



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos Factores Associados
à Mortalidade Infantil
em Moçambique em 2011**

Autor: Jacinto Pedro Cumbana

Maputo, Setembro de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos Factores Associados
à Mortalidade Infantil
em Moçambique em 2011**

Autor: Jacinto Pedro Cumbana

Supervisor: Zacarias Bernardo Mutombene, Msc, Universidade Pedagógica

Maputo, Setembro de 2025

Declaração de Honra

Declaro por minha honra que o presente Trabalho de Licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de **Licenciatura em Estatística**, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Fevereiro de 2024

(Jacinto Pedro Cumbana)

Dedicatória

Eu dedico este trabalho:

Ao meu pai **Pedro Jacinto Silva** (em memória), à minha mãe **Matilde Aires Macamo**, a toda a minha família (**Glória Pedro Fernando, Aschily da Glória Pedro e Jacinto Pedro Cumbana Jr.**), aos meus avós: **Jacinto Silva** (em memória) e **Rachel Zacarias Sumburane**, por todo o carinho e atenção dedicados durante toda a minha vida.

Agradecimentos

O meu percurso académico teve apoio de várias pessoas que fizeram parte desse processo lindo para que o mesmo tornar – se realidade, destacando – se entre outros os seguintes:
À Deus pelo dom gratuito da vida.

Ao meu supervisor **Zacarias Bernardo Mutombene**, pela orientação, bem como, pela entrega, pelas críticas e sugestões valiosas, para que eu pudesse concluir o trabalho de Licenciatura em estatística.

A minha profunda gratidão vai:

À todos docentes do **Departamento de Matemática e Informática**, em especial aos docentes da **secção de Estatística**, pela dedicação, profissionalismo e pelos diversos ensinamentos transmitidos.

À todos funcionários do **Departamento de Matemática e Informática**, pela dedicação, profissionalismo e pelos diversos ensinamentos transmitidos.

Aos meus colegas e amigos: Vai o meu muito Kanimambo a minha Squad composta por: **John Gailhac Paulo Garepe, Clésio Baltazar Mapande, Martins Pedro Guila, Kevin Baptista Samuel Nhantumbo, Euflásio Henrique Francisco, Telça Mario Massingue, Chande Ernesto Jamine, Neima José Candeia, Bonifácio José Nicolas**, pelas caminhadas feitas durante todo o percurso, pela troca de ideias sempre que encaravamos dificuldades e pelas críticas que contribuíram valiosamente para que hoje chegasse ao fim duma longa jornada e a todos colegas do DMI em especial aos do curso de estatística 2018 e a todos que directa ou indirectamente contribuíram, endereço a minha gratidão.

O meu muito Kwanimambo vai a todas pessoas que orar por mim durante toda jornada academica.

”Education is the most powerful
weapon which you can use to
change the world.”

Nelson Rolihlahla Mandela

Resumo

O estudo sobre os factores associados à mortalidade infantil em Moçambique não só lança luz sobre uma questão crucial de saúde pública, como também oferece uma perspectiva fundamental para informar políticas e intervenções direccionadas a melhorar a saúde e o bem-estar das crianças. Ao analisar uma série de variáveis, desde características demográficas até ao acesso aos cuidados de saúde, este estudo procura desvendar os complexos padrões subjacentes à mortalidade infantil no país. Com base em dados robustos do IDS de 2011, examinamos cuidadosamente como escolaridade da mãe, o contexto socioeconómico, o tamanho e sexo do bebé, entre outros factores, desempenham um papel significativo na determinação do destino das crianças moçambicanas durante o crucial primeiro ano de vida. Este estudo não só identifica esses factores críticos, como também os contextualiza em relação aos desafios e oportunidades específicos enfrentados por diferentes regiões e grupos populacionais dentro de Moçambique. Por meio de uma análise cuidadosa e abrangente, esperamos fornecer critério valiosos que possam orientar políticas e programas voltados para a redução da mortalidade infantil e o fortalecimento da saúde materno-infantil em Moçambique e além.

Palavras-chave: óbito; mortalidade infantil; Moçambique; Regressão logística

.

Abstract

The study on factors associated with infant mortality in Mozambique not only sheds light on a crucial public health issue but also offers a fundamental perspective to inform policies and interventions aimed at improving the health and well-being of children. By analyzing a range of variables, from demographic characteristics to access to healthcare, this study seeks to unravel the complex patterns underlying infant mortality in the country. Based on robust data from the IDS of 2011, we Carefully considering factors such as the mother's education level, socioeconomic context, infant size and gender, among other factors, play a significant role in determining the fate of Mozambican children during the critical first year of life. This study not only identifies these critical factors but also contextualizes them in relation to the specific challenges and opportunities faced by different regions and population groups within Mozambique. Through careful and comprehensive analysis, we hope to provide valuable insights that can guide policies and programs aimed at reducing infant mortality and strengthening maternal and child health in Mozambique and beyond.

Keywords:Death; infant mortality; Mozambique; Logistic regression..

Índice

1	INTRODUÇÃO	1
1.1	Contextualização	1
1.2	Definição de problema	2
1.3	Objectivos da pesquisa	3
1.3.1	O objectivo geral	3
1.3.2	Objectivos específicos	3
1.4	Delimitações do estudo	3
1.5	Relevância do estudo	3
1.6	Estrutura do Trabalho	4
2	REVISÃO DE LITERATURA	5
2.1	Referencial teórico	5
2.1.1	Mortalidade Infantil no Mundo	6
2.1.2	Mortalidade Infantil na África	6
2.1.3	Mortalidade Infantil em Moçambique	6
2.1.4	Idade da mãe	7
2.1.5	Estado civil	7
2.1.6	Escolaridade da mãe ao nascer	8
2.1.7	Número de consultas de pré-natal	9
2.1.8	Área de Residência	10
2.1.9	Ordem de Nascimento	10
2.1.10	Religião	10
2.1.11	Sexo do recém-nascido	13
2.1.12	Peso ao nascer	13
2.1.13	Tipos de parto	14
2.1.14	Quartil de riqueza	14
2.2	Análise de Regressão	15
2.3	Modelos Lineares Generalizados	15
2.4	Modelo de Regressão Logística	17
2.5	Análise de dados amostrais complexos	18
2.6	Vantagens do modelo de regressão logística	19

2.7	Comparação de modelo de Regressão logística e o modelo de Regressão linear	20
2.8	Procedimentos de Análise	20
3	MATERIAL E METÓDOS	22
3.1	Classificação de pesquisa	22
3.1.1	Quanto à natureza da pesquisa	22
3.1.2	Quanto aos procedimentos técnicos	22
3.1.3	Quanto à abordagem ou método	22
3.1.4	Quanto aos objectivos	23
3.2	ÁREA DE ESTUDO	23
3.3	MATERIAL	24
3.3.1	Balanceamento de Classes	25
3.4	MÉTODOS	27
3.5	Tratamento de Dados	27
3.5.1	Teste de independência	27
3.5.2	Teste de significância dos parâmetros	28
3.5.3	Teste de Wald	28
3.5.4	Teste de razão de verosimilhança	29
3.5.5	Matriz de confusão	30
3.5.6	Teste Hosmer e Lemeshow	31
3.5.7	Qualidade de ajuste do modelo logístico classico	32
3.5.8	Métodos de Seleção das Variáveis	33
3.5.9	Interpretação dos parâmetros do modelo	33
4	RESULTADOS E DISCUSSÃO	35
4.1	Análise Descritiva	35
4.1.1	Distribuição de óbitos infantis por sexo da criança	40
4.1.2	Distribuição de óbitos infantis por filiação religiosa da mãe	41
4.2	Análise de Associação entre Variáveis Independentes e dependente (Óbito)	41
4.3	Modelo de Regressão Logística Binária	42
4.3.1	Construção do Modelo de regressão logística binária	42
4.3.2	Modelo logístico que prevê a probabilidade de uma criança morrer	48
4.4	Discussão dos Resultados	48
5	CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES	50
5.1	Conclusões	50
5.2	Limitações do Estudo	50
5.3	Recomendações	51

ÍNDICE

ÍNDICE

REFERÊNCIAS	51
ANEXOS	57
Anexo A - Código de Aplicação da Técnica SMOTE	57

Acrónimos

AIDS	Síndrome da Imunodeficiência Adquirida
DHS	Demographic Health Survey
E.P	Erro Padrão
gl	grau de liberdade
HIV	vírus da imunodeficiência humana
I.C	Intervalo de Confiança
IDS	Inquérito Demográfico e de Saúde
INE	Instituto Nacional de Estatística
MISAU	Ministério de Saúde
MLG	Modelo Linear Generalizado
OMS	Organização Mundial de Saúde
ONGs	Organizações Não Governamentais
PNDS	Pesquisa Nacional de Demografia e Saúde
R.C	Razão de Chance
SPSS	Statistical Package for Social Science
UN IGME	United Nations Inter-Agency Group for Child Mortality Estimation
UNICEF	United Nations International Children's Emergency Fund
UPA	Unidades Primárias de Amostragem

Lista de Figuras

2.1	Modelo conceptual de como a religião pode afectar directamente e indirectamente o comportamento dos adolescentes	11
3.1	Mapa de Moçambique	24
4.1	Distribuição de óbitos infantis por sexo da criança	40
4.2	Distribuição de óbitos infantis por filiação religiosa da mãe	41

Lista de Tabelas

2.1	Comparação de modelo de regressão logística e de regressão linear simples .	20
3.1	Descrição das Variáveis e Codificação das Categorias	26
3.2	Tabela de classificação	31
4.1	Distribuição de frequências absolutas (fa) e relativas (fr%)	35
4.2	Teste Qui-quadrado de associação entre variáveis independentes e Dependente	42
4.3	Estimativas dos coeficientes do modelo	43
4.4	Modelo logístico nulo (Passo 0)	46
4.5	Teste de razão de verossimilhança	46
4.6	Medidas de avaliação do ajuste do modelo final	46
4.7	Teste de Hosmer e Lemeshow	47
4.8	Tabela de classificação do modelo final	47
5.1	Tabela de classificação do modelo nulo	56
5.2	Tabela de Contingência para o Teste de Hosmer e Lemeshow	56
5.3	Coeficientes do Modelo de Regressão	56

Capítulo 1

INTRODUÇÃO

Este capítulo contextualizará o tema de pesquisa, apresentando o problema de pesquisa a ser investigado, os seus objectivos, bem como as justificativas e relevância para a realização deste estudo. A delimitação da pesquisa também será apresentada neste capítulo.

1.1 Contextualização

A mortalidade é uma das variáveis demográficas determinantes da esperança de vida de uma dada população. Estudos sobre níveis e padrões de mortalidade são extremamente importantes, uma vez que mostram a qualidade de vida e bem-estar de saúde da população. Rutstein (1984) aponta que uma sociedade com poucos recursos e bem disseminados é mais provável que apresente taxas de mortalidade mais baixas do que uma com recursos escassos ou mal distribuídos. Há, no entanto, segundo o mesmo autor, países com recursos não distribuídos de forma ampla ou equitativa que atingiram níveis de mortalidade mais baixos do que o previsto, devido aos factores culturais e instituições nos programas de mitigação da pobreza.

Moçambique, bem como os outros países da África Subsaariana, apresenta uma esperança de vida abaixo da média africana (53,7 anos), que se deve, em grande medida, ao índice elevado de mortalidade infantil que vem sendo observado nos últimos anos. Conforme o Censo de 2017, a esperança de vida ao nascer em Moçambique para homens é de 51,0 anos e para as mulheres é de 56,6 anos. A esperança de vida ao nascer observa dispersão ao nível nacional/regional, rural/urbana e também ao nível provincial. Conforme argumenta o INE (1998), Moçambique apresenta um nível de mortalidade infantil elevado e com diferenças expressivas, quando consideradas as desigualdades provinciais ou a situação de residência urbana/rural.

Entre 1997 e 2011, Moçambique registou uma redução significativa nas taxas de mortalidade infantil e infanto-juvenil. Segundo dados do Inquérito Demográfico e de Saúde

(IDS), a taxa de mortalidade infantil (óbitos antes de 1 ano) diminuiu de 135 para 101 por mil nascidos vivos entre 1997 e 2003, e para 64 por mil em 2011. Já a taxa de mortalidade infanto-juvenil (óbitos antes dos 5 anos) caiu de 201 para 122 por mil nascidos vivos em 2003, e para 97 por mil em 2011.

Este facto pode resultar da fraca cobertura de serviços hospitalares e de saúde no país. Não obstante, de acordo com o MISAU (2009), no país verifica-se um aumento de unidades sanitárias e de instituições de consultas em todas as províncias. A mesma fonte adianta que, em Moçambique, observa-se aumento de cobertura de consultas pré-natais, coberturas de partos institucionais e consultas pós-parto. No entanto, apesar de uma relativa desconcentração dos serviços de saúde, o país apresenta ampla desvantagem em relação aos outros países da África Austral, em termos de unidades de saúde e aplicação de recursos. Como aponta o INE (2009), mais de 50% da população moçambicana não tem acesso aos serviços de saúde..

1.2 Definição de problema

A mortalidade infantil é um problema que afecta grande parte da população, sobretudo nos países com muita pobreza, e corresponde à morte de crianças entre zero e doze meses de vida. Visto que a mortalidade infantil ainda é uma realidade em muitos locais no mundo, fica claro que um dos grandes objectivos do milénio é reduzir esse número (composto pelo número de nascidos e a morte de crianças num local e tempo específico), por meio da implementação de políticas públicas em prol da saúde das mulheres e dos bebés, desde o período de gestação, parto, pós-parto e ainda, priorizando o desenvolvimento da criança até os dois primeiros anos de vida. Segundo dados do Inquérito Demográfico e de Saúde (IDS) de 2011, em Moçambique a taxa de mortalidade infantil era de 64 óbitos por 1.000 nascidos vivos, evidenciando a persistência desse desafio e reforçando a necessidade de intervenções eficazes para a sua redução.

Os estudos sobre a taxa de mortalidade infantil são essenciais para medir e avaliar a qualidade de vida de determinada população, uma vez que refletem, de certa forma, as condições socioeconômicas de uma população. Neste contexto, surge a seguinte questão de pesquisa:

Quais são os factores associados à mortalidade infantil em Moçambique em 2011?

1.3 Objectivos da pesquisa

1.3.1 O objectivo geral

Analisar os factores associados à mortalidade infantil em Moçambique em 2011.

1.3.2 Objectivos específicos

- Descrever as características sócio-demográficas dos indivíduos (crianças menores de 1 ano e suas mães) em estudo;
- Identificar os factores que influenciam na mortalidade infantil em Moçambique;
- Encontrar um modelo que melhor prevê a probabilidade de uma criança morrer.

1.4 Delimitações do estudo

O estudo foca na análise dos factores associados à mortalidade infantil em Moçambique com base nos dados de 2011, sendo que as conclusões estão limitadas às condições e variáveis observadas nesse período. A análise será restrita àquelas características socioeconômicas e de saúde disponíveis naquele momento, e não será possível avaliar mudanças recentes ou novas tendências que possam ter surgido até o último inquérito demográfico e de saúde. Portanto, as conclusões extraídas não devem ser interpretadas como uma reflexão exacta da realidade actual, mas sim como uma base histórica que pode orientar futuros estudos e intervenções.

Por fim, cabe ressaltar que os resultados possivelmente obtidos neste trabalho farão menção à realidade moçambicana, não podendo ser generalizados em relação a outros países. Entretanto, novos estudos poderão ser realizados.

1.5 Relevância do estudo

Há muitas razões que determinam a realização de uma pesquisa, uma delas é a razão racional. Embora existam dados mais recentes sobre mortalidade infantil em Moçambique, o uso dos dados de 2011 na presente análise justifica-se pela sua disponibilidade e pela abrangência das informações colectadas, como variáveis sócio-demográficas, econômicas e estruturais. Esses dados representam uma base confiável e detalhada que permite investigar os factores associados à mortalidade infantil naquela época. Além disso, a análise desses dados fornece uma perspectiva histórica relevante para compreender o contexto e identificar possíveis factores de risco persistentes ao longo do tempo.

A ausência de acesso a dados mais recentes não invalida a relevância do estudo, uma vez que os resultados obtidos com os dados de 2011 podem servir como referência para estudos comparativos futuros, orientando novas pesquisas e intervenções no campo da saúde pública em Moçambique.

Adicionalmente, a formulação da questão de pesquisa: "*Quais são os factores associados à mortalidade infantil em Moçambique em 2011?*" surge da necessidade de suprir lacunas no conhecimento existente sobre este fenómeno. Apesar de dados gerais estarem disponíveis, há uma carência de estudos que analisem de forma integrada variáveis sócio-demográficas, económicas e estruturais para compreender melhor os determinantes da mortalidade infantil. Essa questão é relevante tanto para a comunidade académica, que busca aprofundar o entendimento do tema, quanto para as autoridades de saúde, que necessitam de evidências para orientar políticas públicas mais eficazes.

