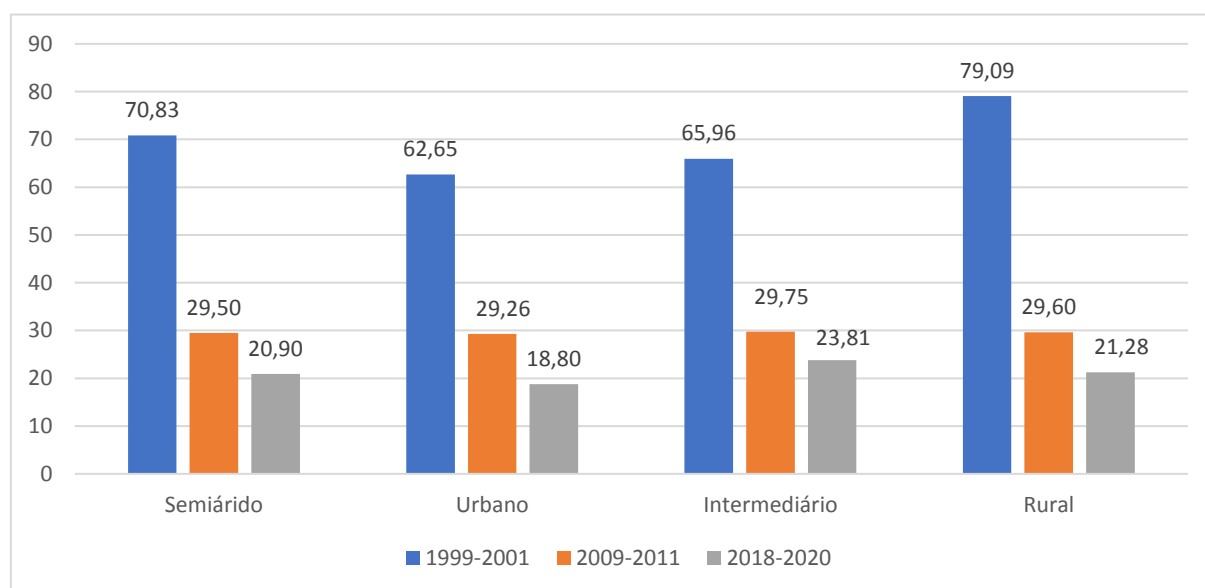
A alta proporção de óbitos por causas mal definidas evidenciada no estudo dos centenários prejudica a análise da distribuição da mortalidade segundo causas e a compreensão fidedigna do perfil de morbimortalidade, o que pode alterar a ordem real de prioridades baseada no impacto das diferentes causas, interferindo na alocação racional de recursos da saúde (OLIVEIRA; MEDEIROS; LIMA, 2015).

Segundo Cunha e colaboradores (2019) a distribuição dos óbitos classificados como causas mal definidas no Brasil é bastante irregular e se caracteriza por profundas variações na escala regional, apresentando a região Nordeste uma das mais altas proporções, principalmente na parte que pertence ao Semiárido. Segundo dados do Sistema de Informações sobre Mortalidade – SIM/DATASUS (2010), em 1980 a região Nordeste respondeu por 58,4% do total de óbitos registrados sob a rubrica de causas mal definidas, 54,5% em 1990, 47,8% no ano 2000, com um declínio considerado em 2010 para 10,7%. Apesar da melhoria e consolidação do SIM a partir da década de 2000, as altas proporções de óbitos por CMD no Brasil ainda são preocupantes, principalmente em regiões específicas como o Semiárido, como descrito na Tabela 4.2.2, cujo percentual de CMD entre os centenários foi de 28,5%, somando todos os óbitos ocorridos ao longo de 20 anos. A nível de comparação é importante destacar que no ano 2000, países como Hungria, Nova Zelândia, Finlândia, Cuba e Romênia já apresentavam

proporção de CMD menor que 1% (MATHERS et al, 2001), evidenciando-se, portanto, a elevada proporção de CMD existente no Brasil e mais preocupante ainda no Semiárido.

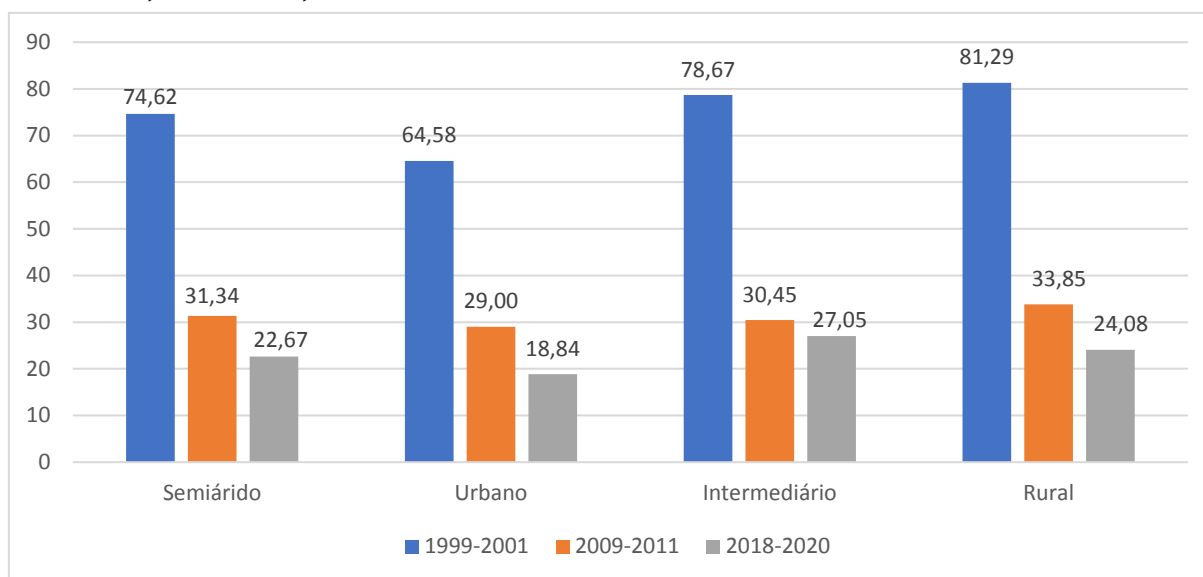
Com o objetivo de aprofundar o estudo das causas mal definidas (CMD) ao longo dos anos, as Figuras 4.2.1 e 4.2.2 representam o percentual das CMD entre os centenários, desagregados por sexo, para os espaços regionalizados do Semiárido brasileiro nos três triênios 1999-2001, 2009-2011, 2018-2020. Esse percentual é de grande importância pois evidencia a qualidade do registro no tocante ao preenchimento da variável Causa Básica do Óbito, imprescindível para o conhecimento do perfil epidemiológico e de mortalidade de uma população. Se a proporção das CMD está elevada, isto sinaliza que a mortalidade pelas outras causas está subestimada, o que pode comprometer seriamente as avaliações e análises da mortalidade por causas dos centenários.

Figura 4.2.1: Distribuição do percentual dos óbitos dos centenários por causa mal definida da Declaração de Óbito, entre os homens, nos espaços regionalizados do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.



Fonte: DATASUS/SIM/MS.

Figura 4.2.2: Distribuição do percentual dos óbitos dos centenários por causa mal definida da Declaração de Óbito, entre as mulheres, nos espaços regionalizados do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.



Fonte: DATASUS/SIM/MS.

Nos espaços regionalizados houve uma predominância de percentuais muito acima de 30% no primeiro triênio, o que realmente é um fato preocupante. Deste modo, segundo a classificação adotada por Chackiel (1987) a qualidade da variável causa básica pode ser categorizada predominantemente como *regular ou pouco confiável* (16 - 30%), porém em alguns períodos a qualidade foi atribuída como *deficiente* (acima 30%). Apenas a área urbana, no último triênio, em ambos os sexos, apresentou um percentual menor que 20%, porém de acordo com o critério adotado ainda foi classificada como *pouco confiável*.

O maior percentual de CMD foi identificado no espaço Rural, no primeiro triênio, entre as mulheres (81,2%) e o menor percentual de CMD foi encontrado no espaço Urbano, entre os homens, no último triênio (18,8%), o que evidencia, como já discutido nas tabelas anteriores, melhores registros vitais nos municípios urbanos, homens e a melhoria da qualidade ao longo dos anos.

Com relação ao Semiárido como um todo, entre os homens, no primeiro triênio a proporção de óbitos por CMD foi de 70,83%, classificado como *deficiente*, nos outros dois triênios foi categorizado como *regular*. Entre as mulheres, dois triênios foram classificados como *deficiente*, apenas o último triênio apresentou uma qualidade melhor da variável, sendo categorizado como *regular*. Esses resultados ajudam a discriminar com bastante clareza os espaços em que os sistemas públicos de captação das informações das causas básicas de morte demandam maiores prioridades em suas políticas de melhoria nas declarações dos óbitos, sem

descuidar, evidentemente, das demais áreas de atuação.

A captação de eventos de mortalidade no Brasil tanto na cobertura como na completude da causa básica de óbito tem apresentado importante avanço nas últimas décadas, pois a partir de 2010 houve uma mudança na legislação que regula o preenchimento no formulário das causas de morte, porém, ainda assim, o país tem sido classificado em perfis *intermediários* em pesquisas sobre cobertura e completude de eventos vitais (IBGE, 2022) e apresenta volume expressivo de registros com causas não definidas e problemas na notificação e no fluxo de informação de óbitos (MORAIS; COSTA, 2017).

Estudo do percentual de óbitos por CMD com centenários, que faleceram entre 2010 e 2019, foi realizado no município de Curitiba-PR, pertencente a região Sul do país, que evidenciou um percentual de apenas 2% de CMD do total de óbitos analisados (SOUZA et al, 2021) Esse resultado apresenta elevada disparidade com os resultados encontrados nos Gráficos 4.2.1 e 4.2.2 referentes aos centenários do Semiárido brasileiro. Já outro estudo realizado por Oliveira, Medeiros e Lima (2015), no estado do Rio Grande do Norte, pertencente a região Nordeste do país, com idosos longevos (80 anos ou mais) que faleceram entre 2001 e 2011, evidenciou um percentual de 17,27% de CMD, ou seja, percentual mais elevado do que foi encontrado no Paraná, mas ainda menor que os resultados encontrados no Semiárido brasileiro.

A situação das CMD dos centenários no Semiárido brasileiro encontra-se em patamar mais preocupante do que o encontrado no Brasil. Analisando-se dados sobre mortalidade de centenários brasileiros em 2010, observou-se que 21,5% dos óbitos referiam-se a causas mal definidas (PEDRO; BÓS, 2017). No Brasil, dentre os óbitos de centenários classificados como CMD no ano de 2010, 34,4% ocorreram sem assistência médica e 28,4% devido à senilidade (PEDRO; BÓS, 2017).

A mortalidade proporcional por CMD aumenta com a idade e tem sido menor na faixa etária de 60-69 anos e maior entre centenários segundo (PEDRO; BÓS, 2017; OLIVEIRA; MEDEIROS; LIMA, 2015; SOUZA et al, 2021). Esses autores ainda relataram que desde os anos 2000 muitos estudos internacionais já confirmavam esse perfil ao comparar causas registradas na DO com resultados de autópsias. Os fatores mais importantes associados ao erro de diagnóstico da causa básica são citados como a idade avançada e a incerteza do diagnóstico prévio. Portanto, existe uma relação considerável entre aumento de CMD e problemas de acesso aos serviços básicos de saúde, em que estados brasileiros com menor gasto per capita em saúde apresenta maiores taxas de mortalidade por CMD em idosos longevos (PEDRO; BÓS, 2017).

Para aprofundar a avaliação da qualidade dos registros de óbitos, analisou-se o

percentual de causas desconhecidas, que traz à tona uma perspectiva mais próxima da realidade do que simplesmente focando o olhar nas causas mal definidas. Deste modo, a Tabela 4.2.5 mostra os percentuais de óbitos dos centenários cuja causa básica é desconhecida de acordo com a definição adotada neste trabalho, nos espaços regionalizados do Semiárido brasileiro, entre homens e mulheres, nos três triênios estudados. Esse percentual procura expressar a totalidade das causas não declaradas pelo médico ou responsável pelo registro, o qual inclui todos os sub-registros e as causas mal definidas daqueles óbitos declarados.

Tabela 4.2.5: Distribuição do percentual de óbitos dos centenários cuja causa básica é desconhecida, segundo sexo, nos espaços regionalizados do Semiárido brasileiro, 1999-2001, 2009-2011, 2018-2020.

| Regiões       | Homens    |           |           | Mulheres  |           |           |
|---------------|-----------|-----------|-----------|-----------|-----------|-----------|
|               | 1999-2001 | 2009-2011 | 2018-2020 | 1999-2001 | 2009-2011 | 2018-2020 |
| Rural         | 83,50     | 37,96     | 29,36     | 85,64     | 42,93     | 33,03     |
| Intermediário | 72,70     | 37,22     | 29,33     | 83,29     | 38,78     | 34,40     |
| Urbano        | 68,90     | 35,18     | 23,85     | 70,82     | 35,84     | 25,02     |
| Semiárido     | 76,40     | 36,75     | 27,16     | 79,91     | 39,43     | 30,27     |

Fonte: DATASUS/SIM/MS.

Observando os resultados apresentados na Tabela 4.2.5 é evidente que os percentuais aumentaram ao serem considerados o conjunto dos dois fatores (sub-registro e causa mal definida), os quais sinalizam para uma deficiência elevada dos registros de óbitos, em ambos os sexos, no Semiárido como um todo e em todos os espaços regionalizados. Os percentuais ficaram acima 40% em todo o Semiárido e regiões, em ambos os sexos, no primeiro triênio, sendo classificadas como registros de óbitos *deficientes*. Essa categoria também foi enquadrada no segundo triênio, entre as mulheres da região Rural (42,93%). Os demais triênios, em ambos os sexos, no Semiárido e regiões, foram categorizados como qualidade dos registros de óbitos *regular ou pouco confiável*, pois apresentaram valores entre (25% e 40%), exceto entre os homens que residiam em área urbana no último triênio, estes apresentaram qualidade de registro *satisfatória* (23,85%).

Esses percentuais comparados com os dados apresentados por Paes (2018), no Brasil em 2010, para adultos, confirmam a elevada deficiência dos registros das causas de óbitos dos centenários no Semiárido brasileiro. Para o Brasil, em 2010, o percentual de óbitos por causas desconhecidas, em adultos, foi de 12,1% entre os homens e 11,5% entre as mulheres, valores muito abaixo do que foram encontrados aqui para o mesmo ano no Semiárido: 36,75% para os homens e de 39,43% para as mulheres.

Estudo realizado por Paes, Silva e Braga (2014), para adultos em geral, no Semiárido brasileiro, em 2010, concluiu que o percentual de causas mal definidas foi de 11,5% e o percentual de óbitos por causas desconhecidas foi de 23%. Esse percentual encontrado no referido estudo de causas mal definidas, em 2010, foi quase três vezes mais baixo do que foi encontrado, no mesmo ano, para os centenários do Semiárido, assim como o percentual de óbitos por causas desconhecidas na população adulta em geral ficou bem abaixo do que foi encontrado nos centenários, o que revela a enorme desvantagem na captação dessa informação dos centenários. Portanto, pode-se dizer que o percentual de causas mal definidas e desconhecidas para os centenários é superior ao dos adultos abaixo dos 100 anos de idade. Porém, analisando a evolução desses percentuais, no decorrer dos anos, para o Semiárido e seus espaços regionalizados, houve redução dos percentuais o que evidencia uma melhora na captação e registro dessas causas ao longo do período estudado.

Os resultados mostraram que houve um importante avanço no registro das informações na Declaração de Óbitos para o Semiárido brasileiro nos triênios estudados para a população dos centenários, fruto dos investimentos dos sistemas de informação de saúde brasileira. No entanto, ainda há muito espaço para melhorias, devido ao percentual elevado de causas desconhecidas, ainda para anos mais recentes.

Especula-se que as coberturas de óbitos dos centenários para o Semiárido brasileiro e seus espaços regionalizados não atingiram níveis acima de 90%. Portanto, para atingir o estado de excelência (100%), ainda há um caminho de evolução e melhorias para percorrer e, espera-se, que não seja em um futuro muito distante.

Portanto, diante do elevado percentual de CMD e para tentar solucionar o problema de incompletude das variáveis da DO e de CMD, na sessão a seguir é apresentada a implementação do método de imputação de dados faltantes e a redistribuição de Causas Mal Definidas.

### **4.3 Análise da completude das variáveis contidas nos registros de óbitos dos centenários do Semiárido brasileiro, nos triênios de 2000 a 2020.**

Nesta sessão, os resultados são apresentados conforme a ordem de tratamento das variáveis em três etapas: A primeira trata-se da análise dos registros de óbitos dos centenários pela completude das variáveis, ao longo de três triênios (2000-2002, 2009-2011, 2018-2020). A segunda se refere à correção da incompletude das variáveis realizando a implementação do método de Imputação Múltipla. Na terceira e última etapa foi realizada a redistribuição de óbitos por causas mal definidas (CMD).

#### **4.3.1 Completude das variáveis dos registros de óbitos**

A completude representa a magnitude ou nível de declaração de uma determinada variável informada quando o registro foi realizado no sistema. Problemas no preenchimento das variáveis são decorrentes de declarações errôneas ou omissões de informações referentes a essas variáveis (PAES, 2018). A completude é traduzida pela proporção de campos preenchidos com resposta não-nulas e a incompletude pelo percentual de informações ignoradas ou não preenchidas, também denominadas de *missings*. Portanto, nesta primeira etapa avaliou-se a qualidade das variáveis contidas nos registros de óbitos dos centenários, ao longo dos anos, verificando o percentual de incompletude das variáveis que foram utilizadas nesta investigação. A Tabela 4.3.1 apresenta um panorama da situação em termos do percentual de informações ignoradas ou não preenchidas das variáveis selecionadas na Declaração de Óbito dos centenários, por sexo, para os triênios de 2000 a 2020.

