



UNIVERSIDADE
E D U A R D O
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

Análise dos Factores Associados à Prevalência
da Desnutrição Crónica, em crianças menores de
5 anos, na Província de Nampula.

Autor: Inês Wiksson Neves Tembe

Maputo, Março de 2025



UNIVERSIDADE
E D U A R D O
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos Factores Associados à Prevalência
da Desnutrição Crónica, em crianças menores de
5 anos, na Província de Nampula.**

Autor: Inês Wiksson Neves Tembe

Supervisor: Mestre Mauro Henrique Langa(UP)

Maputo, Março 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 Licenciada em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Março de 2025

.....

(Inês Wiksson Neves Tembe)

Dedicatória

Dedico este trabalho aos meus pais, **Wiksson Neves Rodrigues Tembe e Isabel José Paulo Machava**, que sempre me incentivaram a me formar e a lutar pelos meus sonhos e nunca desistir mesmo diante das dificuldades.

Agradecimentos

Agradeço à **Deus** acima de tudo, pela vida e por permitir viver a graça de me formar.

Agradeço ao meu supervisor, **Mauro Henriques Langa**, pela disponibilidade, paciência, pelo apoio e pela orientação durante a realização do trabalho.

Aos docentes do DMI, em especial ao **Prof. Doutor Alberto Mulenga**, **Doutor Tiago Devesse** e **Doutor José Moreno** que para além da ciência, ensinaram lições valiosas de vida, o meu muito obrigada.

Aos meus colegas e amigos pela companhia e suporte durante toda a trajetória académica. Em especial ao **dr.Oscar Alexandre**, pelo grande suporte a quando da realização do presente trabalho e por acreditar no sucesso do mesmo, o meu muito obrigada.

A minha família que apoiou-me sempre e acreditou em mim, em particular a minha madrinha **Esperança Machava** e meu tio **Domingos Machava** que mais do que apoio material sempre me incentivaram a lutar pelo que acredito, a não desistir da academia em hipótese alguma, muito obrigada.

A todos que de alguma forma estiveram comigo durante a minha trajetória académica, dando suporte e apoio moral, o meu muito Obrigada!

Resumo

A Nutrição consiste na ingestão de alimentos de forma regular, regrada e saudável. A desnutrição crónica é uma manifestação física de desnutrição que retarda o crescimento e desenvolvimento, é um factor de risco que contribui para a mortalidade infantil e também um marcador de desigualdades no desenvolvimento humano. Este estudo tem como objectivo analisar os factores associados à prevalência da desnutrição crónica em crianças menores de 5 anos na província de Nampula. Foram usados dados do inquérito sobre a avaliação da Segurança alimentar e nutricional (2021-2022) realizado em Moçambique na província de Nampula, pelo Secretariado Técnico de Segurança Alimentar e Nutricional. A regressão logística múltipla foi a técnica aplicada na análise da relação entre a variável dependente e as variáveis independentes que influenciam a ocorrência da desnutrição crónica em Nampula. Constatou-se com o estudo que as variáveis: género da criança, idade da criança, altura da criança, ocorrência de febres, crianças menores de 5 anos no agregado, tamanho do agregado, idade do chefe do agregado e fonte de renda, mostraram-se significativas para explicar a prevalência da desnutrição crónica nas crianças.

Palavras-chave: Criança, Desnutrição crónica, Regressão logística.

Abstract

Nutrition consists of eating food regularly, regulated and healthy. Chronic Malnutrition is a physical manifestation of malnutrition that delays growth and development, is a risk factor that contributes to the infant mortality, is also a marker of inequalities in human development. This study aims to analyze the factors associated with the prevalence of chronic malnutrition in children under 5 years of age in Nampula province. Data from the survey on the assessment of food and nutrition security (2021-2022) carried out in Mozambique in the province of Nampula by the Technical Secretariat for Food and Nutrition Security were used. Multiple logistic regression was the technique applied in the analysis of the relationship between the dependent variable and the independent variables that influence the occurrence of chronic malnutrition in Nampula. The study found that the variables: Child's gender, child's age, child's height, occurrence of fevers, children under 5 years of age in the household, household size, age of head of household and source of income (earnings from fishing and aquaculture) were significant to explain the prevalence of chronic malnutrition in children.

Keywords: Children, Chronic malnutrition, Logistic regression.

Lista de Abreviaturas

AIC	Critério de informação de Akaike
AF	Agregado Familiar
AI	Altura para Idade
BIC	Critério de informação Bayesiano
BP	Baixo Peso
DA	Desnutrição Aguda
DC	Desnutrição Crónica
EP	Erro Padrão
FAO	Food and Agriculture Organization
IC	Intervalo de Confiança
IDS	Inquérito Demográfico e de Saúde
INE	Instituto Nacional de Estatística
IOF	Inquérito ao Orçamento Familiar
MISAU	Ministério da Saúde
ODM	Objectivos de Desenvolvimento do Milénio
OMS	Organização Mundial da Saúde
OR	Odds Ratio
PA	Peso para Altura
PI	Peso paara Idade
ROSC	Civil Society Forum for Child Rights
RC	Razão de chances
SADC	Comunidade de Desenvolvimento da África Austral
SAN	Segurança Alimentar e Nutricional

SETSAN	Secretariado Técnico de Segurança Alimentar e Nutricional
UNICEF	United Nations International Children's Emergency Fund
USA	United States of America
USAID	United States Agency International Development
WHO	World Health Organization
WAHO	West Africa Healt Organization

Índice

1	INTRODUÇÃO	1
1.1	Contextualização	1
1.2	Definição do Problema	2
1.3	Objectivos	3
1.3.1	Objectivo Geral	3
1.3.2	Objectivos Específicos	3
1.4	Relevância de Estudo	4
1.5	Estrutura do Trabalho	4
2	REVISÃO DE LITERATURA	5
2.1	Desnutrição Infantil	5
2.1.1	Tipos de Desnutrição Infantil	6
2.1.2	Causas da Desnutrição Infantil	6
2.1.3	Consequências da Desnutrição Crónica	8
2.1.4	Indicadores de Desnutrição Infantil	9
2.2	Desnutrição Crónica em África	10
2.3	Desnutrição Crónica em Moçambique	11
2.4	Desnutrição Crónica na Província de Nampula	13
2.5	Técnica Estatística	16
2.5.1	Modelos Lineares Generalizados	16
2.5.2	Regressão Logística	17
2.5.3	Regressão Logística Simples	19
2.5.4	Regressão Logística Múltipla	22
2.5.5	Transformação Logít	23
2.6	Métodos de Selecção das Variáveis	24
2.7	Critérios para a Selecção do Modelo	24
2.7.1	Critério de Informação de Akaike	24
2.7.2	Critério de Informação Bayesiano	25
3	Material e Métodos	27
3.1	Material	27

3.2	Métodos	28
3.2.1	Análise Descritiva	28
3.2.2	Testes de Significância dos Coeficientes do Modelo	28
3.2.3	Razão de Chances (Odds Ratio)	30
3.2.4	Diagnóstico e Avaliação do Ajuste do Modelo	31
4	Resultados e Discussão	36
4.1	Resultados da Análise Descritiva	36
4.1.1	Descrição das características das crianças da província de Nampula	36
4.2	Resultados da Regressão Logística	40
4.2.1	Discussão dos resultados	46
5	Conclusões e Recomendações	48
5.1	Conclusões	48
5.2	Recomendações	48
5.3	Limitações	49

Lista de Figuras

2.1	Mapa de Moçambique.	11
2.2	Mapa da Província de Nampula.	14
3.1	Representação gráfica de uma Curva ROC	33
4.1	Distribuição das crianças segundo o género.	36
4.2	Distribuição dos chefes dos agregados segundo a idade	37
4.3	Distribuição dos chefes dos agregados segundo o género	37
4.4	Distribuição dos agregados segundo o nível de acesso a informação	38
4.5	Disribuição dos agragados segundo a fonte de água	38
4.6	Curva ROC do modelo de regressão logística	43
5.1	Resíduos do modelo logístico final	55

Lista de Tabelas

2.1	Variações dos Modelos Lineares Generalizados	17
3.1	Descrição das variáveis em estudo	27
3.2	Tabela de classificação	32
4.1	Distribuição dos AF's segundo a fonte de renda	39
4.2	Estimativas dos coeficientes do modelo	40
4.3	Teste de Razão de Verossimilhança	41
4.4	Medidas de avaliação do ajuste geral do modelo	41
4.5	Teste de Hosmer e Lemeshow	41
4.6	Matriz de classificação do modelo	42
4.7	Critério de selecção do modelo	42
4.8	Área sob a curva (AUC) da curva ROC	42
4.9	Modelo de regressão logístico final	44
4.10	Estimação das probabilidades pelo modelo final	45
5.1	Parâmetros do segundo modelo de regressão logístico	55
5.2	Critério de selecção do modelo	55

Capítulo 1

INTRODUÇÃO

1.1 Contextualização

A Nutrição consiste na Ingestão de alimentos de forma regular, regrada e saudável. Segundo Fernandes e Valois (2017), a nutrição é uma área de conhecimento que estuda a relação do homem com o alimento, considerando o contexto sociocultural no qual se encontra.

Segundo o Fundo das Nações Unidas para a Infância (UNICEF, 2021), a alimentação de baixa qualidade é um dos maiores obstáculos para a sobrevivência, crescimento e desenvolvimento das crianças. No entanto, a alimentação das crianças tem falhado em relação a pontualidade, frequência e diversidade.

Para Monteiro (2003), desnutrição é uma condição clínica que decorre de uma alimentação deficiente em energia e nutrientes ou ainda do inadequado aproveitamento biológico dos alimentos ingeridos, geralmente motivado pela presença de doenças, em particular doenças infecciosas.

De acordo com a Organização Mundial da Saúde et al (OMS et al, 2021), as práticas de alimentação complementar também são motivo de preocupação, 30 % das crianças consomem 4 grupos de alimentos por dia e apenas 13 % das crianças de 6 a 23 meses tem uma dieta minimamente aceitável.

Segundo Silva (2012), a desnutrição infantil pode desencadear problemas que terão suas consequências ao longo de toda vida do ser humano. Uma vez que ela afecta todos os sistemas, impedindo o crescimento e desenvolvimento adequado das crianças e pode impedir que a criança tenha uma vida saudável.

De acordo com o Inquérito ao Orçamento Familiar (IOF, 2019), a desnutrição crónica define-se como baixa estatura para idade, diferindo da desnutrição aguda que refere-se ao baixo peso para altura. A DC é uma manifestação física de uma desnutrição a longo termo que retarda o crescimento e desenvolvimento. É um factor de risco que contribui para a mortalidade infantil, e é também um marcador de inequidades no desenvolvimento humano. Ela reflete a falha em alcançar uma altura óptima, em crianças menores de cinco anos de idade.

1.2 Definição do Problema

A Desnutrição em Moçambique é responsável por mortes infantis e para além disso, ela pode trazer danos irreversíveis a saúde de uma pessoa, podendo colocar as crianças em risco de contrair doenças crónicas tais como diabetes e obesidade. A desnutrição infantil também pode causar impacto prejudicial no aproveitamento escolar, na renda familiar e perpetua o ciclo de privação intergeracional.

De acordo com a UNICEF (2023), a desnutrição crónica em Moçambique manteve-se constante nos últimos 15 anos. Níveis extremamente elevados de desnutrição crónica (38%) em Moçambique afectam uma em cada 5 crianças menores de 5 anos. Estas taxas são mais elevadas que a média da região africana 29 % e entre as mais altas do mundo, situação apontada como crítica.

Os dados actuais apontam uma prevalência de 46,7 %, isto é, em cada 100 crianças 47 menores de idade sofrem da desnutrição crónica na cidade de Nampula, apesar destes dados representarem uma tendência de redução em relação a 2013, que era 51%, entende-se não fazer sentido que a província continue a ser vítima deste mal uma vez que é uma das maiores produtoras agrícolas do país, para além da produção animal. Há uma necessidade de compreender quais são os factores determinantes para redução (se são políticas sociais implementadas, políticas de incentivo agrícola ou melhorias de condições de acesso a água), aliado a isso pretende-se responder a seguinte questão de pesquisa:

Quais são os factores associados à prevalência da desnutrição crónica em crianças menores de 5 anos, na Província de Nampula?

1.3 Objectivos

1.3.1 Objectivo Geral

Analisar os factores associados à prevalência da desnutrição crónica em crianças menores de 5 anos, na Província de Nampula.

1.3.2 Objectivos Específicos

- Descrever as características das crianças da Província de Nampula;
- Identificar os factores associados à prevalência de desnutrição crónica em crianças menores de 5 anos, na Província de Nampula;
- Estimar a probabilidade de uma criança estar em situação de desnutrição crónica.

1.4 Relevância de Estudo

Considerando que as taxas de nutrição estão ainda longe dos níveis desejados, tendo em conta ainda o impacto da desnutrição infantil nas crianças, acredita-se ser de grande relevância o desenvolvimento desse estudo. Uma vez caracterizada e modelada a situação nutricional das crianças em Moçambique, espera-se com esse estudo o desenvolvimento de políticas na área da saúde, a promoção de hábitos de alimentação saudável para mães e crianças menores de 5 anos, assim como para a sociedade no geral.

Na área académica espera-se que com este trabalho efectuem-se investigações relacionadas à desnutrição com vista a despertar o interesse das entidades responsáveis.

1.5 Estrutura do Trabalho

O presente trabalho está dividido em 5 capítulos. Na introdução foi feita contextualização do problema em estudo, é apresentada a sua formulação, os objectivos, bem como a relevância do estudo. Na revisão de literatura foi feita uma abordagem referente a problemática da desnutrição infantil, bem como as técnicas de análise estatísticas usadas neste trabalho, os Modelos Lineares Generalizados, especificamente a regressão logística. No capítulo de material e métodos é apresentada a fonte de obtenção de dados, as variáveis do estudo e os métodos usados no desenvolvimento do trabalho. No capítulo dos resultados apresentou-se as características descritivas dos dados, os resultados do modelo logístico e algumas discussões acerca da sua aplicação. No capítulo 5 foram apresentadas as conclusões e recomendações à luz dos objectivos estabelecidos e resultados obtidos.

Capítulo 2

REVISÃO DE LITERATURA

2.1 Desnutrição Infantil

De acordo com a Organização das Nações Unidas para alimentação e Agricultura (FAO, 2018), os primeiros debates sobre alimentação surgem aos finais do século XIX e início do século XX quando a agricultura começa a perder peso na economia global e começavam a ser enfrentados problemas de fome no mundo. Assim em 1945 foi criada a organização das nações unidas para alimentação e agricultura (FAO).

De acordo com a FAO (2018), em 1992 convocou-se a primeira conferência sobre a nutrição, dedicada exclusivamente a abordar os problemas nutricionais ao nível mundial, organizada pela FAO e a OMS, nesta conferência os governos comprometeram-se com a eliminação fome crónica generalizada, a desnutrição em especial em crianças, mulheres e idosos. Era importante e urgente a necessidade de erradicar a fome e a desnutrição crónica. Em 1976, a FAO convocou uma cúpula mundial sobre alimentação, um grande marco na história pois foi a primeira vez em que os líderes mundiais dirigiam a sua atenção a segurança alimentar. Nesta cúpula foi adoptada a declaração de Roma sobre a segurança alimentar mundial e o plano de acção da cúpula mundial sobre a alimentação.

De acordo com Da Silva et al.(2017), a associação da fome prolongada com o retardo do crescimento em crianças foi percebida nos anos 1800, mas a descrição da desnutrição, como doença associada a deficiência de vitaminas, minerais e proteínas, ocorreu somente nos anos 30. A pobreza e os determinantes sociais foram vinculados como causas principais de desnutrição, tornando-a um problema de ordem social de grande impacto, levando a criança a não atingir todo seu potencial genético, tendo impacto na morbilidade e reflectindo, acima de tudo na redução da produtividade económica.