1.6 Estrutura do Trabalho

Este trabalho compreenderá a seguinte organização: revisão bibliográfica, material e métodos, análise e interpretação de resultados, e por fim a conclusão e as recomendações. No **capítulo II**, na revisão bibliográfica, serão abordadas várias teorias referentes à mortalidade infantil; no **capítulo III**, apresentar-se-á o material e métodos usados para alcançar os objectivos deste estudo; no **capítulo IV**, serão apresentadas as análises e interpretações dos resultados desta pesquisa; e, por último, no **capítulo V**, serão feitas as conclusões e recomendações do estudo.

Capítulo 2

REVISÃO DE LITERATURA

2.1 Referencial teórico

Conforme o INE (2017), define-se a mortalidade infantil como a probabilidade de um nascido vivo morrer antes de atingir o primeiro ano de vida. Este é um dos indicadores adequados para medir o nível de desenvolvimento socioeconómico e o estado de saúde de uma população. Em 2017, a taxa de mortalidade infantil foi de 70,9 óbitos em cada 1 000 nascidos vivos, sendo maior para os homens (73,1 óbitos em cada 1 000 nascidos vivos). Por área de residência, a mortalidade infantil é mais elevada na zona rural (75,0 óbitos em cada 1 000 nascidos vivos) comparativamente à urbana (62,8 óbitos em cada 1 000 nascidos vivos). No período em estudo, ou seja, em 2011, a taxa de mortalidade infantil em Moçambique foi de 71 óbitos por cada 1 000 nascidos vivos, segundo dados do Inquérito Demográfico e de Saúde (IDS) de 2011 (INE, 2012). Este valor representa uma redução significativa em relação a 1997, quando a taxa era de 147 óbitos por 1 000 nascidos vivos, evidenciando avanços nas condições de saúde, embora os níveis ainda se mantenham elevados, sobretudo em zonas rurais.

A mortalidade infantil é definida como a ocorrência de óbitos de menores de um ano de idade em determinada área em um determinado período de tempo. A sua taxa obtém-se dividindo estes óbitos pelo número de nascidos vivos. Para fins de estudo, a mortalidade infantil pode ser separada em duas componentes: mortalidade neonatal e pós-neonatal.

Segundo Bandeira (2004), a mortalidade neonatal refere-se aos óbitos ocorridos entre o nascimento e o vigésimo sétimo dia de vida; pode também ser dividida em neonatal precoce (óbitos ocorridos do nascimento até ao sexto dia de vida) e neonatal tardia (óbitos ocorridos do sétimo ao vigésimo sétimo dia de vida). A mortalidade pós-neonatal refere-se aos óbitos de recém-nascidos que ocorrem entre o vigésimo oitavo dia e antes de completarem um ano de vida.

2.1.1 Mortalidade Infantil no Mundo

O número de crianças que morreram antes de completar cinco anos atingiu um mínimo histórico, caindo para 4,9 milhões em 2022 em todo o mundo, de acordo com as últimas estimativas divulgadas pelo United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME, 2023). O relatório revela que mais crianças estão sobrevivendo hoje no mundo, com a taxa global de mortalidade infantil abaixo dos 5 anos tendo diminuído 51% desde 2000.

Mas os resultados também mostram que, apesar desse progresso, ainda há um longo caminho a percorrer para acabar com todas as mortes de crianças e adolescentes evitáveis. Além dos 4,9 milhões de vidas perdidas antes dos 5 anos – quase metade das quais eram recém-nascidos. A maioria dessas mortes estava concentrada na África Subsaariana e no Sul da Ásia.

2.1.2 Mortalidade Infantil na África

Segundo a UNICEF (2008), a mortalidade infantil na África, especialmente na região subsaariana, continua sendo um desafio significativo de saúde pública. As taxas de mortalidade infantil na África são geralmente mais altas do que em outras regiões do mundo. As principais causas de mortalidade incluem doenças infecciosas como malária, pneumonia, diarreia e HIV/AIDS, além de complicações relacionadas ao parto e prematuridade.

Embora tenha havido progressos notáveis na redução da mortalidade infantil em alguns países africanos, muitos enfrentam desafios persistentes, como acesso limitado a cuidados de saúde de qualidade, nutrição inadequada, falta de saneamento básico e instabilidade política.

Esforços têm sido feitos por organizações internacionais, governos locais e ONGs para abordar esses desafios, incluindo iniciativas para melhorar o acesso a cuidados de saúde básicos, promover a vacinação, fortalecer os sistemas de saúde, promover a educação materna e infantil, bem como melhorar o acesso a água limpa e saneamento básico.

Apesar desses esforços, há uma necessidade contínua de investimento e ação coordenada para enfrentar os determinantes subjacentes da mortalidade infantil na África e garantir que todas as crianças tenham a oportunidade de sobreviver e prosperar.

2.1.3 Mortalidade Infantil em Moçambique

O IDS (2003) revela que a mortalidade infantil em Moçambique situa-se entre 51 por mil em Maputo Cidade e 178 por mil na Província de Cabo Delgado. Outras províncias

com taxas de mortalidade infantil altas são Nampula (164 por mil), Sofala (150 por mil) e Niassa (140 por mil). Os dados mostram que os níveis de mortalidade infantil diferenciam-se de uma província para outra, sendo mais elevados, em média, nas do norte.

2.1.4 Idade da mãe

A idade da mãe é um factor determinante amplamente reconhecido na mortalidade infantil. Estudos indicam que tanto mães adolescentes quanto aquelas com idade avançada apresentam riscos aumentados de desfechos adversos para o recém-nascido, incluindo maior probabilidade de óbito infantil.

Conforme Gage (2008), a idade materna extrema está associada a complicações obstétricas, como parto prematuro e baixo peso ao nascer, que são fatores de risco conhecidos para a mortalidade infantil. Além disso, Gebreegziabher et al. (2023) sugerem que adiar o primeiro parto da adolescência para a idade adulta pode melhorar os resultados do nascimento e reduzir a mortalidade neonatal. A idade da mãe ao ter o filho constitui um importante factor relacionado com os óbitos neonatais e pós-neonatais, sobretudo quando há precocidade ou postergação da maternidade ao longo do período reprodutivo feminino.

Segundo César et al. (2000), a gravidez na adolescência é considerada de alto risco para a mortalidade infantil, em função de aspectos fisiológicos, como peso, estatura, estado nutricional e desenvolvimento do aparelho reprodutivo da mãe.

2.1.5 Estado civil

Segundo o INE (2017), o estado civil ou conjugal é uma característica sócio-demográfica básica das pessoas, que abrange aspetos biológicos, sociais, económicos, legais e, em muitos casos, religiosos. A composição da população segundo esta variável é o resultado de três eventos vitais.

1. O primeiro evento é a união, relativamente permanente, de duas pessoas de sexo oposto, com o propósito de constituir uma família. Quando a união tem um carácter legal, designa-se por casamento ou matrimónio e, quando a união é de facto, designa-se por união marital, ou seja, matrimónio tradicional. Neste sentido, não pode ser considerada como uma simples união consensual, mas sim como um matrimónio tradicional.
2. O segundo evento é a dissolução da união por decisão de um dos cônjuges ou de ambos. Quando a dissolução é legal, designa-se por divórcio e, quando é de facto, por separação.

3. O terceiro evento é o óbito de um dos cônjuges. O cônjuge sobrevivente torna-se viúvo ou viúva. Estes eventos mudam, permanentemente, a distribuição das pessoas por estado civil.

A presença de um companheiro pode representar um suporte emocional e social para gestante. Em relação ao óbito neonatal, a ausência do pai durante a gestação representa um factor de risco na determinação desse desfecho. No estudo realizado no sul do município de São Paulo, os autores verificaram uma forte associação entre a variável sem compromisso e o tempo de relacionamento (menos de um ano) com a mortalidade neonatal precoce (Schoeps et al., 2007).

2.1.6 Escolaridade da mãe ao nascer

A escolaridade da mãe é apontada como indicador socioeconómico para o estudo da mortalidade. O nível de escolaridade é, provavelmente, a variável mais utilizada na análise de diferenciais de mortalidade. De facto, a educação em geral desempenha um papel direto na formação do comportamento em relação às crianças, uma vez que está associada ao nível socioeconómico.

Segundo Sgroi (2008), a baixa escolaridade materna está fortemente associada a indicadores negativos de saúde, incluindo maior risco de mortalidade materna, fetal e infantil, prematuridade, menor intervalo entre gestações, menor duração da amamentação, multiparidade, menor número de consultas pré-natais e baixo peso ao nascer. O índice de baixo peso ao nascer diminui com o aumento do nível de instrução dos pais, especialmente em mães mais instruídas e com boa condição financeira.

Conforme aponta Fuentes (1990), a escolaridade tem influência sobre a mortalidade infantil de uma forma muito característica, pois tem influência tanto direta quanto indireta. Diretamente, contribui para o aumento do conhecimento geral e dos cuidados necessários com a criança. Indiretamente, possibilita a obtenção de emprego com maior renda e conhecimento em planeamento familiar, o que permite a redução da fecundidade e do casamento precoce.

Em relação ao casamento precoce, Mahy et al. (2001), utilizando dados do Inquérito Demográfico e de Saúde de oito países da África Subsaariana (Burkina Faso, Costa do Marfim, Gana, Quênia, Mali, Senegal, Tanzânia e Zimbabué), constataram que as meninas que tinham alcançado o nível secundário eram menos propensas a ter um filho antes dos 18 anos de idade. Em um estudo com mulheres mexicanas, Lindstrom e Brambila Paz (2001) verificaram que um aumento no número de anos de escolaridade reduz a probabilidade de dar à luz ao primeiro filho.

2.1.7 Número de consultas de pré-natal

Segundo o IDS (2011), o acompanhamento pré-natal é definido como o número de visitas pré-natais que a mulher grávida realizou na unidade de saúde, o número de meses de gravidez, a data da primeira visita e o número de doses da vacina antitetânica que a mulher recebeu. Por sua vez, a assistência ao parto é definida pelo tipo de profissional de saúde que assistiu ao parto e pelo local onde ocorreu. Essas informações, combinadas com as taxas de mortalidade neonatal e infantil, podem ser utilizadas para identificar subgrupos de mulheres cujos filhos nascidos vivos estão em risco devido ao não uso desses e outros serviços de saúde. Portanto, essas informações são importantes para o planejamento da ampliação da cobertura dos serviços de saúde.

Segundo a OMS (2016), a assistência pré-natal é um conjunto de procedimentos e medidas que visam diagnosticar, tratar e prevenir situações indesejáveis à saúde da mulher - durante a gravidez, parto, pós-parto - e também, ao bebê. Actualmente, essa assistência é reconhecida como importante estratégia para prevenir ou reduzir o risco de morbimortalidade para a mulher e para a criança.

Segundo a UNICEF (2008), a realização do pré-natal de qualidade aumenta as condições para gestação, parto e pós-parto saudáveis. Porém, a baixa cobertura do pré-natal e da qualidade dos serviços de saúde em algumas regiões do país compromete a saúde da mãe e da criança. O número de consultas pré-natais, também conhecidas como consultas de pré-natal, refere-se ao total de visitas que uma mulher grávida faz a um profissional de saúde durante o período de gestação para receber cuidados médicos e acompanhamento. Essas consultas são vitais para garantir a saúde da mãe e do bebê, bem como para detectar e tratar quaisquer problemas ou complicações que possam surgir durante a gravidez. O número de consultas pré-natais pode variar de acordo com as diretrizes médicas do país e da organização de saúde, mas geralmente inclui várias visitas ao longo dos nove meses de gestação.

O objectivo das consultas pré-natais é acompanhar o desenvolvimento do feto, garantir que a mãe esteja saudável, rastrear problemas de saúde, discutir preocupações da gestante e fornecer orientações sobre dieta, exercícios e preparação para o parto.

Recomenda-se que as mulheres grávidas sigam o plano de cuidados pré-natais recomendado por seus médicos ou profissionais de saúde, que inclui o número e a programação das consultas pré-natais. Normalmente, isso envolve cerca de 10 a 15 consultas ao longo da gravidez, com consultas mais frequentes no início e menos frequentes à medida que a gravidez avança. O objectivo é garantir que a mãe e o bebê recebam a atenção e os cuidados necessários durante toda a gestação.

2.1.8 Área de Residência

A área de residência (urbana ou rural) da mulher afecta indirectamente a mortalidade infantil pela determinação do seu acesso a serviços de atenção à saúde (Mahy, 2003). Na área urbana, a mulher tem maior probabilidade de encontrar um meio de transporte aos postos de serviços de saúde, de adquirir recursos monetários para custear medicamentos, assim como maior oportunidade para a educação e habilidade de alocar, no domicílio, recursos necessários para a saúde das crianças. A "área de residência" se refere à localização geográfica onde uma pessoa vive ou reside. É a região, bairro, cidade ou qualquer outra área geográfica específica onde uma pessoa tem sua casa ou moradia. A área de residência de uma pessoa é o lugar onde ela estabelece sua base e onde geralmente passa a maior parte do seu tempo.

A área de residência de uma pessoa pode variar amplamente, dependendo de factores como preferências pessoais, trabalho, família e outros compromissos. É um conceito importante em várias áreas, incluindo planeamento urbano, demografia, pesquisa de mercado, educação e assistência médica, uma vez que a localização geográfica de uma pessoa pode afectar vários aspectos de sua vida, incluindo acesso a serviços, qualidade de vida e oportunidades.

2.1.9 Ordem de Nascimento

Ordem do nascimento é definida como a posição ocupada por uma pessoa conforme a sua idade entre seus irmãos e irmãs. As crianças de primeira ordem e as de ordem superior têm uma maior chance de morrer em relação as outras, tal como para a idade da mãe, também a relação entre a mortalidade da criança e a paridade, se pode traduzir numa curva em U assimétrica, onde a mortalidade é mais superior nos valores extremos de paridade, (Alberto, 2010).

A "ordem de nascimento" refere-se à posição que uma pessoa ocupa em relação aos seus irmãos em uma família, com base na sequência de seus nascimentos. Geralmente, as pessoas são categorizadas como o filho mais velho, o segundo filho, o terceiro filho, e assim por diante. A ordem de nascimento é um factor que pode influenciar diversos aspectos do desenvolvimento e da personalidade de uma pessoa, de acordo com alguns estudos e teorias.

2.1.10 Religião

Segundo INE (2007), define religião como sendo um conjunto de sistemas culturais e de crenças que estabelecem os símbolos que relacionam a humanidade com a espiritualidade e os valores morais.

Segundo Wood et al. (2007) em estudo realizado na Região Nordeste do Brasil, identificaram que filhos de mães protestantes tradicionais (batistas e presbiterianas) e pentecostais possuíam chances menores de experimentarem mortalidade na infância do que filhos de mães católicas. Nesse estudo, em comparação com mães católicas, as mães protestantes tradicionais apresentavam níveis mais elevados de educação, renda familiar, além de serem mais propensas a estarem em união estável e a terem acesso a certas facilidades domésticas, como água encanada. Nesse mesmo sentido, Verona et al (2010) avaliaram a associação entre envolvimento religioso materno e mortalidade infantil no Brasil, segundo os dados da Pesquisa Nacional de Demografia e Saúde (PNDS), e constatou-se que níveis mais altos de religiosidade estavam relacionados com a redução das taxas de mortalidade infantil (Verone et al, 2010).

A Figura 2.1 apresenta um modelo baseado na teoria de Smith (2003) e adaptado por Verona (2011) de como a influência da religião pode afectar directamente e indirectamente o comportamento sexual de adolescentes, por meio de efeitos directos e indirectos.

