



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

PROPESQ
Pró-Reitoria de Pesquisa UFPB

RELATÓRIO DE EXECUÇÃO PIBIC

(Setembro de 2023 – Agosto de 2024)

TÍTULO DO PROJETO DE PESQUISA:

Abordagem espaço-temporal dos fatores associados ao comportamento da mortalidade e da fecundidade no semiárido brasileiro

TÍTULO DO PLANO DE TRABALHO:

Fatores associados ao comportamento da mortalidade no semiárido brasileiro

Orientadora

Everlane Suane de Araújo da Silva – Docente do Departamento de Estatística

Bolsista

Schênia Taynna Medeiros Silva – Graduanda em Estatística

João Pessoa - PB

Agosto, 2024

Everlane Suane de A. Silva¹
Schênia Taynna Medeiros Silva²

1 - Resumo

O estudo tem em vista avaliar e ajustar a qualidade dos registros de óbitos no Semiárido Brasileiro, construindo indicadores de mortalidade pré e Pós a pandemia (COVID-19), além de se propor a investigar suas relações com variáveis sociodemográficas e de saúde para os triênios de 2018 e 2021. Para alcançar isso, foram realizados ajustes nas estatísticas de óbitos por meio de métodos demográficos, avaliação das causas de morte e relação com indicadores de desenvolvimento humano. A metodologia envolve a coleta e organização de dados do Semiárido Brasileiro, que atualmente compreende onze estados contendo 1477 municípios. Os Dados de mortalidade foram extraídos do Sistema de Informações sobre Mortalidade (SIM), e a base de dados foi organizada seguindo uma nova tipologia proposta pelo IBGE que usa a densidade demográfica para classificar os municípios. Um aspecto fundamental deste estudo consistiu na redistribuição das “Causas Mal Definidas” (CMD) – óbitos cuja causa não é claramente identificada, os coeficientes de redistribuição foi aplicado baseados na distribuição proporcional das causas bem definidas notificadas com exclusão das causas externas.

Palavras-chave: Mortalidade, Semiárido brasileiro, COVID-19.

¹Docente do Departamento de Estatística da Universidade Federal da Paraíba (UFPB); Pesquisadora do Laboratório de Estudos Demográficos (LED) do Departamento de Estatística da UFPB e coordenadora do Projeto de Iniciação Científica: “Abordagem espaço-temporal dos fatores associados ao comportamento da mortalidade e da fecundidade no Semiárido brasileiro.”, período 01 de setembro de 2023 a 31 de agosto de 2024.

²Discente do curso de graduação em Estatística da UFPB e membro do Projeto de Iniciação Científica: “Fatores associados ao comportamento da mortalidade no semiárido brasileiro”, período 01 de setembro de 2023 a 31 de agosto de 2024.

1 - Resume

The study aims to evaluate and adjust the quality of death records in the Brazilian semi-arid region, constructing mortality indicators before and during the COVID-19 pandemic. It also proposes to investigate their relationships with sociodemographic and health variables for the trienniums of 2018 and 2021. To achieve this, adjustments were made to death statistics using demographic methods, evaluation of causes of death, and their relation to human development indicators. The methodology involves collecting and organizing data from the Brazilian semi-arid region, which currently comprises eleven states containing 1,477 municipalities. Mortality data were extracted from the Mortality Information System (SIM), and the database was organized following a new typology proposed by IBGE, which uses population density to classify municipalities. A fundamental aspect of this study consisted of the redistribution of "Poorly Defined Causes" (PDC) – deaths whose cause is not clearly identified. Redistribution coefficients were applied based on the proportional distribution of well-defined reported causes, excluding external causes.

Keywords: Mortality, Brazilian semi-arid region, COVID-19.

2 - Introdução

A doença causada pelo coronavírus SARS-CoV-2, mais conhecida como COVID-19, surgiu no final de 2019 e rapidamente se espalhou pelo mundo, forçando a Organização Mundial da Saúde (OMS) em março de 2020 declarar estado de emergência mundial. A pandemia caracterizada pela crise respiratória, variou desde casos leves com os assintomáticos até casos mais graves de UTI, sendo necessário desencadear medidas de distanciamento social e campanhas de vacinação em massa.

