UNIVERSIDADE
EDUARDO
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

Análise da Prevalência e dos Factores Associados à Anemia em Crianças menores de 5 Anos em Moçambique

Autor: Edilson Neilly Raúl André

Maputo, Agosto de 2025

UNIVERSIDADE
EDUARDO
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

Análise da Prevalência e dos Factores Associados à Anemia em Crianças menores de 5 Anos em Moçambique

Autor: Edilson Neilly Raúl André

Supervisor: Zacarias Bernardo Mutombene, MSc, UP

Maputo, Agosto 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 Licenciado em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Agosto de 2025

(**Edilson Neilly Raúl André**)

Dedicatória

Dedico este trabalho a toda minha família, em especial à minha mãe, **Fátima Sadique**.

Agradecimentos

Agradeço primeiramente à **Deus** que, por sua infinita graça e misericórdia, deu-me amparo em todos momentos, força e ânimo nas dificuldades e sabedoria para lidar com os momentos de ansiedade e medo durante toda a minha trajetória acadêmica.

Ao meu supervisor, **Zacarias Bernardo Mutombene, MSc, UP**, pela disponibilidade, pelo apoio e orientação dispensada durante a realização do presente trabalho.

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

Aos meus pais, **Raúl Anselmo André** e **Fátima Sadique**, não só por me terem dado a vida, mas também por me ensinarem princípios e valores que me tornaram o homem que sou hoje. Agradeço-vos por cada ensinamento, por cada sacrifício e por acreditarem sempre em mim.

À minha irmã, **Karina Nelsa Raúl André**, por estar sempre ao meu lado, confortando-me com as suas palavras encorajadoras e fazendo-me acreditar mais em mim. A nossa união torna tudo melhor e sou imensamente grato por tê-la na minha vida.

À minha tia **Cecília Uresse**, pelo apoio ao longo desta caminhada. A sua presença, incentivo e palavras de sabedoria foram fundamentais nos momentos desafiantes. Agradeço profundamente por todo o carinho e suporte que sempre me proporcionou.

Aos meus colegas **Joaquim Muiambo**, **Mércio Caniço** e **Stélio Andela** , pelo companheirismo, partilha de conhecimento e apoio ao longo desta caminhada. Foram fundamentais nos desafios, tornando cada etapa mais leve e enriquecedora.

A **todos** aqueles que não mencionei, mas que, directa ou indirectamente, contribuíram nas diversas fases da minha caminhada acadêmica e a todos que tornaram possível a realização do presente trabalho.

Meu Muito Obrigado!!!

"Mesmo desacreditado e ignorado por todos,
não posso desistir, pois para mim, vencer é
nunca desistir."

— Albert Einstein

Lista de Abreviaturas

AAS	Amostragem Aleatória Simples
AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
DHS	Demographic and Health Survey
EP	Erro Padrão
FGEPA	Factor Generalizado do Efeito do Plano Amostral
G/dL	Gramas por decilitro
I.C	Intervalo de Confiança
IDS	Inquérito Demográfico e de Saúde
IMC	Índice de Massa Corporal
INE	Instituto Nacional de Estatística
MISAU	Ministério da Saúde
MLG	Modelos Lineares Generalizados
MPV	Máxima Pseudo-Verossimilhança
MV	Máxima Verossimilhança
OMS	Organização Mundial da Saúde
OR	Odds Ratio
UEM	Universidade Eduardo Mondlane
UNICEF	United Nations International Children's Emergency Fund
UPA	Unidade Primária de Amostragem
WHO	World Health Organization

Resumo

A anemia é uma condição caracterizada pela insuficiência de glóbulos vermelhos no sangue e, consequentemente, pela redução da capacidade de transporte de oxigénio para os tecidos do organismo. Em Moçambique, a anemia em crianças menores de 5 anos mantém-se entre as mais elevadas do mundo. O presente estudo propôs-se analisar a prevalência e os factores associados à anemia em crianças menores de 5 anos em Moçambique. Para este fim, foram usados dados referentes ao Inquérito Demográfico e de Saúde (IDS 2022-2023), realizado em Moçambique, obtidos no site da Demographic Health Survey (DHS), mediante um pedido de autorização efectuado pelo autor. Foram usados procedimentos de análise de dados complexos por forma a acomodar o desenho de amostragem do IDS 2022-2023. Para a análise de dados foram utilizados gráficos, tabelas de frequência e para verificar a influência das variáveis independentes sobre a variável de interesse foi usada a técnica de regressão logística binária. A amostra estudada é composta por 3.388 crianças menores de 5 anos, onde cerca de 73.0% foram classificadas como anémicas e apenas 27.0% como não anémicas. Observou-se que a maioria das crianças vive em domicílios pertencentes ao quintil de riqueza mais baixo (26.6%), 19.7% pertencem ao quintil médio e apenas 12.4% ao quintil de riqueza mais elevado. Com base na análise bivariada, foram seleccionadas as variáveis a incluir no modelo multivariado. Estão estatisticamente associadas ao diagnóstico de anemia na criança as seguintes variáveis: idade da criança, área de residência, província de residência, quintil de riqueza, idade da mãe, estado civil da mãe, nível de escolaridade da mãe, estado nutricional da mãe, diagnóstico de anemia da mãe, intervalo entre os nascimentos, duração da amamentação, principal fonte de abastecimento de água, local de parto, acesso a rádio e acesso a televisão. Verificou-se que o modelo final ajustado aos dados apresentou uma capacidade de classificações correctas de 73.3%, evidenciando um desempenho preditivo aceitável.

Palavras-chave: Anemia em Crianças, Inquérito Demográfico e de Saúde, Moçambique, Regressão Logística Binária.

Abstract

Anemia is a condition characterized by an insufficient number of red blood cells, resulting in a reduced capacity to transport oxygen to body tissues. In Mozambique, anemia among children under the age of five remains among the highest in the world. This study aimed to analyze the prevalence and associated factors of anemia in children under five years of age in Mozambique. Data were obtained from the 2022–2023 Demographic and Health Survey (DHS), accessed through an authorization request submitted by the author. Statistical procedures suitable for complex sampling designs were applied to account for the survey's sampling structure. The analysis included graphs, frequency tables, and binary logistic regression to identify factors associated with childhood anemia. The sample consisted of 3,388 children under five years of age, approximately 73.0% of whom were classified as anemic. Most children lived in households in the lowest wealth quintile (26.6%). Based on the bivariate analysis, variables were selected for the multivariate model. Factors statistically associated with childhood anemia included: child's age, area and province of residence, wealth quintile, mother's age, marital status, education level, nutritional status and anemia diagnosis, birth interval, duration of breastfeeding, main source of drinking water, place of delivery, and access to media (radio and television). The final model showed a correct classification rate of 73.3%, indicating acceptable predictive performance.

Keywords: Anemia in Children, Demographic and Health Survey, Mozambique, Binary Logistic Regression.

Índice

1 Introdução **1**

1.1 Contextualização 1
1.2 Formulação do Problema 2
1.3 Objectivos 3
1.3.1 Objectivo Geral 3
1.3.2 Objectivos Específicos 3
1.4 Justificativa 3
1.5 Breve descrição dos capítulos do trabalho 4

2 Revisão da Literatura **5**

2.1 Enquadramento Teórico sobre a Anemia 5
2.2 Prevalência de Anemia em Crianças 6
2.3 Factores associados à anemia em crianças 8
2.3.1 Factores Socioeconómicos 8
2.3.2 Factores relacionados à Saúde Materno-Infantil 8
2.3.3 Factores Demográficos e Factores Ambientais 9
2.4 Modelos Lineares Generalizados 10
2.4.1 Família Exponencial 11
2.4.2 Características do Modelo Linear Generalizado 11
2.4.3 Estimatição dos Parâmetros 13
2.4.4 Selecção do Modelo 14
2.5 Modelos de Regressão Logística 15
2.6 Regressão Logística Binária 16
2.6.1 Transformação Logit 17
2.6.2 Regressão Logística Simples 18
2.6.3 Regressão Logística Múltipla 19
2.6.4 Estimação dos Parâmetros na Regressão Logística Binária 19
2.6.5 Multicolinearidade 20
2.6.6 Razão de Chances (Odds) 22
2.7 Análse de Dados Amostrais Complexos 23

3 Material e Métodos **24**
3.1 Material 24
3.2 Métodos 26
3.2.1 Classificação do estudo 26
3.2.2 Associação entre as Variáveis do Estudo 27
3.2.3 Testes de Significância dos Parâmetros do Modelo 28
3.2.4 Qualidade de Ajuste na Regressão Logística Binária 30
3.2.5 Poder de classificação (ou discriminação) do modelo 32
3.2.6 Curva ROC 33

4 Resultados e Discussão **35**
4.1 Resultados 35
4.1.1 Caracterização da amostra em estudo 35
4.1.2 Análise de Associação entre as Variáveis do Estudo 39
4.2 Análise dos pressupostos do Modelo 40
4.2.1 Verificação da Multicolinearidade 40
4.3 Modelo de Regressão Logística Binária 40
4.3.1 Estimação do modelo 40
4.3.2 Avaliação da Qualidade do Ajuste do Modelo 41
4.3.3 Interpretação das Estimativas para a Razão de Chances 42
4.3.4 Tabela de classificação do modelo final 43
4.4 Discussão 45

5 Conclusões e Recomendações **46**
5.1 Conclusões 46
5.2 Recomendações 48
5.3 Limitações 48

Referências **49**

Anexos **57**

Lista de Figuras

2.1 Distribuição percentual da prevalência de anemia em crianças de 6 a 59 meses por província em Moçambique. 7

4.1 Prevalência de anemia em crianças menores de 5 anos em Moçambique . . . 35
4.2 Distribuição percentual das crianças de acordo com o quintil riqueza (a) e área de residência (b) 36
4.3 Curva ROC 44

5.1 Distribuição percentual das crianças segundo sexo da criança 55
5.2 Distribuição percentual das crianças segundo o nível de escolaridade da mãe 55

Lista de Tabelas

2.1 Ligações canônicas mais comuns nos MLG 12

3.1 Descrição das variáveis do estudo 25
3.2 Tabela de classificação 32
3.3 Valores indicativos da área sobre a curva de ROC 34

4.1 Percentagem de anemia de acordo com as características socioeconómicas, demográficas e ambientais 38
4.2 Teste F de Thomas e Rao para análise de associação entre as variáveis 39
4.3 Teste de Multicolinearidade 40
4.4 Teste de Razão de Verossimilhança de Rao-Scott para as variáveis no modelo. 41
4.5 Teste F-ajustado de Archer e Lemeshow para o modelo final 41
4.6 Modelo de regressão logística binária 43
4.7 Tabela de classificação 44

5.1 Percentagem de anemia de acordo com as características socioeconómicas, demográficas e ambientais (Continuação da Tabela (4.1)) 56

Capítulo 1

Introdução

1.1 Contextualização

Segundo Benoist *et al.* (2009), a anemia é um problema de saúde pública que acfeta tanto países desenvolvidos quanto países em desenvolvimento, gerando consequências negativas à saúde humana, bem como para o desenvolvimento socioeconômico, com maior prevalência nas mulheres em idade reprodutiva e crianças menores de 5 anos devido à maior demanda nutricional nestes períodos.

De acordo com a World Health Organization (WHO, 2017), a anemia é uma condição médica caracterizada pela redução da concentração sanguínea ou hemoglobina abaixo dos limites considerados normais, o que resulta em uma quantidade insuficiente dessa proteína vital para atender às demandas fisiológicas do organismo.

Segundo a WHO (2021), em 2019 a prevalência global de anemia foi de 22,8%, afectando aproximadamente 1,74 bilhões de pessoas com maior incidência entre crianças menores de 5 anos (39,8%). Em termos de gravidade, a maioria dos casos de anemia foi classificada como ligeira (54,1%), seguida por moderada (42,5%) e severa (3,4%). As regiões que enfrentaram a maior carga de anemia foram o Sul da Ásia, a África Ocidental e Central.

De acordo com Ministerio da Saúde (MISAU *et al.*, 2015), em Moçambique, cerca de 64% das crianças entre 6-59 meses de idade foram identificadas com anemia, dentre estas 34% sofrem de anemia moderada e 3% sofrem de anemia severa. A prevalência de anemia em crianças menores de 5 anos diminuiu ligeiramente de 69% para 64% entre 2011 e 2015. No entanto, apesar dessa redução, a anemia continua a afectar de forma preocupante essa faixa etária, especialmente as crianças menores de 18 meses, onde a prevalência permanece elevada.

Segundo Balarajan *et al.* (2011), a anemia traz consequências que podem repercutir negativamente em todos os estágios da vida. Nas crianças, por exemplo tem sido associada ao atraso no crescimento, atraso no desenvolvimento cognitivo e psicomotor, dificuldades de aprendizagem dificuldades e concentração e comprometimento da imunidade, com aumento da predisposição à infecções. Durante a gravidez, pode estar associada ao aumento da mortalidade materna e infantil, bem como à maior incidência de partos prematuros e baixo peso à nascença, que são factores de risco para a ocorrência de anemia em crianças.

Segundo Magalhães *et al.* (2018), vários factores contribuem para a diminuição da concentração de hemoglobina e o aumento da prevalência de anemia em crianças menores de 5 anos. Além das deficiências nutricionais, como a falta de ferro e uma dieta com baixa biodisponibilidade desse mineral, factores socioeconómicos como a precariedade do saneamento básico, acesso limitado a serviços de saúde e o nível de escolaridade dos pais desempenham um papel crucial na prevalência de anemia. No contexto biológico, a idade da mãe, a idade da criança e o peso ao nascer também têm impacto significativo.

1.2 Formulação do Problema

Globalmente, crianças menores de 5 anos são as mais afectadas pela anemia quando comparadas com outros grupos populacionais. Apesar das diversas medidas adotadas para enfrentar este problema de saúde, a anemia continua a ser um dos desafios de saúde pública mais prevalentes em vários países do mundo, com destaque para o continente africano que possui as mais altas prevalências de anemia em crianças menores de 5 anos. Em 2019, a Organização Mundial da Saúde estimou que a prevalência global de anemia em crianças menores de 5 anos era de 39,8%, o que corresponde a aproximadamente 269 milhões de crianças, sendo a maior taxa registada na região africana (60,2%). Embora a prevalência global continue elevada, registou-se uma ligeira redução em relação aos 41,7% observados em 2016. No contexto Africano, Moçambique destaca-se por apresentar uma das mais altas prevalências de anemia em crianças, atingindo 68,9% em 2019 e ocupando a 13ª posição no ranking global (WHO, 2021) .

De acordo com o Instituto Nacional de Estatística (INE *et al.*, 2024), Moçambique, é um país com altos índices de pobreza e desafios nutricionais, a anemia em crianças menores de 5 anos é uma preocupação de saúde pública crítica, a prevalência de anemia em criança é alarmante, com quase todas as províncias apresentando taxas superiores a 50%, exceto a Cidade de Maputo, que tem uma prevalência de 44%. Este cenário sugere que diversos factores subjacentes, como a desnutrição, doenças infecciosas, condições socioeconômicas precárias e acesso limitado aos serviços de saúde, contribuem para a persistência deste problema no país.

Diante do problema descrito faz-se a seguinte questão: **Quais são os factores associados à alta prevalência de anemia em crianças menores de 5 anos em Moçambique?**

1.3 Objectivos

1.3.1 Objectivo Geral

Analisar a prevalência e os factores associados à anemia em crianças menores de 5 anos em Moçambique.

1.3.2 Objectivos Específicos

- Descrever as características da amostra referente à anemia em crianças menores de 5 anos em Moçambique;
- Verificar a existência de associação entre as demais variáveis do estudo e a variável de interesse;
- Identificar os factores de risco associados à anemia em crianças menores de 5 anos em Moçambique;
- Estimar um modelo probabilístico que melhor relacione os potenciais factores à anemia em crianças menores de 5 anos em Moçambique.

1.4 Justificativa

Diante da elevada prevalência e das possíveis repercussões a longo prazo na saúde das crianças, é imprescindível compreender os factores que influenciam no desenvolvimento da anemia em crianças menores de 5 anos em Moçambique. Tendo em vista que a sintomatologia é expressa quando a doença já está cronicamente instalada, investigar factores como condições socioeconômicas, hábitos alimentares, as condições de saneamento, a assistência à saúde insuficiente e factores biológicos relacionados às crianças torna possível estabelecer vínculos precoces com a doença e reduzir as chances de comprometimento cognitivo e formulação de políticas de saúde pública mais eficazes e direcionadas à melhorar a saúde e o desenvolvimento infantil em Moçambique.

1.5 Breve descrição dos capítulos do trabalho

No **primeiro capítulo** foi abordada a introdução ao tema, onde foi apresentada a contextualização do problema em estudo, a formulação do problema, os objectivos do estudo, bem como a sua justificativa. Na secção a seguir, será apresentado o **segundo capítulo**, onde é feita a revisão da literatura, abordando os principais conceitos relacionados à anemia em crianças menores de 5 anos e conceitos básicos das principais as técnicas utilizadas para a concretização do trabalho. No **terceiro capítulo**, é apresentada a fonte de obtenção de dados, as variáveis do estudo e é feita uma abordagem referente à metodologia do trabalho. No **quarto capítulo**, são apresentados os resultados obtidos e a discussão dos mesmos de forma comparativa com alguns estudos disponíveis. No **quinto capítulo**, são apresentadas as conclusões e algumas recomendações à luz dos objectivos estabelecidos e dos resultados obtidos.

Capítulo 2

Revisão da Literatura

2.1 Enquadramento Teórico sobre a Anemia

A anemia é uma condição em que a quantidade de células vermelhas no sangue e, consequentemente, a capacidade de transporte de oxigénio são insuficientes para atender às necessidades do organismo. No caso das crianças, essas necessidades são particularmente elevadas devido ao rápido crescimento e desenvolvimento, tornando-as mais vulneráveis aos efeitos da anemia, que pode comprometer o desenvolvimento cognitivo, a imunidade e o bem-estar geral.

