UNIVERSIDADE FEDERAL DO AMAZONAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA-PPGM
MESTRADO EM MATEMÁTICA NA ÁREA DE CONCENTRAÇÃO EM ESTATÍSTICA
INSTITUTO DE CIÊNCIAS EXATAS

HEVELEN FRANCISCA SILVA E SILVA

MODELAGEM DE RISCO E PADRÕES ESPACIAIS DA MORTALIDADE POR CÂNCER DE MAMA E DO COLO DO ÚTERO NO AMAZONAS

Manaus-AM
2026

HEVELEN FRANCISCA SILVA E SILVA

MODELAGEM DE RISCO E PADRÕES ESPACIAIS DA MORTALIDADE POR CÂNCER DE MAMA E DO COLO DO ÚTERO NO AMAZONAS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Jeremias da Silva Leão

Manaus-AM

2026

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S586m Silva, Hevelen Francisca Silva e
Modelagem de risco e padrões espaciais da mortalidade por câncer de mama e do colo do útero no Amazonas / Hevelen Francisca Silva e Silva. - 2026.
109 f. : il., p&b. ; 31 cm.

Orientador(a): Jeremias da Silva Leão.
Dissertação (mestrado) - Universidade Federal do Amazonas, Programa de Pós-Graduação em Matemática, Manaus, 2026.

1. Neoplasias. 2. Mortalidade Feminina. 3. Modelos Lineares Generaliza- dos. 4. Análise Espacial. 5. Amazonas. I. Leão, Jeremias da Silva. II. Universidade Federal do Amazonas. Programa de Pós-Graduação em Matemática. III. Título

---

MODELAGEM DE RISCO E PADRÕES ESPACIAIS DA MORTALIDADE POR CÂNCER DE MAMA E DO COLO DO ÚTERO NO AMAZONAS

HEVELEN FRANCISCA SILVA E SILVA

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas como requisito parcial para obtenção do título de Mestre em Matemática, com área de concentração em Estatística.

Aprovada por:

Prof. Dr. Jeremias da Silva Leão
Universidade Federal do Amazonas - UFAM
(Orientador)

Prof. Dr. Antônio Alcirley da Silva Balieiro
Fundação Oswaldo Cruz - Fiocruz - Manaus
(Membro)

Prof. Dr. Danilo Gilberto de Oliveira Valadares
Universidade Federal do Amazonas - UFAM
(Membro)

Manaus-AM, 23 de Março de 2026.

À minha mãe, que me ensinou, antes de qualquer fórmula, o verdadeiro significado de resiliência.

Ao meu bisavô e à minha tia, que perderam a luta contra o câncer e me ensinaram que, por trás de cada estatística, existem histórias, afetos e vidas que jamais serão apenas números.

AGRADECIMENTOS

A realização deste trabalho não teria sido possível sem o apoio e a presença de muitas pessoas ao longo desta jornada. Minha gratidão a todos que, de alguma forma, tornaram este trabalho possível.

À minha mãe, Milena Rocha, por ter sido mãe e pai, por me incentivar em todos os momentos da minha vida e nunca deixar que eu desistisse. À minha família, pelo amor, pelo apoio incondicional e por compreenderem minhas ausências ao longo dessa jornada.

À minha melhor amiga, Glenda, família que a vida me permitiu escolher, por me fazer rir nos momentos difíceis e por também ter sido revisora e confidente. Ao meu amigo e primo, Breno, por tudo. Aos meus amigos do mestrado, pelas atividades, pelo conhecimento partilhado e pelos nossos lanches — em especial a Lucas, Carlos e Larissa.

À minha namorada, Taciane de Souza, que conheci ao longo desta jornada acadêmica, por me mostrar que este caminho pode trazer muito mais do que realizações profissionais, tornando minha caminhada mais leve e feliz e contribuindo com reflexões e apoio ao longo da construção desta dissertação.

À Samantha Martins, pelo incentivo constante aos estudos e por ter me motivado a me inscrever neste programa.

Ao meu orientador, professor Jeremias Leão, pela confiança, orientação e entusiasmo durante toda esta jornada e por ter acreditado na relevância deste estudo desde o início, incentivando, com abertura e rigor científico, a investigação de um tema desafiador. À professora Themis da Costa Abensur Leão, por todo o conhecimento partilhado durante meu estágio em docência.

À Universidade Federal do Amazonas, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro, por possibilitarem a realização deste curso de mestrado.

"Os números têm uma história importante para contar. Eles dependem de você para lhes dar uma voz clara e convincente."

(Stephen Few).

RESUMO

A presente dissertação analisa as taxas brutas e padronizadas de mortalidade por câncer de mama e câncer do colo do útero no estado do Amazonas, no período de 2010 a 2023, considerando a relevância dessas neoplasias no contexto da mortalidade feminina e da heterogeneidade sociodemográfica do território. Trata-se de um estudo ecológico, de abordagem quantitativa, no qual as taxas padronizadas por idade foram modeladas por meio de Modelos Lineares Generalizados com distribuição Gamma e função de ligação logarítmica, adequados à natureza contínua e assimétrica do desfecho. A identificação de conglomerados espaciais foi realizada por meio da estatística de varredura espacial de Kulldorff, implementada no software SaTScan. Os resultados evidenciaram aumento progressivo da mortalidade com o avanço etário, bem como associação positiva e estatisticamente significativa entre extrema pobreza e maiores taxas padronizadas por idade, reforçando a influência dos determinantes sociais. Observou-se ainda associação com a variável ausência de filiação religiosa, interpretada como possível marcador indireto de suporte social. A análise espacial identificou predominantemente conglomerados de baixo risco, sugerindo que a mortalidade elevada apresenta distribuição ampla no território estadual. Conclui-se que fatores etários e socioeconômicos exercem influência significativa na dinâmica da mortalidade por essas neoplasias no Amazonas, indicando a necessidade de estratégias voltadas à equidade no acesso ao rastreamento e à atenção oncológica.

Palavras-chave: Neoplasias; Mortalidade Feminina; Modelos Lineares Generalizados; Análise Espacial; Amazonas.

ABSTRACT

This thesis analyzes the crude and age-standardized mortality rates from breast and cervical cancer in the state of Amazonas, Brazil, from 2010 to 2023, considering the relevance of these neoplasms within the context of female mortality and the socio-demographic heterogeneity throughout the state. This is an ecological study with a quantitative approach, in which age-standardized rates were modeled using Generalized Linear Models with a Gamma distribution and a logarithmic link function, appropriate for the continuous and skewed distribution of the outcome. Spatial clusters were identified using Kulldorff's spatial scan statistic, implemented in SaTScan. The results showed a progressive increase in mortality with advancing age, as well as a positive and statistically significant association between extreme poverty and higher age-standardized rates, highlighting the influence of social determinants. An association was also observed with the absence of religious affiliation, interpreted as a possible indirect marker of social support. The spatial analysis predominantly identified low-risk clusters, suggesting that elevated mortality is broadly distributed throughout the state. These results indicate that age-related and socioeconomic factors significantly influence the dynamics of mortality from these neoplasms in Amazonas, thereby reinforcing the need for strategies aimed at equity in access to screening and oncological care.

Keywords: Neoplasms; Female Mortality; Generalized Linear Models; Spatial Analysis; Amazonas.

LISTA DE ILUSTRAÇÕES

1 Regiões Geográficas do Amazonas segundo o IBGE (2017). 24

2 Superestimação e Subestimação do conglomerado. 45

3 Taxa bruta de mortalidade por capítulos CID-10 – Brasil, 2010 a 2023. . . . 51

4 Taxa padronizada de mortalidade por câncer de mama – Brasil *versus* Amazonas, 2010 a 2023. 55

5 Taxa padronizada de mortalidade por câncer de colo de útero – Brasil *versus* Amazonas, 2010 a 2023. 55

6 Taxa padronizada de mortalidade por câncer de mama e colo de útero – Amazonas, 2010 a 2023. 56

7 Taxa padronizada de mortalidade câncer de mama por faixa etária – Amazonas, 2010 a 2023. 58

8 Taxa de mortalidade câncer padronizada por faixa etária de colo de útero por faixa etária – Amazonas, 2010 a 2023. 58

9 Matriz de correlação entre variáveis sociodemográficas associadas à mortalidade por neoplasias – Amazonas, 2010 a 2023. 63

10 Distribuição da taxa padronizada de mortalidade por neoplasias – Amazonas, 2010 a 2023. 65

11 Renda média municipal – Amazonas, 2022. 72

12 Taxas média bruta e padronizada de mortalidade por câncer de mama – Amazonas, 2010 a 2023. 74

13 Taxas média bruta e padronizada de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023. 75

14 Risco relativo (RR) bruto e padronizado de mortalidade por câncer de mama – Amazonas, 2010 a 2023. 77

15 Risco relativo (RR) bruto e padronizado de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023. 78

16 Conglomerados espaciais de alto e baixo risco da mortalidade por câncer de mama, considerando 5% a 15% da população sob risco – Amazonas, 2010–2016. 83

17 Conglomerados espaciais de alto e baixo risco da mortalidade por câncer de mama, considerando 5% a 15% da população sob risco – Amazonas, 2017–2023. 84

18 Conglomerados espaciais de alto e baixo risco da mortalidade por câncer do colo do útero, considerando 5% a 15% da população sob risco – Amazonas, 2010–2016. 87

19 Conglomerados espaciais de alto e baixo risco da mortalidade por câncer do colo do útero, considerando 5% a 15% da população sob risco – Amazonas, 2017–2023. 88

20 Taxa de incidência e mortalidade padronizada por idade (mundial) por 100 mil/mulheres – Mundo, 2022. 99

21 Diagnóstico do ajuste do modelo Gamma com ligação log para mortalidade por câncer de mama – Amazonas, 2010 a 2023. 100

22 Diagnóstico do ajuste do modelo Gamma com ligação log para mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023. 101

23 Taxa média padronizada anual de mortalidade por câncer de mama – Amazonas, 2010 a 2023. 102

24 Taxa média padronizada anual de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023. 103

25 Risco relativo padronizado anual de mortalidade por câncer de mama – Amazonas, 2010 a 2023. 104

26 Risco relativo padronizado anual de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023. 105

LISTA DE TABELAS

1 Macrorregiões, Regiões de Saúde e Indicadores Demográficos do Amazonas 25

2 Taxa bruta de mortalidade por regiões e capítulos CID-10 – Brasil, 2010 a 2023. 50

3 Métricas das taxas brutas de mortalidade por capítulo da CID-10 – Brasil, 2010-2023. 51

4 Taxa bruta de mortalidade por capítulos CID-10 – Região Norte, 2010 a 2023. 52

5 Taxa bruta de mortalidade por capítulos CID-10 – Amazonas, 2010 a 2023. 52

6 Métricas das taxas brutas de mortalidade por capítulo da CID-10 – Amazonas, 2010-2023. 53

7 Número de óbitos por sexo – Amazonas, 2010 a 2023. 53

8 Taxas brutas de mortalidade por câncer de mama e do colo do útero – Amazonas, 2010 a 2023. 53

9 Taxas de mortalidade padronizada por 100 mil mulheres-ano por CID – Brasil e Amazonas, 2010 a 2023. 54

10 Taxas padronizadas de mortalidade por faixa etária e CID-10 – Amazonas, 2010 a 2023. 57

11 Métricas das taxas padronizadas de mortalidade por faixa etária – Amazonas, 2010-2023. 60

12 Resumo estatístico das variáveis independentes incluídas nos modelos de mortalidade por câncer de mama e colo do útero – Amazonas, 2010-2023. 62

13 Comparação entre erros-padrão convencionais e clusterizados 67

14 Estimativas do modelo de regressão Gamma com erros robustos para a taxa padronizada de mortalidade. 67

15 Interpretação dos coeficientes do modelo Gamma com ligação logarítmica para mortalidade por neoplasias. 68

16 Estimativas do modelo Gamma com ligação log com erros-padrão robustos clusterizados para mortalidade por câncer de mama (C50) com razão de taxas e interpretação percentual. 69

17 Estimativas do modelo Gamma com ligação log com erros-padrão robustos clusterizados para mortalidade por câncer do colo do útero (C53) com razão de taxas e interpretação percentual. 70

18 Conglomerados espaciais de baixo risco detectados segundo ano, tamanho da janela e nível para o câncer de mama 80

19 Conglomerados espaciais de alto risco detectados segundo ano, tamanho da janela e nível para o câncer do colo do útero 85

A1 Classificação dos capítulos da CID-10 93

A2 População padrão mundial da OMS (2000–2025) repadronizada para faixas etárias acima de 20 anos. 94

A3 Regionalização do Amazonas: Macrorregiões, Regiões de Saúde e Municípios 95

A4 Municípios do Amazonas segundo macrorregião, código IBGE e identificador 97

A5 Diagnóstico de multicolinearidade (VIF) para o modelo ótimo do Modelo Geral. 98

A6 Diagnóstico de multicolinearidade (VIF) para o om erros-padrão robustos clusterizados para mortalidade por câncer de mama (C50). . . . 98

A7 Diagnóstico de multicolinearidade (VIF) para o om erros-padrão robustos clusterizados para mortalidade por câncer do colo do útero (C53). 98

LISTA DE ABREVIATURAS E SIGLAS

UFAM : Universidade Federal do Amazonas

PPGM : Programa de Pós-Graduação em Matemática

DE : Departamento de Estatística

ICE : Instituto de Ciências Exatas

DATASUS O Departamento de Informação e Informática do Sistema Único de Saúde

IBGE : Instituto Brasileiro de Geografia e Estatística

IDH : Índice de desenvolvimento humano

INCA : Instituto Nacional do Câncer

SES-AM : Secretaria de Estado de Saúde do Amazonas

SIM : Sistema de Informação sobre Mortalidade

SUS : Sistema Único de Saúde

WHO : WHO HEALTH ORGANIZATION, Organização Mundial de Saúde

OMS : Organização Mundial da Saúde

MS : Ministério da Saúde

IARC : International Agency for Research on Cancer, Agência Internacional de Pesquisa sobre o Câncer

CAPES : Coordenação de Aperfeiçoamento de Pessoal de Novel Superior

PNUD : Programa das Naçøes Unidas para o Desenvolvimento

IPEA : Instituto de Pesquisa Econômica Aplicada

FJP :Fundação João Pinheiro

Fiocruz : Fundação Oswaldo Cruz

APC : Annual Percentage Change, Variações Percentuais Anuais

IIA : Estágio IIA

IIIB : Estágio IIIB

SUMÁRIO

1 INTRODUÇÃO **17**

1.1 Contextualização 17

1.2 Objetivos 19

1.2.1 Objetivo geral 19

1.2.2 Objetivos específicos 19

1.3 Organização da Dissertação 20

2 REVISÃO BIBLIOGRÁFICA **21**

3 MATERIAL E MÉTODOS **23**

3.1 Considerações Éticas 23

3.2 Tipo de estudo 23

3.2.1 Divisões territoriais 23

3.2.2 População 25

3.3 Fonte de Dados 26

3.4 Padronização 29

3.5 Variáveis do estudo 30

3.5.1 Variáveis desfecho 31

3.5.2 Variáveis explicativas 31

3.6 Modelos 33

3.6.1 Modelos Lineares Generalizados 34

3.6.1.1 Estrutura dos Modelos Lineares Generalizados . . . 34

3.6.1.2 Estimação dos Parâmetros nos Modelos Lineares Generalizados 36

3.6.1.3 Interpretação dos Coeficientes e Identificação de Fatores de Risco . . . 38
3.6.1.4 Avaliação do Ajuste e Diagnóstico do Modelo . . . 38
3.6.1.5 Medidas descritivas . . . 38
3.6.1.6 Ajuste do Modelo . . . 39
3.6.1.7 Avaliação da Dispersão dos Dados . . . 40
3.7 Estatística Espacial . . . 40
3.7.1 Método de varredura circular de kulldorf . . . 43
3.8 Softwares . . . 47
3.8.1 SaTScan . . . 47
3.8.2 RStudio . . . 48
3.8.3 TabWin . . . 48

4 RESULTADOS 50

4.1 Análise descritiva . . . 50
4.1.1 Taxas brutas . . . 50
4.1.2 Taxas padronizadas . . . 54
4.2 Análise modelos . . . 60
4.2.0.1 Interpolação das variáveis socioeconômicas . . . 60
4.2.0.2 Seleção das covariavéis . . . 61
4.2.0.3 Analise descritiva covariaveis . . . 63
4.2.1 Resultaldo dos modelos . . . 65
4.2.1.1 Modelo linear generalizado . . . 65
4.2.1.2 Modelo Geral . . . 66
4.2.1.3 Modelo câncer de mama . . . 69
4.2.1.4 Modelo câncer do colo do útero . . . 70

4.3 Análise espacial 71

4.3.1 Análise descritiva temporal 71

4.3.2 Análise puramente espacial 79

5 CONCLUSÕES **90**

6 APÊNDICE **93**

6.1 Apêndice A - Tabelas Complementares 93

6.2 Apêndice B - Figuras Complementares 99

Referências **106**

1 INTRODUÇÃO

1.1 Contextualização

O câncer compreende um conjunto de doenças que se caracterizam pelo crescimento anormal e descontrolado de células que podem invadir outros tecidos e formar tumores (Brasil, Ministério da Saúde, 2025). Os tumores malignos podem ser categorizados conforme o tecido de origem. Assim, a transformação celular ocorre em tecidos epiteliais, denominados carcinomas e sarcomas, quando estes se originam de tecidos conjuntivos que atingem cartilagens, ossos ou musculos (Teixeira, 2021).

Epidemiologicamente, o câncer constitui um dos principais problemas de saúde pública no mundo. Em 2020 o câncer foi estatisticamente considerado uma das causas de óbitos (WHO, 2025) no mundo. Nesse período, as taxas padronizadas por idade de incidência e mortalidade foram de, respectivamente, 196 e 91 mortes/100 mil habitantes, indicando sua importância global (IARC, 2025).

Em 2022, os valores de incidência e mortalidade dos cinco principais tipos de neoplasias no mundo para ambos os sexos foram, respectivamente, de próstata: 76,3 e 13,5; mama: 63,1 e 13,9; colorretal: 19,8 e 9,2; pulmão: 14,6 e 12,5 e colo do útero: 12,7 e 6,5(IARC, 2025). Por outro lado, entre as mulheres, a relevância desses agravos torna-se ainda mais evidente. O câncer de mama torna-se o mais incidente e o principal responsável pela mortalidade feminina, enquanto o câncer do colo do útero ascende para a quarta posição em incidência e em óbitos no mundo conforme podemos verificar na Figura 20, apresentada no Apêndice B.

Neste contexto, o câncer de mama e o câncer do colo do útero destacam-se por suas elevadas taxas de incidência e mortalidade, tanto em países desenvolvidos quanto naqueles em desenvolvimento (Oliveira et al., 2020). Em 2016 dados nacionais indicam que o câncer de mama apresentou-se como a principal causa de óbito por neoplasias entre as mulheres, enquanto o câncer do colo do útero ocupou a quarta posição

(Bermudi, 2020). Informações mais recentes mostram que, em 2023, a incidência e a mortalidade por câncer no país permaneceram elevadas, destacando-se o câncer de mama como principal neoplasia em termos de ocorrência e óbitos entre mulheres e o câncer do colo do útero como o terceiro mais incidente e quarto mais letal nesse grupo populacional (INCA, 2023; INCA, 2023).

A ocorrência e a mortalidade por câncer estão diretamente ligados aos determinantes de saúde, que correspondem a fatores sociais, econômicos, culturais e ambientais que influenciam a condição de saúde dos indivíduos e comunidades. De forma complementar, temos que os fatores de risco se referem a aspectos de hábitos pessoais ou de exposição ambiental, que aumentam a probabilidade de ocorrência de alguma doença (Bonita, Beaglehole e Kjellström, 2010). Dessa forma, a epidemiologia do câncer está associada a diferenças socioeconômicas, étnicas e raciais que ocorrem em populações específicas e são determinantes das disparidades de saúde definidas nos processos de adoecimentos e mortes por essas doenças (Oliveira et al., 2020).

Especificamente, a incidência e a mortalidade por câncer do colo do útero apresentam distribuição heterogênea em escala global, estando associadas às desigualdades no acesso aos serviços de saúde, bem como a determinantes de natureza biológica, ambiental, socioeconômica e comportamental (Oliveira et al., 2020). Ademais, temos que, para o desenvolvimento do câncer do colo do útero, a infecção persistente por subtipos do vírus do papiloma humano (HPV) é considerada uma causa necessária (Bermudi, 2020). Além disso, fatores associados à subutilização dos serviços de rastreamento e diagnóstico precoce, como o teste citopatológico (Papanicolaou), bem como a elevada prevalência de infecções cervicovaginais (Oliveira et al., 2020).

Por sua vez, o câncer de mama apresenta maior ocorrência em países e regiões com níveis mais elevados de desenvolvimento econômico, possivelmente associado a características do estilo de vida, bem como a fatores comportamentais e reprodutivos das mulheres, onde destacam-se a nuliparidade, a idade avançada na primeira gestação e a menopausa tardia (Oliveira et al., 2020; Bermudi, 2020).

No Brasil, evidenciam-se discrepâncias territoriais. Observa-se uma tendência geral de redução das taxas de mortalidade por câncer do colo do útero nas regiões Sudeste, Sul e Centro-Oeste, tanto nas capitais quanto nos demais municípios. Contudo, as regiões Norte e Nordeste apresentam diminuição das taxas apenas nas capitais, com aumento nos municípios do interior (Girianelli, Gamarra e Silva, 2014). Dessa forma,

a Região Norte apresenta um cenário desigual em diversos aspectos, em oposição às tendências observadas nas regiões mais desenvolvidas do Sul e Sudeste.

Diante desse cenário, a análise conjunta dos padrões espaciais e dos fatores de risco associados à mortalidade por câncer de mama e do colo do útero no estado do Amazonas torna-se fundamental. A utilização de modelos de regressão e técnicas de análise espacial permite não apenas quantificar os riscos associados a características sociodemográficas, mas também identificar áreas de maior vulnerabilidade, contribuindo para o planejamento de políticas públicas mais equitativas e orientadas às especificidades territoriais.

1.2 Objetivos

1.2.1 Objetivo geral

Analisar os padrões espaciais da mortalidade por câncer de mama e por câncer do colo do útero no Amazonas, no período de 2010 a 2023, identificando os fatores de risco associados por meio de modelos de regressão e mapeando as desigualdades territoriais com o uso de técnicas de análise espacial.

1.2.2 Objetivos específicos

1. Estimar taxas de mortalidade padronizadas por faixa etária e ano, proporcionando uma visão detalhada da evolução temporal e geográfica da mortalidade;

2. Aplicar modelos de regressão para quantificar os riscos associados a variáveis sociodemográficas, permitindo entender melhor os fatores que influenciam a mortalidade;

3. Identificar clusters espaciais de mortalidade por câncer de mama e colo do útero, por meio de análise puramente espacial, de modo a evidenciar as desigualdades territoriais.

1.3 Organização da Dissertação

Esta dissertação encontra-se estruturada em 5 capítulos, partindo da contextualização do câncer, seus impactos e determinantes, principais fatores de risco, a análise de sua distribuição espacial e por fim os resultados obtidos.

Inicialmente, o Capítulo 1 introduz o tema da pesquisa, apresentando a motivação do estudo, o problema investigado e a definição dos objetivos geral e específicos que orientam a análise.

Por seguinte, no Capítulo 2, desenvolve-se a revisão bibliográfica, na qual são discutidos padrões epidemiológicos dos cânceres de mama e do colo do útero e seus principais fatores demográficos e socioeconômicos, com foco nas desigualdades que contribuem para o óbito.

Além disso, no Capítulo 3, são descritos os procedimentos metodológicos adotados, detalhando-se o delineamento do estudo, as divisões territoriais consideradas, a população analisada, as fontes de dados utilizadas e as variáveis selecionadas, bem como os modelos estatísticos empregados, as técnicas de estatística espacial utilizadas para a identificação de padrões territoriais e os softwares que subsidiaram as análises.

Por conseguinte, no Capítulo 4, são apresentados os resultados, que contemplam a análise descritiva das taxas de mortalidade, os desfechos dos modelos estatísticos e a identificação de padrões e aglomerados espaciais da mortalidade por câncer de mama e por câncer de colo do útero no Amazonas.

Por fim, o Capítulo 5 reúne as conclusões do estudo, destacando as principais percepções, suas implicações para a compreensão das desigualdades territoriais em saúde e as proposições para investigações futuras. Os apêndices complementam o trabalho com tabelas e figuras adicionais que auxiliam na interpretação dos resultados.

2 REVISÃO BIBLIOGRÁFICA

Desenvolvemos, neste capítulo, uma revisão de literatura voltada ao entendimento dos padrões epidemiológicos dos cânceres de mama e do colo do útero e seus principais fatores demográficos e socioeconômicos, com foco nas desigualdades que contribuem para o óbito.

As doenças crônicas não transmissíveis (DCNT) — como doenças cardiovasculares, respiratórias crônicas, neoplasias e diabetes — são responsáveis por cerca de 70% de todas as mortes no mundo (Malta et al., 2017). Dentre elas, o câncer se destaca como um dos maiores desafios de saúde pública devido à sua grande morbimortalidade (WHO, 2025; INCA, 2025). Além disso, entre as mulheres no Brasil, em 2023, a maior incidência ocorreu para o tumor maligno de mama, enquanto o câncer do colo do útero ocupou a terceira posição (INCA, 2025). Acerca da mortalidade, o câncer de mama apresentou a maior taxa de óbitos e o câncer do colo do útero na quarta posição (INCA, 2014). Assim, a relevância desses dois agravos é reforçada pelas desigualdades associadas aos determinantes sociais da saúde, que influenciam profundamente padrões de ocorrência, acesso aos serviços e desfechos na população feminina, tanto no contexto global quanto no brasileiro (Girianelli, Gamarra e Silva, 2014; Oliveira et al., 2020).

No país, segundo Girianelli, Gamarra e Silva (2014), acerca da neoplasia maligna do colo do útero observa-se uma tendência de queda na mortalidade entre 1980 e 2006, com oscilações de 2006 a 2010. Contudo, essa redução ocorre de forma desigual, visto que nas regiões Norte e Nordeste, somente as capitais apresentam esse comportamento, em contrapartida com as residentes dos municípios do interior que apresentam taxas elevadas. Além disso, os mesmos autores apontam um cenário distinto para o câncer de mama, com tendência de aumento das taxas de óbitos, entretanto com declínio a partir de 1990 para todas as regiões, exceto para a região Norte e interior do Nordeste.

Dessa forma, temos que o câncer do colo do útero parece ter forte relação com

contextos de vulnerabilidade socioeconômica e desigualdade social. Taxas elevadas de mortalidade estão estatisticamente associadas a maiores valores do Índice de Gini e a menores níveis do Índice de Desenvolvimento Humano (IDH). A letalidade dessa neoplasia é mais acentuada entre mulheres com menor escolaridade e residentes em regiões menos desenvolvidas, evidenciando o acesso limitado aos serviços de rastreamento e diagnóstico precoce (Oliveira et al., 2020).

Em contrapartida, para o câncer de mama verifica-se cenário inverso, uma vez que as taxas de mortalidade mostram correlação direta com indicadores positivos, como maior densidade de médicos generalistas e maior oferta de centros de oncologia licenciados (Oliveira et al., 2020). Observa-se, entretanto, tendência de aumento em regiões do interior (Girianelli, Gamarra e Silva, 2014). Além disso, a região norte apresenta "dupla carga" de neoplasias femininas, devido a prevalência dos cânceres de mama e do colo do útero (Oliveira et al., 2020). A idade se destaca como um fator determinante e de risco para a mortalidade por câncer de mama e do colo do útero (Bermudi, 2020). Observa-se que a mortalidade aumenta com o avanço da idade, sendo mais elevada nas faixas etárias mais avançadas. De modo geral, a mortalidade entre mulheres com 70 anos ou mais pode ser até 31,4 vezes superior à observada entre mulheres jovens (20–39 anos) para o câncer de mama, e 11,4 vezes maior para o câncer do colo do útero (Bermudi, 2020).

Entretanto, embora menos frequente, o câncer de mama em mulheres jovens (≤ 40 anos) apresenta características clínicas mais agressivas e está associado a prognóstico menos favorável (Pereira, 2016). No contexto regional, enquanto a mortalidade por câncer do colo do útero apresenta tendência de redução na maior parte do território brasileiro, municípios do interior da Região Norte demonstram aumento significativo dessas taxas. Sendo, o maior crescimento anual da mortalidade por câncer de mama no país (APC = 5,0%) foi registrado no interior da Região Norte no período de 1999 a 2010 (Girianelli, Gamarra e Silva, 2014). Na região amazônica, a faixa etária jovem possui fator negativo para o diagnóstico precoce. Primeiro, pela densidade mamária em mulheres jovens, que dificulta a detecção precoce da doença. Por seguinte, em relação ao rastreamento, uma vez que aproximadamente 33,6% das pacientes jovens na região identificam a doença por meio do autoexame ou exame clínico, frequentemente já em estágios mais avançados, como IIA (27,5%) e IIIB (21,3%) (Pereira, 2016).

3 MATERIAL E MÉTODOS

3.1 Considerações Éticas

Este estudo utilizou exclusivamente dados secundários, de domínio público, disponivéis principalmente pelo Ministério da Saúde no site <https://datasus.saude.gov.br/>. Vale ressaltar que não é possível a identificação individual, não sendo necessária a submissão ao Comitê de Ética em Pesquisa.