Tabela 4.3.1 Número e percentual de informações ignoradas ou não preenchidas segundo variáveis selecionadas da Declaração de Óbito dos centenários, por sexo, nos espaços regionalizados do Semiárido brasileiro, nos triênios de 2000 a 2020.

| Ano do óbito               |      | 2000-2002 |      |               |      |        |      | 2009-2011 |      |               |      |        |      | 2018-2020 |      |               |      |        |      |
|----------------------------|------|-----------|------|---------------|------|--------|------|-----------|------|---------------|------|--------|------|-----------|------|---------------|------|--------|------|
| Variável                   | Sexo | Rural     |      | Intermediário |      | Urbano |      | Rural     |      | Intermediário |      | Urbano |      | Rural     |      | Intermediário |      | Urbano |      |
|                            |      | n         | %    | n             | %    | n      | %    | n         | %    | n             | %    | n      | %    | n         | %    | n             | %    | n      | %    |
| <i>Faixa Etária</i>        | H    | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  |
|                            | M    | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  | 0         | 0,0  | 0             | 0,0  | 0      | 0,0  |
| <i>Raça/cor</i>            | H    | 21        | 17,8 | 8             | 19,5 | 32     | 31,7 | 88        | 13,5 | 27            | 11,2 | 76     | 15,2 | 60        | 4,0  | 20            | 4,1  | 45     | 5,6  |
|                            | M    | 41        | 19,0 | 12            | 17,4 | 34     | 22,2 | 126       | 11,5 | 52            | 12,9 | 125    | 12,2 | 96        | 4,0  | 23            | 2,8  | 65     | 3,5  |
| <i>Situação Conjugal</i>   | H    | 12        | 10,2 | 2             | 4,8  | 6      | 5,9  | 50        | 7,7  | 19            | 7,9  | 44     | 8,8  | 133       | 8,8  | 39            | 8,0  | 95     | 10,1 |
|                            | M    | 11        | 5,1  | 6             | 8,7  | 6      | 3,9  | 97        | 8,9  | 35            | 8,6  | 113    | 10,0 | 237       | 9,9  | 79            | 9,5  | 177    | 9,7  |
| <i>Escolaridade</i>        | H    | 36        | 30,5 | 10            | 24,4 | 45     | 44,6 | 150       | 23,0 | 64            | 26,0 | 145    | 29,1 | 251       | 16,6 | 91            | 18,7 | 178    | 19,0 |
|                            | M    | 73        | 33,7 | 28            | 40,6 | 52     | 34,0 | 247       | 22,5 | 93            | 23,0 | 266    | 26,0 | 405       | 16,9 | 143           | 17,3 | 397    | 21,7 |
| <i>Assistência médica?</i> | H    | 96        | 81,4 | 39            | 95,1 | 75     | 74,3 | 554       | 85,0 | 198           | 82,0 | 418    | 83,8 | 830       | 54,8 | 271           | 55,6 | 527    | 56,3 |
|                            | M    | 172       | 79,6 | 52            | 75,4 | 119    | 77,8 | 902       | 82,3 | 345           | 85,0 | 830    | 81,1 | 1362      | 56,8 | 464           | 56   | 992    | 54,2 |
| <i>Local de ocorrência</i> | H    | 4         | 3,4  | 0             | 0,0  | 3      | 3,0  | 1         | 0,2  | 0             | 0,0  | 1      | 0,2  | 1         | 0,1  | 0             | 0,0  | 0      | 0,0  |
|                            | M    | 2         | 0,9  | 0             | 0,0  | 1      | 0,6  | 1         | 0,1  | 0             | 0,0  | 1      | 0,1  | 1         | 0,0  | 2             | 0,2  | 0      | 0,0  |

Fonte: Sistema de Informações sobre Mortalidade – Ministério da Saúde.

Legenda: H: Homens M: Mulheres.

Observa-se na Tabela 4.3.1 que a variável *Faixa Etária* não apresentou dados ausentes ou ignorado, o que se deve ao fato que durante a coleta de microdados do SIM foi utilizado como filtro os registros de óbitos dos indivíduos com 100 anos ou mais declarados ou calculados pelo sistema após verificação ou checagem.

Avaliando a variável *Raça/cor* foi possível observar que os valores de dados ausentes ou ignorados variaram entre 4,0-31,7% entre os homens e 2,8-22,2% entre as mulheres, portanto a incompletude dessa variável foi maior entre os homens, principalmente aqueles que residiam em áreas urbanas, porém para ambos os sexos houve um importante descenso da incompletude ao longo dos anos.

A variável *Situação Conjugal* também apresentou maior percentual de *missings* entre os homens com variação de 4,8-10,2 e 3,9-10,0% entre as mulheres. Um fato preocupante e que merece destaque é o aumento do percentual de registros ausentes dessa variável ao longo dos anos, para ambos os sexos, exceto os homens residentes em área rural entre os dois primeiros triênios quando houve uma diminuição no percentual. Em todas as outras situações o acréscimo de pontos percentuais foi marcante no decorrer do período, sinalizando um piora na captação dessa informação, sendo a única, de um modo geral, que andou na contramão da tendência das demais.

O registro da *Escolaridade*, apesar de elevados percentuais sobretudo nos primeiros triênios, melhorou de forma expressiva ao longo dos anos. O percentual de dados ausentes também foi maior entre os homens, com uma variação de 16,6-44,6%, já entre as mulheres a variação foi 16,9-40,6%.

Ainda avaliando a variável *Escolaridade* os municípios classificados como Urbanos, apresentaram maiores percentuais de dados ausentes, situação análoga encontrada nas variáveis *Raça/cor* e *Situação Conjugal*. A presença de *missings* foi superior ou próximo a 30% nos dois primeiros triênios, em ambos os sexos, o que dificulta uma análise com mais consistência, originada pela negligência com as informações socioeconômicas, em especial o nível de instrução quando do preenchimento da DO pelo médico, profissional responsável por esse registro.

As deficiências no preenchimento do formulário são essencialmente um fenômeno sociocultural, que depende do contexto onde o médico está inserido e trabalha (MENDONÇA; DRUMOND; CARDOSO, 2010). Nesse sentido, a escolaridade é um dos itens mais negligenciados no seu preenchimento e de difícil recuperação (BARBUSCIA; RODRIGUES-

JÚNIOR, 2011). Estudos evidenciam que os patologistas negligenciam mais esse preenchimento que o clínico geral (STUQUE; CORDEIRO; CURY, 2003).

A variável *Assistência Médica* (durante a doença que ocasionou o óbito) apresentou os percentuais mais elevados de incompletude durante o período analisado e em comparação a todas as outras variáveis estudadas. Os percentuais de *missings* foram acima de 50% em todas as situações analisadas. O maior percentual encontrado foi entre os homens que residiam em municípios caracterizados como Intermediário (95,1%) no primeiro triênio, já os menores valores foram encontrados entre as mulheres, residentes em área urbana, no último triênio (54,2%). Apesar dos elevados percentuais, houve melhora do registro dessa variável ao longo dos anos. Sem embargo, os percentuais de dados ausentes nessa variável, no triênio mais recente, foram tão elevados, acima de 50%, que inviabiliza qualquer tentativa de recuperação desses dados. Sendo assim, essa variável ficou excluída do processo de imputação dos dados na etapa seguinte e não será introduzida no método de imputação múltipla abordado na Sessão 4.3.2.

Por fim, uma variável bem declarada, com valores quase equivalentes a 0 “zero” e com percentuais menores comparadas as demais variáveis sociodemográficas, foi *Local de ocorrência* do óbito que apresentou poucos dados ausentes ou ignorados. O percentual diminuiu ao longo dos anos, tendo as mulheres menores percentuais de dados ausentes, principalmente aquelas que residiam em área urbana ou intermediária.

Estudo realizado com idosos longevos no Nordeste brasileiro, entre 2001 e 2015, também evidenciou maior incompletude das variáveis sociodemográficas entre os homens, assim como a variável *Escolaridade* apresentou maiores percentuais de dados ausentes, comparadas às variáveis *Raça/cor* e *Estado Civil* (MENDES, 2018). Medeiros (2015) realizou uma análise da mortalidade em idosos longevos no Brasil, entre 2000 e 2010, cujos resultados corroboram com os percentuais de dados ausentes tanto para a variável *Assistência Médica*, com percentual de incompletude também acima de 50%, como a variável *Local de ocorrência* do óbito com os menores percentuais analisados. Em ambos os estudos, houve melhora da completude das variáveis ao longo dos anos, assim como foi evidenciado na maioria das variáveis estudadas entre os centenários.

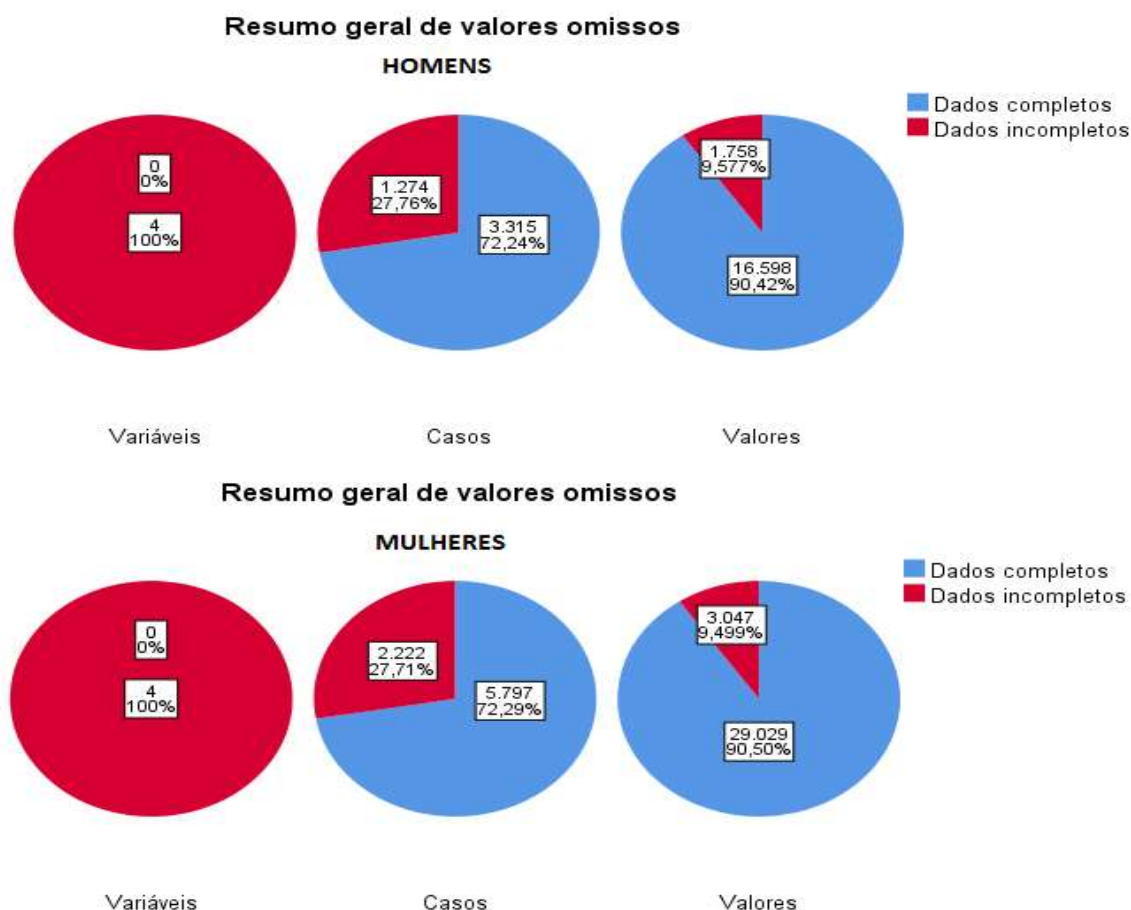
#### **4.3.2 Implementação do Método de Imputação Múltipla**

Após a avaliação da completude das variáveis sociodemográficas dos registros de óbitos dos centenários e a identificação da existência de dados ausentes no período de 2000 a 2020

verificou-se que a variável *Faixa Etária* teve uma incompletude de 0,0% e, por sua vez, a variável *Assistência Médica* apontou uma incompletude acima de 50,0%, ambas em todo o período analisado. Desta forma, essas duas variáveis foram excluídas da aplicação da imputação múltipla, cujo propósito consiste na obtenção de resultados mais verossímeis e dados completamente preenchidos. O processo de imputação múltipla se sustenta em métodos estatísticos inferenciais, já descrito na Sessão 3.4.2.

As variáveis submetidas a imputação dos dados foram: *Raça/cor*, *Situação conjugal* (Estado Civil), *Escolaridade* e *Local de ocorrência do óbito*. Porém, antes de proceder com o método de imputação é necessário compreender o padrão e o mecanismo de dados ausentes das variáveis estudadas. A Figura 4.3.1 apresenta a proporção de dados faltantes, segregados por sexo, nos três triênios estudados do Semiárido brasileiro, para as variáveis entre os casos, que são os indivíduos, e para os valores, que representam as variáveis. Todas as variáveis possuem dados faltantes, correspondendo, aproximadamente, entre homens e mulheres, a 27,7% do total de casos, sendo 9,5% dos valores.

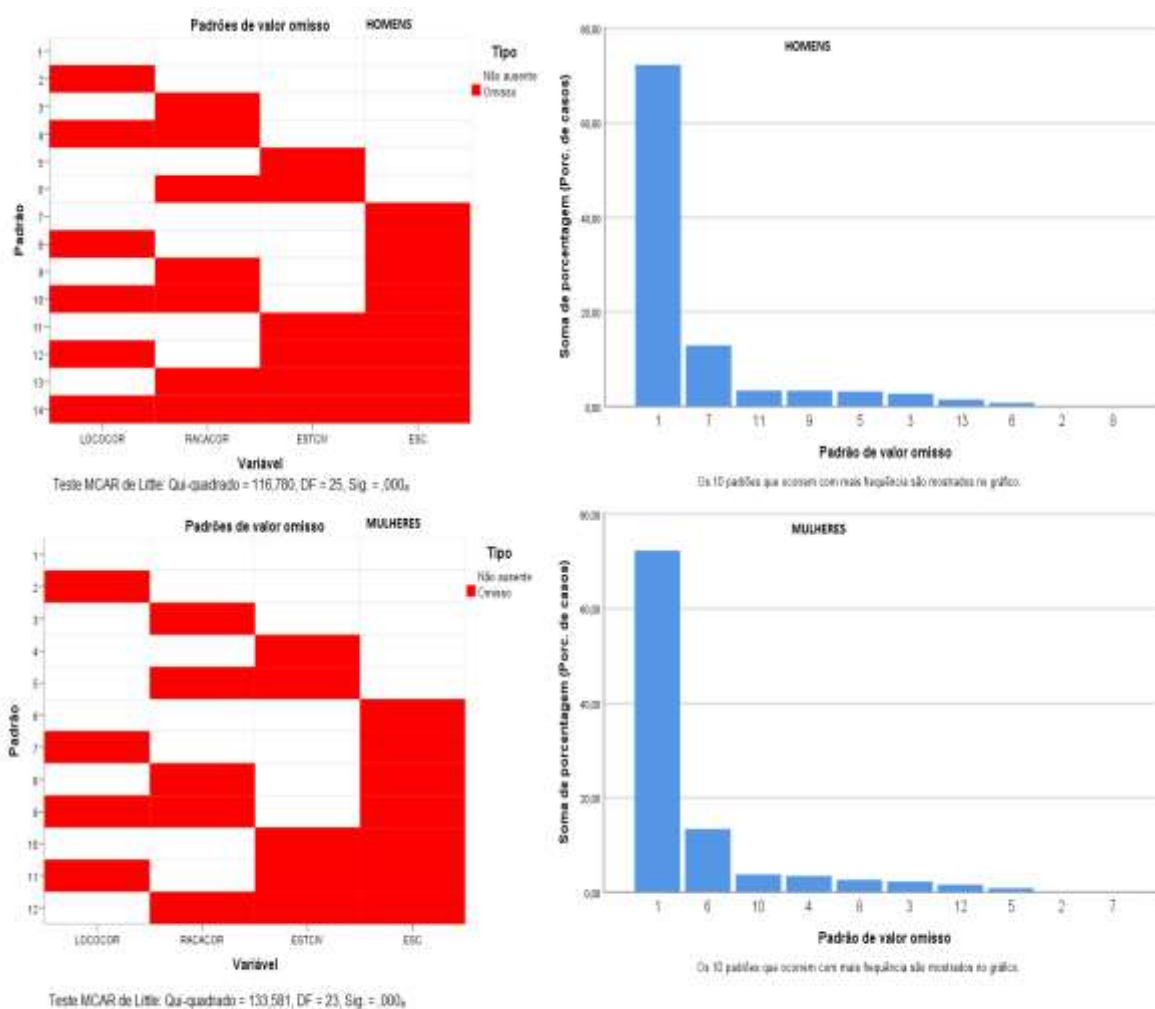
Figura 4.3.1 – Número e Proporção de dados faltantes em diferentes aspectos do banco de dados.



Fonte: Sistema de Informações sobre Mortalidade (SIM).

Foram identificados 14 padrões de não resposta diferentes entre os homens e 12 padrões de não resposta diferentes entre as mulheres, nas variáveis da base de dados analisada, como mostra a Figura 4.3.2. Como complemento a esta informação, exibe-se o percentual de casos em cada padrão e o teste de Little que analisa o Mecanismo Completamente Não Aleatório (MCAR). Nesse caso, não foi observado que houvesse uma tendência que sugerisse a existência de um padrão específico em cada variável, ou que o fato da não resposta em algum item gere também a não resposta em outro item.

Figura 4.3.2 - Padrões de não resposta das variáveis do estudo, entre homens e mulheres, gráficos de barras com percentagem de casos em cada padrão, Teste MCAR de Little.



Fonte: Sistema de Informações sobre Mortalidade (SIM).

De acordo com a classificação proposta por Little e Rubin (2019), ficou evidente que o *padrão* dos dados faltantes foi do tipo “monotônico”, já que dados faltantes foram observados

em todas as variáveis estudadas. Segundo Enders (2011) esse *padrão* é considerado o mais frequente na prática. Quanto ao *mecanismo* gerador dos dados ausentes no banco de dados estudado, conforme valores do p-valor = 0,000 do teste Qui-Quadrado, em ambos os sexos, apresentou-se como significativo ao nível de 5%, indicando que o mecanismo de não-resposta foi considerado como completamente aleatório (MCAR), isto é, os valores ausentes não estavam relacionados com seus valores anteriores ou posteriores, e nem com qualquer outra variável da amostra (ENDERS, 2011).

Para aplicar a técnica de imputação múltipla fez-se necessário decidir pelo número de imputações que foram estabelecidas por meio do cálculo da Eficiência Relativa (ER). Na Tabela 4.3.2 são observados os valores da Eficiência Relativa das variáveis com observações faltantes que foram utilizadas na modelagem estatística para os três tipos de espaços regionalizados por sexo. Optando-se por cinco imputações, a Eficiência Relativa para todas as variáveis desse estudo apresentou no mínimo um percentual de 91,8% (Homens, Espaço regionalizado – Urbano, triênio 2000-2002) para a variável *Escolaridade* e um valor máximo de 99,9% para a variável *Local de Ocorrência* do óbito (entre homens e mulheres, no triênio 2018-2020). Já para as variáveis com ER de 100%, conseqüentemente, não foram necessárias realizar a imputação. Os altos percentuais da Eficiência Relativa fundamentaram a opção pelas cinco imputações, conforme proposto por Rubin (1987).

Tabela 4.3.2 - Eficiência Relativa segundo variáveis sociodemográficas selecionadas de acordo com a escolha de cinco imputações, espaços regionalizados do Semiárido brasileiro, 2000 a 2020.

| Ano do óbito               |      | 2000-2002 |         |        | 2009-2011 |         |        | 2018-2020 |         |        |
|----------------------------|------|-----------|---------|--------|-----------|---------|--------|-----------|---------|--------|
| Variável                   | Sexo | Rural     | Interm. | Urbano | Rural     | Interm. | Urbano | Rural     | Interm. | Urbano |
|                            |      | ER        | ER      | ER     | ER        | ER      | ER     | ER        | ER      | ER     |
| <i>Raça/cor</i>            | H    | 96,6      | 96,2    | 94,0   | 97,4      | 97,8    | 97,0   | 99,2      | 99,2    | 98,9   |
|                            | M    | 96,3      | 96,6    | 95,7   | 97,8      | 97,5    | 97,6   | 99,2      | 99,4    | 99,3   |
| <i>Situação Conjugal</i>   | H    | 98,0      | 99,0    | 98,8   | 98,5      | 98,4    | 98,3   | 98,3      | 98,4    | 98,0   |
|                            | M    | 99,0      | 98,3    | 99,2   | 98,3      | 98,3    | 97,8   | 98,1      | 98,1    | 98,1   |
| <i>Escolaridade</i>        | H    | 94,3      | 95,3    | 91,8   | 95,6      | 95,1    | 94,5   | 96,8      | 96,4    | 96,3   |
|                            | M    | 93,7      | 92,5    | 93,6   | 95,7      | 95,6    | 95,1   | 96,7      | 96,7    | 95,8   |
| <i>Local de ocorrência</i> | H    | 99,3      | 100,0   | 99,4   | 99,9      | 100,0   | 99,9   | 99,9      | 100,0   | 100,0  |
|                            | M    | 99,8      | 100,0   | 99,9   | 99,9      | 100,0   | 99,9   | 100,0     | 99,9    | 100,0  |

Fonte: Elaboração própria.