A anemia é definida, no caso de crianças com menos de 5 anos de idade, como uma condição caracterizada pela concentração de hemoglobina (Hb) no sangue abaixo de 11 g/dL (WHO, 2015). Esse parâmetro é amplamente utilizado como referência global para a identificação e o monitoramento da prevalência de anemia em populações pediátricas, permitindo a avaliação do impacto de intervenções e programas de saúde pública (WHO, 2015).

De acordo com o INE *et al.* (2024), a anemia em crianças é definida com base nos níveis de hemoglobina no sangue, expressos em gramas por decilitro (g/dL), sendo classificada em diferentes níveis de severidade. Considera-se anemia em qualquer caso em que a hemoglobina esteja abaixo de 11,0 g/dL. Quando os valores variam entre 10,0 e 10,9 g/dL, a condição é classificada como anemia ligeira. Para níveis entre 7,0 e 9,9 g/dL, trata-se de anemia moderada, enquanto valores inferiores a 7,0 g/dL caracterizam a forma severa da doença.

Segundo Oliveira *et al.* (2007), embora a cadeia etiopatogénica da anemia seja um processo biológico decorrente do balanço negativo entre as necessidades orgânicas e a ingestão ou absorção de ferro, ou de outros nutrientes essenciais à hematopoiese, estudos populacionais demonstram que sua ocorrência está fortemente associada a factores socioeconómicos e ambientais, sendo mais frequente em crianças menores de 5 anos, gestantes, mulheres em idade reprodutiva e idosos.

2.2 Prevalência de Anemia em Crianças

A anemia continua a ser um dos principais desafios de saúde pública global, afectando desproporcionalmente crianças menores de 5 anos, sobretudo nos países em desenvolvimento. Segundo a WHO (2021), a prevalência global de anemia nessa faixa etária é estimada em 39.8%, com variações significativas entre regiões. O continente africano apresenta a maior prevalência, atingindo 60.2%, seguido pelo Sudeste Asiático (49.0%) e a região do Mediterrâneo Oriental com 42.7%. Por outro lado, a América (16.5%) e a Europa (20.3%) registam os valores mais baixos.

No continente africano, os países com maior prevalência de anemia são o Mali (79.0%), Burkina Faso (76.6%), Guiné (73.8%), República Centro-Africana (73,6%) e Serra Leoa (73.4%). Por outro lado, os níveis mais baixos foram registados nas Seicheles (30.6%), Maurícia (30.7%), Argélia (34,3%), Zimbábue (37,8%) e Ruanda (37.9%). O continente africano, além de apresentar maior carga total, também regista elevadas proporções de anemia moderada (32.0%), ligeira (25.5%) e severa com 2.7% (WHO, 2021).

Em Moçambique, de acordo com o MISAU *et al.* (2015), crianças mais novas são as mais afectadas, com a maior prevalência observada no grupo etário de 6 a 8 meses (77.6%). Entre as crianças deste grupo etário, 40.5% apresentam anemia moderada e 8% anemia severa. À medida que a idade avança, verifica-se uma tendência de redução da prevalência, atingindo o seu menor valor entre crianças de 48 a 59 meses (53.5%), sendo que apenas 1.5% desse grupo apresenta anemia severa.

No que diz respeito ao sexo da criança, não se verifica diferenças expressivas na prevalência de anemia, com valores de 66.4% entre crianças do sexo masculino e 61.2% entre crianças do sexo feminino. Crianças residentes em áreas rurais apresentam uma prevalência maior (66.2%) em comparação com as de áreas urbanas (56.4%). Além disso, observa-se que a gravidade da anemia também varia conforme a área de residência. A anemia severa é mais prevalente entre crianças residentes em áreas rurais (3.7%) em comparação com crianças residentes em áreas urbanas (1.4%).

A distribuição percentual da prevalência de anemia por província revelou variações entre regiões, evidenciando um padrão preocupante no centro e norte do país. As províncias de Zambézia (76,7%), Cabo Delgado (73,2%), Sofala (63,7%) e Nampula (67,6%) apresentam as maiores prevalências, sugerindo a influencia de factores estruturais, como o acesso limitado a serviços de saúde, deficiências nutricionais e condições socioeconómicos desfavoráveis. Em contrapartida, Maputo Província (45,0%) e Maputo Cidade (49,6%) registam as menores prevalências de anemia, seguidas por Inhambane (54,3%), refletindo possivelmente melhores

condições socioeconómicos e maior acesso a serviços de saúde (MISAU *et al.*, 2015).

A prevalência de anemia nas crianças varia conforme a idade da mãe. Crianças de mães com menos de 20 anos apresentam a maior prevalência de anemia (69,3%), sendo a maioria dos casos de anemia moderada (44,5%) e uma maior proporção de anemia severa (4,2%). Entre as crianças de mães com 35 anos ou mais, a prevalência é mais baixa (59,0%), com menor ocorrência de anemia severa (3,4%). No que se refere ao nível de escolaridade, a prevalência de anemia é mais elevada entre crianças de mães sem escolaridade (66,7%), seguida pelas que possuem ensino primário (65,4%). A menor prevalência é observada entre crianças de mães com ensino secundário ou superior (55,7%). Em relação à gravidade da anemia, a forma ligeira é mais frequente em mães com ensino secundário (29,0%), enquanto a anemia severa apresenta maior percentagem entre aquelas cujas mães não possuem escolaridade (3,5%).

Observa-se uma tendência semelhante em relação ao quintil de riqueza. Crianças provenientes de famílias no quintil mais baixo de riqueza apresentam uma prevalência de anemia de 73,6%, enquanto aquelas no quintil mais elevado têm uma prevalência de 50,4%. Além disso, a prevalência de anemia severa é quase seis vezes maior entre as crianças provenientes de famílias pertencentes ao quintil mais baixo de riqueza (6,55%) em comparação com crianças de famílias pertencentes ao quintil mais elevado (0,8%) (MISAU *et al.*, 2015).

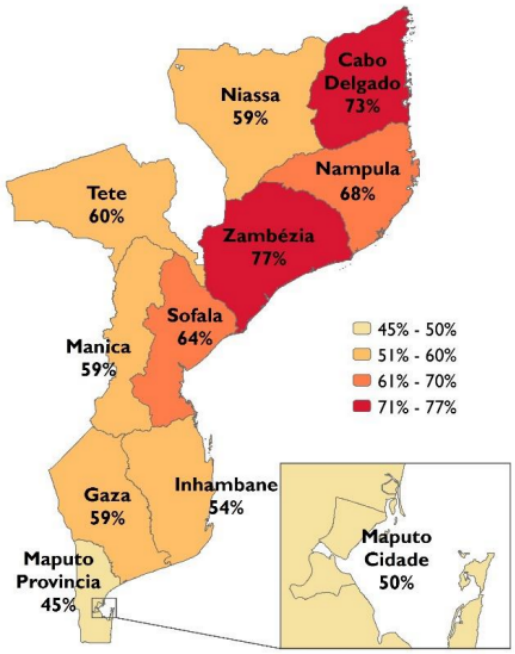

Figura 2.1: Distribuição percentual da prevalência de anemia em crianças de 6 a 59 meses por província em Moçambique. **Fonte:** MISAU *et al.*, 2015.

2.3 Factores associados à anemia em crianças

De acordo com Netto *et al.* (2006), a anemia não resulta somente de uma alimentação inadequada ou de deficiências no cuidado infantil e nas práticas alimentares, mas sim de uma combinação de diversos factores de ordem socioeconómica, demográficos e ambientais.

2.3.1 Factores Socioeconómicos

Os factores socioeconómicos relacionados à anemia em crianças menores de 5 anos englobam aspectos como a área de residência (urbana ou rural), a província de residência, o nível educacional dos pais, a exposição à meios de comunicação como a rádio, a televisão, jornais ou revistas, o quintil de riqueza, entre outros.

Segundo a UNICEF (2014), no que concerne a área de residência, em Moçambique persistem diferenças nos níveis de pobreza, de desenvolvimento humano e de bem-estar infantil, estando quase todos os indicadores em maior desvantagem nas áreas rurais, para o Norte e Centro do país em comparação com as áreas urbanas e a região Sul.

De acordo com Silva (2007), o nível socioeconómico esta directamente relacionado a anemia e afecta a sua prevalência de variavs formas. A pobreza, enquanto determinante social da saúde, está associada a condições de vida e de trabalho precárias, incluindo o acesso limitado a água potável, saneamento e higiene, factores que favorecem o surgimento e agravamento de doenças. Além disso, a pobreza contribui para a insegurança alimentar, a baixa qualidade da dieta e a dificuldade no acesso aos serviços de saúde, factores que, em conjunto, comprometem o estado nutricional e aumentam o risco de anemia, especialmente entre populações vulneráveis, como as crianças menores de 5 anos.

Segundo Vasconcelos (2020), o nível de escolaridade dos responsáveis está directamente associado à qualidade dos cuidados com as crianças, influenciando tanto no acesso a informações sobre saúde e nutrição quanto a adopção de práticas adequadas. Este factor está igualmente relacionado a melhores oportunidades de emprego, o que pode impactar positivamente a renda familiar e, consequentemente, o acesso a uma alimentação adequada.

2.3.2 Factores relacionados à Saúde Materno-Infantil

A saúde materno-infantil desempenha um papel crucial no estado nutricional e hematológico das crianças, influenciando directamente o risco de anemia em menores de 5 anos. Factores como o estado nutricional da mãe durante a gestação, a prática do aleitamento materno, a qualidade da alimentação complementar e a frequência de consultas pré-natal e pós-natal podem impactar significativamente os níveis de hemoglobina das crianças.

De acordo com Silva *et al.* (2015), assistência pré-natal e ao parto devem ser eficientes no sentido de evitar e corrigir os principais problemas relacionados à saúde e nutrição da gestante, que podem desencadear o baixo peso ao nascer e a prematuridade que são factores de risco para a anemia. Neste contexto, mães que apresentam anemia ou deficiências nutricionais tendem a ter crianças com reservas reduzidas de ferro, aumentando a probabilidade de desenvolver anemia nos primeiros meses de vida.

Segundo Melo (2020), o baixo peso ao nascer está fortemente associado ao risco de desenvolver anemia em recém-nascidos, pois reflete deficiências nutricionais maternas que comprometem as reservas de ferro do recém-nascido. A anemia materna, frequentemente associada a uma alimentação inadequada durante a gravidez, pode prejudicar o desenvolvimento fetal, resultando em menores níveis de hemoglobina nos primeiros meses de vida e, consequentemente, aumentar o risco de contrair anemia.

De acordo com Leite *et al.* (2013), a anemia materna é frequentemente associada à ocorrência de anemia na criança. Essa associação pode ser explicada pela exposição simultânea da mãe e da criança a factores de risco, como condições nutricionais deficitárias, elevada exposição a infeções e acesso limitado a serviços de saúde de qualidade. Além disso, a presença de anemia durante a gestação aumenta o risco de complicações obstétricas, podendo comprometer o desenvolvimento fetal e reduzir as reservas de ferro no recém-nascido, o que contribui para uma maior probabilidade de anemia nos primeiros meses de vida.

Outro aspecto fundamental é a prática do aleitamento materno. O aleitamento materno exclusivo, recomendado até os seis primeiros meses de vida, não apenas garante o fornecimento de nutrientes essenciais, mas também protege contra infecções que podem agravar a deficiência de ferro. Por outro lado, a introdução precoce de alimentos complementares ou a interrupção do aleitamento materno pode comprometer a absorção do ferro, contribuindo para o desenvolvimento de anemia (Tadesse *et al.*, 2022).

2.3.3 Factores Demográficos e Factores Ambientais

Os factores demográficos associados à anemia em crianças menores de 5 anos incluem uma combinação de características individuais, maternas e do agregado familiar. A idade da criança é um dos principais determinantes na vulnerabilidade à anemia. Crianças com idade entre 6 e 24 meses de idade se caracterizam como grupo de risco para a anemia, devido ao curto período de aleitamento materno, que muitas vezes não se dá de maneira exclusiva, baixa ingestão de alimentos ricos em ferro, crescimento acelerado, além da substituição do leite materno por leite de vaca e/ou outros alimentos com menor biodisponibilidade de ferro (Silva *et al.*, 2015).

Segundo Amaral *et al.* (2021), mães jovens, entre 15 e 24 anos de idade, tendem a ter menos experiência e conhecimento sobre práticas alimentares adequadas, o que pode comprometer a nutrição dos filhos.

As condições ambientais referem-se ao meio em que as crianças são criadas e influenciam a prevalência da desnutrição infantil. Um bom ambiente de higiene dentro da habitação constitui um factor importante, pois permite reduzir a frequência de ocorrência de doenças como, por exemplo, a diarreia, a qual está relacionada com a prevalência da desnutrição, condição que constitui um factor de risco para a anemia.(Massango, 2018).

Em Moçambique, o uso de fontes melhoradas de água e de meios sanitários seguros ainda é baixo, com práticas como o fecalismo a céu aberto persistindo. Além disso, uma parte reduzida da população tem acesso a fontes de água potável, principalmente nas áreas rurais e nas famílias mais desfavorecidas, ou seja, nas famílias dos quintis de riqueza mais baixos (MISAU *et al.*, 2015).

Segundo Moraes *et al.* (2019), a anemia está fortemente associada a factores ambientais, como saneamento básico inadequado, higiene precária e condições habitacionais desfavoráveis. Esses factores aumentam a incidência de infecções e comprometem a absorção de nutrientes essenciais, agravando o quadro anémico, especialmente em crianças, que são mais vulneráveis devido ao contacto frequente com ambientes contaminados. Essa relação evidencia a importância de intervenções voltadas para a melhoria do saneamento, da nutrição e da educação em saúde como estratégias fundamentais para a prevenção e controle da anemia.

A aglomeração no agregado familiar é uma característica comum em muitas regiões menos desenvolvidas, onde a alta taxa de natalidade e as condições de moradia precária, são comuns. Esta configuração pode ter implicações directas na saúde das crianças, incluindo a prevalência de anemia. A presença de muitos moradores, especialmente de crianças menores de 5 anos, está associada a uma maior vulnerabilidade nutricional e dificuldade de acesso a cuidados de saúde adequados, factores que favorecem o desenvolvimento de anemia (Castro *et al.*, 2023).

2.4 Modelos Lineares Generalizados

Segundo Turkman e Silva (2000), em muitos estudos estatísticos, quer sejam de natureza experimental ou observacional, surgem frequentemente problemas em que o objectivo principal é o estudar a relação entre variáveis, ou mais particularmente, analisar a influência que uma ou mais variáveis explicativas, medidas em indivíduos ou objectos, têm sobre uma variável de interesse, denominada variável resposta.

O modelo de regressão clássico é, provavelmente, o mais utilizado para modelar a relação entre uma variável resposta e as variáveis explicativas. Este modelo assume que o valor esperado da variável resposta é uma combinação linear das variáveis explicativas e que a distribuição da variável resposta é normal. No entanto, esta abordagem apresenta limitações, pois não é adequada quando a variável resposta segue uma distribuição diferente da normal, como é o caso de variáveis binárias (Silva, 2014).

Para resolver este problema, Nelder e Wedderburn (1972), propuseram os Modelos Lineares Generalizados (MLGs), que são uma generalização dos modelos de regressão linear clássicos. A ideia básica envolvida na criação destes modelos consiste em estender as opções da distribuição da variável resposta para englobar todas aquelas que pertencem à família exponencial, permitindo também flexibilizar a relação funcional entre a variável resposta e as variáveis explicativas do modelo de regressão (Paula, 2013).

2.4.1 Família Exponencial

Segundo Alvarenga (2015), a variável aleatória Y tem distribuição pertencente à família exponencial de dispersão (ou simplesmente família exponencial) se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) se puder escrever na forma:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \tag{2.1}$$

Onde:

- Y é a variável de interesse;
- θ é o parâmetro de localização;
- ϕ é o parâmetro de dispersão;
- ϕ é o parâmetro de dispersão associado à variância;
- $a(.)$, $b(.)$ e $c(.)$ são funções reais conhecidas ou também denominadas específicas.

2.4.2 Características do Modelo Linear Generalizado

Segundo Agresti (2015), os MLG podem ser usados quando se tem uma variável aleatória Y associada a um conjunto de variáveis explicativas $x_1, \ldots, x_p$. Para uma amostra de n observações (y_i, x_i), em que $x_i = (x_{i1}, \ldots, x_{ip})^T$ representa o vector coluna de variáveis explicativas, os MLG contêm três componentes:

1. **Componente aleatória**: Dado o vector de covariáveis x_i, as variáveis aleatórias Y_i são condicionalmente independentes com distribuição pertencente à família Exponencial, e, portanto, o seu valor médio é dado por:

$$f(Y_i; x_i) = \mu_i = b(\theta), \quad i = 1, 2, 3, \ldots, n \tag{2.2}$$

2. **Componente sistemática**: Define-se o preditor linear i como combinação linear das variáveis explicativas, ou seja, a componente sistemática é formada pelas variáveis explicativas, que entram sob forma de uma soma linear de seus efeitos, dando origem a um vector de preditores lineares, representado pela expressão dado por:

$$\eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{2.3}$$

   Onde, $x = (x_1, \ldots, x_k)$ - representa a matriz do modelo, $\beta = (\beta_1, \ldots, \beta_k)$ - é o vector de parâmetros desconhecidos .

3. **Função de ligação**: representa uma função que relaciona o componente aleatório ao componente sistemático, relacionando a média ao preditor linear, isto é

$$\eta_i = g(.) \tag{2.4}$$

   Onde $g(.)$ é função real, monótona e diferencial.