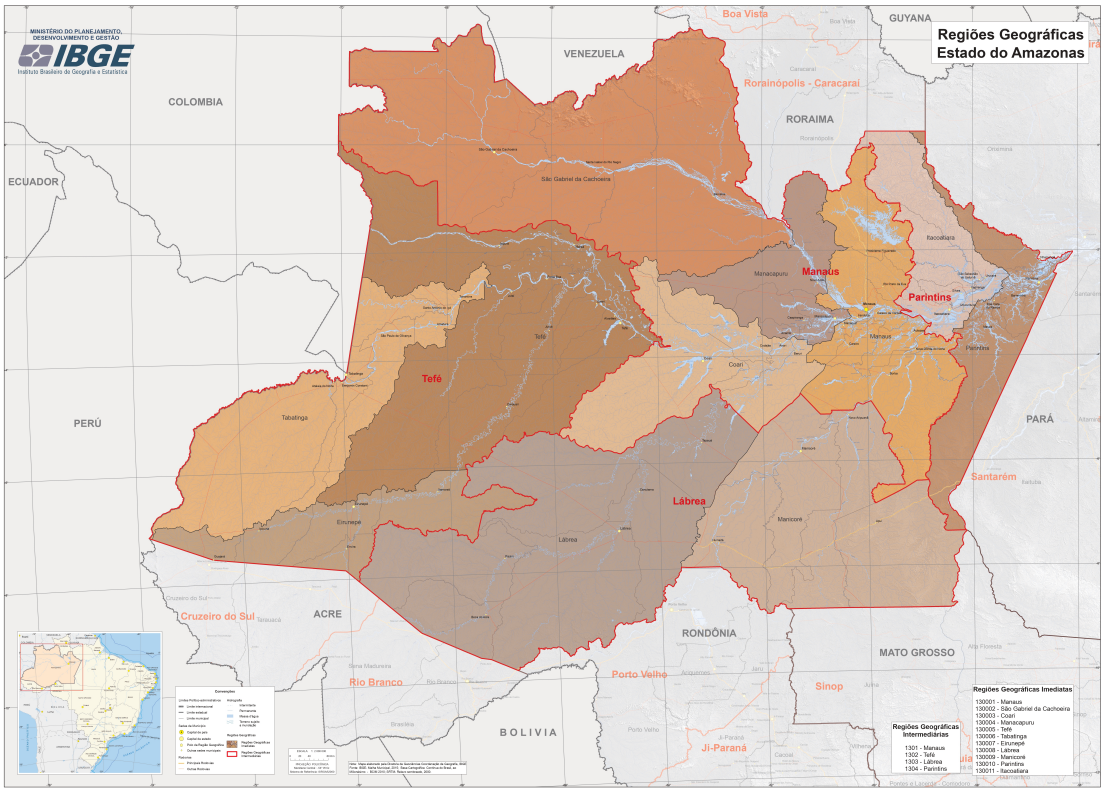
3.2 Tipo de estudo

Apresentamos neste trabalho um estudo caracterizado como quantitativo, descritivo e analítico, obtido através de dados secundários, com delineamento ecológico, utilizando macrorregiões no período de 2010 a 2016. A pesquisa abrange os óbitos ocorridos no Amazonas por câncer do colo do útero (C53) e câncer de mama feminino (C50), analisados em sete faixas etárias possuindo intervalos de dez anos ($20-29, \ldots, 70-79$) até o agrupamento de 80 anos ou mais. Além disso, foram utilizadas técnicas de análise espacial e modelagem estatística para aprofundar a compreensão dos padrões geográficos e associações das variaveis observadas.

3.2.1 Divisões territoriais

O amazonas possui uma área territorial de 1.558.706,127 km², composto de 62 municípios, representando 18% do país, com população de 3.941.613, com uma das menores densidades demográficas, com 2,53 hab/km² (2022), rendimento per capita de 1.238R$ (2024) e com o IDH de 0,7 (2021) (IBGE, 2025). A saber, a distribuição territorial e a localização dos municípios podem ser observadas na Figura 1.

Figura 1: Regiões Geográficas do Amazonas segundo o IBGE (2017).



Fonte: IBGE (2017).
Disponível em: <https://geoftp.ibge.gov.br/organizacao_do_territorio/divisao_regional/divisao_regional_do_brasil/divisao_regional_do_brasil_em_regioes_geograficas_2017/mapas/13_regioes_geograficas_amazonas.pdf>

O estado é dividido em formas distintas de organização territorial variando conforme a finalidade, político-administrativos ou regiões de saúde. Dessa forma, considerando a oferta e acesso aos serviços assistenciais foram institucionalizadas as nove regiões e ampliado para as três macrorregiões de assistência em saúde (SES-AM, 2023).

Neste estudo, optou-se por utilizar as divisões territoriais por estado e macrorregiões de saúde na análise descritiva, devido a disponibilidade de variáveis em nível agregado e ao grande número de municípios do Amazonas, conforme apresentado na Tabela A3 (Apêndice A). A agregação territorial nesse nível permitiu maior estabilidade das estimativas descritivas, reduzindo oscilações decorrentes de pequenos números e garantindo comparabilidade. Por seguinte, para as análises de modelagem estatística e abordagem espacial, utilizou-se o nível municipal, a menor unidade administrativa com disponibilidade das informações de óbitos, por possibilitar maior detalhamento na identificação de padrões espaciais. Portanto, esta escolha se adequa ao escopo métodológico do projeto sem prejuízo com outras formas de agregação territo-

rial.

Tabela 1: Macrorregiões, Regiões de Saúde e Indicadores Demográficos do Amazonas

Macrorregião	Região de Saúde	Número de Municípios	População	% Pop.	Área (Km^2)	Densidade (hab/Km^2)
Central	Manaus, Entorno e Alto Rio Negro	12	2.401.411	61%	365.274	6,57
	Rio Negro e Solimões	8	273.152	7%	157.197	1,74
	Purus	5	136.736	3%	250.416	0,55
	Macrorregião Central	**25**	**2.811.299**	**71%**	**772.887**	**3,64**
Leste	Madeira	5	188.932	5%	221.083	0,85
	Médio Amazonas	6	179.565	5%	58.335	3,08
	Baixo Amazonas	5	232.545	6%	68.393	3,40
	Macrorregião Leste	**16**	**601.045**	**15%**	**347.863**	**1,73**
Oeste	Juruá	6	128.161	3%	93.191	1,38
	Triângulo	6	139.129	4%	132.203	1,05
	Alto Solimões	9	262.269	7%	213.311	1,23
	Macrorregião Oeste	**21**	**529.269**	**13%**	**438.507**	**1,21**
Total		**62**	**3.941.613**	**100%**	**1.559.257**	**2,53**

Fonte: Adaptado de IBGE/Diretoria de Pesquisas – DPE. Coordenação de População e Indicadores Sociais.

A Tabela 1 reúne informações resumidas sobre essas áreas geograficas. Assim, apesar da divisão de municipíos ser relativamente proporcional, verificamos que a maior concentração ocorre na macrorregião central, visto que Manaus possui quase 53% da população Amazonense (IBGE, 2025).

3.2.2 População

As taxas brutas e padronizadas de mortalidade foram calculadas para cada ano do período analisado, utilizando a população correspondente a cada período, município, faixa etária e sexo. Para os anos de 2010 e 2022, adotou-se a população proveniente do Censo Demográfico (IBGE, 2010; IBGE, 2022). Nos anos intercensitários (2011–2021 e 2023), foram empregadas as estimativas populacionais disponibilizadas pelo (Brasil. Ministério da Saúde, 2024). Além disso, para padronização foram utilizadas a população padrão da OMS (NCI, 2013). Essas populações serviram como denominadores

para o cálculo das taxas utilizadas nas análises subsequentes.

3.3 Fonte de Dados

No desenvolvimento desta dissertação, foram utilizadas bases de dados públicas oficiais, sendo estas: do Ministério da Saúde, instituições brasileiras e internacionais. Disponibilizando informações sobre mortalidade, demografia, indicadores socioeconômicos e produção assistencial. A seguir, apresenta-se um resumo de cada fonte, destacando seu conteúdo e relevância para a pesquisa.

Foram utilizadas como fontes de dados o Sistema de Informações sobre Mortalidade (SIM), o Atlas do Desenvolvimento Humano no Brasil e o Instituto Brasileiro de Geografia e Estatística (IBGE).

Os dados de mortalidade foram obtidos a partir do SIM, base oficial do Ministério da Saúde alimentada pelas Declarações de Óbito (DO), preenchidas por profissionais médicos e posteriormente consolidadas nas esferas municipal, estadual e federal. O acesso foi realizado por meio do sistema TabNet do DATASUS, plataforma online que permite a tabulação das informações segundo variáveis como período, local de residência, idade e causa básica do óbito (CID-10).

As informações socioeconômicas foram extraídas do Atlas do Desenvolvimento Humano no Brasil, que disponibiliza indicadores derivados dos Censos Demográficos, organizados por diferentes recortes territoriais (Brasil, estados e municípios), com acesso por meio de tabelas interativas disponíveis na plataforma online.

Adicionalmente, foram utilizados dados do IBGE, obtidos por meio do sistema SIDRA, que permite a consulta e extração de tabelas estatísticas com recorte por variáveis, período e unidade geográfica, abrangendo informações demográficas e socioeconômicas.

Plataforma DATASUS

O DATASUS é um sistema provedor de softwares para as secretarias estaduais e municipais de saúde. Tem como principal ferramenta fornecer aos órgãos do SUS sistemas de informação e suporte de informática necessários para o planejamento, operação e controle. (Brasil, Ministério da Saúde, 2025). Para este estudo, foram utilizados dados demográficos, de mortalidade e produção ambulatorial obtidas através do TABNET – ferramenta de tabulação online – ou TABWIN – tabulação de dados offline. A utilização dessas informações foi fundamental para consolidação do banco final utilizado nas análises no território estudado.

Sistema de Informação sobre Mortalidade

O SIM é a base oficial de registro que reúne informações sistematizadas sobre mortalidade no país. Tem como principal ferramenta a Declaração de Óbito (DO) que contém informações sobre a causa básica (CID-10) e variáveis demográficas e geográficas (Brasil. Ministério da Saúde, 2025). Para este estudo, destacam-se as referentes à idade, ao município de residência e à causa básica do óbito. Dessa forma, a utilização do SIM permite a construção de taxas de mortalidade brutas e padronizadas por câncer de mama e do colo do útero no Amazonas.

Sistema de Informações Ambulatoriais

O SIA é um sistema para coletar, processar e gerenciar dados de atendimentos ambulatoriais de complexidade básica, média e alta. Tem como principal ferramenta o Boletim de Produção Ambulatorial que registra informações acerca dos procedimentos, quantidades e estabelecimento executante (Brasil. Ministério da Saúde, 2025). Para este estudo, foram utilizados dados relativos aos procedimentos relacionados ao câncer de mama e ao câncer do colo do útero, organizados por competência e localidade. A utilização dessas informações é fundamental para a caracterização da produção assistencial vinculada a exames e procedimentos de rastreamento e diagnóstico, contribuindo para a análise da oferta de serviços de saúde no território estudado.

Instituto Brasileiro de Geografia e Estatística

O IBGE é a principal fonte de informações sobre a população, território e economia no país. Tem como principal ferramenta a coleta através de pesquisas, como o Censo Demográfico (IBGE, 2025). Para este estudo, destacam-se a utilização de dados populacionais do IBGE, incluindo estimativas intercensitárias, projeções e indicadores demográficos. Dessa forma, o uso dessas informações são fundamentais para o cálculo das taxas brutas e padronizadas por idade. Ademais, também foram utilizados indicadores municipais, como densidade demográfica, razão de sexo e estrutura etária, que auxiliam na interpretação das desigualdades sociodemográficas.

Atlas do Desenvolvimento Humano no Brasil

O Atlas é uma plataforma responsável por divulgar informações sobre indicadores de desenvolvimento humano para diferentes recortes territoriais do país. Tem como principal ferramenta a divulgação de dados consolidados de forma descomplicada (PNUD,IPEA e FJP, 2025). Para este estudo, destacam-se a utilização de dados como o IDH municipal (IDH-M) e suas dimensões — educação, renda e longevidade. Dessa forma, o uso dessas informações são fundamentais para auxiliar na interpretação das desigualdades sociodemográficas referentes à mortalidade por câncer de mama e do

colo do útero no Amazonas.

Outras Bases Complementares

Por fim, também foram utilizados relatórios técnicos, boletins epidemiológicos e publicações MS, FIOCRUZ, INCA, WHO e dados populacionais intercesitárdo DATASUS para contextualização e embasamento teórico dos fatores determinantes e tendências epidemiológicas.

3.4 Padronização

Segundo o INCA (2014), no Brasil, entre 2010 e 2023, as taxas de mortalidade específicas por câncer de mama e do colo do útero até 14 anos são de 0 óbitos/100 mil mulheres-ano e, para a faixa etária de 15 a 19 anos, respectivamente, de 0,02 e 0,03 óbitos/100 mil mulheres-ano. Assim, a análise foi restrita a mulheres com 20 anos ou mais, uma vez que a mortalidade é rara em idades precoces.

Além disso, a fim de permitir comparação entre municípios com diferentes estruturas etárias, as taxas de mortalidade por faixa etária foram padronizadas utilizando a população padrão mundial da OMS (2000–2025) (NCI, 2013), conforme utilizado em estudos nacionais, tais como Teixeira (2021), Bermudi (2020), Girianelli, Gamarra e Silva (2014).

Dessa forma, tendo em vista que a análise inclui apenas a população com 20 anos ou mais a população padrão da OMS foi repadronizada. Para isso, foram selecionadas as proporções correspondentes a cada faixa etária ($\geq$ 20 anos) e, em seguida, divididas pela soma total dessas proporções, de forma a manter o total igual a 1 (ou 100%), conforme apresentado na Tabela A2 (Apêndice A). Assim, os novos pesos padronizados foram obtidos por meio de

$$w_i^* = \frac{w_i}{\sum_{j \in F} w_j}, \quad \text{para } i \in F,$$

em que

- w_i^* representa o peso padronizado ajustado da faixa etária i;
- w_i é o peso original da OMS para a faixa etária i;

- F é o conjunto das faixas etárias consideradas ($\geq$ 20 anos).

Assim, o cálculo da taxa específica (T_e) por faixa etária foi obtido por

$$T_e = \frac{O_{i,j,t}}{P_{i,j,t}} \times 100.000,$$

em que

- i representa a faixa etária, j a localidade e t representa o ano;
- $O_{i,j,t}$ representa o número de óbitos;
- $P_{i,j,t}$ representa a população feminina da faixa etária i.

A taxa padronizada de mortalidade (T_p) foi então obtida pela média ponderada das taxas específicas por faixa etária, conforme a fórmula:

$$T_p = \sum_{i \in F} w_i^* \times T_e,$$

em que

- T_e é a taxa de mortalidade específica da faixa etária i;
- w_i^* são os pesos padronizados da população padrão (acima de 20 anos).

3.5 Variáveis do estudo

As covariáveis utilizadas nos modelos foram obtidas a partir de diferentes bases secundárias, cada uma com sua própria periodicidade de atualização. Assim, os valores apresentados correspondem ao ano de 2022, referente ao Censo mais atualizado, ou, quando não disponível, ao ano mais próximo e métodologicamente compatível.

A saber, indicadores como IDH-M possuem atualização periódica lenta, cujos últimos dados oficiais são referentes a 2021. Logo, foi utilizado o último ano disponível. Da mesma forma, variáveis populacionais, econômicas e de composição demográfica foram obtidas da fonte mais recente e confiável, preservando a coerência temporal necessária para comparação entre municípios.

Essa escolha garante que todas as covariáveis representem, de forma consistente, o cenário socioeconômico e demográfico mais atual possível, minimizando distorções

decorrentes de desatualização e mantendo a comparabilidade entre os municípios do Amazonas no período analisado (2010–2023).

3.5.1 Variáveis desfecho

As **taxas de mortalidade feminina** por câncer do colo do útero e por câncer de mama foram definidas como as variáveis desfecho (dependentes) do estudo, calculadas conforme apresentado a seguir:

$$T_{pad} = \sum_{i=1}^{k} T_i \times w_i*, \tag{3.1}$$

em que

- T_{pad} = taxa de mortalidade padronizada por idade;
- T_i = taxa específica de mortalidade na faixa etária i;
- w_i* = peso da faixa etária i;
- k = número total de faixas etárias consideradas.

3.5.2 Variáveis explicativas

Todas as variáveis descritas foram analisadas nos níveis: municipal e estadual, tendo como referência o ano de 2022. As exceções correspondem às variáveis que abrangem períodos mais extensos ou mais próximos do ano de referência, as quais foram explicitados na própria descrição. A seguir estão descritas as variáveis **Populacionais, Territoriais e Demográficas** consideradas no estudo:

A **população residente** representa o total de indivíduos residentes no período de 2010 a 2023. Sua unidade de medida é número absoluto de pessoas, categorizada por faixas etárias acima de 20 anos, organizadas em intervalos de 10 anos até a faixa 80 ou mais anos. Dessa forma, foi utilizada para o cálculo de taxas e verificação da estrutura etária.

A **área da unidade territorial** corresponde à extensão territorial oficial. Sua unidade de medida é quilômetros quadrados (km^2). Dessa forma, foi utilizada como variável descritiva do espaço geográfico.

A **densidade demográfica** expressa a relação entre a população residente e a área territorial habitada. Sua unidade de medida é de habitantes por quilômetro quadrado (hab/km^2). Dessa forma, foi utilizada como indicador demográfico e territorial.

A **População por raça/cor feminina** representa a distribuição populacional segundo raça/cor autodeclarada. Sua unidade de medida é o número absoluto de pessoas. Dessa forma, foi utilizada para caracterizar a composição étnico-racial.

A **Taxa de crescimento geométrico da população** representa a velocidade de crescimento ou redução populacional ao longo do período. Sua unidade de medida é a porcentagem (%). Dessa forma, foi utilizada como indicador da dinâmica populacional.

A **Taxa de Envelhecimento** representa a relação entre a população idosa e a população jovem de um determinado território, indicando o grau de envelhecimento populacional. Sua unidade de medida é expressa em percentual (%), conforme disponibilizado no Atlas Brasil. Dessa forma, foi utilizada como variável demográfica descritiva, permitindo caracterizar a estrutura etária da população analisada.

A **cobertura do exame citopatológico e de mamografia** representa o rastreamento do câncer de colo do útero e mamografia, com foco em mulheres de 25 a 64 e 50 a 69 anos, respectivamente. Sua unidade de medida é a razão entre os exames e a população na faixa de interesse. Dessa forma, foi utilizada para contextualizar o público-alvo das políticas de detecção precoce. A seguir temos a forma de cálculo:

$$\text{Cobertura Citopatológica}_{(m,t)} = \frac{3\left(\text{Freq}_{0203010019\,(m,t)} + \text{Freq}_{0203010086\,(m,t)}\right)}{\text{População feminina 25–64 anos}_{(m,t)}} \tag{3.2}$$

$$\text{Cobertura Mamográfica}_{(m,t)} = \frac{2 \cdot \text{Freq}_{0204030188\,(m,t)}}{\text{População feminina 50–69 anos}_{(m,t)}} \tag{3.3}$$

A seguir são descritas as variáveis **socioeconômicas** consideradas no estudo:

O **Índice de Gini** é uma medida de desigualdade de renda. Sua unidade de medida varia de 0 a 1, sendo que valores próximos de 0 indicam maior igualdade de renda, enquanto valores próximos de 1 indicam maior desigualdade. Dessa forma, foi utilizado como medida de desigualdade de renda.

O **IDHM** é uma medida resumo que avalia o progresso de longo prazo em três

pilares: vida longa e saudável, o acesso ao conhecimento e um padrão de vida decente. Sua unidade de medida varia de 0 a 1, sendo que valores próximos de 0 indicam menores condições de renda, educação e longevidade. Dessa forma, foi utilizado para captar diferenças no nível de desenvolvimento humano.

O **Rendimento médio** representa o valor médio do rendimento da população residente. Sua unidade de medida é reais (R$). Dessa forma, foi utilizado como indicador econômico.

A Porcentagem de Extremamente Pobres representa a proporção da população residente em determinado território que se encontra em situação de extrema pobreza, conforme critério adotado pelo Atlas Brasil. Sua unidade de medida é expressa em percentual (%). Dessa forma, foi utilizada como variável socioeconômica descritiva, permitindo caracterizar as condições de vulnerabilidade social da população analisada.

3.6 Modelos

A análise estatística é o processo que se caracteriza pela transformação de dados em conhecimento. Inicia-se na coleta das observações que permitem a obtenção de fatos. Em seguida, são geradas informações por meio da análise exploratória e, por fim, essas informações auxiliam a tomada de decisões que se consolidam como conhecimento (Silva, H. F. S., 2021). Embora a análise exploratória permita a descrição inicial dos dados, as relações entre variáveis e a estimativa de efeitos associados a fenômenos de interesse demandam o uso de modelos estatísticos mais estruturados. Assim, os modelos de regressão são essenciais para analisar a relação em que a variável de resposta (ou dependente) pode ser estimada por uma combinação linear das variáveis explicativas (Silva, Peres and Boscarioli, 2016). Dessa forma, a seleção de modelos é uma parte essencial e ocorre através da procura de um modelo que descreva de forma satisfatória as observações e seja o mais simples possível (Cordeiro and Demétrio, 2007).

Assim, temos que, segundo Casella and Berger (2001), uma regressão linear simples é dada por

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \tag{3.4}$$

onde Y_i é uma variável aleatória, x_i é uma variável observável. As quantidades α e β são, respectivamente, o intercepto e a inclinação da regressão, assumidas como

parâmetros fixos e desconhecidos, e ε_i é outra variável aleatória. Aqui supomos que $E(\varepsilon_i) = 0$ e $\text{Var}(Y_i) = \sigma^2$.

3.6.1 Modelos Lineares Generalizados

Os Modelos Lineares Generalizados (MLGs) descrevem a relação entre a média de uma variável resposta Y e um conjunto de variáveis explicativas, permitindo que essa relação seja mais flexível do que aquela assumida nos modelos lineares clássicos (Casella e Berger, 2022).

Diferentemente da regressão linear tradicional, sob condições específicas, como erro com média zero e relação direta entre média e preditor, na qual se assume que

$$\mathbb{E}(Y_i) = \alpha + \beta x_i.$$

Nos Modelos Lineares Generalizados a relação entre a média da variável resposta e as covariáveis é estabelecida por meio de uma função de ligação $g(\cdot)$, tal que

$$g(\mathbb{E}(Y_i)) = \alpha + \beta x_i,$$

em que $g(\cdot)$ é uma função monótona e diferenciável que possibilita que a média da variável resposta esteja relacionada ao preditor linear para diversos modelos estatísticos, tais como os modelos Binomial, Poisson ou Normal (Casella e Berger, 2022).

3.6.1.1 Estrutura dos Modelos Lineares Generalizados

Para melhor compreensão acerca dos MLGs temos que são compostos por três elementos fundamentais: o componente aleatório, o componente sistemático e a função de ligação. O componente aleatório refere-se à distribuição da variável resposta, assumida como pertencente à família exponencial. Em seguida, o componente sistemático consiste em um preditor linear que relaciona as covariáveis aos parâmetros do modelo. Por fim, a função de ligação estabelece a conexão entre a média da variável resposta e o preditor linear (Casella e Berger, 2022).

Segundo Cordeiro and Demétrio (2007) o componente aleatório é representado por um conjunto de variáveis aleatórias independentes $Y_1, \ldots, Y_n$, com distribuição comum

e pertencente à família exponencial, com médias $\mu_1, \ldots, \mu_n$, isto é,

$$\mathbb{E}(Y_i) = \mu_i, \quad i = 1, \ldots, n.$$

A função densidade de probabilidade de Y_i pode ser escrita como

$$f(y_i; \theta_i, \phi) = \exp\left\{\phi^{-1}\left[y_i\theta_i - b(\theta_i)\right] + c(y_i, \phi)\right\},$$

em que θ_i é o parâmetro canônico, $\phi > 0$ é o parâmetro de dispersão e $b(\cdot)$ e $c(\cdot)$ são funções conhecidas. Nessas condições, tem-se que

$$\mathbb{E}(Y_i) = \mu_i = b'(\theta_i) \quad \text{e} \quad \text{Var}(Y_i) = \phi b''(\theta_i) = \phi V_i,$$

sendo $V_i = V(\mu_i) = \mathrm{d}\mu_i/\mathrm{d}\theta_i$ a função de variância, que depende exclusivamente da média. O parâmetro canônico pode ser expresso como

$$\theta_i = \int V_i^{-1}\,\mathrm{d}\mu_i = q(\mu_i),$$

onde $q(\cdot)$ é uma função conhecida da média. Dada uma forma funcional para a função de variância $V(\mu)$, o parâmetro canônico é determinado e, consequentemente, a distribuição da variável resposta fica especificada dentro da família exponencial. Essa classe de distribuições é particularmente importante por permitir a modelagem de dados assimétricos, discretos ou contínuos, bem como dados restritos a intervalos específicos do conjunto dos reais, como o intervalo $(0, 1)$.

O componente sistemático descreve a forma como as variáveis explanatórias entram no modelo por meio de uma combinação linear de seus efeitos, definida por

$$\eta_i = \sum_{r=1}^{p} x_{ir}\beta_r = \mathbf{x}_i^T\boldsymbol{\beta},$$

ou, de forma matricial,

$$\boldsymbol{\eta} = X\boldsymbol{\beta},$$

em que $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ é a matriz do modelo, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ é o vetor de parâmetros e $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$ representa o preditor linear. Quando algum termo da estrutura linear possui valor conhecido, este é denominado *offset*.

Por fim, a função de ligação estabelece a relação entre o componente aleatório e o componente sistemático, conectando a média da variável resposta ao preditor linear,

isto é,

$$\eta_i = g(\mu_i),$$

em que $g(\cdot)$ é uma função monótona e diferenciável. Dessa forma, observa-se que, na especificação de um MLG, os parâmetros θ_i associados à família exponencial não são de interesse direto, uma vez que há um para cada observação. O interesse central recai sobre um conjunto reduzido de parâmetros $\beta_1, \ldots, \beta_p$, cuja combinação linear está relacionada a uma função do valor esperado de Y_i.

3.6.1.2 Estimação dos Parâmetros nos Modelos Lineares Generalizados

Na regressão linear, em que utilizamos um modelo tal como $Y_i = \alpha + \beta x_i + \varepsilon_i$, a técnica de mínimos quadrados era uma opção para calcular estimativas de α e β. Entretanto, nos modelos lineares generalizados não há mais uma relação direta entre a média da variável resposta e o preditor linear Y_i e $\alpha + \beta x_i$, sendo essa relação estabelecida por meio de uma função de ligação. Portanto, o método de estimação mais comumente utilizado é a máxima verossimilhança, implementada por meio de procedimentos iterativos (Casella e Berger, 2022).

Os parâmetros dos Modelos Lineares Generalizados (MLGs), em particular os coeficientes associados ao preditor linear, são estimados pelo método da Máxima Verossimilhança. Esse método consiste em determinar os valores dos coeficientes que maximizam a função de verossimilhança associada à distribuição da variável resposta, o que conduz à resolução de um sistema de equações não lineares, geralmente sem solução analítica fechada (Cordeiro and Demétrio, 2007).

Sob condições usuais de regularidade, os estimadores de máxima verossimilhança dos coeficientes dos MLGs são consistentes e assintoticamente normais, o que fornece a base teórica para a inferência estatística por meio de intervalos de confiança e testes de hipóteses. Além disso, os coeficientes estimados correspondem diretamente aos parâmetros do preditor linear, enquanto os parâmetros do modelo englobam todas as quantidades desconhecidas que necessitam ser estimadas (Cordeiro and Demétrio, 2007).

Segundo Cordeiro and Demétrio (2007), para estimar os parâmetros lineares $\beta_1, \ldots, \beta_p$ do modelo, procede-se inicialmente à construção do vetor escore, o qual sintetiza a informação contida nos dados acerca dos parâmetros desconhecidos. Para isso tomamos o logaritmo da função de verossimilhança, considerado como função

apenas do vetor de parâmetros $\boldsymbol{\beta}$ e assumindo o parâmetro de dispersão ϕ conhecido, dado o vetor de observações $\mathbf{y}$, definido por $\ell(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}; \mathbf{y})$. Utilizando a expressão da família exponencial, tem-se que

$$\ell(\boldsymbol{\beta}) = \frac{1}{\phi} \sum_{i=1}^{n} [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i, \phi), \tag{3.1}$$

em que $\theta_i = q(\mu_i)$ representa o parâmetro canônico, $\mu_i = g^{-1}(\eta_i)$ é a média da variável resposta e

$$\eta_i = \sum_{r=1}^{p} x_{ir} \beta_r$$

denota o preditor linear do modelo. A partir da expressão do logaritmo da função de verossimilhança, é possível obter, por meio da regra da cadeia, o vetor escore $\mathbf{U}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, de dimensão p, cujos elementos correspondem às derivadas parciais de primeira ordem da log-verossimilhança em relação aos parâmetros do modelo.

Por conseguinte, a estimação por máxima verossimilhança é realizada resolvendo o sistema de equações obtido ao igualar o vetor escore a zero, sendo, em geral, necessário o uso de métodos numéricos iterativos devido à natureza não linear dessas equações.

Dessa forma, torna-se necessário o uso de algoritmos iterativos para a obtenção dos estimadores de máxima verossimilhança. Entre os métodos mais utilizados destacam-se o algoritmo de Newton–Raphson e o método do Escore de Fisher, ambos baseados em aproximações de Taylor da função escore que se baseiam na resolução da equação escore $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$.

No algoritmo de Newton–Raphson, a atualização do vetor de parâmetros na iteração $(t + 1)$ é dada por

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \mathbf{H}^{-1}(\boldsymbol{\beta}^{(t)}) \mathbf{U}(\boldsymbol{\beta}^{(t)}),$$

em que $\mathbf{H}(\boldsymbol{\beta}) = \partial^2 \ell(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ denota a matriz Hessiana da log-verossimilhança, também denominada informação observada.

O método do Escore de Fisher consiste em substituir a matriz Hessiana pela matriz de informação de Fisher,

$$\mathcal{I}(\boldsymbol{\beta}) = \mathbb{E}[-\mathbf{H}(\boldsymbol{\beta})],$$

resultando na atualização

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \mathcal{I}^{-1}(\boldsymbol{\beta}^{(t)}) \mathbf{U}(\boldsymbol{\beta}^{(t)}).$$

Em Modelos Lineares Generalizados com função de ligação canônica, a informação observada coincide com a informação esperada, de modo que os algoritmos de Newton–Raphson e Escore de Fisher tornam-se equivalentes, levando ao procedimento conhecido como Regressão Ponderada Iterativa por Mínimos Quadrados.

3.6.1.3 Interpretação dos Coeficientes e Identificação de Fatores de Risco

A identificação de fatores de risco está diretamente associada à interpretação dos coeficientes estimados do preditor linear. Cada coeficiente β_j representa o efeito marginal da covariável correspondente sobre a função de ligação da média da variável resposta, mantendo-se as demais variáveis constantes. Dessa forma, a interpretação dos coeficientes depende da função de ligação adotada, uma vez que os parâmetros atuam linearmente sobre o preditor linear e não diretamente sobre a média da variável resposta (Cordeiro and Demétrio, 2007).

3.6.1.4 Avaliação do Ajuste e Diagnóstico do Modelo

Esta dissertação possui como principal interesse a interpretação dos efeitos das covariáveis, embora os MLGs possam ser utilizados para predição. Nesse contexto, o adequado ajuste e diagnóstico do modelo é essencial, uma vez que a especificação da distribuição da variável resposta, da função de ligação e das covariáveis incluídas influência diretamente os coeficientes estimados e, consequentemente, a interpretação dos efeitos associados. Dessa forma, devemos verificar se o modelo descreve adequadamente os dados com as suposições básicas atendidas, sendo elas medida de ajuste global e análise de resíduos.