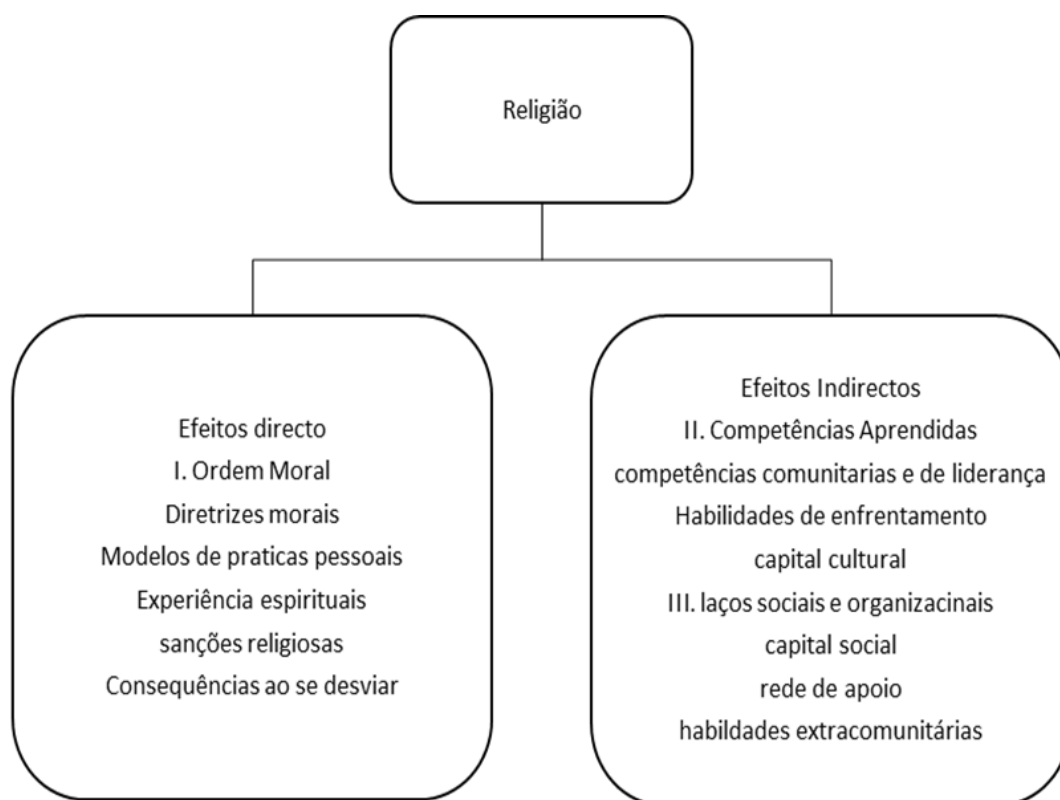


Figura 2.1: Modelo conceptual de como a religião pode afectar directamente e indirectamente o comportamento dos adolescentes

Fonte: Adaptado de Verona, 2011.

Os efeitos directos de ordem moral estão compostos pelas:

1. Diretrizes morais: as religiões promovem normas morais e culturais específicas de

autocontrole e virtude pessoal, no intuito de transmitir e internalizar as ordens morais aos membros para que usem como um guia na escolha de vida e no compromisso moral (Smith, 2003).

2. **Modelos de práticas pessoais:** as religiões podem fornecer aos membros exemplos de práticas de vida daqueles que seguem o padrão das ordens morais religiosas (Smith, 2003).
3. **Experiências espirituais:** as religiões proporcionam as experiências espirituais concebidas pelos membros na conversão, na resposta à uma oração, a sensação de profunda paz espiritual (Smith, 2003).
4. **Sanções religiosas:** consiste na repreensão ou disciplina por parte do representante religioso aos membros que não seguirem as normas religiosas (Verona, 2011).
5. **Consequências ao se desviar:** os membros que violam as normas religiosas poderão sentir ou gerar consequências psicológicas pelo desvio, por exemplo, o sentimento de culpa, constrangimento público, ou a expectativa de punição divina (Verona, 2011).

Já os efeitos indirectos de competências aprendidas e laços sociais e organizacionais estão compostos pelas:

1. **Competências comunitárias e de liderança:** as religiões fornecem contextos organizacionais para que os membros possam aprender e praticar habilidades valiosas da vida comunitária e de liderança (Smith, 2003).
2. **Habilidade de enfrentamento:** as religiões promovem uma variedade de crenças e práticas aos membros que auxiliam a lidarem melhor com o estresse provocado por situações psicológicas difíceis, para processar emoções difíceis e resolver conflitos interpessoais, na busca de aumentar o bem-estar (Smith, 2003).
3. **Capital cultural:** as religiões proporcionam oportunidades alternativas (além da família, da escola e da mídia) para adquirir elementos de capital cultural que podem melhorar directamente o bem-estar (Smith, 2003).
4. **Capital social:** a religião é uma das poucas instituições sociais que não há separação por idade e que enfatiza as interações pessoais ao longo do tempo, proporcionando ligações de rede entre gerações. O que possibilita uma troca de informações, recursos e oportunidades úteis entre as gerações (Smith, 2003).
5. **Rede de apoio:** as congregações religiosas podem fornecer redes relativamente densas de laços relacionais, por exemplo, pessoas que tem a função de desencorajar práticas negativas e ou encorajar práticas positivas dos membros mais jovens (Smith, 2003).

6. **Habilidades extracomunitárias:** as religiões incluem encontros para organizações religiosas nacionais e transnacionais, que proporciona aos membros experiências positivas e eventos além de suas comunidades locais (Smith, 2003).

Esse modelo realizado por Smith (2003) e com adaptação de Verona (2011) pode ser utilizado para explicar a influência que a religião pode exercer sobre a mortalidade na infância, sendo considerados apenas os efeitos directos (diretrizes morais e modelos de práticas pessoais) e os efeitos indirectos (habilidades de enfrentamento, capital cultural, capital social e rede de apoio).

2.1.11 Sexo do recém-nascido

O sexo feminino apresenta o amadurecimento mais rápido dos pulmões, sendo considerado um factor protector em relação às complicações respiratórias, que são consideradas menores em relação aos recém-nascidos do sexo masculino (Ribeiro et al., 2016).

Segundo Araújo et al. (2000), o risco do recém-nascido do sexo masculino morrer no período neonatal precoce foi de 4,16 vezes maior do que os do sexo feminino.

O sexo da criança é um dos factores que deve ser considerado quando se aborda a questão da mortalidade de crianças menores de 5 anos. As estatísticas sobre mortalidade, publicadas em vários países europeus no século XX mostraram evidências da influência do sexo do recém-nascido na mortalidade infantil (Remoaldo, 2002).

O sexo de um recém-nascido refere-se à sua classificação biológica como masculino ou feminino, com base em características sexuais primárias, como genitália externa, cromossomos sexuais e características secundárias associadas ao desenvolvimento sexual.

2.1.12 Peso ao nascer

O peso ao nascer é um importante indicador de saúde neonatal e reflecte o estado de nutrição fetal, crescimento intrauterino e o resultado de uma série de factores que podem influenciar o desenvolvimento do recém-nascido. (OMS, 2014)

Segundo Barker et al. (1993), o peso ao nascer, aferido na primeira hora após o nascimento, é um parâmetro usado mundialmente para avaliar as condições de saúde do recém-nascido, alertando os profissionais de saúde sobre seu risco de morbimortalidade. Ele reflecte as condições nutricionais da gestante e do neonato e tem influência directa no crescimento e desenvolvimento da criança e nas condições de saúde do indivíduo na vida adulta. O peso ao nascer é um parâmetro que é usado para avaliar as condições de saúde do recém-nascido. O peso ao nascer é também um dos factores que tem sido focado nos estudos da mortalidade infantil. Existe a preocupação com os recém-nascidos com peso igual ou superior a 4000 gramas, mas esses casos não são muito comuns. A influência deste fator tem sido condicionada pelo nível de desenvolvimento de cada região ou país.

2.1.13 Tipos de parto

O parto, que é considerado o momento mais importante de toda a gravidez, deve ser precedido de um pré-natal bastante cuidadoso. Conhecer as condições de desenvolvimento do bebê é essencial para saber como o nascimento poderá ser conduzido. De evento fisiológico, familiar e social, o parto/nascimento transforma-se em acto médico, no qual o risco de patologias e complicações se torna a regra e não a exceção (Maia, 2010)

Existem dois tipos principais de parto: **o normal**, em que tem-se a forma natural de se dar à luz, e a **cesariana**, em que se realiza um corte no abdômen para a retirada do bebê. O parto normal é o procedimento no qual o feto nasce por via vaginal facilitando o contacto entre mãe e filho, resultando em menores taxas de desconforto respiratórios e menor risco de infecção puerperal (Ferrari, Carvalhes e Parada, 2016).

A indicação do tipo de parto deve ser baseada em motivos clínicos consistentes e em situações específicas. **A cesárea** é um procedimento cirúrgico e tem como finalidade intervir quando riscos são maiores diante dos benefícios do parto normal, devendo ser indicada, apenas em casos necessários (OMS, 2015).

2.1.14 Quintil de riqueza

É uma divisão da população em cinco partes iguais (20% cada), organizadas de acordo com a distribuição de riqueza ou renda. Essa categorização permite analisar a desigualdade econômica, identificando desde os 20% mais pobres até os 20% mais ricos. Cada quintil representa um segmento específico da população:

- **Primeiro Quintil (20% mais pobres):** Este grupo inclui os indivíduos com menor acesso a recursos financeiros, muitas vezes vivendo em situação de pobreza ou extrema pobreza. Eles enfrentam dificuldades para atender a necessidades básicas, como alimentação, moradia e saúde. Políticas públicas de transferência de renda e programas sociais são frequentemente direcionados a esse grupo para reduzir desigualdades e melhorar suas condições de vida.
- **Segundo Quintil (20% a 40%):** O segundo quintil representa pessoas com renda um pouco acima dos mais pobres, mas que ainda são consideradas vulneráveis. Embora consigam atender a necessidades básicas, elas têm pouco ou nenhum acesso a bens duráveis e enfrentam dificuldades em caso de imprevistos financeiros. Esse grupo pode ascender à classe média com melhores oportunidades de emprego e educação.
- **Terceiro Quintil (40% a 60%):** Conhecido como a classe média baixa ou intermediária, esse grupo tem uma renda mais estável e consegue acessar serviços básicos

com maior facilidade. Eles conseguem pagar por educação, saúde e lazer, mas ainda têm limitações para acumular riqueza. Esse quintil é fundamental para a estabilidade econômica, pois representa uma grande parcela da população consumidora.

- **Quarto Quintil (60% a 80%):** O quarto quintil inclui indivíduos com renda acima da média, que conseguem adquirir bens duráveis, como carros e imóveis, e têm capacidade de poupar parte de sua renda. Esse grupo é considerado parte da classe média alta e contribui significativamente para o consumo e o investimento na economia.
- **Quinto Quintil (20% mais ricos):** O quinto quintil representa os 20% mais ricos da população, que concentram a maior parte da riqueza e da renda. Eles têm acesso a bens de luxo, investimentos e uma alta capacidade de poupança. Esse grupo é frequentemente analisado para entender a concentração de riqueza e a desigualdade econômica, já que, em muitos países, os mais ricos detêm uma parcela desproporcional dos recursos.

2.2 Análise de Regressão

Segundo Mestre (2009), o termo “regressão” foi proposto pela primeira vez por Sir Francis Galton em 1885, num estudo onde demonstrou que a altura dos filhos não tende a reflectir a altura dos pais, mas tende sim a regredir para a média da população.

O termo “análise de regressão” define um conjunto vasto de técnicas estatísticas usadas para modelar as relações entre variáveis e prever o valor de uma ou mais variáveis dependentes (ou de resposta) a partir de um conjunto de variáveis independentes (ou explicativas). Existem várias metodologias estatísticas que permitem explicar ou descrever a relação entre uma variável de interesse (resposta) e uma ou mais variáveis (explicativas). O modelo linear é talvez o mais utilizado para modelar esta relação. Este modelo assume, entre outras, que o valor esperado da variável resposta é uma combinação linear das variáveis explicativas e que a variável resposta segue a distribuição normal (Hill, 2005).

2.3 Modelos Lineares Generalizados

De acordo com Montenegro (2009), o Modelo Linear Generalizado (MLG) é um desenvolvimento dos modelos lineares para acomodar a não normalidade na distribuição da variável resposta e transformações para linearidade.

Segundo Figueira (2006), MLG é especificado por três componentes: uma componente aleatória, a qual identifica a distribuição de probabilidade da variável dependente, uma componente sistemática, que especifica uma função linear entre as variáveis independentes e uma função de ligação que descreve relação matemática entre a componente sistemática e o valor esperado da componente.

De acordo com Montenegro (2009), a componente aleatória de um MLG consiste nas observações da variável aleatória Y , com observações independentes $(y_1, \dots, y_n)$ de uma distribuição na família exponencial, Essa família tem função densidade de probabilidade na forma:

$$f(y_i, \theta_i) = a(\theta_i)b(y_i) \exp[y_i Q(\theta_i)] \quad (2.1)$$

O valor do parâmetro θ_i pode variar de 1 até n , dependendo do valor das variáveis explanatórias, o termo $Q(\theta_i)$ é chamado de parâmetro natural.

A componente sistemática do MLG é definida através de um vector $\eta = X\beta$, onde X é uma matriz que consiste nas variáveis independentes das n observações e β é um vector dos parâmetros do modelo.

A função de ligação conecta os valores esperados das observações as variáveis explanatórias. Isso porque $\mu_i = E(Y_i|x_i)$, com $i = 1, \dots, n$ então $\eta_i = g(\mu_i)$. E,

$$g(\mu_i) = \beta_0 + \sum_{j=1}^n \beta_j x_{ij} \quad (2.2)$$

Segundo Figueira (2006), existem duas classes importantes de modelos lineares generalizados (MLG): os baseados na distribuição Bernoulli, que originam os modelos *logit* (regressão logística) quando a variável resposta é binária; e os baseados na distribuição de Poisson, que originam os modelos *log-linear* quando a variável resposta é de contagem.

Modelo logístico (logit). Seja $\pi(x) = P(Y = 1 | X = x)$. O logit é dado por

$$\text{logit}(\pi(x)) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x. \quad (2.3)$$

Daí

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}. \quad (2.4)$$

As odds em x são

$$\text{odds}(x) = \frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0 + \beta_1 x}, \quad (2.5)$$

e a razão de odds para um acréscimo unitário em X é

$$\text{OR}_{x \rightarrow x+1} = \frac{\text{odds}(x+1)}{\text{odds}(x)} = e^{\beta_1}. \quad (2.6)$$

Note que $OR = 1 \Leftrightarrow \beta_1 = 0$ indica independência entre X e Y .

Modelo log-linear (Poisson). Para contagens, suponha $Y \sim \text{Poisson}(\mu(x))$ com

$$\log(\mu(x)) = \beta_0 + \beta_1 x. \quad (2.7)$$

Logo,

$$\mu(x) = e^{\beta_0 + \beta_1 x}, \quad (2.8)$$

e a razão de taxas para um aumento unitário em X é

$$RR_{x \rightarrow x+1} = \frac{\mu(x+1)}{\mu(x)} = e^{\beta_1}. \quad (2.9)$$

Ao modelar taxas com diferentes exposições t_i , inclua o offset $\log(t_i)$:

$$\log(\mu_i) = \log(t_i) + \beta_0 + \beta_1 x_i, \quad (2.10)$$

de modo que $\mu_i/t_i = \exp(\beta_0 + \beta_1 x_i)$.

De acordo com Farhat (2003), podemos notar que quando o X esta variando de $-\infty$ a $+\infty$, $g(x_i)$ apresenta a mesma variação.

2.4 Modelo de Regressão Logística

Segundo Hill (2005), a regressão logística é uma técnica de estatística multivariada que se enquadra no MLGs, que é usado quando se deseja explicar uma variável resposta categórica em função de uma ou mais variáveis explicativas quantitativas ou qualitativas.

Segundo Paula (2004) e Dobson (1990), modelo de regressão logístico é uma generalização dos modelos lineares para variáveis de duas categorias ou que de alguma maneira foram dicotomizadas assumindo valores 0 para o fracasso ou 1 sucesso.

Gonçalves (2013), concordando com estes autores, acrescenta ainda que a regressão logística é utilizada em situações nas quais pretende-se saber se o indivíduo possui ou não determinada característica em estudo.

Segundo Hosmer e Lemeshow (1989), regressão logística é um método padrão de análise de regressão para variáveis medidas de forma dicotômica. Especialmente nas áreas de saúde. Os métodos de regressão são uma componente integral de qualquer análise de dados quando se trata de descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas. Quando a variável resposta é categórica com duas ou mais categorias os modelos de regressão logística tornam se o método padrão na análise desse tipo de

situação.

2.5 Análise de dados amostrais complexos

Segundo Morais (2008), a inferência estatística, em sua abordagem clássica, fundamenta-se na amostra aleatória simples, este que é um método que requer que cada membro da população tenha uma chance igual e independente de ser seleccionado. Entretanto, a maioria dos inquéritos de saúde não usa amostragem aleatória simples, em parte por restrições orçamentais, em parte por limites de tempo associados à colecta de uma grande quantidade de informações ao longo de um território geográfico grande. Em decorrência disso, outros métodos de amostragem probabilísticos são geralmente utilizados nos inquéritos de base populacional, como a amostragem estratificada e a amostragem por conglomerados em múltiplos estágios com probabilidades desiguais de selecção, para prover uma amostra representativa da população em tempo hábil e de acordo com o orçamento previsto.

Segundo Lumley (2010), a combinação de métodos probabilísticos de amostragem para selecção de uma amostra representativa da população, é chamada de desenho complexo de amostragem. Resultante deste processo, a preocupação subsequente é a análise de dados provenientes de amostras complexas.

Cabe ressaltar ainda que para efectuar a análise de dados provenientes de inquéritos amostrais complexos, é preciso incorporar as informações do plano amostral da pesquisa, informando os pesos das unidades da amostra e as informações estruturais da pesquisa, identificando o estrato e pelo menos as unidades primárias de amostragem.