Na Tabela 2.1, a seguir, são apresentadas as ligações canônicas mais comuns na classe dos modelos lineares generalizados.

Tabela 2.1: Ligações canônicas mais comuns nos MLG

Distribuição	**Normal**	**Binomial**	**Poisson**	**Gama**	**N. Inversa**
Ligação	$\mu = \eta$	$\log\left(\frac{\mu}{1-\mu}\right) = \eta$	$\log \mu = \eta$	$\mu^{-1} = \eta$	$\mu^{-2} = \eta$

Fonte: Paula, 2013

De modo geral, um MLG é definido por uma distribuição da família exponencial (2.2), uma estrutura linear (2.3) e uma função de ligação (2.4), sendo importante na escolha do modelo definir a distribuição da variável resposta (i); a matriz do modelo, contendo o conjunto de variáveis explicativas (ii); e a função de ligação (iii).

2.4.3 Estimatição dos Parâmetros

De acordo com Fatoretto (2016), diversos métodos podem ser usados para a obtenção das estimativas dos β's, sendo o de máxima verossimilhança e o bayesiano os principais métodos de estimação.. O método de estimação de máxima verossimilhança é o mais usado, pois, o seu procedimento de estimação é relativamente simples e, ainda, por este possuir algumas propriedades importantes como, por exemplo, a consistência de seus estimadores.

Considerando-se uma amostra aleatória $y = (y_1, y_2, \ldots, y_n)$ com n observações de uma distribuição pertencente à família (2.1), a função de verossimilhança é dada pela equação (2.5)

$$L = L(y, \theta, \phi) = \prod_{i=1}^{n} f(y_i; \theta_i, \phi) = \exp\left(\sum_{i=1}^{n}\left[\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right]\right). \tag{2.5}$$

O logaritmo da função de verossimilhança é representado pela equação (2.6)

$$l = l(y, \theta, \phi) = \log L(y, \theta, \phi) = \sum_{i=1}^{n}\left[\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right]. \tag{2.6}$$

A partir da expressão em (2.6), procura-se obter o vector β de parâmetros que maximizam a função $l(y; \theta, \phi)$, pela regra da cadeia, derivando-se a função dada pela equação (2.7)

$$U_r = \frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial u_i}\frac{\partial u_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \phi^{-1}\sum_{i=1}^{n}(y_i - u_i)\frac{1}{V_i}\frac{du_i}{d\eta_i}x_{ij} \tag{2.7}$$

De acordo com Martinho (2023), a estimativa de máxima verossimilhança $\hat{\beta}$ é calculada igualando-se a zero a denominada função score U_j, para $j = 1, \ldots, p$. No entanto, em geral, estas equações não são lineares devendo ser resolvidas por processos iterativos como, por exemplo, o método de Newton-Raphson ou o método score de Fisher.

Pelo método de Newton-Raphson, de acordo com a equação em (2.8), tem-se:

$$\beta^{(m+1)} = \beta^{(m)} + [K^{(m)}]^{-1}U^{(m)} \tag{2.8}$$

$U^{(m)}$ é o vector score, com elementos $\left(\frac{\partial l}{\partial \beta_j}\right)$. $K^{(m)}$ é a matriz de informação observada, com elementos $\left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_j'}\right)$, em que se estabelece a igualdade $\beta = \beta^{(m)}$.

Substituindo a matriz de informação observada pela matriz de informação esperada de Fisher, definida como $I = -\mathbb{E}\left(\frac{\partial^2 l}{\partial\beta\,\partial\beta'}\right)$, que representa a matriz de covariância da função score, para um bom estimador de Máxima Verossimilhança, de acordo com a equação em (2.9), é possível obter:

$$\hat{\beta}^{(m+1)} = (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^{(m)}, \tag{2.9}$$

X é a matriz do modelo, W = diag$\{w_1, \ldots, w_h\}$ é uma matriz diagonal de pesos que capta a informação sobre a distribuição e $z^{(m)} = X\beta^{(m)} + \Delta^{(m)}(y-\mu)^{(m)}$ é o vector da variável dependente ajustada na etapa m, em que $\Delta = \text{diag}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$.

2.4.4 Selecção do Modelo

A selecção de modelos é uma parte importante de toda a investigação em modelação estatística e envolve a procura de um modelo que seja o mais simples possível e que descreva bem os dados observados. Na prática há geralmente um elevado número de variáveis que podem ser potencialmente importantes para explicar a variabilidade da variável resposta. Isto implica a existência de vários modelos com combinações diferentes das variáveis explicativas para explicar o fenómeno em causa, o que torna o processo da selecção mais difícil e mais moroso (Alvarenga, 2015).

Segundo Massango (2017), entre os diversos procedimentos para a selecção de variáveis e critérios de comparação da qualidade de ajuste dos modelos, destacam-se os métodos Stepwise, Forward, Backward, o Akaike's Information Criterion (AIC) e o Bayesian Information Criterion (BIC).

- **Selecção Forward:** Este método inicia com um modelo nulo, ou seja, sem nenhuma variável explicativa incluída. Em cada passo, é adicionada ao modelo a variável que apresenta maior contribuição estatística, desde que respeite o nível de significância previamente definido. Após a inclusão de cada variável, o modelo é ajustado novamente com a nova variável incluída, e este processo é repetido até que nenhuma das variáveis restantes cumpra os critérios estabelecidos para inclusão (Gonçalves, 2013).

- **Selecção Backward:** Este método inicia com o modelo completo, ou seja, contendo todas as variáveis explicativas. Em cada passo, é avaliada a contribuição estatística de cada variável, e aquela que apresentar a menor relevância, segundo os critérios definidos, é excluída. O modelo é ajustado novamente, agora sem esta variável, e o processo é repetido até que todas as variáveis remanescentes atendam aos critérios estabelecidos para permanência no modelo (Gonçalves, 2013).

- **Selecção Stepwise:** Este método é uma combinação dos dois métodos anteriores, inclui-se e/ou exclui-se uma variável explicativa dependendo do poder que a mesma tem na explicação da variável resposta (Gonçalves, 2013).

De acordo com Dias (2014), o critério de selecção AIC consiste em selecionar um modelo que seja parcimonioso, ou seja, que esteja bem ajustado aos dados e que utilize um número reduzido de parâmetros. Essa função se baseia na função de log verossimilhança acrescida de uma penalidade pelo número de parâmetros do modelo. Entre vários modelos candidatos, deve-se escolher aquele que apresentar o menor valor da função AIC.

A estimativa de AIC para um determinado modelo é dado por:

$$\text{AIC} = -2\ln(L) + 2k \tag{2.10}$$

Onde:

- $\ln(L)$ é o logaritmo da função de verossimilhança;
- k é o número de parâmetros no modelo.

Outro critério que possui o mesmo raciocínio que o AIC é o critério de Informação BIC. Neste caso, a selecção do melhor modelo também é baseada no menor BIC.

$$\text{BIC} = -2\ln(L) + k\ln(n) \tag{2.11}$$

sendo, $\ln(L)$ é o logaritmo da função de verossimilhança, k é o número de parâmetros no modelo e n é o tamanho da amostra.

2.5 Modelos de Regressão Logística

De acordo com Figueira (2006), a regressão logística pode ser considerada como sendo uma extensão da regressão linear, pois assim como na regressão linear, ela estuda a relação entre variáveis, identificando as variáveis que podem influênciar de alguma forma a variável resposta ou dependente. No entanto, enquanto a regressão linear fornece uma resposta métrica, a regressão logística trabalha com uma variável resposta categórica, expressando os resultados em termos de probabilidades ou chances de ocorrência do evento de estudo.

Segundo Alvarenga (2015), o modelo regressão logística é um dos casos mais importantes dos modelos lineares generalizados, usado quando se deseja explicar uma variável resposta categórica com dois ou mais resultados possíveis em função de uma ou mais variáveis explicativas quantitativas ou qualitativas.

Os modelos de regressão logística são amplamente utilizados para modelar variáveis dependentes categóricas, permitindo prever probabilidades associadas a diferentes categorias. A regressão logística binária é aplicada quando a variável resposta possui apenas duas categorias, como presença ou ausência de um determinado evento. A regressão logística ordinal, por sua vez, é utilizada quando as categorias da variável dependente seguem uma ordem hierárquica, como os diferentes níveis de gravidade de uma doença e quando não existe essa ordenação entre as categorias da variável dependente assume-se o contexto de regressão logística nominal (Maroco, 2007).

O Modelo de Regressão Logística permite:

1. Modelar a probabilidade de um evento ocorrer dependendo dos valores das variáveis independentes que podem ser categóricas ou contínuas;
2. Estimar a probabilidade de um evento ocorrer para uma observação selecionada aleatoriamente contra a probabilidade do evento não ocorrer;
3. Prever o efeito do conjunto de variáveis sobre a variável dependente binária;
4. Classificar observações, estimando a probabilidade de uma observação estar em uma categoria determinada.

Segundo Belfiore (2015), a variável dependente Y na regressão logística é frequentemente binária, logo, nestes casos ela segue a distribuição de Bernoulli, tendo uma probabilidade desconhecida p. Lembrando que a distribuição de Bernoulli é apenas um caso especial da distribuição binomial onde $n = 1$ (considera a realização de um único experimento).

$$Y = \begin{cases} 1, & \text{se } ocorrer\ sucesso \\ 0, & \text{se } ocorrer\ fracasso \end{cases}$$

A probabilidade de sucesso é $0 \leq p \leq 1$ e a probabilidade de fracasso é $q = 1-p$. Na regressão logística, é feita a estimação da probabilidade desconhecida p, dado uma combinação linear de variáveis independentes.

2.6 Regressão Logística Binária

A regressão logística é uma técnica estatística que tem como objectivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por variáveis categóricas, frequentemente binãria, em função de uma ou mais variáveis independentes continuas e/ou binarias.

Segundo Hosmer e Lemeshow (2000), a regressão logística binária é um modelo de regressão usado para descrever a relação entre uma variável resposta de carácter binário ou dicotómico e uma ou mais variáveis explicativas. A regressão logística é utilizada em situações nas quais pretende-se saber se o indivíduo possui ou não determinada característica em estudo (Goncalves, 2013).

De acordo com Fávero *et al.*, (2009), a regressão logística binária é uma técnica utilizada para descrever o comportamento entre uma variável dependente binária ou dicotómica e uma ou mais variáveis explicativas, destinando-se a investigar o efeito das variáveis pelas quais os indivíduos, objectos ou sujeitos estão expostos sobre a probabilidade de ocorrência de determinado evento de interesse.

2.6.1 Transformação Logit

De acordo com Silveira (2021), nos modelos de regressão logística binária, a variável resposta é dicotômica, assumindo apenas dois possíveis valores, representados como "1" a ocorrência do evento de interesse e "0" a sua ausência. No entanto, modelar directamente essa variável em um contexto de regressão linear não é adequado, pois os valores previstos poderiam ultrapassar o intervalo $0 \leq Y \leq 1$, dificultando sua interpretação como probabilidades.

Na regressão logística tem-se interesse em modelar a probabilidade de ocorrência do evento de interesse em função das variáveis explicativas. Para tal, seria inapropriada a aplicação do modelo regressão linear simples dado que para qualquer valor de x, $0 \leq \pi(x) \leq 1$, isto é, a média condicional de Y para o caso de variáveis dicotômicas deve assumir valores entre 0 e 1.

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} \tag{2.12}$$

β_0 e β_1 são os parâmetros do modelo, $\pi(x)$ representa a probabilidade de ocorrência ou não do evento de interesse em razão dos valores assumidos pela variável explicativa x.

A regressão logística assume a forma dada em (2.12). Contudo, essa relação não é linear nos parâmetros, ocasionando dificuldades na interpretação das variações na probabilidade de ocorrência do evento de interesse como resultado das variações nas variáveis explicativas. Deste modo, a regressão logística recorre ao conceito de *chances*, que representa o razão entre a probabilidade de sucesso e a sua complementar, por forma a relacionar a mudança na probabilidade de ocorrência do evento de interesse com as variáveis observadas em x (Gonzalez, 2018).

De acordo com Agresti (2002), para uma probabilidade de sucesso π, a *chance* é definida de acordo com a equação (2.13), como sendo:

$$\Omega = \frac{\pi}{1-\pi} \tag{2.13}$$

Se a probabilidade do evento ocorrer é definida de acordo com a expressão em (2.12), então a probabilidade de que o evento não ocorra é dada pela equação (2.14):

$$1 - \pi(x) = 1 - \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{\beta_0+\beta_1 x}} \tag{2.14}$$

O rácio entre a probabilidade do evento ocorrer e de que o mesmo não ocorra fica definido pela equação (2.15):

$$\frac{\pi(x)}{1-\pi(x)} = e^{\beta_0+\beta_1 x} \tag{2.15}$$

Aplicando o logaritmo natural na expressão (2.15), obtém-se a expressão em (2.16):

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{2.16}$$

2.6.2 Regressão Logística Simples

Segundo Figueira (2006), a regressão logística simples é uma técnica estatística utilizada para modelar a relação entre uma variável dependente dicotómica, ou seja, que assume apenas dois possíveis resultados e uma única variável independente. O modelo estima a probabilidade de ocorrência do evento de interesse em função de uma variável explicativa.

Seja uma variável Y com distribuição de Bernoulli de parâmetro $E(Y) = \pi$, assumindo dois resultados possíveis, geralmente representados por 1 ou 0, indicando a ocorrência ou não do evento de interesse, com probabilidades π e $1 - \pi$.

A probabilidade de sucesso do modelo logístico simples é definida pela equação (2.17)

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}, \tag{2.17}$$

e a probabilidade de fracasso

$$1 - \pi(x) = \frac{1}{1+e^{\beta_0+\beta_1 x}}, \tag{2.18}$$

β_0 e β_1 são os parâmetros do modelo, $\pi(x)$ representa a probabilidade de ocorrência ou não do evento de interesse em razão dos valores assumidos pela variável explicativa x (Kutner et al., 2005; Hosmer e Lemeshow, 2000).

2.6.3 Regressão Logística Múltipla

Segundo Hosmer e Lemeshow (2000), a regressão logística múltipla representa o contexto de regressão logística, em que a variável dependente Y é binária ou dicotômica, ou seja, tem duas categorias e que há mais de uma variável independente.

De acordo com Kutner et al. (2005), geralmente é necessário mais de uma variável explicativa para obter uma adequada descrição do fenômeno em estudo e para a obtenção de previsões úteis. O modelo de regressão logística múltiplo é uma generalização do modelo logístico simples para incluir múltiplas variáveis explicativas. Realizando uma transformação logit, o modelo multiplo é dado por:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \tag{2.19}$$

No contexto da regressão logística múltipla, a probabilidade de sucesso π_i, ou seja, a probabilidade de que o evento de interesse ocorra ($Y = 1$), dado um conjunto de variáveis explicativas $x_1, x_2, \ldots, x_p$, é expressa pela seguinte fórmula:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}} \tag{2.20}$$

Analogamente, a probabilidade de fracasso $1-\pi_i$, ou seja, a probabilidade de que o evento não ocorra ($Y = 0$), é dada por:

$$1-\pi(x) = \frac{1}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}} \tag{2.21}$$

2.6.4 Estimação dos Parâmetros na Regressão Logística Binária

Supondo-se uma amostra de n observações independentes (x_i, y_i), onde y_i representa o valor da variável resposta dicotômica e x_i é o valor da variável independente da i-ésima observação em que $i = 1, 2, \ldots, n$. A estimação dos parâmetros da regressão logística binária pode ser feita com o método de máxima verossimilhança. O método de máxima verossimilhança consiste em encontrar o valor do parâmetro que maximiza a função de verossimilhança, apresentando um estimador razoável para o parâmetro avaliado (Souza, 2006).

A função de verossimilhança é dada por

$$L(\beta) = \prod_{i=1}^{n} [\pi(x_i)]^{y_i} [1-\pi(x_i)]^{1-y_i} \tag{2.22}$$

Onde $\pi(x_i)$ é o parâmetro da distribuição Bernoulli, representando a probabilidade de sucesso para o indivíduo i, e y_i é a variável aleatória que assume os valores 0 ou 1, condicionalmente às variáveis explicativas x_i.

Para encontrar os valores de β que maximizam $L(\beta)$, deriva-se a função em relação aos parâmetros de regressão, igualando o resultado a zero. A derivada de $L(\beta)$ em função de β, também conhecida como equação de verossimilhança, é dada pela expressão abaixo:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} \left[y_i - \pi(x_i) \right] = 0, \quad \text{para } j = 0, 1, 2, \ldots, p. \tag{2.23}$$

De modo a considerar as características do plano amostral é necessário incorporar os pesos amostrais no processo de estimação. Para tal, pode ser usado o método de Pseudo-Máxima Verossimilhança, sendo a contribuição de cada observação dada pela expressão:

$$\left[\pi(x_{ji})\right]^{w_{ji} \cdot y_{ji}} \left[1 - \pi(x_{ji})\right]^{w_{ji} \cdot (1 - y_{ji})} \tag{2.24}$$

A função de pseudo-máxima verossimilhança, construída como o produto da contribuição individual para a verossimilhança ao longo dos m conglomerados contendo n_m observações dentro de um dado conglomerado, é dada pela equação em (2.25).

$$L(\beta_0, \beta_1) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \left[\pi(x_{ji})\right]^{w_{ji} \cdot y_{ji}} \left[1 - \pi(x_{ji})\right]^{w_{ji} \cdot (1 - y_{ji})} \tag{2.25}$$

Onde, $\pi(x_{ji})$ representa a probabilidade de que o indivíduo i, da unidade amostral j, apresente o evento de interesse, y_{ji} indica se esse evento foi observado, w_{ji} corresponde ao peso amostral do indivíduo, m representa o número de conglomerados e n_j representa o número de indivíduos observados dentro do conglomerado j.

