

MODELAÇÃO DE RESULTADOS EM SAÚDE USANDO DADOS DE INQUÉRITOS, AJUSTANDO PARA O DESENHO AMOSTRAL

SHEYLA RATAN RODRIGUES CASSY

Licenciada em Estatística

DOUTORAMENTO EM ESTATÍSTICA E GESTÃO DO RISCO

Universidade NOVA de Lisboa
Maio, 2022

MODELAÇÃO DE RESULTADOS EM SAÚDE USANDO DADOS DE INQUÉRITOS, AJUSTANDO PARA O DESENHO AMOSTRAL

SHEYLA RATAN RODRIGUES CASSY

Licenciada em Estatística

Orientadora: Maria do Rosário Oliveira Martins

Professora Catedrática, Instituto de Higiene e Medicina Tropical da Universidade NOVA de Lisboa

Coorientadores: Filipe José Gonçalves Pereira Marques

Professor Associado, Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa

Pedro Luis do Nascimento Silva

Professor Titular, Instituto Brasileiro de Geografia e Estatística

Júri

Presidente: Doutor Carlos Manuel Agra Coelho

Professor Catedrático da Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa

Arguentes: Doutora Joana de Lima Bastos

Professora Auxiliar Convidada da Faculdade de Medicina da Universidade de Coimbra

Doutora Célia Maria Pinto Nunes

Professora Auxiliar e Vice-Presidente da Faculdade de Ciências da Universidade da Beira Interior

Co-orientador: Doutor Pedro Luís do Nascimento Silva

Investigador Titular do Instituto Brasileiro de Geografia e Estatística, Escola Nacional de Ciências Estatísticas, Brasil

Vogais: Doutor João Tiago Praça Nunes Mexia

Professor Catedrático Jubilado da Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa

Doutor Samuel Osmonde Makande Manda

Professor e Diretor do Departamento de Estatística da Universidade de Pretoria, África do Sul

Doutora Sara Alexandra Fonseca Marques Simões Dias

Professora Auxiliar Convidada da Faculdade de Ciências Médicas da Universidade NOVA Lisboa

Modelação de resultados em saúde usando dados de inquéritos, ajustando para o desenho amostral

Copyright © Sheyla Ratan Rodrigues Cassy, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Ao meu filho, Tiago Luan.
Aos meus pais Bhangy Cassy e Alexandra Rodrigues, e a minha
tia-mãe Paula.
Ao meu querido Tio Jorge (in memoria) que não está mais entre
nós, mas continua a ser a minha força e alento. A sua lembrança
para além de me inspirar, faz-me persistir.
À minha querida Tia Mena (in memoria) que me apoiou
fisicamente no início deste percurso, tendo até viajado para
Portugal para acalantar os meus sonhos e objetivos de vida.

AGRADECIMENTOS

A vida é um conjunto de obstáculos e superações que nos fazem crescer e evoluir! Muitas foram as pessoas que me ajudaram a vencer os obstáculos com que me deparei nesta caminhada e a superar várias dificuldades.

Cumpra-me agora demonstrar o meu eterno apreço a quem foi determinante para que pudesse cumprir esta missão.

À Professora Maria do Rosário Martins, minha orientadora, a quem agradeço, em primeiro lugar, o interesse pelo tema que despertou em mim, o desejo de continuar a pesquisa e aprofundamento dos meus conhecimentos. Agradeço a confiança, inspiração e contribuição na transmissão dos seus conhecimentos para a elaboração desta tese.

Aos meus coorientadores, Professor Filipe Marques e Professor Pedro Silva, por aceitarem orientar o meu trabalho, pelo seu apoio permanente, pelas suas críticas e preciosas sugestões até nos mínimos detalhes, pela paciência e dedicação que foram fundamentais para concretização desta tese.

Ao Professor Samuel Manda pela permanente transmissão de conhecimento, paciência, disponibilidade ilimitada e ajuda necessária para a realização desta tese.

Os professores são aqueles mestres do saber com quem se aprende muito sobre a vida, dentro e fora da sala de aula. Muito obrigada ao Professor Manuel Esquível, coordenador do Programa Doutoral de Estatística e Gestão do Risco, por ter sido capaz estar presente em todas as etapas desta formação e constituir-se no suporte principal para a concretização deste processo de formação.

A Professora Isabel Natário por todo o conhecimento transmitido e contribuições nesta conquista, incentivo e contribuição na revisão deste trabalho.

A todos os professores do Programa Doutoral de Estatística e Gestão do Risco que contribuíram para a minha formação, crescimento pessoal e profissional. Em especial à Professora Marta Faias, Professor Frederico Caieiro, Professor Agra Coelho, Professor Miguel Fonseca.

À Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL), com destaque para a Secretaria, Divisão Académica dos estudos avançados e para a Vanda Martins pelo suporte prestado durante estes anos.

No Centro de Matemática e Aplicações (CMA) passei horas sem conta onde partilhei inesquecíveis momentos de incentivo e saber, com amigos a quem não posso deixar de agradecer por tal. Entre eles está o Alberto Mulenga, Ivanilda Semedo, Ricardo Moura, Elsa Moreira, Sandra, Gabriel Cossa, José Moniz, Abdul Chafim e o Belchior Miguel.

Da Gulbenkian tenho a obrigação de destacar a disponibilidade da bolsa de estudos, a atenção, interação e apoio de toda uma equipa que me deu o devido suporte e consideração, e permitiu a minha inclusão na grande “Família Gulbenkian”. Um agradecimento especial é dirigido à Dra. Margarida Abecassis e à Dra. Margarida Cunha pela participação neste percurso da minha vida académica.

Aos meus pais, Bhangy Cassy e Alexandra Rodrigues, pelo conforto nas horas de angústia, pelo carinho, amor e confiança. Ao meu pai, por ter cultivado em mim o gosto pela matemática, pela força, compreensão e pelo exemplo na minha vida sempre participando e estimulando-me a ir mais além. A minha mãe pelo exemplo de mulher e de mãe.

Ao meu filho Tiago Luan, pelo seu amor incondicional e por conseguir compreender a minha dedicação a este trabalho demonstrada pelas frequentes perguntas "o que falta saber sobre amostragem complexa para terminares o estudo? Queres ajuda?".

Aos meus irmãos Bachir, Danilo e Annette por toda força, apoio e por manterem firme neste trajeto recheado de desafios e por estarem sempre presentes na minha vida.

Sou grata à minha família e amigos pelo apoio e atenção com que sempre me acari-nharam.

Aos meus colegas do Departamento de Matemática e Informática da Universidade Eduardo Mondlane vai o meu reconhecimento pelo apoio incondicional na prossecução dos meus objetivos de formação. A todos os que direta ou indiretamente fizeram parte desta fase da minha formação, o meu muito obrigada!

«Aproveitando dados para moldar o futuro.» (Autor desconhecido)

RESUMO

Os Inquéritos Demográficos e de Saúde, com representatividade a nível nacional, têm vindo a ser cada vez mais utilizados por investigadores da área das ciências da saúde especialmente nos países de África Subsariana (SSA). De entre as principais aplicações destacam-se os estudos de prevalência e fatores associados, os estudos de modelação espacial e mapas de risco, nomeadamente nas denominadas doenças da pobreza (HIV, Malária, Tuberculose, outras Doenças Tropicais Negligenciadas). Embora sejam dados secundários, a elevada cobertura e representatividade (nomeadamente por sexo, classe etária, regional e provincial) tem tornado estes inquéritos muito populares, sobretudo na ausência de outros estudos de base populacional. Assim, tem-se assistido nesta última década a um aumento substancial de publicações em revistas indexadas que utilizam estes dados nas mais diversas áreas da saúde.

No entanto, estes inquéritos geralmente utilizam delineamentos amostrais complexos, onde as unidades amostrais são selecionadas com diferentes probabilidades, através de múltiplas etapas; desta forma é necessário que as análises estatísticas reflitam este processo de amostragem, não só atribuindo a cada indivíduo o correspondente peso, mas também através da correta estimação das variâncias dos estimadores obtidos, sob pena de se sobrestimar ou subestimar os erros padrão associados aos estimadores empregues. Assim sendo, torna-se imprescindível o ajuste dos métodos estatísticos convencionais para obter estimativas mais precisas. Só desta forma se poderá fornecer evidências de apoio à decisão confiáveis, nomeadamente para a orientação de políticas de saúde pública que visam prevenir e combater determinadas doenças e/ou atingir os objetivos do desenvolvimento sustentável.

Usando dados de Inquéritos Demográficos e de Saúde de Moçambique e do Malawi, esta tese modelou diversos resultados em saúde tendo em conta o desenho completo da pesquisa, em particular a estrutura multinível dos dados, a amostragem não proporcional (exemplo: os pesos) e a informação geográfica.

Mais especificamente foi ajustado a nível micro e macro (i.e., a nível do indivíduo e a nível agregado) um modelo para determinar os fatores de risco associados ao uso de redes mosquiteiras em mulheres em idade reprodutiva em Moçambique, como uma

forma de prevenir a Malária. Foram usados Modelos Lineares Generalizados e Modelos Lineares Generalizados Mistos. Adicionalmente, estimou-se a distribuição geográfica da desnutrição infantil no Malawi ao nível distrital, e da febre e diarreia infantil ao nível provincial em Moçambique. Por fim, foi feita a modelação conjunta a da prevalência de duas doenças (febre e a diarreia em Moçambique) que se sabem ser interligadas, recorrendo ao Modelo de Análise de Dados Espaciais Conjuntos.

Em geral, as estimativas pontuais de média, proporção e os coeficientes de regressão foram semelhantes comparando-se amostragem complexa e amostragem aleatória simples. Os erro padrão ficaram subestimados ao usar métodos convencionais não ajustando para o plano amostral. Isto faz com que os coeficientes estimados para modelos de regressão sejam considerados significativos quando na verdade não o são. Adicionalmente, os métodos ajustados ao peso foram os mais adequados. Embora o padrão espacial estimado tenha sido semelhante, esta consideração dos pesos amostrais produziu estimativas de risco de doença que tiveram um nível de confiança maior, tanto na análise univariada quanto na análise bivariada. O uso dos métodos que considerem os pesos amostrais no mapeamento de doenças para estimar a distribuição espacial de riscos de doenças com base em dados complexos de pesquisas de saúde forneceu estimativas mais precisas, e também evidências de apoio confiáveis para orientar as políticas de saúde pública na focalização de recursos nas áreas de maior necessidade.

Palavras-chave: Modelos Lineares Generalizados, Amostragem Complexa, Modelos Multinível, Mapas de Risco, Modelos Espaciais, Malária, Desnutrição, Febre e Diarreia Infantil.