Legenda: H: Homens, M: Mulheres; ER: Eficiência Relativa

Após a verificação do *padrão* e do *mecanismo* de dados ausentes procedeu-se a implementação da imputação múltipla gerando cinco bancos de dados completos utilizando o

Método da Regressão Linear para gerar os valores imputados e ao final da imputação realizou-se a combinação dos resultados por meio das Regras de Rubin (RUBIN, 1987). Nas tabelas 4.3.3 e 4.3.4 estão apresentadas o banco completo das categorias das variáveis do estudo, por sexo, nos triênios estudados, com a distribuição do percentual antes e depois da imputação múltipla.

Observando a Tabela 4.3.3, entre homens centenários, que faleceram no Semiárido brasileiro, a variável *Raça/Cor parda*, após a imputação, foi a mais frequente nas áreas e triênios estudados, seguida da *branca e preta*, exceto na área Urbana do primeiro triênio, no qual a *Raça/cor* mais frequente foi a *branca* (51,5%), seguida da *parda* (39,6%) e *preta* (8,9%). A variável *Situação Conjugal*, após a imputação, apresentou maiores registros de óbitos entre os *viúvos*, seguidos dos *casados e solteiros*, em todas as áreas e triênios. Já as variáveis *Escolaridade e Local de Ocorrência* do óbito dos centenários as categorias *Sem Escolaridade* e que faleceram no próprio *Domicílio* foram maioria entre os registros após a imputação, em todos os triênios e áreas estudadas.

Observando a Tabela 4.3.4, entre mulheres centenárias, que faleceram no Semiárido brasileiro, a variável *Raça/Cor parda*, após a imputação, foi a mais frequente nas áreas e triênios estudados, seguida da *branca e preta*, exceto na área Intermediária e Urbana, do primeiro triênio, nas quais a *Raça/cor* mais frequente, respectivamente, foi a *branca* (49,3 e 49,7%), seguida da *parda* (40,6 e 39,2%) e *preta* (10,1 e 11,1%).

A variável *Situação conjugal*, após a imputação, apresentou maiores registros de óbitos entre as *viúvas*, seguidas das *casadas e solteiras*, na maioria das áreas e triênios, – exceto entre as mulheres residentes da área intermediária, do primeiro triênio, na qual após a imputação os registros mais frequentes foram as *solteiras* (52,2%), seguidas das *viúvas* (42%) e *casadas* (5,8%).

Já as variáveis *Escolaridade e Local de ocorrência do óbito*, assim como foi observado entre os homens, as centenárias *Sem escolaridade* e que faleceram no próprio *Domicílio* foram maioria entre os registros após a imputação, em todos os triênios e áreas estudadas.

Tabela 4.3.3 - Distribuição do percentual das categorias das variáveis sociodemográficas, antes e depois da imputação de dados, entre os homens, segundo espaços regionalizados, nos triênios de 2000 a 2020.

| Ano do óbito               | 2000-2002              |        |               |        |        |        | 2009-2011 |        |               |        |        |        | 2018-2020 |        |               |        |        |        |
|----------------------------|------------------------|--------|---------------|--------|--------|--------|-----------|--------|---------------|--------|--------|--------|-----------|--------|---------------|--------|--------|--------|
|                            | Espaços Regionalizados |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
|                            | Rural                  |        | Intermediário |        | Urbano |        | Rural     |        | Intermediário |        | Urbano |        | Rural     |        | Intermediário |        | Urbano |        |
| Variáveis                  | Antes                  | Depois | Antes         | Depois | Antes  | Depois | Antes     | Depois | Antes         | Depois | Antes  | Depois | Antes     | Depois | Antes         | Depois | Antes  | Depois |
| <b>Raça/cor</b>            |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Branca                     | 33,9                   | 39,0   | 39,0          | 43,9   | 39,6   | 51,4   | 32,5      | 34,2   | 33,5          | 37,2   | 32,7   | 36,7   | 28,2      | 28,9   | 32,6          | 33,3   | 29,1   | 30,1   |
| Preta                      | 7,6                    | 16,1   | 9,8           | 19,5   | 1,0    | 8,9    | 8,1       | 9,0    | 7,0           | 9,1    | 6,8    | 9,2    | 8,8       | 9,2    | 7,3           | 7,6    | 7,4    | 7,9    |
| Amarela                    | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,3       | 4,9    | 0,0           | 0,0    | 0,8    | 2,6    | 0,3       | 0,9    | 0,6           | 2,3    | 0,5    | 1,6    |
| Parda                      | 40,7                   | 44,9   | 31,7          | 36,6   | 27,7   | 39,7   | 45,4      | 47,1   | 48,3          | 53,7   | 44,3   | 49,9   | 58,2      | 59,7   | 55,2          | 56,1   | 58,1   | 60,0   |
| Indígena                   | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,2       | 4,8    | 0,0           | 0,0    | 0,2    | 1,6    | 0,5       | 1,3    | 0,2           | 0,8    | 0,2    | 0,3    |
| <b>Situação Conjugal</b>   |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Solteiro                   | 15,3                   | 17,8   | 22,0          | 22,0   | 19,8   | 21,8   | 21,8      | 23,6   | 19,4          | 19,8   | 20,1   | 22,0   | 12,8      | 14,1   | 15,4          | 16,8   | 19,8   | 21,6   |
| Casado                     | 27,1                   | 30,5   | 29,3          | 31,7   | 27,7   | 28,7   | 23,3      | 25,6   | 25,2          | 26,0   | 22,2   | 23,6   | 21,4      | 23,1   | 19,7          | 21,4   | 20,3   | 22,3   |
| Viúvo                      | 45,7                   | 48,3   | 41,5          | 41,5   | 45,5   | 46,5   | 46,4      | 49,5   | 46,7          | 48,8   | 48,1   | 50,3   | 53,3      | 57,8   | 52,4          | 55,4   | 46,5   | 50,4   |
| Separado/Divorciado        | 0,0                    | 0      | 0,0           | 0,0    | 0,0    | 0,0    | 0,0       | 0,0    | 0,4           | 3,3    | 0,4    | 2,2    | 0,7       | 1,5    | 1,2           | 2,1    | 1,1    | 2,2    |
| União Estável              | 1,7                    | 3,4    | 2,4           | 4,9    | 1,1    | 3,0    | 0,8       | 1,2    | 0,4           | 2,1    | 0,4    | 1,8    | 3,0       | 3,4    | 3,3           | 4,3    | 2,2    | 3,4    |
| <b>Escolaridade</b>        |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Sem escolaridade           | 52,5                   | 59,3   | 53,7          | 63,4   | 31,7   | 40,6   | 59,5      | 68,7   | 57,9          | 68,2   | 51,7   | 63,9   | 64,3      | 73,5   | 60,8          | 70,6   | 54,8   | 65,6   |
| 1 – 3 anos                 | 12,7                   | 16,1   | 19,5          | 24,4   | 16,7   | 27,7   | 13,8      | 17,2   | 12,0          | 15,3   | 14,0   | 19,8   | 13,3      | 15,5   | 13,1          | 15,4   | 15,6   | 18,8   |
| 4 – 7 anos                 | 2,5                    | 6,7    | 0,0           | 0,0    | 2,0    | 9,9    | 2,9       | 7,8    | 3,3           | 11,6   | 3,8    | 8,4    | 4,6       | 5,6    | 6,6           | 8,8    | 7,9    | 9,8    |
| 8-11 anos                  | 0,8                    | 10,2   | 0,0           | 0,0    | 4,0    | 15,8   | 0,6       | 3,8    | 0,0           | 0,0    | 1,2    | 5,2    | 0,9       | 3,7    | 0,6           | 3,3    | 1,8    | 3,3    |
| 12 anos ou mais            | 0,8                    | 7,6    | 2,4           | 12,2   | 1,0    | 5,9    | 0,2       | 2,5    | 0,4           | 5,0    | 0,2    | 2,6    | 0,3       | 1,8    | 0,2           | 1,8    | 0,9    | 2,5    |
| <b>Local de Ocorrência</b> |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Hospital                   | 13,6                   | 14,4   | 9,8           | 9,8    | 18,8   | 19,8   | 23,0      | 23,0   | 22,7          | 22,7   | 35,1   | 35,1   | 31,3      | 31,3   | 39,4          | 39,4   | 40,6   | 40,6   |
| Outros est. de saúde       | 0,0                    | 0,0    | 0,0           | 0,0    | 1,0    | 2,0    | 0,8       | 1,0    | 76,9          | 76,9   | 0,8    | 0,8    | 0,9       | 0,9    | 0,4           | 0,4    | 2,1    | 2,1    |
| Domicílio                  | 82,2                   | 83,1   | 82,9          | 82,9   | 77,2   | 78,2   | 75,2      | 75,2   | 0,0           | 0,0    | 62,9   | 62,9   | 66,7      | 66,7   | 59,5          | 59,5   | 55,2   | 55,2   |
| Via pública                | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,5       | 0,5    | 0,0           | 0,0    | 0,2    | 0,4    | 0,3       | 0,4    | 0,0           | 0,0    | 0,3    | 0,3    |
| Outros                     | 0,8                    | 2,5    | 7,3           | 7,3    | 0,0    | 0,0    | 0,3       | 0,3    | 0,4           | 0,4    | 0,8    | 0,8    | 0,7       | 0,7    | 0,6           | 0,6    | 1,7    | 1,7    |
| Total                      | -                      | 100    | -             | 100    | -      | 100    | -         | 100    | -             | 100    | -      | 100    | -         | 100    | -             | 100    | -      | 100    |

Fonte: Elaboração própria

Tabela 4.3.4 - Distribuição do percentual das categorias das variáveis sociodemográficas, antes e depois da imputação de dados, entre as mulheres, segundo espaços regionalizados, nos triênios de 2000 a 2020.

| Ano do óbito               | 2000-2002              |        |               |        |        |        | 2009-2011 |        |               |        |        |        | 2018-2020 |        |               |        |        |        |
|----------------------------|------------------------|--------|---------------|--------|--------|--------|-----------|--------|---------------|--------|--------|--------|-----------|--------|---------------|--------|--------|--------|
|                            | Espaços Regionalizados |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
|                            | Rural                  |        | Intermediário |        | Urbano |        | Rural     |        | Intermediário |        | Urbano |        | Rural     |        | Intermediário |        | Urbano |        |
| Variáveis                  | Antes                  | Depois | Antes         | Depois | Antes  | Depois | Antes     | Depois | Antes         | Depois | Antes  | Depois | Antes     | Depois | Antes         | Depois | Antes  | Depois |
| <b>Raça/cor</b>            |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Branca                     | 31,9                   | 38,4   | 40,6          | 49,3   | 41,8   | 49,7   | 29,9      | 33,0   | 34,9          | 36,6   | 33,2   | 36,5   | 31,4      | 32,5   | 35,8          | 36,2   | 35,2   | 35,9   |
| Preta                      | 9,7                    | 14,4   | 7,2           | 10,1   | 7,2    | 11,1   | 9,2       | 10,9   | 5,9           | 10,4   | 6,7    | 8,7    | 8,8       | 9,2    | 8,0           | 8,1    | 7,4    | 7,5    |
| Amarela                    | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,5       | 1,7    | 0,2           | 3,7    | 0,5    | 3,7    | 0,7       | 1,0    | 0,2           | 1,1    | 0,3    | 1,0    |
| Parda                      | 39,4                   | 47,2   | 34,8          | 40,6   | 28,8   | 39,2   | 48,8      | 53,5   | 45,8          | 47,5   | 47,4   | 51,1   | 54,8      | 56,5   | 52,7          | 53,1   | 53,5   | 54,7   |
| Indígena                   | 0,0                    | 0,0    | 0             | 0,0    | 0,0    | 0,0    | 0,1       | 0,8    | 0,2           | 1,7    | 0,0    | 0,0    | 0,3       | 0,8    | 0,5           | 1,4    | 0,1    | 0,9    |
| <b>Situação Conjugal</b>   |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Solteiro                   | 35,2                   | 37,0   | 47,8          | 52,2   | 30,7   | 31,4   | 33,7      | 35,8   | 31,2          | 33,7   | 34,0   | 35,8   | 24,6      | 26,9   | 24,0          | 26,1   | 26,7   | 28,9   |
| Casado                     | 8,8                    | 10,2   | 4,3           | 5,8    | 7,2    | 7,8    | 6,7       | 8,0    | 6,2           | 6,9    | 3,5    | 6,0    | 3,9       | 4,5    | 3,6           | 4,3    | 2,9    | 3,7    |
| Viúvo                      | 50,9                   | 52,8   | 39,2          | 42,0   | 58,2   | 60,8   | 50,4      | 52,6   | 53,5          | 56,4   | 51,2   | 55,0   | 60,5      | 66,2   | 61,4          | 65,5   | 59,3   | 63,7   |
| Separado/Divorciado        | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,1       | 1,0    | 0,5           | 3,0    | 0,2    | 2,1    | 0,3       | 0,9    | 0,6           | 1,7    | 0,6    | 1,7    |
| União Estável              | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,2       | 2,6    | 0,0           | 0,0    | 0,1    | 1,1    | 0,8       | 1,5    | 0,9           | 2,4    | 0,8    | 2,0    |
| <b>Escolaridade</b>        |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Sem escolaridade           | 60,2                   | 69,0   | 56,6          | 75,4   | 47,7   | 58,8   | 67,8      | 79,7   | 65,6          | 73,3   | 58,0   | 69,8   | 70,9      | 82,4   | 68,0          | 74,8   | 58,9   | 66,8   |
| 1 – 3 anos                 | 4,2                    | 13,9   | 1,4           | 11,6   | 13,7   | 20,3   | 8,0       | 11,5   | 8,2           | 14,1   | 10,8   | 14,3   | 8,3       | 9,9    | 9,2           | 11,5   | 12,1   | 19,3   |
| 4 – 7 anos                 | 0,9                    | 8,8    | 1,4           | 13,0   | 3,3    | 12,4   | 1,4       | 6,0    | 3,0           | 9,2    | 3,7    | 7,7    | 3,1       | 4,4    | 4,8           | 7,1    | 4,6    | 7,0    |
| 8-11 anos                  | 0,5                    | 3,7    | 0,0           | 0,0    | 1,3    | 8,5    | 0,3       | 2,8    | 0,2           | 3,5    | 0,8    | 4,7    | 0,7       | 2,3    | 0,6           | 5,0    | 2,5    | 4,8    |
| 12 anos ou mais            | 0,5                    | 4,6    | 0,0           | 0,0    | 0,0    | 0,0    | 0         | 0,0    | 0,0           | 0,0    | 0,7    | 3,5    | 0,1       | 1,0    | 0,1           | 1,7    | 0,2    | 2,1    |
| <b>Local de Ocorrência</b> |                        |        |               |        |        |        |           |        |               |        |        |        |           |        |               |        |        |        |
| Hospital                   | 5,6                    | 6,0    | 5,8           | 5,8    | 19,0   | 19,0   | 17,9      | 17,9   | 21,0          | 21,0   | 30,6   | 30,6   | 27,6      | 27,6   | 32,9          | 33,0   | 36,4   | 36,4   |
| Outros est. de saúde       | 0,0                    | 0,0    | 0,0           | 0,0    | 0,0    | 0,0    | 0,5       | 0,5    | 0,2           | 0,2    | 0,4    | 0,4    | 1,5       | 1,5    | 0,8           | 0,8    | 3,5    | 3,5    |
| Domicílio                  | 92,1                   | 92,1   | 94,2          | 94,2   | 80,4   | 81,0   | 80,9      | 81,0   | 77,5          | 77,5   | 68,0   | 68,1   | 70,1      | 70,1   | 64,9          | 64,9   | 59,1   | 59,1   |
| Via pública                | 0,9                    | 1,4    | 0,0           | 0,0    | 0,0    | 0,0    | 0,4       | 0,4    | 0,5           | 0,5    | 0,2    | 0,2    | 0,2       | 0,2    | 0,2           | 0,4    | 0,2    | 0,2    |
| Outros                     | 0,5                    | 0,5    | 0,0           | 0,0    | 0,0    | 0,0    | 0,3       | 0,3    | 0,7           | 0,7    | 0,8    | 0,8    | 0,6       | 0,6    | 1,0           | 1,0    | 0,8    | 0,8    |
| Total                      | -                      | 100    | -             | 100    | -      | 100    | -         | 100    | -             | 100    | -      | 100    | -         | 100    | -             | 100    | -      | 100    |

Fonte: Elaboração própria

Ao serem comparados os valores da distribuição das frequências para cada categoria das variáveis sociodemográficas antes e depois da imputação de dados, observou-se que não foram exatamente as mesmas, conforme mostrado nas Tabelas 4.3.3 e 4.3.4. Por exemplo, a maior diferença de percentual antes e depois da imputação para os homens, da variável *Raça/cor* foi a categoria *parda*, no triênio 2000-2002, na área Urbana, que respondeu por 27,7% da frequência antes da imputação, enquanto que depois essa frequência passou para 39,7%.

Da variável *Situação conjugal* foi a categoria *viúvo*, no triênio 2018-2020, na área Rural, que respondeu por 53,3% da frequência antes da imputação, enquanto que depois essa frequência passou para 57,8%.

A variável *Escolaridade* foi a que apresentou maiores diferenças nos percentuais antes e depois da imputação, devido ao elevado número de *missing* evidenciado antes da imputação, sendo a categoria *sem escolaridade*, no triênio 2009-2011, na área Urbana, que apresentou maior percentual de diferença, respondendo por 51,7% da frequência antes da imputação, enquanto que depois essa frequência passou para 63,9%.

A variável *Local de ocorrência* do óbito foi a que apresentou menos diferenças antes e depois da imputação, devido ao reduzido número de *missing*.

A maior diferença de percentual antes e depois da imputação para as mulheres, da variável *Raça/cor* foi a categoria *parda*, no triênio 2000-2002, na área Urbana, que respondeu por 28,8% da frequência antes da imputação, enquanto que depois essa frequência passou para 39,2%.

Da variável *Situação conjugal* foi a categoria *viúva*, no triênio 2018-2020, na área Rural, que respondeu por 60,5% da frequência antes da imputação, enquanto que depois essa frequência passou para 66,2%.

A variável *Escolaridade* também apresentou maiores diferenças de percentuais entre as mulheres, sendo a categoria *sem escolaridade*, no triênio 2009-2011, na área Intermediária, que apresentou maior percentual de diferença, respondendo por 56,6% da frequência antes da imputação, enquanto que depois essa frequência passou para 75,4%.

A variável *Local de ocorrência* do óbito também foi a que apresentou menos diferenças antes e depois da imputação, devido ao reduzido número de *missing*.