Assim como no método de máxima verossimilhança, a estimação pelo método de pseudo-máxima verossimilhança consiste em encontrar valores que maximizam o valor da função de pseudo-máxima verossimilhança (Archer e Lemeshow, 2006).)

2.6.5 Multicolinearidade

Segundo Maia (2017), o modelo logístico pressupõe que as variáveis explicativas sejam linearmente independentes, isto é, não se deve verificar multicolinearidade. A multicolinearidade ocorre quando as observações amostrais das variáveis independentes, ou combinações lineares delas são altamente correlacionadas.

A utilização de variáveis independentes altamente correlacionadas pode levar a diversos problemas, tais como, aumento dos erros-padrão, menor eficiência dos estimadores, estimativas menos precisas, instabilidade dos coeficientes perante pequenas variações nos dados e dificuldade em isolar os efeitos individuais de cada variável explicativa (Oliveira, 2014).

No entanto, a intensidade da multicolinearidade é avaliada principalmente através de indicadores estatísticos, como a correlação entre variáveis independentes, factor de inflação da variância (VIF – Variance Inflation Factor) e a tolerância (Maroco, 2007).

Factor inflacionário da variância (VIF) e Tolerância (TOL)

Segundo Gujarati (2011), o factor inflacionário da variância (VIF) representa o quanto a variância do estimador $\hat{\beta}_j$ está a ser inflacionada devido à presença de multicolinearidade entre a variável explicativa X_j e as demais variáveis independentes do modelo. Em situações nas quais não existe correlação entre as variáveis explicativas (ou seja, $R_j^2 = 0$), o valor de VIF_j será igual a 1. No entanto, à medida que a correlação entre X_j e as demais variáveis aumenta e se aproxima de uma relação linear perfeita ($R_j^2 = 1$), o valor de VIF_j tende ao infinito.

O VIF é calculado a partir do coeficiente de determinação R_j^2, obtido ao ajustar uma regressão da variável explicativa X_j em função das restantes variáveis independentes. Matematicamente, tem-se:

$$R_j^2 = 1 - \frac{SSE_j}{SST_j} \tag{2.26}$$

Onde SSE_j é a soma dos quadrados dos resíduos e SST_j a soma total dos quadrados da regressão da variável X_j sobre as demais. O valor do VIF é então dado por:

$$VIF_j = \frac{1}{(1 - R_j^2)} \tag{2.27}$$

Onde, R_j^2 corresponde ao coeficiente de determinação entre X_j e as restantes variáveis independentes.

A tolerância é uma medida estatística utilizada na análise de regressão múltipla com o objecivo de avaliar a presença de multicolinearidade entre as variáveis independentes do modelo. Segundo Vaz (2020), a tolerância indica a proporção da variabilidade de uma determinada variável explicativa que não pode ser explicada linearmente pelas demais variáveis independentes incluídas no modelo. Matematicamente, a tolerância de uma variável X_j é dada como:

$$TOL_j = 1 - R_j^2 \tag{2.28}$$

A tolerância varia entre 0 e 1. Valores próximos de 1 indicam baixa correlação entre a variável em questão e as restantes variáveis explicativas, ou seja, ausência de multicolinearidade. Por outro lado, quanto mais próximo de zero, maior será a multicolinearidade.

2.6.6 Razão de Chances (Odds)

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 o aumento ou redução da probabilidade do evento de interesse (Mangas, 2019).

De acordo com Mesquita (2014), na regressão logística, probabilidade e chance odds, são conceitos diferentes para descrever o mesmo fenómeno. Por definição, probabilidade é o risco de determinado evento acontecer, chance é a razão entre a probabilidade de determinado evento acontecer e a probabilidade desse evento não acontecer.

Segundo Vaz (2020), a razão de chances (odds ratio) é a razão entre as chances de um evento acontecer devido a influência de determinado factor, e a chance do evento acontecer sem a interferência deste factor.

Segundo Montenegro (2009), a razão de chance é denotada por OR, e é definida por:

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} \tag{2.29}$$

O logaritmo da *Odds ratio*, "log odds", é expressa da seguinte forma:

$$\ln(OR) = \ln\left\{\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}\right\} \tag{2.30}$$

assim, o logaritmo da razão das chances é dado por:

$$\ln(OR) = \ln[e^{\beta_1}] = \beta_1 \tag{2.31}$$

Deste modo, β_1 pode ser interpretado como sendo o logaritmo da razão de chances em condições acima referenciadas. Consequentemente, a razão das chances do sucesso versus fracasso quando a variável independente x aumenta uma unidade versus manter-se constante é a exponencial do respetivo coeficiente de regressão do logit.

- $OR > 1$ indica que um aumento em x aumenta a probabilidade do evento ocorrer.
- $OR < 1$ indica que um aumento em x reduz a probabilidade do evento ocorrer.
- $OR = 1$ indica que x não tem efeito sobre Y.

2.7 Análse de Dados Amostrais Complexos

Diversos métodos estatísticos são utilizados para analisar dados provenientes de pesquisas amostrais. Tais métodos frequentemente pressupõem que os dados são provenientes de uma amostra aleatória simples, que é uma técnica de amostragem probabilística em que se assume que as observações são independentes e cada elemento na amostra tem a mesma probabilidade de ser seleccionado (Casela e Berger, 2002).

Existem situações em que diferentes técnicas de amostragem podem ser consideradas quando, por exemplo, são levadas a cabo pesquisas para populações de grande dimensão e o mesmo ou maior grau de precisão obtido através da recolha de dados por Amostragem Aleatória Simples (AAS) pode ser obtido, com menor custo, dividindo a população em grupos chamados estratos. Portanto em diferentes situações podem ser aplicadas técnicas de amostragem como, por exemplo, amostragem estratificada, amostragem por clusters e amostragem em estágios múltiplos e que à estas, geralmente, estão associadas diferentes probabilidades de selecção para cada elemento da amostra (Lumley, 2017).

De acordo com Pessoa e Silva (1998), quando os dados a serem analisados foram obtidos a partir de planos amostrais complexos, considerados complexos por incluírem aspectos como estratificação, conglomeração e pesos amostrais distintos, o pressuposto de independência entre as observações é normalmente violado. Portanto, para a estimação dos parâmetros devem-se adoptar métodos que não levem em consideração tal pressuposto e que considerem as características do plano amostral.

Para estimar os valores dos parâmetros β considerando as características do plano amostral, o método de estimação mais adotado é o de Máxima Pseudo-Verossimilhança, que tem como base a incorporação dos pesos amostrais ω de cada indivíduo i, que são definidos como o inverso das probabilidades de inclusão π (Pessoa e Silva, 1998).

As estimativas pontuais são influenciadas por pesos distintos, associados as unidades da amostra, enquanto que as estimativas da variância são influenciadas pela estratificação, pela conglomeração e pelos pesos amostrais. Os aspectos do plano amostral, quando ignorados podem influenciar as medidas de razões de chance e as medidas de precisão das estimativas dos parâmetros dos modelos estimados (Cunha, 2010).

Capítulo 3

Material e Métodos

3.1 Material

O presente trabalho foi realizado com recurso à uma base dados secundária, referente aos dados do Inquérito Demográfico e de Saúde IDS 2022 -2023, realizado em Moçambique. A base de dados foi obtida no site da Demographic Health Survey (DHS), https://dhsprogram.com, mediante um pedido de autorização efectuado no âmbito da realização do presente trabalho.

A recolha de dados do IDS de 2022 -2023 compreendeu um desenho de amostragem complexo, isto é, os dados foram recolhidos através de uma amostragem probabilística, estratificada e multi-etápica. Na primeira etapa, foram selecionadas um total de 619 áreas de enumeração com probabilidade proporcional à dimensão, sendo a medida de tamanho, o número de agregados familiares em cada estrato explícito. Das 619 AEs, 232 eram de áreas urbanas e 387 de áreas rurais. Na segunda etapa, foram seleccionados 26 agregados familiares com probabilidades iguais de cada área de enumeração. Com base nesse procedimento, foram seleccionados 16 045 agregados familiares. Na etapa final, foram recolhidos dados demográficos e de saúde de todas as mulheres dos 15 aos 49 anos de idade e de todas as crianças menores de 5 anos encontradas nos agregados selecionados.

A variável **principal fonte de abastecimento de água** encontra-se dividida em fontes melhoradas e não melhoradas. Consideraram-se como fontes melhoradas, a água proveniente de torneiras públicas ou fontenárias, canalizada do vizinho, dentro de casa ou no quintal, poços protegidos, poços com bomba manual, água engarrafada ou água da chuva. Consideraram-se como fontes não melhoradas, a água proveniente de poços não protegidos, de tanques em camiões ou carregada em tambores e a água da superfície

Para a variável **tipo de instalação sanitária**, consideraram-se como melhoradas as casas de banho com latrinas com e sem autoclismo, latrinas melhoradas e latrinas melhoradas tradicionais. Por outro lado, classificaram-se como não melhoradas as casas de banho com latrinas não melhoradas, sem latrinas ou sem instalações sanitárias, ou seja, famílias que realizam as suas necessidades ao ar livre.

A variável **local de parto** encontra-se dividida em quatro categorias: domicílio (residência do respondente ou outro domicílio), setor público (hospital governamental, geral, provincial, rural, centro de saúde ou outra unidade do setor público), setor privado (hospital privado, clínica, médico privado ou outra unidade do setor privado) e outro (locais que não se enquadram nas categorias anteriores).

Tabela 3.1: Descrição das variáveis do estudo

Variáveis	Descrição	Categorias	Classificação
$\mathbf{X_1}$	Sexo da criança	0: Feminino; 1: Masculino	Qualitativa nominal
$\mathbf{X_2}$	Idade da criança	0: 6 - 23 meses; 1: 24 - 59 meses	Qualitativa ordinal
$\mathbf{X_3}$	Peso ao nascer	0: $<$ 2500g; 1: $\geq$ 2500g	Qualitativa ordinal
$\mathbf{X_4}$	Área de residência	0: Urbano; 1: Rural	Qualitativa nominal
$\mathbf{X_5}$	Província residência	0: Niassa; 1: Cabo Delgado; 2: Nampula; 3: Zambézia; 4: Tete; 5: Manica; 6: Sofala; 6: Sofala; 7: Inhambane; 8: Gaza; 9: Maputo Província; 10: Maputo Cidade	Qualitativa nominal
$\mathbf{X_6}$	Quintil de riqueza	0: Quintil mais baixo; 1: Segundo; 2: Terceiro ; 3: Quarto; 4: Quintil mais alto	Qualitativa ordinal
$\mathbf{X_7}$	Idade da mãe	0: $<$ 20 anos; 1: $[20-34]$ anos; 2: $\geq$ 35 anos	Qualitativa ordinal
$\mathbf{X_8}$	Estado civil da mãe	0: Solteiro; 1: Casado/União; 2: Divorciado/Viúvo	Qualitativa nominal
$\mathbf{X_9}$	Nível de escolaridade da mãe	0: Sem escolaridade; 1: Primário; 2: Secundário; 3: Superior	Qualitativa ordinal
$\mathbf{X}_{10}$	Estado nutricional da mãe	0: Abaixo do peso; 1: Normal; 2: Acima do peso	Qualitativa ordinal
$\mathbf{X}_{11}$	Diagnóstico de anemia da mãe	0: Não anémica; 1: Anémica	Qualitativa nominal
$\mathbf{X}_{12}$	Número de crianças no agregado	0: $<$ 3 crianças; 1: $\geq$ 3 crianças	Qualitativa ordinal
$\mathbf{X}_{13}$	Intervalo entre nascimentos	0: $<$ 24 meses; 1: $24-47$ meses; 2: $\geq$ 48 meses	Qualitativa ordinal
$\mathbf{X}_{14}$	Frequência consultas pré-natal	0: $<$ 4 consultas; 1: $\geq$ 4 consultas	Qualitativa ordinal
$\mathbf{X}_{15}$	Duração da amamentação	0: $<$ 6 meses; 1: $6-23$ meses; 2: $\geq$ 24 meses; 3: Não atualmente; 4: Nunca amamentou	Qualitativa ordinal
$\mathbf{X}_{16}$	Ocupação da mãe	0: Não trabalha; 1: Trabalha	Qualitativa nominal
$\mathbf{X}_{17}$	Local de parto	0: Domicílio; 1: Sector público; 2: Sector privado; 3: Outro	Qualitativa nominal
$\mathbf{X}_{18}$	Principal fonte de abastecimento de água	0: Fonte não melhorada; 1: Fonte melhorada	Qualitativa nominal
$\mathbf{X}_{19}$	Tipo de instalação sanitária	0: Não melhorada; 1: Melhorada	Qualitativa nominal
$\mathbf{X}_{20}$	Incidência de diarreia	0: Não; 1: Últimas 24h; 2: Últimas 2 semanas	Qualitativa nominal
$\mathbf{X}_{21}$	Incidência de febres	0: Não; 1: Sim	Qualitativa nominal
$\mathbf{X}_{22}$	Acesso a rádio	0: Não; 1: Sim	Qualitativa nominal
$\mathbf{X}_{23}$	Acesso a TV	0: Não; 1: Sim	Qualitativa nominal
$\mathbf{X}_{24}$	Suplementação de ferro na gravidez	0: Não ; 1: Sim; 2: Não sabe	Qualitativa nominal

A variável **estado nutricional da mãe** foi criada a partir do seu Índice de Massa Corporal (IMC), sendo que se o IMC for menor a 18,5 a mãe é considerada abaixo do peso; normal, se o IMC estiver entre 18,5 e 25; e acima do peso se o IMC for maior que 25.

3.2 Métodos

3.2.1 Classificação do estudo

- **Quanto à natureza do estudo** - Quanto à natureza, este estudo enquadra-se na pesquisa aplicada, pois o estudo é aplicado para a resolução de problemas específicos. Segundo Nascimento (2016), a pesquisa aplicada é dedicada à geração de conhecimento para a solução de problemas específicos, sendo dirigida à busca da verdade para uma determinada aplicação prática em situação particular.

- **Quanto aos procedimentos técnicos** - é caracterizada como um estudo bibliográfico, fundamentado em materiais como livros, artigos científicos e relatórios, para explorar os conceitos teóricos relacionados ao tema estudado. Segundo Gil Nascimento (2016), a pesquisa bibliográfica é um trabalho de natureza exploratória, que propicia bases teóricas ao pesquisador para auxiliar na exercício reflexivo e crítico sobre o tema em estudo

- **Quanto à abordagem ou método** - De acordo com Nascimento (2016), quanto à abordagem ou método, a pesquisa obedece duas modalidades distintas: abordagem quantitativa e qualitativa. Para este estudo foi adoptada a abordagem quantitativa, fundamentada pela aplicação de métodos estatísticos na colecta e análise de dados.

- **Quanto aos objectivos** - Em relação aos objectivos, o presente estudo é classificado como explicativo, pois, de acordo Fundo de Desenvolvimento de Educação (2009), a pesquisa explicativa é aquela que tem como objectivo básico a identificação dos factores que determinam ou que contribuem para ocorrência de um fenómeno.

Criação da Variável Resposta

Para a estimação do modelo de regressão logística binária, a variável dependente foi definida como o *diagnóstico de anemia na criança*, sendo uma variável dicotómica (anémica e não anémica), de acordo com os critérios estabelecidos pela OMS. A categorização foi realizada com base nos níveis de hemoglobina (Hb), conforme descrito a seguir:

$$Y = \begin{cases} 0, & \text{se a criança não apresenta anemia (Hb} \geq 11 \text{ g/dL);} \\ 1, & \text{se a criança apresenta anemia (Hb} < 11 \text{ g/dL),} \\ & \text{independentemente do nível de severidade.} \end{cases}$$

Dessa forma, todas as formas de anemia (ligeira, moderada e severa) foram agrupadas na categoria "anémico". Essa categorização permite a aplicação da regressão logística binária para avaliar os factores associados à prevalência de anemia nad crianças menores de 5 anos de idade.

3.2.2 Associação entre as Variáveis do Estudo

A estimação dos modelos de regressão logística foi precedida por uma análise bivariada, em que efectuou-se o cruzamento entre as demais variáveis explicativas e a variável de interesse.

- **Teste de independência do Qui-Quadrado**

Geralmente é usado o teste *Qui-Quadrado* por forma a analisar a existência de relação entre duas variáveis categóricas. Segundo Ferreira (2013), a partir de uma distribuição de frequências, o teste qui-quadrado estabelece o grau de correspondência entre as frequências observadas (O_i) e esperadas (E_i) em cada categoria, e portanto indica se existe uma diferença significativa entre os valores observados e esperados.

O valor do qui-quadrado pode ser calculado pela Equação (3.1), e é representado por χ^2.

$$\chi^2 = \sum_{i}^{r}\sum_{j}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)} \tag{3.1}$$

Onde:

- O_{ij} - número de casos observados na linha "i" da coluna "j";
- E_{ij} - número de casos esperados, na linha "i" da coluna "j";
- r - número de linhas;
- c - número de colunas.

Sendo que a hipótese nula do teste Qui-Quadrado consiste em afirmar que não existe relação entre as variáveis, ou seja, são independentes e a hipótese alternativa consiste em afirmar que as variáveis são correlacionadas, ou seja, são dependentes e a regra de decisão consiste em:

- Se $\chi^2 \leq \chi^2_{(r-1)(c-1),\alpha}$ ou *P-value* $\geq \alpha$, não rejeita-se H_0;
- Se $\chi^2 > \chi^2_{(r-1)(c-1),\alpha}$ ou *P-value* $< \alpha$, rejeita-se H_0.

De acordo com Heeringa *et al.* (2010), Rao e Scott (1984) propuseram um ajuste do teste Qui-Quadrado, de modo a incorporar as características do plano amostral, que consiste em dividir o valor da sua estatística por um factor generalizado do efeito do plano amostral (FGEPA), $\hat{\chi}^2_{R-S} = \dfrac{\chi^2}{FGEPA}$, que representa o efeito do plano amostral.