Em 1976, a FAO definiu a desnutrição como sendo o estado patológico, originado do consumo deficiente de alimentos e de uma ingestão calórica inferior aos requerimentos durante período prolongado; manifesta-se principalmente por alterações físicas, psicológicas e bioquímicas, alterações do peso e de estatura média das crianças. Geralmente o segmento populacional utilizado se restringe ao infantil, é usual que o diagnóstico da deficiência nutricional em crianças se baseie na identificação de déficit de peso ou de altura para a idade, não sendo avaliados e quantificados, por exemplo, a mortalidade ou o estado sub-nutricional de crianças decorrente de infecções que têm na má alimentação, a causa primária (FAO, 2018).

2.1.1 Tipos de Desnutrição Infantil

De acordo com Júnior (2014), existem 3 tipos de desnutrição infantil, nomeadamente: desnutrição crónica, desnutrição aguda e desnutrição aguda crónica.

Desnutrição crónica – também chamado de nanismo, é uma relação entre idade e a estatura física ou seja ocorre quando a criança apresenta crescimento deficitário que não corresponde a sua idade (baixa Estatura para idade).

Desnutrição aguda – é a relação entre o baixo peso para estatura, a desnutrição aguda leva a rápida perda de peso ou incapacidade de ganhar peso normalmente.

Desnutrição aguda crónica – quando a criança sofre de desnutrição aguda e crónica, ela fica abaixo do peso, além de experimentar numerosos problemas de saúde. A desnutrição aguda por sua vez pode ser classificada em:

- Kwashiorkor;
- Marasmo;
- Kwashiorkor - Marasmo.

2.1.2 Causas da Desnutrição Infantil

As práticas alimentares, desmame precoce, a higienização deficiente na preparação dos alimentos, falta de vitaminas e minerais na dieta, incidência de repetidas infecções ou doenças diarreicas e parasitoses, as fomes causadas por secas ou outros desastres naturais ou guerras, a desigualdade social, a falta de terra para cultivar e problemas de distribuição intra-familiar destacam-se como as causas mais comuns da desnutrição infantil.

Para Da Silva et al (2013), a nível mundial consideram-se vários factores associados à desnutrição infantil, não se limitando apenas em questões de práticas alimentares. Entre vários factores destacam-se, factores socio-económicos, biológicos, demográficos e ambientais e factores relativos a saúde materno - infantil.

Factores Sócio-económicos

De acordo com Barros et al (1993), os factores sócio - económicos relacionados a desnutrição infantil, incluem: renda familiar, nível de escolaridade dos pais, estado ocupacional da mãe, área de residência, exposição a meios de comunicação como, a rádio, a televisão, jornais ou revistas, entre outros.

Factores Biológicos

Os factores biológicos incluem a idade e o sexo. A OMS considera a idade e o sexo quando se pretende fazer medições relativas ao estado nutricional de uma criança, ou seja existe uma variação que depende da idade e sexo da criança para poder classificá-la como desnutrida ou não. Consideram-se também variáveis antropométricas como o peso, altura e comprimento da criança.

Factores Demográficos e ambientais

De acordo com Barros et al (1993), os factores ambientais que influenciam a desnutrição infantil, são condições de moradia, número de pessoas por dormitório, número de crianças menores de 5 anos no agregado, disponibilidade de água canalizada, tipo de sanitário, bairro, etc.

Segundo Da Silva et al (2013), os factores demográficos incluem, idade materna, número de moradores na casa e sexo do chefe do agregado familiar. As condições ambientais e económicas deficientes limitam a aquisição de bens alimentares essenciais, contribuindo para um deficiente estado de nutrição, particularmente em mulheres e crianças. Observa-se ainda um elevado índice do analfabetismo das mães nos países mais pobres.

Perlito (2014), acrescenta o deficiente saneamento básico, que contribui para a propagação de epidemias como a malária, cólera, principalmente no período das chuvas, comprometendo o estado de nutrição da criança e elevando a mortalidade neste grupo etário, aos factores demográficos.

Factores relatativos à Saúde Materno-infantil

Para Massango (2018), o estado nutricional das crianças é influenciado por factores relacionados à saúde materna e infantil, pois, o bem-estar das crianças para além de ser reflexo do seu próprio estado de saúde também reflexo do estado de saúde de suas progenitoras. Para além dos hábitos alimentares inadequados, a desnutrição crónica inicia antes mesmo da criança nascer e é influenciada pela nutrição materna inadequada, assim como por infecções frequentes, especialmente durante a gravidez e nos primeiros anos de vida.

Segundo Perlito (2014), o baixo peso da mãe no início da gestação, doenças, tabagismo e estresse durante a gestação, falta ou deficiência de assistência pré-natal, antecedentes reprodutivos desfavoráveis e a ocorrência de gravidezes múltiplas (intervalos intergenésicos muito curtos) podem afectar o peso ao nascer. O peso ao nascer tem uma forte influência no estado de saúde e sobrevivência das crianças, um adequado estado nutricional da mãe antes e durante a gravidez são de extrema importância para a saúde da criança diminuindo as chances de uma criança nascer ou entrar em estado de desnutrição.

2.1.3 Consequências da Desnutrição Crónica

Júnior (2014), aponta como consequências da desnutrição, as seguintes:

- Dificuldades em ganhar peso ou altura;
- Prejuízo no desenvolvimento neuro psicomotor;
- Déficits neurológicos e intelectuais irreversíveis;
- Vulnerabilidade à doenças ou infecções que podem culminar em morte.

O Fórum da Sociedade Civil para os Direitos da Criança (ROSC, 2014), sustenta os aspectos acima mencionados, sugerindo que as consequências podem ser numa dimensão individual onde as crianças apresentam um desenvolvimento cognitivo deficitário, dificuldades no seu desenvolvimento intelectual e físico o que pode dificultar o aprendizado e a educação das crianças, contribuindo ainda para a desistência escolar o que pode limitar o seu futuro e comprometer o desenvolvimento socioeconómico de vários países, sem contar com o nível de exposição as infecções e doenças.

A nível económico a desnutrição crónica é apontada como indicador chave da qualidade do capital humano. Além do alto custo para a nação, a elevada prevalência da desnutrição

crónica compromete o alcance dos compromissos nacionais e internacionais sobre desenvolvimento social e económico, assumidos pelo país e tem uma influência directa no alcance das metas dos Objectivos de Desenvolvimento do Milénio (ODM).

2.1.4 Indicadores de Desnutrição Infantil

De acordo com o Inquérito ao Orçamento Familiar (IOF, 2019) os indicadores nutricionais são classificados de acordo com estado nutricional das crianças. Consideram-se a Desnutrição Crónica (DC), a Desnutrição Aguda (DA), o Baixo peso (BP) e o Sobrepeso e Obesidade quando se fazem avaliações antropométricas para classificar o estado nutricional das crianças.

De acordo com Perlito (2014), 3 índices antropométricos são usados como indicadores de desnutrição, nomeadamente: Altura para Idade (AI), Peso para Altura (PA), Peso para Idade (PI).

- **Altura-para-Idade (A/I):** usado para estimar a **Desnutrição Crónica (DC)**, que reflete a falha em alcançar uma altura óptima devido à nutrição inadequada a longo prazo. Esse índice compara a altura observada de uma criança com a altura mediana para a idade, utilizando os padrões de crescimento da OMS. A DC é classificada em desnutrição crónica moderada e grave.
- **Peso-para-Altura (P/A):** utilizado para diagnosticar a **Desnutrição Aguda (DA)** em crianças menores de cinco anos de idade. O índice P/A é calculado comparando o peso observado de uma determinada criança com a mediana do peso de uma criança, em referência aos padrões de crescimento da população da OMS. A DA é soma da desnutrição Aguda Grave e Desnutrição Aguda Moderada.
- **Peso-para-Idade (P/I):** usado para identificar o **Baixo Peso (BP)** que é a manifestação física tanto da Desnutrição Crónica (DC) como da Desnutrição Aguda (DA), Este índice avalia crianças menores de cinco anos, comparando o peso de uma criança com o peso médio de crianças da população de referência para determinada idade, baseado nos padrões da OMS. O BP inclui tanto o baixo peso grave quanto o moderado.

O déficit em um desses é indicado como desnutrição embora este por si só não é apontado como um processo específico que leva a criança à desnutrição. São considerados outros factores como: o déficit estrutural, falta de peso e insuficiência nutricional.

2.2 Desnutrição Crónica em África

Segundo Nedel (2020), a África é um continente com cerca de 30 milhões de km^2 , uma população equivalente a 1 bilhão de habitantes, sendo o 3º mais populoso entre os continentes. Localizado nos hemisférios Norte, Sul, Leste e Oeste, sendo o único que se estende pelos 4 hemisférios, e fica entre o oceano indico, Atlântico, o mar mediterrâneo e o mar vermelho. África é formada por 54 Países e considerada a região mais economicamente pobre do mundo.

De acordo com a Comunidade de Desenvolvimento da África Austral (SADC, 2022), os níveis elevados e generalizados de pobreza na região continuam a ser reforçados por baixas taxas de crescimento económico, elevado nível de desemprego e subemprego, aumento da desigualdade social, aumento da frequência e intensidade de choques, fracos sistemas de protecção social e má prestação de serviços básicos, incluindo saúde, água, higiene e saneamento. Isto dá à insegurança alimentar e nutricional na região a sua natureza crónica.

Segundo Marcon (2017), as mudanças climáticas bem como as crises económicas e alimentar nos últimos anos, provocaram um agravamento da fome e desnutrição nos países da África, o que acabou se tornando uma situação crónica em muitos lugares. Apesar do grande potencial agrícola da África, muitas regiões ainda sofrem com a fome, estima-se que uma em cada quatro pessoas não tem acesso a alimentação e nutrição adequadas na África subsariana.

Para Barreto et al (2021), a desnutrição é particularmente grave na África, onde estima-se que quase 256.1 milhões de crianças, das quais 90% estavam na África subsariana, sofriam de desnutrição em 2018, sendo a prevalência de baixa estatura e de baixo peso em crianças até 5 anos de 30% e 7,1%, respectivamente.

Segundo Humbwaveli (2016), a prevalência da desnutrição infantil ainda é alta nos países da África subsariana. O tipo mais prevalente de desnutrição corresponde a baixa estatura, que merece destaque não só como indicador de desnutrição mas também de pobreza e condições de vida.

Segundo a West África Health Organization et al (WAHO, 2020), a desnutrição é uma questão de saúde pública que aumenta substancialmente o risco de mortalidade e morbidade. A saúde e nutrição das raparigas e mulheres em Idade reprodutiva (particularmente em idade reprodutiva) são críticas e tornam-se uma prioridade, uma vez que estão ligadas a saúde e sobrevivência das crianças.

2.3 Desnutrição Crónica em Moçambique

De acordo com Da Barca e Dos Santos (1992), Moçambique situa-se na faixa sul-oriental do continente africano. Sendo limitado ao norte por Tanzânia, ao oeste por Malawi, Zâmbia, Zimbabwe e Swazilândia, e ao sul por África de Sul. O país está dividido em 11 províncias: ao norte, Niassa, Cabo Delgado e Nampula; ao centro, Zambézia, Tete, Manica e Sofala; ao sul, Inhambane, Gaza, Maputo e Maputo Cidade.

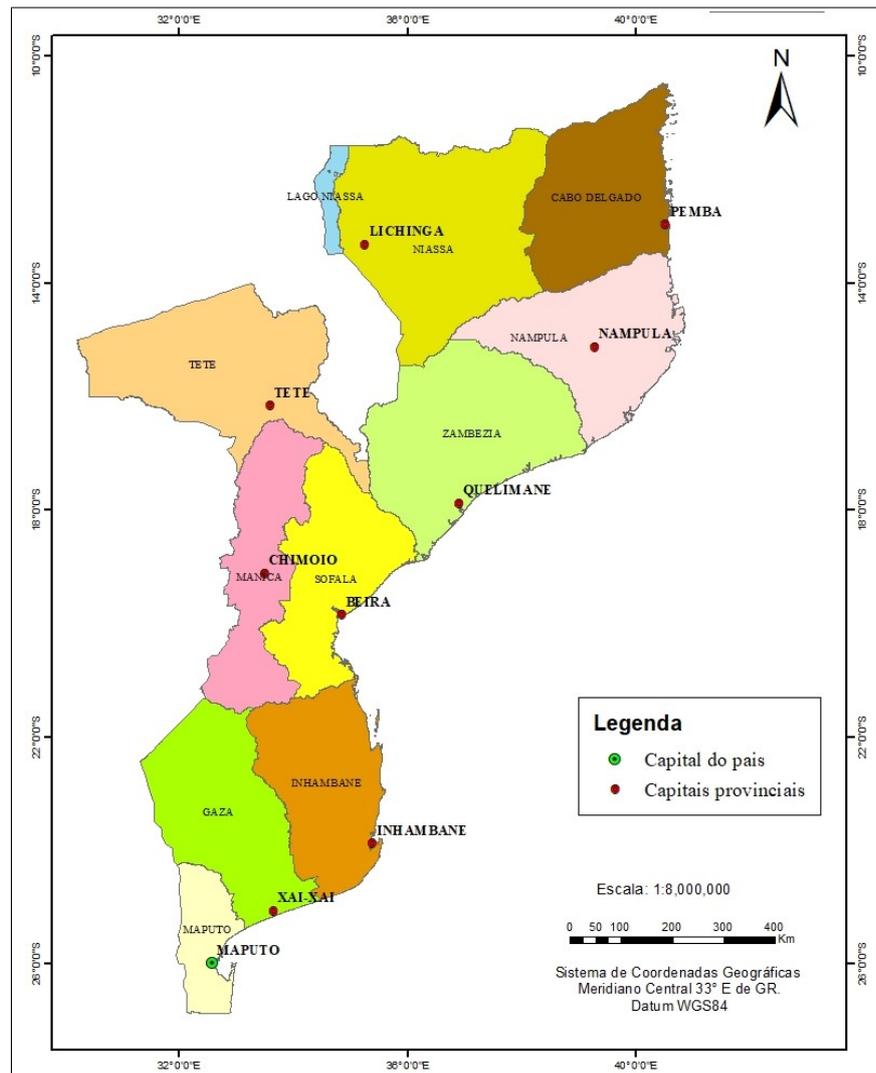


Figura 2.1: Mapa de Moçambique.

Fonte: Chilombe, (2024).

De acordo com o Instituto Nacional de Estatística (INE, 2017), Moçambique é composto por cerca de 27.909.798 milhões de habitantes, sendo 46.6% crianças numa faixa etária dos 0-14.

De acordo com a UNICEF (2023), o recurso mais precioso de Moçambique são as suas crianças. O seu bem-estar e prosperidade são inseparáveis das do país. A população de Moçambique é jovem: dois terços dos habitantes do país têm menos de 25 anos de idade e mais de metade são crianças. Foram feitos grandes progressos ao serviço das crianças, mas há ainda muito trabalho a fazer para proteger os seus direitos e assegurar o seu presente e futuro.

De acordo com Balegamire et al. (2016), a saúde é um direito do Homem, no entanto, metade da população mundial carece de serviços básicos de saúde. Em vários países do mundo, sobretudo os países em vias de desenvolvimento, como Moçambique, existem vários obstáculos no acesso aos serviços de saúde, que constituem uma barreira para cobertura universal e contribuem para a privação do direito a saúde das populações. Em Moçambique um dos grandes desafios tem sido, melhorar as elevadas taxas de desnutrição crónica sobretudo em crianças, alta insegurança alimentar, níveis insuficientes de saneamento básico, etc.