3.6.1.5 Medidas descritivas

Para uma avaliação inicial de modelos de regressão, utilizaremos a correlação, pois permite verificar o grau de associação linear entre as variáveis envolvidas, descrita a seguir conforme apresentado em (Morettin and Bussab, 2010):

Dados n pares de valores $(x_1, y_1), \ldots, (x_n, y_n)$, chamaremos de covariância entre as duas variáveis X e Y:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} \tag{3.5}$$

Ou seja, a média dos produtos dos valores centrados das variáveis. Essa medida permite avaliar como duas variáveis variam conjuntamente, indicando a direção da associação linear entre elas.

Com essa definição, o coeficiente de correlação pode ser escrito como:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{dp(X)\, dp(Y)} \tag{3.6}$$

O coeficiente de correlação corresponde a uma forma padronizada da covariância, assumindo valores no intervalo $[-1, 1]$, o que permite avaliar simultaneamente a intensidade e a direção da associação linear entre as variáveis.

No contexto da modelagem estatística, essas medidas constituem uma etapa preliminar relevante na avaliação de modelos de regressão, pois auxiliam na identificação de possíveis relações entre variáveis explicativas e a variável resposta. Dessa forma, contribuem para a compreensão inicial da estrutura dos dados e para a seleção de variáveis candidatas à composição dos modelos.

3.6.1.6 Ajuste do Modelo

Após a estimação dos parâmetros do modelo linear generalizado, procedeu-se à avaliação da qualidade do ajuste por meio de diferentes medidas estatísticas e análises gráficas. Dentre elas, as medidas de ajuste descrevem a variabilidade dos dados através do modelo ajustado. Para isso, utilizamos o deviance, que compara a verossimilhança do modelo ajustado com a do modelo saturado que se trata do modelo que possui um parâmetro por observação. A maximização da verossimilhança é equivalente à minimização da deviance, de modo que menores valores indicam melhor ajuste do modelo aos dados.

Outra medida considerada foi a **estatística de Pearson** (X^2), que avalia a qualidade do ajuste a partir da soma dos quadrados dos resíduos de Pearson. Essa estatística é amplamente utilizada para verificar a adequação da função de variância assumida pelo modelo.

Adicionalmente, realizou-se a **análise dos resíduos**, utilizando resíduos padroni-

zados para identificar possíveis observações influentes, presença de padrões sistemáticos, não linearidade ou heterogeneidade da variância, os quais podem indicar inadequação do modelo.

O **Teste da Razão de Verossimilhança** foi empregado para comparação entre modelos aninhados, por meio da diferença entre suas deviances. Esse teste permite avaliar se a inclusão de novas variáveis melhora significativamente o ajuste do modelo.

Também foram utilizados os **Critérios de Informação de Akaike (AIC)** e o **Critério de Informação Bayesiano (BIC)**. O AIC busca selecionar o modelo que melhor equilibra qualidade de ajuste e parcimônia, penalizando o excesso de parâmetros.

Procedimentos automáticos de seleção, como o **Stepwise**, incluindo os métodos *Forward Selection* (seleção progressiva) e *Backward Elimination* (eliminação regressiva), foram utilizados como ferramentas exploratórias auxiliares na identificação do modelo mais adequado.

3.6.1.7 Avaliação da Dispersão dos Dados

A verificação da adequação da função de variância assumida pelo modelo foi realizada por meio da análise do parâmetro de dispersão (ϕ). Em modelos da família Gamma, esse parâmetro permite avaliar a variabilidade residual dos dados em relação ao valor esperado pelo modelo.

Valores elevados do parâmetro de dispersão podem indicar variabilidade não explicada ou inadequação da estrutura do modelo. Nesses casos, alternativas métodológicas podem ser adotadas, tais como: quase-verossimilhança, correção dos erros padrão e estimador robusto do tipo “Sanduíche”.

3.7 Estatística Espacial

O estudo dos padrões de distribuição geográfica das doenças e suas relações com fatores socioambientais se dá pela Epidemiologia Geográfica. Assim, estudos epidemiológicos do tipo ecológico mostram-se conceitualmente adequados para analisar as diferenças na distribuição do risco de adoecimento entre distintos grupos populacionais (Brasil et al., 2007).

Os mapas possibilitam, além da verificação da distribuição das doenças, as locali-

dades com alto e baixo risco relativo. A partir disso, a área na estatística que analisa a localização espacial de eventos denomina-se estatística espacial. Através dela, torna-se viável modelar a ocorrência dos eventos, a estrutura de distribuição ou identificar padrões. Dentre os tipos de análises espaciais destacamos mapeamento de doenças, os estudos ecológicos, a identificação de aglomerados espaciais (cluster) e o monitoramento de problemas ambientais. O primeiro método descreve a distribuição espacial focando na variação geográfica da ocorrência, o segundo trata-se do estudo da relação entre incidência de doenças e potenciais fatores etiológicos, o terceiro avalia o aglomerado de eventos não casuais e o quarto a distribuição espacial para acompanhar potenciais fontes ambientais de problemas de saúde. Para o estudo em questão, a análise de aglomerados espaciais mostra-se mais adequada, uma vez que sua finalidade consiste em verificar os fatores que causam aglomerados, tais como agentes infecciosos, contaminação ambiental localizada ou efeitos colaterais de tratamentos (Brasil et al., 2007).

Dependendo do tipo de dado analisado, as técnicas estatísticas diferem quanto à sua adequação. Assim, temos dois tipos de dados: atributos pontuais e atributos agregados por áreas. O primeiro é o mais simples e trata-se da localização pontual, por exemplo através de coordenadas partindo da premissa acerca da localidade, independente de outras covariáveis serem medidas. O segundo trabalha com áreas delimitadas, usualmente divisões político-administrativas, tais como macrorregiões, sendo mais utilizado para contagem de casos ou óbitos e seus respectivos riscos relativos.

Em seguida, ao se trabalhar com registros de saúde para avaliar riscos, deve-se estimar a probabilidade de um evento ocorrer, ponderada pela distribuição da população. Assim, os principais métodos são: área de influência (buffer), estimativa de kernel dual, métodos bayesianos para suavização de riscos e por fim a estatística espacial scan (SaTScan).

O primeiro método descreve a técnica aplicada a partir de uma fonte (que pode ser um ponto, linha ou polígono), para realizar contagens e cálculos em uma área delimitada e cria uma zona em torno de uma fonte suspeita de risco para calcular a taxa de incidência dentro e fora desse limite. O segundo baseia-se em dados de pontos e é uma técnica de interpolação que calcula a razão entre duas superfícies de densidade: a densidade de casos (eventos) e a densidade da população, e o resultado é uma superfície contínua que representa a taxa de incidência ou o risco relativo, permitindo identificar "áreas quentes"que não sejam apenas reflexo da alta concentração populacional. Por

seguinte, o terceiro é aplicado a dados de áreas e têm como objetivo principal corrigir a instabilidade das taxas em pequenas áreas ou com eventos raros, onde a flutuação aleatória pode distorcer a realidade e o Estimador Bayesiano Empírico realiza uma média ponderada entre a taxa observada na área e uma taxa média (global ou da vizinhança); quanto menor a população da área, maior será o peso da vizinhança na correção, estabilizando o indicador para refletir o risco real. Por fim, o último método utiliza dados agregados por áreas, através de seus centróides e utiliza uma janela geográfica circular móvel que percorre o território, alterando seu raio e posição. Ela faz uma varredura (scan) para localizar áreas onde o número de casos observados é significativamente maior do que o esperado em relação à população sob risco.

Nesse contexto, ao se trabalhar com dados agregados por áreas e análise do risco, o método de estatística espacial scan é mais assertivo. Assim, necessitamos da modelagem da ocorrência de eventos discretos, como óbitos, ponderada pela população sob risco.

Para esse tipo de dados, a distribuição de Poisson mostra-se adequada, uma vez que descreve a contagem de eventos raros em uma população definida e em um intervalo fixo de espaço ou tempo.

Além disso, um procedimento comum é utilizar o risco relativo para mapear as áreas (Souza Junior, G. L., 2005). Assim, assume-se que o número de casos observados c_z na área z segue uma distribuição de Poisson, dada por

$$c_z \sim \text{Poisson}(e_z \xi_z), \tag{3.7}$$

em que e_z representa o número esperado de casos na área z e ξ_z denota o risco relativo associado a essa área, com $z = 1, \ldots, Z$.

O número esperado de casos e_z é obtido a partir da padronização indireta, sendo calculado como

$$e_z = N_z \times \frac{\sum_{z=1}^{Z} c_z}{\sum_{z=1}^{Z} N_z}, \tag{3.8}$$

em que N_z corresponde à população sob risco na área z e $\sum_{z=1}^{Z} c_z / \sum_{z=1}^{Z} N_z$ representa a taxa média global do evento na região de estudo.

Sob essa formulação, o risco relativo ξ_z constitui um parâmetro latente que expressa a razão entre o risco da área z e o risco médio da população de referência. O

estimador clássico do risco relativo é dado por

$$\hat{\xi}_z = \frac{c_z}{e_z}. \tag{3.9}$$

Entretanto, quando as áreas de investigação apresentam pequenas populações ou quando o evento analisado é raro, as estimativas diretas do risco relativo tornam-se instáveis, uma vez que estão fortemente sujeitas à variabilidade aleatória. Além disso, tais estimativas são calculadas de forma independente para cada área, desconsiderando possíveis estruturas espaciais e dependências entre regiões vizinhas.

Diante dessas limitações, torna-se necessário empregar métodos capazes não apenas de estimar riscos, mas de identificar padrões espaciais não aleatórios. Nesse sentido, a estatística de varredura espacial proposta por Kulldorff apresenta-se como uma abordagem robusta para a detecção de conglomerados espaciais de áreas com riscos significativamente elevados ou reduzidos, sendo implementada por meio do software *SaTScan*.

3.7.1 Método de varredura circular de kulldorf

A exemplo de Souza Júnior (2005), descreveremos o método de varredura circular de Kulldorff. Contudo, serão necessárias algumas definições iniciais. Seja S uma região dividida em m áreas. Para cada uma dessas áreas, é necessário conhecer as coordenadas geográficas, o número de indivíduos e o número de casos. Definimos uma zona z como uma coleção de áreas, Z como o conjunto de todas as possíveis zonas circulares criadas por este método, π_z como a probabilidade de um indivíduo pertencente à zona z vir a ser um caso e π_0 como a probabilidade de um indivíduo fora da zona z vir a ser um caso.

Sob a hipótese nula H_0, o risco de ocorrência é constante em toda a região, de forma que o número de casos é proporcional à população sob risco, isto é,

$$H_0 : \lambda_z = \lambda_0.$$

Sob a hipótese alternativa $H_1(z)$, admite-se a existência de uma zona z com risco elevado em relação ao restante da região, ou seja,

$$H_1(z) : \lambda_z > \lambda_0.$$

Seja N o número total de indivíduos sob risco na região S, C o número total de casos na região S, n_z o número de indivíduos sob risco na zona z e c_z o valor observado da variável aleatória C_z, que representa o número de casos na zona z. Admitindo que o modelo de Poisson é apropriado para o número de casos, a função de verossimilhança para a zona z será dada por:

De acordo com Kulldorff, M. (1997), a função de verossimilhança sob o modelo de Poisson, para uma zona fixa z, pode ser expressa como

$$L(z) = \left(\frac{C_z}{E_z}\right)^{C_z} \left(\frac{C - C_z}{E - E_z}\right)^{C - C_z} \mathbb{I}\left(\frac{C_z}{E_z} > \frac{C - C_z}{E - E_z}\right),$$

em que $\mathbb{I}(\cdot)$ é a função indicadora, garantindo que apenas zonas com risco maior dentro da janela sejam consideradas candidatas a conglomerados.

O número esperado de casos na zona z é obtido por

$$E_z = E \times \frac{P_z}{P},$$

em que P_z representa a população sob risco na zona z e P representa a população total da região em estudo.

A estatística utilizada para testar a presença de conglomerados é a razão de verossimilhança, definida por

$$\Lambda(z) = \frac{L(z)}{L_0},$$

em que L_0 corresponde à verossimilhança sob a hipótese nula, a qual depende apenas do número total de casos observados e não de sua distribuição espacial.

A zona mais provável de constituir um conglomerado espacial é aquela que maximiza a razão de verossimilhança,

$$z^* = \arg\max_{z \in Z} \Lambda(z).$$

Eliminando termos constantes e considerando apenas zonas para as quais $C_z > E_z$, a razão de verossimilhança pode ser escrita de forma equivalente como

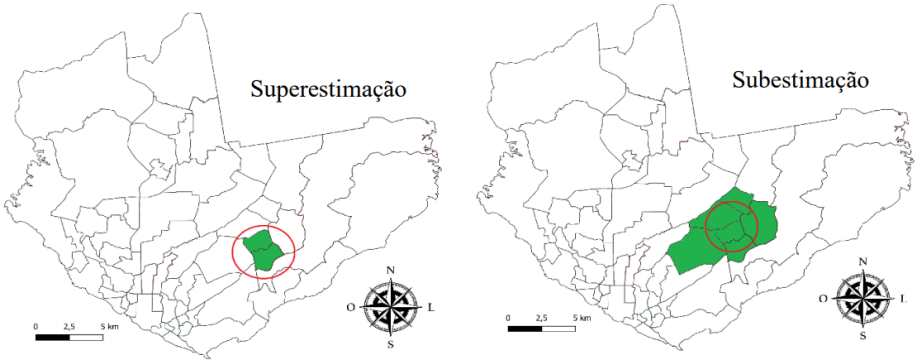
$$\Lambda(z) = \left(\frac{C_z}{E_z}\right)^{C_z} \left(\frac{C - C_z}{C - E_z}\right)^{C - C_z}, \tag{4.13}$$

sendo esta a estatística maximizada no método de varredura espacial para identificar o conglomerado mais provável.

Como a distribuição exata da estatística de teste sob a hipótese nula não é conhecida, a significância estatística é obtida por meio de simulações de Monte Carlo. Nessas simulações, os casos são redistribuídos aleatoriamente na região segundo o modelo de Poisson, permitindo a comparação entre o valor observado da estatística e a distribuição empírica dessa estatística sob H_0.

No desenvolvimento original da estatística de varredura espacial, Kulldorff, M. (1997) emprega uma janela de forma circular que se desloca e altera seu raio para avaliar todas as possíveis regiões candidatas a conglomerados na área de estudo. Essa escolha é adotada por sua simplicidade conceitual e por ser a forma mais compacta. Entretanto, segundo Ronald e Murray (2001), existe a possibilidade de identificarmos um cluster maior ou menor que seu tamanho real, o primeiro indicando um problema de superestimação e o segundo de subestimação, conforme a Figura 2 a seguir.

Figura 2: Superestimação e Subestimação do conglomerado.



Fonte: Elaboração própria

A superestimação ocorre quando a análise identifica erroneamente um aglomerado onde não existe ou exagera sua importância. Os principais motivos são a densidade populacional, a variância em pequenas populações, testes múltiplos e áreas grandes. Por sua vez, a subestimação ocorre quando o método falha em detectar aglomerados reais ou minimiza sua relevância estatística. Os principais motivos são: limitações geográficas, ajustes estatísticos, tamanho ou incerteza nos dados.

Na superestimação, o primeiro problema ocorre pela alta concentração de população em uma área com densidade não homogênea. Na sequência, o segundo problema acontece quando selecionamos aglomerados apenas com a maior taxa de incidência, pois zonas com populações muito pequenas apresentam alta variabilidade. Além disso,

em terceiro lugar, temos que, ao testar milhares de círculos sobrepostos individualmente, a probabilidade de encontrar um resultado "significativo"puramente ao acaso aumenta quando não se utilizam métodos com uma estatística de teste única para todo o mapa. Por fim, temos que círculos que abrangem quase toda a região de estudo podem ser identificados como aglomerados de forma inadequada.

Na subestimação o primeiro problema ocorre com métodos que utilizam fronteiras administrativas como municípios, pois falham em detectar aglomerados cujos limites não participam dessa delimitação. Por seguinte, quando os testes são altamente correlacionados, utilizamos o procedimento de correção de Bonferroni para resolver o problema de testes múltiplos. Além disso, utilizar raios ou tamanho populacional fixo para os círculos de busca pode ter baixo poder estatístico, pois o aglomerado pode ser maior ou menor que o definido. Por fim, a incerteza ocorre quando a localização exata dos casos é desconhecida e eles são distribuídos proporcionalmente à população em uma área. Dessa forma, o método da razão de verossimilhança busca equilibrar esses dois extremos, utilizando simulações de Monte Carlo para garantir que a significância não seja nem inflada (superestimada), nem mascarada (subestimada).

O método funciona gerando uma multidão de zonas circulares que servem como “quadrados” de amostragem (quadrats). Em seguida, é estabelecida uma localidade sobre a região de estudo pelo usuário e em cada ponto são centrados círculos. O raio de cada círculo varia continuamente de zero até um limite superior definido. Isso permite que o teste detecte clusters de qualquer tamanho e em qualquer local da região estudada. Para evitar a detecção de “aglomerados negativos” (áreas vastas que parecem clusters apenas porque o restante do mapa tem poucos casos), recomenda-se que o raio do círculo não exceda 50% da população total em risco (Kulldorff, M. and Nagarwalla, N., 1995). Desse modo, o modelo escolhido, neste caso Poisson, no qual o quantitativo de casos da localidade e a população sob risco também são indicados pelo usuário. A partir das informações imputadas sobre população, casos e coordenadas geográficas, o software procura aglomerados que ocorrem em uma ou mais variáveis de casos imputadas no programa em questão. Por fim, são feitas varreduras para conglomerados com baixas taxas que representa o conjunto de dados com valores inferiores ao número de casos esperados e de altas taxas, que representam o conjunto de dados com valores superiores ao número de casos esperados, sob a hipótese nula da inexistência de clusters.

Por fim, a escolha do modelo Poisson deu-se pela natureza dos dados analisados,

que consistem em contagens de óbitos associadas a uma população em risco conhecida, situação em que a distribuição de Poisson é classicamente empregada. Nesse contexto, assume-se que o número esperado de eventos em cada área é proporcional ao tamanho da população, permitindo a comparação entre valores observados e esperados sob a hipótese nula de risco homogêneo no espaço. A significância estatística dos conglomerados identificados é avaliada por simulações de Monte Carlo, o que confere robustez ao procedimento mesmo diante de desvios moderados da suposição de equidispersão, embora tal limitação deva ser considerada na interpretação dos resultados.

Embora modelos alternativos, como a distribuição binomial negativa, sejam frequentemente utilizados em análises de regressão para lidar com superdispersão, essa abordagem não é implementada na estatística de varredura espacial clássica, o que limita a modelagem de superdispersão nesse contexto. Além disso, o objetivo principal do método de Kulldorff não é a modelagem explicativa completa da variabilidade dos dados, mas sim a detecção de áreas com risco significativamente elevado em relação ao esperado. Dessa forma, o uso do modelo de Poisson permanece como a escolha metodologicamente apropriada e amplamente consolidada na literatura para a detecção de conglomerados espaciais em estudos de mortalidade.

3.8 Softwares

3.8.1 SaTScan

O SaTScan é um aplicativo gratuito desenvolvido por Martin Kulldorff, que baseado no teste da Razão de Verossimilhança, implementa o ajuste dos modelos Binomial ou Poisson usando o método de varredura circular para análise de dados espaciais, temporais e espaço-temporais (Kulldorff, M. – SaTScan User Guide, 2018). Este software está disponível gratuitamente no site <http://www.satscan.org/>.

O software SaTScan utiliza um método baseado em estatísticas de varredura espacial (*spatial scan statistics*) para a detecção de aglomerados espaciais. Inicialmente, cada centróide da área em estudo é considerado como centro de uma janela de varredura, sendo calculadas as distâncias para os demais centróides. A partir desse ponto, janelas circulares são construídas ao redor de cada centróide, com raios progressivamente aumentados, de modo que diferentes configurações espaciais sejam avaliadas.

À medida que novos centróides passam a integrar a janela de varredura, os valores

correspondentes ao número de casos observados e à população sob risco são atualizados. Para cada janela considerada, calcula-se a estatística de teste, baseada na razão de verossimilhança, segundo o modelo probabilístico adotado, sendo os mais comuns os modelos de Poisson e de Bernoulli. A janela que maximiza a razão de verossimilhança é identificada como o aglomerado mais provável.

A significância estatística dos aglomerados detectados é avaliada por meio de simulações de Monte Carlo, nas quais são gerados conjuntos de dados sob a hipótese nula de distribuição aleatória dos casos. O valor-p é estimado pela proporção de simulações cuja estatística excede a estatística observada em relação à distribuição empírica das estatísticas simuladas (Kulldorff, M., 1997).

3.8.2 RStudio

O Rstudio é um ambiente de desenvolvimento integrado para R, possuindo um console, ferramentas para plotagem, histórico, depuração e gerenciamento de espaço de trabalho, desenvolvido por (RStudio Team, 2026). Este software está disponível gratuitamente no site <https://www.rstudio.com/> e, dentre suas inúmeras funcionalidades, gerar metricas de modelos. Como indicado inicialmente em (Olinda, R. A., 2020), o programa necessitará da importação dos dados, para assim os modelos serem ajustados no software R, por meio da implementação de Modelos Lineares Generalizados (MLG), que estendem o modelo linear clássico ao permitir distribuições não normais para a variável resposta. Esse ambiente computacional foi utilizado tanto para a estimação dos parâmetros, via máxima verossimilhança, quanto para a análise de resíduos, avaliação do ajuste e verificação das suposições do modelo.

3.8.3 TabWin

O TabWin, software desenvolvido pelo DATASUS, foi utilizado para tabular, organizar e visualizar espacialmente dados de saúde. A ferramenta permite integrar bases epidemiológicas com arquivos cartográficos, possibilitando a construção de tabelas, mapas temáticos e indicadores agregados por unidades geográficas. No estudo, foi empregada para consolidar informações de sistemas de saúde e gerar mapas por município, auxiliando na análise exploratória da distribuição espacial dos dados e garantindo padronização territorial, conforme divisões administrativas oficiais, uma vez que possibilita a agregação de eventos e populações segundo diferentes níveis territo-

riais. Dessa forma, o uso do TabWin contribuiu para a organização dos dados e para a compreensão da distribuição espacial do fenômeno em estudo, fornecendo subsídios iniciais para análises espaciais mais aprofundadas realizadas em ambientes estatísticos específicos.

4 RESULTADOS

4.1 Análise descritiva

Segundo o Sistema de Informação sobre Mortalidade (Brasil, Ministério da Saúde, 2025), no período de 2010 a 2023, as neoplasias representaram a segunda maior causa de óbitos no Brasil. Observou-se que três capítulos da CID-10 — Doenças do aparelho circulatório (IX), Neoplasias (II) e Causas externas de morbidade e mortalidade (XX) — concentram, em conjunto, 53% de todos os óbitos, considerando as 21 causas distintas analisadas, conforme apresentado na Tabela A1 (Apêndice A). Enquanto as demais 18 causas apresentam participação individualmente menor. Esse comportamento, de manter as neoplasias na segunda posição, corrobora os estudos de (Teixeira, 2021);(Bermudi, 2020) ou (Oliveira et al., 2020), apesar de considerarem períodos diferentes, indicando estabilidade na posição dessa causa ao longo do tempo.

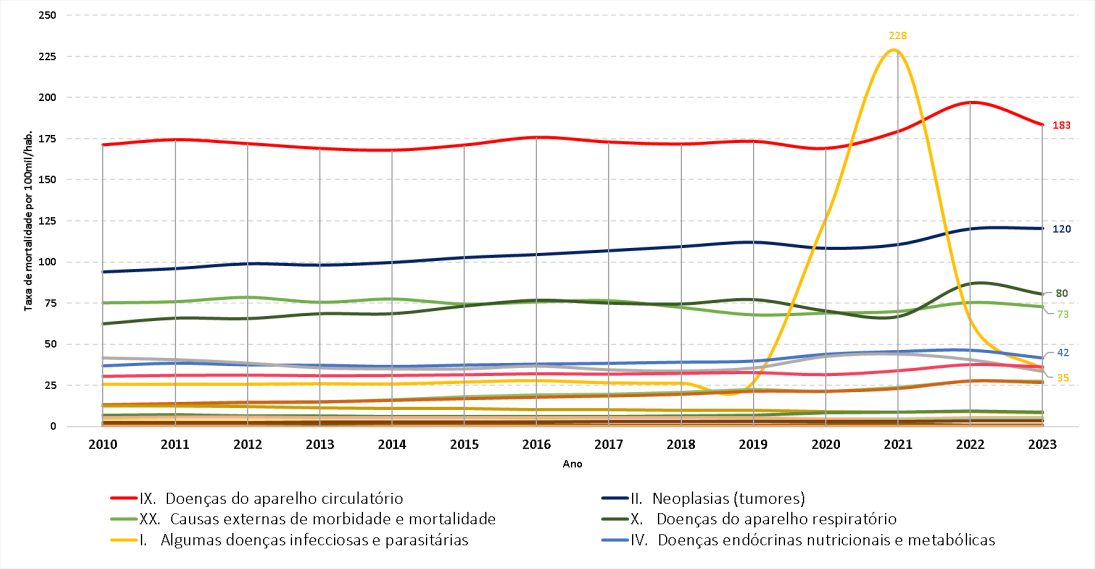
4.1.1 Taxas brutas

Tabela 2: Taxa bruta de mortalidade por regiões e capítulos CID-10 – Brasil, 2010 a 2023.

Capítulo CID-10	**Regiões**					
	Brasil	Sudeste	Sul	Nordeste	Centro-Oeste	Norte
IX. Doenças do aparelho circulatório	175	195	188	167	146	105
II. Neoplasias (tumores)	106	120	141	85	90	62
XX. Causas externas de morbidade e mortalidade	74	64	73	86	83	80
3 maiores causas	**355**	**379**	**403**	**337**	**319**	**247**
Outras causas	306	340	307	289	257	234
Total	**661**	**719**	**710**	**626**	**576**	**481**

Fonte: Adpatado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População e Censo Demográfico (2010,2022);
Nota: Taxa = $\frac{\sum_{i=1}^{14} \text{Óbitos}_i}{\sum_{i=1}^{14} \text{Pop}_i}$, onde i representa os anos.

Figura 3: Taxa bruta de mortalidade por capítulos CID-10 – Brasil, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Na Figura 3, podemos verificar que, no período de 2010 a 2023, observou-se uma taxa estável e crescente de mortalidade por Doenças do aparelho circulatório (IX) e Neoplasias (II). Apesar do leve decréscimo em 2020, possivelmente em razão da pandemia da COVID-19, dado que o Brasil era o terceiro país em número de óbitos, citado em (Brasil, Ministério da Saúde, 2022), manteve-se como segunda causa mais frequente e crescente. Além disso, a partir de 2019, observa-se uma flutuação com aumentos ou reduções em algumas causas, entretanto as Neoplasias após o déscrimo de 2020 tornam a crescer e mantém-se estável em 2022 e 2023. Assim, a taxa de mortalidade por neoplasias aumentou de 94 para 120 óbitos/100 mil habitantes. Correspondendo a um crescimento de 27,7% ao longo dos 14 anos, como pode ser observado na tabela a seguir.

Tabela 3: Métricas das taxas brutas de mortalidade por capítulo da CID-10 – Brasil, 2010-2023.

CID-10	Taxa óbitos 2010	Taxa óbitos 2023	Variação % Total	Variação Média Anual	Crescimento Anual Composto
IX. Doenças do aparelho circulatório	171	183	+7,02%	+0,54%	+0,52%
II. Neoplasias (tumores)	94	120	+27,66%	+2,13%	+1,90%
XX. Causas externas de morbidade e mortalidade	75	73	-2,67%	-0,21%	-0,21%
X. Doenças do aparelho respiratório	62	80	+29,03%	+2,23%	+1,98%
I. Algumas doenças infecciosas e parasitárias	26	35	+34,62%	+2,66%	+2,31%

Nota 1: Média anual = Variação % Total / 13 anos;

Nota 2: Nota 2: Variação Média Anual = $\left(\frac{\text{Valor Final}}{\text{Valor Inicial}}\right)^{\frac{1}{\text{Nº de Anos}}} - 1 \times 100$.

Tabela 4: Taxa bruta de mortalidade por capítulos CID-10 – Região Norte, 2010 a 2023.

Capítulo CID-10	**Região Norte**						
	AP	AM	RR	AC	PA	RO	TO
IX. Doenças do aparelho circulatório	80	82	91	98	110	117	151
XX. Causas externas de morbidade e mortalidade	80	72	97	69	80	85	92
II. Neoplasias (tumores)	55	65	61	62	58	72	71
3 maiores causas	**215**	**219**	**248**	**230**	**248**	**274**	**313**
Outras causas	209	243	214	241	237	236	214
Total	**424**	**461**	**463**	**471**	**484**	**510**	**527**

Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População e Censo Demográfico (2010,2022);
Nota: Taxa = $\frac{\sum_{i=1}^{14} \text{Óbitos}_i}{\sum_{i=1}^{14} \text{Pop}_i}$, onde i representa os anos.

Tabela 5: Taxa bruta de mortalidade por capítulos CID-10 – Amazonas, 2010 a 2023.

Capítulo CID-10	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
IX. Doenças do aparelho circulatório	68	72	73	71	77	79	79	77	78	82	85	94	102	102
XX. Causas externas de morbidade e mortalidade	65	71	75	68	70	76	70	74	72	73	65	77	82	73
II. Neoplasias (tumores)	54	60	65	62	63	63	65	65	69	66	63	63	73	72
I. Algumas doenças infecciosas e parasitárias	24	24	23	25	25	25	28	25	24	24	166	242	49	32
XVIII. Sintomas, sinais e achados anormais	58	56	59	53	49	51	52	53	53	56	66	51	42	37
5 maiores causas	**269**	**282**	**294**	**278**	**285**	**294**	**294**	**294**	**296**	**300**	**444**	**527**	**347**	**317**
Outras causas	113	120	119	119	125	129	126	131	138	142	145	154	164	163
Total	**382**	**402**	**413**	**397**	**410**	**423**	**420**	**425**	**434**	**442**	**589**	**681**	**511**	**481**

Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População e Censo Demográfico (2010,2022);
Nota: Taxa = $\frac{\sum_{i=1}^{14} \text{Óbitos}_i}{\sum_{i=1}^{14} \text{Pop}_i}$, onde i representa os anos.

Ao analisar as regiões brasileiras, observa-se que a região Norte registra a menor taxa de mortalidade geral, mas mantém as neoplasias entre as principais causas de óbitos. No contexto regional e estadual, a taxa de óbitos por neoplasias nos estados da Região Norte e no estado do Amazonas ocupa a terceira posição entre as causas de morte, atrás apenas das Doenças do Aparelho Circulatório e das Causas externas de morbidade e mortalidade. Além disso, a taxa de mortalidade por câncer aumentou de 54 para 72 óbitos/100 mil habitantes, correspondendo a um crescimento de 33,3%, indicando uma tendência de expansão desse grupo de causas no estado.