Estudo realizado por Paes, Santos e Coutinho (2021), fez uso da combinação das técnicas de *linkage* determinístico e da imputação múltipla para atingir a completude das informações do banco de dados sobre a mortalidade neonatal nos municípios rurais adjacentes da Paraíba. Esse estudo também evidenciou maiores diferenças de percentuais antes e depois

da imputação na variável *Escolaridade*. Portanto, sugeriu-se o uso do *linkage* (quando há bancos de dados disponíveis para realizar interação) e da imputação múltipla na rotina dos serviços de vigilância à saúde, para permitir a recuperação de informações nos sistemas de estatísticas vitais, facilitando estudos de planejamento, monitoramento da prevalência e sua magnitude na população.

#### 4.3.3 Redistribuição de óbitos por causas mal definidas (CMD)

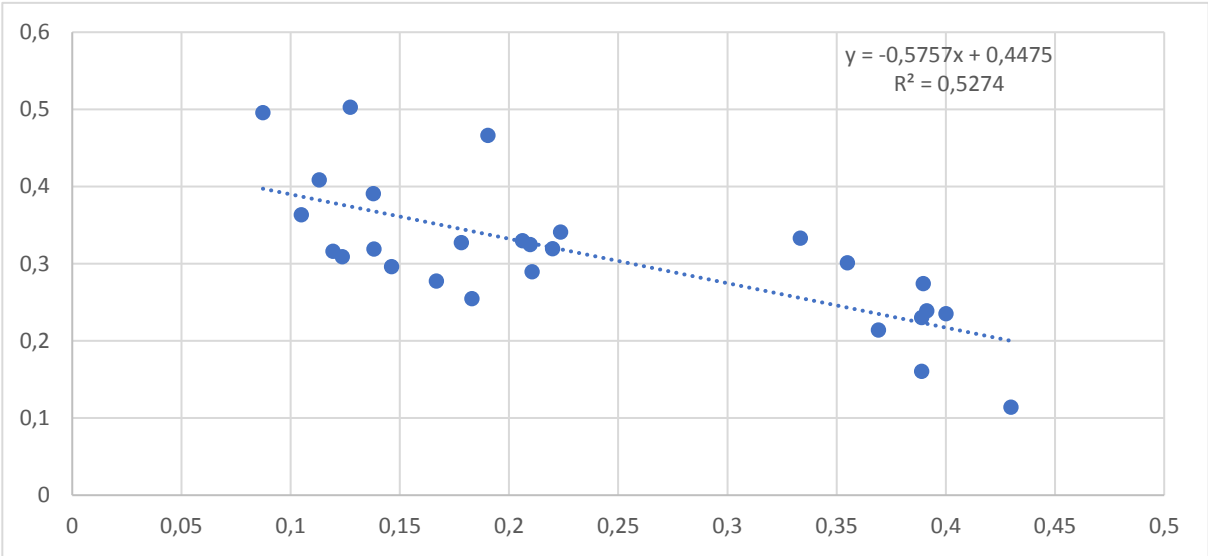
Em sessões anteriores foi analisada a proporção de CMD registradas entre os óbitos de centenários que residiam no Semiárido brasileiro ao longo de 20 anos. Essa análise evidenciou uma proporção elevada de CMD nesses registros. Por conseguinte, surge o seguinte questionamento: Qual seria a solução imediata ou o método mais apropriado para redistribuir as mortes por CMD, em cada um dos capítulos, com causas bem definidas? Distribuir proporcionalmente não parece ser uma solução justa, pois o percentual de CMD entre os centenários revelou-se elevada, para tanto foi escolhido um método mais teórico e inferencial, proposto por Ledermann (1955), que faz uso da regressão linear simples como ferramenta de redistribuição dos óbitos por CMD.

A Tabela 4.3.5 mostra os coeficientes de  $\beta$  usando o método de Ledermann e o número absoluto de óbitos antes e depois da redistribuição das CMD em causas definidas. Para a viabilização da aplicação do método e devido à escassez de óbitos de centenários em algumas regiões específicas do Semiárido brasileiro, foram utilizados os registros dos anos de 2010 a 2020, assim como foram divididos em 9 (nove) estados que compõem o Semiárido os quais foram segregados de acordo com os 3 (três) espaços regionalizados (Urbano, Intermediário e Rural). Deste modo, foram utilizadas 27 unidades de referência para a análise de regressão linear simples. Esse modelo foi aplicado para cada um dos principais grupos de causas específicas da Tabela 4.3.5, sendo estimados, assim, os valores para  $\beta$ . As causas externas não foram incluídas na redistribuição das CMD porque foram consideradas como causas relativamente bem informadas.

Como ilustração do método foi apresentado nos Gráficos 4.3.3 e 4.3.4 a dispersão dos pontos, a linha de tendência e a equação da reta, com a definição de  $\beta$ ,  $\alpha$  e  $R^2$ , segregados por sexo, para o grupo de causa básica do óbito: Doenças do aparelho circulatório. Para o sexo masculino  $\alpha=0,4475$  e  $R^2=0,5274$ , o  $\beta=-0,5757$  utilizado no método de Ledermann como coeficiente de redistribuição. Assim também foi construído para o sexo feminino,  $\alpha=0,4572$  e  $R^2=0,4804$ , o  $\beta=-0,501$ . Após o conhecimento dos valores de  $\beta$ , ou seja, dos coeficientes de

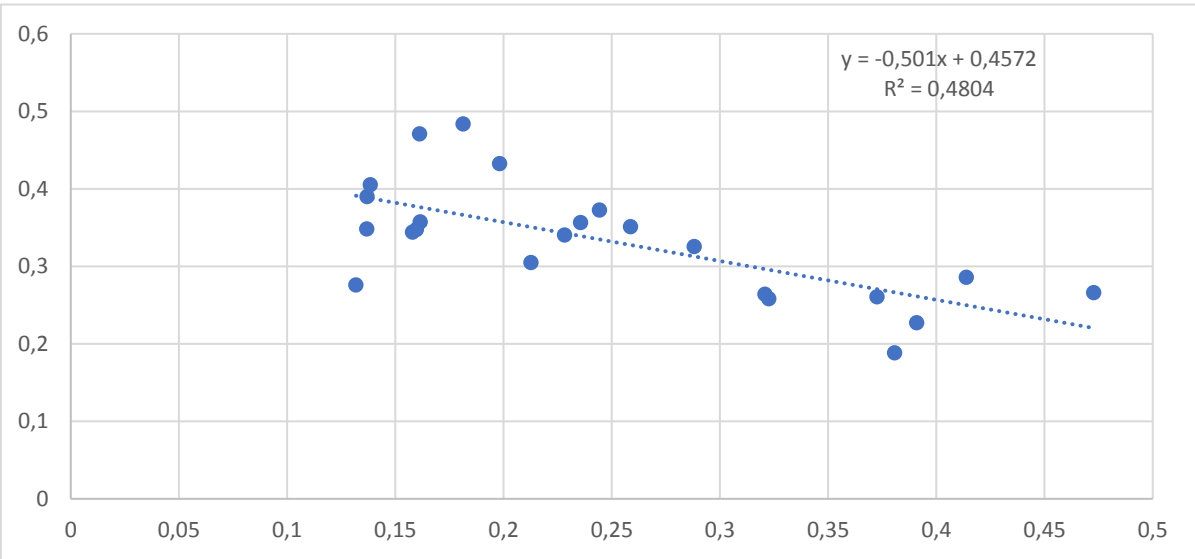
redistribuição de óbitos para os 6 (seis) principais grupos de causas de óbitos por sexo, procedeu-se a redistribuição dos óbitos utilizando a equação de redistribuição proposta no método Ledermann. Os resultados da redistribuição estão apresentados na Tabela 4.3.5.

Gráfico 4.3.3 – Representação da dispersão dos pontos, linha de tendência e a equação da reta, método de Ledermann, para sexo masculino, grupo de causa básica do óbito: Doenças do aparelho circulatório, entre os centenários, Semiárido brasileiro, 2010 – 2020.



Fonte: Elaboração própria

Gráfico 4.3.4 – Representação da dispersão dos pontos, linha de tendência e a equação da reta, método de Lederman, para o sexo feminino, grupo de causa básica do óbito: Doenças do aparelho circulatório, entre os centenários, Semiárido brasileiro, 2010 – 2020.



Fonte: Elaboração própria

Tabela 4.3.5 - Coeficientes de redistribuição e óbitos para os principais grupos de causas de óbitos por sexo, Semiárido brasileiro, 2010-2020.

| Grupo de Causa                    | Coef. de redistribuição<br>( $\beta_i$ ) |         | Óbitos |        |        |        |
|-----------------------------------|------------------------------------------|---------|--------|--------|--------|--------|
|                                   |                                          |         | Antes  |        | Depois |        |
|                                   | M                                        | F       | M      | F      | M      | F      |
| Algumas doenças inf. e parasit.   | -0,0279                                  | -0,0150 | 421    | 772    | 477    | 828    |
| Neoplasias [tumores]              | -0,0920                                  | -0,0410 | 425    | 437    | 610    | 591    |
| Doenças endócrinas nutric. metab. | -0,0809                                  | -0,1026 | 425    | 859    | 588    | 1.242  |
| Doenças do aparelho circulatório  | -0,5757                                  | -0,5010 | 2.570  | 4.706  | 3.726  | 6.578  |
| Doenças do aparelho respiratório  | -0,1729                                  | -0,2486 | 1.631  | 2.445  | 1.979  | 3.374  |
| Outras causas*                    | -0,0506                                  | -0,0918 | 715    | 1.283  | 817    | 1.626  |
| Mal definidas                     | -                                        | -       | 2.010  | 3.737  | -      | -      |
| Total                             | -1,0000                                  | -1,0000 | 8.197  | 14.239 | 8.197  | 14.239 |

Fonte: Elaboração própria.

\*Não inclui as causas externas

O método utilizado na Tabela 4.3.5 supõe que a proporção de uma causa entre todos os óbitos é subestimada, tendo em vista que parte dos óbitos é atribuído às causas de óbitos mal definidos; pois a medida que aumenta a proporção dos óbitos definidos, diminui a dos óbitos mal definidos (PAES, 2018). Desde de 1987, Vallin já utilizava extensivamente o método de Ledermann em diversas regiões da França. Na última década, diversos autores utilizaram o referido método em estudos realizados no Brasil, principalmente em regiões e unidades federativas, bem como o Semiárido brasileiro, os quais apresentaram elevados percentuais de CMD (PAES; GOUVEIA, 2010; KANSO; MONTILA, 2012; PAES; SILVA; BRAGA 2014; CARVALHO; PAES, 2019).

De acordo com a Tabela 4.3.5, após a implementação do método Ledermann, houve incremento no percentual de óbitos, conforme as seguintes causas básicas definidas, para o sexo masculino e feminino, respectivamente: Algumas doenças infecciosas e parasitárias (0,7% e 0,4%); Neoplasias (2,6% e 1,1%); Doenças endócrinas nutricionais e metabólicas (2,0% e 2,7%); Doença do aparelho circulatório (14,1% e 13,1%); Doença do aparelho respiratório (4,2% e 6,5%); Outras causas (1,2% e 2,4%). Portanto, a Doença do aparelho circulatório foi a causa básica que recebeu maior percentual de óbitos na redistribuição, entre os centenários do Semiárido brasileiro, em ambos os sexos, já as Doenças infecciosas e parasitárias foi a causa básica que recebeu menor número de óbitos na redistribuição entre os centenários do Semiárido brasileiro, em ambos os sexos.

Estudo realizado por Paes e Gouveia (2010), no Nordeste brasileiro, com dados de óbitos da população geral, para ambos os sexos, apresentou o perfil e a magnitude dos ganhos em termos percentuais para as principais causas de morte antes e depois da redistribuição das

causas mal definidas, pelo método de Ledermann. De acordo com o referido estudo, os maiores ganhos ocorreram para as doenças do aparelho circulatório e os menores para as doenças infecciosas e parasitárias, resultados que corroboram com os resultados encontrados na Tabela 4.3.5, entre os centenários do Semiárido brasileiro.

Por fim, enfatizando o caráter inovador do presente estudo, em termos de explorar os registros de óbitos dos centenários, é necessário ressaltar que não há registro na literatura acadêmica ou de órgãos oficiais, uma análise sobre a qualidade dos registros de óbitos e tão pouco a redistribuição das CMD entre os centenários, situação esta tão comum entre esses indivíduos que vão a óbito em algumas regiões do Brasil e que não se sabe a causa que os levou a esse desfecho. Certamente que a ausência de estudos dessa natureza pode comprometer seriamente as avaliações e as análises de mortalidade por causas de mortes, dificultando a extração de conclusões confiáveis para esse grupo etário tão chamativo, dentro do contexto atual de envelhecimento da população brasileira e regional, sendo primordial para o estudo da longevidade humana.

#### 4.4 Associação das causas mal definidas com as variáveis da DO

Dando continuidade ao estudo dos óbitos por causas mal definidas e após a evidência de sua elevada proporcionalidade entre os centenários do Semiárido brasileiro, surgiu o interesse de aprofundar acerca dessa problemática com o propósito de conhecer quais fatores estariam associados ao registro de CMD nesse grupo etário específico. Portanto, essa sessão apresenta o resultado do estudo de associação entre as variáveis da DO - *Faixa etária, Tipologia municipal da residência, Raça/cor, Situação conjugal, Escolaridade, Recebeu assistência médica* durante a doença que ocasionou o óbito e *Local de ocorrência do óbito* – e o registro de causas mal definidas (desfecho), no último triênio estudado (2018-2020), utilizando o Modelo de Regressão Logística,

Inicialmente, foram verificadas quais das variáveis possuíam associação significativa ao nível de 20% com o desfecho indicado. Para essa etapa foram utilizadas as variáveis descritas acima. A partir dessa verificação foi construído um modelo de regressão logística considerando as variáveis que se apresentaram significativas para dar início ao ajuste realizado, a partir das combinações possíveis, entre elas, através dos métodos *backward* e *forward* de seleção.

O modelo que melhor se ajustou aos dados da pesquisa apontou as seguintes variáveis independentes estatisticamente significativas para explicar a classificação da ocorrência dos óbitos por causas mal definidas, conforme mostra o Quadro 4.1, por categorização das variáveis: *Tipologia municipal, Situação conjugal, Local da ocorrência do óbito e Assistência médica*.

Quadro 4.4.1 – Variáveis selecionadas para o ajuste do modelo de regressão logística segundo categorias.

| Variáveis           | Categorias                                                                                |
|---------------------|-------------------------------------------------------------------------------------------|
| TIPOLOGIA MUNICIPAL | Urbano<br>Rural<br>Intermediário                                                          |
| SITUAÇÃO CONJUGAL   | Solteiro (a)<br>Casado (a)<br>Viúvo (a)<br>Separado (a) / Divorciado (a)<br>União estável |
| LOCAL DA OCORRÊNCIA | Hospital<br>Outro Estabelecimentos de saúde<br>Domicílio<br>Via pública<br>Outros         |
| ASSISTÊNCIA MÉDICA  | Sim<br>Não                                                                                |

Fonte: Elaboração própria

Na fase de ajuste do modelo as variáveis descritas no Quadro 4.2 foram recategorizadas, com o propósito de melhoria do ajuste final, da seguinte forma:

Quadro 4.4.2 – Recategorização das variáveis inseridas no ajuste do modelo de regressão logística proposto.

| Variáveis           | Categorias                                                                              |
|---------------------|-----------------------------------------------------------------------------------------|
| TIPOLOGIA MUNICIPAL | Urbano<br>Rural/Intermediário                                                           |
| SITUAÇÃO CONJUGAL   | Solteiro (a) / Separado (a) / Divorciado (a)<br>Viúvo (a)<br>Casado (a) / União estável |
| LOCAL DA OCORRÊNCIA | Hospital ou outro estabelecimento de saúde<br>Domicílio/Via pública/Outros              |
| ASSISTÊNCIA MÉDICA  | Sim<br>Não                                                                              |

Fonte: Elaboração própria

Desta forma a equação gerada pelo modelo de regressão logística proposto ficou definida por:

$P(\text{óbito por causas mal definidas})$

$$= \frac{\exp(-5,93 + 1,29(\text{local ocorrência}) + 0,93(\text{assist. méd.}) + 0,19(\text{est. civil}) + 0,24(\text{tipo. muni}))}{1 + \exp(-5,93 + 1,29(\text{local ocorrência}) + 0,93(\text{assist. méd.}) + 0,19(\text{est. civil}) + 0,24(\text{tipo. muni}))}$$

#### 4.4.1 Avaliação do ajuste do modelo

Uma vez encontrado o modelo final foi realizada a avaliação do ajuste deste modelo. Foi aplicado o teste de qualidade de ajuste de Hosmer e Lemeshow, o qual indicou a adequabilidade do modelo,  $[\chi^2 = 5,23; df = 8; p\text{-valor} = 0,7324]$ .

Além do teste de Hosmer e Lemeshow, o modelo respondeu aos demais critérios de ajuste, o qual foram executados na avaliação, tais como: o teste da Razão de Verossimilhança, que avalia se a diferença entre o modelo proposto e o nulo produziu um resultado estatisticamente significativo  $[\chi^2 = 699,1; p\text{-valor} < 0,001]$ , e o  $R^2_{\text{Nagelkerke}} = 0,28$ .

Desta forma, uma vez que o modelo final atendeu aos critérios de adequabilidade, considerando as variáveis *Situação conjugal*, *Tipologia municipal da residência*, *Local de ocorrência do óbito*, *Assistência médica*, pode-se considerar o modelo estatisticamente significativo para explicar os registros de ocorrência de óbitos por causas mal definidas.

#### 4.4.2 Estimação dos parâmetros do modelo

Para esclarecer o efeito das variáveis independentes na variável resposta do modelo, utilizou-se a razão de chances (*Odds Ratio*), que é dada pelo exponencial dos parâmetros estimados pelo modelo e seus respectivos intervalos de confiança (IC<sub>95%</sub>).

A Tabela 4.4.1 detalha o modelo de regressão ajustado com as variáveis independentes, bem como as estimativas dos parâmetros do modelo ( $\beta$ ), Erro padrão, p-valor, *odds ratio* (OR) e os respectivos intervalos de confiança (IC).

Tabela 4.4.1- Modelo ajustado, variáveis sociodemográficas associadas a CMD, Centenários, Semiárido brasileiro, 2018-2020.

| Variáveis independentes            | Parâmetros estimados ( $\beta$ ) | Erro padrão | p-valor        | ODDS  | IC 95% ODDS |       |
|------------------------------------|----------------------------------|-------------|----------------|-------|-------------|-------|
|                                    |                                  |             |                |       | LI          | LS    |
| Intercepto                         | -5,631                           | 0,220       | < <b>0,001</b> | -     | -           | -     |
| <b>Local de ocorrência</b>         |                                  |             |                |       |             |       |
| Hospital/Estabelecimentos de saúde | -                                | -           | -              | 1     | -           | -     |
| Domicílio/Via pública/Outros       | 1,289                            | 0,073       | < <b>0,001</b> | 3,629 | 3,145       | 4,187 |
| <b>Assistência Médica</b>          |                                  |             |                |       |             |       |
| Sim                                | -                                | -           | -              | 1     | -           | -     |
| Não                                | 0,927                            | 0,067       | < <b>0,001</b> | 2,528 | 2,228       | 2,869 |
| <b>Situação conjugal</b>           |                                  |             |                |       |             |       |
| Casado/União estável               | -                                | -           | -              | 1     | -           | -     |
| Viúvo                              | 0,041                            | 0,094       | 0,659          | 1,042 | 0,865       | 1,255 |
| Solteiro/Separado/Divorciado       | 0,305                            | 0,099       | <b>0,002</b>   | 1,357 | 1,117       | 1,648 |
| <b>Tipologia Municipal</b>         |                                  |             |                |       |             |       |
| Urbano                             | -                                | -           | -              | 1     | -           | -     |
| Rural/Intermediário                | 0,243                            | 0,061       | < <b>0,001</b> | 1,275 | 1,131       | 1,437 |

Fonte: Elaboração própria

Legenda: LI: Limite inferior; LS: Limite superior.