Segundo Perlito (2014), o estado de nutrição assume-se como sendo um dos melhores indicadores de saúde, tanto a nível individual como da comunidade, com destaque para a criança, uma vez que o seu crescimento e desenvolvimento estão dependentes de uma adequada alimentação e nutrição.

De acordo com a UNICEF (2021), em Moçambique antes da emergência decorrente dos ciclones, os indicadores nutricionais já eram preocupantes, onde: 43% das crianças menores de 5 anos de idade sofriam de desnutrição crónica, sem diminuição desse facto nos últimos 10 anos, 6% sofrem de desnutrição aguda e apenas 55% das crianças com menos de seis meses de idade são exclusivamente amamentadas.

Segundo Brandão e Pereira (2022), a prevalência da desnutrição crónica em crianças moçambicanas deve-se, sobretudo, às condições ambientais (mudanças climáticas que resultam em cheias e secas frequentes) que dificultam a subsistência das famílias que dependem em 90% da agricultura, condições de saneamento e higiene precárias que causam diarreias nas crianças, sendo uma das principais causas de morte infantil e associação à casos de desnutrição crónica e as práticas alimentares inadequadas de mães e crianças em fase de amamentação, dado o fraco exercício de acções integradas de comunicação para a mudança do comportamento, resultando numa fraca consciência e falta de conhecimento do problema.

De acordo com o INE et al. (2013), a proporção de crianças que sofrem de subnutrição crónica é elevada nas crianças residentes nas áreas rurais 46%, que nas zonas urbanas 35%. As províncias da região norte de Moçambique são as que registam taxas de prevalência mais elevadas, sendo as províncias de Nampula e Cabo Delgado as que se evidenciam com 55% e 52%, respectivamente. Em contrapartida as províncias da região sul com excepção da

província de Inhambane, apresentam as taxas mais baixas, onde Maputo província e Maputo cidade se destacam, ambas com 23%.

De acordo com o INE et al. (2015), a prevalência da desnutrição diminui conforme o nível de escolaridade dos pais, sobretudo em crianças com mães relativamente mais instruídas e com quintil de riqueza mais alto. O nível de exposição aos meios de comunicação, tais como a rádio e a televisão que são meios de disseminação da informação relativa a temas como saúde e nutrição, o que influencia nas práticas de cuidados infantis adaptados.

De acordo com a UNCEF (2014), a amamentação exclusiva até aos 6 meses de idade é uma das práticas-chave para reduzir os riscos de desnutrição, infecção e mortalidade infantil. Em Moçambique, a proporção de crianças dos 0 aos 5 meses exclusivamente amamentadas, passou de 30% em 2003 para 37% em 2008 e para 43% em 2011, em que a duração média da amamentação exclusiva passou de 0.7 para 1.3 meses entre 2003 e 2011 .

De acordo com o WorldBank (2020), em Moçambique, mais de 70% das famílias pobres vive em áreas rurais e depende principalmente da agricultura para a obtenção de alimentos e geração de renda. Moçambique possui cerca de 36 milhões de hectares de terras férteis e aráveis, mas actualmente, apenas 16% das terras adequadas para agricultura são cultivadas.

Segundo o Secretariado Técnico de Segurança Alimentar e Nutricional (SETSAN, 2014), os níveis de produtividade e produção agrícola em Moçambique, continuam baixos. O acesso a produção de alimentos e conseqüentemente o acesso aos alimentos a nível do país continua limitado, particularmente na região norte é fraca. Os altos níveis de pobreza (54.7%) que prevalecem no país, continuam a afectar negativamente a situação de SAN e contribuem para níveis de insegurança alimentar crónica (24%)e desnutrição crónica (43%) em crianças menores de 5 anos.

2.4 Desnutrição Crónica na Província de Nampula

De acordo com Da Barca e Dos Santos (1992), Nampula está situada na região norte de Moçambique, é a província com a maior densidade populacional. Composta por 23 distritos, 3 cidades e 7 municípios. Faz fronteira a Norte através do rio Lúrio, com as províncias de Cabo Delgado e Niassa. A Sudoeste esta separada pelo rio Ligonha da Zambézia, encontrando-se a Este com o oceano índico.

Segundo De Araújo (2005), a cidade de Nampula situa-se, aproximadamente, no centro do espaço geográfico do distrito que leva o mesmo nome (distrito de Nampula) e ocupa uma área de 404 km². Administrativamente, a cidade constitui um município, mas que não pertence ao distrito rural de Nampula que, territorialmente a cerca. 43.5% Dos 6102867 habitantes da província de Nampula tem menos de 15 anos e apenas 1.2% tem 65 ou mais anos, o que sugere uma população essencialmente jovem.

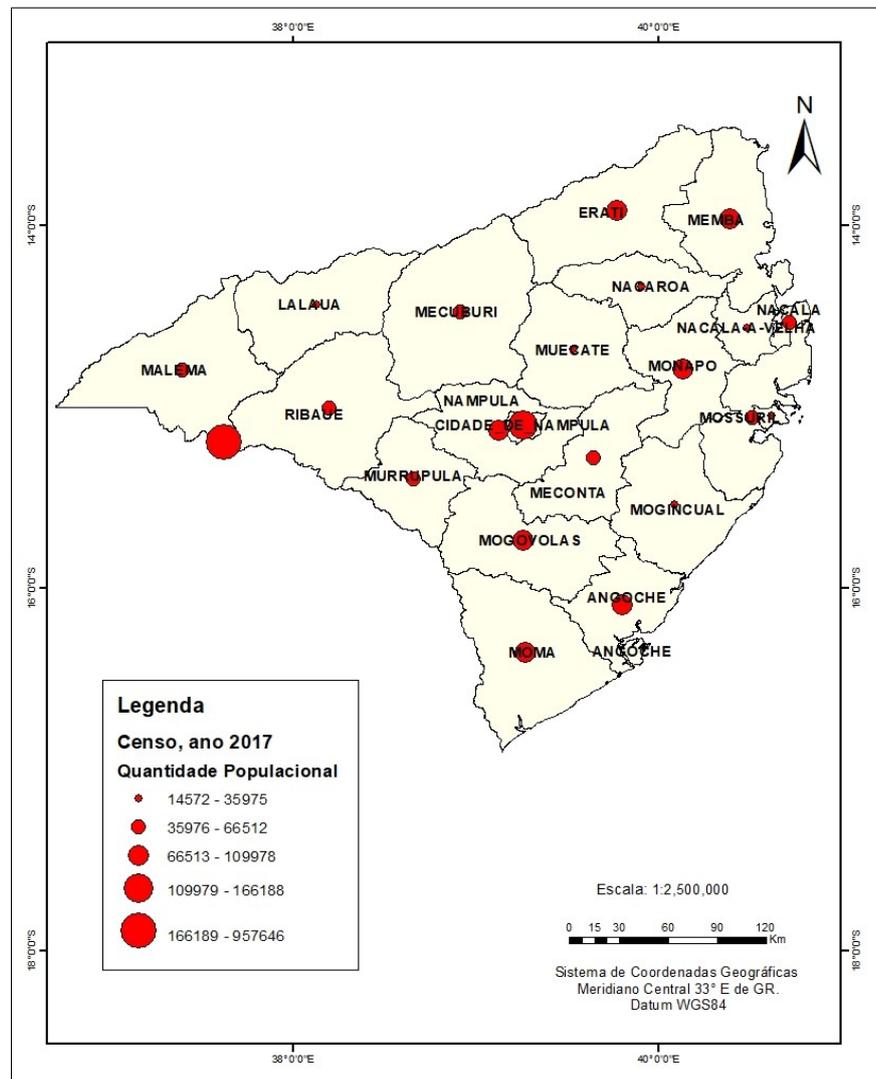


Figura 2.2: Mapa da Província de Nampula.

Fonte: Chilombe, 2024.

Segundo Abbas e Mosca (2021), a província de Nampula destaca-se entre as províncias com maior superfície utilizada para agricultura no país, sendo a de destaque na zona norte. Esta é a que mais área dedica as culturas de mandioca cerca de 39% da área nacional, amendoim 32%, feijão nhemba 23% e algodão 51%.

De acordo com o INE et al (2013), entre as 11 províncias do país, Nampula tem a maior população e é economicamente importante devido à produtividade agrícola e ao seu posicionamento como corredor de transporte para as regiões do interior e outros países. No entanto, a prevalência de atraso de crescimento entre crianças com menos de cinco anos é mais elevada em Nampula (47%) e a prevalência de desperdício é a terceira mais elevada 5%, apesar dos inúmeros investimentos em programas de Nutrição, Saúde, HIV, Água, Saneamento, Agricultura e Crescimento económico.

De acordo com a UNICEF (2022), a situação do bem-estar das crianças em diferentes partes do país varia significativamente. Em Nampula, mais de 2 milhões de crianças das cerca de 3,4 milhões, encontram-se numa situação de pobreza multidimensional. Assim 27% das crianças que vivem na pobreza multidimensional em Moçambique vivem na província de Nampula. As crianças de Nampula ficam acima da média nacional na maioria dos indicadores sociais, tais como desnutrição crónica, taxa de conclusão escolar, taxa de analfabetismo e acesso aos serviços básicos.

De acordo com a USAID (2023), o número de prestadores de cuidados de saúde disponíveis para servir os doentes é insuficiente e a maioria dos prestadores de cuidados de saúde, os profissionais, não têm formação nutricional específica. Para combater a subnutrição, coordenação multisectorial é crucial, mas subsistem lacunas críticas na província de Nampula e no país.

De acordo com Ibramugy (2013), Nampula é uma das províncias moçambicanas com taxas elevadas de desnutrição, sendo os distritos de Ribaué, Murrupula, Naca-velha, Muecate e Mogincual os que lideram a lista dos mais afectados pelo problema da desnutrição crónica. Dados que aumentam a preocupação e questionamentos, principalmente porque os distritos de Ribaué e Murrupula, são dos mais produtivos de Nampula.

Segundo Zano (2017), a contínua urbanização mundial está de certa forma, a comprometer a alimentação dos cidadãos. Não ficando de fora, Moçambique também está a ser afectado pelo fenómeno da urbanização mundial. A província de Nampula, é considerada a mais populosa do País. O crescimento da cidade de Nampula pode ser atribuído ao processo de expansão da urbanização com a absorção de áreas rurais pela cidade. Os espaços antes utilizados para lazer e recreação e outras áreas verdes onde se praticava a agricultura, em bairros como Namicopo, Murrapaniua, Napipine, Muatala e Merrene, foram dando espaço a construções desordenadas, na sua maioria de baixo custo assim como infra-estruturas, tais como barracas para o comércio e estabelecimentos de diversão noturna. O volume de produção alimentar local não permite o abastecimento da população da cidade de Nampula, o que pode estar contribuindo para os elevados índices de desnutrição observados nesta província.

2.5 Técnica Estatística

Segundo Amaral et al (2009), análise de regressão é uma técnica estatística utilizada para investigar a relação existente entre variáveis através da construção de uma equação (um modelo). De modo geral, essa técnica pode ser utilizada com vários objectivos, entre os quais pode-se destacar: descrever a relação entre variáveis para entender um processo ou fenómeno; prever o valor de uma variável a partir do conhecimento dos valores das outras; substituir a medição de uma variável pela observação dos valores de outras variáveis; controlar os valores de uma variável em uma faixa de interesse.

2.5.1 Modelos Lineares Generalizados

Uma das formas de analisar o padrão de associação e interação entre uma variável de interesse, denominada variável resposta, e outras variáveis, denominadas variáveis explicativas, é por meio de modelos estatísticos. Os parâmetros determinam a intensidade e a importância dos efeitos e inferências podem ser feitas sobre esses parâmetros para avaliar quais variáveis realmente afectam a resposta. Os valores preditos pelo modelo melhoram a estimativa da média da variável resposta nos possíveis valores das variáveis explicativas.

De acordo com Da Rocha (2014), os modelos Lineares generalizados são caracterizados por três componentes:

i) Componente Aleatório - é a especificação da distribuição de probabilidade da variável resposta Y_i . As observações (y_1, \dots, y_n) geralmente são consideradas independentes.

ii) Componente Sistemático - especifica as variáveis explicativas, de forma linear. Sejam (x_1, \dots, x_k) as k variáveis explicativas. A combinação linear das variáveis explicativas, denominada preditor linear, é dado por:

$$\eta_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.1)$$

iii) Função de Ligação - que descreve como a média $E(Y_i) = \mu$ depende do preditor linear, isto é, conecta os componentes aleatório e sistemático:

$$g(\mu) = \eta_i \quad (2.2)$$

onde μ representa a expectativa de Y e $g(\mu)$ é a função de ligação.

Tabela 2.1: Variações dos Modelos Lineares Generalizados

Componente aleatória	Tipos de covariáveis	Funções de Ligação	Tipo de modelo
Normal	Contínuas	Identidade	Regressão Linear
Normal	Categorizadas	Identidade	Análise da Variância
Normal	Mistas	Identidade	Análise da Covariância
Binomial	Mistas	logit	Regressão Logística
Poisson	Mistas	logaritmica	Log-linear

Estimação dos Parâmetros

Segundo Vaz (2020), o método mais usual em estatística para a estimação de parâmetros dos modelos lineares generalizados é o método da máxima verossimilhanças. Proposto por Ronald A. Fisher, este método busca encontrar os valores dos parâmetros que maximizam a função log-verossimilhança. Isso é feito através da otimização da função de verossimilhança, que é uma medida da concordância entre os dados observados e os valores previstos pelo modelo. A máxima verossimilhança é amplamente utilizada devido à sua robustez e eficiência na estimação de parâmetros em modelos lineares generalizados.

Método de Máxima Verossimilhança

Segundo Pessoa (2013), a ideia básica do método, é encontrar um conjunto de estimadores que maximizem os parâmetros populacionais à luz da amostra dos dados observados. Para isso, a função de probabilidade conjunta da amostra, sob o modelo especificado, é avaliada para cada uma das observações da variável resposta, sendo tratada como uma função dos parâmetros do modelo.

2.5.2 Regressão Logística

Segundo Anderson et al (2009), a regressão logística é uma forma especializada de regressão que é formulada para prever e explicar uma variável categórica binária (dois grupos). A forma da variável estatística de regressão logística é semelhante à da variável estatística da regressão múltipla. Assim o objectivo fundamental da regressão logística é prever uma variável dependente dicotômica a partir de um conjunto de variáveis independentes contínuas ou não, mediante um modelo que apresenta um bom ajuste.

Segundo Figueira (2006), a regressão logística binária ou univariada, representa casos de regressão logística em que a variável dependente Y é binária ou dicotômica, ou seja, tem duas categorias e apenas uma variável independente.

A Regressão Logística é utilizada para prever e modelar a probabilidade de ocorrência de um evento binário, como “sim” ou “não”, “passou” ou “não passou”, “doente” ou “não doente”, com base em variáveis explicativas ou preditoras (Gonçalves, 2023). Ao contrário da Regressão Linear, que prevê valores contínuos e assume uma relação linear entre as variáveis independentes e a variável dependente, essa suposição só faz sentido para variáveis contínuas, pois variáveis categóricas (como “sim” ou “não”) não apresentam uma progressão linear. A Regressão Linear pode gerar valores fora do intervalo de 0 a 1, o que não é apropriado para variáveis categóricas. Além disso, ela pressupõe que os erros residuais seguem uma distribuição normal, algo inadequado para variáveis binárias.

A Regressão Logística tem uma ampla aplicação, sendo uma ferramenta valiosa para a tomada de decisões e a identificação de fatores importantes que influenciam a ocorrência de um evento específico, permitindo a previsão e compreensão de eventos binários com base em variáveis explicativas. Ela pode ser aplicada em diversas áreas, como medicina, marketing e finanças, entre outras.