Tabela 6: Métricas das taxas brutas de mortalidade por capítulo da CID-10 – Amazonas, 2010-2023.

CID-10	Taxa óbitos 2010	Taxa óbitos 2023	Variação % Total	Variação Média Anual	Crescimento Anual Composto
IX. Doenças do aparelho circulatório	68	102	+50,00%	+3,85%	+3,17%
XX. Causas externas de morbidade e mortalidade	65	73	+12,31%	+0,95%	+0,90%
II. Neoplasias (tumores)	54	72	+33,33%	+2,56%	+2,24%
X. Doenças do aparelho respiratório	24	32	+33,33%	+2,56%	+2,24%
I. Algumas doenças infecciosas e parasitárias	58	37	-36,21%	-2,79%	-3,40%

Nota 1: Média anual = Variação % Total / 13 anos;

Nota 2: Nota 2: Variação Média Anual = $\left(\frac{\text{Valor Final}}{\text{Valor Inicial}}\right)^{\frac{1}{\text{Nº de Anos}}} - 1 \times 100$.

Tabela 7: Número de óbitos por sexo – Amazonas, 2010 a 2023.

Causa (CID10 3C)	M		F		Total	
	n	%	n	%	n	%
C16 Neopl malig do estômago	2.818	7,90%	1.363	3,82%	4.181	11,72%
C34 Neopl malig dos brônquios e dos pulmões	2.347	6,58%	1.616	4,53%	3.963	11,11%
C53 Neopl malig do colo do útero	0	0,00%	3.747	10,50%	3.747	10,50%
C50 Neopl malig da mama	27	0,08%	2.335	6,55%	2.362	6,62%
C61 Neopl malig da próstata	2.354	6,60%	0	0,00%	2.354	6,60%
C22 Neopl malig fígado vias biliares intra-hepáticas	1.132	3,17%	708	1,98%	1.840	5,16%
6 maiores causas	**8.678**	**24,33%**	**9.769**	**27,39%**	**18.447**	**51,72%**
Outras causas	9.388	26,32%	7.835	21,97%	17.223	48,28%
Total	**18.066**	**50,65%**	**17.604**	**49,35%**	**35.670**	**100,00%**

Fonte: tabwin

Ao analisarmos os tumores malignos no estado do Amazonas, o câncer do colo do útero é a 3ª maior causa individual no ranking, seguido pelo câncer de mama como 4ª. Assim, dentre as seis principais neoplasias, os agravos de mama (C50) e do colo do útero (C53) representam, em conjunto, aproximadamente um quarto das principais neoplasias no estado, evidenciando sua relevância epidemiológica e sanitária.

Tabela 8: Taxas brutas de mortalidade por câncer de mama e do colo do útero – Amazonas, 2010 a 2023.

CID	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Média
C50	9	11	13	14	12	13	13	13	15	15	17	14	18	19	14
C53	22	22	24	26	26	24	22	24	23	23	22	20	22	21	23
Total	31	32	38	40	38	38	35	38	38	37	38	34	40	40	37

A taxa bruta estadual de mortalidade foi calculada a partir do total de óbitos registrados no estado dividido pela população feminina total do Amazonas no respectivo

ano, representando o risco médio de morte por câncer de mama e câncer do colo do útero na população estadual.

4.1.2 Taxas padronizadas

A análise das taxas padronizadas permite comparações entre populações com diferentes estruturas etárias, eliminando distorções relacionadas ao envelhecimento populacional.

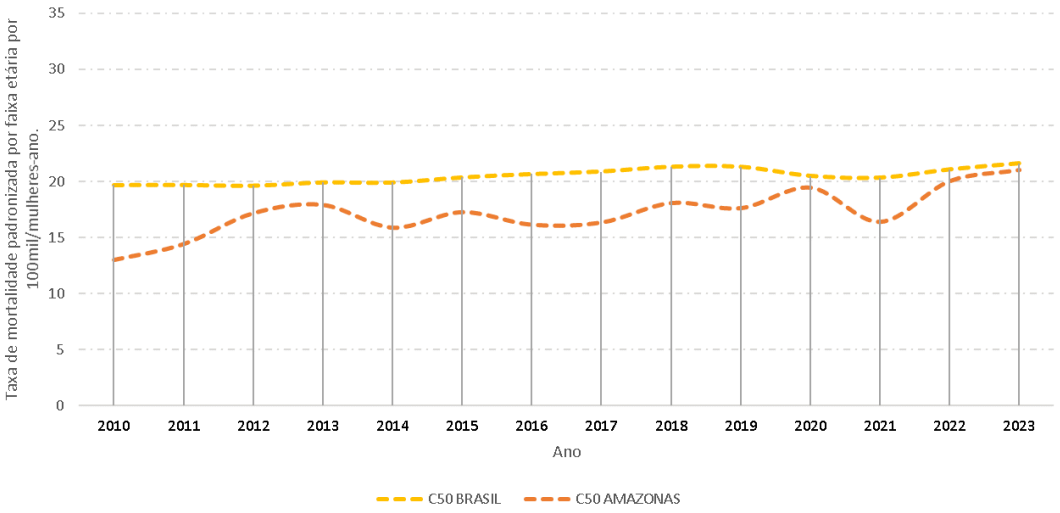
Assim, temos que a média da taxa padronizada de mortalidade por faixa etária no Brasil para o câncer de mama e colo do útero foi, respectivamente, de 20 e 8 óbitos/100 mil habitantes. Por seguinte, no Amazonas o resultado é, respectivamente, 17 e 27 óbitos/100 mil habitantes. Dessa forma, podemos verificar que o câncer de mama aproxima-se da média do Brasil, porém o câncer de colo de útero é quase 4x maior que a média brasileira.

Tabela 9: Taxas de mortalidade padronizada por 100 mil mulheres-ano por CID – Brasil e Amazonas, 2010 a 2023.

CID-10	LOCAL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	MÉDIA
	AMAZONAS	40	43	48	50	47	47	42	45	45	44	44	39	44	44	45
C50	AMAZONAS	13	14	17	18	16	17	16	16	18	18	19	16	20	21	17
C53	AMAZONAS	27	28	31	32	32	29	26	29	27	26	25	23	24	23	27
	BRASIL	28	28	28	28	27	28	28	29	29	29	28	28	29	30	28
C50	BRASIL	20	20	20	20	20	20	21	21	21	21	20	20	21	22	20
C53	BRASIL	8	8	8	8	7	8	8	8	8	8	8	8	16	8	8

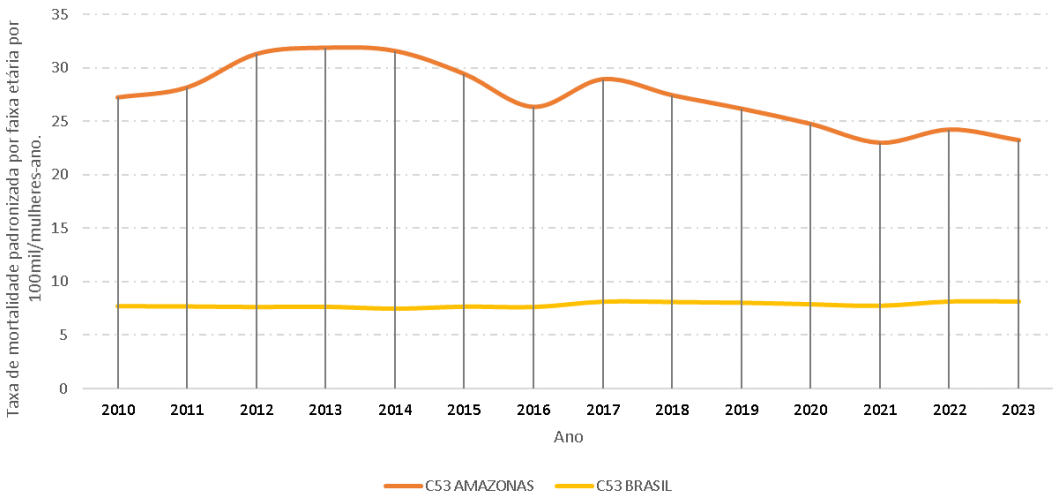
Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População e Censo Demográfico (2010,2022);

Figura 4: Taxa padronizada de mortalidade por câncer de mama – Brasil *versus* Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

Figura 5: Taxa padronizada de mortalidade por câncer de colo de útero – Brasil *versus* Amazonas, 2010 a 2023.

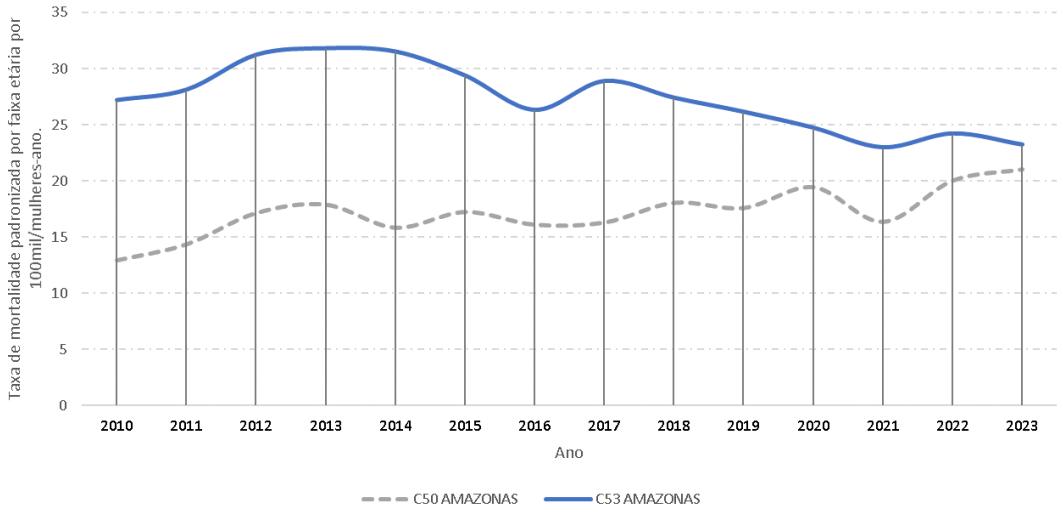


Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025;
Nota: elaboração própria da autora.

Nas Figuras 4 e 5, acerca das taxas de mortalidade por câncer de mama e colo do útero, verificamos discrepância entre o comportamento no Amazonas e Brasil. Em relação ao câncer de mama, apesar do estado possuir valores menores que o país, este apresenta crescimento a partir de 2016, com variação de 61% em contrapartida com

10% no território nacional. Por outro lado, para o câncer do colo de útero verificamos que, no contexto estadual, possui valores 4x acima da média brasileira e, apesar de apresentar decréscimo a partir de 2012, de 14%, representa uma redução pouco acentuada no Amazonas, ao passo que o Brasil permanece estável ao longo do período analisado.

Figura 6: Taxa padronizada de mortalidade por câncer de mama e colo de útero – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

Na Figura 6, ao compararmos o câncer de mama *versus* o câncer do colo do útero, o Amazonas possui disparidade nas taxas no período de 2010 a 2015, com o câncer do colo do útero com maiores taxas de mortalidade. Entretanto, a partir de 2016 essas taxas apresentam valores próximos, ao passo que convergem para uma média de 20 óbitos/100 mil mulheres-ano.

Tabela 10: Taxas padronizadas de mortalidade por faixa etária e CID-10 – Amazonas, 2010 a 2023.

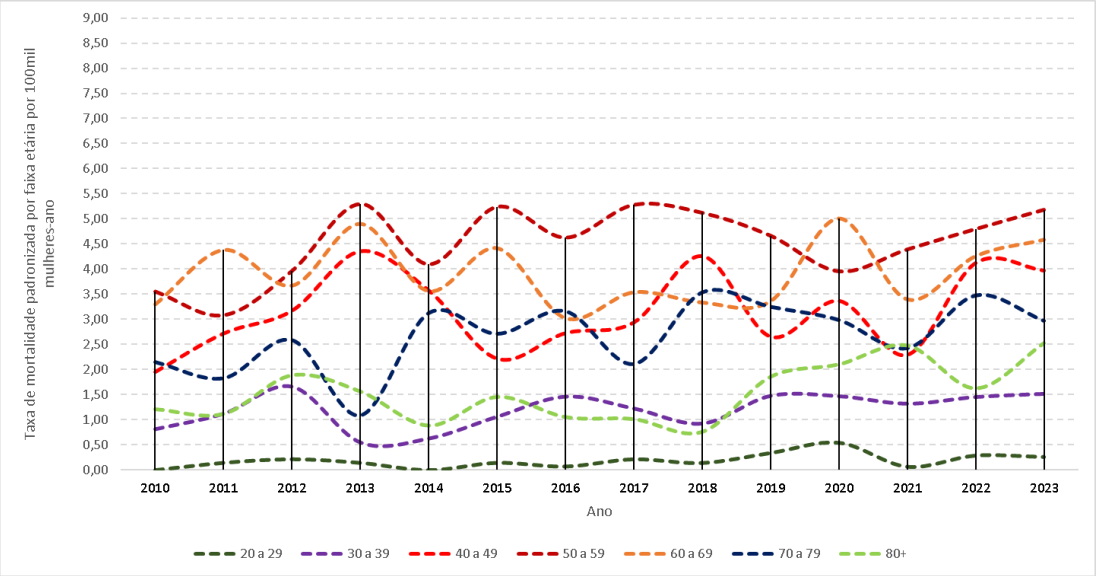
CID-10	FAIXA ETÁRIA	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
	20 a 29	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	30 a 39	1	1	2	1	1	1	1	1	1	1	1	1	1	2
	40 a 49	2	3	3	4	4	2	3	3	4	3	3	2	4	4
C50 MAMA	50 a 59	4	3	4	5	4	5	5	5	5	5	4	4	5	5
	60 a 69	3	4	4	5	4	4	3	4	3	3	5	3	4	5
	70 a 79	2	2	3	1	3	3	3	2	4	3	3	2	4	3
	80+	1	1	2	2	1	1	1	1	1	2	2	2	2	3
	20 a 29	1	0	1	1	1	1	0	1	0	1	1	1	1	1
	30 a 39	3	3	3	4	4	4	2	3	3	3	3	2	2	3
	40 a 49	6	6	7	6	6	5	6	6	5	5	6	5	6	5
C53 ÚTERO	50 a 59	6	6	5	8	7	6	7	6	7	5	5	6	6	4
	60 a 69	5	5	8	5	7	5	5	6	6	5	4	4	4	4
	70 a 79	4	5	4	5	4	5	3	4	4	5	4	3	4	4
	80+	2	3	3	2	3	3	3	2	2	2	2	2	1	3

Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População e Censo Demográfico (2010,2022);

Nota: Taxa = $\frac{\sum_{i=1}^{14} \text{Óbitos}_i}{\sum_{i=1}^{14} \text{Pop}_i}$, onde i representa os anos.

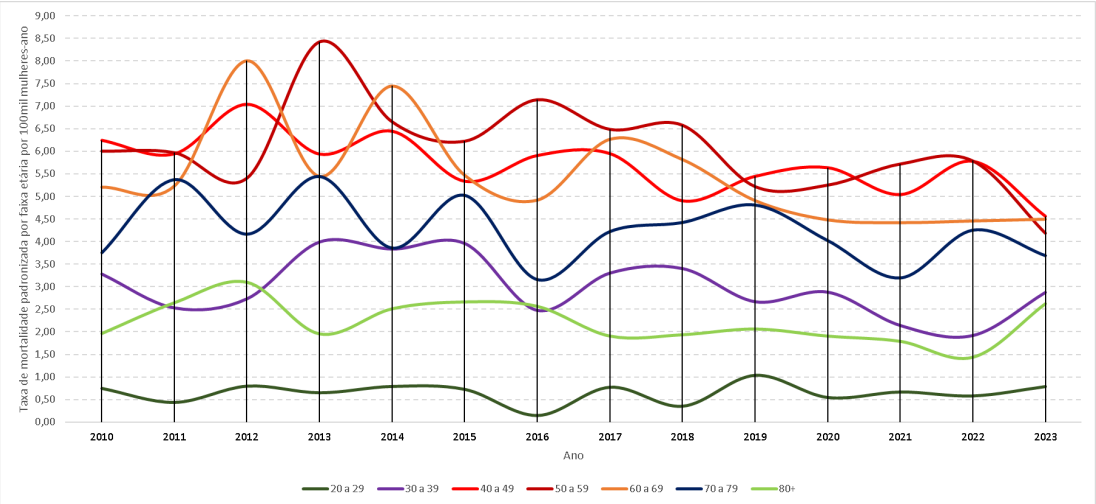
Nas Figuras 7 e 8, acerca das taxas de mortalidade padronizadas por faixa etária, verificamos discrepância nas faixas mais jovens e mais avançadas. Nas faixas intermediárias, entretanto, observamos estabilidade nas taxas de óbitos. No caso do câncer de mama, a partir dos 40 anos a taxa de mortalidade aumenta de forma expressiva, superando a faixa de 60 a 69 anos em alguns anos. Resultado este que é preocupante, uma vez que as recomendações de rastreamento concentram-se entre 50 e 69 anos de idade.

Figura 7: Taxa padronizada de mortalidade câncer de mama por faixa etária – Amazonas, 2010 a 2023.



Fonte: MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025;
Nota: elaboração própria da autora.

Figura 8: Taxa de mortalidade câncer padronizada por faixa etária de colo de útero por faixa etária – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

Nas Figuras 7 e 8 observou-se ainda o surgimento de óbitos na faixa etária de 20 a 29 anos, que passou de 0 para 0,26 óbito por 100 mil mulheres-ano. Por apresentar valor inicial nulo, não foi possível calcular a variação percentual, mas verificou-se

um incremento absoluto equivalente a 0,26 óbito por 100 mil mulheres-ano. Entre as demais faixas etárias, o crescimento anual composto foi de 5,59% para o grupo de 40 a 49 anos, perdendo apenas para o grupo de 80 anos ou mais, que apresentou 5,85%, embora este último abranja um intervalo etário superior a dez anos. Considerando o período de 2010 a 2023, a faixa de 40 a 49 anos apresentou variação total de 102,8%, com aumento da taxa de 2 para 4 óbitos/100 mil mulheres-ano. A faixa de 20 a 29 anos foi a única com menos de 2 óbitos/100 mil mulheres-ano durante todo o período.

No caso do câncer do colo do útero, o perfil é distinto. Embora a faixa de 50 a 59 anos também apresente a maior taxa de óbitos, apenas o grupo de 20 a 29 anos apresentou taxa inferior a 2 óbitos por 100 mil mulheres-ano. A segunda maior ocorrência de óbitos alterna entre as faixas de 40 a 49 e 60 a 69 anos. Em relação às variações percentuais, apenas as faixas de 20 a 29 anos e 80 anos ou mais apresentaram crescimento, com 5,36% e 33,7%, respectivamente, correspondendo a um crescimento anual composto de 0,40% e 2,26%. As demais faixas etárias apresentaram decréscimo, sendo o mais acentuado na faixa de 50 a 59 anos (-2,74% ao ano).

As Figuras 7 e 8 apresentam as taxas de mortalidade por faixa etária, diferenciadas por cores, o que possibilita comparar o comportamento temporal entre os grupos. De modo geral, as maiores taxas concentram-se nas faixas etárias mais elevadas; entretanto, em determinados anos, a faixa de 40 anos apresentou valores superiores aos observados na faixa de 60 anos em ambos os cânceres, indicando que a mortalidade não se restringe apenas às idades mais avançadas. Ressalta-se que, para o câncer de mama, o rastreamento é recomendado principalmente entre 50 e 69 anos, enquanto para o câncer do colo do útero a faixa indicada é de 25 a 64 anos. Quanto à magnitude das taxas, observa-se que o câncer de mama não ultrapassou 6 óbitos por 100 mil mulheres, ao passo que o câncer do colo do útero atingiu valores máximos superiores a 8 óbitos por 100 mil mulheres. Nas faixas mais jovens, a mortalidade por câncer de mama entre 20 e 29 anos apresentou diversos períodos com taxas zeradas, enquanto para o câncer do colo do útero o valor mínimo aproximou-se de zero em apenas um ano da série, sugerindo ocorrência mais persistente ao longo do período analisado. A partir de 2014, verifica-se tendência de redução nas taxas da faixa de 60 a 69 anos para o câncer do colo do útero, embora a faixa de 30 anos tenha mantido valores superiores a 3 óbitos por 100 mil mulheres. Já para o câncer de mama, observa-se maior oscilação das taxas ao longo do período analisado, evidenciando variações importantes entre os anos e faixas etárias.

Tabela 11: Métricas das taxas padronizadas de mortalidade por faixa etária – Amazonas, 2010-2023.

CID-10	Taxa óbitos 2010	Taxa óbitos 2023	Variação % Total	Variação Média Anual	Crescimento Anual Composto
Câncer de mama					
20 a 29	0	0	-	-	-
30 a 39	1	2	87,28%	6,71%	4,94 %
40 a 49	2	4	102,80%	7,91%	5,59 %
50 a 59	4	5	45,95%	3,53%	2,95 %
60 a 69	3	5	39,26%	3,02%	2,58 %
70 a 79	2	3	38,27%	2,94%	2,52 %
80+	1	3	109,50%	8,42%	5,85 %
Câncer do colo do útero					
20 a 29	1	1	5,36%	0,41%	0,40%
30 a 39	3	3	-12,36%	-0,95%	-1,01 %
40 a 49	6	5	-26,98%	-2,08%	-2,39 %
50 a 59	6	4	-30,31%	-2,33%	-2,74 %
60 a 69	5	4	-13,67%	-1,05%	-1,12 %
70 a 79	4	4	-1,70%	-0,13%	-0,13 %
80+	2	3	33,70%	2,59%	2,26 %

Nota 1: Média anual = Variação % Total / 13 anos;

Nota 2: Nota 2: Variação Média Anual = $\left(\frac{\text{Valor Final}}{\text{Valor Inicial}}\right)^{\frac{1}{\text{Nº de Anos}}} - 1 \times 100$.

4.2 Análise modelos

Os dados disponibilizados pelo (Brasil, Ministério da Saúde, 2025) para o período de 2010 a 2023 registram 6.082 óbitos por câncer de colo do útero e de mama para o estado do Amazonas. Posteriormente, ao extrairmos pelo TabWin, esse total se reduz para 6.042 mortes devido à supressão de dados individuais em tabulações com frequências muito baixas, seja este número reduzido de óbitos ou de população, para que não sejam identificados em nível individual (DATASUS, 2025). Dessa forma, as diferenças entre as taxas totais e as recalculadas refletem a influência da completude das informações nos sistemas de registro, e não necessariamente mudanças reais na mortalidade.

4.2.0.1 Interpolação das variáveis socioeconômicas

Algumas variáveis socioeconômicas utilizadas neste estudo, como o Índice de Desenvolvimento Humano (IDH), não estão disponíveis em nível municipal para o período de referência (2022) ou para anos mais recentes. Dessa forma, para evitar a utilização de informações defasadas e possibilitar a inclusão dessa variável como ex-

plicativa nos modelos de regressão, foi realizada uma estimativa do IDH municipal para 2021 com base nos valores observados em 2010.

O procedimento adotado consistiu na aplicação de um método de interpolação baseada no crescimento relativo, utilizando como referência a evolução do IDH em nível estadual. Para isso, foi calculado o Fator de Crescimento Estadual, definido como a razão entre o IDH do estado do Amazonas em 2021 e em 2010. Esse fator foi então aplicado ao IDH municipal observado em 2010, permitindo projetar um valor estimado para 2021. Essa abordagem assume que os municípios seguiram, em média, a trajetória de crescimento observada para o estado ao longo do período.

Formalmente, seja IDH_{2010}^{mun} o valor do IDH do município em 2010 e IDH_{2010}^{AM} e IDH_{2021}^{AM} os valores do IDH do estado do Amazonas nos anos de 2010 e 2021, respectivamente, o IDH municipal estimado para 2021 foi obtido por meio da seguinte expressão:

$$IDH_{2021}^{mun} = IDH_{2010}^{mun} \times \frac{IDH_{2021}^{AM}}{IDH_{2010}^{AM}}$$

Esse procedimento permite a incorporação de uma medida socioeconômica atualizada nos modelos de regressão, preservando as diferenças relativas entre os municípios e mantendo a coerência com a evolução observada em nível estadual.

4.2.0.2 Seleção das covariavéis

Variáveis com elevado percentual de dados ausentes foram excluídas das análises, uma vez que sua inclusão implicaria redução drástica do tamanho amostral e perda de poder estatístico. A seleção das covariáveis priorizou indicadores amplamente utilizados na literatura e com maior completude dos dados.

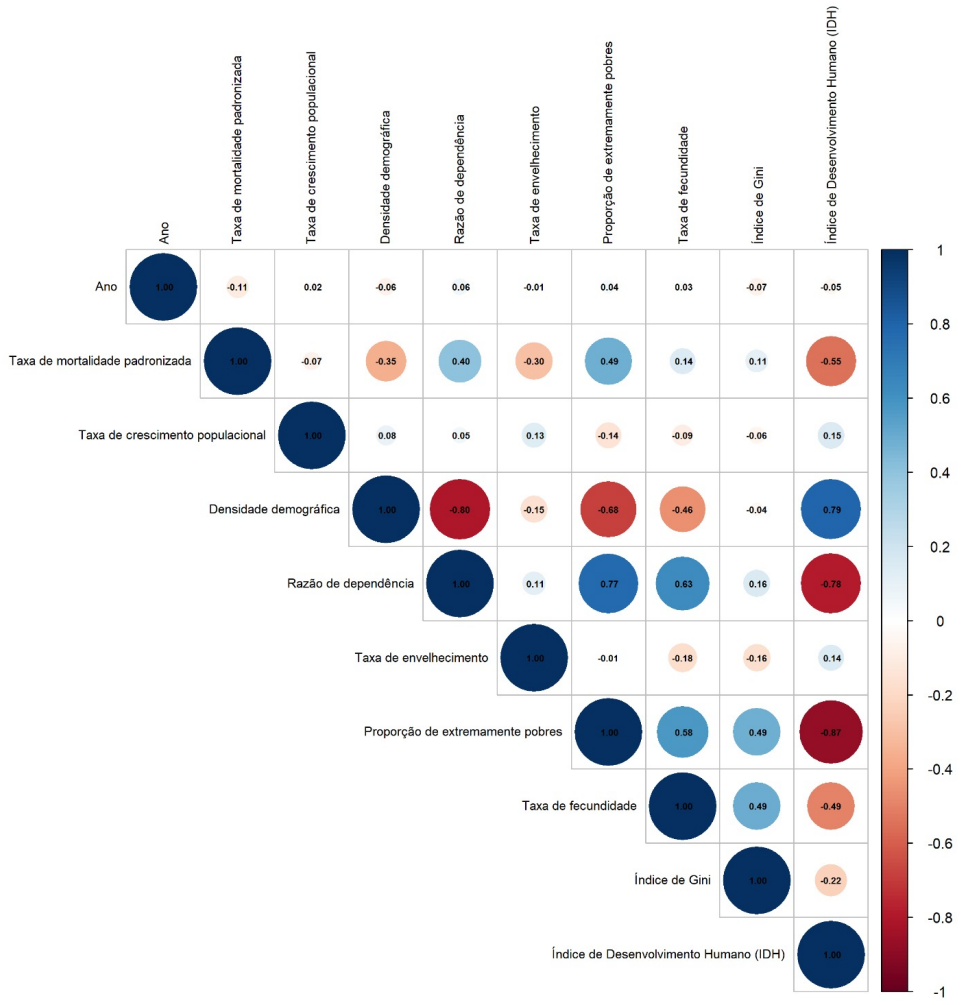
Tabela 12: Resumo estatístico das variáveis independentes incluídas nos modelos de mortalidade por câncer de mama e colo do útero – Amazonas, 2010-2023.

Indicador	Origem	Média	Mediana	Min-Máx	Unidade
Câncer de mama					
Taxa crescimento geométrico	IBGE	0,98	1,14	-2,56 – 3,93	%
Densidade demográfica	IBGE	43,82	6,45	0,15 – 181,01	hab/(km^2)
Razão de dependência	Atlas Brasil	53,03	55,45	37,36 – 75,65	%
Taxa de envelhecimento	Atlas Brasil	6,37	6,31	3,82 – 8,38	%
Condição socioeconômica	Atlas Brasil	14,51	14,94	2,55 – 37,96	%
Fecundidade	DATASUS/IBGE	71,30	69,97	42,18 – 118,91	%
Desigualdade social	Atlas Brasil	0,53	0,53	0,46 – 0,70	Índice (0-1)
Desenvolvimento humano	Atlas Brasil	0,65	0,64	0,47 – 0,77	Índice (0-1)
Câncer do colo do útero					
Taxa crescimento geométrico	IBGE	1,00	1,14	-2,56 – 3,93	%
Densidade demográfica	IBGE	24,22	3,11	0,15 – 181,01	hab/(km^2)
Razão de dependência	Atlas Brasil	56,43	57,53	37,36 – 75,65	%
Taxa de envelhecimento	Atlas Brasil	6,25	6,09	3,82 – 8,38	%
Condição socioeconômica	Atlas Brasil	17,78	17,56	2,55 – 37,96	%
Fecundidade	DATASUS/IBGE	76,75	74,33	42,18 – 118,91	%
Desigualdade social	Atlas Brasil	0,54	0,54	0,46 – 0,70	Índice (0-1)
Desenvolvimento humano	Atlas Brasil	0,62	0,62	0,47 – 0,77	Índice (0-1)
Razão Colo do útero (cobertura)	DATASUS/IBGE	0,72	0,70	0,22 – 1,54	-
Câncer de mama + colo do útero					
Taxa crescimento geométrico	IBGE	1,00	1,14	-2,56 – 3,93	%
Densidade demográfica	IBGE	30,57	4,29	0,15 – 181,01	hab/(km^2)
Razão de dependência	Atlas Brasil	55,33	56,57	37,36 – 75,65	%
Taxa de envelhecimento	Atlas Brasil	6,29	6,13	3,82 – 8,38	%
Condição socioeconômica	Atlas Brasil	16,72	15,51	2,55 – 37,96	%
Fecundidade	DATASUS/IBGE	74,99	74,23	42,18 – 118,91	%
Desigualdade social	Atlas Brasil	0,54	0,54	0,46 – 0,70	Índice (0-1)
Desenvolvimento humano	Atlas Brasil	0,63	0,63	0,47 – 0,77	Índice (0-1)
Razão Mamografia (cobertura)	DATASUS/IBGE	0,03	0,01	0,00 – 0,20	-

Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

4.2.0.3 Analise descritiva covariaveis

Figura 9: Matriz de correlação entre variáveis sociodemográficas associadas à mortalidade por neoplasias – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

Na Figura 9 a matriz de correlação foi construída utilizando apenas as variáveis incluídas no modelo final, com o objetivo de avaliar possíveis relações de multicolinearidade entre os preditores. Além disso, as tabelas do Fator de Inflação da Variância (VIF) encontram-se disponíveis no Apêndice A, contemplando todos os modelos ajustados neste estudo[1].