Assim, observando os resultados da Tabela 4.4.1, constatou-se que com exceção da categoria *Situação conjugal Viúvo* todas as demais categorias das variáveis *Local da ocorrência do óbito*, *Assistência médica*, *Situação conjugal* e *Tipologia municipal* foram estatisticamente significativas em relação aos registros de ocorrências de óbito por causas mal definidas de idosos centenários do Semiárido brasileiro, incluídos neste estudo.

A partir da interpretação da OR é possível inferir que os óbitos que ocorreram em domicílios, vias públicas ou outros locais que não são estabelecimentos de saúde têm aproximadamente 3,6 vezes mais chances de serem registrados como óbitos por causas mal

definidas do que os óbitos que ocorreram em hospitais e/ou estabelecimentos de saúde; bem como os idosos que não tiveram assistência médica durante a doença que ocasionou o óbito, apresentaram aproximadamente 2,5 vezes mais chances de serem registrados da mesma forma.

De maneira similar, considerando a recategorização das demais variáveis, observou-se também que os registros de óbito dos centenários com estado civil de solteiro (a) ou separado (a) / divorciado (a) têm aproximadamente 35% de chances a mais de ter classificação de óbito por causas mal definidas do que os centenários com registro civil de casado (a) ou união estável; bem como os óbitos que ocorreram em área rural ou intermediária têm também aproximadamente 27% de chances a mais de serem registrados como óbitos por causas mal definidas do que os óbitos que ocorreram em zona urbana.

Na análise da variável *Local de ocorrência do óbito*, a qual apresentou maior associação com o desfecho CMD comparada as outras variáveis, a categoria óbitos que ocorreram em *Domicílio/Via pública/Outros* apresentou valor do OR (3,63). Estudo realizado com idosos que faleceram no Brasil, analisou a variável *Local de ocorrência* entre os óbitos registrados como CMD, dentre esses, destacaram-se a categoria ‘Domicílio’ em proporções superiores a 70%, sendo 78,8% em 1996 e 71,6% em 2005, contudo, as mortes registradas como CMD ocorridas em ‘Hospital’ variaram entre 15 e 23%, ao ser acrescida a categoria ‘Outros estabelecimentos de saúde’ (JORGE et al, 2008). Análise de proporção semelhante ao estudo anterior, com idosos, foi realizada em quatro capitais brasileiras (São Paulo, Belo Horizonte, Rio de Janeiro e Porto Alegre) e observou-se que, no Rio de Janeiro, o percentual de óbitos por esse grupo de CMD ocorrido em domicílio, em 2007, foi de quase 74%, superior as demais capitais. Evidenciou-se também pela análise do OR que a classificação da causa do óbito como CMD mostrou-se negativamente associada à ocorrência em hospital, cujas OR variaram de 0,04, em Porto Alegre, a 0,79, no Rio de Janeiro, no período estudado (1996-2007), não se observando diferencial segundo sexo. (ABREU; SAKURAI; CAMPOS, 2010)

Kanso e colaboradores (2011), calcularam a razão de chance de um idoso ter em seu registro de óbito causa mal definida, no Brasil, em 2007, encontrando que a categoria *Domicílio* da variável *Local de ocorrência* apresentou um OR de 8,42 comparados aos registros de óbitos que ocorreram em estabelecimentos de saúde (incluídos hospitais). Portanto, as referidas análises já evidenciavam, em anos anteriores a 2010, elevadas proporções e força de associação de registros por CMD nos óbitos ocorridos em domicílio, corroborando com a análise dos óbitos dos centenários. Estudos mais recentes realizados na região Norte do Brasil, também evidenciou elevadas chances de registros óbitos com CMD entre os óbitos ocorridos em domicílios ou

outros locais fora de estabelecimentos de saúde, chegando a uma OR de quase 9,0 no estado do Amazonas em 2014 (BALIEIRO, et al, 2020).

O tipo de atenção à morte dos idosos também se mostrou associado ao registro de causa mal definida na DO, situação já esperada diante da compreensão de que a variável *Recebeu assistência médica* já é definida com relação a dificuldade de saber a causa básica do óbito quando não houve assistência médica durante a doença que ocasionou o óbito. Estudo realizado por Kanso e colaboradores (2011), evidenciou, no ano de 2007, que aqueles idosos que não tiveram assistência médica durante a doença que ocasionou o óbito, apresentaram aproximadamente 14 vezes mais de chance de ter causas mal definidas do que aqueles que tiveram assistência, chances bem mais elevadas do que foi encontrado entre os centenários do Semiárido brasileiro entre 2018 e 2020, evidenciando a melhora dos registros ao longo dos anos. Na região Norte, em 2014, o registro de assistência médica reduziu em 84% a chance de causas mal definidas (BALIEIRO, et al, 2020).

Com relação a variável “Assistência médica” é necessário refletir que nem todos os médicos têm o mesmo entendimento do que seja realmente a variável “Assistência médica” na DO. Para muitos deles, trata-se da “assistência médica no momento da morte” e não da menção de uma situação “Sem assistência médica na doença que levou à morte”, conforme preconiza o Ministério da Saúde na própria DO (MENDONÇA; DRUMOND; CARDOSO, 2010). Não haver assistência médica no momento da morte, para esses médicos, poderia equivaler a “Sem assistência médica”, mesmo que tenha havido hospitalização anterior, com prestação de toda assistência possível. Apesar desse erro de interpretação do profissional responsável pelo preenchimento da DO, a proporção de óbitos com a anotação de que não houve assistência médica vem decrescendo entre os idosos no país (JORGE et al, 2008; PRESTES, et al, 2018).

Ser solteiro/separado/divorciado foi outra condição associada ao registro de CMD em centenários. Na região Norte, em 2014, a associação entre a condição de solteiro também associou-se ao risco mais elevado de causas mal definidas (BALIEIRO, et al, 2020), situação similar também foi reportada em estudo de coorte ocupacional realizado nos EUA, atribuída ao isolamento frente à inexistência de laços parentais, e intensificada com a ocorrência do óbito em domicílio, sem a presença de outras pessoas (CRAGLE; FLETCHER, 1992). Padrões comportamentais específicos assim como o impacto potencial das consequências psicológicas relacionadas à condição de solteiro foram aventados como possíveis explicações. A evidência dessa associação entre os centenários pode contribuir para o conhecimento da importância que a rede de apoio familiar e o arranjo que a estrutura de cuidados oferece ao médico atestante

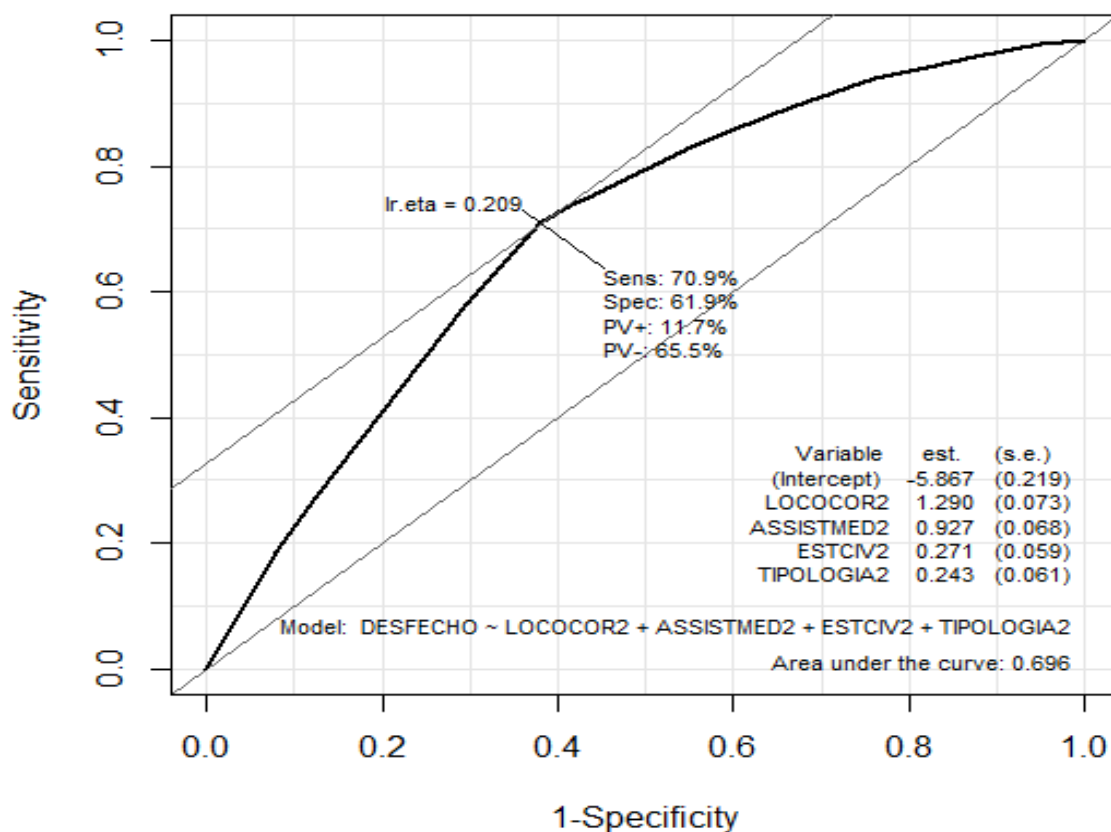
sobre informações de saúde e doença acerca daquele idoso e seu histórico ao longo dos anos. Os informantes mais próximos e a rede de apoio podem facilitar o acesso a informações essenciais sobre as causas de morte, o que estaria deficiente naqueles centenários sem integrantes de núcleos familiares mais próximos (como cônjuges ou filhos), muito frequente na condição de solteiros ou separados. Todavia, a associação entre mortalidade e a situação de solteiro parece não se restringir às causas mal definidas, conforme relatos de outros estudos (BOS et al, 2004; FOUILLET et al, 2006), já em pesquisas conduzidas por Kanso (2010) e Prestes (2018), a variável *Situação conjugal* não se mostrou associada com CMD.

Por fim, a tipologia do município de residência do centenário foi outro fator associado ao registro de CMD, residir em municípios caracterizados como Rural ou Intermediário, ou seja, aqueles denominados como não urbanos, aumentou as chances da ocorrência de registros por CMD. A qualidade da informação da causa de morte apresenta fortes diferenciais geográficos e socioeconômicos, com pior qualidade nas áreas rurais e mais pobres (PAES, 2007; PAES; GOUVEIA, 2010). Estudo realizado em áreas do país consideradas mais desenvolvidas, especialmente as capitais das Regiões Sudeste, Sul e Centro-oeste, apresentaram as menores proporções de causas mal definidas (KANSO, et al, 2010), em consonância com pesquisa voltada para registros de óbitos em menores de um ano (SZWARCOWALD, 2014), o que poderia ser explicado pelo maior acesso aos serviços de saúde em centros urbanos e a disponibilidade de profissionais médicos para determinar a causa do óbito, situação não evidenciada em áreas rurais, distantes de centros urbanos.

#### **4.4.3 Curva Roc**

Para compreender a adequabilidade do modelo proposto neste estudo de associação, o modelo final apresentou área sob a curva igual a 0,696, isto é, o modelo tem capacidade de classificar o indivíduo corretamente cerca de 70% (69,6%) das vezes em que é utilizado, com maior taxa de verdadeiros positivos (sensibilidade de 70,9%) e menor taxa de falsos positivos (Figura 4.4.1).

Figura 4.4.1 - Curva ROC para o modelo final (área sob a curva = 0,696).



Dessa forma, concluiu-se que o modelo final tem uma boa taxa de acerto, tanto para os registros de óbitos por causas mal definidas quanto para os demais registros. Após a construção da curva ROC, foi escolhido o ponto de corte com melhor taxa de verdadeiros positivos e menor taxa de falsos positivos. Este ponto é fundamental para a construção da tabela de contingência (ou matriz de confusão), estabelecendo as taxas de acerto e de erro do modelo final (Tabela 4.4.2). Para a construção da tabela, foi estabelecido o ponto de corte 0,209. Desta forma, o modelo utiliza a seguinte regra de decisão para classificar os registros: registros de óbitos que apresentam probabilidade acima do ponto de corte terão valor predito  $Y=1$  e registros de óbitos com probabilidade abaixo do ponto de corte terão valor predito  $Y=0$ . O modelo final obtido demonstrou a capacidade de prever corretamente os registros de óbitos por causas mal definidas com taxa de acerto de 63,9% (48,3% +15,6%).

Tabela 4.4.2 - Tabela de Contingência para o Modelo de Regressão.

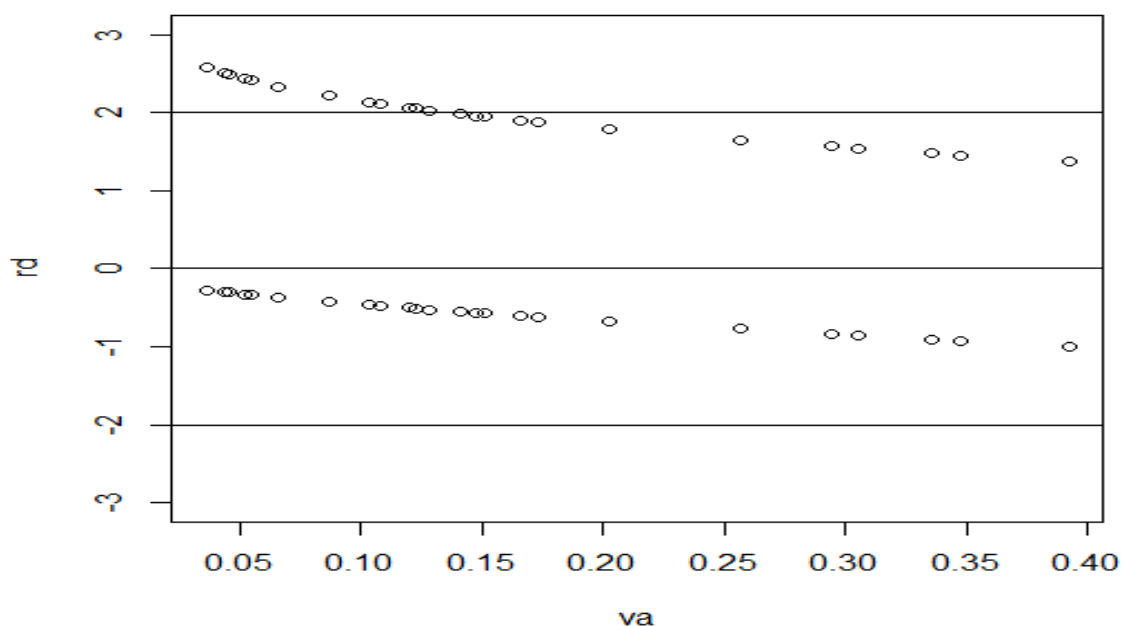
| Valores observados | Valores preditos |                | Total | Acerto | Erro   |
|--------------------|------------------|----------------|-------|--------|--------|
|                    | Y=0              | Y=1            |       |        |        |
| Y=0                | 3.857 (48,25%)   | 2.375 (29,71%) | 6.232 | 61,89% | 38,11% |
| Y=1                | 5.120 (6,40%)    | 1.249 (15,62%) | 1.761 | 70,90% | 29,07% |
|                    | 4.369            | 3.624          | 7.993 |        |        |

Fonte: Elaboração própria

#### 4.4.4 Análise De Resíduos

Os resultados apresentados na Tabela 4.4.1 só serão válidos se as premissas do modelo logístico forem atendidas. Tais premissas se referem às características dos resíduos, como: valor esperado dos resíduos padronizados entre o intervalo de  $[-2;2]$ ; ausência de heterocedasticidade; e ausência de multicolinearidade. A Figura 4.4.2 mostra que os resíduos do modelo logístico proposto sugerem aleatoriedade com padrão de variabilidade constante para os níveis das variáveis explicativas do modelo, porém com alguns pontos aberrantes fora do limite estabelecido, ou seja, os nove primeiros que ficaram acima do limite de tolerância, mas espera-se que não comprometa a viabilidade do modelo utilizado, uma vez que os demais pressupostos foram satisfeitos, e a eliminação desses pontos poderia descompensar, ou não satisfazer os demais pressupostos.

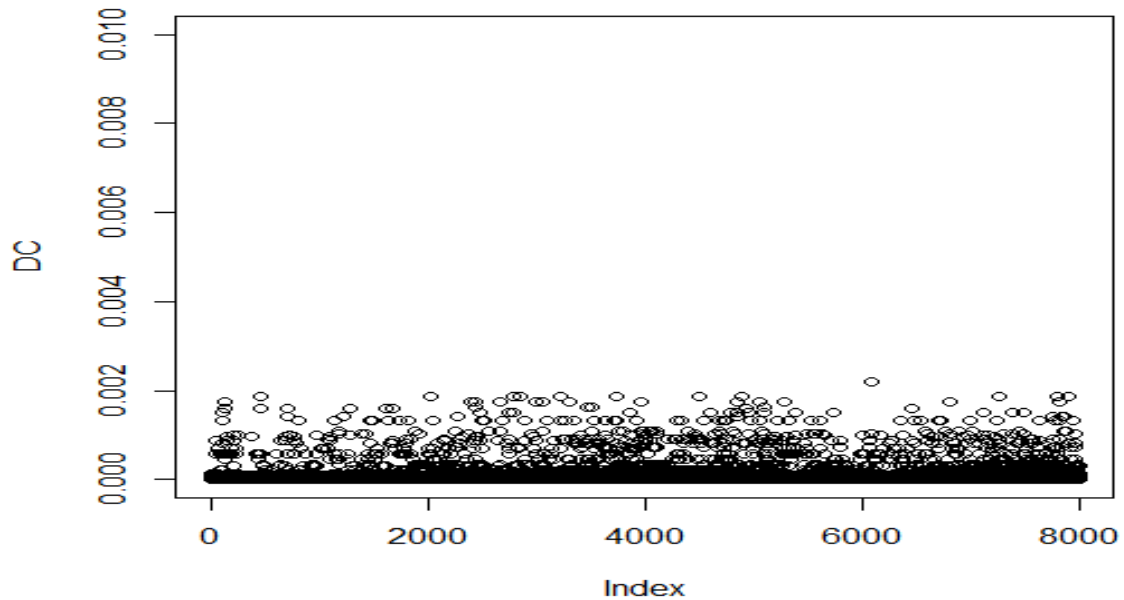
Figura 4.4.2 – Gráfico dos resíduos padronizados do modelo final.