- **Medicina**

Pesquisadores da área de medicina desenvolvem tratamento e cuidados preventivos estimando a probabilidade de doença em pacientes. Eles usam modelos de regressão logística para comparar o impacto do histórico familiar ou dos genes nas doenças.

- **Marketing**

As ferramentas de publicidade *on-line* usam o modelo de regressão logística para prever se os usuários aderem. Como resultado, os profissionais de marketing podem analisar as respostas dos usuários usando diferentes palavras e imagens e criar anúncios de alto desempenho que terão interação dos clientes.

- **Finanças**

As empresas financeiras precisam analisar as transações financeiras em busca de fraude e avaliar os pedidos de empréstimo e os pedidos de seguro quanto a riscos. Esses problemas são adequados para um modelo de regressão logística porque tem resultados discretos, como alto risco ou baixo risco e fraudulentos ou não fraudulentos.

Vantagens do Modelo Logístico

Segundo Mesquita (2014), há várias razões para o uso da regressão logística, entre elas pode-se citar:

- Facilidade no uso de variáveis independentes categóricas;
- Utilização de variáveis métricas e não métricas simultaneamente;
- Fornece resultados em termos de probabilidade;
- Fornece um bom ajuste, com menos parâmetros do que os modelos lineares;
- Facilidade na interpretação de resultados, dada a sua semelhança com a regressão linear.

Limitações do Modelo Logístico

Segundo Bittencourt e Clarke (2001), no modelo logístico não se evidenciam desvantagens contudo podem ser observadas algumas limitações tais como:

- Obrigatoriedade na utilização de métodos numéricos para a obtenção da solução de máxima verossimilhança;
- Sensibilidade à colinearidade, sendo suas consequências: erros padrão extremamente elevados que não permitem a realização de testes de significância para os coeficientes \mathbf{b} .

2.5.3 Regressão Logística Simples

Segundo Da Rocha (2014), considera-se casos em que a variável dependente Y é categórica, que a única a variável explicativa seja quantitativa. O modelo de regressão logística tem forma linear para o logito dessa probabilidade:

$$\text{logit}(\pi(i)) = \ln \left(\frac{\pi_i}{1 + \pi_i} \right) = \beta_0 + \beta_1 x_i \quad (2.3)$$

Estimação dos parâmetros no modelo de regressão logística simples

Para Martins (2012), enquanto no modelo de regressão linear a estimação dos parâmetros é efectuada através do método dos mínimos quadrados, no modelo de regressão logística a estimação é feita pelo método da máxima verossimilhança.

Segundo Vaz (2020), na regressão não linear o método da máxima verossimilhança é utilizado de forma iterativa para que sejam encontradas as estimativas mais prováveis dos parâmetros. Ao invés de minimizar os desvios quadrados, a regressão não linear maximiza a probabilidade de que um evento ocorra.

Sabendo que os dados seguem uma distribuição de bernoulli, e dado que as observações de conjunto de dados são independentes, a função de máxima verossimilhança é dada por :

$$L(\beta) = \prod_{i=1}^n [\pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}] \beta \in R^2 \quad (2.4)$$

onde $\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$ e $L(\beta)$ é função de verossimilhança

O princípio da máxima verossimilhança é estimar o valor de β que maximiza $L(\beta)$. Aplicando o logaritmo, a função fica definida conforme a equação 2.5 :

$$\begin{aligned} l(\beta) &= \ln \left(\prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} \right) \\ l(\beta) &= \sum_{i=1}^n [Y_i \ln(\pi_i) + (1 - Y_i) \ln(1 - \pi_i)] \end{aligned} \quad (2.5)$$

onde $l(\beta)$ é a verossimilhança logarítmica. Substituindo o π_i pela sua forma exponencial, obtem-se:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n \left[Y_i \ln \left(\frac{1}{1 + e^{-\beta_0 + \beta_1 x_i}} \right) + (1 - Y_i) \ln \left(\frac{e^{-\beta_0 + \beta_1 x_i}}{1 + e^{-\beta_0 + \beta_1 x_i}} \right) \right] \\ l(\beta) &= \sum_{i=1}^n Y_i \left[\ln \left(\frac{1}{1 + e^{-\beta_0 + \beta_1 x_i}} \right) - \ln \left(\frac{e^{-\beta_0 + \beta_1 x_i}}{1 + e^{-\beta_0 + \beta_1 x_i}} \right) \right] + \ln \left(\frac{e^{-\beta_0 + \beta_1 x_i}}{1 + e^{-\beta_0 + \beta_1 x_i}} \right) \\ l(\beta) &= \sum_{i=1}^n Y_i \left[\log(e^{\beta_0 + \beta_1 x_i}) + \log \left(\frac{e^{-\beta_0 + \beta_1 x_i}}{1 + e^{-\beta_0 + \beta_1 x_i}} \cdot \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i}} \right) \right] \end{aligned}$$

$$l(\beta) = \sum_{i=1}^n \left[Y_i(\beta_0 + \beta_1 x_i) + \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right]$$

$$l(\beta) = \sum_{i=1}^n [Y_i(\beta_0 + \beta_1 x_i) - \ln(e^{\beta_0 + \beta_1 x_i})] \quad (2.6)$$

A expressão 2.6 é então a função de verossimilhança usada para encontrar o valor de β que maximiza $L(\beta)$. O valor desses parâmetros é obtido derivando $l(\beta)$ em relação a cada parâmetro (β_0, β_1) , conforme as equações 2.7 e 2.8.

$$\frac{\partial \ln L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(y_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = \sum_{i=1}^n (y_i - \pi_i) \quad (2.7)$$

$$\frac{\partial \ln L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(y_i x_i - \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = \sum_{i=1}^n x_i (y_i - \pi_i) \quad (2.8)$$

onde $i = 1, 2, \dots, n$ e $\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$

Igualando as equações 2.7 e 2.8, obtém-se o valor do parâmetro β que maximiza a função de verossimilhança $l(\beta)$.

Por serem não-lineares, estas equações são resolvidas recorrendo métodos iterativos, como por exemplo Newton-Raphson, e o resultado desta aplicação são incluídos na matriz denominada de Informação de Fisher. A matriz de informação de Fisher, para o modelo logístico com uma variável, tem a seguinte forma:

$$L(\beta) = \begin{bmatrix} \sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^n x_i \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\ \sum_{i=1}^n x_i \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^n x_i^2 \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \end{bmatrix}$$

$$L(\beta) = \sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix}$$

Após obter as estimativas dos parâmetros do modelo pode-se calcular as probabilidades estimadas:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (2.9)$$

2.5.4 Regressão Logística Múltipla

Segundo Da Rocha (2014), em caso em que se tem mais de uma variável explicativa, a regressão é usualmente chamada de regressão múltipla. O modelo de regressão logística múltiplo é uma extensão do modelo logístico simples.

Segundo Moura (2018), no modelo de regressão múltiplo, a probabilidade de sucesso é dada por:

$$\pi(i) = \frac{e^{g(\mu)}}{1 + e^{g(\mu)}} \quad (2.10)$$

e a probabilidade de insucesso é dada por :

$$1 - \pi(i) = 1 - \frac{e^{g(\mu)}}{1 + e^{g(\mu)}} \quad (2.11)$$

onde $g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$.

Dependendo das áreas de aplicação da regressão logística, existem várias possibilidades de escolha para as covariáveis, como por exemplo, sexo, cor dos olhos, grupo de tratamento etc. Assim, surge a necessidade de atribuir valores numéricos, meramente identificativos, a cada categoria da variável, contudo, é necessário utilizar as variáveis dummy. Estas variáveis são definidas da seguinte forma :

$$\begin{cases} 0, \textit{Característica Ausente} \\ 1, \textit{Característica Presente} \end{cases}$$

Estimação dos parâmetros no modelo de regressão logística múltipla

Segundo Moura (2018), a estimação dos parâmetros pelo método da máxima verossimilhança é dada de modo que os estimadores maximizem o logaritmo da função de verossimilhança.

Para estimar os parâmetros da regressão logística múltipla por máxima verossimilhança encontra-se o valor de β que maximiza $L(\beta)$, o que exige um processo iterativo e que seja necessário derivar $L(\beta)$ em relação a cada parâmetro:

$$\frac{\partial[l(\beta)]}{\partial(\beta_0)} = \sum_{i=0}^n [y_i x_{ij} - \frac{e^{g(\mu)}}{e^{g(\mu)} + 1} x_{ij}] \quad (2.12)$$

Onde $l(\beta)$ é a função (2.5).

A matriz de covariância dos coeficientes estimados é obtida a partir das derivadas parciais de segunda ordem do logaritmo da função de verossimilhança:

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j^2} = - \sum_{i=1}^n [x_{ij}^2 \pi_i (1 - \pi_i)] \quad (2.13)$$

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n [x_{ij} x_{ik} \pi_i (1 - \pi_i)] \quad (2.14)$$

Onde $j, k = 0, 1, 2, 3, \dots, p$, e p é o número de parâmetros desconhecidos do modelo

Se for formada uma matriz quadrada de dimensão $(p + 1)$ constituída pelo simétrico dos valores médios dos termos referidos nas equações (2.16) e (2.17)) obtém-se $l(\beta)$, a chamada Matriz de Informação. A matriz de informação de Fisher é dada por:

$$l(\beta) = E\left(-\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T}\right) = X^T Q X \quad (2.15)$$

em que $Q = \text{diag}[\pi_i(1 - \pi_i)]$, $i=1,2,3,\dots,n$ e X a matriz dos dados, e sua inversa $l^{-1}(\beta)$ é a matriz de variância e covariância das estimativas de máxima verossimilhança dos parâmetros.

2.5.5 Transformação Logít

Considere-se um vector de variáveis independentes $X = [x_1, \dots, x_n]$ e uma variável aleatória real Y binária, tal que $P(Y = 1) = \pi_i$, é a probabilidade de ter sucesso, e $P(Y = 0) = 1 - \pi_i$, a probabilidade de ter insucesso. Verifica-se que Y segue uma distribuição de Bernoulli.

Em qualquer modelo de regressão pretende-se determinar $E(Y_i)$. Na regressão linear, esta esperança é dada como uma equação linear em x .

$$E(Y) = \beta_0 + \sum_{i=1}^{p-1} \beta_i x_i = \mu \quad (2.16)$$

Nesta expressão, $E(Y)$ e X podem tomar qualquer valor entre $-\infty$ e $+\infty$. Entretanto, na regressão logística Y_i é uma variável binária o que implica que o valor de $E(Y_i)$ deve variar no intervalo $[0, 1]$. Da definição de esperança tem-se $E(Y_i) = \pi_i$.

Ao aplicar a transformação logit à função $\pi(i)$, onde logit é a função de ligação para o modelo de regressão logística que se denota $g(\mu)$, obtém-se:

$$g(\mu) = \ln \left(\frac{\frac{e^\mu}{1 + e^\mu}}{1 - \frac{e^\mu}{1 + e^\mu}} \right) = \beta_0 + \sum_{i=1}^{p-1} \beta_i x_i = \mu \quad (2.17)$$

2.6 Métodos de Seleção das Variáveis

De acordo com Moura (2018), a seleção das variáveis do modelo é baseada em algum algoritmo que verifica a importância de dada variável e a sua inclusão ou não no modelo, nomeadamente: método enter, método forward, método backward e método stepwise.

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

2.7 Critérios para a Seleção do Modelo

Escolher o melhor modelo é controverso, mas um bom modelo deve conseguir equilibrar a qualidade do ajuste e a complexidade, sendo esta, em geral, medida pelo número de parâmetros presentes no modelo, quanto mais parâmetros, mais complexo o modelo, sendo mais difícil interpretar o modelo. A seleção do “melhor” modelo torna-se então necessária.

2.7.1 Critério de Informação de Akaike

Para Al-rubaie et al (2006), um dado pode ser testado utilizando-se critérios de informação construídos para avaliar o viés e variância da função. Quando muito simples o modelo não consegue prever a irregularidade dos dados e, por tanto tem mais viés, em contrapartida quando muito complexo o modelo tende a distribuir dados em muitos parâmetros, resultando em uma má estimação de cada parâmetro e numa maior variância. O melhor modelo é então aquele que consegue um equilíbrio entre viés e variância, avaliados pelo segundo Critério de Informação de Akaike .

Um critério de selecção que tem sido muito utilizado na selecção de modelos é o critério de informação Akaike(AIC), conforme a equação :

$$AIC = -2 \sum_{i=1}^n \ln L(\hat{\theta}_i, y_i) + 2 * (p) \quad (2.18)$$

onde é (p) é o número de parâmetros, y_i é o i-ésimo valor da resposta e $\hat{\theta}$ é uma estimativa de y_i .

De acordo com Burnham e Anderson (2002), o AIC foi desenvolvido a partir da divergência de Kullback-Leibler (K-L), que mensura a diferença entre os dois modelos. É recomendada a utilização do AIC apenas quando a razão entre o número de observações e variáveis preditoras (n/p) for maior ou igual a 40. A expressão AIC pode ser simplificada, conforme a equação 2.19.

$$AIC = n \ln(\hat{\sigma}_p^2) + 2 * (p + 1) \quad (2.19)$$

onde $\hat{\sigma}_p^2$ é o estimador de máxima verossimilhança da variância do erro, conforme a equação (2.20). Esse estimador pode ser obtido pela seguinte expressão:

$$\hat{\sigma}_p^2 = \frac{\sum_{i=1}^n (y_i - \hat{\mu}_i)^2}{n} \quad (2.20)$$

Segundo Ferreira (2021), conforme podemos observar Critério de Informação de Akaike realiza um processo de minimização que não envolve testes estatísticos e pode ser expresso em função do desvio do modelo, e é baseado na função de verossimilhança. Fundamenta-se no conceito de entropia, oferecendo uma medida relativa das informações perdidas, quando um determinado modelo é usado para descrever a realidade.

2.7.2 Critério de Informação Bayesiano

O Critério de informação Bayesiano (BIC), também chamado de Critério de Schwarz, foi proposto por Schwarz (1978), e é um critério de avaliação de modelos definido em termos da probabilidade a posteriori. Schwarz (1978) definiu o Critério de Informação de Bayes (BIC) como:

$$BIC = -2(L) + 2(K.\ln(n)) \quad (2.21)$$

onde n é o número de observações. e K o número de parâmetros estimados no modelo.

Assim, os melhores modelos serão aqueles que apresentarem valores baixos destes critérios.

- **Teorema de Bayes**

De acordo com Bolfarine e Sandoval (2000), quando dois eventos de um mesmo espaço amostral são levados em consideração conjuntamente, passa a haver sentido (conjeturar) se a ocorrência ou não de um afecta a ocorrência ou não do outro, isto é, se são independentes ou não. Intuitivamente, chega-se a definição de que dois eventos são independentes se, $P[A \cap B] = P[A]P[B]$. Entretanto, se há dependência entre os eventos, passa a haver sentido falar na probabilidade de que um evento ocorra dada a ocorrência ou não do outro. Esta dependência motiva a definição de probabilidade condicional. Assim, os conceitos de independência e probabilidade condicional levarão ao teorema de Bayes.

Terorema . *Se $(\Omega, \mathcal{A}, P[\cdot])$ é um espaço de probabilidade e B_1, B_2, \dots, B_n é uma coleção de eventos mutuamente exclusivos em \mathcal{A} . satisfazendo $\Omega = \cup_{i=1}^n B_j$ e $P[B_j] > 0$, para $j = 1, 2, \dots, n$, então para todo $A \in \mathcal{A}$, que $P[A] > 0$ tem - se :*

$$P[B_k | A] = \frac{P[A | B_k]P[B_k]}{\sum_{j=1}^n P[A | B_j]P[B_j]}, \quad (2.22)$$

sendo Ω o espaço amostral e \mathcal{A} o espaço paramétrico.