A análise da matriz de correlação foi realizada para verificar a associação linear entre as variáveis sociodemográficas e a taxa padronizada de mortalidade, classificadas

como correlações fracas ($|r| < 0,30$), moderadas ($0,30 \leq |r| < 0,70$) e fortes ($|r| \geq 0,70$), independentemente do sinal.

Assim, observamos que a taxa padronizada de mortalidade apresentou correlação positiva moderada com a proporção de extremamente pobres (r = 0,49) e com a razão de dependência (r = 0,40), sugerindo que municípios com maior vulnerabilidade socioeconômica e maior proporção de população dependente tendem a apresentar maiores taxas de mortalidade.

Por seguinte, identificamos correlação positiva fraca com o Índice de Gini (r = 0,11) e com a taxa de fecundidade (r = 0,14), indicando, respectivamente, heterogeneidade populacional e transição demográfica tendem a apresentar discretamente maiores taxas.

Em contrapartida, verificaramos correlações negativas moderadas entre a taxa padronizada de mortalidade e o Índice de Desenvolvimento Humano – IDH (r = -0,55), com a densidade demográfica (r = -0,35) e com a taxa de envelhecimento (r = -0,30), sugerindo que melhores condições de desenvolvimento, maior urbanização e maior proporção de idosos estão associadas à redução da mortalidade. Mas também, observadamos correlações negativas fracas com a proporção de população branca (r = -0,25) e com outras variáveis sociodemográficas.

Além disso, observou-se correlação negativas com a taxa de crescimento populacional (r = -0,07) e, principalmente, com o Índice de Desenvolvimento Humano – IDH (r = -0,55), sugerindo que melhores condições socioeconômicas estão associadas a menores taxas de mortalidade.

De modo geral, os resultados indicam que determinantes sociais relacionados à desigualdade e à vulnerabilidade econômica apresentaram maior associação com o aumento da mortalidade, enquanto indicadores de desenvolvimento e estrutura demográfica mostraram associação com a redução do risco.

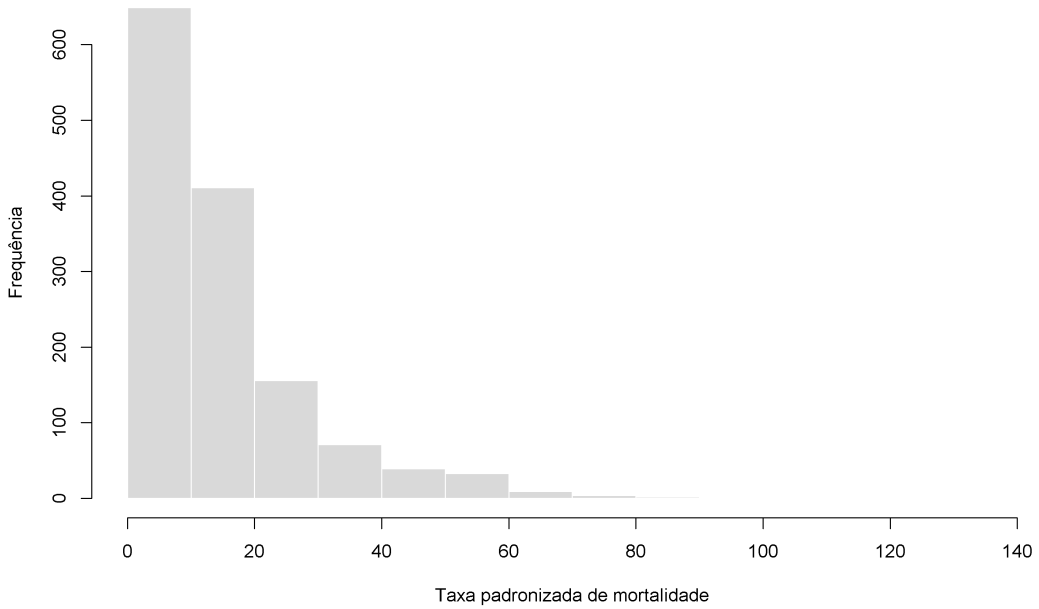
[1] Apesar da presença de variáveis potencialmente correlacionadas, como extrema pobreza e IDH, foi realizada a avaliação do fator de inflação da variância (VIF), indicando no máximo multicolinearidade moderada. Ressalta-se ainda que o IDH não apresentou significância estatística no modelo final.

4.2.1 Resultaldo dos modelos

Inicialmente foi ajustado um modelo de regressão linear múltipla. Entretanto, a avaliação dos pressupostos do modelo indicou violações de normalidade e homocedasticidade dos resíduos, além de inadequação da distribuição da variável resposta. Tais limitações sugerem inadequação da suposição de distribuição normal dos erros e da variância constante, comprometendo a validade inferencial do modelo. Dessa forma, optou-se pela utilização de Modelos Lineares Generalizados.

4.2.1.1 Modelo linear generalizado

Figura 10: Distribuição da taxa padronizada de mortalidade por neoplasias – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População; WHO: World Health Organization, 2025.

Na Figura 10 evidencia-se que a variável resposta apresenta distribuição assimétrica à direita, com predominância de valores positivos e presença de cauda longa. Esse comportamento viola o pressuposto de normalidade dos resíduos exigido pelo modelo de regressão linear clássico.

Diante dessa característica, optou-se pela utilização de Modelos Lineares Gene-

ralizados com distribuição Gamma e função de ligação logarítmica. A distribuição Gamma é indicada para modelagem de variáveis contínuas, positivas e assimétricas, sendo amplamente empregada em estudos epidemiológicos envolvendo taxas e medidas de risco. A função de ligação log permite interpretar os coeficientes estimados como efeitos multiplicativos, possibilitando a obtenção de razões de taxa por meio da exponenciação dos coeficientes do modelo.

Sua função densidade de probabilidade é dada por

$$f(y;\alpha,\theta) = \frac{y^{\alpha-1}e^{-y/\theta}}{\Gamma(\alpha)\theta^{\alpha}}, \quad y > 0,$$

com esperança $E(Y) = \alpha\theta$ e variância proporcional ao quadrado da média. No contexto dos Modelos Lineares Generalizados, utilizou-se a função de ligação logarítmica, definida por $\log(\mu) = X\beta$, garantindo estimativas positivas, uma vez que taxas de mortalidade não podem assumir valores negativos.

4.2.1.2 Modelo Geral

Para a dependência entre observações pertencentes ao mesmo município, foram estimados erros-padrão robustos clusterizados por município utilizando o estimador robusto do tipo sandwich[2]. Essa abordagem permitiu observar o aumento consistente dos erros-padrão quando comparados aos estimadores convencionais, indicando presença de correlação intra-município. Dessa forma, a interpretação dos resultados foi baseada nos estimadores robustos clusterizados.

.

[2]Foram utilizados erros-padrãos robustos do tipo sandwich para reduzir viés na inferência estatística diante da presença de heterocedasticidade e possível depência entre observaçõesm garantindo estimati-

Tabela 13: Comparação entre erros-padrão convencionais e clusterizados

Variável	**EP Convencional**	**EP Cluster**	**Impacto**
Intercepto	0.352	1.054	Aumento grande
cid C53	0.034	0.052	Aumento moderado
Ano (Spline 1)	0.064	0.072	Aumento leve
Ano (Spline 2)	0.133	0.157	Aumento moderado
Ano (Spline 3)	0.047	0.060	Aumento moderado
Faixa 30–39	0.079	0.228	Aumento grande
Faixa 40–49	0.076	0.275	Aumento grande
Faixa 50–59	0.076	0.280	Aumento grande
Faixa 60–69	0.078	0.269	Aumento grande
Faixa 70–79	0.081	0.241	Aumento grande
Faixa 80+	0.084	0.189	Aumento moderado
Densidade	0.00046	0.00125	Aumento grande
Razão de dependência	0.00397	0.00981	Aumento grande
Taxa de envelhecimento	0.017	0.062	Aumento grande
Extremamente pobres	0.0037	0.011	Aumento grande
Taxa de fecundidade	0.0014	0.0039	Aumento grande
Gini	0.597	1.953	Aumento grande

Tabela 14: Estimativas do modelo de regressão Gamma com erros robustos para a taxa padronizada de mortalidade.

Variável	Estimativa	Erro Padrão	z	p-valor
Intercepto	4,4305	1,0546	4,2012	$< 0,001$***
CID C53	0,1385	0,0521	2,6576	0,0079**
ns(ano) 1	-0,2189	0,0721	-3,0353	0,0024**
ns(ano) 2	-0,4436	0,1571	-2,8239	0,0047**
ns(ano) 3	-0,2781	0,0603	-4,6099	$< 0,001$***
Faixa 30–39	0,4204	0,2288	1,8371	0,0662
Faixa 40–49	0,8066	0,2774	2,9074	0,0036**
Faixa 50–59	0,9752	0,2808	3,4732	0,0005***
Faixa 60–69	1,0063	0,2692	3,7386	$< 0,001$***
Faixa 70–79	0,9624	0,2413	3,9875	$< 0,001$***
Faixa 80+	0,7167	0,1899	3,7742	$< 0,001$***
Densidade	-0,0034	0,0013	-2,6708	0,0076**
Razão de dependência	0,0147	0,0098	1,4944	0,1351
Taxa de envelhecimento	-0,2948	0,0626	-4,7119	$< 0,001$***
Extrema pobreza	0,0574	0,0116	4,9673	$< 0,001$***
Taxa de fecundidade	-0,0169	0,0040	-4,2641	$< 0,001$***
Gini	-2,2066	1,9531	-1,1298	0,2586

* p<0,05; ** p<0,01; *** p<0,001

O modelo ajustado corresponde a uma regressão do tipo Gamma com função de ligação logarítmica, apropriada para modelagem de taxas positivas e assimétricas. O

vas mais consistentes dos erros-padrãos.

modelo apresentou redução expressiva da deviance residual em relação à deviance nula, indicando melhora do ajuste aos dados. O Critério de Informação de Akaike (AIC = 8.640,13) foi utilizado para avaliação comparativa do desempenho do modelo. O parâmetro de dispersão estimado foi de 0,33, sugerindo variabilidade compatível com a distribuição assumida.

Tabela 15: Interpretação dos coeficientes do modelo Gamma com ligação logarítmica para mortalidade por neoplasias.

Variável	β	**exp(β)**	**Interpretação**
CID C53	0,1385	1,14	Aumento de 14%
ns(ano, df = 3) 1	-0,2189	0,80	Redução de 20%
ns(ano, df = 3) 2	-0,4436	0,64	Redução de 36%
ns(ano, df = 3) 3	-0,2781	0,75	Redução de 25%
Faixa etária 30–39 anos	0,4204	1,52	Aumento de 52%
Faixa etária 40–49 anos	0,8066	2,24	Aumento de 124%
Faixa etária 50–59 anos	0,9752	2,65	Aumento de 165%
Faixa etária 60–69 anos	1,0063	2,74	Aumento de 174%
Faixa etária 70–79 anos	0,9624	2,62	Aumento de 162%
Faixa etária 80 anos ou mais	0,7167	2,05	Aumento de 105%
Densidade demográfica	-0,0034	0,997	Redução de 0,3%
Razão de dependência	0,0147	1,02	Aumento de 2,2%
Taxa de envelhecimento	-0,2948	0,74	Redução de 26%
Proporção de extremamente pobres	0,0574	1,06	Aumento de 6,1%
Taxa de fecundidade	-0,0169	0,98	Redução de 2,1%
Índice de Gini	-2,2066	0,15	Redução de 85%

A análise de regressão linear generalizada evidenciou que o modelo otimizado apresentou capacidade explicativa alta, com pseudo coeficiente de determinação ajustado de 0,7168, ou seja, 71,68% da variabilidade da taxa padronizada de mortalidade foi explicada pelas variáveis incluídas no modelo. A comparação entre os modelos evidenciou redução do Critério de Informação de Akaike, equilibrando o ajuste dos dados com a complexidade do modelo.

Por seguinte, em relação ao câncer analisado e faixas etárias, obtivemos associação estatisticamente significativa entre a mortalidade e o tipo de neoplasia, sendo o CID C53 associado ao aumento médio de 1,14 vezes na taxa padronizada em relação à categoria de referência C50. A faixa etária apresentou forte influência na mortalidade, evidenciando aumento progressivo do risco com o avanço da idade, com aumento médio entre 2,67 e 2,74 nas faixas de 50 a 69 anos em relação a categoria de 20 a 29 anos, e por fim, com descréscimo nas faixas 30 a 39 e 80 anos ou mais com aumento de, respectivamente, 1,55 e 2,11.

Além disso, para o período e demografia, temos que a variável ano apresentou tendência de redução das taxas. Entre os indicadores demográficos, a razão de dependência apresentou associação positiva com a mortalidade, enquanto a taxa de envelhecimento demonstrou associação inversa.

Por fim, no campo socioeconômico, a proporção de indivíduos em extrema pobreza mostrou associação positiva com a mortalidade, reforçando as desigualdades sociais na determinação do risco. Por fim, a avaliação da multicolinearidade indicou ausência de colinearidade relevante entre as variáveis independentes, assegurando estabilidade e confiabilidade das estimativas obtidas.

4.2.1.3 Modelo câncer de mama

Tabela 16: Estimativas do modelo Gamma com ligação log com erros-padrão robustos clusterizados para mortalidade por câncer de mama (C50) com razão de taxas e interpretação percentual.

Variável	β	exp(β)	Interpretação	p-valor
Intercepto	5.5586	259.37	Valor quando covariáveis são 0	
ns(ano, df = 3)1	-0.2172	0.805	Redução de 19,5%	0,1464
ns(ano, df = 3)2	-0.2478	0.781	Redução de 21,9%	0,3320
ns(ano, df = 3)3	-0.2541	0.776	Redução de 22,4%	0,0192*
Faixa 30 a 39	0.4290	1.536	Aumento de 53,6%	0,1884
Faixa 40 a 49	0.7770	2.175	Aumento de 117,5%	0,0771.
Faixa 50 a 59	0.9879	2.685	Aumento de 168,5%	0,0389*
Faixa 60 a 69	1.1311	3.099	Aumento de 209,9%	0,0123*
Faixa 70 a 79	0.9654	2.626	Aumento de 162,6%	0,0189*
Faixa 80 ou mais	0.7241	2.063	Aumento de 106,3%	0,0238*
Taxa de crescimento	-0.0467	0.954	Redução de 4,6%	0,4352
Densidade demográfica	-0.0046	0.995	Redução de 0,45%	$< 0,001$***
Taxa de envelhecimento	-0.2785	0.757	Redução de 24,3%	$< 0,001$***
Extremamente pobres	0.0857	1.089	Aumento de 8,9%	$< 0,001$***
Taxa de fecundidade	-0.0172	0.983	Redução de 1,7%	$< 0,001$***
Índice de Gini	-4.1640	0.016	Redução de 98,4%	0,1228
Razão mama	2.1936	8.968	Aumento de 796,8%	0,0289*

. p<0,10; * p<0,05; ** p<0,01; *** p<0,001

A análise específica para mortalidade por câncer de mama manteve-se semelhante ao observado no modelo geral, apresentou evidências de variação temporal ao longo do período analisado, com significância estatística em um dos componentes da função spline para o ano, sugerindo possível redução em parte do intervalo estudado. Observamos aumento do risco com o avanço da idade, com resultados superiores à observada

no modelo conjunto, com destaque para as faixas etárias acima de 40 anos apesar de não significativa. Entre os indicadores demográficos, a razão de dependência foi desconsiderada pelo modelo ótimo, após considerar simultaneamente as outras variáveis, ela não contribuiu significativamente para explicar a variação da mortalidade, enquanto a taxa de envelhecimento manteve associação inversa, reforçando a complexidade da estrutura etária na determinação do risco. No campo socioeconômico, a proporção de indivíduos em extrema pobreza permaneceu associada ao aumento da mortalidade, porém com valor superior ao observada no modelo geral. A taxa de fecundidade manteve associação inversa com a mortalidade, corroborando o padrão geral. De forma geral, os resultados demonstram consistência entre os modelos, com pequenas variações na magnitude das associações, sugerindo que os determinantes sociodemográficos apresentam influência diferenciada conforme o tipo de neoplasia analisada.

4.2.1.4 Modelo câncer do colo do útero

Tabela 17: Estimativas do modelo Gamma com ligação log com erros-padrão robustos clusterizados para mortalidade por câncer do colo do útero (C53) com razão de taxas e interpretação percentual.

Variável	β	exp(β)	Interpretação	p-valor
Intercepto	3.7726	43.49	Valor quando covariáveis são 0	
ns(ano, df = 3)1	-0.2009	0.818	Redução de 18,2%	**
ns(ano, df = 3)2	-0.5425	0.581	Redução de 41,9%	***
ns(ano, df = 3)3	-0.3204	0.726	Redução de 27,4%	***
Faixa 30 a 39	0.4863	1.63	Aumento de 62,6%	*
Faixa 40 a 49	0.8830	2.42	Aumento de 141,8%	***
Faixa 50 a 59	1.0008	2.72	Aumento de 172,0%	***
Faixa 60 a 69	1.0266	2.79	Aumento de 179,1%	***
Faixa 70 a 79	0.9936	2.70	Aumento de 170,1%	***
Faixa 80 ou mais	0.7690	2.16	Aumento de 115,8%	***
Densidade demográfica	-0.0028	0.997	Redução de 0,27%	*
Razão de dependência	0.0203	1.02	Aumento de 2,1%	*
Taxa de envelhecimento	-0.2782	0.757	Redução de 24,3%	***
Extremamente pobres	0.0505	1.05	Aumento de 5,2%	***
Taxa de fecundidade	-0.0161	0.984	Redução de 1,6	***
Índice de Gini	-1.1321	0.322	Redução de 67,8%	
Razão colo do útero	-0.2254	0.798	Redução de 20,2%	

* p<0,05; ** p<0,01; *** p<0,001

A análise específica para mortalidade por câncer do colo do útero manteve-se semelhante ao observado no modelo geral, indicando tendência temporal de redução das taxas ao longo do período analisado. Observamos aumento do risco com o avanço

da idade, com resultados superiores à observada no modelo conjunto, com destaque para as faixas etárias acima de 50 anos. Entre os indicadores demográficos, a razão de dependência permaneceu associada ao aumento da mortalidade, enquanto a taxa de envelhecimento manteve associação inversa, reforçando a complexidade da estrutura etária na determinação do risco.

No campo socioeconômico, a proporção de indivíduos em extrema pobreza permaneceu associada ao aumento da mortalidade, porém com valor inferior à observada no modelo geral. A taxa de fecundidade manteve associação inversa com a mortalidade, corroborando o padrão geral. De forma geral, os resultados demonstram consistência entre os modelos, com pequenas variações na magnitude das associações, sugerindo que os determinantes sociodemográficos apresentam influência diferenciada conforme o tipo de neoplasia analisada.

A avaliação diagnóstica das neoplasias em conjunto e individuais encontram-se no apêndice de figuras complementares. Observamos no geral associação positiva entre valores observados e preditos e a análise dos resíduos pelo pacote DHARMa não identificou desvios significativos da distribuição teórica, nem problemas de dispersão. Embora tenha sido detectada presença de valores extremos, comportamento é esperado em dados epidemiológicos, mas não comprometeram o ajuste global dos modelos.

4.3 Análise espacial

4.3.1 Análise descritiva temporal

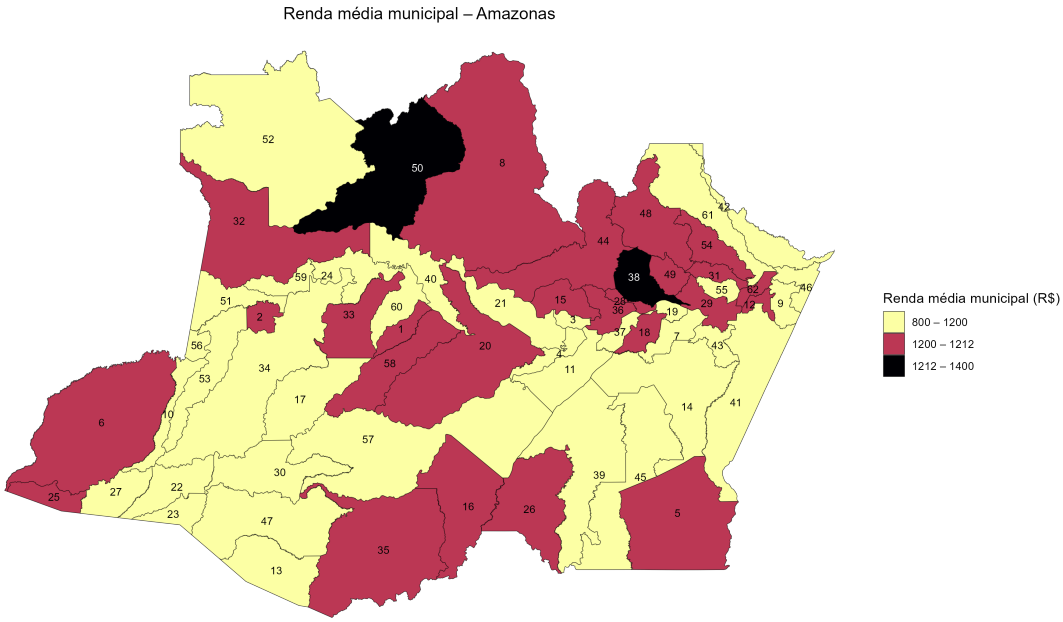
As taxas brutas municipais na análise espacial, foram calculadas utilizando como denominador a população feminina residente em cada município no ano correspondente, permitindo identificar a heterogeneidade do risco entre os municípios, inclusive aqueles sem registro de óbitos no período, os quais foram mantidos na análise com taxa igual a zero. Dessa forma, diferentemente do último resultado apresentado na análise descritiva verificaremos que existe discrepância entre a taxa estadual e algumas taxas municipais elevadas, uma vez com o número reduzido populacional de determinados municípios, nos quais a ocorrência de poucos óbitos resultou em taxas brutas altas, demonstrando a variabilidade.

Assim, apresentamos a representação cartográfica através de mapas coropléticos dividido em quatro classes. A classificação das taxas médias com ou sem padroniza-

ção foi realizada por meio do método dos quartis. Seja $x_1, x_2, \ldots, x_n$ o conjunto das taxas municipais, ordenadas de forma crescente ($x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$). Os pontos de corte foram definidos pelos quantis de ordem $p = \{0, 0{,}25, 0{,}50, 0{,}75, 1\}$, correspondentes ao primeiro quartil (Q_1), mediana (Q_2) e terceiro quartil (Q_3). Dessa forma, cada classe representa aproximadamente 25% dos municípios, permitindo identificar a distribuição das taxas entre os grupos com menores, intermediários e maiores níveis de mortalidade. Por fim, Os municipios sem óbitos no período de 2010 a 2023 estão representados pela cor branca. Além disso, somente os valores maiores que zero foram classificados como quartis.

Os identificadores numéricos (ID), conforme a Figura 11, atribuídos aos municípios foram definidos com base na ordenação crescente dos códigos do IBGE, do menor para o maior, garantindo uma padronização única e consistente para fins de análise. Além disso, para verificação anual dos mapas de taxas padronizadas e risco relativo consultar apêndice B.

Figura 11: Renda média municipal – Amazonas, 2022.



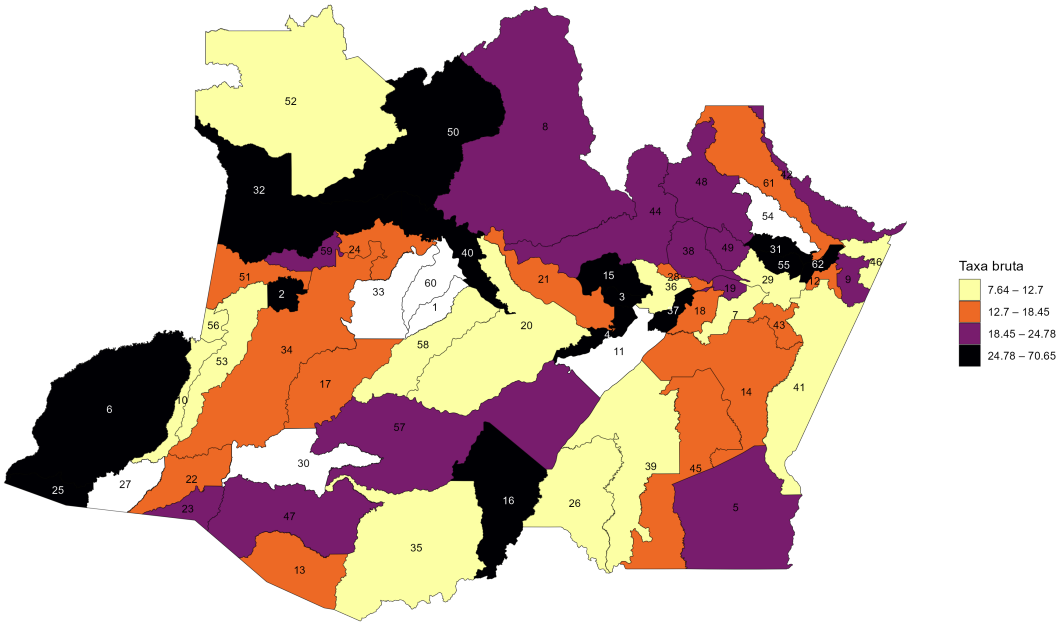
Fonte: Adaptado de IBGE – Censo Demográfico.

Os intervalos de classificação foram definidos a partir da distribuição das taxas médias padronizadas por município no período total (2010–2023). Esses pontos de

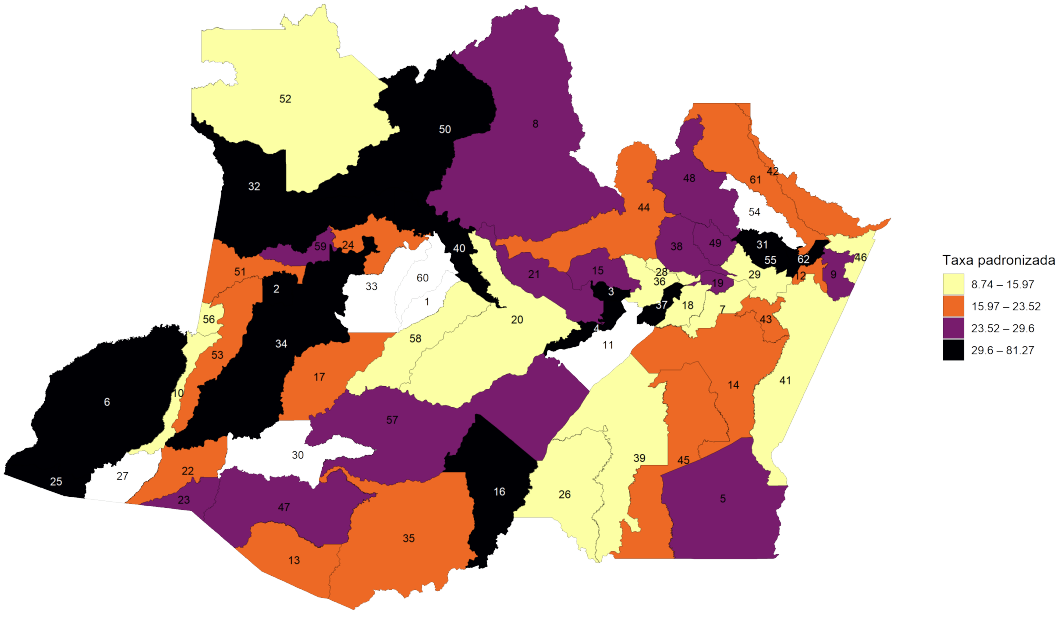
corte foram mantidos constantes nos mapas anuais, permitindo a comparação temporal das taxas em relação a um padrão epidemiológico de referência.

Figura 12: Taxas média bruta e padronizada de mortalidade por câncer de mama – Amazonas, 2010 a 2023.

Taxa bruta /100 mil mulheres



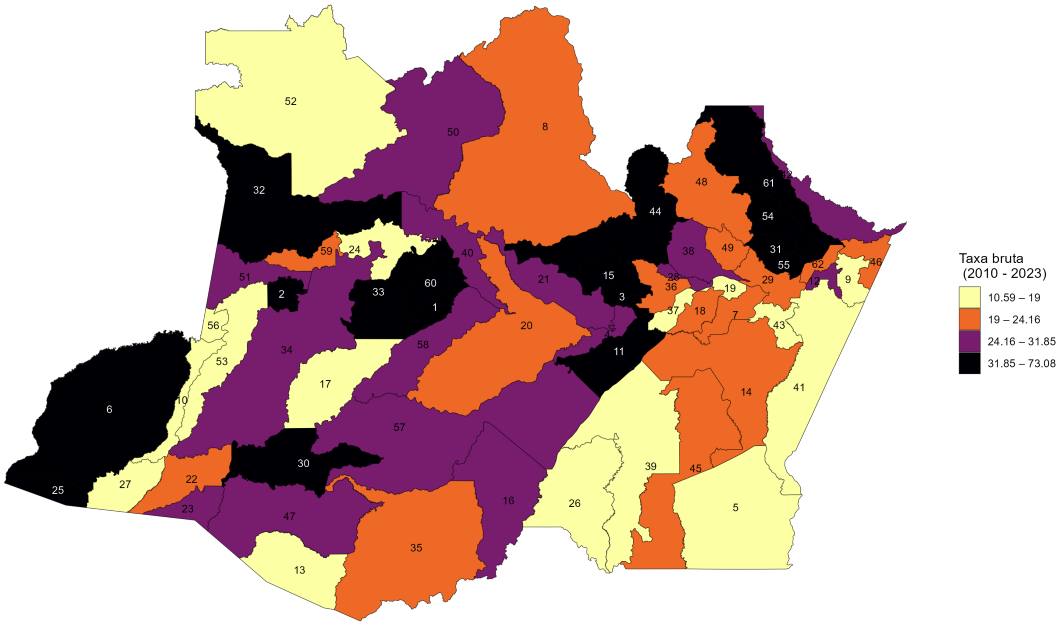
Taxa padronizada /100 mil mulheres



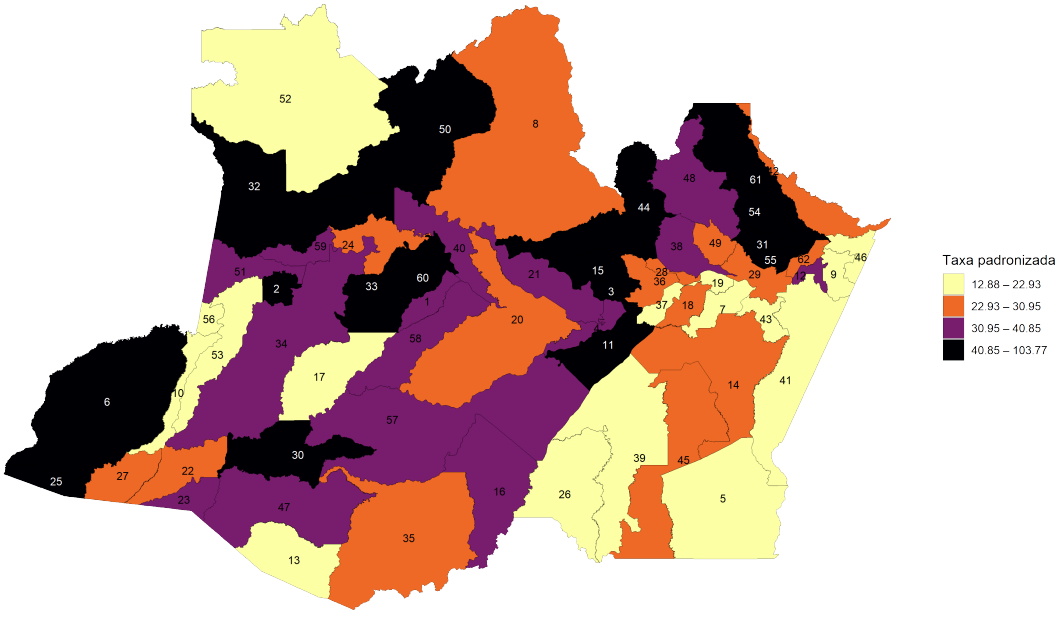
Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

Figura 13: Taxas média bruta e padronizada de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023.