Segundo Morais (2008), as estimativas pontuais são influenciadas por pesos distintos associados às unidades da amostra, enquanto as estimativas de variância são influenciadas pela estratificação, conglomeração e pesos amostrais. No caso da regressão logística, o método de máxima pseudo-verossimilhança produz estimativas levando em consideração as características do plano amostral.

Mesmo que o estimador encontrado na máxima pseudo-verossimilhança possa ser em algumas situações igual ao de máxima verossimilhança, a estimativa da variância pelo máxima pseudo-verossimilhança é preferível em relação ao método que não leva em consideração o plano amostral. Este procedimento fornece estimativas mais robustas para a realização de inferência para a população (Pessoa e Silva, 1998).

Segundo Pessoa e Silva (1998), as dificuldades na aplicação do método de máxima pseudo

verossimilhança é a necessidade de conhecimento das características do plano amostral e o facto de que suas propriedades não são conhecidas para pequenas amostras, o que impossibilita algumas análises estatísticas. Para situações em que os dados são provenientes de amostras complexas, o estimador de máxima pseudo-verossimilhança é obtido a partir de correções que levam em consideração os pesos amostrais. Para o estimador de máxima pseudo-verossimilhança de BETA é necessária uma ponderação. Assim tem-se:

Os parâmetros são estimados por métodos numéricos iterativos como já exemplificado na secção anterior.

Para situações em que os dados são colectados por planos amostrais complexos, não existe normalmente uma forma analítica directa para calcular o erro-padrão de estimadores como de coeficientes de modelos de regressão logística. Desta forma, são utilizados métodos, tais como a linearização de Taylor (também chamado de método delta) e os métodos de replicação de Jackknife e bootstrap (Wolter, 1985).

O método da linearização de Taylor vem sendo utilizado com frequência para se obter aproximações de estimadores não - lineares por estimadores lineares. Este método tem como base, a expansão em séries de Taylor do estimador pontual do parâmetro de interesse, em torno do verdadeiro parâmetro e considera apenas as aproximações de primeira ordem.

2.6 Vantagens do modelo de regressão logística

Segundo Mestre (2009), a regressão logística tem as seguintes vantagens:

- Facilidade para lidar com variáveis independentes categóricas;
- Fornece resultados em termos de probabilidade, o que facilita a interpretação;
- Facilidade de classificação de indivíduos em categorias;
- Requer pequeno número de suposições;
- Alto grau de confiabilidade.

De acordo com Agresti (2007), os modelos de regressão logística podem ser classificados segundo o número de variável(is) independente(s). Quando a variável independente é única sendo esta qualitativa ou quantitativa então pode se estabelecer um **Modelo de Regressão Logística Simples**, mas se o modelo contiver mais de uma variável qualitativa ou quantitativa o modelo passa ser **Modelo de Regressão Logística Múltiplo**.

2.7 Comparação de modelo de Regressão logística e o modelo de Regressão linear

Tabela 2.1: Comparação de modelo de regressão logística e de regressão linear simples

Regressão linear simples	Regressão logística
$E(Y/X = x_i) = \beta_0 + \beta_1 x_i$	$E(Y/X = x_i) = \pi_i$
$-\infty < (Y/X = x_i) < +\infty$	$0 \leq (Y/X = x_i) \leq 1$
$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$	$Y_i = \pi_i + \varepsilon_i$
$\varepsilon_i \sim N(0, \sigma^2)$	$\varepsilon_i = \begin{cases} 1 - \pi_i \text{com } P(Y = 1/X = x_i) \\ -\pi_i \text{com } P(Y = 0/X = x_i) \end{cases}$ $E(\varepsilon_i) = 0$ $Var(\varepsilon_i) = \pi_i(1 - \pi_i)$
$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$	$y_i \sim Ber(\pi_i)$

Fonte: autor

2.8 Procedimentos de Análise

Na análise dos dados, a Príncipe fez-se à **estatística descritiva** que conforme Johnson & Bhattacharyya (2010), é área da estatística que se baseia na organização, resumo e apresentação de dados para descrever suas características principais. Ela inclui técnicas para calcular medidas de tendência central, medidas de dispersão e criar representações visuais dos dados, como gráficos e tabelas de frequência.

De acordo com Mulenga (2018), a **estatística descritiva** é o ramo ou parte da estatística cujo objectivo é a observação de fenómenos de mesma natureza, recolha, organização, classificação, análise e interpretação de dados sem deixar de calcular algumas medidas (estatísticas), que permitem resumidamente descrever o fenómeno estudado. A estatística descritiva tem como objectivo fornecer uma visão clara e concisa dos dados, permitindo que pesquisadores e analistas compreendam melhor a distribuição e a estrutura dos dados colectados.

A seguir estimou-se os coeficientes do modelo, para analisar a influência das variáveis independentes. A posterior o modelo logístico reduzido, que é aquele que contém apenas a variável dependente, e onde usei o teste do wald para verificar se cada parâmetro ou coeficiente estimado é significativamente diferente de zero.

Após a estimação do modelo nulo, procedi com o teste de razão verossimilhança que é usado para estimar os parâmetros do modelo de regressão de forma geral. Com base no

p-valor e nível de significância de 5% verifica-se à existe alguma associação entre a Desfecho e as variáveis independentes, e se os dados estão bons para fazer a análise.

Para a avaliação do ajuste do modelo usei o Pseudo R^2 que é um valor que varia de 0 à 1 e informa o quanto as variáveis independentes explicam as variações da variável dependente, e quanto maior melhor.

Para verificar se o modelo consegue explicar adequadamente a variável dependente usei o teste de Hosmer e Lemeshow e para verificar se o teste explica ou não o p-valor deve ser acima de 5%. e por fim a tabela de classificação para compreender se o modelo consegue explicar bem o comportamento das variáveis independentes.

Capítulo 3

MATERIAL E METÓDOS

Neste capítulo será apresentado a classificação do presente estudo, o material e os procedimentos técnicos aplicados para a obtenção dos resultados.

3.1 Classificação de pesquisa

3.1.1 Quanto à natureza da pesquisa

Tal como se pode depreender a partir dos objectivos propostos a alcançar, quanto à finalidade, esta é uma **pesquisa aplicada**, porque o estudo é aplicado para a resolução de problemas específicos. Segundo Prodanov e Freitas (2013) pesquisa aplicada é aquela cujo objectiva gerar conhecimentos para aplicação prática dirigidos a solução de problemas específicos envolve verdades e interesses locais.

3.1.2 Quanto aos procedimentos técnicos

Do ponto de vista dos Procedimentos Técnicos a pesquisa realizada pode ser classificada como sendo: **pesquisa bibliográfica**.

Segundo Gil (2008), **Pesquisa Bibliográfica** é aquela desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos. Não recomenda-se trabalhos oriundos da internet.

3.1.3 Quanto à abordagem ou método

Quanto ao ponto de vista da forma de abordagem do problema a pesquisa realizada pode ser classificada como **quantitativa**.

De acordo com Minayo (2008), os métodos quantitativos têm o objectivo de mostrar dados, indicadores e tendências observáveis, ou produzir modelos teóricos abstractos com elevada aplicabilidade prática. Suas investigações evidenciam a regularidade dos fenómenos. No

que tange a abordagem, esta pesquisa é quantitativa na medida em que há quantificação na colecta dos dados e no tratamento deles, por meio de procedimentos estatísticos.

3.1.4 Quanto aos objectivos

Em relação aos objectivos, a presente pesquisa é classificada como explicativa, pois, têm como objectivo central identificar os factores que determinam ou que contribuem para que o fenómeno ocorra. É a pesquisa que mais aprofunda o conhecimento da realidade, pois explica a razão e o porquê das coisas. Por isso mesmo é o tipo mais complexo e delicado já que o risco de cometer erros aumenta bastante. (Gil, 2002)

3.2 ÁREA DE ESTUDO

Moçambique é um país da costa oriental da África Austral que tem como limites: a norte, a Tanzânia; a noroeste, o Malawi e a Zâmbia; a oeste, o Zimbábwe, a África do Sul e a Suazilândia; a sul, a África do Sul; a leste, a secção do Oceano Índico designada por Canal de Moçambique. No Canal de Moçambique, os vizinhos são Madagáscar e as Comores (incluindo a possessão francesa de Mayotte). No Oceano Índico, para leste da grande ilha de Madagáscar, situam-se as dependências de Reunião, Juan de Nova e Ilha Europa. No Canal de Moçambique, sensivelmente a meia distância entre o continente e Madagáscar, o atol de Bassas da Índia, igualmente possessão francesa. A capital de Moçambique é Maputo (foi chamada por Lourenço Marques durante a dominação portuguesa). A metade norte (a norte do rio Zambeze) é um grande planalto, com uma pequena planície costeira bordejada de recifes de coral, limitando no interior com maciços montanhosos pertencentes ao sistema do Grande Vale do Rift. A metade sul é caracterizada por uma larga planície costeira de aluvião, coberta por savanas e cortada pelos vales de vários rios, o mais importante dos quais é o rio Limpopo.

A superfície do território moçambicano é de 799,380 km², e uma população total de 27.909.799 habitantes, sendo estes cerca de 52% do sexo feminino e os restantes 48% do sexo masculino (INE, 2017).

O país está dividido em 11 províncias: ao Norte, estão as Províncias do Niassa, Cabo Delgado e Nampula, no Centro encontram-se as de Zambézia, Tete, Manica e Sofala e ao Sul, Inhambane, Gaza, Maputo Província e Maputo Cidade (Veja o Mapa 3.1). O território moçambicano, como toda a região Austral do Continente Africano, não apresenta grande variedade de paisagem. Da costa para o interior podem-se distinguir três tipos de relevos:

- **A planície do litoral** que ocupa a grande parte do território (40 por cento). Esta é a região natural onde se observa a maior concentração da população;
- **Os planaltos:** com altitudes que variam entre 200 e 1.000 metros;

- **Os grandes planaltos e montanhas:** que ocupam uma pequena parte do território nacional, com altitudes superiores a 1.000 metros. Do ponto de vista da distribuição geográfica da população, já que não constituem uma superfície contínua, não oferecem grandes obstáculos para assentamentos humanos.



Figura 3.1: Mapa de Moçambique

3.3 MATERIAL

Para a realização deste estudo, recorreu-se a uma base de dados secundária proveniente do Inquérito Demográfico e de Saúde (IDS) realizado em Moçambique no período

de 2011–2012, identificado como **MZKR62FL**. O inquérito foi conduzido pelo Instituto Nacional de Estatística (INE), em colaboração com o Ministério da Saúde (MISAU), contando com apoio técnico da ICF International no âmbito do programa Demographic and Health Surveys (DHS), e financiamento da Agência dos Estados Unidos para o Desenvolvimento Internacional (USAID).

O ficheiro original compreendia **982** variáveis e **11.102** observações. Após a selecção criteriosa das variáveis relevantes ao objectivo do estudo, foram retidas 13 variáveis principais. Seguiu-se um processo de limpeza, organização e tratamento estatístico dos dados, resultando num conjunto final de 7.076 observações válidas para análise.

A descrição pormenorizada das variáveis seleccionadas encontra-se apresentada na Tabela. No que se refere ao seu planeamento amostral, pode-se considerar como um estudo transversal que utiliza um plano amostral complexo que envolve estratificação, multi-etápica, com probabilidades iguais de selecção que resultam em pesos amostrais distintos para as unidades amostrais. A concepção da amostra é composta por três etapas. Na primeira etapa foram seleccionadas 611 Unidades Primárias de Amostragem (UPA) com probabilidade proporcional à dimensão, sendo medida de tamanho o número de agregados familiares em cada estrato dentro de cada província.

Na segunda etapa de amostra, foram seleccionados com probabilidades iguais 20 agregados familiares nas UPA urbanas e 25 agregados familiares nas Unidades Primárias de Amostragem rurais. Esta selecção foi realizada após uma listagem prévia de agregados familiares. Finalmente, na terceira etapa, foi realizada uma amostragem exaustiva, isto é, foram recolhidos dados demográficos e de saúde de todas as mulheres de 15-49 anos e crianças menores de 5 anos encontradas nos agregados familiares seleccionados.

Em cada área de enumeração seleccionada foi feita uma listagem de estruturas e sua classificação em residenciais e não residenciais. Apenas as estruturas residenciais, foram utilizadas para a selecção da amostra em cada área de enumeração. Em cada província foram seleccionadas 51-55 Unidades Primárias de Amostragem, com a excepção de Nam-pula (60 UPA), Zambézia e Sofala (58 UPA cada), Maputo Província (61 UPA) e Maputo Cidade (65 UPA), devido a maior variabilidade nas características sócio-demográficas e indicadores de cobertura baixos para mulheres 15-49 anos e crianças menores 5 anos para estas províncias.

3.3.1 Balanceamento de Classes

Devido ao desbalanceamento significativo observado na variável dependente com apenas 363 casos de sucesso (5,1%) contra 6.713 casos de insucesso (94,9%) optou-se pela aplicação da técnica SMOTE (Synthetic Minority Over-sampling Technique).

Esta abordagem sintetizou novos casos da classe minoritária através de interpolação entre exemplos reais existentes, resultando numa distribuição balanceada de 6.713 casos para cada classe (50%-50%), totalizando 13.426 observações para análise.

A decisão fundamenta-se na necessidade de evitar viés do modelo contra a classe minoritária e melhorar a capacidade de identificação de padrões preditivos relevantes.

Descrição das Variáveis em Estudo

Tabela 3.1: Descrição das Variáveis e Codificação das Categorias

Descrição da Variável	Codificação / Categoria
Óbito**	0 - Não; 1 - Sim;
Região de nascimento da mãe	1 - Norte; 2 - Centro; 3 - Sul;
Local de residência	1 - Rural; 2 - Urbano;
Religião da mãe	1 - Católica; 2 - Islâmico; 3 - Zione; 4 - Evangélica/Pentecostal; 5 - Anglicano; 6 -Sem religião; 7 - Protestante; 96 -Outras ;
Escolaridade da mãe	0 - Sem educação; 1 - Primário incompleto; 2 - Primário completo; 3 - Secundária incompleta; 4 - Secundária completa; 5 - Superior;
Quintil de riqueza	1 - Primeiro Quintil; 2 - Segundo Quintil; 3 - Terceiro Quintil; 4 - Quarto Quintil; 5 - Quinto Quintil;
Estado civil da mãe	1 - Solteiro(a); 2 - Casado(a); 3 - Cônjuge; 4 - Viúvo(a); 5 - Divorciado(a); 6 - Separado(a)/ nao vive mais junto;
Sexo da criança	1 - Masculino; 2 - Feminino;
Número de consultas Pre-natais	0 - Nenhuma; 1 - 1 a 3; 2 - 4 a 6; 3 - 7+;
Tipo de parto	1 - Normal; 2 - Cesariano ou Vaginal;
Tamanho da criança ao nascer	1 - Pequeno; 2 - Média; 3 - Grande;
Faixa etária da mãe	1 - 15-19; 2 - 20-24; 3 - 25-29; 4 - 30-34; 5 - 35-39; 6 - 40-44; 7 - 45-49;
Ordem de nascimento	1 - Primogênito; 2 - 2 ou 3; 3 - 4 ou ordem superior;

Óbito** é a variável dependente (Y), e as restantes variáveis são independentes (X).

Categorização da variável resposta

Ao estimar o modelo de regressão logística binária, a variável resposta foi categorizada da seguinte forma para estudar os factores associados à morte infantil em Moçambique:

$$\begin{cases} 1 & \textbf{Sim}, \text{ se ocorreu morte infantil durante o período considerado} \\ 0 & \textbf{Não}, \text{ se não ocorreu morte infantil durante o período considerado} \end{cases}$$

Essa categorização permite investigar os factores que aumentam ou diminuem a probabilidade de ocorrência de morte infantil, utilizando dados quantitativos para identificar associações significativas e propor intervenções eficazes.

Todas as análises foram feitas à um nível de significância de 5% e para efeitos de avaliação

da regra de decisão, usou-se um p-valor associado à estatística do teste.

Os dados foram processados através do software estatísticos R versão 3.6.2, Python e SPSS, para a realização de gráficos foi usado Microsoft Excel 2019 o relatório foi produzido com recurso aos programas Latex e Word.