Capítulo 3

Material e Métodos

3.1 Material

O presente trabalho foi realizado com recurso à uma base de dados secundária, fornecida pelo SETSAN, relativa ao inquérito sobre a avaliação da segurança alimentar e nutricional realizado em Moçambique na província de Nampula em 2022. A mesma é composta por 17 variáveis e 1802 observações.

Tabela 3.1: Descrição das variáveis em estudo

Variáveis Demográficas	Variáveis Relativas a Saúde
Género da criança(X_1)	Ocorrência de febres(X_7)
Idade da criança X_2	Ocorrência de diarreias(X_8)
Peso da criança (X_3)	Ocorrência de tosse (X_9)
Altura da criança(X_4)	Ocorrência de diarreia aquosa(X_{10})
Idade do chefe do agregado (X_5)	Ocorrência de desnutrição aguda(X_{11})
Género do chefe do agregado (X_6)	
Variáveis Ambientais	Variáveis Socioeconómicas
Fonte de água no agregado (X_{12})	Fonte de Renda do Agregado (X_{15})
Tamanho do agregado (X_{13})	Nível de acesso a informação (X_{16})
Número de crianças menores de 5 anos no agregado(X_{14})	
Variável Dependente:	Desnutrição Crónica(Y)

Os dados foram processados no software R 4.3.1, para a construção dos gráficos foi utilizado Microsoft Excel 2013 e para a elaboração do trabalho foi usado o software LaTeX 4.7.2 .

Neste trabalho todas as hipóteses foram testadas a um nível de significância de 0.05, onde rejeita-se a hipótese nula nos casos em que o p-value associado a estatística do teste for menor que o nível de significância de 5%.

3.2 Métodos

Convista a alcançar os objectivos do estudo, foram calculadas as estatísticas descritivas para descrever as características das crianças da província de Nampula, usou-se a regressão logística para identificar os factores associados a desnutrição crónica em crianças, usou-se ainda o modelo de regressão logístico para prever a probabilidade de uma criança estar em situação de desnutrição crónica.

3.2.1 Análise Descritiva

De acordo com Mulenga (2018), a estatística descritiva é o ramo ou parte da estatística cujo objectivo é observar fenómenos da mesma natureza, recolha, organização, classificação, análise e interpretação dos dados sem deixar de calcular algumas medidas (estatísticas), que permitem resumidamente descrever o fenómeno estudado.

Para Guedes et al (2005), objectivo básico da análise descritiva é resumir uma série de valores da mesma natureza, permitindo uma visão global da variação desses valores, pode-se representar os dados usando gráficos, tabelas, diagramas e medidas descritivas.

3.2.2 Testes de Significância dos Coeficientes do Modelo

Após a estimação dos coeficientes do modelo, avalia-se a qualidade da estimação, ou seja, testa-se se as variáveis explicativas pertencentes ao modelo são significativas para explicar o comportamento da variável resposta.

De acordo com Hosmer e Lemeshow (2000), há vários testes para avaliar o modelo logístico, destacando-se, nomeadamente: teste de razão de verossimilhança, teste de Wald e as medidas pseudo- R^2 .

- Teste de Razão de Verossimilhança

De acordo com Maroco (2007), uma vez que os erros do modelo logístico não seguem distribuição normal nem apresentam variância constante, não é possível obter estimativas óptimas dos coeficientes do modelo pelo método dos mínimos quadrados usados pela regressão linear. O método de ajustamento usado na regressão logística é o Método da Máxima Verossimilhança. Com este teste pretende-se testar simultaneamente se os coeficientes de regressão associados a β são todos nulos, com excessão de β_0 .

Pela teoria de estimação de máxima verossimilhança, sabe-se que os estimadores de máxima verossimilhança maximizam a função log-verossimilhança. O teste de razão de verossimilhança

avalia se o valor de log-verossimilhança é suficientemente grande para concluir que as variáveis retiradas são importantes para o modelo.

A estatística de teste à significância do modelo compara a verossimilhança de um modelo só com a constante (modelo nulo ou modelo reduzido) com a verossimilhança do modelo com as variáveis explicativas (modelo completo). Como verossimilhança (L) é inferior a 1, e geralmente muito pequena, é usual usar $LL = \ln(L)$, que é um número negativo, pelo que se multiplica por -2 para torná-lo positivo, maior e com distribuição amostral conhecida. O $-2LL$ é um indicador da mediocridade do ajustamento do modelo aos dados: Quanto maior for $-2LL$, pior é o ajustamento; por outro lado, se $-2LL = 0$, o ajustamento é perfeito.

A estatística de teste para testar a significância do modelo de regressão logística é dada pela seguinte expressão:

$$G = -2\ln \left(\frac{\text{Verossimilhança do modelo reduzido}}{\text{Verossimilhança do modelo completo}} \right) \quad (3.1)$$

De forma alternativa, o teste ao ajustamento do modelo pode fazer-se com a estatística Deviance. Essa estatística avalia o valor ajustado na regressão logística, desempenhando o mesmo papel que a soma de quadrados residuais tem na regressão linear. A deviance sempre é positiva e quanto menor, melhor é o ajuste do modelo. Para estimar a significância de uma variável independente, comparam-se o valor 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:

$$D = -2\ln \left(\frac{\text{Verossimilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right) \quad (3.2)$$

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

As hipóteses do teste são:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$ (O modelo não é adequado)

H_1 : Existe pelo menos um $\beta_i \neq 0$ $i=1,2,3,\dots,p$ (O modelo é adequado)

Ao rejeitar-se a hipótese nula pode-se concluir que, pelo menos um dos coeficientes é estatisticamente diferente de 0. Por esta razão, antes de se chegar a conclusão final, deve-se testar se cada um dos coeficientes é significativamente diferente de zero, sendo para isso realizado o teste de Wald.

- **Teste de Wald**

Segundo Cabral (2013), o teste de Wald é um teste estatístico paramétrico que testa se cada coeficiente é significativamente diferente de zero. Desta forma, este teste averigua se uma determinada variável independente apresenta uma relação estatisticamente significativa com a variável dependente. A estatística do teste de Wald é dada pela seguinte expressão:

$$W_i = \frac{\hat{\beta}_i}{EP(\hat{\beta}_i)} \quad (3.3)$$

Onde $\hat{\beta}_i$ é estimador de β_i e $EP(\hat{\beta}_i)$ é o erro padrão de $\hat{\beta}_i$.

As hipóteses referentes a cada parâmetro são:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0,$$

Regra de decisão: Rejeita-se a hipótese nula se o nível de significância associado ao teste (p-valor) for menor ou igual a 5%.

Para analisar a adequação do modelo são utilizados os resíduos de Pearson e os Deviance residuals.

3.2.3 Razão de Chances (Odds Ratio)

Segundo Hosmer e Lemeshow (2000), a razão de chances (odds ratio) é dada pelo quociente entre a odds do acontecimento de interesse ocorrer ($Y = 1$) nos indivíduos com $x = 1$ e a odds desse acontecimento ocorrer nos indivíduos com $x = 0$. A odds do acontecimento de interesse ocorrer nos indivíduos com $x = 1$ é definida como:

$$\text{Odds para } x = 1 : \frac{\pi_1}{1 - \pi_1}.$$

Analogamente, a odds do acontecimento de interesse ocorrer nos indivíduos com $x = 0$ é definida como:

$$\text{Odds para } x = 0 : \frac{\pi_0}{1 - \pi_0}.$$

Portanto, o odds ratio é utilizado para comparar se a probabilidade do acontecimento de interesse ($Y = 1$) é a mesma entre os indivíduos com $x = 1$ e $x = 0$, sendo calculado como:

$$\text{Odds Ratio} = \frac{\frac{\pi_1}{1 - \pi_1}}{\frac{\pi_0}{1 - \pi_0}} \quad (3.4)$$

A medida de associação Odds Ratio (OR), denominada razão de chances, é utilizada usualmente na regressão logística univariável para complementar o teste à significância da covariável (x).

A codificação de x permite a interpretação trivial dos parâmetros. Na prática é usado a tabela de contingência para o cálculo de odds ratio. As probabilidades do evento de interesse ocorrer para as duas categorias de x , são dadas pelas expressões 3.5 e 3.6:

$$\pi_{(1)} = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \quad (3.5)$$

$$\pi_{(0)} = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \quad (3.6)$$

Consequentemente, o valor da razão de chances é dado pela expressão 3.7 :

$$OR = \frac{\pi_{(1)}[1 - \pi_{(0)}]}{\pi_{(0)}[1 - \pi_{(1)}]} = e^{\beta_1} \quad (3.7)$$

Onde :

- $OR > 1$ indica que a probabilidade de o evento ocorrer é maior do que a probabilidade do evento não ocorrer.
- $OR < 1$ indica uma diminuição na ocorrência do evento.
- $OR = 1$ indica que a probabilidade de o evento ocorrer não afecta a probabilidade do evento não ocorrer.

3.2.4 Diagnóstico e Avaliação do Ajuste do Modelo

O diagnóstico do modelo é uma etapa fundamental no ajuste de modelos de regressão. Quando um modelo é ajustado a um conjunto de dados tem que se ter em atenção a análise de medidas das diferenças entre os seus valores observados da variável resposta, Y e os resíduos. Vale destacar que os resíduos têm papel fundamental na verificação do ajuste de um modelo.

- Matriz de Confusão

A Matriz de confusão é formada pelo número de vezes que um classificador atribui a classe **A** a uma amostra da classe **B**. Especificamente, as linhas da matriz de confusão $M = (m_{ij}) \in \mathbb{R}^{L \times L}$ estão associados à classe correcta enquanto as colunas correspondem as

previsões do modelo . Os elementos da matriz de confusão dados pela equação :

$$m_{ij} = \sum_{k=1}^m \chi[Y_k = C_i, \hat{Y}_k = C_j] \quad (3.8)$$

Em que o elemento m_{ij} da matriz de confusão contem o número de vezes que o classificador atribui a classe C_j para uma amostra de classe C_i . De um modo geral a matriz de confusão contem informações sobre todos os acertos e erros do classificador (Valle, 2022).

Tabela de classificação

Um modelo de regressão pode ser estatisticamente significativo e não representar a realidade em estudo. Uma das formas de avaliar a eficiência classificativa do modelo é através de tabela classificativa.

Tabela 3.2: Tabela de classificação

		Estimados	
		Endpoint=0	Endpoint=1
Observados	Endpoint=0	V_p	V_n
	Endpoint=1	F_p	F_n
	Desempenho	Sensibilidade: $V_p/(V_p + F_n)$	Especificidade: $V_n/(V_n + F_p)$

Com a matriz de confusão pode-se calcular o Desempenho geral utilizando a equação abaixo:

$$AC = \frac{V_n + V_p}{V_n + V_p + F_n + F_p} \quad (3.9)$$

Como pode-se ver o desempenho calcula a relação das classificações correctas, verdadeiros positivo (V_p) e verdadeiro negativo (V_n) e as outras classificações que incluem também Falsos Negativos (F_n) e Falsos positivos (F_p). Assim essa Medida pode ser interpretada como a taxa de acertos do modelo (Reis, 2021).

Sensibilidade: revela a taxa de V_p entre todas as classes positivas do conjunto de dados, ou seja avalia quantas observações positivas foram classificadas correctamente. Considerando que o *recall* varia de 0 a 1, quanto maior o seu valor maior o número de classificações correctas.

$$SENS = \frac{V_p}{V_p + F_n} \quad (3.10)$$

Precisão: avalia a taxa de V_p entre todos os valores positivos preditos. Enquanto o recall reduz em situações onde o classificador não interpreta corretamente um cenário de falha, a precisão diminui com o aumento de falsos alarmantes.

$$PREC = \frac{V_p}{V_p + F_p} \quad (3.11)$$

Especificidade: representa a proporção de verdadeiros negativos (V_n) em relação ao total de casos negativos. É importante destacar que a especificidade não fornece necessariamente uma indicação precisa sobre o resultado negativo no teste de classificação porque os resultados negativos do teste podem conter muitos resultados falsos negativos. Essa proporção é dada pela razão entre o V_n e $(V_n + F_p)$.

$$ESP = \frac{V_n}{V_n + F_p} \quad (3.12)$$

- Curva ROC

Segundo Xu (2017), a curva ROC (Receiver Operating Characteristic) é uma ferramenta fundamental para avaliação de testes diagnósticos. É cada vez mais utilizado em muitos domínios, como os dados mineração, pontuação de crédito financeiro, previsão do tempo etc.

Segundo Pereira (2013), curva ROC é a representação gráfica da performance de um modelo de classificação, em que no eixo X se encontra a proporção dos falsos positivos (1-Especificidade) e no eixo Y a proporção de verdadeiros positivos (Sensibilidade). Neste sentido, cada ponto da curva representa um possível ponto de corte e a sua respectiva Sensibilidade e (1-Especificidade).

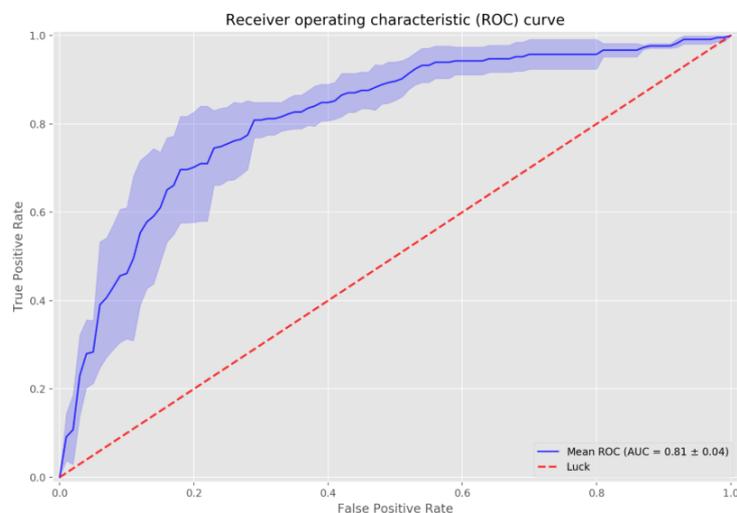


Figura 3.1: Representação gráfica de uma Curva ROC

A AUC (Área Sob a Curva) é uma métrica que resume a performance da curva ROC em um único valor. Ela varia de 0.5 a 1.0:

AUC = 0.5: O modelo não tem poder discriminativo, equivalente a uma escolha aleatória.

$0.5 < \text{AUC} < 0.7$: O modelo tem baixa discriminação.

$0.7 \leq \text{AUC} < 0.8$: O modelo tem discriminação aceitável.

$0.8 \leq \text{AUC} < 0.9$: O modelo tem boa discriminação.

$\text{AUC} \geq 0.9$: O modelo tem excelente discriminação.

- Medida Pseudo - R^2

De acordo com Marôco (2018), em modelos de regressão logística, medidas de qualidade do ajuste são funções dos resíduos definidos como a diferença entre o valor observado e o valor ajustado. Tratando-se de regressão logística não existe uma estatística que seja equivalente ao R^2 , utilizado em regressão linear. A denominação de pseudo R^2 deve-se ao facto de que eles se parecem com o R^2 do modelo de regressão linear, pois estão em uma escala similar, variando de 0 a 1, cujos valores mais próximos de 1 indicam melhor ajuste do modelo, mas apesar dessa similaridade não podem ser interpretados da mesma forma como se interpreta o R^2 . Além disso, diferentes definições de R^2 para um modelo logit (R^2_{Loggit}) pode ser calculado como:

$$R^2_{Loggit} = \frac{-2LL_{nulo} - (-2LL_{modelo})}{-2LL_{nulo}} \quad (3.13)$$

Exactamente como na contraparte da regressão múltipla, o valor R^2 logit varia de 0 a 1. À medida que o modelo proposto aumenta o ajuste, o $-2LL$ diminui. Um ajuste perfeito tem um valor de $-2LL$ igual a 0 e um (R^2_{Loggit}) de 1.