Taxa bruta /100 mil mulheres



Taxa padronizada /100 mil mulheres



Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

Ao analisarmos a Figura 12 verificamos que as taxas brutas e padronizadas de mortalidade por câncer de mama possuem discrepâncias entre os resultados. A saber, as taxas brutas apresentam maior variabilidade em todo território, especialmente em municipios com populaçãor menor que 20 mil-mulheres. Em contrapartida, o mapa de taxas padronizadas apresenta menor variabilidade e distribuição homogênea. Dessa forma, verificamos que as diferenças são atribuiveis a estrutura etária dos municipios. No entanto, alguns municípios, como Atalaia do Norte, Canutama, Maraã e Guajará, permaneceram classificados nas faixas de maior mortalidade mesmo após a padronização, indicando a presença de maior risco relativo não explicado apenas pela composição etária populacional.

Já na Figura 13, acerca da mortalidade por câncer do colo do útero, observamos padrão espacial distinto do identificado para o câncer de mama. As taxas brutas apresentam elevada variabilidade entre os municípios. Entretanto, mesmo após a padronização por idade, persistem áreas com taxas elevadas e distribuição espacial mais concentrada, sugerindo maior heterogeneidade territorial desse agravo. Dessa forma, indicando que fatores além da estrutura etária, como desigualdades, podem contribuir para a manutenção das elevadas taxas em determinados municípios.

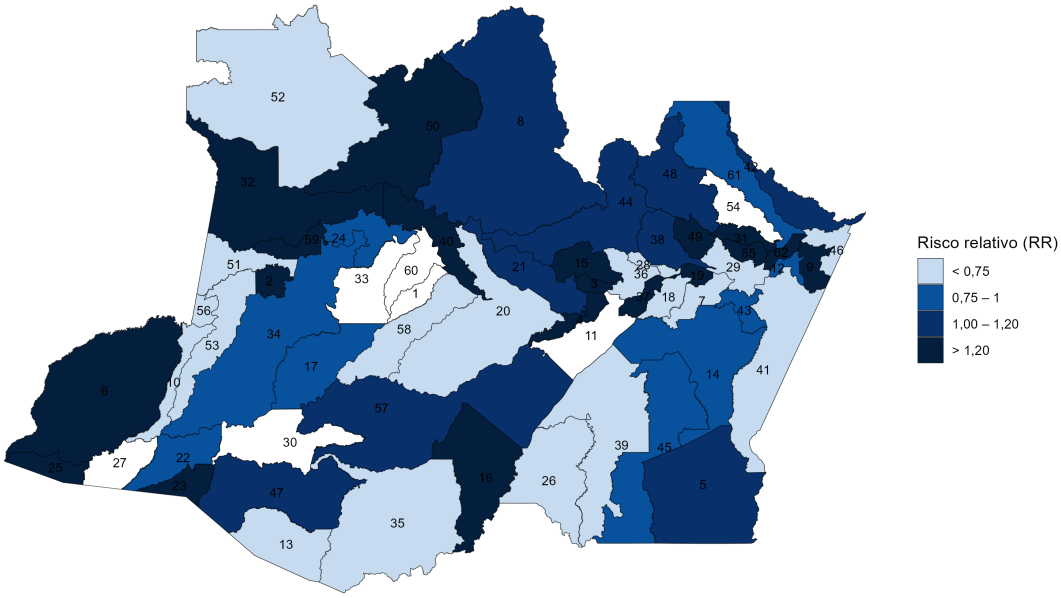
Ao relacionar o mapa de distribuição espacial da renda com os mapas de taxas padronizadas de mortalidade por câncer de mama e câncer do colo do útero, observa-se padrão territorial que sugere associação entre condições socioeconômicas e distribuição dos agravos.

Os municípios com menores níveis de renda concentram-se como classificados nas faixas mais elevadas de mortalidade por câncer do colo do útero, mesmo após a padronização por idade. Indicando que as altas taxas padronizadas nessas áreas constantes não podem ser explicadas exclusivamente pela estrutura etária da população. Por outro lado, a mortalidade por câncer de mama apresenta distribuição espacial mais homogênea e menor concentração em áreas de baixa renda. Indicando maior explicação pela estrutura etária da população.

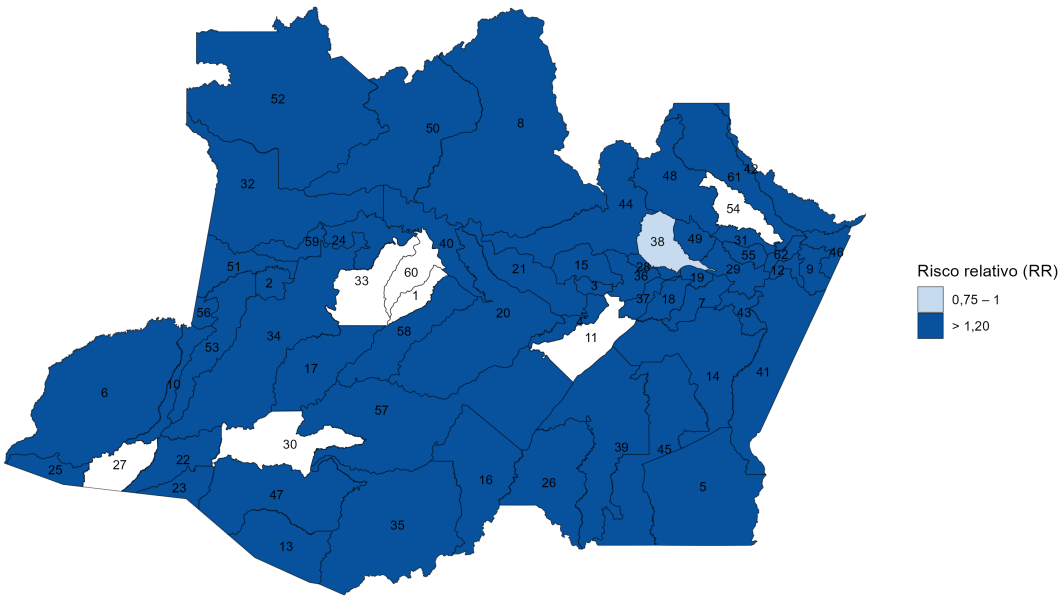
Dessa forma, a sobreposição dos mapas evidencia que as desigualdades socioeconômicas desempenham papel relevante na mortalidade feminina por neoplasias no estado, principalmente para o câncer do colo do útero.

Figura 14: Risco relativo (RR) bruto e padronizado de mortalidade por câncer de mama – Amazonas, 2010 a 2023.

RR bruto



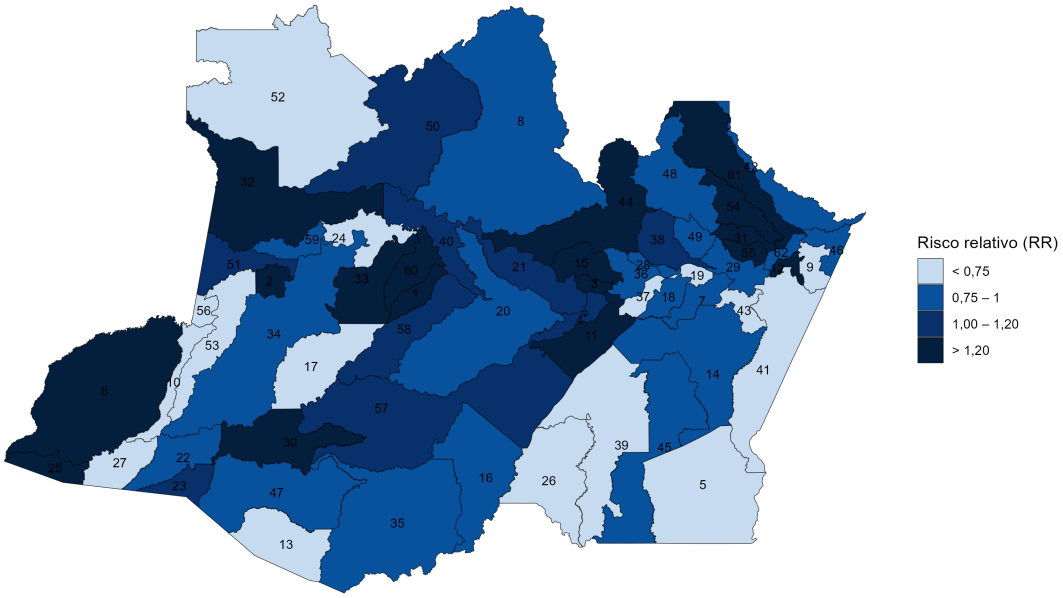
RR padronizado



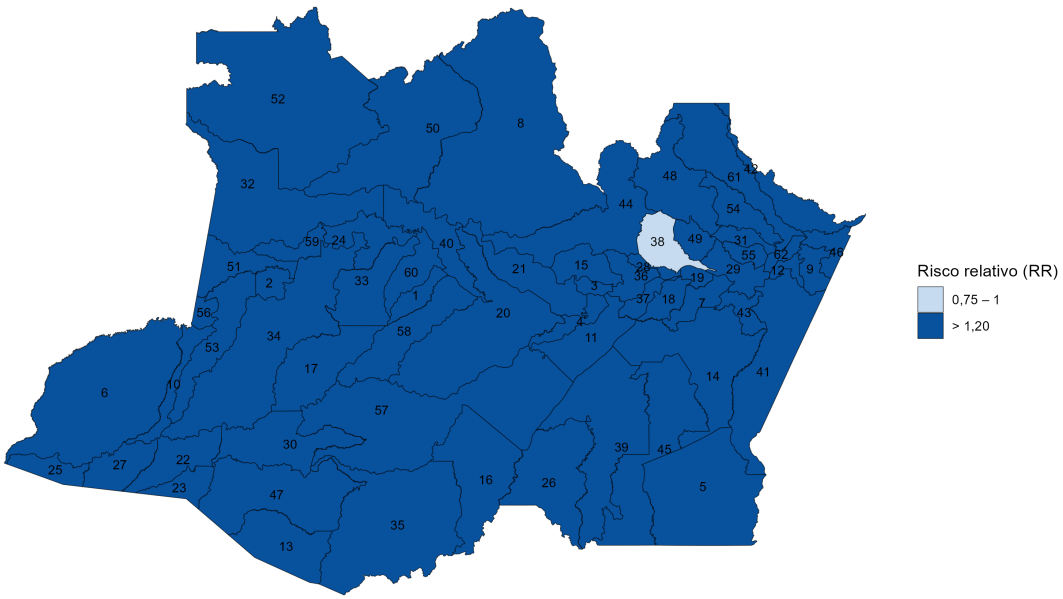
Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

Figura 15: Risco relativo (RR) bruto e padronizado de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023.

RR bruto



RR padronizado



Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

Sob uma formulação mais geral, o número esperado de casos pode ser obtido por meio da padronização indireta estratificada, considerando, por exemplo, faixas etárias e períodos temporais. Nesse caso, o número esperado de casos na área z é dado por:

$$e_z = \sum_a N_{z,a} \times r_a, \tag{4.1}$$

em que $N_{z,a}$ representa a população sob risco na área z pertencente à faixa etária a, e r_a denota a taxa de referência correspondente à faixa etária a, estimada a partir da população total da região de estudo.

Quando a padronização não considera estratos, essa expressão se reduz ao caso particular em que a taxa de referência é a taxa média global do evento, resultando em:

$$e_z = N_z \times \frac{\sum_{z=1}^{Z} c_z}{\sum_{z=1}^{Z} N_z}. \tag{4.2}$$

4.3.2 Análise puramente espacial

Considerando o objetivo de identificar padrões demográficos persistentes ao longo de um período de 14 anos, optou-se pela análise puramente espacial do método de varredura de Kulldorff. Essa abordagem permite identificar aglomerados espaciais estáveis, reduzindo a influência de flutuações temporais aleatórias e evidenciando desigualdades estruturais associadas à composição demográfica da população.

A análise espaço-temporal não foi priorizada, uma vez que o foco do estudo não se concentra na identificação de períodos específicos de risco elevado, mas sim na caracterização de padrões espaciais consolidados ao longo do tempo.

Foi realizada análise espacial pura utilizando o modelo Poisson no SaTScan. A população sob risco foi definida a partir da população residente com idade ≥ 20 anos do Censo Demográfico de 2022, adotada como denominador fixo para todo o período de 2010 a 2023, uma vez que o objetivo do estudo foi exclusivamente a detecção de clusters espaciais, não sendo realizada análise temporal. O uso de população censitária observada, em detrimento de estimativas intercensitárias, visou reduzir viés associado a projeções populacionais.

Os tamanhos máximos das janelas circulares foram fixados em 5%, 10% e 15%

da população em risco, com o objetivo de identificar conglomerados espaciais. A significância estatística de cada conglomerado foi avaliada por meio de 999 simulações de Monte Carlo. As figuras a seguir apresentam os conglomerados de alto e baixo risco para a mortalidade nos municípios do Amazonas, no período de 2010 a 2023.

Os quadros correspondentes reúnem as principais informações dos conglomerados identificados, a saber: (i) ano analisado; (ii) tamanho da janela, definido pela proporção da população em risco considerada; (iii) tipo do conglomerado, classificado como alto ou baixo risco e categorizado em primário, secundário ou terciário; (iv) raio, expresso em quilômetros a partir do centroide do conglomerado; (v) razão observado/esperado, correspondente ao número de casos observados dividido pelo número esperado; (vi) razão de verossimilhança logarítmica (LLR), utilizada para avaliar a força estatística do conglomerado e (vii) p-valor, referente à significância estatística;

Considerando o longo período analisado, optou-se por uma apresentação mais compacta dos resultados. Assim, nos anos e configurações de janela em que não foram identificados conglomerados estatisticamente significativos pelo SaTScan, os campos correspondentes permanecem em branco nas tabelas. Ressalta-se que, ao longo do período estudado, para o alto risco não foram detectados conglomerados ou não eram estatisticamente significativos.

Tabela 18: Conglomerados espaciais de baixo risco detectados segundo ano, tamanho da janela e nível para o câncer de mama

Ano	Janela (%)	Nível	Raio (km)	Obs./Esp.	LRR	*p*-valor
2010	5	1	101.16	0	4,72	0,015
	10	1	709.80	0.11	6,32	0,042
	10	2	245.22	0	6,24	0,047
	15	1	711.31	0.077	1,03	0,001
	15	2	881.34	0.095	7,72	0,010
2011	5	1	417.67	0	5,56	0,016
	5	2	437.54	0	5,31	0,044
	10	1	832.12	0	1,15	0,001
	10	2	555.80	0	8,52	0,003
	15	1	1015.86	0	1,71	0,001
	15	2	555.80	0	8,52	0,003
2012	5	1	328.73	0	6,94	0,009
	5	2	636.61	0	6,87	0,015
	5	3	554.99	0	6,28	0,041
	10	1	639.20	0	1,35	0,001
	10	2	890.48	0.15	8,13	0,009

Continua na próxima página

Tabela 18 – continuação

Ano	Janela (%)	Nível	Raio (km)	Obs./Esp.	LRR	*p*-valor
	15	1	1093.83	0.048	1,84	0,001
2013	5	1	482.84	0	7,54	0,003
	10	1	619.71	0	1,30	0,001
	10	2	570.08	0.16	6,97	0,026
	15	1	1015.55	0.098	1,50	0,001
	15	2	570.08	0.16	6,97	0,032
2014	5	1	636.61	0	6,62	0,007
	10	1	781.59	0.076	1,03	0,002
	15	1	805.19	0.062	1,32	0,001
2015	5	1	417.67	0	7,68	0,002
	5	2	437.54	0	7,34	0,005
	5	3	718.10	0	6,55	0,027
	10	1	784.77	0	1,50	0,001
	10	2	570.08	0.077	9,91	0,001
	10	3	718.10	0	6,55	0,037
	15	1	1141.83	0.096	1,54	0,001
2016	5	1	407.46 km	0	6,93	0,018
	5	2	271.12 km	0	6,41	0,033
	10	1	570.08 km	0.081	9,31	0,003
	10	2	804.32 km	0.15	8,12	0,006
	15	1	1184.71 km	0.19	1,16	0,001
	15	2	609.65 km	0.22	1,09	0,001
2017	5	1	499.45 km	0	8,02	0,004
	5	2	631.72 km	0	7,64	0,010
	5	3	144.61 km	0	7,4	0,011
	10	1	690.47 km	0	1,05	0,003
	10	2	144.61 km	0	7,4	0,022
	15	1	929.89 km	0.17	1,31	0,001
	15	2	609.65 km	0.38	6,63	0,049
2018	5	1	522.89 km	0	8,48	0,002
	10	1	800.13 km	0.11	1,23	0,001
	10	2	254.84 km	0.10	6,49	0,039
	15	1	901.77 km	0.11	1,89	0,001
	15	2	254.84 km	0.10	6,49	0,043
2019	5	1	467.58 km	0	8,75	0,0030
	10	1	832.12 km	0.16	1,07	0,0020
	15	1	832.12 km	0.16	1,1	0,0020
2020	5	1	221.60 km	0	9,96	0,001
	5	2	501.11 km	0	9,33	0,002
	5	3	718.10 km	0	9,1	0,002
	10	1	789.89 km	0.20	1,05	0,001
	10	2	221.60 km	0	9,96	0,001

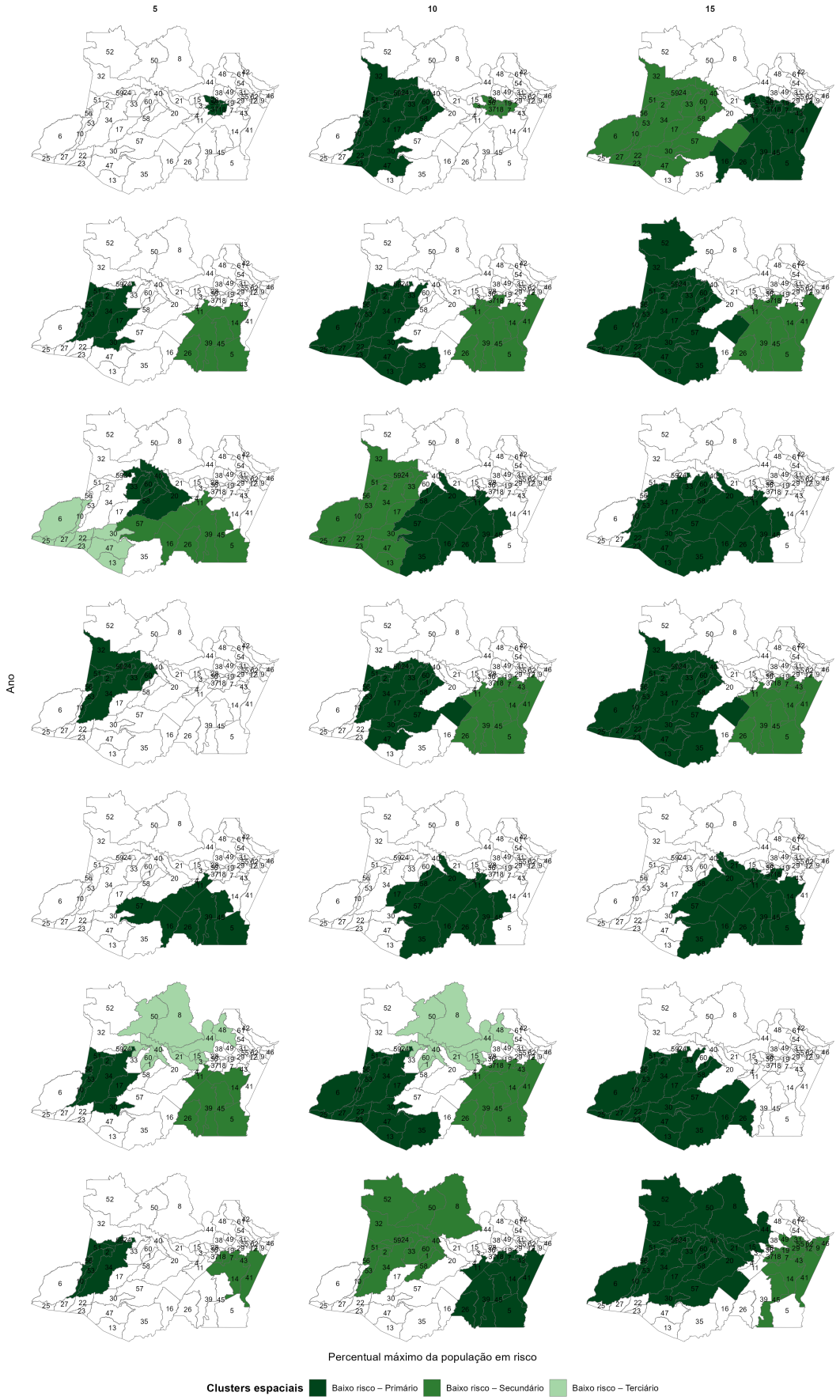
Continua na próxima página

Tabela 18 – continuação

Ano	Janela (%)	Nível	Raio (km)	Obs./Esp.	LRR	*p*-valor
	10	3	804.32 km	0.21	9,09	0,002
	15	1	805.19 km	0.28	1,27	0,001
	15	2	221.60 km	0	9,96	0,001
2021	5	1	631.72 km	0	8,59	0,003
	5	2	107.52 km	0	7,31	0,008
	10	1	879.96 km	0.17	1	0,002
	10	2	107.52 km	0	7,31	0,014
	15	1	1141.83 km	0.20	1,3	0,001
	15	2	443.22 km	0.32	7,47	0,012
2022	5	1	690.47 km	0	1,02	0,003
	10	1	749.32 km	0.068	1,16	0,002
	15	1	1184.71 km	0.22	1,56	0,001
2023	5	1	482.46 km	0	9,15	0,002
	5	2	690.47 km	0.090	7,94	0,008
	5	3	311.61 km	0	7,48	0,014
	10	1	865.90 km	0.16	1,11	0,001
	10	2	555.80 km	0.13	9,24	0,003
	10	3	690.47 km	0.090	7,94	0,009
	15	1	1015.86 km	0.30	1,42	0,001
	15	2	555.80 km	0.13	9,24	0,003

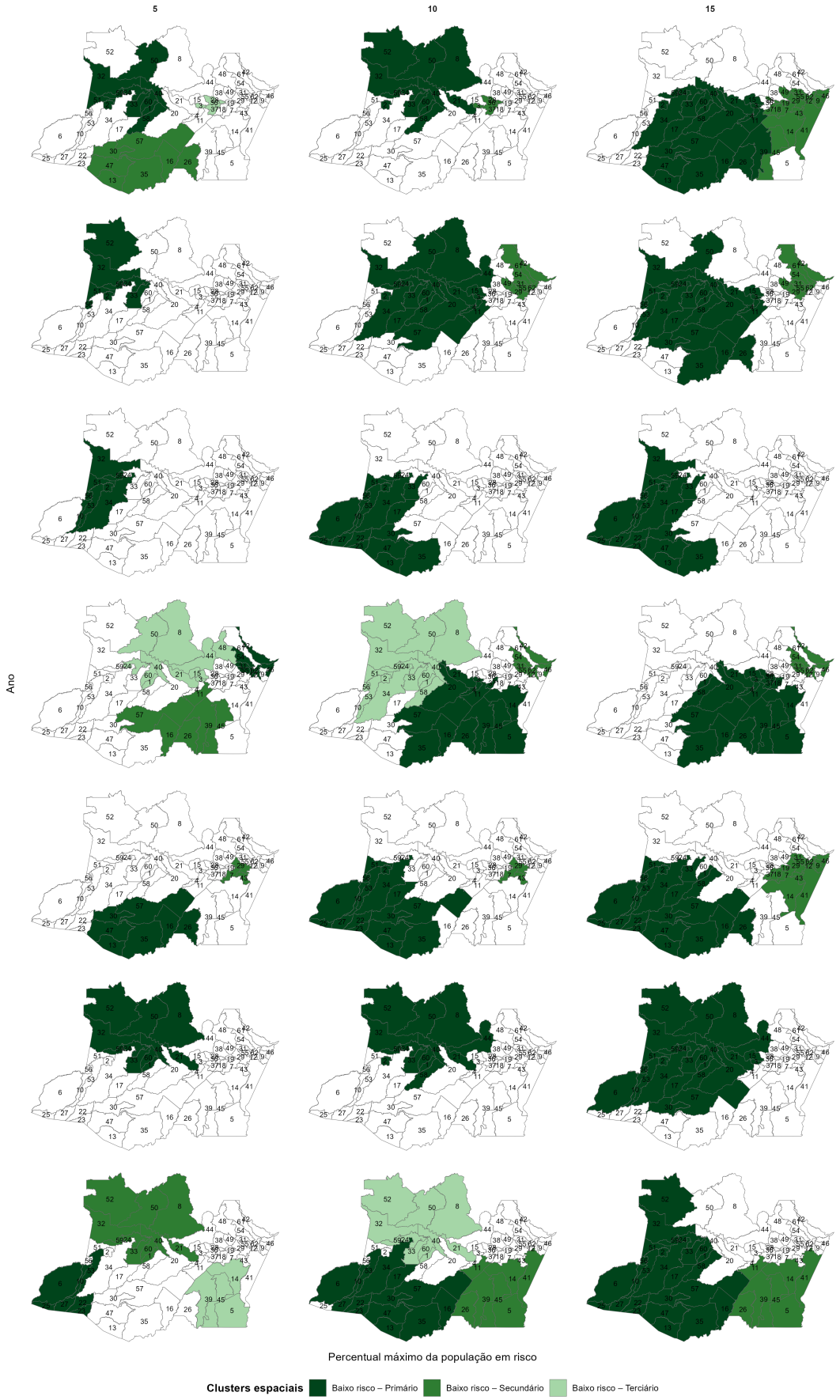
Nota: O nível do conglomerado segue a hierarquia definida pelo SaTScan, em que 1 = Primário, 2 = Secundário e 3 = Terciário.

Figura 16: Conglomerados espaciais de alto e baixo risco da mortalidade por câncer de mama, considerando 5% a 15% da população sob risco – Amazonas, 2010–2016.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Figura 17: Conglomerados espaciais de alto e baixo risco da mortalidade por câncer de mama, considerando 5% a 15% da população sob risco – Amazonas, 2017–2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

As Figuras 16 e 17 apresentam a distribuição espacial dos conglomerados de alto e baixo risco da mortalidade por câncer de mama no estado do Amazonas, considerando diferentes tamanhos de janela populacional (5% a 15%) ao longo do período de 2010 a 2023. A Tabela 18 apresenta as estatísticas dos conglomerados identificados.

Observamos quase em sua totalidade conglomerados de baixo risco ao longo do período analisado, enquanto conglomerados de alto risco foram raramente detectados ou ausentes na maioria dos anos e tamanhos de janela considerados. A ausência de conglomerados de alto risco em determinados anos sugere que não foram identificados concentração excessiva de óbitos.

Ao longo do período analisado, verificou-se a permanência desses conglomerados em diferentes tamanhos de janela, indicando padrão espacial das áreas de menor risco. A extensão territorial desses agrupamentos variou entre os anos, chegando a abranger regiões amplas do estado com as janelas de 15%. Esses resultados podem estar associados à baixa densidade populacional.

Por seguinte, observa-se que, entre 2010 e 2016, os conglomerados de baixo risco apresentaram maior dispersão espacial, enquanto no período de 2017 a 2023 os agrupamentos apresentam para os tamanhos de janelas e períodos permanência dos conglomerados.

De modo geral, os resultados indicam que a mortalidade por câncer de mama no Amazonas apresenta distribuição espacial relativamente homogênea, sem evidência de clusters de alto risco. Esse comportamento corrobora com os fatores associados ao perfil epidemiológico da doença e às características demográficas da população feminina no estado.