3.4 MÉTODOS

Marconi & Lakatos (1985) define método como o conjunto de procedimentos sistemáticos e racionais que permitem alcançar os objectivos da pesquisa, levando em consideração aspectos de segurança, economia e validade. Esse método, seria o responsável por traçar o caminho a ser seguido pelo cientista, ajudando-o nas decisões e na detecção de erros. De acordo com Mesquita (2014), os pressupostos da regressão logística são:

- Ausência de Multicolinearidade.
- A soma dos valores estimados é igual a soma dos valores;
- A soma dos erros é igual a zero;
- A soma das observações multiplicada pelo erro é zero;
- A recta dos mínimos quadrados passa pelo ponto (x,y);

3.5 Tratamento de Dados

3.5.1 Teste de independência

Segundo Agresti & Finlay (2009), utiliza-se o teste de independência quando os dados da pesquisa se apresentam sob forma de frequências em categorias discretas. Para se aplicar este teste de hipótese, recorre-se à estatística de Qui-quadrado (χ^2) para determinar se duas variáveis são independentes, usando a seguinte estatística:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3.1)$$

Onde:

χ^2 : Estatística do teste qui-quadrado

O_{ij} : Frequência observada na célula (i, j)

E_{ij} : Frequência esperada na célula (i, j)

A estatística do teste qui-quadrado é usada para determinar se as duas variáveis são independentes entre si. As hipóteses a serem testadas e as regras de decisão para um teste de independência são as seguintes:

H_0 : As variáveis são independentes (não há associação).

H_1 : As variáveis não são independentes (há associação).

Regra de Decisão:

- Escolha um nível de significância α (por exemplo, $\alpha = 0.05$).
- Calcule a estatística de teste apropriada (por exemplo, estatística qui-quadrado).
- Determine os graus de liberdade (df) da distribuição da estatística de teste.
- Calcule o valor-p associado à estatística de teste.
- Compare o valor-p com α :
Se o valor- $p < \alpha$, rejeita-se a hipótese nula (H_0). Conclua que as variáveis não são independentes.
Se o valor- $p > \alpha$, não há evidência suficiente para rejeitar a hipótese nula. (H_0). Conclua que as variáveis são independentes.

3.5.2 Teste de significância dos parâmetros

Após a estimação dos coeficientes, tem-se interesse em verificar a significância das variáveis no modelo. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada com a variável resposta. Para isso, existem os testes de hipóteses. Neste trabalho, serão utilizados os testes de Razão de Verossimilhança de Rao-Scott e Wald.

3.5.3 Teste de Wald

Segundo Cabral (2013), o teste de Wald testa se cada coeficiente é significativamente diferente de zero. Deste modo, o teste de Wald verifica se uma determinada variável independente apresenta uma relação estatisticamente significativa com a variável dependente. Então testam se as hipóteses:

H_0 : O coeficiente β é igual a zero.

H_1 : O coeficiente β é diferente de zero.

A estatística de teste é dada por:

$$W_j = \frac{\hat{\beta}_j}{\text{var}(\hat{\beta}_j)} \quad (3.2)$$

Onde:

W_j : Estatística do Teste de Wald

$\hat{\beta}_j$: é o vector de estimativas para β

$\text{Var}(\hat{\beta}_j)$: é a matriz de variâncias

3.5.4 Teste de razão de verossimilhança

Uma vez estimado o modelo, é necessário testar a sua significância. Um dos teste comumente usado para o testar essa significância é o teste da razão de verossimilhanças. Este teste avalia a significância dos coeficientes estimados simultaneamente, ou seja verifica se o modelo estimado é globalmente significativo (Cabral, 2013). Com este teste pretende-se testar simultaneamente se os coeficientes de regressão associados a β são todos nulos com excepção de β_0 .

As hipóteses à testar são:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_3 = \beta_p = 0$$

H_1 : pelo menos um dos parâmetros é diferente de zero.

A estatística D tem como objectivo, comparar o modelo em análise e o modelo saturado, ou seja, a comparação dos valores observados e dos valores esperados usando a função de verossimilhança que é baseada na seguinte expressão:

$$D = 2 (\log L(M_s) - \log L(M_0)) \quad (3.3)$$

Onde:

D : Estatística D

$\log L(M_s)$: Valor da função de log-verossimilhança do modelo saturado

$\log L(M_0)$: Valor da função de log-verossimilhança do modelo em análise

A estatística D é usada para testar a adequação do modelo em análise em relação ao modelo saturado, e uma diferença significativa entre esses modelos pode indicar a necessidade de um modelo mais complexo.

Onde: modelo ajustado corresponde ao modelo com apenas as variáveis desejadas para o estudo e o modelo saturado corresponde ao modelo com todas as variáveis e interações.

A estatística do teste de razão de verossimilhanças para testar a significância conjunta dos

coeficientes, isto é, a significância do modelo, é dada pela equação em 3.4.

A estatística do teste de razão de verossimilhança (LRT) é comumente representada como G^2 . Ela é usada para comparar a adequação de dois modelos estatísticos: o modelo restrito (nulo) e o modelo irrestrito (alternativo).

A estatística LRT é baseada na diferença entre os valores de log-verossimilhança dos dois modelos, seguindo uma distribuição qui-quadrado com k graus de liberdade, onde k é o número de restrições impostas pelo modelo restrito.

A expressão para a estatística LRT é a seguinte:

$$G^2 = -2 (\log L_R - \log L_U) \quad (3.4)$$

Onde:

G^2 : Estatística do teste de razão de verossimilhança

$\log L_R$: Valor da função de log-verossimilhança do modelo restrito

$\log L_U$: Valor da função de log-verossimilhança do modelo irrestrito

A regra de decisão para a estatística do teste de razão de verossimilhança (LRT, Likelihood Ratio Test) envolve a comparação da estatística LRT com uma distribuição qui-quadrado. A decisão é baseada no valor-p associado à estatística LRT. A regra geral é a seguinte: Estabeleça as hipóteses nula (H_0) e alternativa (H_1)

H_0 : O modelo restrito é verdadeiro; as restrições impostas são válidas.

H_1 : O modelo irrestrito é verdadeiro; pelo menos uma das restrições é inválida.

O teste de razão de verossimilhanças para amostras complexas é baseado em diferenças entre funções Deviance resultantes da ponderação das funções de pseudo-máxima verossimilhança (Lumley, 2018).

3.5.5 Matriz de confusão

Uma maneira prática de qualificar o ajuste do modelo de regressão logística é pela projeção do modelo na tabela de classificação (ou Matriz de Confusão). Para isto, precisa-se criar uma tabela com o resultado da classificação cruzada da variável resposta, de acordo com uma variável dicotômica em que os valores se derivam das probabilidades logísticas estimadas na regressão (Hosmer e Lemeshow 2000). No entanto, é preciso definir uma regra de predição, que dirá se houve acerto ou não da probabilidade estimada com os valores reais, pois as probabilidades variam de 0 a 1 enquanto os valores reais binários possuem valores fixos de 0 “ou” 1.

É intuitivo supor que se as probabilidades aproximam-se de 1 o indivíduo estimado pode ser classificado como $\hat{Y}_i = 1$, bem como de forma contrária, se o modelo estimar probabilidades perto de 0, classificá-la como $\hat{Y}_i = 0$. (Smolski e Battisti, 2019)

De acordo com Smolski e Battisti (2019), após determinado o ponto de corte, é importante avaliar o poder de discriminação do modelo, pelo seu desempenho portanto em classificar os “eventos” dos “não eventos”. Cria-se a Matriz de Confusão com as observações de Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN).

Tabela 3.2: Tabela de classificação

		Previsto	
		Óbito	
Óbito	Observado	Não	Sim
		VP	FP
	Não		
	Sim		

Fonte: Smolski e Battisti (2019)

Sensibilidade : representa a proporção de verdadeiros positivos, ou seja, a capacidade do modelo em avaliar o evento como $\hat{Y}_i = 1$ (estimado) dado que ele é evento real $Y = 1$.

$$SENS = \frac{VP}{FN} \quad (3.5)$$

Especificidade : a proporção apresentada dos verdadeiros negativos, ou seja, o poder de predição do modelo em avaliar como “não evento” $\hat{Y}_i = 0$ sendo que ele não é evento $Y = 0$

$$ESP = \frac{VN}{VN + FP} \quad (3.6)$$

3.5.6 Teste Hosmer e Lemeshow

Segundo Lassance (2015), o teste de Hosmer-Lemeshow, também conhecido como teste dos decis-de-risco, se trata de um teste que compara a frequência da ocorrência das observações com as estimações teóricas, obtidas através do modelo ajustado. Consiste em realizar um ordenamento das observações e dos valores ajustados em função dos valores ajustados. As observações ordenadas são então separadas em g grupos, sendo que comumente $g = 10$. A estatística do teste de Hosmer-Lemeshow é dada por:

$$\hat{C} = \sum_{k=1}^g \frac{(o_{jk} - \hat{e}_{jk})^2}{\hat{e}_{jk}} \quad (3.7)$$

Sendo o_{1k} o número de sucessos no k -ésimo grupo, o_{0k} o número de fracassos, $\hat{e}_{1k}$ a soma dos valores ajustados para o k -ésimo grupo e $\hat{e}_{0k}$ a soma de valor ajustado subtraindo 1.

Teste de Hipótese:

H_0 : Não há diferenças significativas entre os resultados previstos e observados;

H_1 : Há diferenças significativas entre os resultados previstos e observados.

3.5.7 Qualidade de ajuste do modelo logístico clássico

Segundo Silva (2016), sempre que se estima um modelo de regressão é crucial, antes de passar a extrair conclusões, verificar se o modelo se ajusta efectivamente aos dados usados para estimá-lo. O teste de Hosmer-Lemeshow é largamente utilizado na regressão logística com o objectivo de testar a qualidade do ajuste, isto é, o teste avalia se o modelo obtido pode explicar devidamente os dados observados. Este teste baseia-se na divisão dos dados de acordo com as probabilidades previstas. As observações são separadas em grupos de acordo com as probabilidades previstas. A estatística de teste é dada por: A estatística C , também conhecida como estatística de concordância, é usada para avaliar a qualidade de ajuste de um modelo clássico. Ela é definida como:

$$C = \frac{\text{Número de casos correctamente previstos}}{\text{Total de casos}} \quad (3.8)$$

Onde:

”**Número de casos correctamente previstos**”: representa o número de casos que o modelo previu correctamente.

”**Total de casos**”: é o número total de casos no conjunto de dados.

Em modelos de regressão logística, medidas de qualidade do ajuste são funções dos resíduos definidos como a diferença entre o valor observado e o valor ajustado. Por se tratar de regressão logística, não existe uma estatística que seja equivalente ao R^2 , utilizado em regressão linear.

A denominação de pseudo- R^2 deve-se ao facto de eles se assemelham com um do modelo de regressão linear, pois estão em uma escala similar, variando de 0 à 1, mas apesar dessa similaridade não podem ser interpretados da mesma forma como se interpreta em regressão linear (Mesquita, 2014).

A hipótese nula (H_0) geralmente afirma que o modelo se ajusta bem aos dados, enquanto a hipótese alternativa (H_1) sugere que o modelo não se ajusta adequadamente aos dados. Portanto, as hipóteses são estabelecidas da seguinte forma:

H_0 : O modelo se ajusta bem aos dados

H_1 : O modelo não se ajusta adequadamente aos dados

Para testar essas hipóteses, a estatística C é calculada e comparada com um valor crítico. Se a estatística C for significativamente menor do que o valor crítico, pode-se rejeitar a hipótese nula, o que indica que o modelo não se ajusta adequadamente aos dados.

3.5.8 Métodos de Seleção das Variáveis

De acordo com Moura (2018), a seleção das variáveis do modelo é baseada em algum algoritmo que averigua a importância de dada variável e a sua inclusão ou não no modelo. Assim tem-se de forma bastante difundida estes três métodos a saber:

- **Método enter:** todas as variáveis pré-selecionadas são forçadas a ficar no modelo, não tem exclusão de variável insignificante.
- **Método forward:** cada variável é adicionada individualmente, sendo a primeira a que adiciona maior poder explicativo ao modelo e assim por diante até que nenhuma das variáveis restantes aumente o poder explicativo do modelo.
- **Método backward:** contrário do forward ele começa com todas as variáveis e retira individualmente a variável que adiciona o menor poder explicativo ao modelo até que restem somente as variáveis que expliquem significativamente uma parcela da nossa variável dependente.
- **Método stepwise:** absorve os modelos forward e backward, inicia com o forward. Porém a cada variável adicionada as variáveis anteriores são revisadas e verifica-se se seu poder de explicação do modelo permanece significativo.

3.5.9 Interpretação dos parâmetros do modelo

Hosmer & Lemeshow (1989) advogam que na regressão logística, o interesse não só consiste em modelar a probabilidade de ocorrência de um dado evento de interesse em função de um conjunto de variáveis explicativas, mas também em compreender quais dessas variáveis contribuem para a variação da tal probabilidade e em quanto é essa variação.

Essa análise é feita com base no conceito de razão de chances, que é uma medida de intensidade de associação que proporciona informação sobre a probabilidade de ocorrência do evento de interesse quando sob influência de uma dada variável explicativa.

De acordo com Agresti (2002), para uma probabilidade de sucesso π , a chance é definida, de acordo com a equação 3.7, como sendo:

$$g(x) = \frac{\pi(x)}{1-\pi(x)} \quad (3.9)$$

A interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso com a probabilidade de fracasso, usando a função de Chances.

Assim, ao se tomar dois valores distintos da variável explicativa, X_i e $X_i + 1$, obtém-se: Tem-se ainda que:

Admitindo que $x_j + 1 - x_j = 1$, então:

$$Ln(OR) = \ln(e^{\beta_1}) = \beta_1 \quad (3.10)$$

Um problema que surge frequentemente com as variáveis explicativas em um modelo de regressão logística, é a forma de interpretação dos parâmetros. Variáveis qualitativas que não são ordinais, não têm escala ordenada de medida. Desta forma, se forem codificadas de forma arbitrária as variáveis, o parâmetro estimado não terá interpretação. Variáveis qualitativas, devem ter categorias de referência para representar correctamente os efeitos (ausência ou presença) de tais variáveis em um modelo de regressão logística multivariado. (Queiroz, 2004).

Capítulo 4

RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados a estatística descritiva e os resultados obtidos pelas técnicas de regressão logística binária, seguidos pela discussão.

4.1 Análise Descritiva

Tabela 4.1: Distribuição de frequências absolutas (fa) e relativas (fr%)

Variável	Categoria	OBITO				Total
		Sim		Não		
		fa	fr%	fa	fr%	
Região de nascimento da mãe						
	Norte	72	19.83	1675	24.95	1747
	Centro	167	46.01	2614	38.94	2781
	Sul	124	34.16	2424	36.11	2548
Local de residência						
	Rural	220	60.61	4248	63.28	4468
	Urbano	143	39.39	2465	36.72	2608
Religião da mãe						
	Católica	74	20.39	1591	23.70	1665
	Islâmico	52	14.33	1075	16.01	1127
	Zione	99	27.27	1504	22.40	1603
	Evangélica/Pentecostal	65	17.91	1301	19.38	1366
	Anglicano	3	0.83	77	1.15	80

Continua na próxima página

Variável	Categoria	Sim		Não		Total
		fa	fr%	fa	fr%	
	Protestantes	11	3.03	310	4.62	321
	Outras	11	3.03	160	2.38	171
	Sem religião	48	13.22	695	10.35	743
Escolaridade da mãe						
	Sem educação	87	23.97	2051	30.55	2138
	Primário incompleto	198	54.55	2935	43.72	3133
	Primário completo	27	7.44	580	8.64	607
	Secundária incompleta	43	11.85	931	13.87	974
	Secundária completa	6	1.65	147	2.19	153
	Superior	2	0.55	69	1.03	71
Quintil de riqueza						
	Primeiro Quintil	53	14.60	1013	15.09	1066
	Segundo Quintil	60	16.53	1134	16.89	1194
	Terceiro Quintil	71	19.56	1255	18.70	1326
	Quarto Quintil	79	21.76	1576	23.48	1655
	Quinto Quintil	100	27.55	1735	25.85	1835
Estado civil da mãe						
	Solteiro(a)	24	6.61	488	7.27	512
	Casado(a)	161	44.35	3321	49.47	3482
	Cônjuge	111	30.58	2088	31.10	2199
	Viúvo(a)	12	3.31	168	2.50	180
	Divorciado(a)	8	2.20	101	1.50	109
	Separados / Não vive mais junto	47	12.95	547	8.15	594
Sexo da criança						
	Masculino	195	53.72	3377	50.31	3572
	Feminino	168	46.28	3336	49.69	3504
Número de crianças						
	Nenhuma	0	0.00	4	0.06	4
	1 a 3	60	16.53	1019	15.18	1079
	4 a 6	275	75.76	5000	74.48	5275
	7+	28	7.71	690	10.28	718
Tipo de parto						
<i>Continua na próxima página</i>						

Variável	Categoria	Sim		Não		Total
		fa	fr%	fa	fr%	
	Normal ou Vaginal	339	93.39	6343	94.49	6682
	Cesariano	24	6.61	370	5.51	394
Tamanho da criança ao nascer						
	Pequeno	98	26.99	952	14.18	1048
	Médio	125	34.44	2560	38.13	2685
	Grande	140	38.57	3201	47.69	3341
Faixa etária da mãe						
	15–19	65	17.91	748	11.14	813
	20–24	89	24.52	1680	25.03	1769
	25–29	79	21.76	1618	24.10	1697
	30–34	65	17.91	1253	18.67	1318
	35–39	38	10.47	896	13.35	934
	40–44	19	5.23	375	5.59	394
	45–49	8	2.20	143	2.13	151
Ordem de nascimento						
	Primogênito	105	28.93	1515	22.57	1620
	2 ou 3	110	30.30	2369	35.29	2479
	4 ou ordem superior	148	40.77	2829	42.14	2977

A mortalidade infantil constitui um dos mais relevantes indicadores de saúde pública, refletindo não apenas as condições biológicas do recém-nascido, mas sobretudo as desigualdades sociais, econômicas e territoriais. A tabela apresentada fornece um retrato estatisticamente expressivo da distribuição de óbitos infantis, permitindo identificar padrões que reforçam a compreensão dos determinantes sociais e demográficos deste fenômeno. A seguir, descrevem-se as principais evidências extraídas.