Todas as hipóteses foram testadas a um nível de significância de 5% e, para efeitos de avaliação da regra de decisão, usou-se o p-valor associado à estatística do teste.

O *FGEPA* é tido como a média dos autovalores da matriz $D = V_{\text{design}} V_{\text{SRS}}^{-1}$, em que V_{design} é uma matriz de variância-covariância baseada no plano amostral para o vector de proporções usadas na construção da estatística Qui-Quadrado e V_{SRS} é a matriz de variância-covariância para as proporções estimadas dada uma amostra aleatória simples.

Foi proposta ainda uma correção de segunda ordem que consiste em dividir a estatística de Rao e Scott pela quantidade $(1 + \hat{a}^2)$ de modo a incorporar a variabilidade nos autovalores da matriz D, em que $\hat{a}$ representa o coeficiente de variação dos autovalores. Thomas e Rao (1987), propuseram uma versão F-ajustada da correção de segunda ordem, que melhor controla o erro do tipo I, dada pela expressão em (3.2) (Heeringa *et al.*, 2010).

$$F_{T-R} = \frac{\hat{\chi}^2_{R-S}}{(R-1)(C-1)} \tag{3.2}$$

3.2.3 Testes de Significância dos Parâmetros do Modelo

Após a estimação dos parâmetros do modelo, procede-se à avaliação da qualidade dessa estimação por meio da verificação da significância estatística das variáveis explicativas. Esse procedimento tem como objectivo identificar aquelas variáveis que exercem influência relevante na explicação do comportamento da variável resposta.

- **Teste de Razão de Verossimilhança**

Uma vez ajustado o modelo, é necessário testar a significância do modelo. Isto pode ser feito atravéz do teste da razão de verossimilhança. Segundo Cabral (2013), a razão de verossimilhança avalia a significância dos coeficientes estimados simultaneamente, ou seja, verifica se o modelo estimado é globalmente significativo. Com este teste pretende-se testar simultaneamente se os coeficientes de regressão associados a β são todos nulos com excepção de β_0.

$$D = -2 \ln \left[\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}} \right] \tag{3.3}$$

$$D = -2 \sum_{i=1}^{n} \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \tag{3.4}$$

Onde, y_i é o valor observado da variável dependente para o indivíduo i e $\hat{\pi}_i$ é a estimativa da probabilidade de ocorrência do evento para o indivíduo i, obtida a partir do modelo ajustado.

No presente trabalho foi usado o teste F de Thomas e Rao, sendo a sua regra de decisão análoga à regra de decisão do teste *Qui-Quadrado*.

O modelo é dito saturado se contém todas as variáveis, enquanto o modelo ajustado corresponde ao modelo apenas com as variáveis desejadas para o estudo. Esta função D, também chamada de *deviance* (desvio), sempre é positiva e quanto menor, melhor é o ajuste do modelo (Beserra, 2021) .

Hipóteses a testar:

- $H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$
- $H_a : \exists \beta_k \neq 0, \quad k = q, q+1, \ldots, p-1$

A hipótese nula H_0 afirma que todos os coeficientes são iguais a zero, ou seja, as variáveis explicativas correspondentes não contribuem significativamente para o modelo e a hipótese alternativa H_a afirma que existe pelo menos um coeficiente diferente de zero, indicando que ao menos uma das variáveis explicativas tem efeito significativo sobre a variável dependente.

- Se *P-value* $\geq \alpha$, não se rejeita H_0;
- Se *P-value* $< \alpha$, rejeita-se H_0.

Para estimar a significância de uma variável independente, comparam-se o valor de D com e sem variável independente na equação. A alteração no valor de D esperada pela inclusão da variável independente no modelo é obtida através de:

$$G = D\left(\frac{\text{modelo sem a variável}}{\text{modelo com a variável}}\right) \tag{3.5}$$

De acordo com Lumley (2018), o teste de razão de verossimilhança para amostras complexas é baseado nas diferenças entre funções Deviance resultantes da ponderação das funções de pseudo-máxima verossimilhança.

- **Teste de Wald**

De acordo com Moura (2018), o teste de Wald é utilizado na regressão logística para a determinação da significância dos coeficientes do modelo estimado, ele testa se cada coeficiente é significativamente diferente de zero. Deste modo, o teste de Wald verifica se uma determinada variável independente possui uma relação estatisticamente significativa com a variável dependente.

Se os coeficientes logísticos forem estatisticamente significativos, seu impacto pode ser interpretado em termos da probabilidade estimada e, consequentemente, na predição do objecto de estudo no grupo respectivo, isto é, no grupo do evento de interesse ($Y = 1$) ou no grupo da não ocorrência do evento ($Y = 0$) (Gonzalez, 2018).

A estatística do teste de Wald é dada por:

$$W_j = \frac{\hat{\beta}_j}{\text{EP}(\hat{\beta}_j)} \tag{3.6}$$

Onde $EP(\beta_j)$ é o erro padrão de β_j. O teste de hipóteses referente a cada parâmetro é:

- $H_0 : \beta_j = 0$
- $H_1 : \beta_j \neq 0$

A hipótese nula afirma que o parâmetro β_j é igual a zero, isto é, a variável associada a β_j não tem um efeito estatisticamente significativo no modelo. A hipótese alternativa afirma que o parâmetro β_j é diferente zero, ou seja, a variável tem um efeito estatisticamente significativo no modelo.

- Se o p-value $< \alpha$, rejeita-se H_0.
- Se o p-value $\geq \alpha$, não se rejeita H_0;

- **Teste de Wald Ajustado**

De modo a incorporar a estrutura do plano amostral na avaliação da significância dos parâmetros do modelo, utiliza-se o teste de Wald ajustado.

O teste de Wald ajustado, dado pela equação:

$$F = \frac{s - p + 1}{sp} W \tag{3.7}$$

Onde $s = \left(\sum_{k=1}^{k} m_k\right) - K$ é o número de conglomerados selecionados menos o número de estratos e p é o número de variáveis explicativas e W é a versão multivariada do teste de Wald. A estatística F segue a distribuição F com p com $(s - p + 1)$ graus de liberdade (Cassy *et al.*, 2016).

- Se *p-value* $< \alpha$, rejeita-se H_0: o parâmetro é estatisticamente significativo;
- Se *p-value* $\geq \alpha$, não se rejeita H_0.

3.2.4 Qualidade de Ajuste na Regressão Logística Binária

Estimado o modelo de regressão, é importante verificar a qualidade de ajuste do modelo antes de tirar qualquer conclusão (Alvarenga, 2015). Para isso, existem várias medidas que podem ser utilizadas para quantificar a qualidade de ajuste do modelo logístico dentre elas se destaca o teste baseado na estatística de Hosmer e Lemeshow, o qual será utilizado no presente trabalho.

- **Teste de Hosmer e Lemeshow**

Conforme Hosmer & Lemeshow (2000), a estatística de Hosmer e Lemeshow corresponde a um teste *Qui-Quadrado* que consiste em dividir o numero de observações em aproximadamente dez classes e, em seguida comparar as frequências preditas com as observadas. A finalidade deste teste é verificar se existem diferenças significativas entre as classificações realizadas pelo método e a realidade observada. Antes do cálculo da estatística teste, é necessário estimar a probabilidade de sucesso para cada observação e ordenar as probabilidades preditas em ordem crescente, e em seguida agrupar os dados de acordo com os decis de probabilidades preditas. Em cada decil, dividir os valores observados e os valores preditos para o sucesso e fracasso.

A estatística do teste de Hosmer & Lemeshow é dada pela equação:

$$\hat{C} = \sum_{k=1}^{g} e_k \frac{(o_k - e_k)^2}{e_k \left(1 - \frac{e_k}{n_k}\right)} \tag{3.8}$$

Onde:

- $O_k = \sum_{j=1}^{n_k} y_{kj}$ representa o número de casos registados no k-ésimo decil.
- $e_k = \sum_{j=1}^{n_k} \hat{\pi}_{kj}$ representa o número esperado de casos no k-ésimo decil.

Portanto, testa-se as seguintes hipóteses:

- H_0: Existem diferenças significativas entre os resultados previstos e observados, isto é, o modelo estimado se ajusta aos dados.
- H_1: Não existem diferenças significativas entre os resultados previstos e observados, isto é, o modelo estimado não se ajusta aos dados.

A regra de decisão consiste em:

- Se $\hat{C} \leq \chi^2_{1-\alpha;g-2}$ ou $P(\hat{C} > \chi^2_{1-\alpha;g-2}) \geq \alpha$, não rejeitar H_0.
- Se $\hat{C} > \chi^2_{1-\alpha;g-2}$ ou $P(\hat{C} > \chi^2_{1-\alpha;g-2}) < \alpha$, rejeitar H_0.

- **Teste de Archer e Lemeshow**

De modo a acomodar os aspectos do plano amostral ao verificar o ajuste do modelo, Archer e Lemeshow propuseram uma modificação do teste de Hosmer e Lemeshow através de um teste baseado na média dos resíduos do modelo estimado sob o método de pseudo-máxima verossimilhança considerando as características do plano amostral (Massango, 2018) .

A estatística do teste é dada pela equação:

$$F_{A-L} = \frac{f - g + 2}{(fg)\hat{M}^t \hat{V}(\hat{M})^{-1}\hat{M}} \tag{3.9}$$

Onde f é o número de conglomerados menos o número de estratos, g é o número de agrupamentos dos resíduos médios, geralmente divididos em decis de risco, $\hat{M} = (\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_{10})$, e $\hat{V}(\hat{M})$ é a matriz de variância-covariância (Archer e Lemeshow, 2006).

3.2.5 Poder de classificação (ou discriminação) do modelo

Segundo Mangas (2019), um modelo de regressão pode ser estatisticamente significativo e não representar a realidade em estudo. Uma maneira prática de qualificar o ajuste do modelo de regressão logística é pela projecção do modelo na tabela de classificação.

Para isto, precisa-se criar uma tabela com o resultado da classificação cruzada da variável resposta, de acordo com uma variável dicotómica em que os valores se derivam das probabilidades logísticas estimadas na regressão (Hosmer e Lemeschow 2000).

A análise do poder de discriminação do modelo através da tabela de classificação é feita por meio de algumas medidas de desempenho como a sensibilidade e a especificidade. De acordo com Mangas (2019), a sensibilidade do modelo é a percentagem de classificações correctas no grupo "1-sucesso" da variável dependente (isto é, o modelo prevê correctamente para um dado sujeito a característica que se quer modelar). A especificidade é a percentagem de classificações correctas no grupo "0-insucesso" (isto é, o modelo prevê correctamente para um dado sujeito que não tem a característica que se quer modelar).

Tabela 3.2: Tabela de classificação

	Valores observados	
Valores preditos	**Grupo (1)**	**Grupo (0)**
Grupo (1)	VP	FP
Grupo (0)	FN	VN

Onde:

- Verdadeiro Positivo (VP)
- Falso Positivo (FP)
- Verdadeiro Negativo (VN)
- Falso Negativo (FN)

Assim essas medidas serão definidas como:

- **Capacidade de acerto do Grupo (1) - Sucesso:**

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

- **Capacidade de acerto Grupo (0) - Fracasso:**

$$\text{Especificidade} = \frac{VN}{VN + FP}$$

- **Capacidade de acerto total:**

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}$$

Segundo Gonçalves (2013) num modelo perfeito todos os casos estariam sobre a diagonal principal. Contudo, na prática é muito difícil obter um modelo perfeito e, deve-se classificar a sua capacidade preditora. Esta é considerada boa se a sensibilidade e a especificidade forem superiores a 80%, razoável se estes dois valores estiverem entre 50% e 80% e medíocre se ambos forem inferiores a 50%.

3.2.6 Curva ROC

Segundo Mangas (2019), o problema da utilização da sensibilidade e da especificidade é que estas medidas dependem do valor de corte, cuja escolha é arbitrária, como referido anteriormente, implicando aumentar-se a sensibilidade em detrimento da especificidade, ou vice-versa. O modo que permite resolver este problema, que não depende do valor de corte, consiste em utilizar-se uma análise ROC (Receiver Operating Characteristic).

De acordo com Quintas (2020), a curva ROC é uma ferramenta estatística que permite avaliar o desempenho de um sistema de classificação. Estas curvas são representadas num gráfico bidimensional, com a ordenada a representar a fracção de verdadeiro positivos ou sensibilidade e a abcissa a representar a fracção de falsos positivos ou 1-especificidade. O índice que avalia a medida de exatidão destes gráficos é a área abaixo da curva (AUC) e quanto maior for a área maior é o desempenho do sistema em causa

Segundo Hosmer e Lemeshow (2000), a curva ROC é desenvolvida num plano unitário de abscissa $1 - \beta$ e ordenada α, sobre o qual são traçados os diferentes pontos de corte das probabilidades relativas à sensibilidade e à especificidade. Sobre o plano unitário é traçado também um eixo de bissetriz ($y = x$), quanto mais próxima a curva estiver da bissetriz, menor é o poder discriminatório do modelo.

Na tabela (3.3) são apresentados os valores indicativos da área de discriminação com critério de aplicação geral para a descrição do poder discriminante ROC para regressão logística:

Tabela 3.3: Valores indicativos da área sobre a curva de ROC

Área ROC	**Poder discriminante do modelo**
$[0.5; 0.7[$	Discriminação fraca
$[0.7; 0.8[$	Discriminação aceitável
$[0.8; 0.9[$	Discriminação boa
≥ 0.9	Discriminação excepcional

Por forma a incluir as características do plano amostral no processo de análise, a análise de dados foi realizada no ambiente Complex Samples do software SPSS e através do pacote survey do software R .

Capítulo 4

Resultados e Discussão

4.1 Resultados

4.1.1 Caracterização da amostra em estudo

A amostra analisada no presente estudo é composta por 3.388 crianças menores de 5 anos, conforme consta na base de dados. Na figura(4.1), é apresentada a estimativa da prevalência de anemia nesta faixa etária, destacando a percentagem de crianças anémicas e não anémicas na amostra. Observa-se que a maioria das crianças da amostra são anémicas, correspondendo a 73.0%, enquanto 27.0% não apresentam anemia.

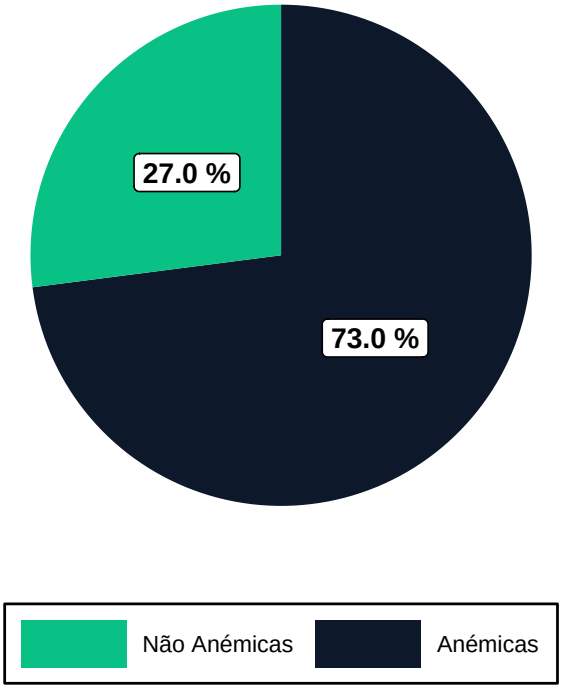

Figura 4.1: Prevalência de anemia em crianças menores de 5 anos em Moçambique

Nas figuras em (4.2), apresenta-se a distribuição das crianças segundo o Quintil de riqueza (a) e por Área de residência (b), em que observou-se que a maioria vive em domicílios pertencentes ao quintil de riqueza mais baixo (26.6%), 22.8% pertencem ao segundo quintil de riqueza, 19.7% ao quintil de riqueza médio, 18.5% ao quarto quintil de riqueza e apenas 12.4% pertencem ao quintil de riqueza mais elevado. Em relação a área de residência, maior parte das crianças residem em áreas rurais 72.2% , enquanto apenas 27.8% vivem em áreas urbanas.

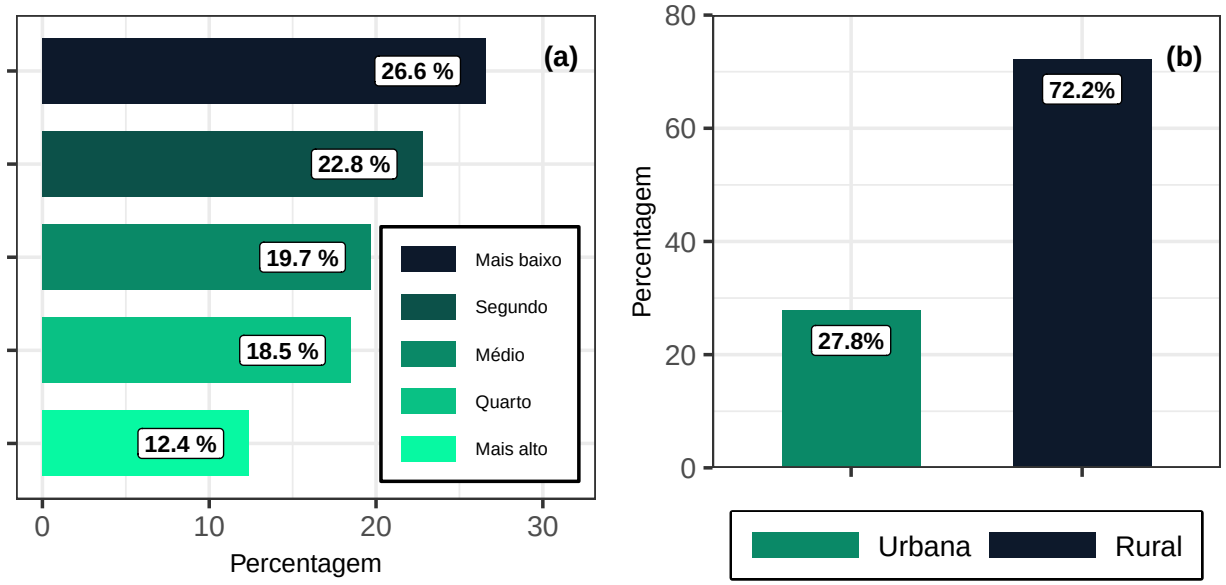

Figura 4.2: Distribuição percentual das crianças de acordo com o quintil riqueza (a) e área de residência (b)

Com base nas tabelas (4.1) e (5.1) (em anexo), verifica-se que a prevalência de anemia em crianças menores de 5 anos apresenta variações conforme diferentes características sociodemográficas, ambientais e relacionadas a saúde materno-infantil. Quanto ao sexo da criança, pode-se verificar que, dentre as 1715 crianças do sexo masculino, 1261 (73.5%) apresentam anemia, enquanto 454 (26.5%) não apresentam anemia. Em relação crianças do sexo feminino, composto por 1859 crianças, 1347 (72.5%) apresentam anemia e 512 (27.5%) não apresentam anemia. Em relação ao nível de severidade, verifica-se uma maior predominância da anemia moderada, com 719 (41.9%) no sexo masculino e 734 (39.5%) no sexo feminino.