Duas outras medidas são semelhantes ao valor pseudo R^2 e são geralmente categorizadas também como medidas pseudo R^2 . A medida R^2 de Cox e Snell opera do mesmo modo, com valores maiores indicando maior ajuste do modelo. No entanto, esta medida é limitada no sentido de que não pode atingir o valor máximo de 1, de forma que Nagelkerke propôs uma modificação que tinha o domínio de 0 a 1. Essas duas medidas adicionais são interpretadas como refletindo a quantidade de variação explicada pelo modelo logístico, com 1 indicando ajuste perfeito.

Teste de Hosmer-Lemeshow

Hosmer e Lemeshow (2000), propuseram um teste para verificar a qualidade de ajuste do modelo de regressão logística, que consiste em agrupar os dados baseado nas probabilidades estimadas. O teste de Hosmer - Lemeshow, testa se o modelo obtido explica de forma adequada os dados observados. Estes dados são separados em g grupos de acordo com as probabilidades previstas.

Hosmer e Lemeshow (1980,1982), citados por Agranonik (2005), propõem que seja utilizado $g = 10$. Os grupos são criados de maneira que o primeiro tenha probabilidade predita entre 0.0 e 0.1, e o segundo, entre 0.1 e 0.2 e assim por diante até que o décimo grupo tenha valores de probabilidade predita entre 0.9 e 1.0. A seguir são calculados os valores esperados para cada grupo e comparados com os valores observados usando a estatística χ^2 de Pearson.

O teste avalia o modelo ajustado através das distâncias entre as probabilidades ajustadas e as probabilidades observadas.

Neste caso, a hipótese a testar é :

H_0 : O modelo se ajusta aos dados

H_1 : O modelo não se ajusta aos dados

A estatística de teste é dada pela seguinte expressão:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - E_k)^2}{(1 - E_k/n_k)} \quad (3.14)$$

Onde :

$o_k = \sum_{j=1}^{n_k} y_{kj}$ representa o número de casos registados no j -ésimo decil, e

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

O teste de Hosmer e Lemeshow segue, aproximadamente, uma distribuição qui-quadrado com $g-2$ g.l., quando o modelo está especificado correctamente.

Capítulo 4

Resultados e Discussão

4.1 Resultados da Análise Descritiva

4.1.1 Descrição das características das crianças da província de Nampula

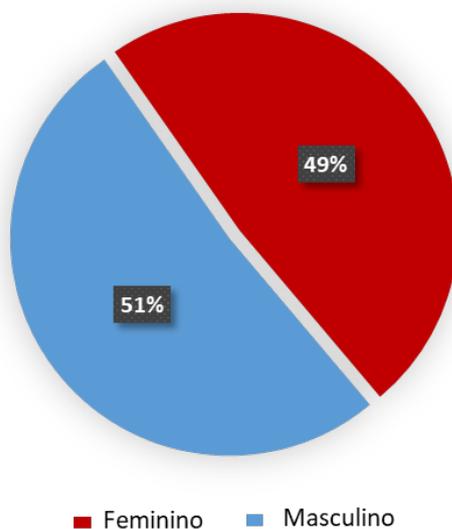


Figura 4.1: Distribuição das crianças segundo o género.

Segundo a figura 4.1, das crianças em estudo, 49% são do sexo feminino e 51% são do sexo Masculino, verifica-se assim uma maioria das crianças do sexo masculino.

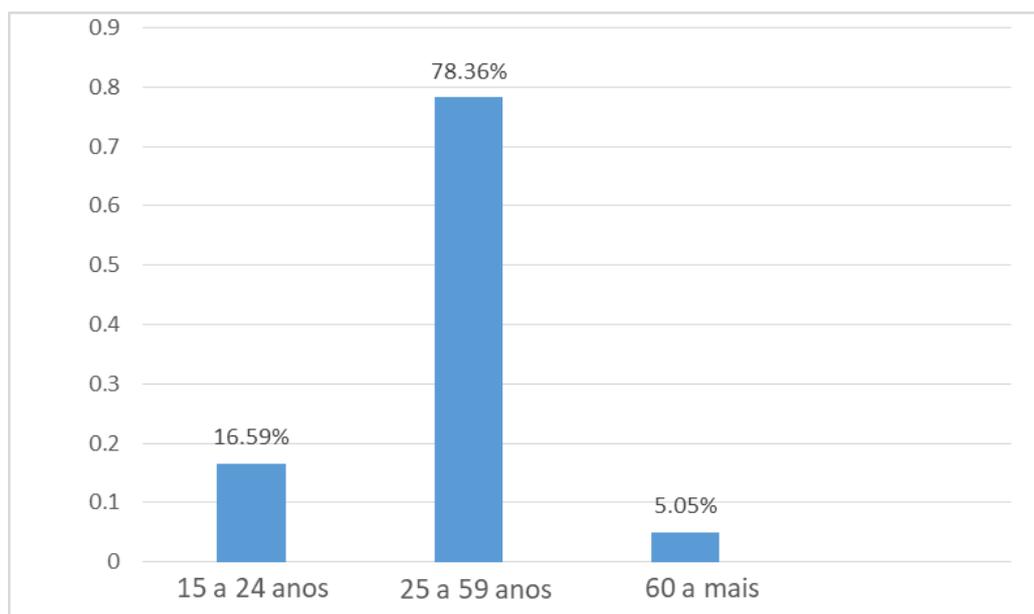


Figura 4.2: Distribuição dos chefes dos agregados segundo a idade

De acordo com a figura 4.2, nota-se que 78,36% dos chefes dos agregados tem idades compreendidas entre 25 e 59 anos, 16,50% tem idades compreendidas entre 15 a 24 anos e 5,05% dos chefes dos agregados tem 60 ou mais anos, sugerindo que a maior parte das crianças em estudo vivem em agregados chefeados por adultos.

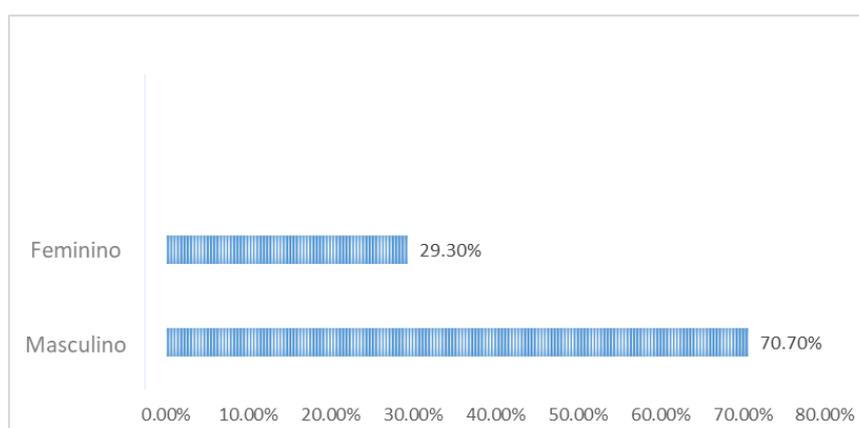


Figura 4.3: Distribuição dos chefes dos agregados segundo o gênero

De acordo com a figura 4.3, mais que a metade dos chefes do agregado são do sexo masculino, com uma proporção de 70,7% e 29,30% são do sexo feminino.

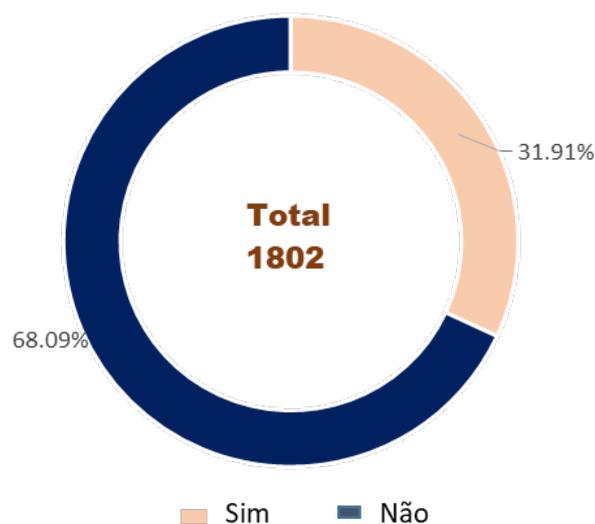


Figura 4.4: Distribuição dos agregados segundo o nível de acesso a informação

De acordo com a figura 4.4, pode-se observar que mais que a metade dos chefes dos agregados não têm acesso a meios de comunicação, ou seja 68,09% e somente 31,91% dos agregados tem acesso a pelo menos um meio de comunicação.

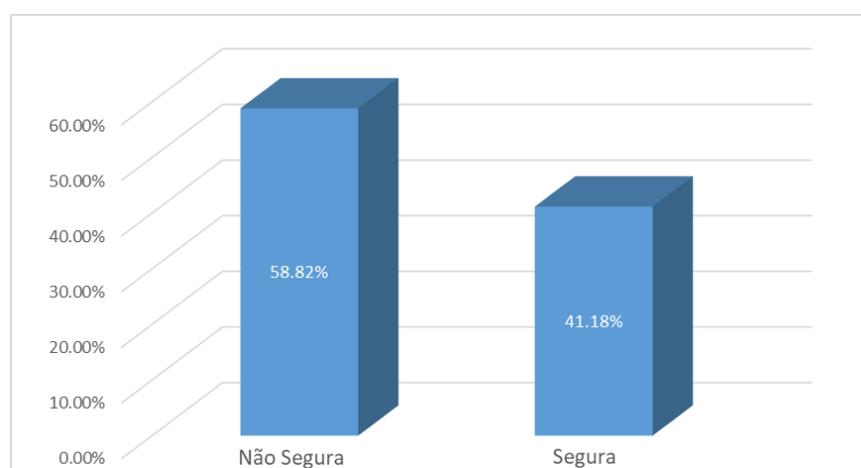


Figura 4.5: Distribuição dos agregados segundo a fonte de água

Segundo a figura 4.5, apenas 41.18 % dos agregados consomem água segura, a maior parte dos agregados têm consumido água não tratada, isto é 58.82% .

Tabela 4.1: Distribuição dos AF's segundo a fonte de renda

Fonte de Renda	Frequência	Percentagens
Produção, venda de culturas e animais	1153	63.98%
Auto emprego/ Trabalha por conta própria (processamento, comércio, serviços)	134	7.44%
Assistência alimentar	15	0.83%
Ganho na agricultura	166	9.21%
Ganho noutros negócios /biscate	168	9.32%
Pesca e aquacultura	25	1.39%
Remessas/ajuda	17	0.94%
Salário e pensões	92	5.11%
Sem fonte de rendimento	32	1.78%
Total Geral	1802	100.00%

Quanto a distribuição dos AF's segundo a fonte de renda, verifica-se que 63.98% dos chefes dos agregados têm como fonte de renda a produção, venda de culturas e animais, 9.32% têm com fonte de rendimento ganho noutros negócios(biscate), 9.21% ganho na agricultura, 7.44% trabalham por contam própria (auto emprego), 5.11% ganham salários e pensões, 1.78% não têm fonte de renda, 1.39% tem ganhos por meio da pesca e aquacultura, 0.94% recebem remessas /ajuda e 0.83% recebem assistência alimentar.

4.2 Resultados da Regressão Logística

Tabela 4.2: Estimativas dos coeficientes do modelo

Variáveis	Coef.	E.P	Wald	gl	P-valor	O.R	I.C-95%
Constant	71.366	4.113	301.100	1	0.000	–	–
Género da Criança							
Feminino	-2.176	0.208	109.400	1	0.000	0.113	[0.075; 0.169]
Masculino(<i>ref</i>)							
Idade da Criança	0.773	0.044	307.300	1	0.000	2.165	[1.994;2.370]
Peso da Criança	0.019	0.082	0.052	1	0.820	1.019	[0.868;1.196]
Altura da Criança	-1.132	0.067	253.000	1	0.000	0.322	[0.281;0.365]
Ocorrência de Desnutrição Aguda							
Sim	0.217	0.550	0.160	1	0.693	1.243	[0.415;3.601]
Não (<i>ref</i>)							
Ocorrência de Diarreia							
Sim	0.019	0.321	0.004	1	0.953	1.019	[0.543; 1.918]
Não (<i>ref</i>)							
Ocorrência de Diarreia Aquosa							
Sim	-0.346	0.344	1.000	1	0.314	0.707	[0.359; 1.386]
Não (<i>ref</i>)							
Ocorrência de Febre							
Sim	0.415	0.214	3.800	1	0.053	1.514	[0.997;2.308]
Não (<i>ref</i>)							
Ocorrência de Tosse							
Sim	0.028	0.211	0.017	1	0.895	1.028	[0.679;1.557]
Não(<i>ref</i>)							
Género do Chefe do AF							
Feminino	-0.276	0.216	1.600	1	0.201	0.759	[0.496;1.158]
Masculino(<i>ref</i>)							
Idade do Chefe do AF							
15 à 24 anos(<i>ref</i>)							
25 à 59 anos	0.090	0.252	0.130	1	0.720	1.094	[0.668;1.795]
60 ou mais	-0.939	0.526	3.200	1	0.075	0.391	[0.137;1.083]
Tem acesso a algum meio de Comunicação							
Sim	0.163	0.216	0.570	1	0.450	1.177	[0.771 1.797]
Não (<i>ref</i>)							
Tamnho do AF							
1 à 5 membros (<i>ref</i>)							
Mais de 5 Membros	0.466	0.223	4.400	1	0.036	1.593	[1.032;2.471]
Número de Crianças Menores de 5 anos no AF	-0.339	0.137	6.100	1	0.014	0.713	[0.544;0.933]
Fonte de Água no AF							
Segura	-0.111	0.192	0.340	1	0.562	0.895	[0.614;1.303]
Não Segura (<i>ref</i>)							
Fonte de Renda no AF							
Auto emprego/ Trabalha por conta própria	-0.422	0.353	1.400	1	0.232	0.656	[0.327;1.308]
Assistência alimentar	0.328	1.163	0.080	1	0.778	1.388	[0.133;15.078]
Ganho ganho na agricultura	-0.237	0.305	0.600	1	0.438	0.789	[0.433;1.433]
Ganho ganho/biscate noutros negócios	0.319	0.309	1.100	1	0.303	1.375	[0.751;2.527]
Pesca e aquacultura	1.980	0.904	4.800	1	0.029	7.245	[1.331;45.283]
Remessas/ajuda	1.301	1.208	1.200	1	0.282	3.672	[0.319;36.506]
Salário e pensões	-0.735	0.449	2.700	1	0.102	0.479	[0.197;1.150]
Sem fonte de rendimento	0.212	0.883	0.057	1	0.811	1.236	[0.203;6.564]
Produção, venda de culturas e animais (<i>ref</i>)							

A tabela 4.2, apresenta as estimativas dos coeficientes do modelo e os respectivos testes. De acordo com o teste de Wald associado a estatística χ^2 apenas as variáveis que apresentarem um *p-value* inferior ao nível de significância de 5%, serão mantidas no modelo. Portanto pode-se notar que as variáveis: Género da criança, Idade da criança, Altura da criança, Tamanho do AF e Fonte de renda a base de Pesca e Aquacultura são significativas.