Com seu crescimento exponencial a pandemia se tornou um evento de saúde pública que marcou a geração atual. A pandemia variou em intensidade e natureza em diferentes partes do mundo, dependendo das medidas adotadas, das características socioeconômicas e da infraestrutura de cada região. Pessoas perdendo seus empregos, empresas fechando suas portas, crianças em fase de desenvolvimento sendo obrigadas a se isolar, professores e alunos sem a estrutura necessária para ensinar e aprender, além de muitas outras áreas de impacto e sequelas que só serão percebidas e mensuradas daqui a alguns anos.

Para o Semiárido brasileiro os impactos da COVID-19 não seriam diferentes, além dos desafios econômicos e climático vivido diariamente pela população, a pandemia trouxe novos desafios para a região, impactando de maneiras diversas. A incessante busca por identificar os problemas e encontrar soluções se tornou ainda mais essencial, à medida que as comunidades enfrentam os efeitos combinados da crise de saúde pública e das condições climáticas características da região.

Nesse sentido, devemos nos perguntar, "Os dados de mortalidade no Semiárido brasileiro trouxe um diferencial importante durante a pandemia e pós-pandemia?", "Ao passar o período mais crítico da pandemia, quais foram as principais mudanças identificadas?". A partir desta perspectiva, este estudo tem o objetivo geral:

- Avaliar e ajustar a qualidade dos registros de óbitos, e ao construir indicadores de mortalidade para o período pré e pós-pandêmico (COVID-19), investigar as suas relações com variáveis sociodemográficas e da saúde para o território do Semiárido Brasileiro, no período de 2019-2022.

E como objetivos específicos:

- Atualização das bases de dados para o ano mais recente disponível no Sistema de Informações sobre Mortalidade (2022). Ajuste nas estatísticas de óbitos, ou seja, correção dos sub-registros, redistribuição dos códigos garbage e correção da incompletude das variáveis consideradas nas declarações de óbitos;
- Avaliar o cenário da mortalidade por causas básicas no Semiárido brasileiro pré e pós a pandemia (COVID-19);
- Avaliar a relação das taxas de mortalidade com a condição de vida da região.

3 - Metodologia

3.1 - Fonte e Organização dos Dados

O Semiárido brasileiro se estende por onze estados contendo 1477 municípios, totalizando 15% de todo território nacional. Nesse sentido, como mostrado na Tabela 1

Tabela 1 – Municípios incluídos no Semiárido

Estado	2017	2021	2024*
Alagoas	38	38	42
Bahia	278	283	287
Ceará	175	171	175
Espírito Santos	0	6	6
Maranhão	2	16	16
Minas Gerais	91	209	217
Paraíba	194	188	198
Pernambuco	123	137	142
Piauí	185	215	216
Rio Grande do Norte	147	141	148
Sergipe	29	23	30
Total	1.262	1.427	1.477

Fonte: Superintendência do Desenvolvimento do Nordeste-(SUDENE).

houve o acréscimo de 50 municípios comparando com o ano de 2021, abrigando cerca de 28 milhões de habitante.

A atualização e extração dos dados foi feita de maneira automatizada através da rapagem de dados utilizando-se da biblioteca Selenium presente na linguagem de programação Python, assim como toda a estrutura da base de dados foi organizada utilizando a mesma linguagem.

A base de dados para o território do Semiárido brasileiro foi construída e organizada considerando uma nova abordagem metodológica proposta pelo Instituto Brasileiro de Geografia e Estatística–IBGE, no ano de 2017, que visa classificar e categorizar os municípios do território brasileiro sob a tipologia Urbana, Rural Remoto, Rural Adjacente, Intermediário Remoto ou Intermediário Adjacente cujo critério fundamental escolhido pelo IBGE para essa aplicação é a densidade demográfica. Dos 1.477 municípios que compõe o território do Semiárido brasileiro, os novos espaços regionalizados estão distribuídos contendo a seguinte quantidade de municípios: Rural Adjacente com 1035 municípios, Intermediário Adjacente com 189 municípios, Intermediário Remoto com 7 municípios, Rural Remoto com 90 municípios e por fim o Urbano contendo 156 municípios.