Do ponto de vista da região de nascimento da mãe, mostra que o **Centro concentra 46,01% dos óbitos**, contra 34,16% no Sul e apenas 19,83% no Norte. O predomínio do Centro não se explica apenas pelo volume de nascimentos, mas sugere fragilidades estruturais na oferta de cuidados de saúde materno-infantil. Estatisticamente, o risco relativo da criança morrer antes de completar um ano é consideravelmente mais elevado nesta região.

O local de residência, nas áreas rurais registam-se 60,61% das mortes, valor superior ao verificado nas zonas urbanas (39,39%). Esta diferença não é trivial: ela revela um gradiente espacial da mortalidade infantil que acompanha o défice de infraestrutura, saneamento básico e acesso a cuidados obstétricos qualificados. Em termos probabilísticos, nascer em meio rural associa-se a maior vulnerabilidade ao óbito.

A Religião da mãe, indica que mães pertencentes a grupos Zione (**27,27%**), Católicos (**20,39%**) e Evangélicos/Pentecostais (17,91%) concentram maior proporção de perdas. Embora os números reflitam em parte a composição populacional, estatisticamente chama a atenção a elevada incidência no grupo Zione, o que pode indicar barreiras culturais ou práticas de saúde distintas que merecem investigação aprofundada.

A escolaridade da mãe, a análise mostra um padrão inequívoco: **78,52%** das mortes concentram-se em mães com baixa escolaridade (**sem educação formal ou apenas primário incompleto**). O efeito da educação materna sobre a sobrevivência infantil é estatisticamente robusto e amplamente documentado: mães com maior instrução apresentam maior probabilidade de procurar serviços de saúde, adotar práticas preventivas e reconhecer sinais de risco precoce. Aqui, a variável escolaridade emerge como um dos mais fortes determinantes da mortalidade infantil.

O quintil de riqueza, embora o maior número absoluto de mortes ocorra nos quintis mais elevados, este efeito é sobretudo proporcional ao tamanho das coortes de nascimentos nesses estratos. Em termos relativos, não há evidência de vantagem substantiva dos mais baixos sobre os mais elevados. Contudo, a interpretação estatística exige cautela: a riqueza material, isoladamente, não garante proteção sem adequada cobertura de serviços de saúde.

O estado civil da mãe, mostra prevalência de óbitos entre casadas (44,35%) e mulheres em união conjugal (30,58%). Apesar de parecer paradoxal, tal resultado reflete a maior representação dessas categorias na população de mães. Do ponto de vista estatístico, o estado civil não se destaca como factor de risco independente, funcionando mais como variável de contexto.

O sexo da criança, **os meninos representam 53,72% dos óbitos** contra 46,28% das meninas. Esta diferença é consistente com a literatura biomédica, que evidencia maior vulnerabilidade biológica masculina no período neonatal. A análise sugere, portanto, uma desvantagem estatisticamente significativa do sexo masculino.

A maioria das mortes infantis ocorreu entre filhos de mães com 4 a 6 crianças (75,76%),

refletindo o padrão predominante na amostra. Embora se observe ligeiro aumento da mortalidade entre mães com 1 a 3 filhos, as diferenças entre os grupos são pequenas, indicando que o número de filhos não apresenta relação directa com a mortalidade infantil, podendo estar influenciado por outros fatores socioeconômicos e de acesso à saúde.

A esmagadora maioria dos óbitos (93,39%) ocorreu em **partos normais ou vaginal**. Este resultado, no entanto, não pode ser interpretado como indicativo de risco, visto que a cesariana é menos prevalente na população. Do ponto de vista estatístico, trata-se de um reflexo da estrutura de distribuição do tipo de parto e não necessariamente de causalidade.

Os resultados mostram que os recém-nascidos pequenos apresentam uma maior proporção de óbitos (26,99%) em comparação aos nascidos vivos (14,18%), evidenciando maior vulnerabilidade e risco de mortalidade nesse grupo. Já os recém-nascidos médios e grandes registraram maiores taxas de sobrevivência, indicando que o tamanho ao nascer está inversamente associado à mortalidade infantil..

A Faixa etária da mãe revela que, as idades de 20–24 anos (24,52%) e 25–29 anos (21,76%) concentram a maioria dos óbitos, o que acompanha o maior volume de nascimentos nessas faixas. Do ponto de vista do risco proporcional, observa-se maior vulnerabilidade em mães muito jovens (15–19 anos), embora em termos absolutos a incidência seja menor.

Em relação a ordem de nascimento, crianças de ordem superior representam 40,77% dos óbitos, contrastando com 28,93% nos primogénitos. O aumento da ordem de nascimento associa-se a maior probabilidade de morte infantil, indicando sobrecarga de recursos familiares e atenção materna diluída. Este padrão é estatisticamente consistente e reforça o impacto da fecundidade elevada sobre a sobrevivência infantil.

Do modo geral, a análise estatística da tabela evidencia que a mortalidade infantil em Moçambique não é um fenómeno aleatório, mas sim fortemente condicionado por determinantes estruturais. A desigualdade territorial (Centro e zonas rurais), a baixa escolaridade materna e a elevada fecundidade configuram-se como os principais factores de risco. Em termos epidemiológicos, o sexo masculino também se revela mais vulnerável. Estes resultados reforçam a necessidade de políticas públicas direccionadas, com foco na educação feminina, redução da fecundidade e melhoria do acesso a cuidados de saúde nas regiões mais desfavorecidas.

4.1.1 Distribuição de óbitos infantis por sexo da criança

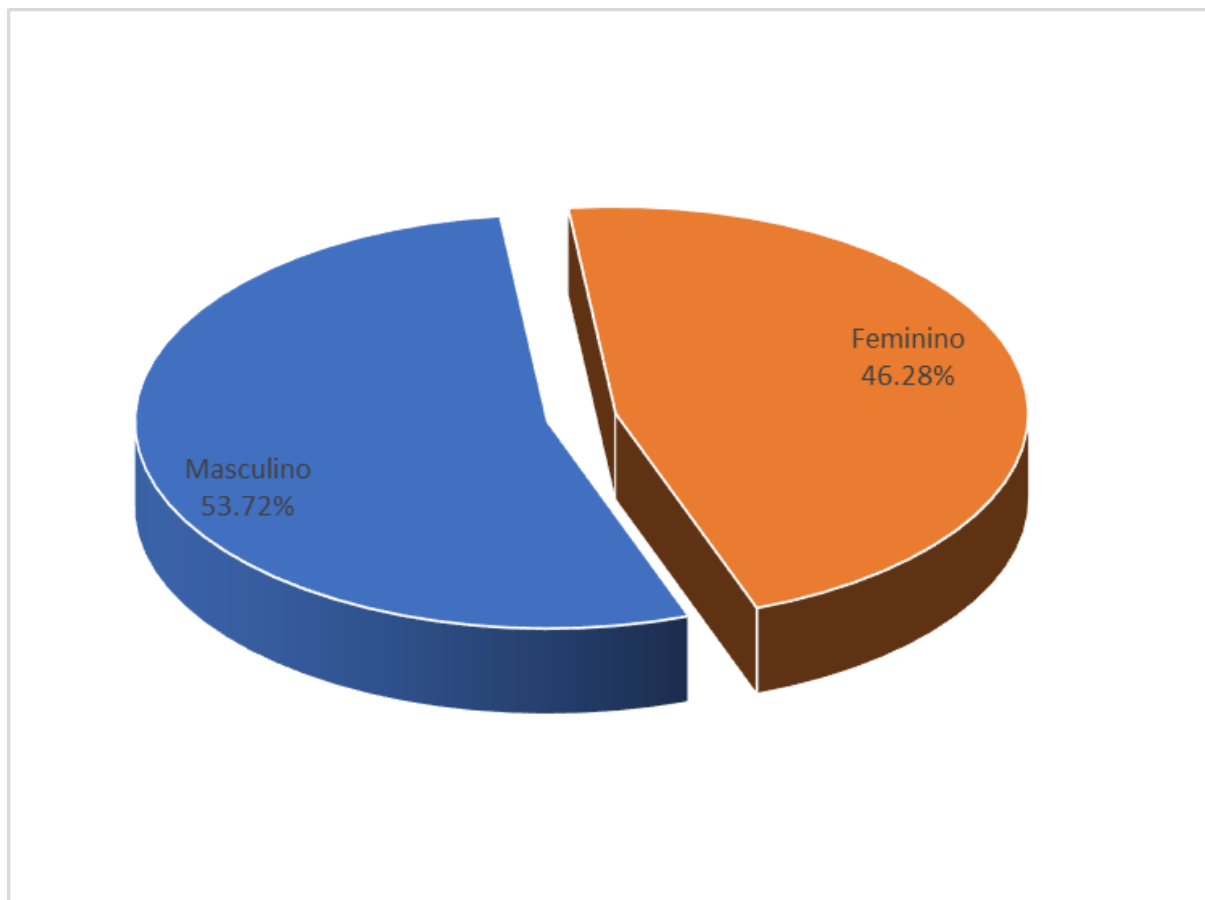


Figura 4.1: Distribuição de óbitos infantis por sexo da criança

Como ilustra gráfico **4.1**: a categoria Masculino representa 195 casos (53,72%) dos óbitos na amostra, indicando que mais da metade dos óbitos infantis registrados ocorreram entre meninos. Essa sobrerrepresentação é consistente com o total de nascidos vivos, onde os meninos também são maioria (3.572 nascimentos, 50,5% do total de 7.076), e se reflete na população geral de óbitos, constituindo 3.377 casos (50,31%).

Por outro lado, a categoria Feminino representa 168 casos (46,28%) na amostra, ou seja, um pouco menos da metade. Essa proporção é ligeiramente inferior à sua representação no total de nascidos vivos (3.504 nascimentos, 49,50%), corroborando a tendência de uma menor mortalidade entre meninas, que também é observada na população de referência, com 3.336 óbitos (49,69%).

4.1.2 Distribuição de óbitos infantis por filiação religiosa da mãe

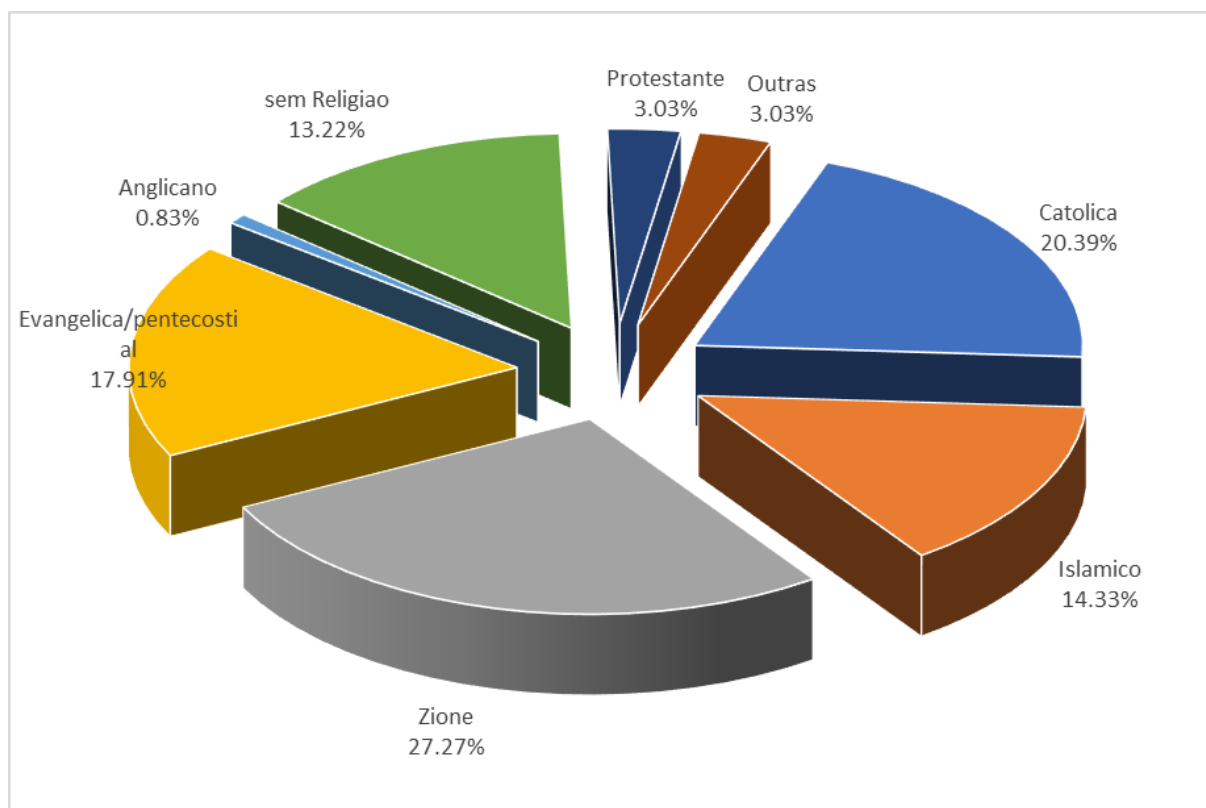


Figura 4.2: Distribuição de óbitos infantis por filiação religiosa da mãe

Na análise da mortalidade infantil segundo filiações religiosas das mães dos bebês, conforme observa-se no gráfico **4.2**, a igreja Zione representa 27,27% dos óbitos, superior à sua proporção nos nascidos vivos (22,40%), e "Sem Religião" representa 13,22% dos óbitos, acima dos 10,35% dos nascidos vivos — indicando sobrerrepresentação e possível maior risco. As religiões Católica (20,39% dos óbitos vs. 23,70% dos nascidos), Evangélica/Pentecostal (17,91% vs. 19,38%) e Islâmica (14,33% vs. 16,01%) têm óbitos proporcionais ao seu tamanho populacional. Já Protestantes (3,03% vs. 4,62%), Anglicano (0,83% vs. 1,15%) e Outras (3,03% vs. 2,38%) apresentam sub-representação nos óbitos, sugerindo factores protectores.

4.2 Análise de Associação entre Variáveis Independentes e dependente (Óbito)

Para a análise de existência de relação entre a variável de interesse e as demais variáveis do estudo foi usado o teste Qui-quadrado (χ^2), cujos resultados encontram-se na tabela **4.2**. Consideraram-se como variáveis significativas aquelas que apresentaram P-valor menor

que nível de significância ($\alpha = 5\%$).

Tabela 4.2: Teste Qui-quadrado de associação entre variáveis independentes e Dependente

Variável	Qui-quadrado (χ^2)	P-Value
Região de nascimento da mãe	8.39	0.02
Local de residência	1.06	0.30
Religião da mãe	11.74	0.11
Escolaridade	16.90	0.00
Quintil de riqueza	1.04	0.90
Estado civil da mãe	13.49	0.02
Sexo do bebê	1.61	0.21
Número de consultas pré-natais	2.99	0.56
Tipo de parto	0.79	0.37
Tamanho da criança ao nascer	50.43	0.00
Idade da mãe	16.89	0.01
Ordem de nascimento	8.68	0.01

De acordo com a tabela 4.2 os resultados do teste Qui-quadrado indicaram que algumas variáveis estão significativamente associadas ao óbito infantil. A **região de nascimento da mãe** apresentou associação significativa ($\chi^2 = 8.39$, $p = 0.02$), sugerindo que diferenças regionais podem influenciar o risco de óbito. O **nível de escolaridade da mãe** também mostrou associação estatisticamente significativa ($\chi^2 = 16.90$, $p = 0.00$), indicando que mães com menor escolaridade podem apresentar um risco mais elevado de óbito infantil. O **estado civil da mãe** foi outro factor significativo ($\chi^2 = 13.49$, $p = 0.02$), evidenciando que mães solteiras ou sem parceiro podem dispor de menos suporte social e apresentar maior risco. O **tamanho da criança ao nascer** apresentou associação altamente significativa ($\chi^2 = 50.43$, $p = 0.00$), confirmando que recém-nascidos com baixo peso ou menor tamanho apresentam maior probabilidade de óbito.