O teste da razão de verossimilhança testa a hipótese de que o modelo com pelo menos um preditor é igual ao modelo sem nenhum preditor, ou seja, se existe pelo menos algum parâmetro significativo no modelo. Segundo a Tabela 4.3 pelo menos um dos parâmetros do modelo logístico final é estatisticamente significativo, uma vez que p-valor = 2.2×10^{-16} associado a estatística de teste da razão de verossimilhança $\chi^2 = 1713.3$ é inferior a um nível de significância de 5%.

Tabela 4.3: Teste de Razão de Verossimilhança

Qui-quadrado	G.L	P- Valor
1713.3	8	2.2×10^{-16}

De acordo com os resultados da Tabela 4.4, a medida de Nagelkerke R², mostra que cerca de 81.9% da variação da desnutrição crônica das crianças pode ser explicada pelo modelo de regressão logística, ou seja as variáveis explicativas do modelo têm impacto significativo na ocorrência da desnutrição crônica em crianças.

Tabela 4.4: Medidas de avaliação do ajuste geral do modelo

Modelo	-2 Log likelihood	Cox e Snell R Square	Nagelkerke R Square
1	-856.65	0.614	0.819

O teste de Hosmer e Lemeshow, avalia a qualidade do ajuste do modelo aos dados. De acordo com os resultados da tabela 4.5, rejeita-se a hipótese de que o modelo se ajuste bem aos dados, a um nível de significância de 0.05. No entanto importa referir que este teste é usado para avaliar a significância global do modelo e está sujeito a muitas restrições, sendo sensível a amostras de tamanho muito grande, agrupamentos dos dados e a distribuição desigual de cada grupo. De um modo geral, considerando todos testes realizados e fixando o nível de significância de 5% o modelo se ajusta bem aos dados.

Tabela 4.5: Teste de Hosmer e Lemeshow

Modelo	Qui-quadrado	G1	P-value
modelo final	1802	8	2.2×10^{-16}

A partir da tabela 4.6, é possível perceber que o modelo classificou correctamente 862 dos 959 casos de crianças com desnutrição crónica, ou seja, uma sensibilidade de 89.9%. Além disso, o modelo classificou correctamente 745 dos 843 casos de crianças que não apresentam desnutrição, ou seja, o modelo retorna uma especificidade de 88.4%. Entretanto, 97 observações foram classificadas pelo modelo como sendo de crianças que não apresentam desnutrição crónica quando na verdade eram de crianças com desnutrição crónica. Por fim, 98 casos foram classificados incorrectamente como sendo de crianças com desnutrição crónica quando na verdade eram de crianças que não têm. De modo geral modelo classifica correctamente 89.2% dos casos.

Tabela 4.6: Matriz de classificação do modelo

		Previsto		% Correcta
		Desnutrição Crónica		
Desnutrição Crónica	Sim(1)	862	97	89.9
	Não(0)	98	745	88.4
Acurácia				89.2

De acordo com a Tabela 4.7, o terceiro modelo é o que descreve melhor a prevalência de desnutrição em crianças menores de 5 anos na província de Nampula, uma vez que apresenta o menor valor para a estatística BIC em relação aos restantes modelos.

Tabela 4.7: Critério de selecção do modelo

Modelos	g.l	BIC
Modelo Completo	25	952.5503
Modelo2	10	848.7661
Modelo Final	9	844.7916

A Tabela 4.8 apresenta o valor da área sob a curva (AUC). A área sob a curva (AUC) foi calculada como 0.9705, indicando uma excelente capacidade de discriminação do modelo.

Tabela 4.8: Área sob a curva (AUC) da curva ROC

Medida	Valor
Área sob a curva (AUC)	0.9705

A figura 4.6, ilustra a Curva de ROC, onde a curvatura está próxima do canto superior esquerdo indicando uma alta sensibilidade e baixa taxa de falsos positivos, sugerindo que o

modelo tem uma boa capacidade de discriminar entre as classes.

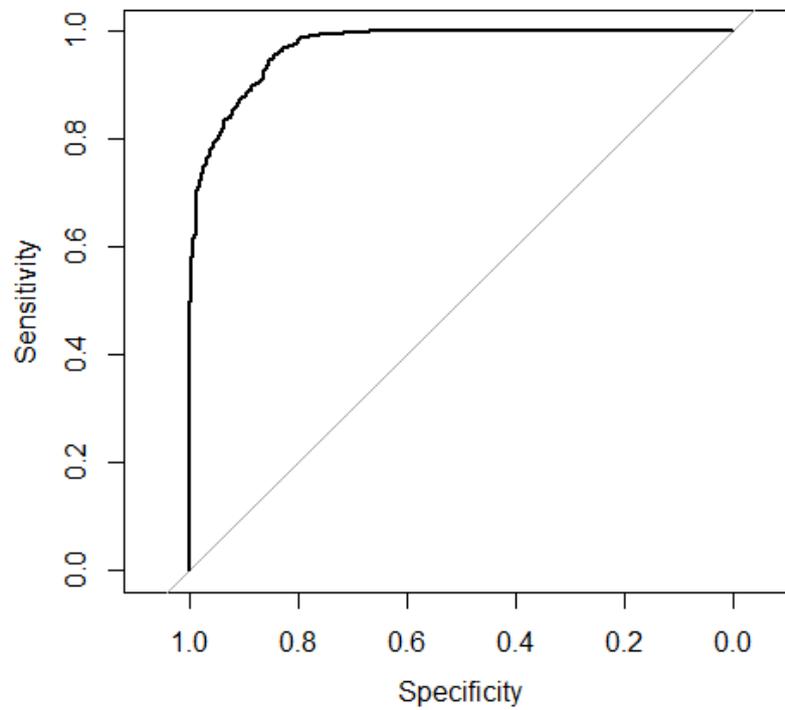


Figura 4.6: Curva ROC do modelo de regressão logística

Tabela 4.9: Modelo de regressão logístico final

Variável	Coefficiente	EP	Wald	gl	p-value	RC	I.C-95%
Constante	70.506	3.955	317.8	1	0.000	–	–
Género da Criança	-2.141	0.204	110	1	0.000	0.118	[0.078;0.174]
Idade da Criança	0.765	0.043	317.7	1	0.000	2.149	[1.982;2.348]
Altura da Criança	-1.114	0.063	255.7	1	0.000	0.328	[0.289;0.369]
Idade do Chefe do AF(60+)	-1.027	0.473	4.7	1	0.030	0.358	[0.139;0.892]
Ocorrência de Febre	0.361	0.183	3.9	1	0.049	1.435	[1.003;2.060]
Tamanho do AF	0.462	0.206	5.1	1	0.025	1.587	[1.063;2.382]
Número de crianças menores de 5 anos no AF	-0.334	0.135	6.1	1	0.013	0.716	[0.549;0.933]
Ganho com Pesca e Aquicultura	1.953	0.889	4.8	1	0.028	7.050	[1.353;43.688]

A partir da tabela 4.9 pelo teste de significância de Wald pode-se concluir que, as variáveis: Idade da criança, Género da criança, altura da Criança, a ocorrência de febres, Número de crianças menores de 5 anos no AF, Tamanho do Agregado, Idade do chefe do Agregado e Fonte de Renda: Ganhos com Pesca e Aquicultura, revelaram -se significativas a um nível de significância de 5%.

Em termos de razão de chances, no modelo logístico da tabela 4.9, conclue-se que aumentando em uma unidade a idade das crianças, as chances de criança estar desnutrida aumenta em 114.9% (RC=2.149-1), as crianças que apresentaram febre têm 43.5% mais chances de ter desnutrição crónica que as crianças que não apresentaram febres (RC=1.435-1), agregados com mais de 5 membros têm 58.7% mais chance de ter desnutrição crónica em relação aos agregados com no máximo 5 membros (RC=1.587-1), agregados com fonte de renda a base da pesca e aquicultura têm 7.050 vezes chances de ter desnutrição crónica em relação aos agregados com fonte de renda à base de produção, venda de culturas e animais.

Quanto ao género da criança, crianças do sexo feminino têm 11.8% menos chances de ter desnutrição em relação a crianças do sexo masculino(RC=0.118), o aumento em uma unidade na altura da criança reduz em 32.8% a chance de uma criança ter desnutrição crónica (RC=0.328), Para a variável idade do chefe do agregado, observou-se que, quando os chefes dos agregados têm 60 anos ou mais, há uma redução de 64.2% nas chances de uma criança apresentar desnutrição crônica em comparação com agregados cujo chefe tem idades entre 15 e 24 anos (RC = 0.358).”

Aplicação do modelo logístico final para previsão

$$\pi_i(x_i) = \frac{e^{70.506 + \sum_{i=1}^8 \beta_1 x_i}}{1 + e^{70.506 + \sum_{i=1}^8 \beta_1 x_i}} \tag{4.1}$$

Tabela 4.10: Estimação das probabilidades pelo modelo final

π_i	X_1	X_2	X_4	X_5	X_7	X_{13}	X_{14}	X_{15}	Probabilidade	Y_i
$\pi_{(3)}$	0	44	99.0	0	0	1	2	0	0.0018	0
$\pi_{(100)}$	0	27	77.6	0	1	0	1	0	0.9913	1
$\pi_{(500)}$	1	48	102.5	0	1	1	2	0	0.0001	0
$\pi_{(720)}$	0	41	83.8	0	1	1	1	0	0.9999	1
$\pi_{(1057)}$	0	47	86.1	0	0	1	1	0	0.9999	1

O modelo estimou correctamente as probabilidades. As observações 3 e 500 correspondem a crianças sem desnutrição crónica e as probabilidades estimadas estão abaixo de 0.5. Na estimação das probabilidades das crianças terem desnutrição crónica, o modelo o fez correctamente, pois as probabilidades são acima de 0.5, isso pode-se verificar também nas observações 100 , 720 e 1057.

Para a observação 720, os valores das variáveis (X) são:

- $X_1 = 0, X_2 = 41, X_4 = 83.8, X_5 = 0,$
- $X_7 = 1, X_{13} = 1, X_{14} = 1, X_{15} = 0.$

Os coeficientes (β) associados são:

- $\beta_1 = -2.141, \beta_2 = 0.765, \beta_4 = -1.114, \beta_5 = -1.027,$
- $\beta_7 = 0.361, \beta_{13} = 0.462, \beta_{14} = -0.334, \beta_{15} = 1.953.$

$$\pi_{(720)} = \frac{e^{70.506+(0)(-2.141)+(41)(0.765)+(83.8)(-1.114)+(0)(-1.027)+(1)(0.361)+(1)(0.462)+(1)(-0.334)+(0)(1.953)}}{1 + e^{70.506+(0)(-2.141)+(41)(0.765)+(83.8)(-1.114)+(0)(-1.027)+(1)(0.361)+(1)(0.462)+(1)(-0.334)+(0)(1.953)}}$$

$$\pi_{(720)} = \frac{e^{70.506+31.365-93.401+0.361+0.462-0.334}}{1 + e^{70.506+31.365-93.401+0.361+0.462-0.334}}$$

$$\pi_{(720)} = \frac{e^{7.54033}}{1 + e^{7.54033}}$$

$$\pi_{(720)} \approx 0.99947$$

Tendo substituído os valores na equação 4.1, tem-se que a probabilidade estimada $\pi_{(720)}$ para a observação 720 é **99.95%**, indicando uma alta chance de a criança estar em situação de desnutrição crónica. Com o valor observado de $Y_{(720)} = 1$, confirmando a precisão do modelo.

4.2.1 Discussão dos resultados

O presente trabalho tem como principal objetivo analisar os fatores que determinam a prevalência da desnutrição crônica em crianças menores de 5 anos, na província de Nampula. Para isso, foram estimados modelos estatísticos apropriados para identificar e compreender a influência desses fatores.

Num universo de 16 variáveis, 8 são as que mostraram-se como os factores que explicam a ocorrência da desnutrição crónica naquela região do País.

O género da criança, fonte de renda, incidência de febre nas crianças, tamanho do agregado, idade da criança, idade do chefe do agregado, são mencionados por Massango (2018), como factores significativos para explicar a desnutrição crónica em crianças menores de 5 anos, o que vai de acordo com os resultados obtidos neste trabalho.

O nível de acesso a informação e fonte de água para o consumo no agregado, são identificados por Cesse et al (2022), como factores relacionados a desnutrição crónica, o que não se verificou no presente trabalho.

Agregados liderados por mulheres, sendo os chefes dos AF's idosos, crianças com infecções respiratórias, como desnutrição aguda, tosse e diarreias estão sujeitas a ocorrência de desnutrição crónica, de acordo com o SETSAN(2015), o que não se verifica neste estudo, de acordo com os resultados deste trabalho, agregados cujos chefes são adultos têm mais probabilidade de ter uma criança com desnutrição crónica, e o género do chefe agregado não tem poder explicativo sobre a desnutrição crónica em crianças menores de 5 anos na província de Nampula.

Em relação ao peso da criança, autores como Galvão(2014) e Barbosa (2008), consideram como um dos determinates primários da desnutrição infantil, facto que é contrariado pelos resultados desse estudo.

Barbosa (2008), considera que a renda familiar menor que 0.25 salários mínimos, é um factor de risco significativamente associado ao déficit nutricional, esperava-se, neste estudo que as outras categorias como: assistência alimentar, remessas e ajuda e a categoria sem fonte de rendimento também tivessem valor significativo para explicar a desnutrição crónica, facto que não se verificou.

Em estudos sobre desnutrição infantil, autores como Massango (2018), De Cerqueira et al (2021) e Arce et al (2018), não consideram o número de crianças menores de 5 anos no AF, como um possível factor para explicar o fenómeno, no entanto neste estudo a variável revela-

se como um dos factores relacionados à desnutrição crónica em crianças menores de 5 anos na província de Nampula.

Reis (2021), aponta que um modelo pode ser estatisticamente significativo e isso não corresponder a realidade em estudo, sugerindo a matriz de confusão para avaliar a eficiência classificativa do modelo. A medida usada para avaliar o poder global de classificação do modelo, é o desempenho com taxas altas indicando um potencial preditivo muito bom do modelo. Neste estudo obteve-se uma taxa de 89.2%, assim pode-se assumir que os resultados obtidos condizem com a realidade do estudo.

Capítulo 5

Conclusões e Recomendações

5.1 Conclusões

Com base nos resultados do trabalho e conforme os objetivos inicialmente estabelecidos, foram encontrados os seguintes resultados:

- Das 1802 crianças analisadas, 49% são do sexo feminino e 51% são sexo masculino. Destas, cerca de 53.2% têm desnutrição crônica, e 46.8% não têm desnutrição crônica, isto é, mais da metade das crianças em estudo sofrem de desnutrição crônica. São crianças cujos chefes dos agregados são do sexo masculino, na sua maioria (70.7%), de idade adulta (21 à 59 anos), onde o nível de acesso a informação é muito baixo (31.91%), a fonte de água nesses agregados não é segura, ou seja, são agregados que consomem água das chuvas, de poços não protegidos ou de lagoas e lagos. A principal fonte de renda nesses agregados é a produção, venda de culturas e animais.
- As variáveis: gênero da criança, idade da criança, altura criança, ocorrência de febres, crianças menores de 5 anos no agregado, tamanho do agregado, idade do chefe do agregado e fonte de renda (ganhos com pesca e aquacultura), mostraram-se significativas para explicar a prevalência da desnutrição crônica nas crianças.
- A probabilidade de uma criança estar em situação de desnutrição crônica é de aproximadamente 99.9%.

5.2 Recomendações

- Recomenda-se que as organizações invistam na capacitação técnica de suas equipes em métodos estatísticos avançados, promovendo treinamentos e iniciativas de formação contínua. Além disso, é essencial que colaborem com especialistas da área, já presentes em algumas dessas organizações, bem como com instituições acadêmicas, para

desenvolver e implementar análises estatísticas robustas que atendam às necessidades locais.