ABSTRACT

Demographic and Health Surveys, with national representativeness, have been increasingly used by health science researchers, especially in sub-Saharan African (SSA) countries. Among the main applications, stands out the prevalence studies and associated factors, spatial modeling studies and risk maps, particularly in the so-called poverty diseases (HIV, Malaria, Tuberculosis, other Neglected Tropical Diseases). Although secondary data are secondary data, high coverage and representativeness (mainly by gender, age group, regional and provincial) has made these surveys very popular, especially in the absence of other population-based studies. Thus, there has been a substantial increase in publications in indexed journals that use these data in various areas of health.

However, these surveys generally use complex sample designs, where sample units are selected with different probabilities, through multiple steps; thus, it is necessary that the statistical analyses reflect this sampling process, not only attributing to each individual the corresponding weight, but also through the correct estimation of the variances of the estimators obtained, under penalty of overestimating or underestimating the standard errors associated with the estimators used. Therefore, it is essential to adjust conventional statistical methods to obtain more accurate estimates. Only in this way can reliable evidence of decision support be provided, in particular for the guidance of public health policies aimed at preventing and combating certain diseases and/or achieving the objectives of sustainable development.

Using data from Demographic and Health Surveys of Mozambique and Malawi, this thesis modeled several health outcomes taking into account the complete design of the research, in particular the multilevel structure of the data, non-proportional sampling (e.g., weights) and geographic information.

More specifically, a model was adjusted at micro and macro level (i.e., at the individual level and at the aggregate level) to determine the risk factors associated with the use of mosquito nets in women of reproductive age in Mozambique as a way to prevent Malaria. Generalized Linear Models and Mixed Generalized Linear Models were used. In addition, the geographical distribution of child malnutrition in Malawi at the district level and of fever and child diarrhea at the provincial level in Mozambique was estimated. Finally,

the joint modeling of the prevalence of two diseases (fever and diarrhea in Mozambique, which are known to be interconnected) was carried out, using joint model of spatial data analysis.

In general, the point estimates of mean, proportion and regression coefficients were similar comparing complex sampling and simple random sampling. The standard errors were underestimated when using conventional methods not adjusting for the sampling plan. This makes the estimated coefficients for regression models considered significant, when in fact they are not. Additionally, the weight-adjusted methods were the most appropriate. Although the estimated spatial pattern was similar, the use of sample weights produced estimates of disease risk that had a higher level of confidence, both in the univariate and bivariate analysis. The use of methods that consider sample weights in disease mapping to estimate the spatial distribution of disease risks based on complex data from health research provided more accurate and reliable estimates that can be used to guide public health policies to focus resources on areas of the greatest need.

Keywords: Logistic Regression Models, Mosquito bednet, Malaria, Multilevel Models, complex survey design, Malnutrition, Fever, Diarrhea, Risk maps, Spatial models .

ÍNDICE

Índice de Figuras	xix
Índice de Tabelas	xxiii
Siglas	xxv
1 Introdução	1
1.1 Introdução	1
1.2 Exemplos de motivação: o panorama da saúde em África	4
1.2.1 Epidemia da malária em Moçambique	4
1.2.2 Febre e diarreia em Moçambique	5
1.2.3 Malnutrição infantil no Malawi	6
1.3 Objetivos	6
1.3.1 Objetivo Geral	6
1.3.2 Objetivos Específicos	7
1.4 Metodologia	7
1.4.1 Modelos Lineares Generalizados e a sua aplicação as doenças . .	7
1.4.2 Modelos Lineares Generalizados Mistos	8
1.4.3 Modelos de Análise de Dados Espaciais agregados e construção de mapas de risco	8
1.5 Estrutura da tese	9
2 Conceitos fundamentais e contexto geral	11
2.1 Descrição da área de estudo e população	11
2.1.1 Moçambique	11
2.1.2 Malawi	12
2.2 Descrição das fontes de dados e considerações éticas	13
2.2.1 Inquérito Demográfico e de Saúde 2011	14
2.2.2 Inquérito de Indicadores de Imunização, Malária e HIV/SIDA 2015	15
2.2.3 Inquérito Demográfico e de Saúde do Malawi	15

2.2.4	Amostragem	16
3	Modelos de Regressão logística, ajustando para o desenho da pesquisa	19
3.1	Introdução	19
3.2	Modelo Clássico de Regressão Logística	19
3.2.1	Estimação dos parâmetros do modelo	20
3.2.2	Testes de hipóteses	21
3.2.3	Avaliação da qualidade do ajuste e da seleção do modelo	22
3.3	Regressão Logística para amostras complexas	22
3.4	Aplicação para os fatores que contribuem para o uso de redes mosquiteiras por mulheres em idade reprodutiva em Moçambique	25
3.4.1	Variáveis do estudo	25
3.4.2	Análise dos dados	26
3.4.3	Resultados	27
3.4.4	Conclusão	33
4	Modelos logísticos multinível, ajustando para o desenho da pesquisa	35
4.1	Introdução	35
4.2	Modelos Logístico Multinível	36
4.2.1	Estimação dos parâmetros do modelo	37
4.2.2	Método de estimação de Máxima Pseudo-Verosimilhança	38
4.3	Aplicação para os fatores que contribuem para o uso de redes mosquiteiras em mulheres em idade reprodutiva em Moçambique	39
4.3.1	Variáveis do estudo	39
4.3.2	Análise dos dados	40
4.3.3	Resultados	41
4.3.4	Conclusão	45
5	Modelação espacial e Mapeamento de doenças, ajustando para o desenho amostral	47
5.1	Introdução	47
5.2	Notação Geral	48
5.2.1	Estimador de Horvitz-Thompson	49
5.3	Modelos Hierárquicos Bayesianos suavizados	49
5.3.1	Opções para ter em conta os pesos da pesquisa nos modelos hierárquicos	50
5.3.2	Inferência Bayesiana, Computação e Avaliação de Modelos	52
5.4	Aplicação da modelação espacial à prevalência da malnutrição infantil no Malawi e da febre e da diarreia infantil em Moçambique	53
5.4.1	Análise dos dados	53
5.4.2	Resultados	53
5.4.3	Conclusão	67

6	Modelação espacial e mapeamento conjunto de doenças ajustando para o desenho amostral	69
6.1	Introdução	69
6.2	Modelo de Componente Compartilhada	71
6.2.1	Inferência Bayesiana, Computação e Avaliação de Modelos	73
6.3	Aplicação da modelação espacial conjunta à prevalência da febre e da diarreia infantil em Moçambique	73
6.3.1	Dados e variáveis do estudo	73
6.3.2	Resultados	73
6.3.3	Conclusão	76
7	Conclusões e Trabalho Futuro	79
7.1	Conclusões	79
7.2	Recomendações	80
	Bibliografia	83

ÍNDICE DE FIGURAS

2.1	Mapa mostrando as regiões geográficas (províncias) de Moçambique.	12
2.2	Mapa mostrando as regiões geográficas (distritos) do Malawi.	13
5.1	Mapas da prevalência observada de baixa estatura para a idade (esquerda), baixo peso para a idade (centro) e baixo peso (direita) por distrito gerados pelo estimador de HT no Malawi usando os dados do MDHS 2015-16.	55
5.2	Mapas da prevalência observada (UW: Não ponderado, HT: Horvitz-Thompson) e espacialmente estimadas da baixa estatura para a idade (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Transformação de raiz quadrada de arco-seno estimador (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho efetivo de amostra usando efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.	57
5.3	Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada do baixo peso para a idade obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit-estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando efeito de desenho (Modelo 6)) em distritos no Malawi usando os dados do MDHS 2015-16.	58

5.4	Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacial estimada da baixa estatura para a idade (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.	59
5.5	Boxplots da prevalência observada (UW: não ponderada e HT: Horvitz Thompson) e prevalência estimada da baixa estatura para a idade obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN:Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando efeito de desenho (Modelo 6)) nos distritos do Malawi usando os dados do MDHS 2015-16.	60
5.6	Mapas da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e espacialmente estimadas do baixo peso (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.	61
5.7	Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada do baixo peso obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN:Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador do tamanho efetivo da amostra usando o efeito do desenho (Modelo 6)) nos distritos do Malawi estimado usando os dados do MDHS 2015-16.	62
5.8	Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacial estimada de febre (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por província em Moçambique usando o IMASIDA 2015.	65

5.9	Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada de febre obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador do tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador do tamanho efetivo da amostra usando efeito de desenho (Modelo 6)) em toda a província em Moçambique estimado a partir do IMASIDA 2015.	65
5.10	Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada de diarreia obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4) e ES: Estimador de tamanho de amostra efetivo (Modelo 5)) em toda a província em Moçambique estimado a partir do IMASIDA 2015.	66
5.11	Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacialmente estimadas de diarreia (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada arcsine (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4) e ES: Estimador de tamanho de amostra efetivo (Modelo 5)) por província em Moçambique usando o IMASIDA 2015.	66
6.1	Mapas da prevalência observada (ponderada) da febre e diarreia por províncias em Moçambique usando IMASIDA 2015.	74
6.2	Mapas da prevalência estimada da febre (não ponderada (a), ponderada (b)) e diarreia (não ponderada (c), ponderada (d)) por províncias em Moçambique usando IMASIDA 2015.	75
6.3	Mapas das estimativas a posteriori dos efeitos espaciais compartilhados da febre e da diarreia por províncias em Moçambique usando IMASIDA 2015 ((a) não ponderadas, (b) ponderadas).	76

ÍNDICE DE TABELAS

2.1	Junção das bases de dados.	15
3.1	Variáveis do estudo.	26
3.2	Caraterísticas sócio-económicas e demográficas das mulheres e idade reprodutiva em Moçambique, usando os dados do MDHS 2011.	27
3.3	Disponibilidade e uso da rede mosquiteira e pulverização intradomiciliar.	29
3.4	Modelo logit final Não Ponderado.	30
3.5	Modelo logit final Ponderado.	31
4.1	Variáveis do estudo	40
4.2	Modelo logístico multinível final não ponderado	42
4.3	Modelo logístico multinível final ponderado	44
5.1	Resumo da prevalência observada (ponderada) da baixa estatura para a idade, baixo peso para a idade e baixo peso por distrito no Malawi.	54
5.2	Comparação de modelos BYM para o mapeamento da baixa estatura para a idade no Malawi usando MDHS 2015-16.	55
5.3	Comparação de modelos BYM para o mapeamento do baixo peso para idade no Malawi usando MDHS 2015-16.	56
5.4	Comparação de modelos BYM para o mapeamento do baixo peso no Malawi usando MDHS 2015-16.	56
5.5	Resumo das taxas de prevalência observadas (ponderadas) de febre e diarreia por província em Moçambique.	63
5.6	Comparação do modelo BYM para mapeamento da febre em Moçambique.	63
5.7	Comparação do modelo BYM para o mapeamento da diarreia em Moçambique.	64
6.1	Tabela cruzada da prevalência da febre e da diarreia infantil em Moçambique, usando os dados do IMASIDA 2015.	74
6.2	Medidas de comparação dos modelos.	75