A coleta dos dados foi feita através do TABNET que facilita o acesso aos dados do Sistema de Informações sobre Mortalidade–SIM, considerando o triênio 2019–2021, que abrange o período pré e pandêmico da COVID-19. Os dados de óbitos por causa básica estão categorizados segundo a Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde na sua 10ª revisão (CID-10), introduzida no Brasil no ano de 1996. A CID, desenvolvida pela OMS, trata-se de códigos com o objetivo de padronizar e classificar não só doenças, como também sintomas, sinais, queixas, entre outros fatores.

Por fim, destaca-se que, ao analisar dados e informações de domínio público, e conforme o inciso III do parágrafo único do artigo 1º da Resolução nº 510/16 do Conselho Nacional de Saúde, este estudo está isento da obrigatoriedade de registro e avaliação pelo Comitê de Ética em Pesquisa (CEP) e pela Comissão Nacional de Ética em Pesquisa (CONEP).

3.2 - Cobertura e Correção dos Dados de Óbitos

O método da Equação do Balanço de Crescimento, proposto por Brass em 1975, foi utilizado para tentar estimar o grau de cobertura no território do Semiárido brasileiro. Este método baseia-se no pressuposto da teoria das populações estáveis, isto é, consideramos que a estrutura etária e o crescimento populacional não mudam com o passar do tempo. Apesar disto, um dos motivos de sua escolha foi sua notoriedade ao ser considerado robusto mesmo descumprindo com este pressuposto. Para sua aplicação é necessário apenas conhecer a distribuição da população e de seus óbitos por faixa etária. Por fim, para estimarmos a cobertura ajustamos um modelo de regressão apresentado a seguir:

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

Onde,

$N(a)$ = População na idade exata a ;

$N(a+)$ = População acumulada da idade exata a ;

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

r = A taxa de crescimento populacional estável;

k = Fator de correção dos óbitos.

Em que, a cobertura é dada pelo quociente de $\frac{1}{k}$.

Apesar da robustez do método, alguns problemas estão associados a ele, por exemplo: efeito dos desvios na estabilidade, efeitos no grau de cobertura dos óbitos ser constante por idade e problemas de enumeração na declaração da idade. Para solucionar alguns destes problemas, a literatura recomenda a seleção de diferentes grupos etários mais estáveis no ajuste do modelo de regressão, isto é, buscarmos faixas etárias onde os dados tendem apresentar a menor quantidade de erro possível.

A respeito da classificação da qualidade dos registros vitais, foi utilizado uma classificação adaptada por Paes (2007), proposta por Chackiel (1987), onde: coberturas acima de 90% são consideradas “muito boas”, de 81% e 90% são “boas”, de 71% e 80% são consideradas “regulares” e abaixo dos 71% consideradas “insuficientes”.

Durante a pesquisa foram aplicados diferentes critérios de seleção das faixas etárias, entretanto nenhum dos grupos selecionados apresentaram um fator de correção que fazia sentido para a correção da base de dados. Por conta disso, para dar continuidade as correções nos dados, utilizamos os resultados mais recentes de cobertura de óbitos dos estados divulgados pelo Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis-DAENT. Onde, a Busca Ativa trata-se de uma estratégia de captação de eventos do Ministério da Saúde que não foram informados ao Sistema de Informações sobre Mortalidade-SIM.

Dessa forma, o fator de correção k para cada um dos estados do Semiárido brasileiro é dado pelo quociente de 1 sobre a taxa de cobertura. No site do DATASUS, foram extraídos os dados de mortalidade do Semiárido, dessa vez considerando a tipologia definida pelo IBGE e desagrupando por estados. Com esse pequeno ajuste pôde-se efetuar a correção dos dados multiplicando pelo fator de correção k e, por fim, agrupando novamente para voltarmos a tipologia definida anteriormente. O método foi aplicado para o triênio 2019-2021, tanto para o sexo masculino quanto para o sexo feminino.