Tabela 19: Conglomerados espaciais de alto risco detectados segundo ano, tamanho da janela e nível para o câncer do colo do útero

Ano	Janela (%)	Nível	Raio (km)	Obs./Esp.	LRR	*p*-valor
2010	5	1	468.16 km	0	6,94	0,016
	5	2	409.29 km	0.10	6,57	0,025
	10	1	619.71 km	0.23	7,99	0,008
	15	1	968.18 km	0.25	1,25	0,001
2011	5	1	466.00 km	0	1,05	0,002
	5	2	550.95 km	0	5,77	0,049
	10	1	528.07 km	0	1,14	0,003
	10	2	570.08 km	0.27	7,49	0,012

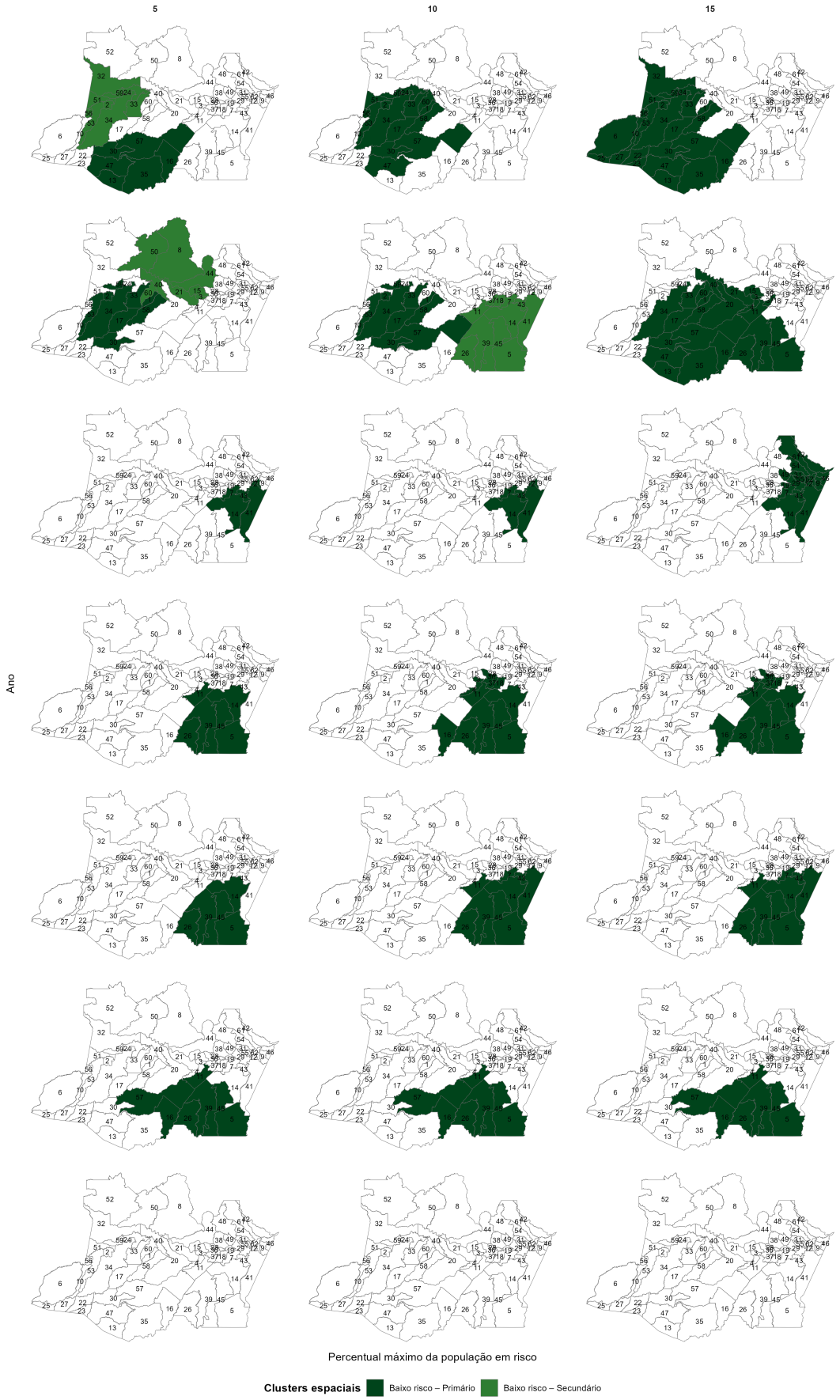
Continua na próxima página

Tabela 19 – continuação

Ano	Janela (%)	Nível	Raio (km)	Obs./Esp.	LRR	*p*-valor
	15	1	1047.15 km	0.28	1,26	0,001
2012	5	1	289.15 km	0	1,02	0,002
	10	1	289.15 km	0	1,02	0,002
	15	1	466.77 km	0.28	1,28	0,001
2013	5	1	437.54 km	0.15	7,5	0,013
	10	1	581.21 km	0.24	1,15	0,001
	15	1	581.21 km	0.24	1,15	0,001
2014	5	1	421.47 km	0.16	6,93	0,022
	10	1	570.08 km	0.23	9,56	0,001
	15	1	570.08 km	0.23	9,56	0,001
2015	5	1	636.61 km	0.15	7,67	0,007
	10	1	636.61 km	0.15	7,67	0,011
	15	1	636.61 km	0.15	7,67	0,011
2018	5	1	630.88 km	0.15	7,78	0,008
	10	1	865.90 km	0.18	1,16	0,001
	15	1	865.90 km	0.18	1,16	0,001
2019	5	1	339.36 km	0.099	6,94	0,019
	10	1	804.32 km	0.30	7,37	0,014
	10	2	339.36 km	0.099	6,94	0,025
	15	1	804.32 km	0.30	7,37	0,015
	15	2	339.36 km	0.099	6,94	0,026
2020	5	1	311.61 km	0	8,2	0,005
	10	1	570.08 km	0.26	9,82	0,001
	15	1	570.08 km	0.26	9,82	0,001
2021	5	1	321.38 km	0.23	5,97	0,04
	10	1	711.31 km	0.27	1,04	0,002
	15	1	711.31 km	0.27	1,04	0,002
2022	10	1	711.31 km	0.41	6,59	0,038
	15	1	901.77 km	0.46	8,34	0,007

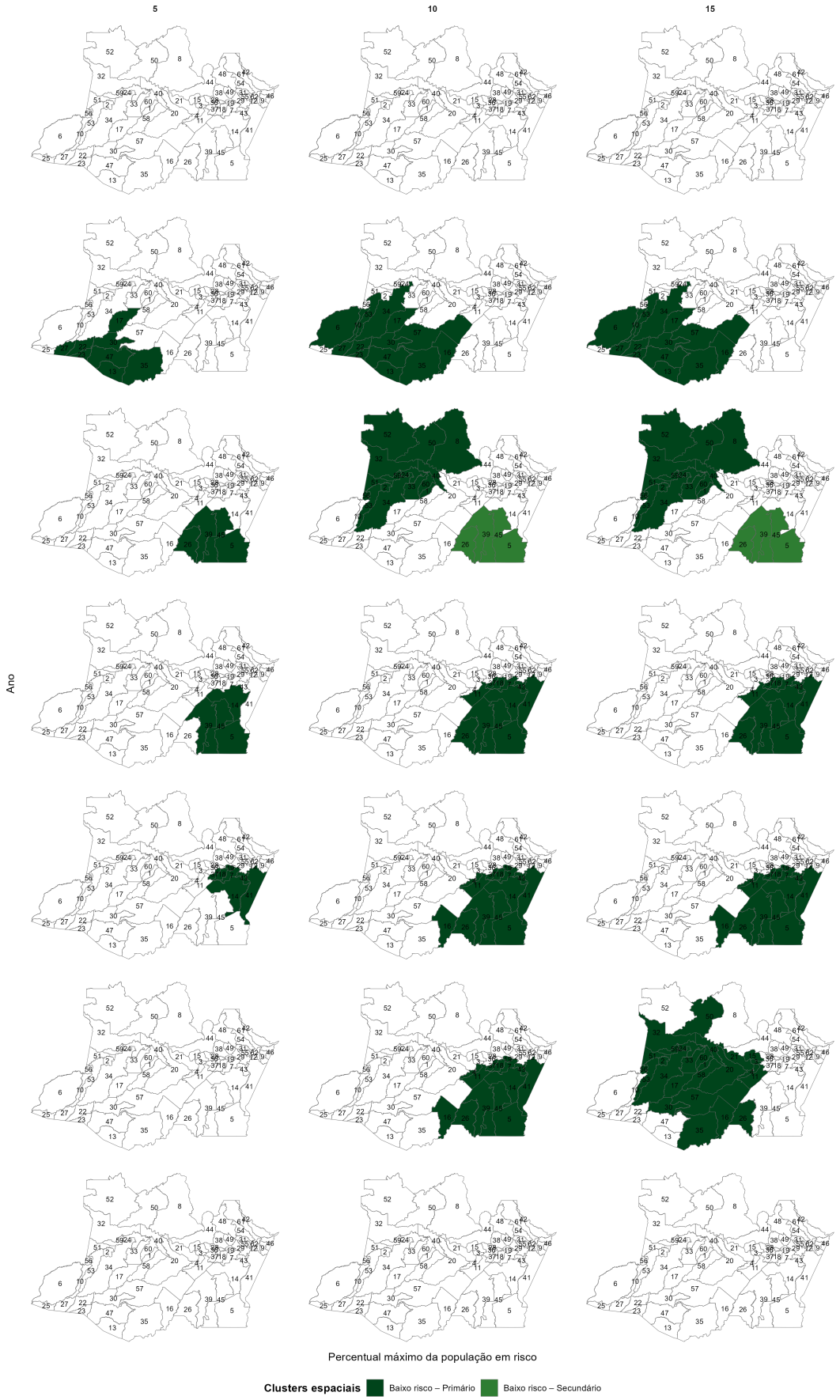
Nota: O nível do conglomerado segue a hierarquia definida pelo SaTScan, em que 1 = Primário, 2 = Secundário e 3 = Terciário.

Figura 18: Conglomerados espaciais de alto e baixo risco da mortalidade por câncer do colo do útero, considerando 5% a 15% da população sob risco – Amazonas, 2010–2016.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Figura 19: Conglomerados espaciais de alto e baixo risco da mortalidade por câncer do colo do útero, considerando 5% a 15% da população sob risco – Amazonas, 2017–2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

As Figuras 18 e 19 apresentam a distribuição espacial dos conglomerados de alto e baixo risco da mortalidade por câncer do colo do útero no estado do Amazonas, considerando diferentes tamanhos de janela populacional (5% a 15%) ao longo do período de 2010 a 2023. A Tabela 19 apresenta as estatísticas dos conglomerados identificados.

Observamos, de forma semelhante ao padrão identificado para o câncer de mama, quase em sua totalidade conglomerados classificados como de baixo risco ao longo do período analisado. Em contrapartida, conglomerados de alto risco foram raramente detectados ou ausentes na maioria dos anos e tamanhos de janela considerados. A ausência de conglomerados de alto risco no período sugere que não foram identificados concentração excessiva de óbitos.

Ao longo do período analisado, verificou-se a permanência de conglomerados classificados como de baixo risco, com exceção dos conglomerados de baixo risco terciário, que não foram identificados em diferentes tamanhos de janela. Indicando padrão espacial das áreas de menor risco.

Além disso, observou-se que, em diversos anos do período entre 2010 e 2023, independentemente do tamanho da janela de análise, foram identificados apenas conglomerados primários de baixo risco, sem detecção de conglomerados secundários ou terciários. Ressalta-se ainda a ocorrência de três anos nos quais não foram identificados conglomerados estatisticamente significativos, sugerindo ausência de concentração espacial relevante nesses períodos.

De modo geral, os resultados indicam que a mortalidade por câncer do colo do útero no Amazonas apresenta predominância de conglomerados classificados como de baixo risco ao longo do período analisado, com ocorrência pouco frequente ou ausência de conglomerados de alto risco. Diferentemente do observado para o câncer de mama, no qual houve variações ao longo dos anos, o câncer do colo do útero apresentou maior estabilidade espacial. Esse comportamento sugere a persistência de áreas com características semelhantes de risco, permitindo a identificação de territórios prioritários para o fortalecimento e direcionamento das ações de prevenção, rastreamento e diagnóstico precoce da doença.

5 CONCLUSÕES

Neste trabalho, foram realizadas três abordagens analíticas complementares: análise descritiva, modelagem estatística e análise espacial. Em conjunto, os resultados permitiram evidenciar a heterogeneidade das taxas de mortalidade, quantificar os riscos associados e identificar a presença de conglomerados espaciais no estado do Amazonas.

Na análise descritiva, verificou-se que, ao longo do período estudado, as taxas de mortalidade padronizadas por faixa etária apresentaram elevada heterogeneidade, com valores anuais que chegam a ser até quatro vezes superiores à média nacional. De modo geral, as taxas de mortalidade por câncer do colo do útero foram superiores às de câncer de mama em praticamente todo o período analisado. Entretanto, observou-se uma convergência entre essas taxas nos anos de 2022 e 2023, decorrente do aumento das taxas de câncer de mama e da redução das taxas de câncer do colo do útero, o que pode indicar uma possível sobrecarga do sistema de saúde no período.

As análises estratificadas por faixa etária mostraram que, embora ocorram óbitos entre mulheres de 20 a 29 anos, o aumento mais expressivo das taxas para ambas as neoplasias concentra-se nas faixas etárias de 40 a 49 anos. Esse achado está em consonância com as decisões recentes do Ministério da Saúde do Brasil, que atualizou as diretrizes de rastreamento, permitindo a realização de mamografia pelo Sistema Único de Saúde (SUS) em mulheres a partir dos 40 anos, mesmo na ausência de sintomas.

Destaca-se, ainda, o comportamento observado na faixa etária de 70 a 79 anos, que apresentou taxas de mortalidade semelhantes às observadas entre mulheres de 20 a 29 anos para ambos os tipos de câncer, indicando um padrão etário não linear da mortalidade e reforçando a importância de análises estratificadas por idade.

Nos modelos ajustados, tanto no modelo geral — que inclui conjuntamente os cânceres de mama e do colo do útero — quanto nos modelos estimados separadamente

para cada neoplasia, observou-se redução das taxas de óbito ao longo dos anos analisados. Em contrapartida, todas as faixas etárias apresentaram aumento da taxa média de óbitos, indicando maior vulnerabilidade com o avanço da idade.

No modelo específico para câncer de mama, a taxa de envelhecimento populacional apresentou associação estatisticamente significativa com o risco de mortalidade, enquanto a razão de dependência não foi significativa, não sendo, portanto, incluída no modelo otimizado. Já no modelo para câncer do colo do útero, todas as faixas etárias foram estatisticamente significativas e apresentaram aumento do risco com magnitude semelhante à observada no câncer de mama; nesse caso, a razão de dependência mostrou-se estatisticamente significativa, indicando influência da estrutura etária sobre a mortalidade por essa neoplasia.

Em ambos os modelos, a covariável que representa a população em situação de extrema pobreza apresentou associação estatisticamente significativa, assim como a densidade demográfica. Esses resultados sugerem que fatores relacionados à estrutura etária, às condições socioeconômicas e às características do território exercem influência relevante sobre as taxas de mortalidade por câncer de mama e câncer do colo do útero no estado do Amazonas.

Apesar da identificação de aglomerados espaciais apenas de baixo risco em todos os períodos e localidades analisados e da ausência de aglomerados em alguns anos, o estado do Amazonas permanece inserido em uma das regiões com as maiores taxas de mortalidade por câncer de mama do Brasil, a Região Norte. Esse resultado indica que, mesmo na ausência de conglomerados de alto risco, a mortalidade permanece elevada, evidenciando que não encontramos no Amazonas nenhuma área que tenha o risco de vir a ser um caso maior que outras, mas amplamente distribuídos indicando desigualdades regionais.

Os resultados permitem identificar áreas que demandam a manutenção e a qualificação das estratégias de prevenção, bem como a ampliação da cobertura do rastreamento e o fortalecimento do acesso ao diagnóstico precoce e ao tratamento oportuno. No caso do câncer de mama, observou-se maior heterogeneidade espacial ao longo dos anos, com variação na distribuição dos risco relativo entre os municípios e períodos analisados, o que indica uma dinâmica territorial mais instável da mortalidade por essa neoplasia. Entretanto, a mortalidade por câncer do colo do útero apresentou um padrão espacial constante, com concentração ao longo do período estudado nas

mesmas localidades.

De forma geral, concluímos que a mortalidade por câncer no Amazonas não se distribui de maneira aleatória, sendo fortemente influenciada por questões estruturais e de faixa etária. A identificação de áreas prioritárias, oferece subsídios relevantes para o planejamento de ações voltadas ao fortalecimento da prevenção, ampliação da cobertura de rastreamento, melhoria do acesso ao diagnóstico precoce e ao tratamento oportuno. Assim, este estudo contribui para o aprofundamento do conhecimento sobre a dinâmica da mortalidade por câncer no estado e fornece evidências que podem apoiar a tomada de decisão em saúde pública, com vistas à redução das iniquidades e da mortalidade evitável.

6 APÊNDICE

6.1 Apêndice A - Tabelas Complementares

Tabela A1: Classificação dos capítulos da CID-10

Capítulo	Descrição	Códigos da CID-10
I	Algumas doenças infecciosas e parasitárias	A00–B99
II	Neoplasmas [tumores]	C00–D48
III	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários	D50–D89
IV	Doenças endócrinas, nutricionais e metabólicas	E00–E90
V	Transtornos mentais e comportamentais	F00–F99
VI	Doenças do sistema nervoso	G00–G99
VII	Doenças do olho e anexos	H00–H59
VIII	Doenças do ouvido e da apófise mastóide	H60–H95
IX	Doenças do aparelho circulatório	I00–I99
X	Doenças do aparelho respiratório	J00–J99
XI	Doenças do aparelho digestivo	K00–K93
XII	Doenças da pele e do tecido subcutâneo	L00–L99
XIII	Doenças do sistema osteomuscular e do tecido conjuntivo	M00–M99
XIV	Doenças do aparelho geniturinário	N00–N99
XV	Gravidez, parto e puerpério	O00–O99
XVI	Algumas afecções originadas no período perinatal	P00–P96
XVII	Malformações congênitas, deformidades e anomalias cromossômicas	Q00–Q99
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	R00–R99
XIX	Lesões, envenenamentos e algumas outras consequências de causas externas	S00–T98
XX	Causas externas de morbidade e de mortalidade	V01–Y98
XXI	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde	Z00–Z99

Tabela A2: População padrão mundial da OMS (2000–2025) repadronizada para faixas etárias acima de 20 anos.

Faixa etária	Peso original	Peso ajustado (>20)
20–24	0,0822	0,1256
25–29	0,0793	0,1212
30–34	0,0761	0,1163
35–39	0,0715	0,1093
40–44	0,0659	0,1007
45–49	0,0604	0,0923
50–54	0,0537	0,0821
55–59	0,0455	0,0696
60–64	0,0372	0,0569
65–69	0,0296	0,0452
70–74	0,0221	0,0338
75–79	0,0152	0,0232
80+	0,0155	0,0237
Total	**0,6542**	**1,0000**

Tabela A3: Regionalização do Amazonas: Macrorregiões, Regiões de Saúde e Municípios

Macrorregião	Região de Saúde	Município de Referência	Município
Macrorregião Central 25 municípios 3 regiões de saúde	13001 Manaus, Entorno e Alto Rio Negro	Manaus	Autazes
			Barcelos
			Careiro
			Careiro da Várzea
			Iranduba
			Manaquiri
			Manaus
			Nova Olinda do Norte
			Presidente Figueiredo
			Rio Preto da Eva
			Santa Isabel do Rio Negro
			São Gabriel da Cachoeira
	13002 Rio Negro e Solimões	Manacapuru	Anamã
			Anori
			Beruri
			Caapiranga
			Coari
			Codajás
			Manacapuru
			Novo Airão
	13006 Purus	Lábrea	Boca do Acre
			Canutama
			Lábrea
			Pauini
			Tapauá
Macrorregião Oeste 21 municípios 3 regiões de saúde	13007 Juruá	Eirunepé	Carauari
			Eirunepé
			Envira
			Guajará
			Ipixuna
			Itamarati
	13008 Triângulo	Tefé	Alvarães
			Japurá
			Juruá
			Maraã
			Tefé
			Uarini
	13009 Alto solimões	Tabatinga	Amaturá
			Atalaia do Norte
			Benjamim Constant
			Fonte Boa
			Jutaí
			Santo Antônio do Içá
			São Paulo de Olivença

			Tabatinga
			Tonantins
	13003 Rio Madeira	Humaitá	Apuí
			Borba
			Humaitá
			Manicoré
			Novo Aripuanã
	13004 Médio Amazonas	Itacoatiara	Itacoatiara
			Itapiranga
			São Sebastião do Uatumã
			Silves
			Urucará
			Urucurituba
Macrorregião Leste 16 municípios 3 regiões de saúde	13005 Baixo Amazonas	Parintins	Barreirinha
			Boa Vista dos Ramos
			Maués
			Nhamundá
			Parintins

Tabela A4: Municípios do Amazonas segundo macrorregião, código IBGE e identificador

Macrorregião	Município	Código IBGE	ID
1302 OESTE	Alvarães	1300029	1
	Amaturá	1300060	2
	Atalaia do Norte	1300201	6
	Benjamin Constant	1300607	10
	Carauari	1301001	17
	Eirunepé	1301407	22
	Envira	1301506	23
	Fonte Boa	1301605	24
	Guajará	1301654	25
	Ipixuna	1301803	27
	Itamarati	1301951	30
	Japurá	1302108	32
	Juruá	1302207	33
	Jutaí	1302306	34
	Maraã	1302801	40
	Santo Antônio do Içá	1303700	51
	São Paulo de Olivença	1303908	53
	Tabatinga	1304062	56
	Tefé	1304203	58
	Tonantins	1304237	59
	Uarini	1304260	60
1303 LESTE	Apuí	1300144	5
	Barreirinha	1300508	9
	Boa Vista do Ramos	1300680	12
	Borba	1300805	14
	Humaitá	1301704	26
	Itacoatiara	1301902	29
	Itapiranga	1302009	31
	Manicoré	1302702	39
	Maués	1302900	41
	Nhamundá	1303007	42
	Novo Aripuanã	1303304	45
	Parintins	1303403	46
	São Sebastião do Uatumã	1303957	54
	Silves	1304005	55
	Urucará	1304302	61
	Urucurituba	1304401	62
	Anamã	1300086	3
	Anori	1300102	4
	Autazes	1300300	7
	Barcelos	1300409	8
	Beruri	1300631	11
	Boca do Acre	1300706	13
	Caapiranga	1300839	15
	Canutama	1300904	16
	Careiro	1301100	18
	Careiro da Várzea	1301159	19
	Coari	1301209	20
	Codajás	1301308	21

Tabela A5: Diagnóstico de multicolinearidade (VIF) para o modelo ótimo do Modelo Geral.

Baixa correlação					
Termo	VIF	IC 95% (VIF)	VIF ajustado	Tolerância	IC 95% (Tol.)
CID	1,06	[1,02; 1,16]	1,03	0,94	[0,86; 0,98]
ns(ano, df=3)	1,04	[1,01; 1,16]	1,02	0,96	[0,86; 0,99]
Faixa etária	1,06	[1,02; 1,16]	1,00	0,94	[0,86; 0,98]
Densidade	3,27	[2,99; 3,57]	1,81	0,31	[0,28; 0,33]
Taxa de envelhecimento	1,18	[1,12; 1,27]	1,09	0,85	[0,79; 0,89]
Extrema pobreza	4,52	[4,12; 4,96]	2,13	0,22	[0,20; 0,24]
Taxa de fecundidade	2,55	[2,35; 2,78]	1,60	0,39	[0,36; 0,43]
Gini	2,37	[2,19; 2,58]	1,54	0,42	[0,39; 0,46]
Correlação moderada					
Termo	VIF	IC 95% (VIF)	VIF ajustado	Tolerância	IC 95% (Tol.)
Razão de dependência	5,32	[4,84; 5,85]	2,31	0,19	[0,17; 0,21]

Tabela A6: Diagnóstico de multicolinearidade (VIF) para o om erros-padrão robustos clusterizados para mortalidade por câncer de mama (C50).

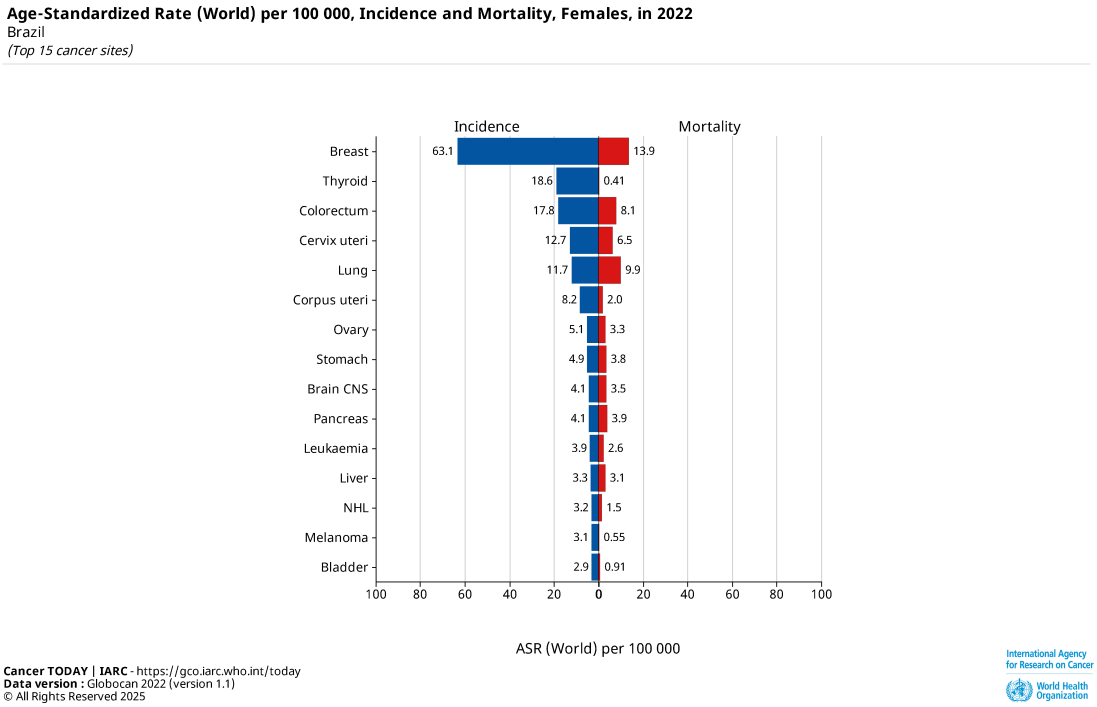
Baixa correlação					
Termo	VIF	IC 95% (VIF)	VIF ajustado	Tolerância	IC 95% (Tol.)
ns(ano, df=3)	1,08	[1,02; 1,29]	1,04	0,92	[0,77; 0,98]
Faixa etária	1,15	[1,07; 1,33]	1,01	0,87	[0,75; 0,94]
Taxa de crescimento	1,10	[1,03; 1,29]	1,05	0,91	[0,77; 0,97]
Densidade	4,73	[4,05; 5,57]	2,18	0,21	[0,18; 0,25]
Taxa de envelhecimento	1,19	[1,10; 1,36]	1,09	0,84	[0,73; 0,91]
Extrema pobreza	4,82	[4,12; 5,68]	2,20	0,21	[0,18; 0,24]
Taxa de fecundidade	1,72	[1,53; 1,98]	1,31	0,58	[0,51; 0,65]
Gini	1,93	[1,71; 2,23]	1,39	0,52	[0,45; 0,59]
Razão mama	4,63	[3,96; 5,45]	2,15	0,22	[0,18; 0,25]

Tabela A7: Diagnóstico de multicolinearidade (VIF) para o om erros-padrão robustos clusterizados para mortalidade por câncer do colo do útero (C53).

Termo	VIF	IC 95% (VIF)	VIF ajustado	Tolerância	IC 95% (Tol.)
ns(ano, df=3)	1,04	[1,01; 1,21]	1,02	0,96	[0,83; 0,99]
Faixa etária	1,08	[1,03; 1,20]	1,01	0,92	[0,83; 0,97]
Densidade demográfica	2,83	[2,56; 3,15]	1,68	0,35	[0,32; 0,39]
Razão de dependência	4,27	[3,83; 4,79]	2,07	0,23	[0,21; 0,26]
Taxa de envelhecimento	1,22	[1,14; 1,33]	1,10	0,82	[0,75; 0,87]
Extremamente pobres	4,00	[3,59; 4,48]	2,00	0,25	[0,22; 0,28]
Taxa de fecundidade	2,47	[2,24; 2,74]	1,57	0,40	[0,36; 0,45]
Índice de Gini	2,47	[2,25; 2,75]	1,57	0,40	[0,36; 0,45]
Razão colo do útero	1,17	[1,10; 1,28]	1,08	0,86	[0,78; 0,91]

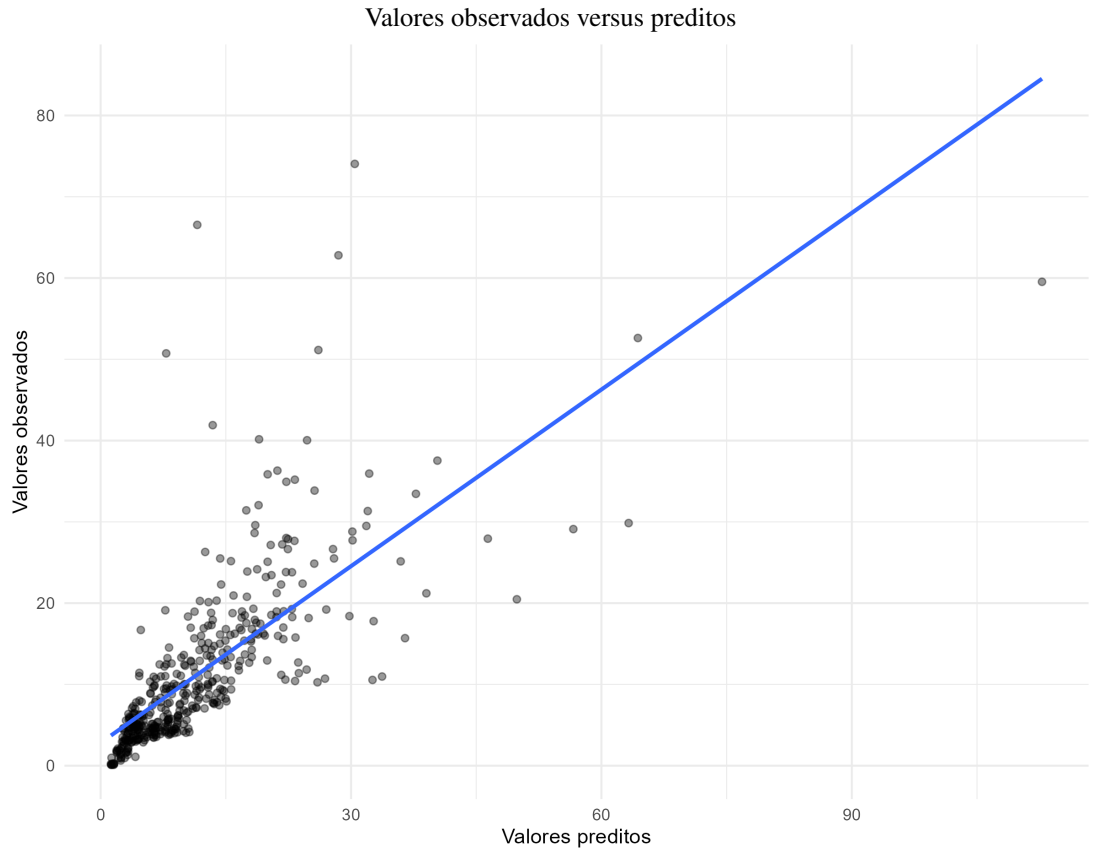
6.2 Apêndice B - Figuras Complementares

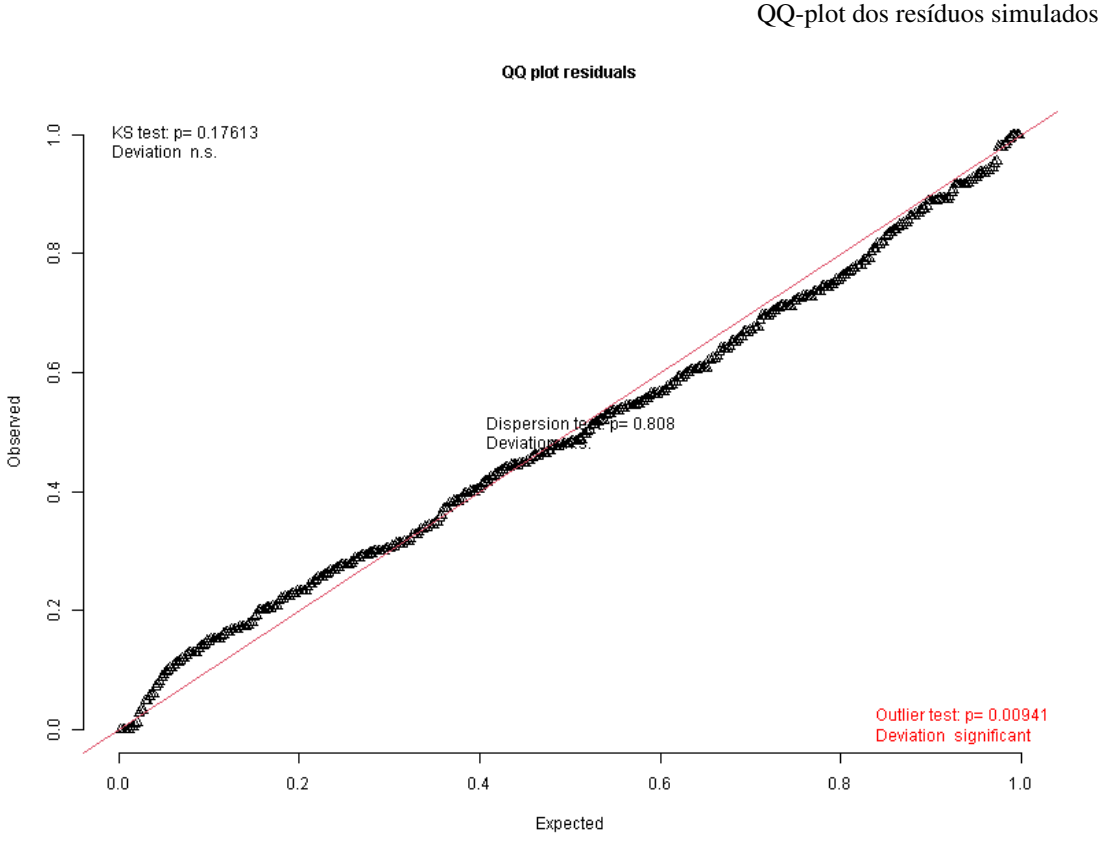
Figura 20: Taxa de incidência e mortalidade padronizada por idade (mundial) por 100 mil/mulheres – Mundo, 2022.