A Figura 4.4.3 mostra a Distância de Cook das observações do modelo. Esta medida mede a diferença entre o modelo de regressão com determinada observação e o modelo sem

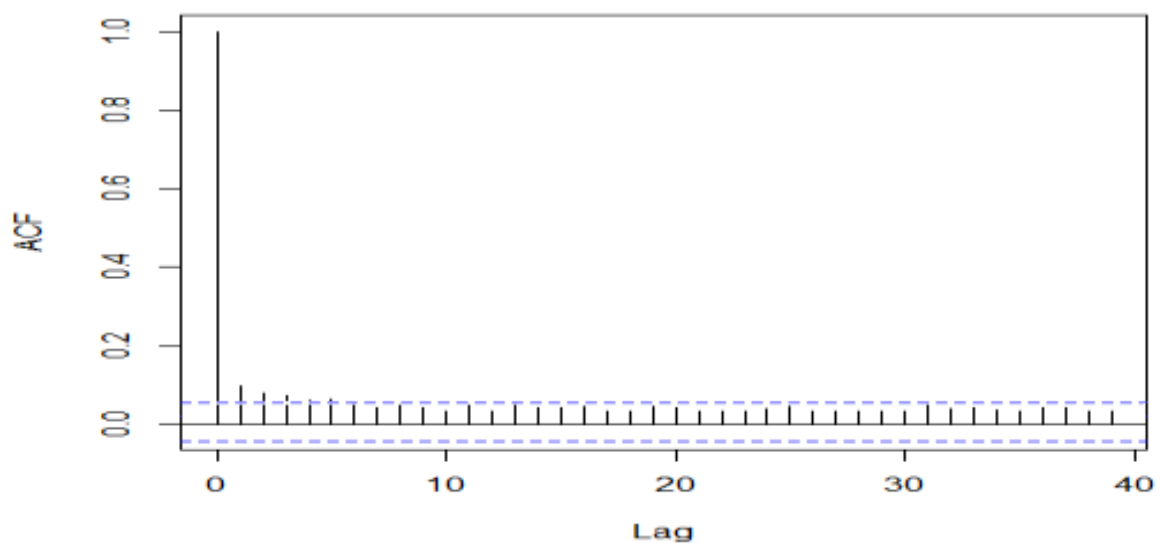
aquela observação, indicando a influência da estimativa daquele ponto do modelo. O ideal é que sejam retiradas as observações com distância de Cook maiores que 1.

Figura 4.4.3 – Gráfico da Distância de Cook do modelo final.



A estatística de Durbin-Watson foi de 1,803 [DW = 1,8027;  $p\text{-valor} < 0,001$ ], bem próximo de 2, ou seja, indicando que os resíduos no modelo são independentes, corroborado pela Figura 4.4.4.

Figura 4.4.4 – Gráfico de Independência dos erros do modelo final.



Assim, de acordo com todas essas demonstrações da análise de resíduos e a taxa de acerto de 63,9% ( $48,3\% + 15,6\%$ ), o modelo logístico apresentado demonstrou uma boa qualidade preditiva.

## 5 CONSIDERAÇÕES FINAIS

Os resultados encontrados nesta Tese fornecem elementos de avaliação sobre qualidade, confiabilidade e fidedignidade dos dados populacionais e dos registros de óbitos dos centenários no contexto da longevidade, entre os anos 2000 a 2020. A longevidade é um fenômeno complexo de ser analisado ou explicado devido a sua heterogeneidade. O Brasil está enfrentando uma rápida transição demográfica e epidemiológica, e o Semiárido, apesar de ser uma das regiões mais iníquas do país, também entrou nesse processo, uma vez que vem apresentando aumento considerável no número de idosos, dentre eles os centenários. Porém, para compreender o real avanço do número de centenários e a essência do fenômeno da longevidade é necessário analisar antes a qualidade dos dados e registros que retratam esse grupo populacional específico, através da análise criteriosa e sistemática dos dados censitários e dos registros de óbito desses indivíduos, ao longo dos anos.

Na análise da contagem censitária dos centenários do Semiárido brasileiro, ao longo de 20 anos, foi observado que o tamanho da população com 100 anos ou mais de idade no Semiárido e em seus espaços regionalizados cresceu no decorrer dos anos, especialmente entre as mulheres, contudo o aumento populacional de centenários, nos Censos demográficos brasileiros, pode estar sendo enviesado pelo registro errôneo do número de pessoas nas referidas idades, seja devido a erros de declaração de idade por parte do respondente – omissão ou declaração errônea – ou por parte das instituições que fornecem os dados – problemas na contagem por omissão ou repetição.

Para realizar tal análise, foi elaborada nesta Tese uma proposta de classificação para avaliar a qualidade dos dados censitários dos idosos centenários por meio de indicadores construídos com dados populacionais, o que refletiu a demanda e a necessidade de manter uma observação sistemática dos dados a fim de assegurar a qualidade do registro censitário. Destaca-se que essa classificação apresentou boa consistência interna, com resultados coerentes com a realidade dos países estudados, satisfatórios níveis de aplicação, de escalabilidade e de reprodutibilidade para diversas localidades. Ademais, a classificação proposta é de fácil compreensão e simplicidade e utiliza poucos indicadores, não demanda programas de computação sofisticados, nem de cálculos elaborados, bastando que se tenha disponíveis os três indicadores básicos sugeridos para aferição da qualidade dos dados censitários dos centenários. Vantagens estas que podem facilmente ser preenchidas por qualquer país ou região do mundo.

De maneira geral e de acordo com a classificação proposta, os dados populacionais referentes aos centenários do Semiárido brasileiro evoluíram com o tempo, passando da classificação *deficiente* para *regular* no decorrer dos censos de 2000 e 2010. Os dados censitários em idades mais avançadas, tendem a melhorar ao longo dos anos, a medida em que as populações de octogenários, nonagenários e centenários se tornam maiores e seus membros passam a ser encontrados com mais frequência nos domicílios brasileiros. Porém, mesmo com o avanço observado, os indicadores do Semiárido ainda são de “baixa” qualidade, principalmente naqueles municípios caracterizados como rurais, sendo ainda distantes dos encontrados nos países mais desenvolvidos, fato esse preocupante, que merece atenção e cuidados na captação desse contingente populacional nos próximos censos demográficos, cujas implicações na contagem errônea trazem implicações que dificultam o planejamento e atenção aos idosos centenários.

Na análise da qualidade dos registros de óbitos, os resultados mostraram que houve um importante avanço no registro das informações na Declaração de Óbitos para o Semiárido brasileiro nos anos estudados para a população dos centenários, fruto dos investimentos dos sistemas de informação de saúde brasileira. No entanto, ainda há muito espaço para melhorias, devido ao percentual elevado de causas mal definidas e desconhecidas, ainda para anos mais recentes.

Especula-se que as coberturas de óbitos de centenários para o Semiárido brasileiro e seus espaços regionalizados não atingiram níveis acima de 90%. Portanto, para atingir o estado de excelência (100%), ainda há um caminho de evolução e melhorias para percorrer e, espera-se, que não seja em um futuro muito distante. O percentual das causas mal definidas, entre os centenários, foi classificado como de qualidade *pouco confiável* principalmente para a maioria dos municípios classificados como rurais ou intermediários, apresentando valores mais elevados entre as mulheres, demandando maiores cuidados nas declarações da causa básica de óbitos. Ademais, a partir do cálculo do percentual das causas desconhecidas, foi possível identificar com maior precisão a preocupante qualidade dos dados de óbitos dos centenários com relação a causa básica, intensificado pelos valores de sub-registros encontrados.

A análise da completude das variáveis contidas na DO permitiu uma percepção mais realista do quadro geral do Semiárido, pois o não preenchimento das variáveis estudadas dificulta ainda mais a compreensão do perfil epidemiológico dos centenários. Dentre as variáveis sociodemográficas, a Escolaridade apresentou maiores percentuais de incompletude, tendo valores mais elevados entre os homens, nos anos iniciais da série história e nos municípios

caracterizados Urbanos. Outra variável ainda mais preocupante, referente especificamente à informação do óbito foi a variável Assistência médica, que apresentou percentuais de *missings* acima de 50% em todas as situações analisadas. O maior percentual encontrado foi entre os homens que residiam em municípios caracterizados como Intermediário (95,1%) no primeiro triênio, para essa variável especificamente, diante de dados ausentes tão elevados, são necessários estudos mais direcionados e de soluções oriundas das certificações médicas e dos sistemas de captação de dados. Para solucionar os problemas de incompletude das variáveis estudadas, foi implementado o método de imputação múltipla para completar os dados das categorias das variáveis sociodemográficas ausentes no banco, já para solucionar o problema das CMD foi implementado o método de redistribuição das causas básicas proposto por Ledermann, ao final o problema da incompletude das variáveis foi solucionado.

Após a redistribuição das CMD, as doenças do aparelho circulatório emergiram como a causa definida mais frequente entre os centenários. Essas somadas as doenças do aparelho respiratório e as doenças endócrinas, nutricionais e metabólicas foram responsáveis por quase 78% dos óbitos dos centenários, em ambos os sexos, nos últimos 10 anos, demonstrando o peso que essas causas têm no perfil de morbimortalidade dos idosos longevos.

Para aprofundar o estudo das CMD e compreender as razões associadas à sua ocorrência, implementou-se a análise de regressão logística a qual evidenciou que os centenários solteiros, separados ou divorciados, que residiam em municípios rurais/intermediários, que faleceram fora de estabelecimentos de saúde e sem assistência médica são categorias associadas ao incremento dos registros por CMD.

É necessário ressaltar que utilizar dados de centenários trata-se de uma tarefa difícil em função do reduzido número de óbitos registrado no SIM em décadas anteriores a 2008, à preocupante qualidade dos dados e a escassez de trabalhos anteriores, principalmente sobre mortalidade, envolvendo esse grupo etário específico. Para tanto, a fonte disponível para trabalhar com a mortalidade é o SIM e se já existem problemas na cobertura dos óbitos e de completude das variáveis em adultos, esses problemas se exacerbam com relação aos centenários. Além dessas limitações, no presente estudo não foi utilizado como variável de mensuração de qualidade o *Garbage Code*, o que fica como sugestão para ser realizado em explorações posteriores, principalmente quando se tratar do estudo de uma causa básica específica. Portanto, sem uma crítica severa aos dados censitários e aos registros de óbitos ficará comprometida a construção de indicadores de mortalidade confiáveis para os centenários.

Ressalta-se a importância da ampliação dos campos de pesquisa no sentido de compreender o comportamento demográfico e epidemiológico dos centenários na região semiárida, contribuindo para o melhor esclarecimento do processo de envelhecimento no Brasil, em especial nas regiões com maior carência de estudos científicos e acadêmicos.

A proposta de classificação para avaliar a contagem dos centenários censitários, os indicadores e os resultados aqui tratados podem colaborar com a compreensão do verdadeiro panorama na distribuição dos idosos longevos e seu perfil de morbimortalidade, indicando áreas prioritárias que necessitam melhorar seus registros e informações, para assim acompanhar mais fidedignamente o processo de envelhecimento da população brasileira e em especial a do Semiárido. As análises realizadas nesse estudo podem servir também de alerta para que sejam estabelecidos mecanismos de captação e controle desses dados de forma a minimizar erros específicos e evitar repetições de erros no próximo Censo Demográfico Brasileiro.

Quanto aos registros de óbitos, cabe aos responsáveis pelos sistemas de informações e órgãos de captação desses dados e informantes médicos mais atenção e compromisso social quanto ao registro e ao preenchimento das Declarações de Óbitos, cujas informações são em vários aspectos comprometidos e inviáveis para fazer, por exemplo, estudos microrregionais para unidades espaciais como a dos municípios, tendo, assim, escasso valor estatístico e epidemiológico, particularmente para uma população de risco como a dos centenários.

Como produtos científicos desta Tese, até a sua defesa, foram publicados os seguintes artigos completos:

Eventos:

- Qualidade da contagem censitária dos centenários no Semiárido brasileiro – Anais do VII Congresso Internacional de Envelhecimento Humano, Campina Grande/PB (ano de publicação 2020).
- Análise das características sociodemográficas e operacionais dos óbitos dos centenários no Semiárido brasileiro: 2000 a 2019 – Anais do VIII Congresso Internacional de Envelhecimento Humano, Campina Grande/PB (ano de publicação 2021).
- Avaliação da qualidade dos registros de óbitos dos centenários dos espaços regionalizados do Semiárido brasileiro – Anais do X Congreso de la Asociación Latinoamericana de Población-ALAP, Valparaíso/Chile (ano de publicação 2022).
- Avaliação da consistência da contagem de centenários recenseados no Semiárido brasileiro – Anais do XXII Congresso da Associação Brasileira de Estudos Populacionais-ABEP, evento online (ano de publicação 2022).

#### Artigos:

- A Qualidade da Contagem Censitária dos Centenários no Semiárido Brasileiro: Uma Análise Comparativa – Revista Desenvolvimento em Questão (ano de publicação 2022).
- Proposta de classificação para avaliar a qualidade dos dados censitários dos centenários – Revista Latinoamericana de Población (ano de publicação 2022).

Diante desse contexto de produção de ativos acadêmicos, limitações inerentes ao trabalho e os resultados obtidos, chama-se a atenção para o caráter contributivo pioneiro desse trabalho e a inexistência de trabalhos acadêmicos e não acadêmicos com essa abordagem no Brasil. Esta Tese proporcionou um procedimento classificatório valioso para avaliar a qualidade censitária da contagem populacional que pode ser utilizado por qualquer região do mundo e em qualquer tempo. O caráter exploratório desse trabalho lança luzes mais assertivas para uma melhor compreensão da situação da contagem dos dados populacionais e dos registros de óbitos, bem como auxiliar as avaliações e tomadas de decisão nas políticas de atenção à saúde, programas e serviços geriátricos e de assistência a essa população tão pouco estudada à despeito do acelerado processo de envelhecimento em que se encontra a população brasileira e em particular a do Semiárido.

## REFERÊNCIAS

AAKER, D. A.; KUMAR, V.; DAY, G. S. **Pesquisa de marketing**, São Paulo: Atlas, 2001. 745p.

ABREU, Daisy Maria Xavier de; SAKURAI, Emília; CAMPOS, Lorenza Nogueira. A evolução da mortalidade por causas mal definidas na população idosa em quatro capitais brasileiras, 1996-2007. **Revista Brasileira de Estudos de População**, v. 27, p. 75-88, 2010.

AFONSO, R. M. et al. Reaching 100 in the countryside: health profile and living circumstances of Portuguese centenarians from the Beira interior region. **Current gerontology and geriatrics research**, v. 2018, 2018.

AGRESTI, A. **An Introduction to Categorical Data Analysis**. 2ª ed. Florida: John Wiley, 2007.

ANDERBERG, M. R. **Cluster analysis for applications**. New York: Acafenic press, 1973, 359p.

ANDRADE, P. G. et al. Qualidade da declaração da idade de idosos no Brasil: uma análise dos censos demográficos das décadas de 1960 a 2010. *In*. Encontro Nacional de Estudos Populacionais. 21., 2018, Poços de Caldas. **Anais**, São Paulo, p. 1-15, 2019.

ARAÚJO. C. A. D. **Ser idoso, sexualidade e cuidados preventivos no atual cenário da longevidade e envelhecimento populacional: estudo de caso no município de Natal/Rio Grande do Norte**. 2018. 185p. Tese (Doutorado em Demografia) - Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais. Belo Horizonte, 2018.

ARAÚJO, J. D. Polarização Epidemiológica no Brasil. **Revista do Sistema Único de Saúde do Brasil**, Brasília, DF, v. 21, n. 4, p. 533-538, 2012.

ARAÚJO, T. B. Economia do semiárido nordestino: a crise como oportunidade. **Revista coletiva**, v. 16, n. 1, 2013.

ASA. ARTICULAÇÃO SEMIÁRIDO BRASILEIRO. **Ações no semiárido**. 2018. Disponível em: < <http://www.asabrasil.org.br/acoes/p1mc>>. Acesso em: 02 abr. 2018.

ASA. ARTICULAÇÃO SEMIÁRIDO BRASILEIRO. **Semiárido: é no Semiárido que a vida pulsa**, 2020. Disponível em: <<https://www.asabrasil.org.br/semiarido#:~:text=No%20Semi%-C3%A1rido%20ocorrem%20dois%20biomas,onde%20>>. Acesso em: 14 mai. de 2020.

BALIEIRO, Patrícia Carvalho da Silva et al. Factors associated with unspecified and ill-defined causes of death in the State of Amazonas, Brazil, from 2006 to 2012. **Ciência & Saúde Coletiva**, v. 25, p. 339-352, 2020.

BAPTISTA, N. Q.; CAMPOS, C. H. Caracterização do semiárido brasileiro. In: CONTI, I. L.; SCHROEDER, E. O. **Convivência com o semiárido brasileiro**: autonomia e protagonismo social. Brasília: IABIS, 2013.

BARBUSCIA, D. M.; RODRIGUES-JÚNIOR, A. L. Completude da informação nas Declarações de Nascido Vivo e nas Declarações de Óbito, neonatal precoce e fetal, da região de Ribeirão Preto, São Paulo, Brasil, 2000-2007. **Cad. Saúde Pública**, v. 27, n. 6, p. 1192-1200, 2011.

BARRETO, A. S. **Modelos de Regressão**: Teorias e Aplicações com o Programa Estatístico R. 1ª Ed. Brasília: Ed. do Autor, 2011.

BATISTA, J. S. **Hábitos Alimentares e Estado Nutricional em Centenários**. 2019. 56p. Dissertação (Mestrado em Nutrição Clínica) - Faculdade de Ciências da Nutrição e Alimentação, Universidade do Porto, Porto, 2019.

BENNETT, N. G.; HORIUCHI, S. Estimating the completeness of death registration in a closed population. **Population index**, p. 207-221, 1981.

BLAND, J. Martin; ALTMAN, Douglas G. The odds ratio. **Bmj**, v. 320, n. 7247, p. 1468, 2000.

BORGES, G. M. Health transition in Brazil: regional variations and divergence/convergence in mortality. **Revista Cadernos de Saúde Pública**, Rio de Janeiro, v. 33, n. 8, p. 1-15, 2017.

BOS, V. et al. Ethnic inequalities in age-and cause-specific mortality in The Netherlands. **International journal of epidemiology**, v. 33, n. 5, p. 1112-1119, 2004.

BOURBEAU, R.; LEBEL, A. Mortality statistics for the oldest-old: An evaluation of Canadian data. **Demographic Research**, v. 2, 2000.

BRASIL. Secretaria de Atenção a Saúde. Departamento de Ações Programáticas e Estratégicas. **Atenção à saúde da pessoa idosa e envelhecimento**. Ministério da Saúde. Secretaria de Atenção a saúde. Departamento de Ações Programáticas e estratégicas. Área Técnica Saúde do Idoso. Brasília: Ministério da Saúde; 2010.

BRASIL. Ministério da Integração Nacional. **Semiárido**. 2015. Disponível em: <<http://www.sudene.gov.br/acesso-a-informa%C3%A7%C3%A3o/institucional/area-de-atuacao-da-sudene/semiario>>. Acesso em: 22 set. 2020.

BRASIL. Ministério da Saúde. **Sistema de Informações sobre Mortalidade – SIM**. Brasília: Ministério da Saúde; 2020

BRASS, W. Methods for estimating fertility and mortality from limited and defective data. **Methods for estimating fertility and mortality from limited and defective data.**, 1975.

BRAVO, J. M. Tábuas de mortalidade contemporâneas e prospectivas: Modelos estocásticos, aplicações actuariais e cobertura do risco de longevidade. **Universidade de Évora, Évora**, 2007.

BUAINAIN, A. M.; GARCIA, J. R. Desenvolvimento rural do semiárido brasileiro: transformações recentes, desafios e perspectivas. **Confins. Revue franco-brésilienne de géographie/Revista franco-brasileira de geografia**, n. 19, 2013.

BUSSAB, W. O.; MIAZAKI, É. S.; ANDRADE, D. F. Introdução à análise de agrupamentos: In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9.,1990, São Paulo. Resumos. São Paulo, 1990.