Em relação à idade da criança, observa-se maior prevalência em crianças pertencentes ao grupo etário de 6 a 23 meses, dentre as 1217 crianças deste grupo, 1038 (85.3%) apresentam anemia, destas 602 (49.5%) apresentam anemia moderada e 7,2% anemia severa, enquanto 179 (14.7%) não apresentam anemia. Verifica-se ainda que, das 2358 crianças com idades entre 24 a 59 meses, 1570 (66.6%) apresentam anemia e 788 (33.4%) não apresentam anemia. Neste grupo, apenas 3,5% das crianças apresentam anemia severa.

Em relação à área de residência, a prevalência de anemia é maior nas áreas rurais. Das 2582 crianças residentes nas áreas rurais, 1967 (76.2%) apresentam anemia, enquanto 615 (23.8%) não apresentam anemia. Das 992 crianças residentes nas áreas urbanas, 641 (64.6%) apresentam anemia e 352 (35.4%) não apresentam anemia. Além disso, observa-se que a severidade da anemia também varia conforme a área de residência, sendo que a anemia severa é mais prevalente entre crianças residentes em área rurais (5,7%) em comparação com crianças residentes em área urbanas (1,0%).

Quanto a província de residência, verifica-se que, as províncias de Nampula (84.3%), Cabo Delgado (80.0%) e Sofala (77.9%) apresentam as maiores percentagens de crianças com anemia. Por outro lado, as províncias de Niassa (58.9%), Província de Maputo (55.3%) e Cidade de Maputo (41.1%) apresentam as menores percentagens de crianças diagnosticadas com anemia. No que diz respeito à severidade da anemia, a província de Nampula regista a maior percentagem de anemia severa (9.0%).

No que se refere à idade da mãe, verifica-se que, das 306 crianças cujas mães têm menos de 20 anos, 253 (82.8%) apresentam anemia e uma maior percentagem de anemia severa (5,7%). Quanto às mães com idades entre 20 e 34 anos, observa-se que, 1752 (71.7%) das 2442 crianças têm anemia. Entre as 827 crianças cujas mães têm 35 anos ou mais, 603 (72.9%) apresentam anemia, com menor percentagem de anemia severa (5,7%) e 224 (27.1%) não apresentam anemia.

Em relação ao nível de escolaridade da mãe, num total de 1145 mulheres sem escolaridade, 873 (76.2%) têm filhos com anemia. Dentre as 1717 mulheres com ensino primário, 1277 (74.4%) têm filhos com anemia, e 440 (25.6%) não presentam anemia. Em relação à gravidade da anemia, a anemia severa apresenta maior percentagem entre aquelas cujas mães não possuem escolaridade (5,3%)

Quanto ao quintil de riqueza, observa-se que a prevalência de anemia é maior entre crianças pertencentes ao quintil mais baixo onde 758 (79,6%) das 952 crianças apresentam anemia. Em relação ao quintil de riqueza médio, de um total de 705 crianças, 501 (71,2%) apresentam anemia. Verifica-se ainda uma redução na prevalência de anemia à medida que o quintil de riqueza, sendo que 258 (58,3%) das 442 crianças apresentam anemia no quintil de riqueza mais elevado.

Quanto ao grau de severidade, observou-se maior prevalência de anemia severa em crianças de famílias pertencentes ao quintil de riqueza mais baixo 77 (8.1%) e uma menor prevalência em crianças de famílias pertencentes ao quintil de riqueza mais elevado 3 (0.6%).

Em relação ao diagnostico de anemia da mãe, em um universo de 1877 crianças com mães anémicas, observa-se uma maior proporção de crianças diagnosticadas com anemia, correspondendo a 1476 (78.6%) e uma maior proporção de anemia severa 114 (6.0%), enquanto apenas 401 (21.4%) das crianças cujas mães eram anémicas não apresentam anemia. Por outro lado, entre as 1697 mães que não eram anémicas, 1132 (66.7%) das crianças apresentam anemia e com menor proporção de anemia severa 43 (2.5%), enquanto 565 (33.3%) não eram anémicas.

Tabela 4.1: Percentagem de anemia de acordo com as características socioeconómicas, demográficas e ambientais

Característica	Anémico				Não Anémico	Total
	Ligeira	Moderada	Severa	Qualquer		
Sexo da criança						
Masculino	462 (26.9%)	719 (41.9%)	80 (4.7%)	1261 (73.5%)	454 (26.5%)	1715
Feminino	536 (28.9%)	734 (39.5%)	76 (4.1%)	1346 (72.5%)	512 (27.5%)	1859
Idade da criança						
6 a 23 meses	348 (28.6%)	602 (49.5%)	87 (7.2%)	1037 (85.3%)	179 (14.7%)	1217
24 a 59 meses	650 (27.6%)	851 (36.1%)	69 (2.9%)	1570 (66.6%)	788 (33.4%)	2358
Área de residência						
Urbana	312 (31.4%)	319 (32.2%)	10 (1.0%)	641 (64.6%)	352 (35.4%)	992
Rural	687 (26.6%)	1134 (43.9%)	147 (5.7%)	1968 (76.2%)	615 (23.8%)	2582
Província de residência						
Niassa	92 (29.5%)	87 (28.0%)	4 (1.4%)	183 (58.9%)	128 (41.1%)	311
Cabo Delgado	59 (25.2%)	117 (49.8%)	12 (5.0%)	188 (80.0%)	47 (20.0%)	235
Nampula	224 (22.8%)	516 (52.5%)	88 (9.0%)	828 (84.3%)	155 (15.7%)	982
Zambézia	164 (26.7%)	270 (43.8%)	31 (5.0%)	465 (75.5%)	151 (24.5%)	616
Tete	120 (31.2%)	126 (32.9%)	6 (1.5%)	252 (65.7%)	132 (34.4%)	383
Manica	112 (40.0%)	77 (27.6%)	1 (0.4%)	190 (68.1%)	89 (32.0%)	279
Sofala	63 (24.9%)	121 (47.7%)	13 (5.0%)	197 (77.9%)	57 (22.4%)	253
Inhambane	38 (31.1%)	42 (34.4%)	1 (0.9%)	81 (66.4%)	41 (33.6%)	122
Gaza	47 (35.5%)	44 (33.4%)	1 (0.4%)	92 (69.7%)	41 (30.6%)	132
Maputo Província	59 (31.6%)	44 (23.7%)	0 (0.0%)	103 (55.3%)	84 (44.8%)	187
Maputo Cidade	21 (29.2%)	9 (11.9%)	0 (0.0%)	30 (41.1%)	43 (58.9%)	73
Idade da mãe						
$<$ 20 anos	88 (28.7%)	148 (48.4%)	17 (5.7%)	253 (82.8%)	53 (17.2%)	306
[20 anos - 34 anos]	673 (27.6%)	969 (39.7%)	110 (4.5%)	1752 (71.7%)	690 (28.3%)	2442
$\geq$ 35 anos	237 (28.7%)	337 (40.7%)	29 (3.5%)	603 (72.9%)	224 (27.1%)	827
Nível de escolaridade da mãe						
Sem Escolaridade	317 (27.7%)	495 (43.2%)	61 (5.3%)	873 (76.2%)	272 (23.8%)	1144
Primário	439 (25.6%)	746 (43.4%)	92 (5.4%)	1277 (74.4%)	440 (25.6%)	1717
Secundário	231 (34.4%)	207 (30.8%)	3 (0.5%)	441 (65.7%)	230 (34.3%)	672
Superior	12 (28.6%)	6 (14.5%)	0 (0.0%)	18 (42.9%)	24 (56.9%)	42
Quintil de riqueza						
Mais Baixo	214 (22.4%)	467 (49.0%)	77 (8.1%)	758 (79.6%)	194 (20.4%)	952
Segundo	201 (24.6%)	385 (47.2%)	48 (5.9%)	634 (77.8%)	182 (22.2%)	816
Médio	228 (32.4%)	254 (36.0%)	19 (2.8%)	501 (71.2%)	203 (28.8%)	705
Quarto	224 (33.9%)	224 (33.9%)	8 (1.3%)	456 (69.1%)	203 (30.9%)	659
Mais Elevado	132 (29.8%)	123 (27.9%)	3 (0.6%)	258 (58.3%)	184 (41.7%)	442
Diagnóstico de anemia da mãe						
Não Anémica	497 (29.3%)	592 (34.9%)	43 (2.5%)	1132 (66.7%)	565 (33.3%)	1697
Anémica	502 (26.7%)	861 (45.9%)	114 (6.0%)	1477 (78.6%)	401 (21.4%)	1877

4.1.2 Análise de Associação entre as Variáveis do Estudo

Para a análise de existência de relação entre a variável de interesse e a demais variáveis do estudo foi usado o teste F de Thomas e Rao, cujos resultados encontram-se na tabela (4.2). Consideraram-se como variáveis significativamente relacionadas aquelas cujo p-valor for menor do que o nível de significância (p-valor <0.05).

Encontram-se estatisticamente relacionadas a variável diagnostico de anemia da criança, a um nível de significância de 5% as seguintes variáveis: idade da criança, área de residência, província de residência, quintil de riqueza, idade da mãe, ocupação da mãe, nível de escolaridade da mãe, estado nutricional da mãe, diagnóstico de anemia da mãe, intervalo entre nascimentos, duração da amamentação, principal fonte de abastecimento de água, local de parto, acesso a rádio, e acesso a TV.

Tabela 4.2: Teste F de Thomas e Rao para análise de associação entre as variáveis

Variável	**Teste F**	**p-valor**
Sexo da criança	0.356	**0.547**
Idade da criança	83.387	0.000
Peso ao nascer	0.325	**0.564**
Área de residência	22.451	0.000
Província de residência	11.833	0.000
Quintil de riqueza	11.320	0.000
Idade da mãe	5.314	0.005
Estado civil da mãe	0.579	**0.559**
Nível de escolaridade da mãe	9.982	0.000
Estado nutricional da mãe	14.822	0.000
Diagnóstico de anemia da mãe	31.382	0.000
Número de crianças no agregado	0.120	**0.729**
Intervalo entre nascimentos	6.962	0.001
Frequência de consultas pré-natal	2.105	**0.147**
Duração da amamentação	9.923	0.000
Ocupação da mãe	26.406	0.000
Local de parto	38.280	0.001
Principal fonte de abastecimento de água	20.918	0.000
Tipo de instalação sanitária	3.260	**0.072**
Incidência de diarreia	1.070	**0.339**
Incidência de febres	1.365	**0.255**
Acesso a rádio	17.190	0.000
Acesso a TV	28.980	0.000
Suplementação de ferro na gravidez	0.740	**0.471**

4.2 Análise dos pressupostos do Modelo

4.2.1 Verificação da Multicolinearidade

Para verificar a presença de multicolinearidade nos dados, foram analisados os valores de VIF e Tolerância de cada variável em estudo. Conforme apresentado na Tabela (4.3), observa-se que os dados não exibem multicolinearidade, uma vez que todos os valores de VIF são inferiores a 10 e os valores de tolerância se aproximam de 1. Isso indica que não há duas ou mais variáveis explicativas que estejam simultaneamente relacionadas à variável resposta "Diagnóstico de anemia na criança".

Tabela 4.3: Teste de Multicolinearidade

Variáveis	**Tolerância**	**VIF**
Idade da criança	0.635	1.575
Área de residência	0.693	1.443
Província de residência	0.942	1.061
Quintil de riqueza	0.784	1.276
Idade da mãe	0.963	1.039
Nível de escolaridade da mãe	0.851	1.175
Estado nutricional da mãe	0.947	1.056
Diagnóstico de anemia da mãe	0.942	1.062
Intervalo entre nascimentos	0.958	1.043
Duração da amamentação	0.878	1.137
Ocupação da mãe	0.887	1.127
Local de parto	0.945	1.058
Principal fonte de abastecimento de água	0.846	1.183
Acesso a rádio	0.923	1.076
Acesso a TV	0.591	1.693

4.3 Modelo de Regressão Logística Binária

4.3.1 Estimação do modelo

Inicialmente, foi construído um modelo logístico com todas as variáveis que possuíam relação significativa com a variável de interesse (ver a tabela (4.2)). Após a estimação do modelo saturado, foi avaliada a significância global de cada variável, excluindo as que não são estatisticamente significativas a um nível de significância de 5%. Para tornar o processo de selecção mais flexível e robusto, utilizou-se o Método Backward, descrito na Secção 2.4.4 , permitindo a identificação das variáveis mais relevantes a serem mantidas no modelo.

A significância dos coeficientes das variáveis inclusas no modelo foi avaliada através do teste de razão de verossimilhança de Rao-Scott. Desta feita, o modelo final inclui as seguintes variáveis: Idade da criança, Província de residência, Quintil de riqueza, Intervalos entre nascimentos, Diagnóstico de anemia da mãe e Local de parto. Os resultados são apresentados na **tabela** (4.4)

Tabela 4.4: Teste de Razão de Verossimilhança de Rao-Scott para as variáveis no modelo.

Covariáveis	**P-valor**
Idade da criança	0.000
Província de residência	0.000
Quintil de riqueza	0.005
Intervalos entre nascimentos	0.027
Diagnóstico de anemia da mãe	0.000
Local de parto	0.000

4.3.2 Avaliação da Qualidade do Ajuste do Modelo

Para avaliar a qualidade de ajuste do modelo, foi aplicado o teste F ajustado proposto por Archer e Lemeshow. Pelo valor da estatística F-ajustada de Archer e Lemeshow $F_A - _L(9; 567) = 0.636$, associado a um p valor igual à 0.766 , comparado ao nível de significância de 5%, não existem evidências suficientes para a rejeição da hipótese nula de adequação do modelo, concluindo-se assim, que o modelo ajusta-se aos dados.

Tabela 4.5: Teste F-ajustado de Archer e Lemeshow para o modelo final

Passo	**Estatística F**	**Graus de Liberdade**	**P-valor**
1	0.636	9, 567	0.766

Posteriormente, a significância do modelo foi testada por meio do teste de Razão de Verossimilhança entre o modelo nulo e o modelo final. Rejeita-se a hipótese nula de que todos os parâmetros sejam estatisticamente iguais a zero ao nível de significância de 5%, com um valor de $p = 0.000$ no teste. Dessa forma, pode-se concluir que pelo menos uma variável independente é estatisticamente significativa para explicar a probabilidade de crianças menores de 5 anos apresentarem anemia.

4.3.3 Interpretação das Estimativas para a Razão de Chances

Os resultados referentes às razões de chances são apresentados na Tabela (4.6). Em relação à idade da criança, verifica-se que crianças com idade entre 24 e 59 meses apresentam menores chances de apresentar anemia ($\hat{\beta}$ = -0.866, OR = 0.420), indicando uma redução de 57,8% no risco em comparação com aquelas com menos de 24 meses, controlando o efeito dos demais factores.

Quanto à província de residência, ajustado pelos restantes factores, verifica-se que crianças residentes nas províncias de Cabo Delgado ($\hat{\beta}$ = 1.026, OR=2.789), Nampula ($\hat{\beta}$ = 1.311, OR = 3.711), Zambézia ($\hat{\beta}$ = 0.633, OR = 1.883), Manica ($\hat{\beta}$ = 0.536, OR = 1.710), Sofala ($\hat{\beta}$ = 0.927, OR = 2.527) e na província de Gaza ($\hat{\beta}$ = 0.881, OR = 2.412) apresentam um risco significativamente maior de apresentar anemia em comparação com aquelas residentes na Cidade de Maputo.

Em relação ao quintil de riqueza , observa-se que crianças pertencentes a famílias com quintil riqueza mais baixo têm 2 vezes mais chances de desenvolver anemia em comparação com aquelas do quintil mais elevado ($\hat{\beta}$ = 0.703 , OR = 2.019). De forma semelhante, crianças pertencentes ao segundo quintil de riqueza e médio também apresentam maior probabilidade de desenvolver anemia, com chances cerca de 1.8 vezes ($\hat{\beta}$ = 0.593, OR = 1.810) e 1.5 vezes ($\hat{\beta}$ = 0.408 , OR = 1.503) superiores, respetivamente, quando comparadas às do quintil mais elevado.