Tabela 2 – Grau de Cobertura da Mortalidade pela Busca Ativa

Estado	Infantis		Adultos	
	2018	2021	2018	2021
Alagoas	85,9	89,3	90,3	92,8
Bahia	87,0	91,2	90,1	93,8
Ceará	86,2	89,6	93,4	93,7
Espirito Santos	100	100	100	100
Maranhão	80,0	86,0	91,3	94,2
Minas Gerais	86,2	92,1	89,8	93,7
Paraíba	84,1	87,9	90,3	91,7
Pernambuco	92,4	98,0	94,2	98,9
Piauí	81,1	86,0	91,2	93,5
Rio Grande do Norte	83,1	86,1	91,3	93,2
Sergipe	94,1	97,0	92,2	94,9

Fonte: Ministério da Saúde/SVSA/DAENT/CGIAE - SIM e Busca Ativa.

3.3 - Redistribuição das Causas Mal Definidas

Definida no capítulo XVIII–Sintomas, sinais e achados anormais de exames clínicos e de laboratório não classificado em outra parte, comumente conhecido como Causas Mal Definidas, é um grupo de causas que ocorre quando não se sabe ao certo o que ocasionou aquele óbito.

Por conta disso, o método de Ledermann (1955) é comumente utilizado para redistribuir esse grupo para outras causas bem definidas de maneira proporcional. Em outras palavras, o método estima a quantidade de óbitos mal definidos que deveriam estar em uma das causas bem definidas por meio de um ajuste de um modelo de regressão linear simples entre, as proporções de cada uma das causas de morte bem definidas como variável resposta e a proporção de mortes mal definidas como variáveis independentes. Entretanto, ao aplicar o método utilizando as tipologias do semiárido brasileira que são apenas cinco, foi observado sub reajuste na regressão por ter poucos pontos e ocorrendo redistribuição de óbitos por causas bem definidas.

Nesse sentido, se faz importante utilizar o método alternativo que consiste em calcular o coeficiente de redistribuição das causas mal definida, baseado na distribuição proporcional de todas as causas definidas notificadas por sexo e faixa etária quinquenais, com exclusão das causas externas.

3.4 - Relação entre a Taxa de Mortalidade Bruta padronizada e condição de vida

As variáveis de condições de vida desempenham um papel crucial na análise e compreensão das disparidades sociais e econômicas em uma população. Compreende as informações sobre as condições de vida da população no seu sentido mais amplo, abrangendo medidas de desigualdade e pobreza; inclusão ou exclusão social; indicadores de situação social, qualidade de vida e de vulnerabilidade ambiental; entre outros aspectos.

Essas variáveis são fundamentais para identificar e entender as desigualdades existentes e seus impactos nas taxas de mortalidade e outros indicadores de saúde, nesse sentido o presente trabalho utilizou as seguintes variáveis relacionar com a taxa de mortalidade bruta padronizada:

- Taxa de alfabetização

- Porcentagem de casas com rede de esgoto

A relação entre a taxa de analfabetismo e a proporção da rede de esgoto com a taxa de mortalidade é crucial para entender como fatores socioeconômicos e de infraestrutura que afetam a saúde pública. A taxa de analfabetismo pode indicar o nível de educação e acesso à informação, enquanto a proporção da rede de esgoto reflete o acesso ao saneamento básico, ambos influenciando diretamente a qualidade de vida e, consequentemente, as taxas de mortalidade.

Dessa forma, para analisar como a taxa de analfabetismo e a proporção de rede de esgoto em relação à taxa de mortalidade, utilizamos uma regressão linear múltipla. Esse método estatístico permite modelar a relação entre uma variável dependente (taxa de mortalidade) e várias variáveis independentes (taxa de analfabetismo e proporção de rede de esgoto).

O modelo de regressão linear múltipla dado como :

$$\text{Taxa de Mortalidade} = \beta_0 + \beta_1 \cdot \text{Taxa de Analfabetismo} + \beta_2 \cdot \text{Proporção de Rede de Esgoto} + \epsilon$$

Onde,

β_0 é Intercepto do modelo

β_1 é Coeficiente da variável "Taxa de Analfabetismo".

β_2 é Coeficiente da variável "Proporção de casas com Rede de Esgoto".

ϵ é o erro do modelo.

A análise de regressão linear múltipla foi realizada no R utilizando dados referentes ao ano de 2022.