CALZADILLA, E. C. et al. Caracterización del estado nutricional de la población centenaria de Cuba. **Revista Cubana de Investigaciones Biomédicas**, v. 32, n. 1, 2013.

CALZADILLA, E. C. et al. Evaluación nutricional de los centenarios cubanos y su relación con biomarcadores del envejecimiento. **Revista Cubana de Investigaciones Biomédicas**, v. 32, n. 1, 2013.

CAMARANO, A. A.; KANSO, S. Envelhecimento da População Brasileira. Uma contribuição demográfica. In: FREITAS, E.V.; PY, L. (Org.). **Tratado de Geriatria e Gerontologia**. Rio de Janeiro: Guanabara Koogan, 2016. p. 141-163.

CARONE, E. Coronelismo: definição histórica bibliografia. **Revista de Administração de Empresas**, v. 11, n. 3, p. 85-92, 1971.

CARVALHO, J. B.; PAES, N. A. Taxas de mortalidade por câncer corrigidas para os idosos dos estados do Nordeste brasileiro. **Ciência & Saúde Coletiva**, v. 24, p. 3857-3866, 2019.

CARVALHO, L. H. A concentração fundiária e as políticas agrárias governamentais recentes. **Revista IDeAS**, v. 4, n. 2, p. 395-428, 2010.

CASCÃO, A. M.; COSTA, A. J. L.; KALE, P. L. Qualidade da informação sobre mortalidade numa coorte de diabéticos-Estado do Rio de Janeiro, 2000 a 2003. **Revista Brasileira de Epidemiologia**, v. 15, p. 134-142, 2012.

CASTRO, J. M.; RODRIGUES-JÚNIOR, A. L. A influência da mortalidade por causas externas no desenvolvimento humano na Faixa de Fronteira brasileira. **Cadernos de Saúde Pública**, v. 28, n. 1, p. 195-200, 2012.

CEVENINI, E. et al. Human models of aging and longevity. **Expert opinion on biological therapy**, v. 8, n. 9, p. 1393-1405, 2008.

CHACKIEL, J. Research on causes of death in Latin America. **Notas de poblacion**, v. 15, n. 44, p. 9-30, 1987.

CHAIMOVICZ, F. Epidemiologia do Envelhecimento no Brasil. FREITAS, E.V.; PY, L. (Org.). **Tratado de Geriatria e Gerontologia**. Rio de Janeiro: Guanabara Koogan, 2016. p. 163-195.

CHUANWAN, S. *et al.* Incompleteness of registration data on centenarians in Thailand. **Journal of Population and Social Studies [JPSS]**, v. 20, n. 2, p. 38-54, 2012.

CLOSS, V. E.; SCHWANKE, C. H. A. Indicadores demográficos relacionados ao envelhecimento. In: SCHWANKE, C. H. A. et. al. (Orgs.). **Atualizações em geriatria e gerontologia IV aspectos demográficos, biopsicossociais e clínicos do envelhecimento**. Porto Alegre: EDIPUCRS, 2017, p. 17-37.

COALE, A. J.; CASELLI, G. Estimation of the number of persons at advanced ages from the number of deaths at each age in the given year and adjacent years. **Genus**, p. 1-23, 1990.

CORDEIRO, G. M.; LIMA NETO, E. A. **Modelos paramétricos**. 1ª ed. São Paulo: ABE, 2004. 256p.

CORMACK, R. M. A review of classification. **Journal of the Royal Statistical Society: Series A (General)**, v. 134, n. 3, p. 321-353, 1971.

CORRÊA, É. R. P.; RIBEIRO, A. M. Ganhos em expectativa de vida ao nascer no Brasil nos anos 2000: impacto das variações da mortalidade por idade e causas de morte. **Ciência & Saúde Coletiva**, v. 22, p. 1005-1015, 2017.

CORREIA, L. O. dos S.; PADILHA, B. M.; VASCONCELOS, S. M. L. Métodos para avaliar a completude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. **Ciência & Saúde Coletiva**, v. 19, p. 4467-4478, 2014.

CORTEZ, A. C. L.; Silva, C. R. L.; Silva, R. C. L.; Dantas, E. H. M. Aspectos gerais sobre a transição demográfica e epidemiológica da população brasileira. **Revista Enfermagem Brasil**, São Paulo, v. 18, n. 5, p. 700-709, 2019.

COSTA, V. et al. The determinants of home and nursing home death: a systematic review and meta-analysis. **BMC palliative care**, v. 15, p. 1-15, 2016.

COURBAGE, Y.; FARGUES, P. A method for deriving mortality estimates from incomplete vital statistics. **Population Studies**, v. 33, n. 1, p. 165-180, 1979.

CRAGLE, D. L.; FETCHER, A. Risk factors associated with the classification of unspecified and/or unexplained causes of death in an occupational cohort. **American journal of public health**, v. 82, n. 3, p. 455-457, 1992.

CUNHA, C. C. et al. Avaliação da investigação de óbitos por causas mal definidas no estado da Bahia, Brasil, em 2010. **Ciência & Saúde Coletiva**, v. 24, p. 1831-1844, 2019.

DINIZ, P. C. O.; PIRAUX, M. Das intervenções de combate à seca às ações de convivência com o semiárido: trajetória de 'experimentalismo institucional' no semiárido brasileiro. **Cadernos de estudos sociais**, v. 26, n. 2, 2011.

ENDERS, C. K. Analyzing longitudinal data with missing values. **Rehabilitation psychology**, v. 56, n. 4, p. 267, 2011.

EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster Analysis**. New York: Oxford University Press, 2001, 237 p.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; Chan, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Campus Elsevier, 2009.

FERLA, A. A.; CECCIM, R. B.; ALBA, R. D. Information, education and health care work: Beyond evidence, collective intelligence. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**; v. 6, n. 2 (2012), v. 24, n. 2, 2012.

FERREIRA C. L.; SAMPAIO, T. L. **Tendência temporal e custos das internações por Diabetes Mellitus no Brasil 2011 a 2019**. Orientador: Betine Pinto Moehlecke. 2022. 96 f. Monografia (Graduação) – Curso de Medicina, Universidade Ânima, Curitiba, 2022.

FIELD, A. **Descobrendo a estatística usando o SPSS**. Tradução: Lorí Viali. 2. ed. Porto Alegre: Artmed, 2009. 684 p.

FOUILLET, A. et al. Excess mortality related to the August 2003 heat wave in France. **International archives of occupational and environmental health**, v. 80, p. 16-24, 2006.

FRANCESCHI, C. et al. Centenarians as a 21st century healthy aging model: A legacy of humanity and the need for a world-wide consortium (WWC100+). **Mechanisms of ageing and development**, v. 165, n. Pt B, p. 55-58, 2017.

FREIRE, F. H. M. A.; JUSTINO, J. R. ASSOCIAÇÃO DO MÉTODO DEMOGRÁFICO: EQUAÇÃO GERAL DE BALANCEAMENTO E ESTIMADOR BAYESIANO EMPÍRICO, PARA ESTIMATIVAS DE MORTALIDADE NA REGIÃO NORDESTE DO BRASIL. **Anais**, p. 1-25, 2016.

FREITAS, Leillimar dos Reis *et al.* **Comparação das funções de ligação logit e probit em regressão binária considerando diferentes tamanhos amostrais**. Mestrado em Estatística aplicada e Biometria, Universidade Federal de Viçosa, 2013.

GALIZONI, F. M. Rural e ruralidades: reflexões para o Programa Nacional de Saneamento Rural. **Relatório técnico do projeto Estudos para concepção, formulação e gestão do Programa Nacional de Saneamento Rural**. 2019.

GOMES, A. A. et al. Doenças respiratórias por influenza e causas associadas em idosos de um município do Nordeste brasileiro. **Cadernos de Saúde Pública**, v. 29, p. 117-122, 2013.

GOMES, M. M. F.; TURRA, C. M. The number of centenarians in Brazil: Indirect estimates based on death certificates. **Demographic Research**, v. 20, p. 495-502, 2009.

HAIR, J. F. et al. **Multivariate Data Analysis**. Prentice Hall: New Jersey, 1998.

HAIR, J. F. et al. PLS-SEM or CB-SEM: updated guidelines on which method to use. **International Journal of Multivariate Data Analysis**, v. 1, n. 2, p. 107-123, 2017.

HARRELL JR, F.E. **Regression modeling strategies: with applications to linear models, logistic regression and survival analysis**. New York: Springer-Verlag, 2001.

HILL, K. Estimating census and death registration completeness. In: **Asian and Pacific population forum/East-West Population Institute, East-West Center**. The Asian & Pacific Population Forum, 1987. p. 8-13, 23.

HORTA, C. J. G. Idade declarada nos censos demográficos e a qualidade da informação. In: Encontro Nacional de Estudos Populacionais, 18., 2012, Águas de Lindóia, **Anais**, Águas de Lindóia, 2012. p. 1-19

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. New York. 2ª ed.: John Wiley, 2000.

IBGE - INSTITUTO BRASILEIRO DE DEMOGRAFIA E ESTATÍSTICA. **Censo 2000**. Rio de Janeiro: IBGE. 2000. Disponível em: [www.ibge.gov.br/home/estatistica/populacao/censo2000](http://www.ibge.gov.br/home/estatistica/populacao/censo2000). Acesso em: 07 abr. 2020.

IBGE -. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo 2010**. Disponível em: <https://censo2010.ibge.gov.br>. Acesso em: 06 abr. 2019.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Indicadores Sociais Municipais**. Uma Análise dos Resultados do Universo do Censo Demográfico 2010. Rio de Janeiro: IBGE. 2010. Disponível em: <<http://biblioteca.ibge.gov.br/visualizacao/livros/liv54598.pdf>>. Acesso em: 14 fev. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Projeção da População do Brasil por Sexo e Idade para o Período 2000 a 2060– Revisão 2013**. Rio de Janeiro: IBGE, 2013. Disponível em: [www.ibge.gov.br/home/estatistica/populacao/projecao\\_da\\_populacao/2013/](http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2013/) Acesso em: 03 jan. 2019.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional de Saúde: 2013: acesso e utilização dos serviços de saúde, acidentes e violências**: Brasil, grandes regiões e unidades da federação Rio de Janeiro: IBGE; 2015.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Classificação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação** / IBGE, Coordenação de Geografia. – Rio de Janeiro: IBGE, 2017. 84p.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Banco de Dados Agregados. 2017**. Rio de Janeiro: IBGE, 2017. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/censo-agropecuario/censo-agropecuario-2017>>. Acesso em: 03 jan. 2019.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistemas de estatísticas vitais no Brasil: avanços, perspectivas e desafios** / Antônio Tadeu Ribeiro de Oliveira, organizador - Rio de Janeiro : IBGE, Coordenação de População e Indicadores Sociais, 2018. 70 p.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Estudo complementar à aplicação da técnica de capturar e captura: estimativas desagregadas dos totais de nascidos vivos e óbitos: 2016- 2019 [Internet]. Rio de Janeiro: IBGE; 2022. 36 p.

Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101927.pdf>>. Acesso em: 22 out. 2022.

INSA - INSTITUTO NACIONAL DO SEMIÁRIDO. Sinopse do Censo Demográfico para o Semiárido brasileiro. Campina Grande, PB: INSA, 2012.

INSA - INSTITUTO NACIONAL DO SEMIÁRIDO. O semiárido brasileiro: riqueza, diversidade e saberes. In: **Coleção (Re)conhecendo o semiárido**. 2014. Disponível em: <<https://portal.insa.gov.br/images/acervocartilhas/O%20Semi%C3%A1rido%20br>>. Acesso em: 06 abr. 2018.

INSA - INSTITUTO NACIONAL DO SEMIÁRIDO. **Sinopse do Censo Demográfico para o Semiárido Brasileiro**. 2017. Disponível em: <<https://portal.insa.gov.br/acervolivros/198-sinopse-do-censo-demografico-para-o-semiarido-brasileiro>>. Acesso em: 15 dez. 2019.

JOLLIVET, Marcel. **Vers un rural postindustriel: rural et environnement dans huit pays européens**. Editions L'Harmattan, 1997.

JOPE, D. S. et al. Physical, cognitive, social and mental health in near-centenarians and centenarians living in New York City: findings from the Fordham Centenarian Study. **BMC geriatrics**, v. 16, n. 1, p. 1-10, 2016.

JORGE, M. H. P. de M. et al. A mortalidade de idosos no Brasil: a questão das causas mal definidas. **Epidemiol. Serv. Saúde**, Brasília, v. 17, n. 4, p. 271-281, 2008.

KANSO, S. et al. Diferenciais geográficos, socioeconômicos e demográficos da qualidade da informação da causa básica de morte dos idosos no Brasil. **Cadernos de Saúde Pública**, v. 27, n. 7, p. 1323-1339, 2011.

KANSO, S.; MONTILLA, D. E. R. Revisão bibliográfica-mortalidade dos idosos no Brasil: análise da estrutura e causalidade e qualidade da informação de causa de morte. **Revista Brasileira de Estudos Populacionais**, v. 30, p. 197-211, 2012.

KAYSER, B. **La Renaissance rurale**, Armand Colin: Paris, 1990.

KOŁŁAJTIS-DOŁOWY, A. et al. The nutritional habits among centenarians living in Warsaw. **Rocz Panstw Zakł Hig**, v. 58, p. 279-286, 2007.

KUMON, M. T.; et al. Centenários no mundo: uma visão panorâmica. **Kairós**, v.12, n. 01, p. 213-232, 2009. Disponível em: <https://revistas.pucsp.br/kairos/article/view/2788/1823>. Acesso em 16 de maio 2020.

KWAK, C. S. et al. Anthropometric index, dietary habits and nutrient intake of the oldest-old population aged 95 and over living in Seoul. **Korean Journal of Community Nutrition**, v. 17, n. 5, p. 603-622, 2012.

LEDERMANN, S. La répartition de décès de causa indeterminée, **Revue de l'institut International de Statistique**, vol 23, 1/3, 1955.

LIMA, A. V.; KONRAD, J. A Transição Demográfica no Brasil e o impacto na previdência social. **Boletim Economia Empírica**, Brasília, DF, v. 1, n. 2, p. 3-12, 2020.

LITTLE, R. JA; RUBIN, D. B. **Statistical analysis with missing data**. John Wiley & Sons, 2019.

MADEIRA, I. R. et al. Ponto de corte do índice Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) avaliado pela curva Receiver Operating Characteristic (ROC) na detecção de síndrome metabólica em crianças pré-púberes com excesso de peso. **Arquivos Brasileiros de Endocrinologia & Metabologia**, v. 52, p. 1466-1473, 2008.

MAHAN, L. K.; ESCOTT-STUMP, S.; RAYMOND, J. L. Krause: alimentos, nutrição e dietoterapia. Rio de Janeiro : Elsevier, 2012.

MALHOTRA, N. K. **Pesquisa de Marketing: uma orientação aplicada**. Editora Bookman, Porto Alegre, 2001.

MALTA, Deborah Carvalho et al. Prevalência e fatores associados com hipertensão arterial autorreferida em adultos brasileiros. **Revista de Saúde Pública**, v. 51, 2017.

MANSUR, Antonio de Padua et al. Tendência do risco de morte por doenças circulatórias, cerebrovasculares e isquêmicas do coração em treze estados do Brasil, de 1980 a 1998. **Arquivos Brasileiros de Cardiologia**, v. 87, p. 641-648, 2006.

MARIN, H. F. Sistemas de informação em saúde: considerações gerais. **Journal of Health Informatics**, v. 2, n. 1, 2010.

MARQUES, L. P.; CONFORTIN, S. C. Doenças do aparelho circulatório: principal causa de internações de idosos no Brasil entre 2003 e 2012. **Revista Brasileira de Ciências da Saúde**, v. 19, n. 2, p. 83-90, 2015.

MARTINS, P. C. R.; PONTES, E. R. J. C.; HIGA, L. T. Convergência entre as Taxas de Mortalidade Infantil e os Índices de Desenvolvimento Humano no Brasil no período de 2000 a 2010. **Interações (Campo Grande)**, v. 19, p. 291-303, 2018.

MATHERS, C. D. et al. National burden of disease studies: a practical guide. **Geneva: World Health Organization**, 2001.

MATHIEU, N. "La notion de rural et les rapports ville-campagne en France". **Économie rurale**, 247, pp. 11-20, 1998.

MAZO, G.Z. et al. Estudo com centenários: Atividade física, estilo de vida e longevidade. **Estudos Interdisciplinares sobre o Envelhecimento**, v. 24, 2019.

MEDEIROS, J. B.; PAES, N. A. Proposta de classificação para avaliar a qualidade dos dados censitários dos centenários. **Revista Latinoamericana de Población**, v. 16, p. 15, 2022.

MEDEIROS, W. R. **Mortalidade em idosos longevos e "mais jovens" no Brasil**. 2015. 108f. Tese (Doutorado em Saúde Coletiva) - Centro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, 2015.

MENDES, E. V. Entrevista: A abordagem das condições crônicas pelo Sistema Único de Saúde. **Ciência & Saúde Coletiva**, v. 23, p. 431-436, 2018.

MENDES, T. C. O. **Perfis da mortalidade de idosos no Nordeste: estudo comparativo entre três faixas etárias e seus fatores contextuais relacionados**. 2018. 160f. Tese (Doutorado em Saúde Coletiva) - Centro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, 2018.

MENDONÇA, F. M.; DRUMOND, E.; CARDOSO, A. M. P. Problemas no preenchimento da Declaração de Óbito: estudo exploratório. **Revista Brasileira de Estudos de População**, v. 27, p. 285-295, 2010.

MEYER, J. Centenarians: 2010, 2010 Census Special Reports. **United States Census Bureau** < <http://www.census.gov/prod/cen2010/reports/c2010sr-03.pdf>, 2012.

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MIRANDA, C.; SILVA, H. **Concepções da ruralidade contemporânea: as singularidades brasileiras**. Instituto Interamericano de Cooperação para a Agricultura-IIICA, 2013.

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

MORAIS, R. M.; COSTA, A. L. Uma avaliação do Sistema de Informações sobre Mortalidade. **Saúde em Debate**, v. 41, p. 101-117, 2017.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. Saraiva Educação SA, 2017.

MORMONT, M. “Le rural comme catégorie de lecture du social”. In M. Jollivet, N. Eizner (dirs.), *L’Europe et ses campagnes*, Paris, **Presses de Sciences**, 1996, pp. 161-176.

MURRAY, C. JL; LOPEZ, A. D. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. **The lancet**, v. 349, n. 9063, p. 1436-1442, 1997.

MUZY, J.; CASTANHEIRA, D.; ROMERO, D. Análise da qualidade da informação da mortalidade prematura por doenças crônicas não transmissíveis e sua utilização nos Objetivos de Desenvolvimento Sustentável. **Cadernos Saúde Coletiva**, v. 29, p. 152-164, 2021.

NAGHAVI, M. et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. **The lancet**, v. 390, n. 10100, p. 1151-1210, 2017.

NATIONAL INSTITUTE ON AGING, U.S. Department of Health and Human Services, National Institutes of Health. **Why Population aging matters: a global perspective**. U.S. Department of State; 2008. Disponível em: <https://www.nia.nih.gov/research/publication/why-population-aging-matters-global-perspective>. Acesso em: 10 jan. 2020.

NEUMAN, M. Regional design: recovering a great landscape architecture and urban planning tradition. **Landscape and Urban Planning**, v. 47, n. 3-4, p. 115-128, 2000.

NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. **Revista Brasileira de Epidemiologia**, v. 13, p. 596-606, 2010.