Além disso, a **idade da mãe** ($\chi^2 = 16.89$, $p = 0.01$) e a **ordem de nascimento** da criança ($\chi^2 = 8.68$, $p = 0.01$) mostraram associação significativa com o óbito infantil, indicando que extremos de idade materna e a posição na sequência de nascimentos podem aumentar o risco.

Em suma, os factores maternos, socioeconómicos e biológicos, tais como **escolaridade**, **estado civil**, **idade**, **região de nascimento**, **ordem de nascimento** e **tamanho ao nascer**, são determinantes importantes para o risco de óbito infantil nesta amostra.

4.3 Modelo de Regressão Logística Binária

4.3.1 Construção do Modelo de regressão logística binária

Tabela 4.3: Estimativas dos coeficientes do modelo

Co-variáveis	β	E.P	Wald	gl	P-Valor	R.C	I.C-95%
Região							
Norte	0.380	0.087	18.900	1	0.000	1.462	[1.232; 1.735]
Centro	0.794	0.061	171.998	1	0.000	2.211	[1.964; 2.490]
Sul (Ref)	—	—	192.361	2	0.000	—	—
Zona de Residência							
Urbano	1.332	0.055	584.779	1	0.000	3.788	[3.400; 4.219]
Rural (Ref)	—	—	—	—	—	—	—
Religião							
Católica	-0.359	0.136	6.920	1	0.009	0.698	[0.535; 0.913]
Islâmica	-0.366	0.145	6.330	1	0.012	0.694	[0.522; 0.922]
Zione	0.195	0.133	2.151	1	0.142	1.215	[0.937; 1.577]
Evangélica/Pentecostal	-0.813	0.137	35.385	1	0.000	0.444	[0.339; 0.580]
Anglicano	0.261	0.214	1.487	1	0.223	1.299	[0.853; 1.977]
Sem religião	-0.192	0.141	1.852	1	0.173	0.825	[0.625; 1.088]
Protestantes	-1.592	0.184	74.700	1	0.000	0.203	[0.142; 0.292]
Outros (Ref)	—	—	355.757	7	0.000	—	—
Escolaridade							
Sem escolaridade	3.111	0.457	46.294	1	0.000	22.441	[9.159; 54.983]
Primário Incompleto	3.064	0.456	45.198	1	0.000	21.406	[8.763; 52.291]
Primário Completo	2.393	0.458	27.291	1	0.000	10.948	[4.461; 26.869]
Secundário Incompleto	1.866	0.456	16.768	1	0.000	6.460	[2.645; 15.776]
Secundário Completo	1.501	0.482	9.692	1	0.002	4.487	[1.744; 11.544]
Superior (Ref)	—	—	337.279	5	0.000	—	—
Quintil de Riqueza							
Primeiro Quintil	0.545	0.093	34.244	1	0.000	1.725	[1.437; 2.070]
Segundo Quintil	0.450	0.090	25.076	1	0.000	1.569	[1.315; 1.871]
Terceiro Quintil	0.356	0.082	18.760	1	0.000	1.428	[1.215; 1.677]
Quarto Quintil	0.325	0.067	23.818	1	0.000	1.384	[1.214; 1.576]
Quinto Quintil (Ref)	—	—	37.614	4	0.000	—	—
Estado Civil							
Solteiro(a)	0.011	0.100	0.013	1	0.909	1.011	[0.832; 1.230]
Casado(a)	-0.016	0.076	0.042	1	0.838	0.985	[0.848; 1.143]
Cônjuge	-0.164	0.079	4.275	1	0.039	0.849	[0.727; 0.992]
Viúvo(a)	0.039	0.143	0.073	1	0.787	1.039	[0.786; 1.375]

Continua na próxima página

Tabela 4.3 – continuação

Co-variáveis	β	E.P	Wald	gl	P-Valor	R.C	I.C-95%
Divorciado(a)	0.679	0.149	20.908	1	0.000	1.973	[1.474; 2.640]
Separados (Ref)	—	—	38.905	5	0.000	—	—
Sexo da criança							
Masculino	0.881	0.041	457.100	1	0.000	2.413	[2.226; 2.616]
Feminino (Ref)	—	—	—	—	—	—	—
Número de consultas							
Nenhuma	-18.077	18560.790	0.000	1	0.999	0.000	[0.000; 0.000]
1 a 3	2.510	0.128	385.823	1	0.000	12.306	[9.580; 15.809]
4 a 6	2.009	0.120	278.157	1	0.000	7.454	[5.887; 9.439]
7+ (Ref)	—	—	394.708	3	0.000	—	—
Tipo de Parto							
Normal/Vaginal	2.021	0.166	147.987	1	0.000	7.548	[5.450; 10.454]
Cesárea (Ref)	—	—	—	—	—	—	—
Tamanho da Criança							
Pequeno	-0.169	0.099	2.922	1	0.087	0.844	[0.696; 1.025]
Médio	-0.259	0.099	6.881	1	0.009	0.772	[0.636; 0.937]
Grande (Ref)	—	—	9.659	2	0.008	—	—
Idade da Mãe							
15–19	1.225	0.188	42.522	1	0.000	3.405	[2.356; 4.922]
20–24	0.695	0.178	15.239	1	0.000	2.004	[1.413; 2.840]
25–29	0.655	0.173	14.328	1	0.000	1.925	[1.371; 2.702]
30–34	0.691	0.172	16.124	1	0.000	1.995	[1.424; 2.794]
35–39	0.489	0.176	7.753	1	0.005	1.630	[1.156; 2.300]
40–44	0.423	0.188	5.039	1	0.025	1.526	[1.055; 2.207]
45–49 (Ref)	—	—	83.738	6	0.000	—	—
Ordem de Nascimento							
Primogênito	0.621	0.082	57.230	1	0.000	1.861	[1.584; 2.185]
2 ou 3	0.183	0.058	9.971	1	0.002	1.201	[1.072; 1.346]
4+ (Ref)	—	—	61.029	2	0.000	—	—
Constante	-9.170	0.556	271.778	1	0.000	0.000	—

Ref** – categoria de referência.

Interpretação do Modelo em Termos da Razão de Chances

Controlando o efeito dos restantes factores no modelo, conclui-se:

- **Sexo da Criança:** Crianças do sexo masculino têm cerca de 2,4 vezes mais chance de morrer antes de completar um ano em relação às do sexo feminino.
- **Região de Nascimento:** Crianças nascidas na região Centro têm cerca de 2,2 vezes mais chance de mortalidade infantil em relação às nascidas na região Sul. Crianças da região Norte têm cerca de 1,5 vezes mais chance do que as da região Sul.
- **Zona de Residência:** Crianças que vivem em área urbana têm cerca de 3,8 vezes mais chance de mortalidade infantil em relação às crianças em área rural.
- **Religião da Mãe:** Controlando outros factores, crianças de mães católicas têm cerca de 0,7 vezes a chance de mortalidade infantil em relação às mães da categoria “Outros”. Crianças de mães islâmicas têm cerca de 0,7 vezes; evangélicas/pentecostais têm cerca de 0,4 vezes; protestantes têm cerca de 0,2 vezes a chance, indicando redução importante do risco.
- **Escolaridade da Mãe:** Crianças cujas mães não têm escolaridade têm cerca de 22 vezes mais chance de mortalidade infantil em relação às mães com escolaridade superior. Mães com primário incompleto têm cerca de 21 vezes mais chance, e mães com primário completo têm cerca de 11 vezes mais chance.
- **Quintil de Riqueza:** Crianças do primeiro quintil têm cerca de 1,7 vezes mais chance de mortalidade infantil em relação ao quinto quintil (mais rico), enquanto o risco diminui progressivamente nos quintis mais altos.
- **Estado Civil da Mãe:** Crianças de mães divorciadas têm cerca de 2 vezes mais chance de mortalidade infantil em comparação com mães “não mais juntos/separados”.
- **Número de Consultas Pré-natais:** Crianças cujas mães tiveram de 1 a 3 consultas pré-natais têm cerca de 12 vezes mais chance de mortalidade infantil em comparação com mães que tiveram 7 ou mais consultas. Mães com 4 a 6 consultas têm cerca de 7,5 vezes mais chance.
- **Tipo de Parto:** Crianças nascidas por parto normal/vaginal têm cerca de 7,5 vezes mais chance de mortalidade infantil em relação a cesarianos.
- **Tamanho da Criança ao Nascer:** Crianças de tamanho médio têm cerca de 0,8 vezes a chance de mortalidade infantil em relação às de tamanho grande. Crianças pequenas têm cerca de 0,8 vezes a chance também.
- **Idade da Mãe:** Crianças de mães entre 15–19 anos têm cerca de 3,4 vezes mais chance de mortalidade infantil em relação às mães entre 45–49 anos. Crianças de

mães de 20–24 anos têm cerca de 2 vezes mais chance, as de mães de 25–29 anos têm cerca de 1,9 vezes, e as de mães de 30–34 anos têm cerca de 2 vezes mais chance.

- **Ordem de Nascimento:** Primogênitos têm cerca de 1,9 vezes mais chance de mortalidade infantil em relação a filhos com ordem superior a 4. Filhos de segunda ou terceira ordem têm cerca de 1,2 vezes mais chance.

Tabela 4.4: Modelo logístico nulo (Passo 0)

	Coefficiente	EP	Wald	gl	p-valor	RC
Passo 0 Constante	0.000	0.017	0.000	1	1.000	1.000

De acordo com o resultado apresentado na Tabela (4.4), o modelo depende apenas da constante ($\beta_0 = 0.000$). No entanto, segundo o p-valor associado à estatística do Teste de Wald ($p = 1,000$), não se rejeita a hipótese nula de que o coeficiente β_0 seja igual a zero.

Tabela 4.5: Teste de razão de verossimilhança

		Qui-quadrado	gl	p-valor
Passo 1	Passo	3525.28	39	0.000
	Bloco	3525.28	39	0.000
	Modelo	3525.28	39	0.000

De acordo com os resultados do teste de Razão de Verossimilhança apresentados na Tabela (4.5), verifica-se que: O passo 1 ($\chi^2_{39} = 3525.28$, $p = 0.000$), o bloco ($\chi^2_{39} = 3525.28$, $p = 0.000$) e o modelo completo ($\chi^2_{39} = 3525.28$, $p = 0.000$) são altamente significativos estatisticamente. A um nível de significância de 5%, rejeita-se fortemente a hipótese nula, concluindo que as variáveis incluídas no modelo possuem poder preditivo substancial sobre a variável dependente.

Tabela 4.6: Medidas de avaliação do ajuste do modelo final

Passo	-2LL	Cox e Snell R^2	Nagelkerke R^2
1	15087.110	0.231	0,308

Com base nos resultados apresentados na Tabela (4.6), observa-se que: O coeficiente de determinação de Nagelkerke de 0,308 (30,8%) indica que as variáveis independentes explicam uma proporção substancial da variabilidade na mortalidade infantil. De forma consistente, o coeficiente de Cox e Snell de 0,231 (23,1%) confirma o **bom poder explicativo** do modelo.

Estes valores, considerados moderados a altos na literatura de modelos logísticos, sugerem que o modelo captura de forma relevante os padrões associados à mortalidade infantil. A estatística de verossimilhança ($-2LL = 15087.110$) deve ser comparada com o modelo

nulo para avaliar a melhoria no ajuste, mas os valores de R^2 já indicam uma capacidade explicativa robusta. O modelo demonstra ser uma ferramenta analítica válida para compreender os fatores associados à mortalidade infantil em Moçambique, explicando aproximadamente um terço da variabilidade do fenómeno estudado.

Tabela 4.7: Teste de Hosmer e Lemeshow

	Qui-quadrado	gl	p-valor
Passo 1	18.573	8	0.017

De acordo com o Teste de Hosmer e Lemeshow apresentado na Tabela 4.7, verifica-se um resultado de $\chi^2_8 = 18.573$ com p-valor = 0.017. Uma vez que o p-valor é inferior ao nível de significância de 5%, rejeita-se a hipótese nula do teste. Isto indica que existem evidências estatísticas de que o modelo não se ajusta adequadamente aos dados, sugerindo diferenças significativas entre as probabilidades previstas pelo modelo e as probabilidades observadas nos dados.

Tabela 4.8: Tabela de classificação do modelo final

		Previsto		% Correcto
		Óbito		
		Não	Sim	
Observado	Não	4.661	2.052	69,4%
	Sim	1.885	4.828	71,9%
% Global				70,7%

Fonte: Elaborado pelo autor

A Tabela 4.8 apresenta a matriz de confusão do modelo logístico final para previsão de óbito infantil. O modelo demonstra uma acurácia global de 70,7%, indicando que aproximadamente sete em cada dez observações são classificadas correctamente.

O modelo apresenta sensibilidade de 71,9%, indicando que consegue identificar correctamente cerca de 72% dos casos reais de óbito infantil. A especificidade de 69,4% demonstra boa capacidade de distinguir casos sem óbito. O equilíbrio entre sensibilidade e especificidade (diferença de 2,5 pontos percentuais) confirma que o modelo não apresenta viés significativo para nenhuma das classes.

Os 1.885 falsos negativos representam o erro mais crítico, pois correspondem a óbitos reais não identificados pelo modelo. Os 2.052 falsos positivos representam casos classificados erroneamente como óbito. A acurácia global de 70,7% sustenta a robustez do modelo para apoio à decisão em contextos de saúde pública moçambicana.

4.3.2 Modelo logístico que prevê a probabilidade de uma criança morrer

A probabilidade estimada de mortalidade infantil é dada por:

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(p)}},$$

onde:

$$\text{logit}(p) = -9.170 + \sum_j \beta_j X_j.$$

O modelo estimado é:

$$\begin{aligned} \text{logit}(p) = & -9.170 + 0.380 R_{\text{Norte}} + 0.794 R_{\text{Centro}} + 1.332 Z_{\text{Urbano}} \\ & - 0.359 \text{Rel}_{\text{Católica}} - 0.366 \text{Rel}_{\text{Islâmica}} + 0.195 \text{Rel}_{\text{Zione}} \\ & - 0.813 \text{Rel}_{\text{Evangélica/Pentecostal}} + 0.261 \text{Rel}_{\text{Anglicano}} - 0.192 \text{Rel}_{\text{Sem religião}} \\ & - 1.592 \text{Rel}_{\text{Protestantes}} + 3.111 \text{Esc}_{\text{Sem escolaridade}} + 3.064 \text{Esc}_{\text{Primário incompleto}} \\ & + 2.393 \text{Esc}_{\text{Primário completo}} + 1.866 \text{Esc}_{\text{Secundário incompleto}} + 1.501 \text{Esc}_{\text{Secundário completo}} \\ & + 0.545 Q_{1^\circ \text{ quintil}} + 0.450 Q_{2^\circ \text{ quintil}} + 0.356 Q_{3^\circ \text{ quintil}} \\ & + 0.325 Q_{4^\circ \text{ quintil}} + 0.011 EC_{\text{Solteiro(a)}} - 0.016 EC_{\text{Casado(a)}} \\ & - 0.164 EC_{\text{Cônjuge}} + 0.039 EC_{\text{Viúvo(a)}} + 0.679 EC_{\text{Divorciado(a)}} \\ & + 0.881 \text{Sexo}_{\text{Masculino}} - 18.077 \text{Consultas}_{\text{Nenhuma}} \\ & + 2.510 \text{Consultas}_{1 \text{ a } 3} + 2.009 \text{Consultas}_{4 \text{ a } 6} + 2.021 \text{Parto}_{\text{Normal}} \\ & - 0.169 \text{Tamanho}_{\text{Pequeno}} - 0.259 \text{Tamanho}_{\text{Médio}} \\ & + 1.225 \text{Idade}_{15-19} + 0.695 \text{Idade}_{20-24} + 0.655 \text{Idade}_{25-29} \\ & + 0.691 \text{Idade}_{30-34} + 0.489 \text{Idade}_{35-39} + 0.423 \text{Idade}_{40-44} \\ & + 0.621 \text{Ordem}_{\text{Primogênito}} + 0.183 \text{Ordem}_{2 \text{ ou } 3}. \end{aligned}$$

Portanto, o estimador final da probabilidade é:

$$\hat{p} = \frac{1}{1 + e^{-(-9.170 + \sum \beta_i X_i)}}$$

4.4 Discussão dos Resultados

No que diz respeito a Escolaridade da mãe, Estudo realizado no Sudão (África), por Elshibly & Schmalisch (2008), com o objectivo de quantificar os efeitos antropométricos das mães, educação e status socioeconômico na gestação e no baixo peso ao nascer das

crianças, apontaram que a variável anos de escolaridade da mãe apresentou efeito significativo do risco de morrer no primeiro ano de vida, com o aumento dos anos de escolaridade da mãe, o risco de morrer da criança no seu primeiro ano de vida diminui. Os resultados encontrados nesse estudo mostraram a mesma tendência em comparação com o estudo feito por Elshibly e Schmalish.