Quanto ao intervalo entre nascimentos, verifica-se que crianças nascidas após um intervalo entre nascimentos maior ou igual a 48 meses apresentam uma probabilidade significativamente menor de desenvolver anemia ($\hat{\beta} = -0{,}439$, OR = 0,644), com uma redução de aproximadamente 35,6% em comparação com aquelas nascidas com menos de 24 meses de intervalo.

Em relação ao diagnóstico de anemia da mãe, podemos verificar uma associação significativa, indicando que crianças cujas mães são anémicas apresentam aproximadamente 1.57 vezes mais chances de desenvolver anemia em comparação com aquelas cujas mães não são anémicas ($\hat{\beta}$ = 0.447, OR = 1.564), controlando o efeito dos demais factores.

Quanto ao local de parto, verifica-se que crianças nascidas no sector privado apresentam uma probabilidade significativamente menor ($\hat{\beta}$ = -1.441, OR = 0.237), com uma redução de aproximadamnete 76,7 no risco de desenvolver anemia em comparação com aquelas nascidas em casa . De forma semelhante, crianças nascidas em outro local também apresentam uma redução do risco, embora em menor magnitude ($\hat{\beta}$ = -0.514, OR = 0.598).

Tabela 4.6: Modelo de regressão logística binária

Variáveis	$\hat{\beta}$	SE	p-valor	OR [IC 95%]
Intercepto	0.921	0.329	0.005	2.511 [1.314 ; 4.796]
Idade da criança	—	—	—	—
De 6 a 23 meses (Ref)	—	—	—	—
De 24 a 59 meses	-0.866	0.139	**0.000**	0.420 [0.320 ; 0.552]
Província de residência	—	—	—	—
Niassa	0.004	0.281	0.989	1.004 [0.578 ; 1.742]
Cabo Delgado	1.026	0.287	**0.000**	2.789 [1.586 ; 4.905]
Nampula	1.311	0.292	**0.000**	3.711 [2.092 ; 6.583]
Zambézia	0.633	0.312	**0.043**	1.883 [1.021 ; 3.475]
Tete	0.411	0.270	0.128	1.509 [0.889 ; 2.563]
Manica	0.536	0.262	**0.041**	1.710 [1.022 ; 2.861]
Sofala	0.927	0.285	**0.001**	2.527 [1.445 ; 4.419]
Inhambane	0.558	0.283	0.060	1.747 [1.001 ; 3.048]
Gaza	0.881	0.269	**0.001**	2.412 [1.423 ; 4.091]
Maputo Província	0.532	0.285	0.062	1.702 [0.973 ; 2.977]
Maputo Cidade (Ref)	—	—	—	—
Quintil de riqueza	—	—	—	—
Mais elevado (Ref)	—	—	—	—
Quarto	0.275	0.166	0.097	1.317 [0.951 ; 1.823]
Médio	0.408	0.163	**0.013**	1.503 [1.092 ; 2.070]
Segundo	0.593	0.195	**0.002**	1.810 [1.234 ; 2.654]
Mais baixo	0.703	0.196	**0.000**	2.019 [1.374 ; 2.967]
Intervalo entre nascimentos	—	—	—	—
Menor que 24 meses (Ref)	—	—	—	—
Entre 24 a 47 meses	-0.294	0.172	0.088	0.746 [0.532 ; 1.045]
Maior ou igual a 48 meses	-0.439	0.165	**0.008**	0.644 [0.466 ; 0.891]
Diagnóstico de anemia da mãe	—	—	—	—
Não anémico (Ref)	—	—	—	—
Anémico	0.447	0.124	**0.000**	1.564 [1.227 ; 1.994]
Local de parto	—	—	—	—
Domicílio (Ref)	—	—	—	—
Sector público	-0.198	0.215	0.357	0.820 [0.537 ; 1.252]
Sector privado	-1.441	0.284	**0.000**	0.237 [0.135 ; 0.414]
Outro	-0.514	0.211	**0.015**	0.598 [0.396 ; 0.905]

4.3.4 Tabela de classificação do modelo final

Para avaliar a capacidade classificativa ou discriminatória do modelo logístico apresentado na Tabela (4.6), foram comparadas as taxas de acerto na classificação de crianças anémicas e não anémicas, bem como a proporção total de indivíduos correctamente classificados, com base na matriz de classificação.

Com base nos resultados apresentados na tabela (4.7), observa-se que no grupo de crianças anémicas, 2181 casos (92.2%) foram classificados correctamente, ou seja uma sensibilidade de 92.2% e no grupo de crianças não anémicos, 304 casos (29.8%) foram classificados correctamente pelo modelo logístico proposto. No geral, o modelo de regressão logística apresentou uma taxa de acerto de 73,3%, considerada aceitável.

Tabela 4.7: Tabela de classificação

	Valores observados		
Valores preditos	**Anémico (1)**	**Não Anémico (0)**	
Anémico (1)	2181	718	**Sensibilidade** = 92.2%
Não Anémico (0)	185	304	**Especificidade** = 29.8%
% Global			**Acurácia** = 73.3%

Esta análise e reforçada pelo gráfico da curva ROC apresentado na figura (4.3), onde verifica-se que o AUC (área sobre a curva de ROC) é de 0.728, demonstrando uma qualidade discriminatória aceitável do modelo.

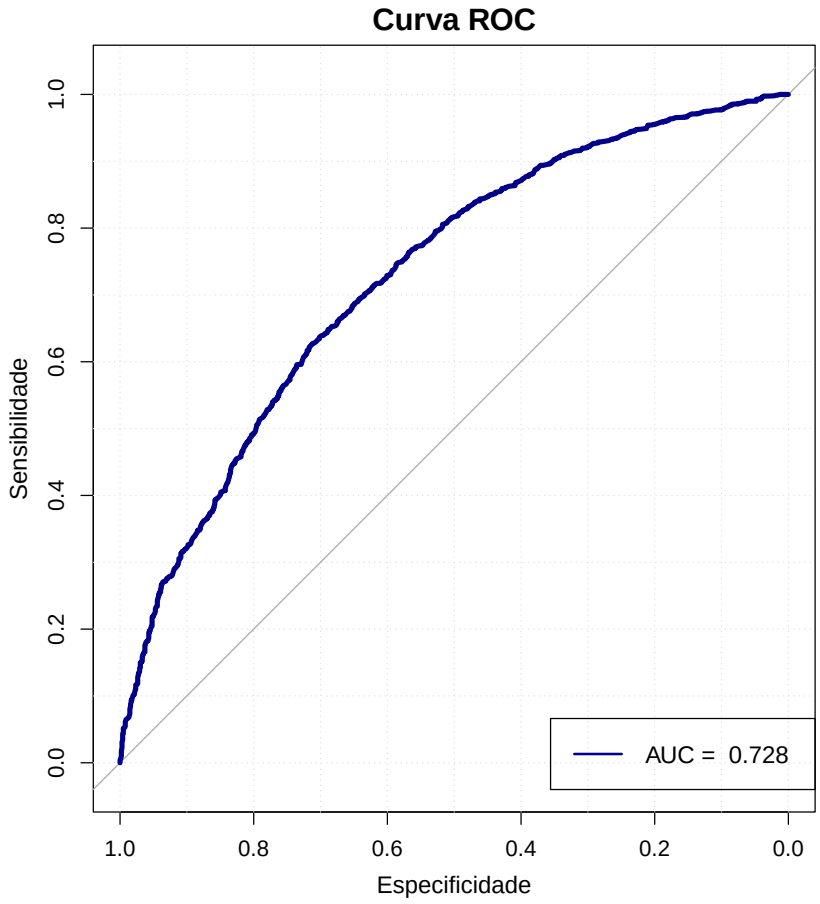

Figura 4.3: Curva ROC

4.4 Discussão

A prevalência de anemia encontrada no presente estudo, superior a 70%, revela um quadro preocupante de saúde pública entre crianças menores de 5 anos em Moçambique. Esse resultado está em concordância com o estudo realizado por Gebreegziabher e Sidibe (2023), que investigaram a prevalência e factores associados à anemia em crianças de 6 a 59 meses no Mali, reportando uma prevalência de 76% nessa faixa etária, o que evidencia a magnitude do problema em contextos semelhantes no continente africano.

Maior parte dos estudos apontam uma existência de associação entre o sexo da criança e e a ocorrência de anemia, indicando que crianças do sexo masculino apresentam maior risco de desenvolver anemia em comparação com crianças do sexo feminino. No entanto, alguns autores, como Ferreira *et al.* (2021) e Kejo *et al.* (2018), não observaram uma diferença estatisticamente significativa entre os sexos. De igual modo, no presente estudo, não foi encontrada associação significativa entre o sexo da criança e o ocorrência de anemia na criança. Esse achado está alinhado com alguns estudos que sugerem que o sexo pode não ser um factor determinante na ocorrência de anemia em crianças.

Observou-se que crianças menores de 24 meses apresentam um risco maior de desenvolver anemia em comparação com aquelas em idade mais avançadas. Este achado está de acordo com a literatura existente. Estudos de Osório *et al.* (2004) e Oliveira *et al.* (2007) destacam a idade da criança como um factor de risco relevante para a ocorrência de anemia. Esses estudos identificaram que a idade influencia significativamente a prevalência de anemia, sugerindo que factores como a transição da amamentação exclusiva para a alimentação complementar, a vulnerabilidade nutricional e maior susceptibilidade a infecções desempenham um papel crucial no risco de desenvolvimento da condição.

Crianças pertencentes a famílias com quintil de riqueza elevado apresentam um risco significativamente reduzido de desenvolver anemia. Esse achado está alinhado com a literatura existente, que aponta uma relação inversa entre condição socioeconômica e prevalência de anemia em crianças. Estudos de Agho *et al.* (2008) e Goswmai *et al.* (2015) destacam que o maior poder aquisitivo das famílias favorece o acesso a uma alimentação mais nutritiva e melhores cuidados de saúde, contribuindo assim, para a redução do risco de anemia entre as crianças.

Neste estudo, verificou-se que crianças cujas mães apresentavam anemia tendem a ter um risco maior de desenvolver anemia em comparação com aquelas cujas mães não eram anémicas. Este achado está de acordo com a literatura existente. Estudo de Agho *et al.* (2008), evidenciam a anemia materna como um factor de risco para a ocorrência de anemia em crianças.

Capítulo 5

Conclusões e Recomendações

5.1 Conclusões

Da analise da prevalência e dos factores associados a anemia em crianças menores de 5 anos em Moçambique, conclui-se que:

- A prevalência de anemia é maior entre crianças com idades entre 6 e 23 meses (85.3%), residentes em áreas rurais (76.2%) e entre crianças de mães sem escolaridade (76.2%). Além disso, observa-se maior prevalência entre crianças que vivem em domicílios com a principal fonte de abastecimento de água não melhorada (78.8%) e mães que não trabalham (76.2%).

- A idade da criança, área de residência, província de residência, quintil de riqueza, idade da mãe, ocupação da mãe, nível de escolaridade da mãe, estado nutricional da mãe, diagnóstico de anemia da mãe, intervalo entre nascimentos, duração da amamentação, principal fonte de abastecimento de água, local de parto, acesso a rádio, e acesso a TV são os factores que estão associados a anemia em crianças menores de 5 anos em Moçambique.

- Os principais factores de risco associados à alta prevalência de anemia em crianças menores de 5 anos em Moçambique são: idade da criança, província de residência, quintil de riqueza, intervalo entre os nascimentos, diagnóstico de anemia da mãe e local de parto.

- Crianças residentes nas províncias de Cabo Delgado, Nampula, Zambézia, Manica, Sofala e Gaza apresentam maiores chances de desenvolver anemia quando comparadas a crianças residentes na Cidade de Maputo.

- Crianças de famílias pertencentes ao quintil de riqueza mais baixo apresentam maior probabilidade de desenvolver anemia, com o quintil de riqueza mais baixo tendo aproximadamente o dobro de chances em comparação ao quintil mais elevado.

- O modelo de regressão logística estimado ajustou-se aos dados. O teste de qualidade do ajuste de Archer e Lemeshow $F_A -_L (9; 567) = 0.636$; p-valor $= 0, 766$) indicou que não há evidências suficientes para rejeitar a hipótese nula de boa adequação, sugerindo que o modelo apresenta um bom ajuste aos dados.

5.2 Recomendações

De acordo com as conclusões obtidas e observações feitas ao longo da realização do presente trabalho, recomenda-se:

- Desenvolver políticas específicas para famílias de baixa renda, garantindo melhor acesso a suplementos nutricionais, alimentação adequada e serviços de saúde
- Promover campanhas de sensibilização para cuidadores e profissionais de saúde sobre a importância da alimentação rica em ferro e outras medidas preventivas contra a anemia;
- Articular esforços entre os sectores da saúde, educação e assistência social para criar estratégias mais abrangentes e sustentáveis no combate à anemia em crianças menores de 5 anos.

5.3 Limitações

O presente trabalho apresenta algumas limitações, entre as quais a indisponibilidade de informações sobre factores relevantes, como o estado nutricional detalhado das crianças e as suas práticas alimentares, que poderiam influenciar a prevalência da anemia. Além disso, a qualidade dos dados foi limitada pela presença de valores omissos. Por fim, houve restrições quanto à disponibilidade de técnicas avançadas para avaliação da qualidade do ajuste dos modelos de regressão logística na análise de dados provenientes de amostras complexas.

Referências

[1] Alvarenga, A. (2015). Modelos lineares generalizados: aplicação a dados de acidentes rodoviários. Dissertação de Mestrado - Lisboa, Portugal.

[2] Amaral, S. M., Costa, S. de J., Pessoa, C. C. M., Pereira, P. L., Feitosa, Á. T. de O., Alves, Y. dos S., Júnior, J. L. P. (2021). Anemia ferropriva na infância: causas e consequências. Revista de Casos e Consultoria.

[3] Agresti, A. (2002). Categorical Data Analysis. 2nd ed. New York: Wiley.

[4] Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. John Wiley & Sons, Hoboken, New Jersey.

[5] Archer, K. J., e Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. The Stata Journal, 6(1), 97–105.

[6] Agho, K. E., K. J. Inder, S. J. Bowe, J. Jacobs, e M. J. Dibley. (2009). Prevalence and risk factors for stunting and severe stunting among under-fives in North Maluku province of Indonesia. BMC Pediatrics.

[7] Balarajan, Y., U. Ramakrishnan, E. Özaltin, A. H. Shankar, e S. V. Subramanian. (2011). Anaemia in Low-income and Middle-income Countries. The Lancet, 378(9809), 2123–2135.

[8] Belfiore, P. (2015). Estatística: aplicada a administração, contabilidade e economia com Excel e SPSS. 1. Ed. Rio de Janeiro.

[9] Benoist, B., McLean, E., Egli, I., & Cogswell, M. (2009). Worldwide prevalence of anaemia 1993–2005: WHO Global Database on Anaemia. World Health Organization.

[10] Beserra, R. S. (2021). Modelagem com Regressão Logística para Análise de Concessão de Crédito. Trabalho de Conclusão de Curso — Universidade Federal de Campina Grande, Campina Grande - PB.

[11] Brito, L. L., Barreto, M. L., Silva, R. de C. R., Assis, A. M. O., Reis, M. G., Parraga, I., e Blanton, R. E. (2003). Fatores de risco para anemia por deficiência de ferro em crianças e adolescentes parasitados por helmintos intestinais. Rev Panam Salud Publica

[12] Cabral, C. I. S. (2013). Aplicação do modelo de regressão logística num estudo de mercado. Dissertação (Mestrado) — Universidade de Lisboa, Lisboa.

[13] Casella, G., e Berger, R. (2002). Statistical inference. 2nd ed. Pacific Grove, CA: Wadsworth.

[14] Cassy, S. R., Natário, I., e Martins, M. R. (2016). Logistic regression modelling for complex survey data with an application for bed net use in Mozambique. Open Journal of Statistics.

[15] Castro, I. R. R., Normando, P., Farias, D. R., Berti, T. L., Schincaglia, R. M., Andrade, P. G., Bertoni, N., Lacerda, E. M. A. (2023). Fatores associados à anemia e à deficiência de vitamina A em crianças brasileiras menores de 5 anos. Cadernos de Saúde Pública

[16] Cunha, I. C. (2010). Análise de dados amostrais complexos da pesquisa do PROALFA de Minas Gerais. Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora, Juiz de Fora.

[17] Dias, B. C. C. (2014). Seleção de variáveis via Backward em Modelo Linear Normal Assimétrico. Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, Departamento de Estatística.

[18] Fávero, L. P., Belfiore, P., Da Silva, F. L., e Chan, B. L. (2009). Análise de dados. Campus.

[19] Fatoretto, M. B. (2016). Modelos Para Dados Categorizados com Efeito Aleatório: Uma Aplicação à Análise Sensorial. Dissertação de Mestrado - Universidade de São Paulo, Piracicaba.

[20] Ferreira, M. C. C. S. (2013). Modelos de regressão: uma aplicação em Medicina Dentária. Dissertação de Mestrado - Universidade Aberta, Lisboa.

[21] Ferreira, C. M. X., Azevedo, A. B., Santos, S. L. S., Kassar, S. B., Cardoso, M. A., & Ferreira, H. D. S. (2021). Evolução da prevalência de anemia em crianças, segundo dois inquéritos de base populacional em Alagoas, Brasil.

[22] Figueira, C. V. (2006). Modelos de Regressão Logística. Dissertação de Mestrado — Universidade Federal do Rio Grande do Sul, Porto Alegre.

[23] Fundo de Desenvolvimento de Educação. (2009). Métodos de pesquisa. Primeira edição. Rio Grande do Sul, Brasil.

[24] Gebreegziabher, T. e Sidibe, S. (2023). Prevalence and contributing factors of anaemia among children aged 6–24 months and 25–59 months in Mali. Journal of Nutritional Science.

[25] Gonçalves, A. M. (2013). Regressão Logística Aplicada à pesquisa de Preditores de Morte. Coimbra, Portugal.