NYATANGA, B. Dying at home: reconciling with patient and family wishes. **British Journal of Community Nursing**, v. 20, n. 8, p. 410-410, 2015.

OLIVEIRA, A. S. Transição demográfica, transição epidemiológica e envelhecimento populacional no Brasil. **Hygeia-Revista Brasileira de Geografia Médica e da Saúde**, v. 15, n. 32, p. 69-79, 2019.

OLIVEIRA, G. M. M. et al. Estatística Cardiovascular–Brasil 2020. **Arquivos brasileiros de Cardiologia**, v. 115, p. 308-439, 2020.

OLIVEIRA, T. C.; MEDEIROS, W. R.; LIMA, K. C. Diferenciais de mortalidade por causas nas faixas etárias limítrofes de idosos. **Revista brasileira de geriatria e gerontologia**, v. 18, p. 85-94, 2015.

OLSEN, J. et al. Global response to non-communicable diseases—the role of epidemiologists. **International journal of epidemiology**, v. 41, n. 5, p. 1219-1220, 2012.

OMRAN, A. R. The epidemiologic transition: a theory of the epidemiology of population change, **Milbank Memorial Fund Quartely**, n. 49, p. 509-538, 1971.

OMRAN, A. R. The Epidemiologic Transition Theory. A Preliminary Update. **Journal of Tropical Pediatrics**, v. 29, p. 305-316, 1983.

OMS – ORGANIZAÇÃO MUNDIAL DE SAÚDE. **Envelhecimento e Saúde**. OMS. 2015. Disponível em: <https://sbgg.org.br/wp-content/uploads/2015/10/OMS-ENVELHECIMENTO-2015- port.pdf> . Acesso em 10 jan. 2020.

PAES, N. A. **Model life representation for Brazilian mortality** [tese de doutorado]. London: London School of Hygiene and Tropical Medicine, University of London; 1993.

PAES, N. A. Qualidade das estatísticas de óbitos por causas desconhecidas dos Estados brasileiros. **Revista de Saúde Pública**, v. 41, p. 436-445, 2007.

PAES, N.A. **Demografia estatística dos eventos vitais: com exemplos baseados na experiência brasileira**. João Pessoa: Editora do CCTA, 2018. 215p.

PAES, N. A. et al. Nascer no Semiárido brasileiro: desigualdades e a qualidade dos dados. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS DA ABEP, 19, 2014, São Pedro. **Anais eletrônicos...** São Pedro: ABEP, 2014.

PAES, N. A.; GOUVEIA, J. F. Recuperação das principais causas de morte do Nordeste do Brasil: impacto na expectativa de vida. **Revista de Saúde Pública**, v. 44, p. 301-309, 2010.

PAES, N. A.; SANTOS, C. S. A.; COUTINHO, T. D. F. Qualidade dos registros de óbitos infantis para espaços regionalizados: um percurso metodológico. **Revista Brasileira de Epidemiologia**, v. 24, 2021.

PAES, N. A.; SILVA, A. S.; BRAGA, J. S. A Qualidade das declarações de óbitos e sua relação com as condições de vida do Semiárido brasileiro. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS DA ABEP, 19., 2014, São Pedro. **Anais eletrônicos...** São Pedro: ABEP, 2014. Disponível em: <[http://www.abep.org.br/~abeporgb/abep.info/files/trabalhos/trabalho\\_completo/TC-5-18-431-261.pdf](http://www.abep.org.br/~abeporgb/abep.info/files/trabalhos/trabalho_completo/TC-5-18-431-261.pdf)>. Acesso em: 12 out. 2020.

PAES, N. A.; SILVA, E. S. A.; MACIEL, K. A. Uma abordagem metodológica para a construção de tábuas de vida para o semiárido brasileiro a partir dos óbitos registrados. **Brazilian Journal of Development**, v. 6, n. 5, p. 31635-31646, 2020.

PATRÍCIO, S. S. **A longevidade excecional: estudo sobre a qualidade de vida das pessoas centenárias da Beira Interior**. 2013. 122p. Dissertação (Mestrado em Gerontologia) - Universidade da Beira Interior, Covilhã, 2013.

PEDRO, R. E. L.; BÓS, Â. J. G. Características epidemiológicas dos centenários. In: SCHWANKE, C. H. A. et. al. (Orgs.). **Atualizações em geriatria e gerontologia IV aspectos demográficos, biopsicossociais e clínicos do envelhecimento**. Porto Alegre: EDIPUCRS, 2017, p. 55-65

PEREIRA, D. M.; CLEMENTE, C. M. S. A utilização do sistema de informação geográfica–SIG na análise da evolução populacional e do índice de desenvolvimento humano–IDH na microrregião de Guanambi (BA)–1991, 2000 E 2010. **Revista Desenvolvimento Social**, v. 13, n. 1, p. 13-22, 2014.

PEREIRA, R. A., ALVES-SOUZA, R. A., VALE, J. S. O processo de transição epidemiológica no Brasil: uma revisão de literatura. **Revista Científica da Faculdade de Educação e Meio Ambiente**, v. 6, n. 1, p. 99-108, 2015.

PERLS, T. et al. What does it take to live to 100?. **Mechanisms of Ageing and Development**, v. 123, n. 2-3, p. 231-242, 2002.

PIETRUSZKA, B. et al. Characteristics of nutrition among centenarians living in Warsaw. **Children**, v. 1, p. 20, 2006.

PINTO, L. F.; FREITAS, M. P. S.; FIGUEIREDO, A. W. S. Sistemas Nacionais de Informação e levantamentos populacionais: algumas contribuições do Ministério da Saúde e do IBGE para a análise das capitais brasileiras nos últimos 30 anos. **Ciência & Saúde Coletiva**, v. 23, p. 1859-1870, 2018.

POLO, T. C. F.; MIOT, H. A. Aplicações da curva ROC em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 19, 2020.

PREGIBON, D. Logistic regression diagnostics. **The annals of statistics**, v. 9, n. 4, p. 705-724, 1981.

- PRESTES, Cristina et al. Tendência da mortalidade por causas mal definidas no estado do Tocantins e na sua capital Palmas, 1998-2014. **Epidemiologia e Serviços de Saúde**, v. 27, p. e2017471, 2018.
- PRESTON, S. H.; ELO, I. T.; PRESTON, SAMUEL H. Effects of age misreporting on mortality estimates at older ages. **Population studies**, v. 53, n. 2, p. 165-177, 1999.
- PRESTON, S. H.; HILL, M. E.; DREVENSTEDT, G. L. Childhood conditions that predict survival to advanced ages among African-Americans. **Social science & medicine**, v. 47, n. 9, p. 1231-1246, 1998.
- PNUD - PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Relatório de Desenvolvimento Humano 2015** – Valores e Desenvolvimento Humano. Brasília, DF: PNUD, 2015. Disponível em <https://esa.un.org/unpd/wpp/Download/Standard/Population>. Acesso em: 29 jun. 2018.
- REIS, C. S.; TURRA, C. M. Distribuição espacial dos centenários no Brasil: uma análise exploratória da qualidade dos dados dos censos de 2000 e 2010. **Revista Espinhaço|UFVJM**, v. 5, n.1, p. 52-61, 2016.
- RIOS, M. A. et al. Completude do sistema de informação sobre mortalidade por suicídio em idosos no estado da Bahia. **Jornal Brasileiro de Psiquiatria**, v. 62, p. 131-138, 2013.
- ROBINE, J. M.; CUBAYNES, S. Demografia mundial de centenários. **Mecanismos de Envelhecimento e Desenvolvimento**, v. 165, p. 59-67, 2017.
- ROCHON, P. A. et al. Demographic characteristics and healthcare use of centenarians: A population-based cohort study. **Journal of the American Geriatrics Society**, v. 62, n. 1, p. 86-93, 2014.
- RODRIGUES, J. F. O rural e o urbano no Brasil: uma proposta de metodologia de classificação dos municípios. **Análise Social**, v. 49, n. 211, 2014.
- ROMERO, D. E.; CUNHA, C. B. Avaliação da qualidade das variáveis sócioeconômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). **Cadernos de Saúde Pública**, v. 22, p. 673-681, 2006.
- ROSENWAIKE, I. On measuring the extreme aged in the population. **Journal of the American Statistical Association**, v. 63, n. 321, mar. 1968.
- ROSENWAIKE, I.; STONE, L. F. Verification of the ages of supercentenarians in the United States: results of a matching study. **Demography**, v. 40, nº. 4, p. 727-739, 2003.
- ROSES, C. F.; LEIS, R. P. Um estudo das condições socioeconômicas de municípios gaúchos através da análise de cluster. FECAP. **Revista Administração On Line**, v. 3, n.3, jul./ago./set. 2002. Disponível em: <[www.fecap.br/adm\\_online/art33/rose1.htm](http://www.fecap.br/adm_online/art33/rose1.htm)>. Acesso em: 29 abr. 2012.
- ROWE, J.; KAHN, R. Human aging: usual and successful. **Science**, v.237, n. 4811, p. 143-9, 1987.

RUBIN, D.B. **Multiple imputation for nonresponse in surveys**. New York: Wiley; 1987.

RUBIN, D. B. Multiple imputation after 18+ years. **Journal of the American statistical Association**, v. 91, n. 434, p. 473-489, 1996.

SCHAFER, J. L. Multiple imputation: a primer. **Statistical methods in medical research**, v. 8, n. 1, p. 3-15, 1999.

SCHRAMM, J. M. A. et al. Transição epidemiológica e o estudo de carga de doença no Brasil. **Ciência & Saúde Coletiva**, v. 9, p. 897-908, 2004.

SOUZA, G. F. et al. Perfil da mortalidade de centenários em Curitiba-Paraná, 2010-2019. **Revista Eletrônica Acervo Saúde**, v. 13, n. 11, p. e9013-e9013, 2021.

SPENCER, G; GOLDSTEIN, A. A.; TAEUBER, C. M. **America's centenarians: Data from the 1980 census**. US Department of Health and Human Services, Public Health Service, National Institutes of Health, 1987.

STATSOFT, INC. **Textbook Estatísticas Eletrônico**. Tulsa, OK: StatSoft, [s.d.]. Disponível em: <<http://www.statsoft.com/textbook/cluster-analysis/>>. Acesso em: 29 abr. 2020.

STUQUE, C. O.; CORDEIRO, J. A.; CURY, P. M. Avaliação dos erros ou falhas de preenchimento dos atestados de óbito feitos pelos clínicos e pelos patologistas. **J. Bras. Patol. Med. Lab.**, v. 39, n. 4, p. 361-364, 2003.

SUDENE - SUPERINTENDÊNCIA DO DESENVOLVIMENTO DO NORDESTE. Nova delimitação do Semiárido. Brasília, 2017. Disponível em: <[http://www.sudene.gov.br/images/arquivos/semiarido/arquivos/Rela%C3%A7%C3%A3o\\_de\\_Munic%C3%A9pios\\_Semi%C3%A1rido.pdf](http://www.sudene.gov.br/images/arquivos/semiarido/arquivos/Rela%C3%A7%C3%A3o_de_Munic%C3%A9pios_Semi%C3%A1rido.pdf)>. Acesso em: 13 set. 2019.

SZWARCWALD, C. L. et al. Correction of vital statistics based on a proactive search of deaths and live births: evidence from a study of the North and Northeast regions of Brazil. **Population health metrics**, v. 12, n. 1, p. 1-10, 2014

TAMAKI, E.M. et al. Metodologia de construção de um painel de indicadores para o monitoramento e a avaliação da gestão do SUS. **Ciência & Saúde Coletiva**, 2012.

TEIXEIRA, L. et al. Centenarians in Europe. **Maturitas**, v. 104, p. 90-95, 2017.

TERBLANCHE, W.; WILSON, T. An evaluation of nearly-extinct cohort methods for estimating the very elderly populations of Australia and New Zealand. **PloS one**, v. 10, n. 4, p. e0123692, 2015.

THUM, M. A.; BALDISSEROTTO, J.; CELESTE, R. K. Utilização do e-SUS AB e fatores associados ao registro de procedimentos e consultas da atenção básica nos municípios brasileiros. **Cadernos de Saúde Pública**, v. 35, p. e00029418, 2019.

TURRA, C. M. Os limites do corpo. **Revista da Universidade Federal de Minas Gerais**, v. 19, n. 1 e 2, p. 156-181, 2012.

UNICEF. FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA. **Avanços e desafios – A Redução da Mortalidade Infantil em Alagoas**. Coordenação-geral de Cristina Albuquerque e Jane Santos. Brasília, DF: UNICEF, 2012.

UN - UNITED NATIONS. *World Population Prospects. The 2004 Revision*. New York: United Nations, 2005. Disponível em: [www.un.org/esa/population/publications/WPP2004/wpp2004.htm](http://www.un.org/esa/population/publications/WPP2004/wpp2004.htm). Acesso em: 08 abr. 2020.

UN - UNITED NATIONS. *World Population Prospects: 2015 Revision*. New York: United Nations, 2015. Disponível em: [www.un.org/en/development/desa/population/.../pdf/ageing/WPA2015\\_Report.pdf](http://www.un.org/en/development/desa/population/.../pdf/ageing/WPA2015_Report.pdf). Acesso em: 08 abr. 2020.

UNSD - United Nations Statistics Division. UNdata: A world of information. United Nations, 2021. Disponível em: <https://data.un.org/>. Acesso em: 08 abr. 2020.

UNPD - United Nations Development. *Programme Human development indices and indicators: 2018 statistical update*. United Nations, 2018. Disponível em: [http://hdr.undp.org/sites/default/files/2018\\_human\\_development\\_statistical\\_update.pdf](http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf). 08 abr. 2020.

VACANTE, M. et al. Centenarians and supercentenarians: a black swan. Emerging social, medical and surgical problems. **BMC surgery**, v. 12, n. 1, p. 1-8, 2012.

VASCONCELOS, A. M. N.; GOMES, M. M. F. Transição demográfica: a experiência brasileira. **Revista do Sistema Único de Saúde do Brasil**, Brasília, DF, v. 21, n. 4, p. 539-548, 2012.

VIDOR, A.C.; FISHER, P.D.; BORDIN, R. Utilização dos sistemas de informação em saúde em municípios gaúchos de pequeno porte. **Revista de Saúde Pública** [Internet]. 2011; 45(1):24-30.

WARD, J. H. Hierarchical grouping to optimize an objective function. **Journal of. American Statistical Association**, v. 58, p. 236-244, 1963.

WILLCOX, D. C. et al. The Okinawan diet: health implications of a low-calorie, nutrient-dense, antioxidant-rich dietary pattern low in glycemic load. **Journal of the American College of Nutrition**, v. 28, n. sup4, p. 500S-516S, 2009.

WILMOTH, J. In search limits. In: WATCHER, K. W.; FINCH, C.E. (Eds.). **Between Zeus and the Salmon**. Washington: National Academy Press, 1997, p. 38-64.

XU, J. Mortality among Centenarians in the United States, 2000–2014. **NCHS Data Brief**, v. 233., p 1-8, 2016.

YU, R.; TAM, W; WOO, J. Trend of centenarian deaths in Hong Kong between 2001 and 2010. **Geriatrics & gerontology international**, v. 17, n. 6, p. 931-936, 2017.

# ANEXO 1 – MODELO DA DECLARAÇÃO DE ÓBITO

| República Federativa do Brasil<br>Ministério da Saúde<br>1ª VIA - SECRETARIA DE SAÚDE |                                                                                                    | Declaração de Óbito |                                                                   |
|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|---------------------|-------------------------------------------------------------------|
| I<br>Identificação                                                                    | 1 Tipo de óbito<br><input type="checkbox"/> Fetal<br><input type="checkbox"/> Não Fetal            | 2 Data do óbito     | 3 Cerião SUS                                                      |
|                                                                                       | 4 Nome da Falecida                                                                                 | 5 Naturalidade      |                                                                   |
|                                                                                       | 6 Nome do Pai                                                                                      | 7 Nome da Mãe       |                                                                   |
|                                                                                       | 8 Data de nascimento                                                                               | 9 Idade             | 10 Sexo                                                           |
| II<br>Residência                                                                      | 11 Raça/Cor                                                                                        |                     | 12 Situação conjugal                                              |
|                                                                                       | 13 Escolaridade (última série concluída)                                                           |                     | 14 Ocupação habitual                                              |
|                                                                                       | 15 Logradouro (rua, praça, avenida, etc.)                                                          |                     | 16 CEP                                                            |
|                                                                                       | 17 Bairro/Distrito                                                                                 |                     | 18 Município de residência                                        |
| III<br>Ocorrência                                                                     | 19 Local de ocorrência do óbito                                                                    |                     | 20 Estabelecimento                                                |
|                                                                                       | 21 Endereço da ocorrência, se fora do estabelecimento ou da residência (rua, praça, avenida, etc.) |                     | 22 CEP                                                            |
|                                                                                       | 23 Bairro/Distrito                                                                                 |                     | 24 Município de ocorrência                                        |
|                                                                                       | 25 Local de ocorrência do óbito                                                                    |                     | 26 Estabelecimento                                                |
| IV<br>Fetal ou menor que 1 ano                                                        | 27 Idade (anos)                                                                                    |                     | 28 Escolaridade (última série concluída)                          |
|                                                                                       | 29 Número de filhos vivos                                                                          |                     | 30 Nº de semanas de gestação                                      |
|                                                                                       | 31 Tipo de parto                                                                                   |                     | 32 Tipo de gravidez                                               |
|                                                                                       | 33 Morte em relação ao parto                                                                       |                     | 34 Peso ao nascer                                                 |
| V<br>Condições e causas do óbito                                                      | 35 Óbito de mulher em idade fértil                                                                 |                     | 36 Assistência médica                                             |
|                                                                                       | 37 Diagnóstico confirmado por:                                                                     |                     | 38 Nasrúptia?                                                     |
|                                                                                       | 39 Causas da morte                                                                                 |                     | 40 Recebeu assist. médica durante a doença que ocasionou a morte? |
|                                                                                       | 41 Causas da morte                                                                                 |                     | 42 Nasrúptia?                                                     |
| VI<br>Médico                                                                          | 43 Nome do Médico                                                                                  |                     | 44 CRM                                                            |
|                                                                                       | 45 Meio de contato (telefone, fax, e-mail, etc.)                                                   |                     | 46 Data do atestado                                               |
|                                                                                       | 47 Assinatura                                                                                      |                     | 48 Óbito atestado por Médico                                      |
|                                                                                       | 49 Município e UF do SVO ou IML                                                                    |                     | 50 Assinatura                                                     |
| VII<br>Causas externas                                                                | 51 Prováveis circunstâncias de morte não natural                                                   |                     | 52 Fonte de informação                                            |
|                                                                                       | 53 Descrição sumária do evento, incluindo o tipo de local de ocorrência                            |                     | 54 Local de ocorrência                                            |
|                                                                                       | 55 SE A OCORRÊNCIA FOR EM VIA PÚBLICA, ANOTAR O ENDEREÇO                                           |                     | 56 Logradouro (rua, praça, avenida, etc.)                         |
|                                                                                       | 57 Cartório                                                                                        |                     | 58 Registro                                                       |
| VIII<br>Cartório                                                                      | 59 Município                                                                                       |                     | 60 UF                                                             |
|                                                                                       | 61 Declarante                                                                                      |                     | 62 Testemunhas                                                    |
|                                                                                       | 63 Assinatura                                                                                      |                     | 64 Assinatura                                                     |
|                                                                                       | 65 Localidade, SI Médico                                                                           |                     | 66 Assinatura                                                     |