SIGLAS

AAS	amostragem aleatória simples (pp. 3, 19, 22, 48, 52)
AC	amostragem complexa (pp. 3, 22, 23, 34, 48)
AE	Área de Enumeração (p. 16)
AF	agregados familiares (pp. 14–16, 26, 28)
AIC	critério de Informação de Akaike (p. 25)
aOR	<i>Adjusted Odds Ratio</i> ou Razão de Chances Ajustada (p. 26)
AS	transformação da raiz quadrada do arco-seno (pp. 51, 55)
BYM	Besag, York e Mollié (p. 50)
Deff	efeito do desenho (p. 52)
DHS	Inquéritos Demográficos e de Saúde (pp. 2, 4, 13, 19, 40)
DIC	critério de informação da deviance (pp. 52, 55, 73, 74)
GLMMs	Modelos Lineares Generalizados Mistos (pp. 35, 36)
GLMs	Modelos Lineares Generalizados (pp. 19, 35, 36)
HIV	Vírus de Imunodeficiência Humana (p. 4)
HT	Horvitz-Thompson (pp. 49, 50, 56, 57, 60)
ICAR	autoregressiva condicional intrínseca (pp. 50, 72)
ICC	coeficiente de correlação intraclasse (p. 40)
iid	independente e identicamente distribuídas (pp. 19, 20, 22)
IMASIDA 2015	Inquérito de Indicadores sobre Imunização, Malária e HIV/SIDA em Moçambique (pp. 13–15, 48, 53, 70, 73)
INE	Instituto Nacional de Estatística (p. 14)
INLA	aproximação de Laplace aninhada e integrada (pp. 52, 73)
LMIC	países de baixa e média renda (pp. 4–6)

LN	Logit Normal (<i>p. 51</i>)
MCMC	Monte Carlo via cadeias de Markov (<i>p. 52</i>)
MDHS 2011	Inquérito Demográfico e de Saúde de Moçambique do ano 2011 (<i>pp. 13, 14, 19, 26</i>)
MDHS 2015-16	Inquérito Demográfico e de Saúde do Malawi do ano 2015-16 (<i>pp. 13–16, 48, 53</i>)
MISAU	Ministério da Saúde (<i>p. 14</i>)
MPHC	Censo da População e habitação do Malawi (<i>p. 16</i>)
MPV	Máxima Pseudo-Verosimilhança (<i>pp. 23, 24</i>)
MPVM	Máxima Pseudo-Verosimilhança Multinível (<i>pp. 35, 38</i>)
MRF	campo aleatório de Markov (<i>p. 52</i>)
MV	Máxima Verosimilhança (<i>pp. 20–23</i>)
ODS	Objetivos de Desenvolvimento Sustentável (<i>pp. 1, 3–6</i>)
OMS	Organização Mundial de Saúde (<i>pp. 4, 5</i>)
PL	pseudo-verosimilhança (<i>p. 51</i>)
PPS	probabilidade proporcional ao tamanho (<i>p. 16</i>)
RL	Regressão Logística (<i>pp. 19, 23</i>)
SSA	África Subsariana (<i>pp. 1, 2, 5, 6, 69, 70, 76, 81</i>)
UB	binomial não ajustado (<i>pp. 55–58</i>)
UPAs	Unidades Primárias de amostragem (<i>pp. 14–16, 23, 24</i>)
USAs	Unidades Secundárias de amostragem (<i>pp. 14–16</i>)

INTRODUÇÃO

1.1 Introdução

Todos os anos, milhões de mulheres e crianças morrem de doenças evitáveis. O sofrimento destas pessoas é inaceitável no século XXI. Globalmente, em 2017 cerca de 5.4 milhões de mortes de crianças no mundo foram por causas evitáveis [91]. Faltando apenas poucos anos para atingir as metas dos Objetivos de Desenvolvimento Sustentável (ODS), o indicador 3.2 garante o fim das mortes evitáveis de recém-nascidos e menores de cinco anos visando reduzir as mortes de menores de cinco anos para no máximo, menos de 25 por 1,000 nascidos vivos até o ano de 2030 [117].

Muitos países da África Subsaariana (SSA), incluindo Moçambique e o Malawi, precisam acelerar o progresso da redução das altas taxas de mortalidade infantil. Os líderes globais devem intensificar os seus esforços para melhorar a saúde das mulheres e das crianças. Reconhece-se cada vez mais que a saúde de mulheres e crianças é a chave para o progresso em todas as metas de desenvolvimento [78]. Este facto exige que todos os parceiros se unam e tomem ações coordenadas. Todos têm um importante papel a desempenhar: governos, sociedade civil, organizações comunitárias, instituições globais e regionais, doadores, fundações filantrópicas, as Nações Unidas e outras organizações multilaterais, bancos de desenvolvimento, o setor privado, força de trabalho em saúde, associações profissionais, académicos e pesquisadores.

Moçambique e Malawi são dois países que têm tentando fazer a sua reconstituição e reorganização desde a independência. Desde essa altura vários progressos têm-se verificado em termos do desenvolvimento das condições de vida dos habitantes, infraestruturas e cuidados de saúde [106]. Embora tenha havido, nos últimos anos, enormes financiamentos na área da saúde por parte da comunidade internacional e esforços por parte dos governos, o estado de saúde da população destes dois países ainda se encontra muito débil, sendo as mulheres e as crianças os grupos mais afetados [78, 92, 93, 106].

Conhecer e caracterizar os principais resultados em saúde, assim como os seus determinantes é fundamental para se monitorizar o estado de saúde das populações assim como para avaliar o impacto de políticas que visem a sua melhoria. Para tal, é necessário que

existam inquéritos que permitam a recolha de dados de base populacional, que gerem estimativas fiáveis sobre estados de saúde e de doença da população, bem como as respetivas determinantes, assim como a sua evolução ao longo do tempo. Por exemplo, em Portugal já foram realizados 6 Inquéritos Nacionais de Saúde (1987, 1995/1996, 1998/1999, 2005/2006, 2014/15 e 2019) utilizando amostras probabilísticas representativas da população de Portugal. No entanto, estes inquéritos têm um custo muito elevado dificilmente suportável pelos países de baixa renda. Neste contexto, a USAID (Agência dos Estados Unidos para o Desenvolvimento Internacional), através do Programa DHS (Inquéritos Demográficos e de Saúde) há mais de 30 anos que financia e fornece assistência técnica para a implementação destas operações estatísticas em países de África, Ásia, América Latina/Caribe e Europa Oriental. Os dados assim recolhidos permitiram aprofundar e transformar o conhecimento que temos sobre questões ligadas à demografia, saúde e nutrição no mundo em desenvolvimento (<https://dhsprogram.com/>).

Para o caso específico de Moçambique já foram implementados 3 Inquéritos Demográficos e de Saúde (DHS) e está em curso um quarto relativo ao ano 2022 (DHS 1997 [29], DHS 2003 [28], DHS 2011 [27], DHS 2022). Também já foi conduzido um inquérito de Comportamentos, Atitudes, Práticas e Prevalência do HIV/SIDA em 2009 (INSIDA 2009 [40]), um Inquérito de Indicadores de Imunização, Malária e HIV/SIDA (IMASIDA 2015 [76]) e um Inquérito Nacional sobre Indicadores de Malária 2018 [41]. Para o caso do Malawi, 5 Inquéritos Demográficos e de Saúde (DHS 1992 [87], DHS 2000 [86], DHS 2004 [84], DHS 2010 [83], DHS 2015-16 [82]), um inquérito de Comportamentos, Atitudes, Práticas em 1996 (KAP 1996 [85]), um Inquérito de Avaliação de Prestação de Serviços 2013-14 (SPA 2013-14 [35]) e um Inquérito Nacional sobre Indicadores de Malária em 2012 [99], 2014 [81] e 2017 [80].

Estes Inquéritos Demográficos e de Saúde têm vindo a ser cada vez mais utilizados por investigadores da área das ciências da saúde especialmente nos países de SSA. De entre as principais aplicações destacam-se os estudos de prevalência e fatores associados, os estudos de modelação espacial e mapas de risco, nomeadamente nas denominadas doenças da pobreza (HIV, Malária, Tuberculose, outras Doenças Tropicais Negligenciadas). Embora sejam dados secundários, a elevada cobertura e representatividade (nomeadamente por sexo, classe etária, regional e provincial) tem tornado estes inquéritos muito populares, sobretudo na ausência de outros estudos de base populacional. Assim, tem-se assistido nesta última década a um aumento substancial de publicações em revistas indexadas que utilizam estes dados nas mais diversas áreas da saúde.

No entanto, estes inquéritos geralmente utilizam delineamentos amostrais complexos, onde as unidades amostrais são selecionadas com diferentes probabilidades, através de múltiplas etapas; desta forma é necessário que as análises estatísticas reflitam este processo de amostragem, não só atribuindo a cada indivíduo o correspondente peso, mas também através da correta estimação das variâncias dos estimadores obtidos, sob pena de se sobrestimar ou subestimar os erros padrão associados aos estimadores empregues [11, 50, 97]. Assim sendo, torna-se imprescindível o ajuste dos métodos estatísticos convencionais

para obter estimativas mais precisas. Só desta forma se poderá fornecer evidências de apoio à decisão confiáveis, nomeadamente para a orientação de políticas de saúde pública que visem prevenir e combater determinadas doenças e/ou atingir os ODS.

Ao contrário da amostragem aleatória simples (AAS), que seleciona aleatoriamente uma amostra de uma população-alvo para garantir que as observações selecionadas sejam independentes umas das outras, no desenho de amostragem complexa (AC), as informações referem-se às unidades de amostragem selecionadas, e que geralmente consistem em estratificar a população em grupos homogêneos, dos quais um número pré-determinado de unidades é selecionado, e de agrupamento, onde a população é dividida em grupos e um subconjunto é amostrado. Sabemos que em delineamentos complexos as unidades amostrais têm probabilidades de seleção diferentes, o que tem de ser tido em conta nas análises estatísticas a realizar, ponderando as unidades de forma diferente, normalmente como o inverso da probabilidade de seleção.