- Expansão dos meios de comunicação para regiões com altos índices nos indicadores de saúde infantil;
- A realização de mais estudos semelhantes, dada a alta taxa de prevalência da doença, para melhor se perceber como estes factores podem influenciar à desnutrição crónica;

5.3 Limitações

- Carência de estudos regionais que abordam sobre a desnutrição crónica, muitos estudos abordam apenas uma forma de desnutrição (desnutrição aguda).
- Ausência excessiva de informação em variáveis importantes como: variáveis relativas à mãe da criança (situação nutricional, nível de escolaridade, peso ,altura,etc), variável local de residência, saneamento do meio, entre outras, levando assim a exclusão de algumas delas.

Referências

- [1] Abbas, M. e Mosca, J. (2021). IAI 2020: *Produção e produtividade agrícola*, Moçambique.
- [2] Acorsi, C. R. L., Guedes, T. A., Janeiro, V. e Martins, A. B. T. (2005). *Estatística Descritiva*, Universidade Estadual de Maribga .
- [3] Agranonik, M. (2005). *Técnicas de diagnósticas aplicadas ao modelo de regressão logística*, Porto Alegre.
- [4] Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. New York : Wiley.
- [5] Al-rubaie, K. S. F., Godefroid, L. B. e Lopes, J. A. M. (2006). Modeling of fatigue crack growth rate in Inconel alloy 600, *International Journal Of Fatigue. Science Direct* p. 931-940.set.
- [6] Amaral , G.D ., Reis , E. A . e Silva , V.L . (2009). *Análise de regressão linear no pacote R*. Universidade Federal de Minas Gerais, Brasil.
- [7] Anderson, R. E., Babin, B. J., Black, W. C., Junior , J. F. H. e Tatham, R. L. (2009). *Análise multivariada de dados* , Sexta edição Brasil.
- [8] Arce, C. B., Da Silva, M. V., Cruz, J. R., Dos Santos, P. F., Ferreira, A. T. S. e Viana, T. C. T. (2018). Factores determinantes da desnutrição infantil em crianças de 0 a 5 anos, cadastradas no programa de crescimento e desenvolvimento infantil na amazônia Legal. *Brazilian Journal of Surgery and Clinical Research*. Vol.23, n.2, pp.58-62.
- [9] Balegamire, J. B., Lima, M. C. R. A. A., Mitano , F., Palha, P. F. e Ventura , C. A. A. (2016). Direito à Saúde: incongruências entre o arcabouço Jurídico e o sistema de saúde. *Rev.Latino-AM. Enfermagem*. 24:2679.
- [10] Barbosa, M. D. DS. (2008). *Risco nutricional e desnutrição: prevalência e factores associados em crianças menores de cinco anos*. Recife.
- [11] Barreto, S. M., COSTA, A. B. P., Giatti, L., Herbert, S., KhanaferII, N. e Mendes, L. L. (2021). *Desnutrição e factores associados em escolas primárias em Lokossa, Benin: um estudo transversal*. Minas Gerais.

- [12] Barros, F. C., Olinto, M. T. A., Victora, C. G. e Tomasi, E. (1993). *Determinantes da desnutrição infantil em uma população de baixa renda: um modelo de análise hierarquizado*, Rio de Janeiro.
- [13] Brandão, N. G. e Pereira, N. D. S. (2022). *As atitudes e comportamentos decorrentes da acção das campanhas de desnutrição em crianças, realizadas pela ONG SD junto da população do Distrito de Monamapo-Nampula*.
- [14] Bittencourt, H. R. e Clarke, R. T. (2001). *Um classificador baseado na discriminação logística: vantagens e desvantagens*, Brasil.
- [15] Burnham, K. e Anderson, D. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.p
- [16] Bolfarine, H. e Sandoval, M. C. (2000). *Introdução à inferência estatística*. São Paulo: Sociedade Brasileira de Matemática.
- [17] Cabral, C. I. S. (2013). *Aplicação do modelo de regressão logística num estudo de Mercado*, Lisboa.
- [18] Cesse, E. A. P., Culpa, M. J., Silva, D. O. e Souza, N. P. (2022). *Desnutrição infantil e seus determinantes imediatos, subjacentes e básicos em Moçambique: uma análise temporal de 1997 a 2015*.Brasil. p.77-92.
- [19] Chilombe, N (2024). *Mapa de Moçambique (mapa)*. Produzido para o trabalho de conclusão de curso, Universidade Eduardo Mondlane.
- [20] Chilombe, N (2024). *Mapa da província de Nampula (mapa)*. Produzido para o trabalho de conclusão de curso, Universidade Eduardo Mondlane.
- [21] Da Barca, A. e Dos Santos, T. (1992). *Geografia de Moçambique*.Vol.I.
- [22] Da Rocha, A. L. M. M. (2014). *Regressão logística multinível: uma aplicação de Modelos Lineares Generalizados Mistos*.
- [23] Da Silva, J. P., Silva, V. R. e Kruguer, G. S. (2017). *Desnutrição infantil: reflexão histórica focada na segurança alimentar e nas políticas sociais e de saúde*, Universidade Federal de Maranhão/São Luís-Brasil.
- [24] De Cerqueira, C. T. de., Macedo, A. C. de O., Macêdo, D. J. do N., Menezes de Lima, I. R., Santana, L. Q. de M. e Viana, L. de A. (2021). Fatores associados à desnutrição crônica de crianças e adolescentes admitidas em um hospital de referência do Nordeste. *Brazilian Journal of Development*, V.7(10), p. 95142-95155.
- [25] De Araújo, M. M. (2005). *Cidade de Nampula: A Rainha do Norte*, Moçambique.

- [26] Fernandes, M. e Valois, J. (2017). *Introdução à nutrição: histórico da nutrição, campos de atuação do técnico e classificação dos alimentos*.
- [27] Ferreira, R. R. (2021). *Regressão logística geograficamente ponderada na análise de risco de crédito*, Porto Alegre.
- [28] Figueira, C. (2006). *Modelos de Regressão Logística*, Universidade Federal rio grande sul
- [29] FAO. (2018). *A caminho da fome Zero: 1945-2030. relatório sobre nutrição na infância*.
- [30] Galvão, M. A. M. (2014). *Factores de desnutrição infantil em crianças de 0 a 05 anos no Cernutri-Boa Vista/Roraima*. Florianópolis.
- [31] Gonçalves, N. (2023). *Regressão logística: Uma Visão Geral*. Disponível em: <https://medium.com/@natalia.gcv.s/regress>. Acesso em 27 de Setembro de 2024
- [32] Guedes, T. A., Janeiro, V., Lonardan, C. R. e Martins, A. B. T.(2005). *Estatística Descritiva*. Brasil.
- [33] Hosmer, D. W e Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd. Canada.
- [34] Hosmer, D. W e Lemeshow, S. (1982). *American Journal of Epidemiology*, Volume 115, Issue 1, January 1982, Pages 92–106, <https://doi.org/10.1093/oxfordjournals.aje.a113284>
- [35] Humbwavali, J. B. (2016). *Situação nutricional e suas tendências em mulheres e crianças da África subsariana e factores associados à desnutrição, em uma população infantil de Luanda-Angola* .Porto Alegre.
- [36] Ibramugy, F. (2013). *Subnutrição aumenta em Nampula*. Disponível em: <https://www.voaportugues.com/amp/subnutricao-aumenta-em-Nampula/1697486.html> . Acesso em 2 de maio de 2024.
- [37] INE. (2017). *IV Recenseamento Geral da População e Habitação: Indicadores Sócio-Demográficos de Moçambique*.
- [38] INE , MISAU. e ICFI. (2013). *Inquérito Demográfico e de Saúde*, Moçambique.
- [39] INE, MISAU. e ICFI. (2015). *Inquérito de indicadores de imunização, Malária e HIV/SIDA em Moçambique*.
- [40] IOF. (2019). *Estado nutricional das crianças menores de cinco anos de Idade, Moçambique*.

- [41] Júnior, A. C. A. (2014). *A desnutrição infantil: trabalho Missionário na área de saúde em Moçambique*.
- [42] Marcon, J. N. (2017). *Brasil-Africa subsariana e cooperação para Segurança Alimentar: o caso do Quênia*.
- [43] Marôco, J. (2007). *Análise estatística Com Utilização do SPSS, 3ª Edição*. Lisboa: Edição Silabo.
- [44] Marôco, J. (2018). *Análise Estatística com o SPSS Statistics, 7ª Edição*. Portugal
- [45] Martins, S. J. (2012). *Modelo de avaliação de risco em Acidentes no ramo automóvel*, Lisboa, Portugal.
- [46] Massango, E. T. (2018). *Análise de factores associados a desnutrição infantil em Moçambique*.
- [47] McCullagh, P e Nelder , J.A. (1989). *Generalized Linear Models*. 2nd ed. York: Chapman Hall/CRC.
- [48] Mesquita, P. S. B. (2014). *Um modelo de regressão logística para avaliação dos programas de Pós - graduação no Brasil*.
- [49] Monteiro, C. A. (2003). *A dimensão da pobreza, da desnutrição e da fome no Brasil, Estudos avançados*.
- [50] Moura, G. M. (2018). *Regressão logística aplicada a análise de risco de crédito*, Brasil.
- [51] Mulenga, A. (2018). *Introdução à estatística* . Editora impressa Universitária, Maputo.
- [52] Nedel, V. (2020). *África: Aspectos Gerais*, Brasil.
- [53] OMS., WFT. e UNICEF. (2021). *Declaração Conjunta: alimentação infantil no contexto de emergência*, Moçambique.
- [54] Perlito, D. (2014). *Avaliação do estado de nutrição e dos hábitos alimentares no 1º ano de vida de crianças dos 0 aos 24 meses da província de Nampula - Moçambique*, Porto.
- [55] Pereira, M. B. (2013). *Estimação da Sensibilidade, da Especificidade e da curva ROC*,
- [56] Pessoa, J. (2013). *Uso de regressão logística para Identificar os factores de risco associados à ocorrência de anomalias congénitas em recém-nascidos*, Brasil.
- [57] ROSC. (2014). *Reflectindo sobre a desnutrição crónica: os desafios da construção de um capital humano sustentável em Moçambique*.

- [58] Silva, R. P. (2012). *Desnutrição infantil: um problema a ser enfrentado*. Corinto-Minas Gerais, Brasil.
- [59] SADC. (2022). *Relatório Síntese Sobre a Situação da Segurança alimentar e Nutricional e Vulnerabilidade na África Austral*.
- [60] SETSAN. (2014). *Relatório do estudo de base de Segurança Alimentar e Nutricional de 2013 em Moçambique*.
- [61] SETSAN (2015). *Relatório da Avaliação da Situação de Segurança Alimentar e Nutricional*. Maputo.
- [62] Schwarz, G. (1978). *Estimating the Dimension of a Model*. Annals of Statistics, 6, 461-464. <http://dx.doi.org/10.1214/aos/1176344136>
- [63] USAID. (2023). *USAID Advancing Nutrition Mozambique Final Report*.
- [64] UNICEF. (2023). *Um apelo a acção: necessidade urgente de uma acção coordenada para a nutrição em Moçambique*.
- [65] UNICEF. (2022). *Principais Indicadores Sociodemográficos da Província de Nampula, Moçambique*.
- [66] UNICEF. (2021). *Alimentação fadada ao fracasso: a crise na alimentação das crianças nos primeiros anos de vida*.
- [67] UNICEF. (2014). *Análise da situação das crianças em Moçambique*.
- [68] Valle, M. E. (2022). *Aprendizado de máquinas: medidas de desempenho em problemas de classificação*, Brasil.
- [69] Vaz, F. E. (2020). *Aplicação da regressão logística no estudo de Câncer da Mama*.
- [70] Xu, J. (2017). *ROC (Receiver Operating Characteristic) Curve Analysis*,
- [71] World Bank. (2020). *Agricultura Nutricionalmente Inteligente em Moçambique*. Disponível em : <https://documents1.worldbank.org/curated/en/119011>. Acesso em 10 de Abril de 2024.
- [72] WAHO, APHHC. (2020). *Desigualdades na cobertura e no estado nutricional entre mulheres, crianças e adolescentes na África Ocidental : uma análise Multipaíses*.
- [73] Zano, F. C. (2017). *Segurança Alimentar Urbana em Moçambique*. V.24 , n.2 , p.180-181 , jul./dez.2017.

ANEXOS

Tabela 5.1: Parâmetros do segundo modelo de regressão logístico

Variável	Coefficiente	EP	Wald	gl	p-value	RC	I.C-95%
Constante	70.742	3.974	316.900	1	0.000	–	–
Gênero da Criança	-2.166	0.206	111.100	1	0.000	0.115	[0.076;0.170]
Idade da Criança	0.766	0.043	313.500	1	0.000	2.151	[1.984;2.351]
Altura da Criança	-1.117	0.063	316.600	1	0.000	0.327	[0.288;0.368]
Idade do Chefe do AF(60+)	-1.029	0.473	4.700	1	0.030	0.358	[0.139;0.891]
Ocorrência de Febre	0.346	0.184	3.500	1	0.060	1.414	[0.987;2.032]
Tamanho do AF	0.482	0.206	5.500	1	0.020	1.620	[1.083;2.434]
Número de crianças menores de 5 anos no AF	-0.335	0.135	6.200	1	0.013	0.716	[0.549;0.932]
Ganho com Pesca e Aquacultura	1.925	0.890	4.700	1	0.031	6.855	[1.311;42.521]
Salários e Pensões	-0.809	0.435	3.500	1	0.063	0.445	[0.188;1.036]

Tabela 5.2: Critério de selecção do modelo

Modelos	g.l	AIC	Deviance
Modelo Completo	25	815.134	765.134
Modelo2	10	793.7995	773.800
Modelo Final	9	795.322	777.322

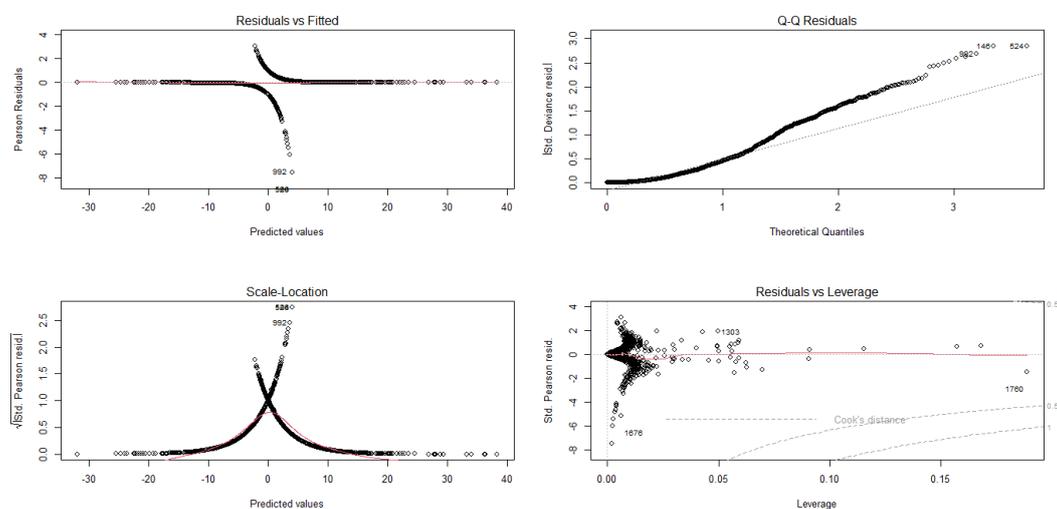


Figura 5.1: Resíduos do modelo logístico final