4 - Resultados

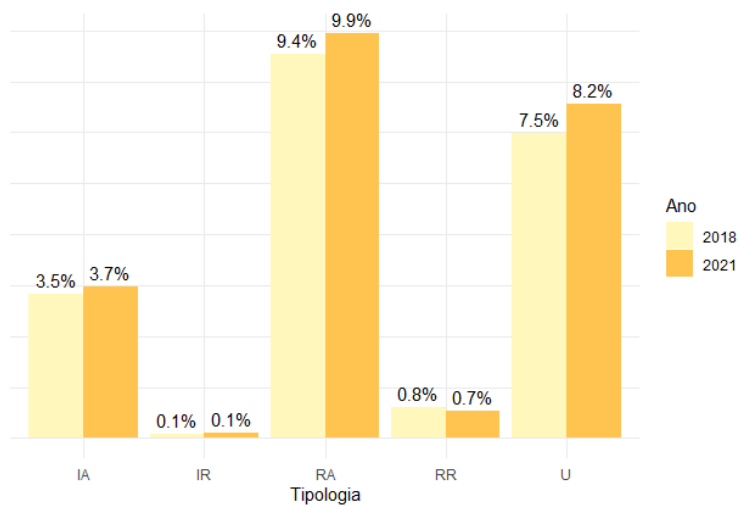
Conforme os dados disponíveis no Sistema de Informações sobre Mortalidade – SIM/DATASUS, o percentual de óbitos identificados com o o capítulo XVIII da CID-10 conhecido como Causas Mal Definidas (CMD) apresentou variações conforme a tipologia e o sexo. Para o sexo feminino, a tipologia Rural Adjacente (RA) apresentou o maior percentual de CMD, com aproximadamente 9,43% em 2018 e 9,92% em 2021. Nas outras tipologias, como Intermediário Adjacente (IA), Intermediário Remoto (IR).), Rural Remoto (RR) e Urbano (U), os percentuais foram significativamente menores, variando de 0,0997% a 8,19% ao longo do período desenvolvido.

Para o sexo masculino, a tipologia Rural Adjacente (RA) também exibiu o maior percentual de CMD, com cerca de 12,2% em 2018 e 12,9% em 2021. Em comparação, nas zonas Intermediário Adjacente (IA), Intermediário Remoto (IR), Rural Remoto (RR) e Urbano (U), os percentuais variaram de 0,114% a 10,4%.

5 - Conclusão

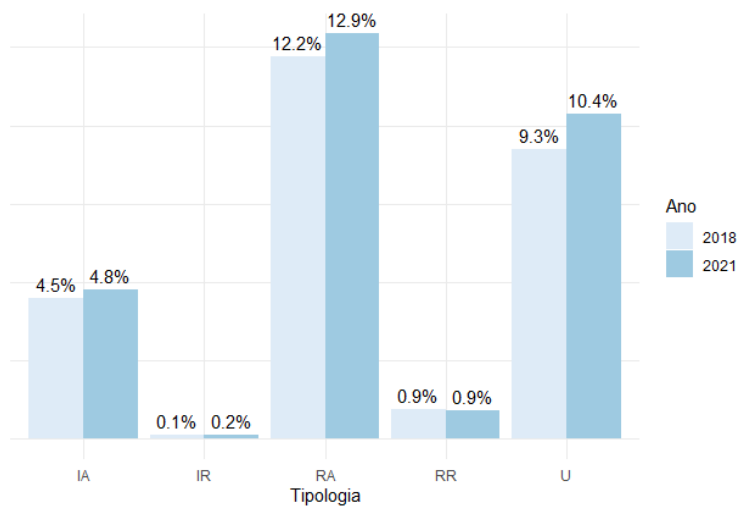
O presente modelo de relatório apresentou uma estrutura padrão de um relatório, bem como apresentou exemplos de utilização de algumas ferramentas essenciais e/ou úteis.

Figura 1 – Proporção das Causas Mal Definidas das Mulheres



Fonte: DATASUS/SIM

Figura 2 – Proporção das Causas Mal Definidas dos Homens



Fonte: DATASUS/SIM

Referências