Os métodos estatísticos convencionais que pressupõem observações independentes não devem ser usados com dados coletados de amostragem por conglomerados devido às possíveis observações não independentes [11, 16, 109]. O uso de métodos estatísticos convencionais em dados não independentes pode resultar em estimativas tendenciosas dos erros padrão e conclusões estatísticas incorretas [11, 16, 50, 109].

No entanto, verifica-se que, na prática, em muitas publicações da área da saúde, mesmo em revistas de impacto elevado, esta questão é ignorada e que os modelos são frequentemente estimados como se o processo de amostragem não fosse complexo. Existem duas estratégias possíveis para lidar com esse problema. Seja uma abordagem baseada no desenho, onde os pesos são incluídos no procedimento de inferência para contabilizar o esquema de amostragem, ou uma abordagem baseada no modelo, onde o esquema de amostragem é contabilizado considerando variáveis de desenho no modelo, que não estão frequentemente disponíveis, ou usando a distribuição (aproximada) da função de pseudo-verosimilhança que incorpora os pesos amostrais.

Desta forma, pode-se então estimar os *outcomes* de saúde (prevalências, entre outros) levando em consideração as características do desenho amostral, obtendo erros padrão, testar a significância dos parâmetros, avaliar a qualidade do ajuste, fazer a seleção do modelo. É ainda possível incorporar espaço [18, 75] nesta estimação. No entanto, na prática, muitas vezes o esquema de amostragem não é levado em consideração e as análises feitas, muitas vezes, simplesmente o ignoram. Esta não é uma questão fácil [30], os desenvolvimentos metodológicos necessários são bastante recentes, ainda não estão completamente concluídos, e nem todos estes procedimentos estão implementados nos *softwares* disponíveis.

Assim, uma questão fundamental é a de saber como modelar o comportamento de *outcomes* de saúde ajustando para o delineamento da pesquisa, de modo a obter estimativas confiáveis para poder tomar as decisões mais acertadas para redução dos problemas de saúde da população. Para além disso, algumas doenças (como a Malária, por exemplo) apresentam um padrão geográfico que é necessário levar em linha de conta aquando da

estimação desses modelos.

1.2 Exemplos de motivação: o panorama da saúde em África

O continente africano enfrenta um grande desafio lidando com múltiplas doenças em simultâneo, como a Malária, a Tuberculose, Vírus de Imunodeficiência Humana (HIV), doenças respiratórias inferiores e doenças diarreicas que representam as principais causas de morbidade e mortalidade em países de baixa e média renda (LMIC) [92, 93].

A região africana da Organização Mundial de Saúde (OMS) é também a região com menor esperança de vida no mundo e também uma das regiões com maior mortalidade infantil e mortalidade materna. Embora estratégias estejam a ser implementadas nos diferentes países da região, o ritmo de melhoria dos indicadores de saúde em África é insuficiente para se alcançar as metas dos ODS [93].

Para melhorar a situação de saúde, o uso de dados é crucial para a tomada de decisão. Em África muitos países realizam inquéritos de saúde de rotina, particularmente os DHS. No entanto, a capacidade dos países para analisar as informações dos inquéritos também é bastante baixa, sendo a maior parte feita por parceiros externos [92].

1.2.1 Epidemia da malária em Moçambique

O impacto da malária em África ainda é um importante problema de saúde pública, particularmente nos países tropicais mais pobres do continente, sendo as mulheres grávidas e as crianças menores de 5 anos os grupos de maior risco [94]. Os efeitos adversos desta doença estão ligados a um ciclo vicioso de pobreza e muitas doenças infecciosas, particularmente nas áreas económicas de baixo crescimento.

Moçambique é um dos países em que esse problema é uma verdadeira dor de cabeça na vida diária, não só por causa do número de casos, mas também pela taxa de mortalidade causada pela doença. Esta doença constitui numa das principais causas de morte, sendo responsável por 40% das consultas externas, 60% das admissões em enfermarias e 30% de mortes nos hospitais [94].

Reconhece-se que em Moçambique, fatores como as mudanças ambientais, particularmente os níveis de precipitação, temperatura e humidade, contribuem para desenvolver as melhores condições para ter mosquitos em todo o país e, mais recentemente, durante todo o ano. Além disso, é difícil, por razões económicas, que os cidadãos obtenham os meios para prevenir a malária. Desta forma, um plano nacional para prevenir e evitar a malária foi desenvolvido e implementado no país ao nível da comunidade, que oferece a possibilidade de que, gratuitamente, as comunidades recebam redes mosquiteiras tratadas com inseticidas e periodicamente as casas são tratadas com diferentes tipos de inseticidas, predominantemente do tipo *spray* através da pulverização intradomiciliar. Também são usadas formas de educar a população, de modo a reduzir as condições físicas propícias para criar o mosquito.

No entanto, apesar destes métodos de prevenção eficazes serem de distribuição gratuita para grupos de risco e subsidiada para o resto da população, a sua taxa de utilização é ainda muito reduzida, pela generalidade da população em Moçambique [27]. Como tal é importante conhecer e avaliar os principais fatores associados ao uso das redes mosquiteiras.

1.2.2 Febre e diarreia em Moçambique

Em todo o mundo, o número de mortes de menores de cinco anos caiu de 12.7 milhões para 5.9 milhões de mortes entre 1990 e 2016, o que significa que diariamente 16,000 crianças morrem com menos de 5 anos em comparação com 35,000 mortes em 1990, enquanto a mortalidade neonatal passou de 5.1 milhões de mortes em 1990 para 2.7 milhões em 2015. A maioria das mortes de crianças ocorreu na SSA, com uma taxa de redução anual de apenas 1.6% em 1990–2000 e 4.1% em 2000–2015.

Globalmente, cerca de 6.3 milhões de mortes de crianças e adolescentes ocorreram em 2017 e a maioria, cerca de 5.4 milhões de mortes no mundo foram por causas evitáveis. O indicador dos ODS, 3.2, propõe o fim das mortes evitáveis de recém-nascidos e menores de cinco anos com o objetivo de reduzir as mortes de menores de cinco anos para níveis inferiores a 25 por 1,000 nascidos vivos até o ano de 2030.

Segundo a OMS a principal causa de febre em crianças menores de cinco anos são as infeções respiratórias agudas, sendo outras causas as gastroenterites, infeções localizadas como a meningite e infeções de pele ou tecidos e infeções do trato urinário. Porém, em áreas endémicas a malária pode ser uma das principais causas de febre.

Os Grupos Inter-agências da ONU em 2015 relataram que de 195 países, apenas 62 países (24 LMIC) atingiram a meta de desenvolvimento do milénio (ODM) 4 com metas de reduzir as taxas de mortalidade de menores de cinco anos em dois terços entre 1990 e 2015. Muitos países da SSA, incluindo Moçambique, precisam acelerar o progresso da redução das altas taxas de mortalidade infantil. A taxa de mortalidade na SSA é 14 vezes maior que a razão média de 1 em 185 nos países desenvolvidos.

Estimativa global da carga de doenças em 2017, as maiores taxas de mortalidade por diarreia foram observadas na SSA e no Sul da Ásia, variando de 50 a 150 por 100,000. Embora um número substancial de mortes de crianças por diarreia e malária possa ser evitado pelas intervenções existentes, muitos LMIC têm cobertura abaixo do ideal dessas intervenções e enfrentam uma grave escassez da força de trabalho necessária para fornecer serviços essenciais de saúde.

Neste contexto, compreender a distribuição geográfica da febre e da diarreia infantil é fundamental para definir e priorizar intervenções que podem contribuir para acelerar a redução da morbi-mortalidade em crianças menores de 5 anos.

1.2.3 Malnutrição infantil no Malawi

A malnutrição constitui um dos problemas de saúde pública com importância a nível mundial, especialmente nos LMIC, onde as crianças menores de 5 anos de idade são mais propensas a sofrer desta condição do que os adultos.

A malnutrição é definida como a falta de nutrição adequada. Essa condição é observada quando um indivíduo não recebe uma quantidade suficiente de vitaminas, minerais e outros nutrientes. Desastres naturais como terremotos, secas, inundações, deslizamentos de terra e ciclones expõem crianças e adultos à desnutrição.

De acordo com [23, 116], existem várias formas da malnutrição se manifestar, sendo a destacar: a desnutrição crónica ou baixa estatura para a idade (*stunting* ou nanismo nutricional); a desnutrição aguda ou baixo peso para a estatura (*wasting*); o baixo peso para a idade ou *underweight*, a deficiência de micronutrientes.

Em todo o mundo, 248 milhões de crianças menores de cinco anos são afetadas pela malnutrição [118]. A África Austral (onde se incluem Moçambique e Malawi), África Central, África Oriental e África Ocidental têm uma prevalência maior do que a prevalência média global (28%, 33%, 37% e 31%, respetivamente) [116].

A malnutrição contribui para a maior percentagem de causas de mortalidade em menores de cinco anos. Estatísticas divulgadas pelo UNICEF (2018) [119] revelam que a taxa de mortalidade de menores de cinco anos diminuiu mais da metade desde 1990. No entanto, a SSA ainda tem a taxa mais alta.

Embora o Malawi tenha apresentado melhorias substanciais nos indicadores de saúde e progresso na saúde infantil nas últimas décadas, as mortes infantis evitáveis continuam a ser inaceitavelmente altas para atingir os ODS [7, 116, 47]. Compreender a epidemiologia local de doenças é fundamental para definir e priorizar intervenções que podem contribuir para acelerar a redução da morbi-mortalidade em crianças menores de 5 anos.

1.3 Objetivos

1.3.1 Objetivo Geral

Usando dados de Inquéritos Demográficos e de Saúde, este estudo teve como objetivo modelar um número de *outcomes* da saúde tendo em conta o desenho completo da pesquisa, em particular a estrutura multinível dos dados; a amostragem não proporcional (exemplo: os pesos) e a informação geográfica. Reconhece-se que alguns das *outcomes* de saúde estão ligadas etiologicamente (compartilhando alguns ou semelhantes fatores de risco). Em particular, os métodos utilizados envolveram modelação de efeitos mistos ponderados com ambos termos aleatórios não estruturados e estruturados em nível apropriado de agregação geográfica, relevantes para o planeamento local de saúde.