Fonte: WHO. Disponível em: <https://gco.iarc.who.int/today/en/dataviz/bars?mode=cancer&key=asr&group_populations=1&types=0_1&sexes=2&sort_by=value1&populations=76&multiple_populations=0&values_position=out&cancers_h=39>

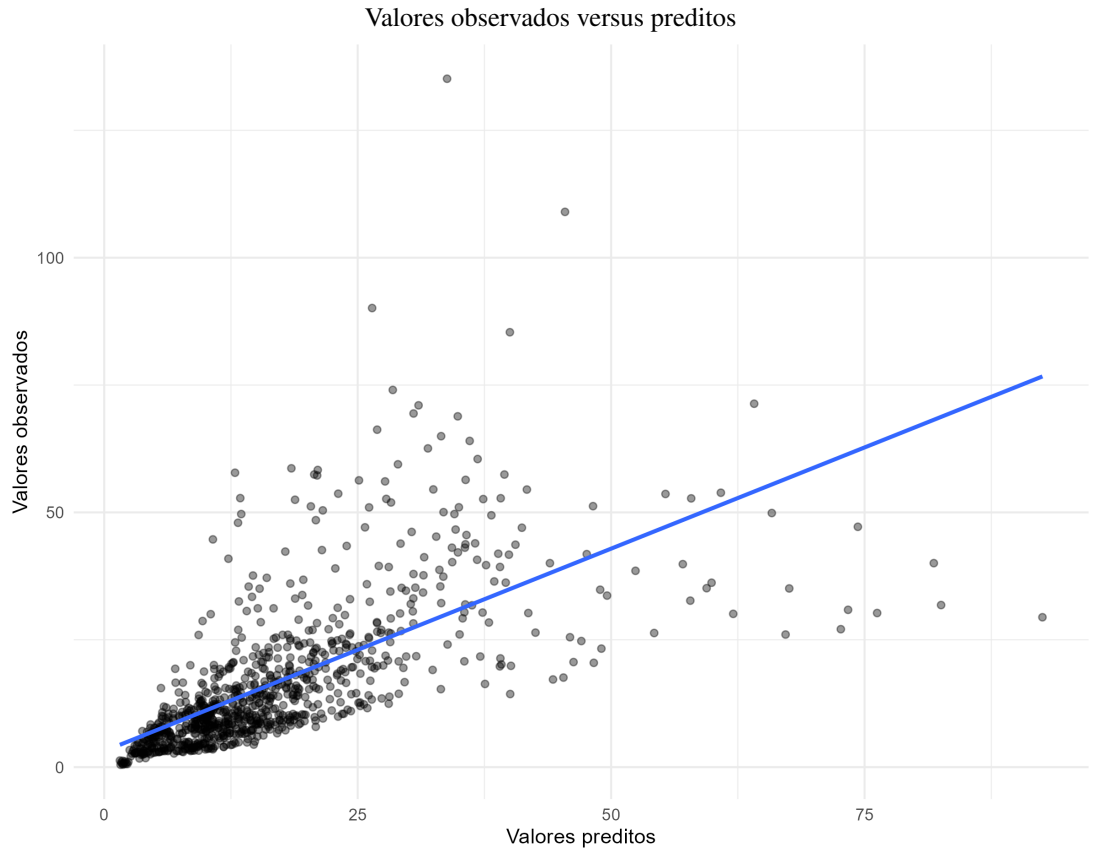
Figura 21: Diagnóstico do ajuste do modelo Gamma com ligação log para mortalidade por câncer de mama – Amazonas, 2010 a 2023.

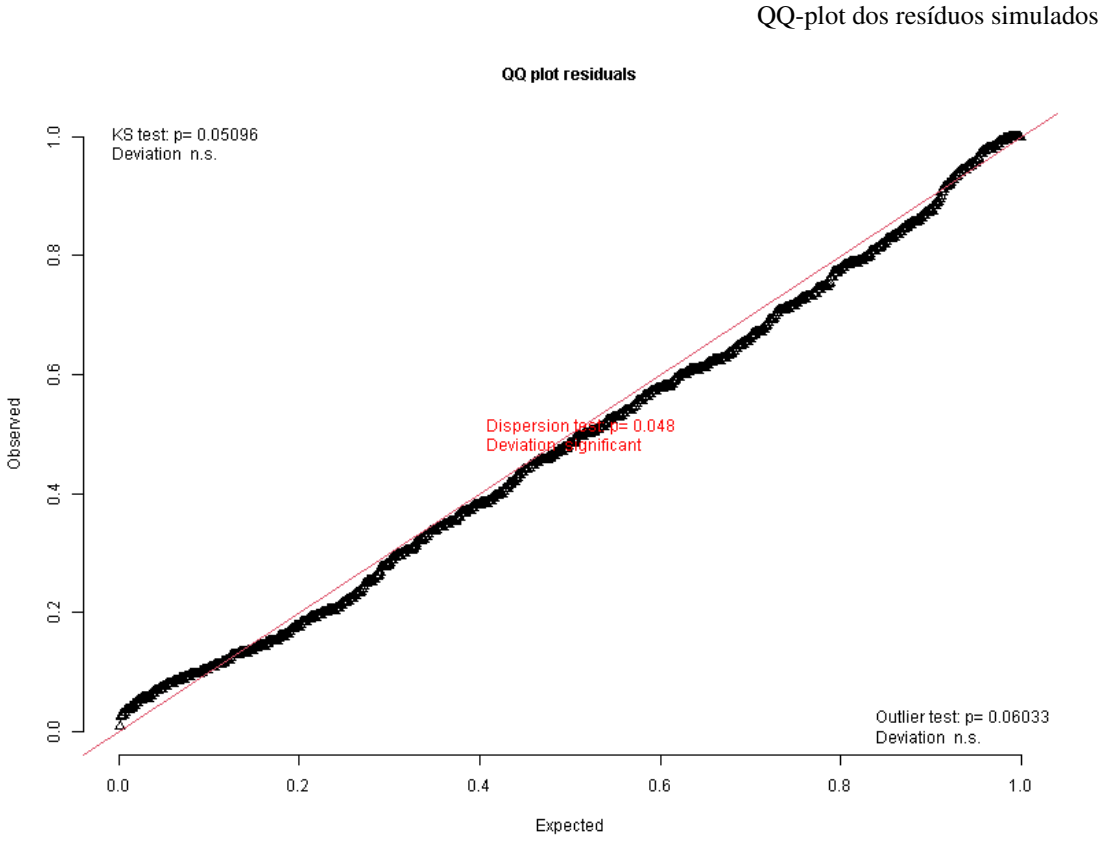




Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

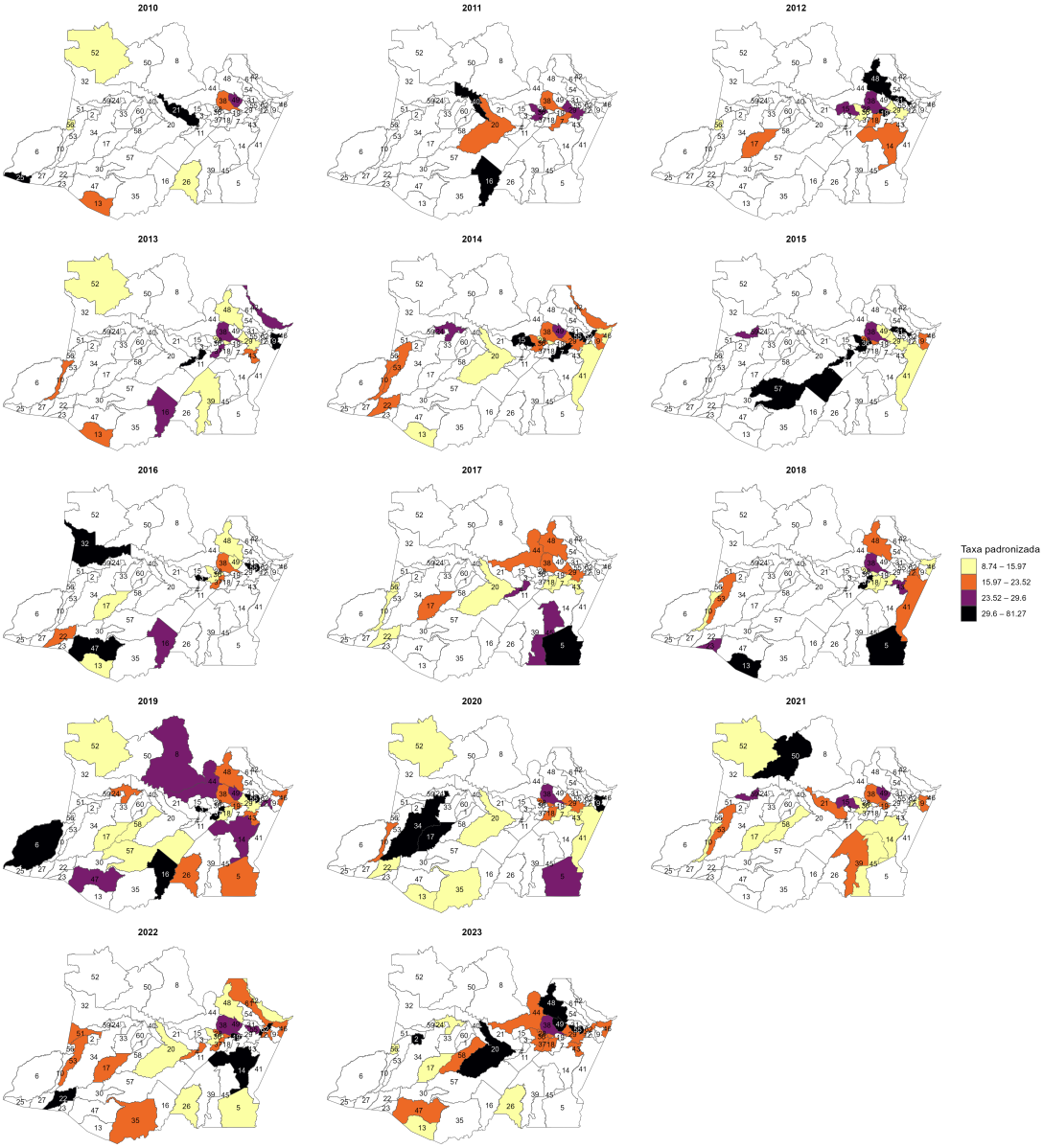
Figura 22: Diagnóstico do ajuste do modelo Gamma com ligação log para mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023.





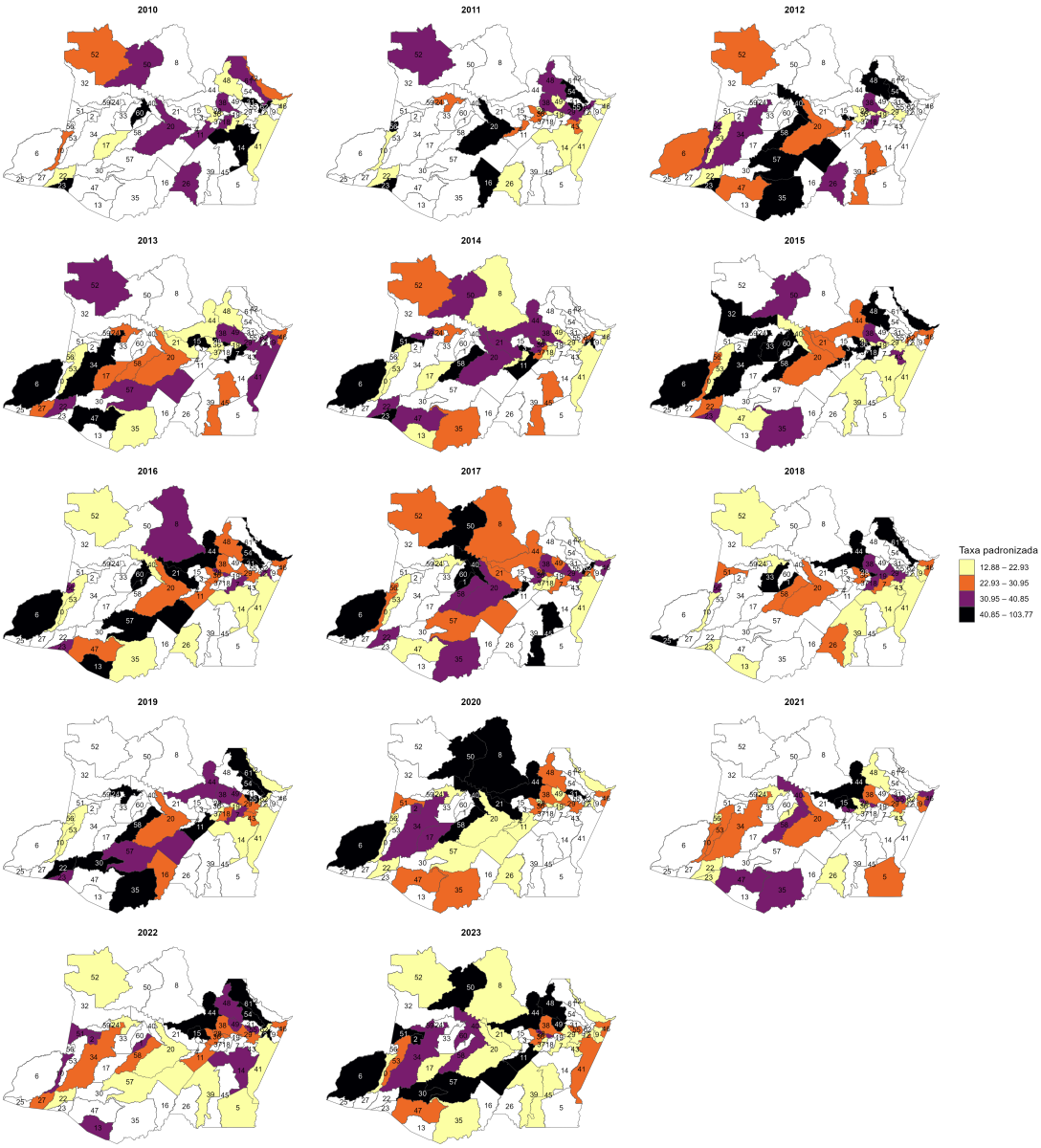
Fonte: Adaptado de MS/SVS/CGIAE – Sistema de Informações sobre Mortalidade (SIM); IBGE – Estimativas de População.

Figura 23: Taxa média padronizada anual de mortalidade por câncer de mama – Amazonas, 2010 a 2023.



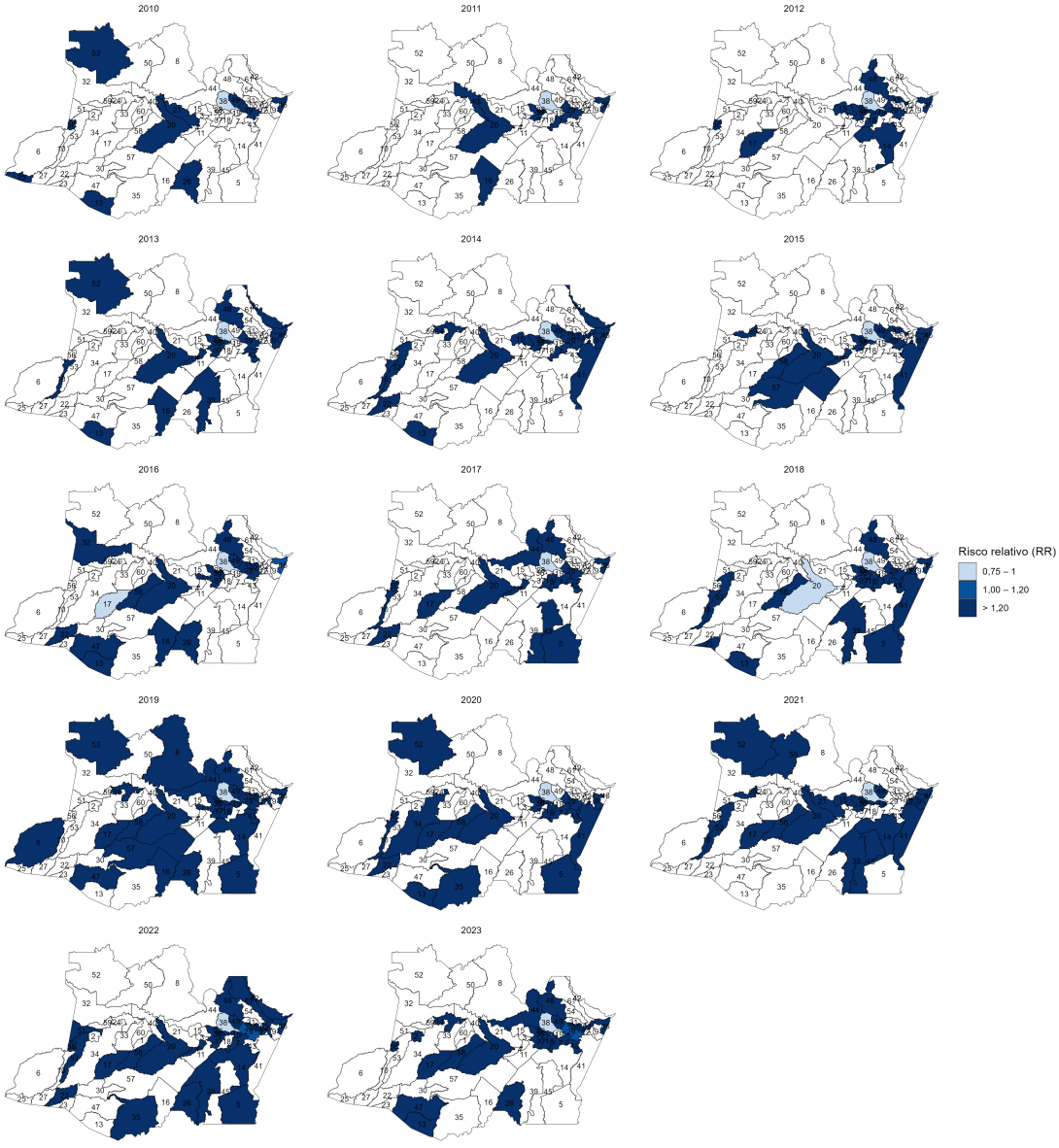
Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Figura 24: Taxa média padronizada anual de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Figura 25: Risco relativo padronizado anual de mortalidade por câncer de mama – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

Figura 26: Risco relativo padronizado anual de mortalidade por câncer do colo do útero – Amazonas, 2010 a 2023.



Fonte: Adaptado de MS/SVS/CGIAE - Sistema de Informações sobre Mortalidade – SIM; IBGE - Estimativas de População;

REFERÊNCIAS

Bermudi. **Análise espacial e espaço-temporal dos óbitos por câncer de mama e do colo do útero, município de São Paulo, 2000 a 2016.** 162 p. Dissertação (Dissertação (Mestrado)) — Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, 2020. Orientador: Francisco Chiaravalloti Neto
Coorientadora: Alessandra Cristina Guedes Pellini.

Bonita, Beaglehole e Kjellström. **Epidemiologia Básica**. 2. ed. São Paulo: Santos, 2010. 213 p. Revisão de português: Elvira Castanon. Diagramação: Rodrigo S. dos Santos. Capa: Gilberto R. Salomão. Tradução de: Basic Epidemiology, 2nd edition. ISBN 978-85-7288-839-4.

Brasil et al. **Introdução à Estatística Espacial para a Saúde Pública**. Brasília: Ministério da Saúde, 2007. 120 p. (Série B. Textos Básicos de Saúde; Série Capacitação e Atualização em Geoprocessamento em Saúde, 3). Ilustrado. ISBN 978-85-334-1427-3.

Brasil, Ministério da Saúde. **Saúde Brasil 2020/2021 : uma análise da situação de saúde diante da pandemia de covid-19, doença causada pelo coronavírus SARS-CoV-2**. 2022. Acesso em: 09 out. 2025. Disponível em: <https://bvsms.saude.gov.br/bvs/publicacoes/saude_brasil_2020-2021_analise_pandemia_covid-19.pdf>.

Brasil. Ministério da Saúde. **Sistema de Informações sobre População residente**. 2024. Acesso em: 20 set. 2025. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popsvs2024br.def>.

Brasil, Ministério da Saúde. **Câncer**. 2025. Acesso em: 04 dez. 2025. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer>.

Brasil. Ministério da Saúde. **Sistema de Informações Ambulatoriais – SIA/SUS**. 2025. Acesso em: 17 dez. 2025. Disponível em: <https://wiki.saude.gov.br/sia/index.php/P%C3%A1gina_principal>.

Brasil, Ministério da Saúde. **Sistema de Informações sobre Mortalidade - SIM**. 2025. Acesso em: 20 jun. 2025. Disponível em: <https://datasus.saude.gov.br/mortalidade-desde-1996-pela-cid-10>.

Brasil. Ministério da Saúde. **Sistema de Informações sobre Mortalidade (SIM)**. 2025. Acesso em: 17 dez. 2025. Disponível em: <https://www.gov.br/saude/pt-br/composicao/svsa/sistemas-de-informacao/sim>.

Brasil, Ministério da Saúde. **Sobre o DATASUS**. 2025. Acesso em: DD Mês AAAA. Disponível em: <https://datasus.saude.gov.br/sobre-o-datasus/>.

Casella and Berger. **Statistical Inference**. 2. ed. Pacific Grove: Duxbury Press, 2001. ISBN 9780534243128.

Casella e Berger. **Inferência estatística**. 2. ed. São Paulo: Cengage Learning, 2022. Tradução de Solange Aparecida Visconte. ISBN 978-85-221-0894-7.

Cordeiro and Demétrio. **Modelos Lineares Generalizados**. Santa Maria, RS: [s.n.], 2007. Minicurso apresentado no 12º SEAGRO e na 52ª Reunião Anual da RBRAS. 10 de julho de 2007.

DATASUS. **Metodologia de Administração de Dados (MAD)**. 2025. Acesso em: 06 fev. 2026. Disponível em: <https://datasus.saude.gov.br/wp-content/uploads/2025/06/MAD_240625.pdf>.

Girianelli, Gamarra e Silva. Os grandes contrastes na mortalidade por câncer do colo uterino e de mama no brasil. **Revista de Saúde Pública**, v. 48, n. 3, p. 459–467, 2014.

IARC. **All cancers fact sheet**. 2025. Acesso em: 24 nov. 2025. Disponível em: <https://gco.iarc.who.int/media/globocan/factsheets/cancers/39-all-cancers-fact-sheet.pdf>.

IARC. **Cancer Today — GLOBOCAN 2022 (Version 1.1)**. 2025. Acesso: 04 dez. 2025. Disponível em: <https://gco.iarc.who.int/today>.

IBGE. **Censo Demográfico 2010: Resultados do Universo**. 2010. Acesso em: 20 set. 2025. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?edicao=10503&t=resultados>.

IBGE. **Censo Demográfico 2022: Resultados Preliminares**. 2022. Acesso em: 20 set. 2025. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-demografico-2022.html>.

IBGE. **IBGE – Portal do IBGE**. 2025. Acesso em: 17 dez. 2025. Disponível em: <https://www.ibge.gov.br/>.

IBGE. **Manaus (AM) — Cidades e Estados**. 2025. Acesso: em 28 de nov. de 2025. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/am/manaus.html>.

IBGE. **Panorama do Estado do Amazonas**. 2025. Acesso: em 26 de nov. de 2025. Disponível em: <https://cidades.ibge.gov.br/brasil/am/panorama>.

INCA. **Ministério da Saúde - Atlas on-line de Mortalidade.** 2014. Acesso em: 24 nov. 2025. Disponível em: <https://www.inca.gov.br/MortalidadeWeb/pages/Modelo04/consultar.xhtml#panelResultado>.

INCA. **Ministério da Saúde - Atlas on-line de Mortalidade**. 2014. Acesso em: 20 set. 2025. Disponível em: <https://www.inca.gov.br/MortalidadeWeb/pages/Modelo03/consultar.xhtml#panelResultado>.

INCA. **Dados e números sobre câncer de mama: Relatório anual 2023**. 2023. Acesso em: 15 dez. 2025. Disponível em: <https://www.inca.gov.br/publicacoes/relatorios/dados-e-numeros-sobre-cancer-de-mama-relatorio-anual-2023>.

INCA. **Dados e números sobre câncer do colo do útero: Relatório anual 2023**. Rio de Janeiro: [s.n.], 2023. Acesso em: 15 dez. 2025. Disponível em: <https://www.inca.gov.br/publicacoes/relatorios/dados-e-numeros-sobre-cancer-do-colo-do-utero-relatorio-anual-2023>.

INCA. **Estatísticas de câncer**. 2025. Acesso em: 24 nov. 2025. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>.

Kulldorff, M. A spatial scan statistic. **Communications in Statistics – Theory and Methods**, v. 26, p. 1481–1496, 1997.

Kulldorff, M. and Nagarwalla, N. Spatial disease clusters: Detection and inference. **Statistics in Medicine**, v. 14, p. 799–810, 1995.

Kulldorff, M. – SaTScan User Guide. **SaTScan User Guide**. 2018. Acesso em 28 de nov. de 2025. Disponível em: <https://www.satscan.org/>.

Malta et al. Doenças crônicas não transmissíveis e a utilização de serviços de saúde: análise da pesquisa nacional de saúde no brasil. **Revista de Saúde Pública**, 2017.

Morettin and Bussab. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010. ISBN 978-85-02-08177-2.

NCI. **National Cancer Institute - World (WHO 2000-2025) Standard Populations**. 2013. Acesso em: 20 set. 2025. Disponível em: <https://seer.cancer.gov/stdpopulations/world.who.html>.

Olinda, R. A. **Modelos Lineales Generalizados: Análisis de datos provenientes de un ensayo en un DCL y dosis-respuesta**. 2020. Última atualização há cerca de 5 anos. Acesso em: 18 fev. 2026. Disponível em: <https://rpubs.com/ricardo_olinda/ensayo_DCL>.

Oliveira et al. Association of cervical and breast cancer mortality with socioeconomic indicators and availability of health services. **Cancer Epidemiology**, v. 64, p. 101660, 2020. Disponível em: <https://doi.org/10.1016/j.canep.2019.101660>.

Pereira. **Perfil epidemiológico e clínico de mulheres jovens com câncer de mama no Amazonas: estudo de onze anos**. 82 p. Dissertação (Dissertação (Mestrado em Ciências da Saúde)) — Universidade Federal do Amazonas, Manaus, 2016.

PNUD,IPEA e FJP. **Atlas do Desenvolvimento Humano no Brasil**. 2025. Acesso em: 17 dez. 2025. Disponível em: <http://www.atlasbrasil.org.br/>.

RStudio Team. **Rstudio**. 2026. Acesso em: 18 fev. 2026. Disponível em: <https://www.rstudio.com/>.

SES-AM. **Regionalização**. 2023. <https://www.saude.am.gov.br/regionalizacao/>. Acesso: em 26 de nov. de 2025.

Silva, H. F. S. **Análise da produção ambulatorial das Unidades Básicas de Saúde de Manaus: aplicação multivariada de clusters e uma abordagem puramente espacial**. Dissertação (Trabalho de Conclusão de Curso (Bacharelado em Estatística)) — Universidade Federal do Amazonas, Manaus, AM, 2021.

Silva, Peres and Boscarioli. **Introdução à mineração de dados: com aplicações em R**. 1. ed. Rio de Janeiro: Elsevier, 2016. ISBN 978-85-352-8446-1.

Souza Junior, G. L. **Subnotificação de casos de câncer de mama: um estudo de conglomerados espaciais no estado de São Paulo**. Dissertação (Dissertação (Mestrado em Estatística)) — Universidade Federal de Minas Gerais, Belo Horizonte, dez. 2005.

Teixeira. **Aglomerados espaciais de alta mortalidade por câncer no Brasil: uma abordagem de Machine Learning**. 88 p. Dissertação (Dissertação (Mestrado)) — Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, 2021. Orientador: Alexandre Dias Porto Chiavegatto Filho.

WHO. **Cancer**. 2025. Acesso em: 24 nov. 2025. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/cancer>.