[26] Gonzalez, L. A. (2018). Regressão Logística e suas Aplicações. Monografia de Graduação — Universidade Federal do Maranhão, São Luís.

[27] Goswmai, S., e K. K. Das. (2015). Socio-economic and demographic determinants of childhood anemia. Revista Brasileira de Hematologia e Hemoterapia.

[28] Gujarati, D. N e Porter, D. C. (2011). Econometria Básica. 5 Edição. New York. EUA

[29] Herringa, S. G., West, B. T., & Berglund, P. A. (2010). Applied Survey Data Analysis. Chapman & Hall/CRC.

[30] Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons.

[31] Instituto Nacional de Estatística (INE) e ICF. (2024). Inquérito Demográfico e de Saúde em Moçambique 2022–23. Maputo, Moçambique e Rockville, Maryland, EUA: INE e ICF.

[32] Kejo, D., T. C. E. Mosha, P. Petrucka, H. Martin, e M. E. Kimanya. (2018). Prevalence and predictors of undernutrition among underfive children in Arusha District, Tanzania. Food Sci Nutr.

[33] Kutner (2005). The Statistical Analysis of Failure Time Data. John Wiley Sons.

[34] Leite, M. S., A. M. Cardoso, C. E. Coimbra Jr., J. R. Welch, S. A. Gugelmin, P. C. Lira, B. L. Horta, R. V. Santos, e A. L. Escobar. (2013). Prevalence of anemia and associated factors among indigenous children in Brazil: results from the First National Survey of Indigenous People's Health and Nutrition.

[35] Lucena, S. L. L. (2018). Comparação de Modelo Classico e Bayesiano Para Dados de Óbitos Perinatais no ISEA. Trabalho de Conclusão do Curso - Universidade Estadual da Paraíba, Campina Grande.

[36] Lumley, T. (2017). survey: analysis of complex survey samples. R package version 3.32.

[37] Lumley, T. (2018). Analysis of complex survey samples: Package survey. UTC.

[38] Magalhães, E. I. S., Maia, D. S., Netto, M. P., Lamounier, J. A., & Rocha, D. S. (2018). Análise hierarquizada dos fatores associados à anemia em lactentes. Revista Paulista de Pediatria.

[39] Maia, A. G. (2017). Econometria: conceitos e aplicações. Capítulo 13. Brasil.

[40] Mangas, E. J. (2019). Introdução à regressão categorial: aplicação a dados de escolas angolanas, com utilização do SPSS. Brasil.

[41] Massango, E. T. (2018). Análise dos factores associados à desnutrição infantil em Moçambique. Maputo.

[42] Maroco, J. (2007). Análise Estatística com Utilização do SPSS. 5ª edição. Lisboa. Edições Sílabo.;

[43] Melo, M. A. (2020). Factores Associados a Anemia em Crianças dos 6 aos 24 meses. dissertação de Mestrado - Universidade Federal de Pernambuco.

[44] Mesquita, P. S. B. (2014). Um modelo de regressão logística para avaliação dos cursos de pós-graduação no Brasil. Dissertação de Mestrado — Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes.

[45] Ministério da Saúde (MISAU), Instituto Nacional de Estatística (INE) e ICF. (2015). Inquérito de Indicadores de Imunização, Malária e HIV/SIDA em Moçambique 2015. Maputo, Moçambique e Rockville, Maryland, EUA: INS, INE e ICF.

[46] Montenegro, S. G. (2009). Modelo de regressão logística ordinal em dados categóricos na área de ergonomia experimental. Dissertação de Mestrado — Universidade Federal da Paraíba, João Pessoa.

[47] Moraes, L. J. R., Andrade, L. da S., Farias, C. B. P., e Pinto, L. C. (2019). Prevalência de anemia associada a parasitoses intestinais no território brasileiro: uma revisão sistemática. Revista Pan-Amazônica de Saúde.

[48] Moura, G. L. (2018). Regressão logística aplicada à análise de risco de crédito. Brasil.

[49] Nascimento, F. P. (2016). Classificação da pesquisa e abordagens metodológicas. Brasil.

[50] Netto, M. P., S. E. Priore, H. M. P. Sant'Ana, M. D. C. G. Peluzio, C. M. Sabarense, D. G. da Silva, e S. D. C. C. Franceschini. (2006). Prevalência e fatores associados à anemia e deficiência de ferro em crianças de 18 a 24 meses. Departamento de Nutrição e Saúde - Universidade Federal de Viçosa, Universidade Federal de São Paulo. Viçosa, MG, Brasil.

[51] Oliveira, M. A. A., Osório, M. M., & Raposo, M. C. F. (2007). Socioeconomic and dietary risk factors for anemia in children aged 6 to 59 months. J Pediatric.

[52] Oliveira, T. M. de. (2014). Distância psíquica e seus efeitos sobre o fluxo de exportações dos estados brasileiros. Tese de Doutoramento - Faculdade de Economia da Universidade de Coimbra.

[53] Osório, M. M., P. I. Lira, M. Batista-Filho, e A. Ashworth. (2004). Prevalence of anemia in children 6–59 months old in the state of Pernambuco, Brasil.

[54] Paula, G. A. (2013). Modelos de Regressão com Apoio Computacional. São Paulo: Universidade de São Paulo.

[55] Pereira, R. C., Ferreira, L. O. C., Diniz, A. S., Batista Filho, M. (2007). Eficácia da suplementação de ferro associado ou não à vitamina A no controlo da anemia em escolares. Cadernos de Saúde Pública.

[56] Pessoa, D. G. C., e Silva, P. L. (1998). Análise de dados amostrais complexos. São Paulo: ABE.

[57] Quintas, J. P. S. (2020). Análise através da curva ROC: que ferramentas utilizar?. Dissertação de Mestrado - Universidade do Minho.

[58] Silva, D. G., S. E. Priore, e S. C. C. Franceschini. (2007). Risk factors for anemia in infants assisted by public health services: the importance of feeding practices and iron supplementation. J Pediatr, Rio Janeiro.

[59] Silva, A. J. d. (2014). Estudo Teórico Sobre Modelos Lineares Generalizados com Aplicação a Dados Genéticos. Trabalho de Conclusão do Curso - Universidade Estadual da Paraíba, Campina Grande.

[60] Silva, M. A., C. A. de Carvalho, P. C. de A. Fonsêca, S. A. Vieira, A. Q. Ribeiro, S. E. Priore, e S. D. C. C. Franceschini. (2015). Prevalência e fatores associados à anemia ferropriva e hipovitaminose A em crianças menores de um ano. Caderno de Saúde Coletiva, Rio de Janeiro.

[61] Silveira, M. B. G. da. (2021). Aplicação da regressão logística na análise dos fatores de risco associados à hipertensão arterial. Trabalho de Conclusão de Curso – Universidade Estadual da Paraíba, Campina Grande.

[62] Souza, É. C. de. (2006). Análise de influência local no modelo de regressão logística. Tese (Doutorado em Agronomia) — Escola Superior de Agricultura Luiz de Queiroz.

[63] Tadesse, S. E., Zerga, A. A., Mekonnen, T. C., Tadesse, A. W., Hussien, F. M., Feleke, Y. W., Anagaw, M. Y. (2022). Burden and determinants of anemia among under-five children in Africa: Systematic review and meta-analysis. Journal of Nutrition and Metabolism

[64] Turkman, M. A. A. & Silva, G. L. (2000). Modelos Lineares Generalizados: da teoria à prática. Lisboa: Universidade de Lisboa, 151 pp.

[65] UNICEF. (2014). Situação das Crianças em Moçambique.

[66] Vasconcelos, P. N. de. (2020). Fatores associados à anemia ferropriva em crianças de 6 a 24 meses do município de Viçosa - Alagoas. Tese de Doutorado - Universidade Federal de Pernambuco, Recife.

[67] Vaz, F. E. C. R. (2020). Modelação Linear e Extensões: Aplicação da Regressão Logística no Estudo de Câncer da Mama. Trabalho de Mestrado em Estatística, Universidade Aberta.

[68] World Health Organization. (2015). The global prevalence of anaemia in 2011. Geneva.

[69] World Health Organizatin. (2017). Nutritional anaemias: tools for effective prevention and control. Geneva.

[70] World Health Organization . (2021). Anaemia in women and children. In: Global Health Observatory. Geneva.

Apêndice

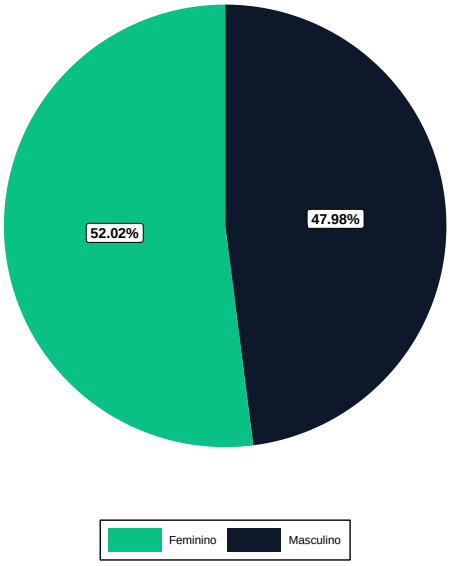

Figura 5.1: Distribuição percentual das crianças segundo sexo da criança

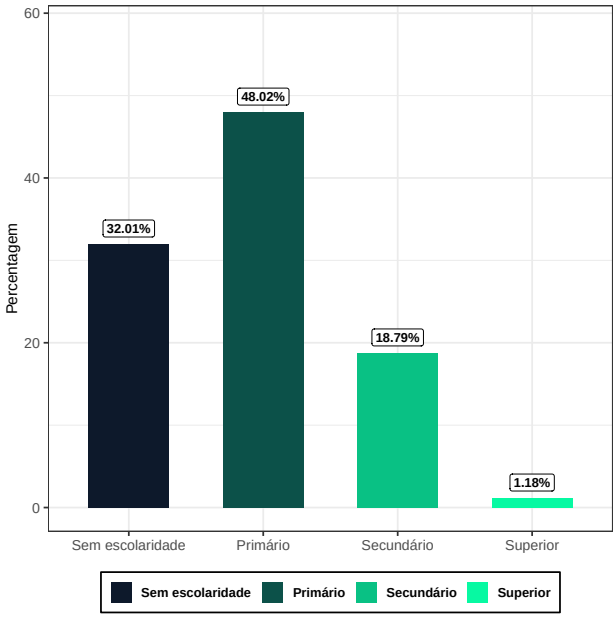

Figura 5.2: Distribuição percentual das crianças segundo o nível de escolaridade da mãe

Tabela 5.1: Percentagem de anemia de acordo com as características socioeconómicas, demográficas e ambientais (Continuação da Tabela (4.1))

Característica	Anémico				Não Anémico	Total
	Ligeira	Moderada	Severa	Qualquer		
Estado civil da mãe						
Solteira	54 (33.5%)	60 (37.1%)	2 (1.1%)	116 (71.6%)	46 (28.4%)	161
Casada/União	835 (28.1%)	1216 (40.9%)	131 (4.4%)	2182 (73.4%)	792 (26.6%)	2974
Divorciada/Viúva	109 (24.8%)	178 (40.5%)	24 (5.4%)	311 (70.8%)	129 (29.2%)	440
Nº de crianças						
<3 crianças	406 (29.8%)	532 (39.1%)	61 (4.5%)	999 (73.4%)	362 (26.6%)	1361
≥ 3 crianças	592 (26.8%)	922 (41.7%)	95 (4.3%)	1609 (72.7%)	604 (27.3%)	2214
Peso ao nascer						
<2500g	41 (28.0%)	72 (48.3%)	5 (3.3%)	118 (79.7%)	30 (20.3%)	148
≥ 2500g	359 (31.4%)	466 (40.8%)	55 (4.8%)	880 (77.1%)	262 (22.9%)	1142
Estado nutricional da mãe						
Abaixo do peso	52 (26.5%)	88 (44.3%)	15 (7.8%)	155 (78.6%)	42 (21.4%)	198
Normal	724 (27.5%)	1139 (43.3%)	121 (4.6%)	1984 (75.5%)	644 (24.5%)	2628
Acima do peso	211 (29.9%)	214 (30.3%)	20 (2.8%)	445 (63.0%)	262 (37.0%)	707
Intervalo entre nascimentos						
≥ 48 meses	350 (30.0%)	405 (34.7%)	43 (3.7%)	798 (68.4%)	369 (31.6%)	1167
24 a 47 meses	535 (27.5%)	844 (43.5%)	80 (4.1%)	1459 (75.1%)	485 (24.9%)	1943
<24 meses	113 (24.4%)	204 (44.0%)	33 (7.2%)	350 (75.6%)	113 (24.4%)	464
Ocupação da mãe						
Não trabalha	693 (28.1%)	1052 (42.7%)	132 (5.4%)	1877 (76.2%)	584 (23.7%)	2462
Trabalha	305 (27.4%)	401 (36.1%)	24 (2.1%)	730 (65.6%)	382 (34.4%)	1113
Principal fonte de abastecimento de água						
Fonte não melhorada	402 (25.7%)	735 (47.0%)	95 (6.1%)	1232 (78.8%)	331 (21.2%)	1563
Fonte melhorada	589 (29.7%)	705 (35.6%)	59 (3.0%)	1353 (68.3%)	630 (31.7%)	1983
Tipo de instalação sanitária						
Não melhorada	712 (27.0%)	1115 (42.3%)	124 (4.7%)	1951 (74.0%)	687 (26.0%)	2638
Melhorada	280 (30.8%)	325 (35.8%)	30 (3.3%)	635 (69.9%)	273 (30.1%)	908
Número de consultas pré-natal						
Menos de 4 consultas	230 (26.8%)	413 (48.1%)	71 (8.3%)	714 (83.1%)	145 (16.9%)	859
4 ou mais consultas	309 (30.6%)	450 (44.5%)	45 (4.5%)	804 (79.6%)	206 (20.4%)	1010
Incidencia de diarreia recentemente						
Não	904 (27.9%)	1304 (40.2%)	142 (4.4%)	2350 (72.5%)	891 (27.5%)	3241
Sim, últimas 2 semanas	90 (27.9%)	146 (45.1%)	14 (4.4%)	250 (77.4%)	73 (22.6%)	324
Incidencia febre nas últimas duas semanas						
Não	885 (28.2%)	1258 (40.1%)	129 (4.1%)	2350 (72.3%)	869 (27.7%)	3140
Sim	113 (26.7%)	189 (44.5%)	28 (6.5%)	220 (77.7%)	95 (22.3%)	424
Acesso a rádio no agregado						
Não	720 (28.0%)	1074 (41.8%)	134 (5.2%)	1944 (75.0%)	641 (25.0%)	2568
Sim	272 (27.8%)	366 (37.5%)	21 (2.1%)	659 (67.4%)	319 (32.6%)	978
Acesso a TV no agregado						
Não	725 (26.7%)	1178 (43.5%)	150 (5.5%)	2053 (75.8%)	656 (24.2%)	2709
Sim	267 (31.9%)	262 (31.3%)	4 (0.5%)	533 (63.7%)	304 (36.3%)	837
Local de parto						
Domicílio	166 (23.8%)	367 (52.7%)	68 (9.8%)	601 (86.2%)	96 (13.8%)	697
Setor público	414 (30.0%)	585 (42.4%)	62 (4.5%)	1061 (76.8%)	319 (23.1%)	1381
Setor privado	62 (26.6%)	32 (13.7%)	0 (0.0%)	94 (40.3%)	139 (59.7%)	233
Outro	355 (28.1%)	469 (37.1%)	26 (2.1%)	850 (67.3%)	412 (32.6%)	1263
Total	**27.9%**	**40.7%**	**4.4%**	**73.0%**	**27.0%**	—

Anexo

Jun 05, 2024

Edilson André
Universidade Eduardo Mondlane
Mozambique
Request Date: 06/05/2024

Dear Edilson André:

This is to confirm that you are approved to use the following Survey Datasets for your registered research paper titled: "Análise da Prevalência e dos Factores Associados à Anemia em Crianças menores de 5 Anos em Moçambique":

Mozambique

To access the datasets, please login at: https://www.dhsprogram.com/data/dataset_admin/login_main.cfm. The user name is the registered email address, and the password is the one selected during registration.

The IRB-approved procedures for DHS public-use datasets do not in any way allow respondents, households, or sample communities to be identified. There are no names of individuals or household addresses in the data files. The geographic identifiers only go down to the regional level (where regions are typically very large geographical areas encompassing several states/provinces). Each enumeration area (Primary Sampling Unit) has a PSU number in the data file, but the PSU numbers do not have any labels to indicate their names or locations. In surveys that collect GIS coordinates in the field, the coordinates are only for the enumeration area (EA) as a whole, and not for individual households, and the measured coordinates are randomly displaced within a large geographic area so that specific enumeration areas cannot be identified.

The DHS Data may be used only for the purpose of statistical reporting and analysis, and only for your registered research. To use the data for another purpose, a new research project must be registered. All DHS data should be treated as confidential, and no effort should be made to identify any household or individual respondent interviewed in the survey. Also, be aware that re-distribution of any DHS micro-level data, either directly or within any tool/dashboard, is not permitted. Please reference the complete terms of use at: https://dhsprogram.com/Data/terms-of-use.cfm.

The data must not be passed on to other researchers without the written consent of DHS. However, if you have coresearchers registered in your account for this research paper, you are authorized to share the data with them. All data users are required to submit an electronic copy (pdf) of any reports/publications resulting from using the DHS data files to: references@dhsprogram.com.

Sincerely,

Bridgette Wellington

Bridgette Wellington
Data Archivist
The Demographic and Health Surveys (DHS) Program

530 Gaither Road, Suite 500, Rockville, MD 20850 USA +1.301.407.6500 +1.301.407.6501 fax icf.com