1.3.2 Objetivos Específicos

Para a concretização do objetivo geral definiram-se os seguintes objetivos específicos:

- Ajustar/Estimar o modelo de regressão logística para analisar os fatores que influenciam no uso da rede mosquiteira em mulheres em idade reprodutiva em Moçambique, ajustando ao delineamento amostral;
- Ajustar/Estimar o modelo de regressão logística multinível para analisar os fatores que influenciam no uso da rede mosquiteira em mulheres em idade reprodutiva em Moçambique, ajustando ao delineamento amostral;
- Definir e analisar abordagens inovadoras de ajuste de pesos amostrais nos modelos espaciais de mapeamento de doenças e produzir mapas de risco da prevalência dos indicadores da malnutrição no Malawi e da diarreia e da febre em Moçambique;
- Modelar conjuntamente doenças que se supõe estar interligadas e incorporar os pesos amostrais no mapeamento de doenças (febre e diarreia em Moçambique).

1.4 Metodologia

Para a prossecução destes objetivos recorreu-se essencialmente a três abordagens estatísticas importantes para a pesquisa complexa, apresentadas de forma resumida nas subsecções seguintes.

1.4.1 Modelos Lineares Generalizados e a sua aplicação as doenças

Na especificação de um Modelo Linear Generalizado (MLG) existe uma variável aleatória de interesse Y , designada por *outcome*, variável dependente ou variável de resposta, e um conjunto de k variáveis explicativas ou co-variáveis (designadas usualmente por $X_1, X_2, \dots, X_k$) que se admite poderem explicar parte da variabilidade inerente de Y . A variável Y pode ser contínua ou discreta e as variáveis explicativas, determinísticas ou estocásticas podem ser de qualquer natureza.

Exemplos de MLG incluem o Modelo de Regressão Linear, o Modelo de Regressão Logística, o Modelo de Regressão de Poisson e o Modelo Probit, entre outros [79, 115]. Os MLG têm sido aplicados a diversos domínios da área da saúde, nomeadamente à modelação de doenças (modelos de epidemiologia), à modelação de tempos de internamento hospitalares, à modelação de acesso aos cuidados de saúde e aos ensaios clínicos. O modelo de regressão logística detém um dos papéis centrais em estudos de prevalência de doenças [38], sendo dos modelos mais empregues neste tipo de análises. Esta escolha deriva essencialmente de dois fatores: a natureza da variável dependente (qualitativa e dicotómica) e a interpretação dos coeficientes de regressão que estão intimamente ligados ao conceito de *Odds Ratio*, conceito este fundamental na área das doenças infecciosas, uma

vez que pode ser visto como uma boa aproximação (sob determinadas hipóteses) do Risco Relativo.

1.4.2 Modelos Lineares Generalizados Mistos

Um modelo linear generalizado misto (MLGM) é um tipo particular de modelo misto. É uma extensão do modelo linear generalizado em que as variáveis explicativas, para além dos efeitos fixos habituais, incluem igualmente efeitos aleatórios. Geralmente assume-se que estes efeitos aleatórios seguem uma distribuição normal. Estimar estes modelos pela máxima verosimilhança implica calcular integrais nestes efeitos aleatórios. No entanto, em geral, é muito difícil escrever analiticamente estes integrais. Têm vindo a ser desenvolvidos na literatura vários métodos de aproximação, mas nenhum deles é suficientemente geral, para que se consiga ter boas propriedades estatísticas, qualquer que seja a natureza dos dados. Como tal, desenvolveram-se outros tipos de métodos, baseados, por exemplo, no algoritmo EM, entre outros, que permitem estimar este tipo de modelos [73]. A abordagem com modelos mistos é particularmente adequada para estimar modelos com uma estrutura multinível [56, 32], como é o caso de muitos modelos usados para descrever doenças infecciosas que se baseiam nos inquéritos demográficos e de saúde. No entanto, as aplicações destes modelos, nomeadamente, na modelação de doenças infecciosas no Continente africano, especialmente nos países de língua portuguesa, ainda escasseiam.

1.4.3 Modelos de Análise de Dados Espaciais agregados e construção de mapas de risco

Os mapas de risco de doença não devem simplesmente ser construídos com base nas taxas de mortalidade observadas ou prevalência observada, sob pena do resultado poder levar a conclusões espúrias por não levar em consideração a variabilidade populacional e a natural dependência espacial que é de admitir. Assim, estes devem ser estimados com base em modelos para o número de casos de doença ou prevalência de doença em cada divisão (área) da região de interesse, por exemplo, que possam levar esses aspetos em consideração.

Nesta construção, tipicamente, assume-se que a variável resposta é condicionalmente independente entre áreas. Contudo, a dependência espacial é introduzida através dos seus valores esperados condicionais. Por este motivo os modelos que se utilizam, como, por exemplo, os modelos CAR (do inglês *conditional auto-regressive*) e SAR (do inglês *simultaneous auto-regressive*), são muitas vezes formulados sob uma moldura hierárquica Bayesiana. É ainda relativamente incipiente a utilização desta metodologia em mapas de risco de doenças ajustando para o desenho da pesquisa. Pretende-se fazê-lo para as várias doenças em análise, avaliando a real existência de padrões espaciais associados, estimando o risco de doença em cada área e construindo os correspondentes mapas de risco. Adicionalmente, pretende-se estimar e construir mapas de risco das doenças em conjunto.

1.5 Estrutura da tese

Para uma melhor organização e apresentação do trabalho, este encontra-se dividido em 7 capítulos.

O Capítulo 2 fornece uma explicação detalhada dos dados, a fonte e o desenho amostral. Os próximos quatro capítulos detalham as principais contribuições desta tese.

No Capítulo 3 descreve-se o modelo de regressão logística da pesquisa e ajustar-se o modelo ao desenho da pesquisa. A seguir ajusta-se o modelo para identificar os fatores que influenciam no uso da rede mosquiteira em mulheres em idade reprodutiva em Moçambique.

No Capítulo 4 é apresentada uma extensão dos modelos lineares generalizados que relaxam a suposição de independência entre as observações, os Modelos lineares generalizados mistos, dando particular destaque aos modelos de regressão logística multinível. Estes modelos são então ajustados ao delineamento amostral e aplicados para a identificação dos fatores de risco associados ao uso da rede mosquiteiras por mulheres em idade reprodutiva em Moçambique.

O Capítulo 5, serve como uma introdução a análise espacial, dando recentes desenvolvimentos no mapeamento de doenças usando dados de pesquisa para doenças, modelos espaciais que incorporam pesos de amostragem são descritos. Os modelos são então aplicados à modelação da variação geográfica da prevalência da malnutrição infantil no Malawi e da febre e da diarreia infantil em Moçambique.

O Capítulo 6, estende os modelos de mapeamento de doenças para múltiplas doenças e faz-se o ajuste para o desenho da pesquisa. Descrevem-se os modelos comumente encontrados para modelação espacial conjunta de doenças e adotada para modelos logísticos espaciais bivariados. Fez-se uma aplicação desta metodologia para modelar em conjunto a prevalência da diarreia e da febre em Moçambique. Por fim, no Capítulo 7 são apresentadas de forma resumida as conclusões que derivam dos capítulos anteriores e sugestões para trabalhos futuros.

CONCEITOS FUNDAMENTAIS E CONTEXTO GERAL

2.1 Descrição da área de estudo e população

2.1.1 Moçambique

Moçambique encontra-se situado na faixa sul-oriental do Continente Africano, entre os paralelos $10^{\circ}27'$ e $26^{\circ}52'$ de latitude Sul e entre os meridianos $30^{\circ}12'$ e $40^{\circ}51'$ longitude Este. Ao Norte faz fronteira com a Tanzânia; ao Oeste com o Malawi, Zâmbia, Zimbabwé e Essuatíni; e ao Sul com a África do Sul.

Toda a faixa Este, banhada pelo Oceano Índico numa extensão de 2,740 km, tem um significado vital tanto para Moçambique como para os países vizinhos situados no interior que só têm ligação com o oceano através dos portos moçambicanos. A superfície do seu território é de $799,380\text{km}^2$, uma faixa costeira com quase 2,700 km de norte a sul.

O país está dividido em 10 Províncias e a Cidade de Maputo: 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). Cada província é dividida em distritos; cada distrito é constituído por postos administrativos; subdivididos por localidades ou bairros. O território moçambicano, como toda a região Austral do Continente Africano, não apresenta grande variedade de paisagem, podendo-se verificar três tipos de relevos: a planície litoral que ocupa grande parte do território (40%), esta é a região natural onde se observa a maior concentração da população; os planaltos com altitudes que variam entre 200 e 1000 metros; os grandes planaltos e montanhas que ocupam uma pequena parte do território nacional, com altitudes superiores a 1000 metros. Sob o ponto de vista da distribuição geográfica da população, já que não constituem uma superfície contínua, não oferecem grandes obstáculos para assentamentos humanos.

O clima é do tipo tropical húmido, com duas estações, a estação quente e chuvosa que decorre entre novembro a março e uma estação seca que vai de Abril a Outubro. No entanto, as condições climáticas dependem da altitude. As chuvas são fortes ao longo da costa e diminuem de norte a sul.



Figura 2.1: Mapa mostrando as regiões geográficas (províncias) de Moçambique.

2.1.2 Malawi

O Malawi está localizado na África Austral, estendendo-se entre as latitudes, $9^{\circ}22'S$ e $17^{\circ}03'S$ e longitudes, $33^{\circ}40'E$ e $35^{\circ}55'E$. Tem uma área de $118,480\text{km}^2$, 900 km de comprimento e cerca de 250 km de largura [74]. Faz fronteira com Moçambique ao sul, sudeste e sudoeste, Zâmbia ao centro-oeste e noroeste, e Tanzânia ao norte e nordeste. O país tem três regiões principais, a saber: região norte, região central e região sul. Administrativamente, há um total de vinte e oito distritos e a região norte tem seis distritos, ou seja, Karonga, Rumphi, Chitipa, Mzimba, Nkhata-bay e Likoma. A região centro tem nove distritos (Kasungu, Dowa, Ntchisi, Nkhota-kota, Mchinji, Salima, Dedza, Lilongwe e Ntcheu) e por sua vez a região sul conta com treze distritos (Mulanje, Chiradzulu, Blantyre, Balaka, Mangochi, Machinga, Phalombe, Chikwawa, Nsanje, Mwanza, Thyolo, Neno e Zomba). Existem quatro grandes cidades no Malawi que são: Blantyre, Lilongwe, Mzuzu e Zomba, ver Figura 2.2. O tamanho total da população com base no último censo realizado em 2018 é de 17,563,749, sendo que as regiões sul e centro são relativamente mais populosas do que a região norte com os tamanhos populacionais 7,750,629, 7,523,340 e 2,289,780 respetivamente [88].