Alguns estudos têm encontrado associação não significativa da educação com a mortalidade infantil. Por exemplo, Silva et al. (2006) observaram que o risco de óbito diminuía à medida que aumentava o grau de escolaridade da mãe, mas o efeito não foi estatisticamente significativo.

Em relação a escolaridade da mãe, apartir dos resultados do estudo feito comparado com estudos feito por Silva há um consenso na educação materna e o óbito infantil, isto é, quanto maior for a escolaridade da mãe menos são as chance de mortalidade.

De acordo com Remoaldo (2002) o sexo de um recém-nascido é um dos factores que deve ser considerado quando se aborda a questão da morbidade e mortalidade infantil. os resultados desse estudo mostraram que os bebés do sexo masculino morrem mais em comparação com as do sexo feminino.

Quanto a ordem de nascimento do bebé, para Bicego & Ahmad (1996), as chances de sobrevivência variam de acordo com a ordem de nascimento da criança. Os estudos de Rutstein (2000), Ahmad, Lopez & Inoue (2000) destacam que os primeiros nascimentos e aqueles a partir da oitava ordem em diante apresentam um aumento no risco de mortalidade infantil (Bicego & Ahmad, 1996; Mahy, 2003).

A análise por filiação religiosa revelou padrões distintos, com as religiões Zione (27,27% dos óbitos) e o grupo Sem Religião (13,22%) apresentando sobre-representação nos óbitos. Estes resultados introduzem uma dimensão sociocultural pouco explorada na literatura, sugerindo que crenças e práticas religiosas podem influenciar comportamentos em saúde, acesso a serviços ou condições socioeconômicas que impactam a sobrevivência infantil.

A sobrevivência do primeiro filho de uma mulher pode estar intrinsecamente relacionada a factores biológicos. Muitos dos primeiros nascimentos ocorrem antes de a mulher atingir a maturidade física e reprodutiva, aumentando assim o risco perinatal e introduzindo maiores desafios durante o parto (Bicego & Ahmad, 1996; Bai et al., 2002).

Em resumo, ao discutir a ordem de nascimento neste estudo, destaca-se que o risco mais elevado de mortalidade em crianças menores de um ano é observado com maior intensidade para o primogénitos.

Capítulo 5

CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES

5.1 Conclusões

Com base nos objectivos traçados no presente estudo, conclui-se que:

- Os resultados evidenciam que a mortalidade infantil em Moçambique resulta de múltiplos factores interligados, incluindo determinantes maternos (escolaridade, idade e estado civil), factores biológicos (tamanho do bebé e ordem de nascimento) e desigualdades regionais. Aspectos socioeconómicos, demográficos e relacionados ao cuidado pré-natal especialmente a escolaridade materna, o número de consultas pré-natais, a região e a zona de residência, assim como o sexo da criança exercem influência significativa. Estes resultados reforçam a necessidade de políticas públicas integradas que promovam a educação materna, garantam acompanhamento pré-natal de qualidade e reduzam desigualdades regionais, como forma de diminuir a mortalidade infantil.

5.2 Limitações do Estudo

Uma limitação significativa deste trabalho reside na aplicação da técnica SMOTE para balanceamento dos dados. A decisão de transformar a base original (com apenas 5,1% de casos de sucesso) para uma distribuição equilibrada (50% – 50%) introduz implicações estatísticas que devem ser reconhecidas:

- **Artificialidade dos padrões:** A geração de casos sintéticos pode criar relações não existentes na população real, reflectindo-se no resultado do Teste de Hosmer-Lemeshow [$\chi^2_{(8)} = 18,573$; $p = 0,017$]

- **Compromisso metodológico:** Optou-se por reduzir o viés de classe às custas de potencial sobreajuste, priorizando a detecção de padrões preditivos em detrimento da perfeita calibração probabilística
- **Validade externa:** As probabilidades geradas pelo modelo reflectem a distribuição artificial, não a realidade epidemiológica onde a mortalidade infantil corresponde a aproximadamente 5% da população

Esta abordagem representou uma solução pragmática para o problema de desbalanceamento, permitindo identificar variáveis preditivas relevantes, ainda que com o custo estatístico evidenciado pelo teste de ajuste.

5.3 Recomendações

A mortalidade em Moçambique, especialmente a mortalidade infantil, é uma preocupação significativa e abordar esse problema requer uma abordagem multipla. portanto mediante as conclusões apresentadas, recomenda-se que:

- Promover programas de educação e sensibilização dirigidos às mães e famílias;
- Melhorar o acesso à educação feminina;
- Reduzir desigualdades regionais no acesso a cuidados de saúde;
- Promoção do planeamento familiar;
- Integração intersetorial e monitoramento contínuo.

Referências

- [1] Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). New York: Wiley.
- [2] Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). New York: John Wiles & Sons.
- [3] Agresti, A., & Finlay, B. (2009). *Statistical Methods for Social Sciences* (4th ed.). Prentice Hall.
- [4] Ahmad, O. B.; Lopez, A. D.; Inque, M. (2000), The decline in child mortality: a reappraisal. *Bulletin of the World Health Organization*, New York.
- [5] Bai, J. et al. (2002), Parity and pregnancy outcomes. *American Journal of Obstetrics and Gynecology*, Saint Louis.
- [6] Bessinger, R., C. Katende, e N. Gupta. (2004). "Multi-media Campaign Exposure Effects on Knowledge and Use of Condoms for STI and HIV/AIDS Prevention in Uganda." *Evaluation and Program Planning*.
- [7] Bicego, G.; Ahmad, O. B. (1996), *Infant and child mortality*. Calverton, Maryland: Macro International Inc.
- [8] Cabral, C. (2013). *Aplicação do Modelo de Regressão Logística num Estudo de Mercado*. Lisboa, Portugal.
- [9] César, C. C; Miranda-Ribeiro, P e Abreu, D. M. X. (2002). Efeito-idade ou efeito-pobreza? Mães adolescentes e mortalidade neonatal em Belo Horizonte. *Revista Brasileira de Estudos de População*, Campinas.
- [10] Elshibly, E. M.; Schmalisch, G. (2008), The effect of maternal anthropometric characteristics and social factors on gestational age and birth weight in Sudanese newborn infants. *BMC Public Health*, Genebra.
- [11] Farhat, C. A. (2003). *Análise de diagnóstico em regressão logística*. São Paulo.
- [12] Ferrari, A. P., Carvalhes, M. A., & Parada, C. M. (2016). Associação entre pré-natal e parto na rede de saúde suplementar e cesarea eletiva. *Revista Brasileira de Epidemiologia*.

-
- [13] Figueira, C. V. (2006). Modelos de Regressão Logística. Porto Alegre.
- [14] Fuentes, V. L. P. (1990). Condicionantes Sócio-Econômicas da Mortalidade Infantil: Estado de São Paulo 1960-1984. Dissertação de Mestrado – Universidade de São Paulo, São Paulo. Brasil
- [15] Gil, A. C. (2002). Como elaborar projetos de pesquisa. São Paulo: Atlas.
- [16] Gil, A. C. (2008). Como elaborar projetos de pesquisa (4th ed.). São Paulo: Atlas.
- [17] Gonsçaves, A. M. (2013). Regressão Logística Aplicada à preditores de Morte. Tese Mestrado. Univesidade de Coimbra.
- [18] Hill, M. (2005). Investigação por questionário. (L. Edições Silaba, Ed.) Lisboa.
- [19] Hosmer, J., & Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley e Son Inc.
- [20] IDS. (2011). Moçambique Inquerito Demográfico e de Saúde. MEASURE DHS/ICF International.
- [21] Johnson, R. A., & Bhattacharyya, G. K. (2010). Estatística básica e elementar (vol.2). Pearson Education.
- [22] Lumley, T. (2018). analysis of complex survey samples: Package survey. UTC.
- [23] Maia, M. B. (2010). Humanização do parto: políticas públicas, comportamento organizacional e. (E. Fiocruz, Ed.) Rio de Janeiro.
- [24] Marconi, M. A., & Lakatos, E. M. (1985). fundamentos de metodologia científica (5th ed.). São Paulo: Atlas.
- [25] Mesquita, C. (2014). Um modelo de regressão Logística para avaliação dos programas de pós-graduação no Brasil. Rio de Janeiro, Brasil.
- [26] Mestre, P. (2009). Aplicação dos Modelos Lineares Generalizados às telecomunicações Móveis: Caracterização dos clientes que desactivam os seus serviços. Lisboa, Portugal.
- [27] Minayo, M. C. (2008). O desafio do (11frs ed.). (Hucitec, Ed.) São Paulo.
- [28] Montenegro, S. G. (2009). Modelo de Regressão Logistica Ordinal em dados categoricos na area de ergonomia experimental. João Pessoa.
- [29] Moura, G.M. (2018). Regressão Logística aplicada à análise de risco de crédito. Brasil.
- [30] Mulenga. A (2018). Introdução à Estatística. Editora impressa Universitária, Maputo.

-
- [31] Lassance, R. F. (2015). Comparação dos Modelos Lineares Generalizados Logístico e Log-Binomial. Brasília.
- [32] Lindstrom, D. P. e Brambila Paz. C. (2001). Alternative theories of the relationship of schooling and work to family formation: The Mexican paradox. *Social Biology*.
- [33] OMS. (2016), Recomendações sobre Cuidados Pré-natais para uma Experiência Positiva da Gravidez da OMS. Genebra, Suíça.
- [34] OMS. (2015). Declaracao da OMS sobre Taxas de Cesareas.
- [35] Prodanov, C. C., & Freitas, E. C. (2013). metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico. Novo Hamburgo – Rio Grande do sul, Brasil: Universidade FEEVALE.
- [36] Queiroz, N. (2004). Regressão Logística - Uma estimativa Bayesiana aplicada na identificação de Factores de Risco para HIV em doadores de Sangue. Recife, Brasil.
- [37] Remoaldo, P. C. A. (2002), Desigualdades territoriais e sociais subjacentes à mortalidade infantil em Portugal. Textos universitários de ciências sociais e humanas. Fundação Calouste Gulbenkian, Fundação para a ciência e tecnologia.
- [38] RIBEIRO J.F., et al., (2009), O prematuro em unidade de terapia intensiva neonatal: a assistência do enfermeiro. *Journal of Nursing UFPE / Revista de Enfermagem UFPE*.
- [39] Rutstein, S. O. (2000), Factors associated with trends in infant and child mortality in development countries during the 1990s. *Bulletin of the World Health Organization*, New York.
- [40] Rutstein, S. O. (1984). *Infant and Child Mortality: Levels, Trends and Demographic Differentials*. Revised edition. WFS Comparative Studies no 43. Voorburg, Netherlands: International Statistical Institute.
- [41] Silva, C. F. et al. Fatores de risco para mortalidade infantil em município do Nordeste do Brasil: linkage entre bancos de dados de nascidos vivos e óbitos infantis – 2000 a 2002. *Revista Brasileira de Epidemiologia*.
- [42] Sgroi J.L. (2008). Relação de Factores de risco na gestação e peso ao nascer em crianças atendidas no centro de Saude Escola de Butana, Cidade de Sao Paulo. São Paulo, Brasil.
- [43] Schoeps, D.; et al., (2007), Fatores de risco para mortalidade neonatal precoce. *Revista de Saúde Pública*.

-
- [44] Silva, P. (2016). Modelos de Regressão Linear e Logística utilizando os software R. Porto, Portugal.
- [45] Smith, C. (2003) Theorizing religious effects among american adolescents. Journal for the Scientific Study of Religion.
- [46] Smolski, F. M., e Battisti, I. D. (2019). Software R: curso avançado. UFFS.
- [47] UNICEF. (2008), Manual e aleitamento materno. Edição revista.
- [48] UNIGME. (2023). Levels and Trends in Child Mortality. [<https://brasil.un.org/sites/default/files/2024-03/UNIGME-2023-Child-Mortality-Report.pdf>].
- [49] Verona, A. P. A., et al., (2010), Infant mortality and mothers' religious involvement in Brazil. Revista Brasileira de Estudos da População.
- [50] Wood, C. H.; Williams, P.; Chijiwa, K. (2007), Protestantism and Child Mortality in Northeast Brazil, 2000. Journal for the Scientific Study of Religion.
- [51] World Health Organization. (2014). Low Birth Weight: Country, Regional and Global Estimates.

ANEXOS

Tabela 5.1: Tabela de classificação do modelo nulo

		Previsto		
		Óbito		
		Sim	Não	% Correcto
Observado	Sim	0	6.713	0,0%
	Não	0	6.713	100,0%
% Global				50,0%

Tabela 5.2: Tabela de Contingência para o Teste de Hosmer e Lemeshow

Step	A criança está viva = Não		A criança está viva = Sim		Total
	Observed	Expected	Observed	Expected	
1	1234	1236.676	109	106.324	1343
2	1074	1058.902	271	286.098	1345
3	949	928.006	394	414.994	1343
4	784	805.850	556	534.150	1340
5	687	708.674	655	633.326	1342
6	646	604.390	697	738.610	1343
7	478	488.609	862	851.391	1340
8	359	401.183	988	945.817	1347
9	324	311.206	1019	1031.794	1343
10	178	169.503	1162	1170.497	1340

Tabela 5.3: Coeficientes do Modelo de Regressão

Modelo	<i>B</i>	E.P	β	t	Sig.	I.C-95.0% para <i>B</i>		Estatísticas de Colinearidade	
						IC _{inf}	IC _{sup}	Tolerância	VIF
Constante	1.943	0.034		57.223	0.000	1.877	2.010		
Região de Nascimento da mãe	-0.015	0.006	-0.022	-2.325	0.020	-0.027	-0.002	0.699	1.430
Local de Residência	-0.252	0.010	-0.250	-25.348	0.000	-0.272	-0.233	0.636	1.571
Religião da mãe	0.000	0.000	0.006	0.745	0.456	0.000	0.001	0.944	1.059
Escolaridade da mãe	-0.093	0.005	-0.193	-20.322	0.000	-0.102	-0.084	0.687	1.456
Quintil de riqueza	-0.039	0.004	-0.109	-9.401	0.000	-0.047	-0.031	0.460	2.174
Estado civil da mãe	0.003	0.003	0.009	1.067	0.286	-0.003	0.009	0.966	1.035
Sexo da criança	-0.183	0.008	-0.179	-22.658	0.000	-0.199	-0.167	0.989	1.011
Número de consultas pré-natais	-0.167	0.008	-0.161	-20.055	0.000	-0.183	-0.151	0.960	1.042
Tipo de parto	-0.332	0.023	-0.115	-14.439	0.000	-0.377	-0.287	0.970	1.031
Tamanho da criança ao nascer	-0.003	0.007	-0.004	-0.473	0.636	-0.017	0.010	0.972	1.029
Faixa etária da mãe	-0.032	0.004	-0.097	-8.248	0.000	-0.040	-0.025	0.452	2.215
Ordem de nascimento	-0.058	0.007	-0.093	-7.752	0.000	-0.072	-0.043	0.429	2.332

ANEXOS

Anexo A - Código de Aplicação da Técnica SMOTE

Código Python para Balanceamento com SMOTE:

```
# APLICAÇÃO DA TÉCNICA SMOTE
import pandas as pd
import numpy as np
import pyreadstat
from imblearn.over_sampling import SMOTE
caminho_arquivo = "Base25.sav"
df, meta = pyreadstat.read_sav(caminho_arquivo)
df = df.astype(int)
variaveis = [
    'Regiao', 'Zona_Resid', 'Religiao', 'Escolaridade',
    'Quintil_Riqueza', 'Estado_Civil', 'Sexo_Bebe',
    'Nr_Consultas', 'Tipo_Partto', 'Tamanho_Bebe',
    'Idade_Mae', 'Ordem_Nasc']
X = df[variaveis]
y = df['Obito']
X_const = pd.get_dummies(X, drop_first=False)
smote = SMOTE(random_state=42)
X_res, y_res = smote.fit_resample(X_const, y)
df_res = pd.DataFrame(X_res, columns=X_const.columns)
df_res['Obito'] = y_res
df_res.to_excel("base_balanceada.xlsx", index=False)
df_res.to_csv("base_balanceada.csv", index=False)
pyreadstat.write_sav(df_res, "base_balanceada.sav")
print("Base original:", df.shape)
print("Base balanceada:", df_res.shape)
print("Arquivo criado com sucesso!")
```

Descrição do Processo:

- **Base Original:** 7.076 observações (5,1% óbitos vs 94,9% não-óbitos)
- **Variáveis:** 12 preditoras categóricas convertidas para formato dummy

CONTACTOS

Contactos:

823315932/842950005/878835094

Jacinto Pedro Cumbana

E-mail:

jacintopedro.cumbana@gmail.com

jacintopedro.cumbane@yahoo.com