O clima do país é tropical, mas a influência da sua alta elevação significa que as temperaturas são relativamente frias. No inverno as temperaturas caem para cerca de $18 - 19^{\circ}\text{C}$, e nos meses mais quentes (Setembro a Janeiro) as temperaturas variam de $22 - 27^{\circ}\text{C}$. A estação chuvosa pode variar no seu tempo e intensidade de ano para ano.



Figura 2.2: Mapa mostrando as regiões geográficas (distritos) do Malawi.

No sul do Malawi, a estação chuvosa normalmente dura de novembro a fevereiro, mas a chuva continua em março e abril no norte do país.

2.2 Descrição das fontes de dados e considerações éticas

Para apoiar os conceitos metodológicos apresentados nos capítulos seguintes desta tese, usou-se dados dos seguintes inquéritos: Inquérito Demográfico e de Saúde de Moçambique do ano 2011 (MDHS 2011), Inquérito de Indicadores sobre Imunização, Malária e HIV/SIDA em Moçambique (IMASIDA 2015) e Inquérito Demográfico e de Saúde do Malawi do ano 2015-16 (MDHS 2015-16), cuja descrição encontra-se nas seguintes subsecções. Estes dados foram cedidos pelo programa DHS sob o compromisso de que só devem ser utilizados para fins da investigação ou estudo registado. Para usar estes dados para outra finalidade, deve-se apresentar uma nova solicitação de projeto de pesquisa. Os dados são confidenciais, mas estão disponíveis ao público mediante autorização do DHS *program* e podem ser obtidos em <https://dhsprogram.com/>. Nenhum esforço foi feito para identificar qualquer família ou indivíduo entrevistado na pesquisa.

O DHS desenvolveu procedimentos padrão, metodologias e manuais para orientar os processos de planeamento da pesquisa por amostragem, desenho e coleta de dados

da pesquisa por amostragem para obter dados de qualidade e refletir a saúde e a representação demográfica da população comparável entre países. O procedimento padrão de amostragem usado no MDHS 2011, IMASIDA 2015 e MDHS 2015-16 é descrito na subsecção 2.2.4.

2.2.1 Inquérito Demográfico e de Saúde 2011

O MDHS 2011 foi realizado pelo Instituto Nacional de Estatística (INE) e Ministério da Saúde (MISAU), com apoio técnico e financeiro da *Macro Internacional Inc.* Os dados do MDHS 2011 são de um estudo transversal, nacional, com base populacional [27].

A amostra do MDHS 2011 foi desenhada de modo a obter estimativas representativas a nível nacional, provincial, regional e de área de residência (urbana e rural) de mulheres de 15 a 49 anos e de homens de 15 a 64 anos. Com relação ao delineamento amostral, o MDHS 2011 segue um delineamento amostral complexo, ou seja, combinou estratificação e conglomeração em duas etapas, com probabilidades de seleção desiguais resultando em pesos amostrais distintos para os elementos da amostra. Os estratos considerados foram ao nível de província e de área de residência, resultando em 21 estratos. A primeira etapa de seleção consistiu na seleção de 611 Unidades Primárias de amostragem (UPAs), sendo as áreas de enumeração com probabilidade de seleção proporcional ao número de agregados familiares (AF) em cada estrato dentro de cada província. Na segunda etapa, foram selecionadas as Unidades Secundárias de amostragem (USAs), sendo os AF (20 nas UPAs urbanas e 25 nas UPAs rurais).

Assim, o MDHS 2011 reúne informação de 13,919 AF, tendo sido entrevistadas 13,745 mulheres de 15 a 49 anos e 4,035 homens de 15 a 64 anos, e esta coleta dos dados foi feita entre julho e novembro de 2011 por entrevistas domiciliares, com aplicação de três tipos de questionários (de agregados familiares, de mulheres e de homens) [27].

Os dados do MDHS 2011 foram usados para o propósito dos estudos nos Capítulos 3 e 4, onde se considerou o subconjunto de mulheres em idade reprodutiva (15 a 49 anos) e fez-se uma junção com aos dados dos AF para obter algumas variáveis relevantes, como a pulverização intradomiciliar, tipo de rede mosquiteira no agregado e o número de redes mosquiteiras por agregado, que não constavam na base de dados das mulheres. A concatenação da base de dados foi feita conforme a Tabela 2.1 abaixo.

2.2. DESCRIÇÃO DAS FONTES DE DADOS E CONSIDERAÇÕES ÉTICAS

Tabela 2.1: Junção das bases de dados.

código da variável	descrição	Tipo de base de dados
V001	número do <i>cluster</i>	mulheres
V002	número do agregado no <i>cluster</i>	mulheres
HV001	número do <i>cluster</i>	agregados
HV002	número do agregado no <i>cluster</i>	agregados
código das variáveis do plano amostral		
V021	unidade primária de amostragem	mulheres
V005	peso amostral	mulheres
V025	Local de residência	agregados

2.2.2 Inquérito de Indicadores de Imunização, Malária e HIV/SIDA 2015

O IMASIDA 2015 é de um estudo transversal, nacional, com base populacional [76]. Este é o segundo inquérito cujo objetivo é fornecer estimativas atualizadas de indicadores demográficos e de saúde, determinar a prevalência do HIV, SIDA, malária e outros indicadores relevantes de saúde da mulher e criança no país.

O IMASIDA 2015 reúne informação de 7,169 AF, tendo sido entrevistadas 7,749 mulheres de 15 a 59 anos e 5,283 homens de 15 a 59 anos, e esta coleta dos dados foi feita entre junho e setembro de 2015 por entrevistas domiciliares, com aplicação de quatro tipos de questionários: de AF, de mulheres de 15-59 anos, de homens de 15-59 anos e outro de biometria para homens e mulheres de 15-59 anos e crianças dos 6-59 meses [76].

A amostra do IMASIDA 2015 foi definida com base no 3º Recenseamento Geral da População e Habitação (RGPH) de Moçambique, conduzido pelo Instituto Nacional de Estatística, em 2007, desenhada de modo a obter estimativas representativas a nível nacional, provincial e de área de residência (urbana e rural) de mulheres e homens de 15 a 59 anos. Com relação ao delineamento amostral, o IMASIDA 2015 seguiu um desenho estratificado e conglomerado em 2 etapas, onde os estratos considerados foram ao nível de província e de área de residência, resultando em 21 estratos. Assim, a primeira etapa de amostragem consistiu na seleção sistemática de 307 UPAs com probabilidade de seleção proporcional ao tamanho da UPA. Na segunda etapa, todos os AF foram enumerados em cada uma das 307 UPAs sendo selecionadas as USAs, os AF (24 AF em cada UPA nas áreas urbanas e rurais).

Este conjunto de dados foi usado no Capítulo 5 para os quais consideramos o subconjunto das crianças menores de 5 anos.

2.2.3 Inquérito Demográfico e de Saúde do Malawi

O MDHS 2015-16 é uma pesquisa transversal, nacional, de base populacional realizada entre dezembro de 2015 e fevereiro de 2016. O MDHS 2015-16 empregou uma amostragem por conglomerados em duas etapas, desenhada para produzir uma amostra representativa

a nível nacional, nível de residência (urbana e rural) ao nível do distrito. A estratificação foi feita em dois níveis: a nível distrital (32 distritos) e em área urbana e rural. Na primeira etapa, com base no Censo da População e habitação do Malawi (MPHC) conduzido em 2008 no Malawi e atualizado com o Censo Geral da Agricultura de 2009 foram selecionadas 850 UPAs, que eram as Área de Enumeração (AE), com uma probabilidade proporcional ao seu tamanho (i.e. o número de AF em cada AE). Destas UPAs, 173 estavam localizadas em áreas urbanas e 677 estavam em áreas rurais. A segunda etapa de amostragem envolveu uma seleção sistemática de um total de 27,516 AF. A taxa de resposta para os AF foi de 99%. A metodologia usada no MDHS 2015-16 pode ser encontrada em detalhe em [82].

Este conjunto de dados foi usado nos Capítulos 5 e 6 para os quais consideramos o subconjunto das crianças menores de 5 anos.

2.2.4 Amostragem

A amostra do DHS foi desenhada de modo a obter estimativas representativas a nível nacional, provincial ou distrital e de área de residência (urbana e rural) de mulheres e de homens em idade reprodutiva. Com relação ao delineamento amostral, o DHS segue um delineamento amostral complexo, ou seja, combinou estratificação, conglomeração em duas etapas, com probabilidades de seleção desiguais resultando em pesos amostrais distintos para os elementos da amostra.

Os estratos considerados foram ao nível de província ou distrito e de área de residência. Assim, a primeira etapa de amostragem consistiu na seleção de UPAs, sendo as AE, probabilidade proporcional ao tamanho (PPS) (o número de AF) em cada estrato dentro de cada província ou distrito. A probabilidade de seleção de uma UPA j no estrato h é dada por:

$$\pi_{1hj} = \frac{n_h \times M_{hj}}{\sum_j M_{hj}} \quad (2.1)$$

onde $j = 1, \dots, N_h$, N_h é o número total de UPAs no estrato h , n_h é o número de UPAs selecionadas no estrato h , M_{hj} é o número de AF dentro da UPA j com base nos dados do censo; $\sum_j M_{hj}$ é o número total de AF do estrato h .

Na segunda etapa, foram selecionadas com base na amostragem sistemática de igual probabilidade as USAs, os AF. Selecionaram-se m_{hj} AF do total de agregados da UPA j do estrato h (M_{hj}), aquando da listagem dos AF, com probabilidade de seleção dada por:

$$\pi_{2i|hj} = \frac{m_{hj}}{M_{hj}^*} \quad (2.2)$$

onde $i = 1, \dots, M_{hj}^*$, M_{hj}^* é o número atualizado de AF da UPA j do estrato h por ocasião da pesquisa. Assim, a probabilidade de selecionar um AF i na UPA j do estrato h é:

$$\pi_{hji} = \pi_{1hj} \times \pi_{2i|hj} \quad (2.3)$$

Por fim, foram recolhidos dados de todas as mulheres e homens em idade reprodutiva que se encontravam em cada agregado selecionado. Como a repartição da amostra não é proporcional, o ponderador de amostragem é o inverso da probabilidade global com eventuais correções das não resposta:

$$w_{hji} = \frac{1}{\pi_{hji}}. \quad (2.4)$$

As ponderações (pesos do desenho) foram ajustadas tanto para a não resposta da unidade (uma unidade amostral não é entrevistada) quanto para a ausência de resposta de um item (não resposta para uma pergunta) de agregados familiares e individuais, para obter as ponderações de amostragem (pesos de amostragem ou também chamados de pesos da pesquisa) para agregados familiares e para os homens e mulheres, respetivamente [42]. A diferença de ponderação de amostragem de agregados familiares e as ponderações de amostragem individuais é introduzida pela ausência de resposta (não resposta) individual.

As ponderações finais são normalizadas para obter um número total de casos não ponderados igual ao número total de casos ponderados utilizando ponderações normalizadas ao nível nacional, para o número total de agregados familiares, mulheres e homens. As ponderações normalizadas são ponderações relativas, válidas para estimação de proporções, índices e taxas, mas não são válidas para estimar totais de população ou para dados agrupados e também para estimação de variâncias. Para mais detalhes consulte [42].

MODELOS DE REGRESSÃO LOGÍSTICA, AJUSTANDO PARA O DESENHO DA PESQUISA

3.1 Introdução

Os Modelos de Regressão Logística têm sido amplamente usados em muitas áreas de pesquisa, nomeadamente nas ciências da saúde, por exemplo com o intuito de estudar os fatores de risco associados às doenças. Inquéritos de base populacional sobre saúde, como os Inquéritos Demográficos e de Saúde (DHS) usados nesta tese, são construídos empregando amostragem complexa, ou seja, amostragem probabilística, estratificada e multietápica, com pesos desiguais nas observações.

Os modelos de regressão logística usuais, formulados para descrever as amostras, são inadequados para analisar dados de pesquisas complexas, e resultariam em estimativas incorretas dos erros-padrão e consequentemente, dos testes de hipóteses. Isto porque o pressuposto de que as observações provêm de AAS, independente e identicamente distribuídas (iid), é quase sempre violado. Assim, alguns ajustes que levam em conta os elementos do desenho da pesquisa são necessários para fazer inferências válidas a partir dos dados da pesquisa [11, 16, 109].

O objetivo deste capítulo é descrever os modelos de regressão logística, e em particular demonstrar como estimá-los ajustando para o delineamento amostral complexo na implementação desta metodologia usando o pacote `survey` do *software* R. Mais especificamente, usou-se os dados do MDHS 2011 para ilustrar como corrigir para o efeito de desenho da amostra no caso particular estimando os fatores de risco associados à probabilidade de uso de redes mosquiteiras. Este capítulo é baseado na seguinte publicação [14] e está relacionado com a seguinte publicação [25].

3.2 Modelo Clássico de Regressão Logística

O modelo de Regressão Logística (RL) é um caso particular dos Modelos Lineares Generalizados (GLMs) e é um modelo adequado para situações em que se pretende estudar a relação entre a variável resposta (*outcome*) binária, Y , sendo geralmente representada por

sucesso ($Y = 1$) ou fracasso ($Y = 0$), e um conjunto de covariáveis $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ [2, 73, 79]. A função massa de probabilidade dada por:

$$f(Y_i = y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \quad (3.1)$$

onde π_i é a probabilidade associada à ocorrência do fenômeno em estudo, $\pi_i = P(Y_i = 1 | \mathbf{x}'_i)$ e $1 - \pi_i = P(Y_i = 0 | \mathbf{x}'_i)$ a probabilidade de fracasso, $i = 1, \dots, n$. O modelo de regressão logística pode ser descrito como [38]:

$$E(Y_i | \mathbf{x}_i) = P(Y_i = 1 | \mathbf{x}_i) = \pi_i = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})}, \quad (3.2)$$

e assim,

$$P(Y_i = 0 | \mathbf{x}_i) = 1 - \pi_i = \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})}, \quad (3.3)$$

onde $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ são parâmetros desconhecidos do modelo a serem estimados. Aplicando o logaritmo na razão de π_i por $(1 - \pi_i)$, tem-se a seguinte função de ligação logit

$$g(\mathbf{x}_i) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \mathbf{x}'_i \boldsymbol{\beta}. \quad (3.4)$$

3.2.1 Estimação dos parâmetros do modelo

Para ajustar o modelo de regressão logística a um conjunto de dados é necessário estimar os valores dos coeficientes $\boldsymbol{\beta}$. A estimação dos parâmetros é feita pelo método de Máxima Verosimilhança (MV), onde se assume que as observações são iid.

Assume-se que a variável resposta Y segue uma distribuição de Bernoulli com função massa de probabilidade dada por (3.1), então, a função de verosimilhança dos dados sob o modelo logístico é:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}. \quad (3.5)$$

Aplicando-se o logaritmo em ambos os lados da equação (3.5), tem-se a expressão:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^n \left[y_i \ln \pi_i + (1 - y_i) \ln(1 - \pi_i) \right] \\ &= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \\ &= \sum_{i=1}^n \left[y_i \mathbf{x}'_i \boldsymbol{\beta} - \ln[1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})] \right]. \end{aligned} \quad (3.6)$$

Para determinar os valores de $\boldsymbol{\beta}$ que maximizam $l(\boldsymbol{\beta})$, deriva-se a função (3.6) em relação aos elementos de $\boldsymbol{\beta}$ e iguala-se a zero, e assim obtêm-se as equações de log-verosimilhança que se seguem:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} = 0 \quad (3.7)$$

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} = 0. \quad (3.8)$$

Os valores de β que solucionam as equações (3.7) e (3.8) são os estimadores de MV dos parâmetros. Como as equações (3.7) e (3.8) são não lineares, e não existe uma solução direta, para a sua solução recorre-se a métodos iterativos, como os de Newton-Raphson [2, 20, 79, 115].

Um estimador consistente da matriz de variância-covariância do vetor de estimativas $\hat{\beta}$ é dado por:

$$\widehat{Var}(\hat{\beta}) = (X' \hat{V} X)^{-1}, \quad (3.9)$$

onde V é a matriz diagonal $n \times n$ com elemento $\hat{\pi}_i(1 - \hat{\pi}_i)$, X é a matriz do modelo [2].

3.2.2 Testes de hipóteses

Aquando do ajuste do modelo é de interesse saber se as covariáveis consideradas são relevantes ou não na análise, sendo os testes mais usados para o efeito, o teste de razão de verosimilhanças e o teste de Wald.

3.2.2.1 Teste de Razão de Verosimilhanças

O teste de razão de verosimilhanças pode ser utilizado para avaliar a significância global dos parâmetros e é baseado na função de verosimilhança. Este teste consiste em comparar a verosimilhança de um modelo, M_1 , com a verosimilhança de outro modelo, M_2 , que lhe está encaixado. Para tal é necessário utilizar a chamada *deviance* ou função desvio, D , que tem uma distribuição aproximada Qui-Quadrado e é dada por:

$$D = -2 \ln \left[\frac{\text{verosimilhança do modelo ajustado, } M_1}{\text{verosimilhança do modelo saturado, } M_2} \right]. \quad (3.10)$$

Assim, sob a hipótese $H_0 : \beta_1 = \dots = \beta_p = 0$ de que os p coeficientes das covariáveis no modelo são nulos, a estatística de teste é dada por:

$$G = D(\text{modelo sem as } p \text{ variáveis, } M_2) - D(\text{modelo com as } p \text{ variáveis, } M_1), \quad (3.11)$$

onde G segue uma distribuição Qui-Quadrado com $p - 1$ graus de liberdade. A rejeição da hipótese nula significa que pelo menos um dos coeficientes é estatisticamente diferente de zero.

3.2.2.2 Teste de Wald

O teste de Wald avalia a significância dos parâmetros, ou seja, se cada coeficiente é significativamente diferente de zero, onde a hipótese nula é

$$H_0 : \beta_j = 0, \quad \text{versus} \quad H_1 : \beta_j \neq 0, \quad j = 1, \dots, p$$

que indica que a variável explicativa X_j não é relevante para o modelo ou que β_j não é estatisticamente significativo. A estatística do teste é dada por:

$$W = \hat{\beta}' [\widehat{Var}(\hat{\beta})]^{-1} \hat{\beta} = \hat{\beta}' (X \hat{V} X) \hat{\beta}, \quad (3.12)$$

onde $\hat{\beta}$ denota o vetor dos p coeficientes e $\widehat{Var}(\hat{\beta})$ a sub-matriz $p \times p$ obtida da matriz completa $(p+1) \times (p+1)$ da equação (3.9). O valor-p é encontrado usando uma distribuição Qui-Quadrado com p graus de liberdade como valor- $p = P[\chi^2(p) \geq W]$.

3.2.3 Avaliação da qualidade do ajuste e da seleção do modelo

Depois de ajustar um modelo aos dados, é necessária a avaliação da qualidade do ajuste do modelo, ou seja, a capacidade do modelo para representar os dados observados. Uma das estatísticas que é utilizada para o efeito é a *deviance*, D , cujo valor se obtém a partir da expressão (3.10), onde se compara a verosimilhança do modelo ajustado com a verosimilhança do modelo saturado, ou seja, do modelo que se obtém substituindo $E[Y_i|x_i] = \pi_i$ pela correspondente observação y_i [38].

Um valor pequeno obtido da *deviance* indica que, para um número menor de parâmetros, obtemos um ajuste tão bom quanto o ajuste com o modelo saturado.

Outro método usado para a escolha adequada do melhor modelo é o critério de Informação de Akaike (AIC) que é, na verdade, valores para os logaritmos das funções de log-verosimilhança penalizadas pelo número de parâmetros estimados. Outra vantagem do uso deste critério é que eles podem ser usados para comparar modelos que não são aninhados. O AIC é dado por [38]:

$$AIC = -2l(\beta) + 2p$$

onde l é o logaritmo da função de verosimilhança maximizada sob o modelo, p é o número de parâmetros estimados pelo modelo. Quanto menor for o valor obtido do AIC, melhor será o ajuste do modelo.

3.3 Regressão Logística para amostras complexas

De acordo com [38, 55], o modelo de regressão logística atrás apresentado é inapropriado para produzir estimativas de parâmetros válidas quando os dados são provenientes de AC, i.e., dados onde está presente a estratificação, conglomeração e com pesos amostrais desiguais em cada observação. Isto porque o pressuposto de que as observações provêm de AAS, iid, é normalmente violado.

A estimação dos parâmetros por MV não é mais direta por razões como probabilidade desiguais de seleção das observações e exigindo a incorporação de pesos amostrais. Mais ainda, a estratificação e a conglomeração das observações viola a suposição de iid que é crucial para a abordagem MV padrão para estimar a variância amostral dos parâmetros do modelo.

Assim, a estimação por MV levaria a estimativas incorretas dos erros-padrão e consequentemente, dos testes de hipóteses. É necessário fazer alguns ajustes nos métodos da regressão logística usual para obter inferências válidas [38, 55, 97].

A regressão logística baseada no desenho da pesquisa permite a incorporação das informações do desenho complexo da amostra incluindo a estratificação, conglomeração e os pesos desiguais de seleção.

Supondo-se que a população possa ser dividida em $h = 1, 2, \dots, H$ estratos, $j = 1, 2, \dots, n_h$ UPAs em cada estrato e $i = 1, 2, \dots, n_{hj}$ elementos na UPA j do estrato h ; Supondo-se também que os dados observados referem-se a n_{hj} elementos das n_h UPAs do estrato h . O número total de observações é dado por $n = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} n_{hji}$, designem-se os pesos amostrais conhecidos da hji -ésima observação por $w_{hji} = \frac{1}{n_{hji}}$, o vetor de covariáveis $\mathbf{x}_{hji}$ e a variável resposta binária por y_{hji} .

$$\text{logit}(\pi_{hji}) = \ln \left(\frac{\pi_{hji}}{1 - \pi_{hji}} \right) = \mathbf{x}'_{hji} \boldsymbol{\beta}. \quad (3.13)$$

Sob a hipótese de AC, o vetor $\boldsymbol{\beta}$ para o modelo de RL são estimados pelo método de Máxima Pseudo-Verosimilhança (MPV), ou também designado por método de máxima verosimilhança ponderada, que incorpora o delineamento amostral e os pesos distintos na estimação dos $\boldsymbol{\beta}$ [4, 38, 55, 62, 103].

A ideia principal do método de MPV é definir uma função que aproxima a função de verosimilhança da população finita amostrada com uma função de verosimilhança formada pela amostra observada e os pesos amostrais conhecidos.

Assim, a função de MPV e log MPV são dadas por:

$$L_w(\boldsymbol{\beta}) = \prod_{h=1}^H \prod_{j=1}^{n_h} \prod_{i=1}^{n_{hj}} \pi(\mathbf{x}_{hji})^{w_{hji} \times y_{hji}} (1 - \pi(\mathbf{x}_{hji}))^{w_{hji} \times (1 - y_{hji})}, \quad (3.14)$$

$$l_w(\boldsymbol{\beta}) = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} w_{hji} \{ y_{hji} \times \ln[\pi(\mathbf{x}_{hji})] + (1 - y_{hji}) \times \ln[1 - \pi(\mathbf{x}_{hji})] \}. \quad (3.15)$$

Substituindo o valor de π_{hji} na função (3.15) obtém-se

$$l_w(\boldsymbol{\beta}) = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} w_{hji} \times \left[y_{hji} \times \ln \left(\frac{e^{\mathbf{x}'_{hji} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{hji} \boldsymbol{\beta}}} \right) + (1 - y_{hji}) \times \ln \left(1 - \frac{e^{\mathbf{x}'_{hji} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{hji} \boldsymbol{\beta}}} \right) \right]. \quad (3.16)$$

Os coeficientes de regressão e os seus estimadores de variância-covariância podem ser obtidos a partir das equações score $\frac{\partial l_w(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$ e $\text{Cov}(\boldsymbol{\beta})_w = I_w^{-1}(\boldsymbol{\beta}) = -E \left(\frac{\partial^2 l_w(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right)$.

Derivando a função de pseudo log-verosimilhança em relação a $\boldsymbol{\beta}$ e igualando a zero, obtém-se o estimador de MPV dado pela solução das seguintes equações:

$$\frac{\partial l_w(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0} = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} w_{hji} \left[y_{hji} - \frac{\exp^{\mathbf{x}'_{hji} \boldsymbol{\beta}}}{1 + \exp^{\mathbf{x}'_{hji} \boldsymbol{\beta}}} \right] = 0 \quad (3.17)$$

$$\frac{\partial l_w(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} w_{hji} \left[y_{hji} x_{hji} - x_{hji} \frac{\exp^{x'_{hji} \boldsymbol{\beta}}}{1 + \exp^{x'_{hji} \boldsymbol{\beta}}} \right] = 0 \quad (3.18)$$

onde w_{hji} é o peso amostral.

No contexto de delineamentos amostrais complexos, não existe uma forma analítica direta para se calcular as variâncias destes estimadores. Assim sendo, para a obtenção dos estimadores das variâncias de MPV recorre-se a métodos, tais como a linearização de Taylor (também chamado como método delta) e os métodos de replicação de Jackknife e bootstrap [55]. Neste trabalho, utilizou-se o método de linearização de Taylor que é o método mais frequentemente usado e que está implementado no pacote *survey* do *software* R [62].

A partir do método de linearização de Taylor, obtém-se a seguinte expressão da variância do estimador de MPV de $\boldsymbol{\beta}$ [38, 62]:

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1} \mathbf{S} (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1}, \quad (3.19)$$

onde $\mathbf{D}$ é a matriz diagonal $n \times n$ com elemento $w_{hji} \times \hat{\pi}_{hji}(1 - \hat{\pi}_{hji})$, $\mathbf{S}$ é o estimador agrupado intra-estrato da matriz de covariâncias (3.19). Este estimador é dado por:

$$\mathbf{S} = \sum_{h=1}^H (1 - f_h) \frac{n_h}{n_h - 1} \sum_{j=1}^{n_h} (\mathbf{z}_{hj} - \bar{\mathbf{z}}_{h..})(\mathbf{z}_{hj} - \bar{\mathbf{z}}_{h..})', \quad (3.20)$$

onde $\mathbf{z}'_{hji} = \mathbf{x}'_{hji} w_{hji} (y_{hji} - \hat{\pi}_{hji}(x_{hji}))$, o somatório para todas as n_{hj} unidades selecionadas na UPA j do estrato h como $\mathbf{z}_{hj} = \sum_{i=1}^{n_{hj}} \mathbf{z}_{hji}$ e a sua média específica do estrato como $\bar{\mathbf{z}}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} \mathbf{z}_{hj}$. O fator de correção é dado por $(1 - f_h)$, onde f_h é o quociente do número de UPAs observadas (n_h) pelo número total de UPAs no estrato h (N_h).

Do mesmo modo que o método de estimação dos parâmetros do modelo sofre adaptações de forma a incorporar o plano amostral complexo na análise, os procedimentos dos testes de hipóteses usuais precisam ser ajustados. Assim, para obter inferências válidas usando este tipo de delineamento, introduzem-se correções na estatística de teste de Pearson, tais como os ajustes de Rao-Scott, ou alternativamente usam-se outras estatísticas de teste que já incorporem o plano amostral, tais como a estatística de Wald ajustada [38, 55]. A estatística de Wald ajustada é dada por [38, 55, 64]:

$$F = \frac{s - p + 1}{sp} W \quad (3.21)$$

onde $s = (\sum_{h=1}^H n_h) - H$ é o número total de UPAs selecionadas menos o número de estratos. A estatística de Wald ajustada definida em (3.21) tem como distribuição aproximada a distribuição F com p e $(s - p + 1)$ graus de liberdade. O valor- p é determinado da seguinte forma $\text{valor-}p = P[F(p, s - p + 1) \geq F_{obs}]$.

Conforme [64], as estatísticas envolvidas nos testes de razão de verossimilhanças não têm distribuição χ^2 . As suas distribuições têm de ser aproximadas e para isso têm se usado as aproximações de Rao e Scott.

Para avaliar a qualidade do ajuste dos modelos de regressão logística para dados recolhidos com base em delineamentos amostrais complexos, o conhecido teste de Hosmer-Lemeshow foi adaptado por [4].

Ainda a seleção de modelos através de critérios de seleção de modelos tais como os critério de Informação de Akaike (AIC) e BIC foram recentemente adaptados para dados de delineamentos amostrais complexos por [63].

3.4 Aplicação para os fatores que contribuem para o uso de redes mosquiteiras por mulheres em idade reprodutiva em Moçambique

3.4.1 Variáveis do estudo

A variável cujo comportamento se pretende explicar é aquela que representa se a rede mosquiteira é usada para dormir e, para tal, perguntou-se a pessoa entrevistada se dormiu ou não com a rede mosquiteira na noite passada (V461). Assim, o uso da rede mosquiteira (1=usa, 0=não usa) foi a variável de interesse. As possíveis variáveis preditoras estão apresentadas na tabela que se segue (Tabela 3.1).

Tabela 3.1: Variáveis do estudo.

variável (código da variável na base de dados)	descrição
Idade (V013)	(1) 15-19, (2) 20-24, (3) 25-29, (4) 30-34, (5) 35-39, (6) 40-44, (7) 45-49
Quinto de Riqueza (V190)	(1) mais baixo, (2) segundo, (3) médio, (4) quarto, (5) mais elevado
Área de residência (V025)	(1) urbana, (2) rural
Província (V024)	(1) Niassa, (2) Cabo Delgado, (3) Nampula, (4) Zambézia, (5) Tete, (6) Manica, (7) Sofala, (8) Inhambane, (9) Gaza, (10) Maputo Província, (11) Maputo Cidade
Nível de escolaridade (V106)	(1) analfabeta, (2) primário, (3) secundário, (4) superior
N. de membros do agregado familiar (V136)	-
Sexo do chefe do agregado (V130)	(1) masculino, (2) feminino
Estado civil (V501)	(1) solteira, (2) casada, (3) vive maritalmente, (4) viúva, (5) divorciada, (6) separada
Se está grávida (V213)	(1) não, (2) sim
Se está a trabalhar (V714)	(1) não, (2) sim
Posse de rede mosquiteira (V459)	(0) não, (1) sim
Número de redes mosquiteiras (HML1)	-
Tipo de rede mosquiteira que usou (ML101)	(1) não tem rede, (2) somente redes tratadas, (3) redes tratadas e não tratadas, (4) somente redes não tratadas
Pulverização intradomiciliar (HV253) nos últimos 12 meses	(1) não, (2) sim, (3) não sei

3.4.2 Análise dos dados

Para propósitos deste estudo, usou-se a base de dados do MDHS 2011 referentes as mulheres e aos AF conforme descrito na secção 2.2.1. Para investigar alguns determinantes do uso da rede mosquiteira, e identificar aqueles que estão fortemente associados com as chances das mulheres usarem a rede mosquiteira, foram ajustados modelos de regressão logística, considerando as covariáveis apresentadas na tabela 3.1 na secção anterior. Na análise usou-se o teste de Pearson com o ajuste de Rao-Scott para testar as possíveis associações entre as covariáveis e a variável de interesse. O modelo final considerou apenas as variáveis que estavam associadas com o uso da rede mosquiteira ao nível de significância de 5%. Os resultados da análise de regressão logística são apresentados como coeficientes com intervalo de confiança de 95% (IC). Após o ajuste, pelo teste de Wald foram calculados e ajustados os *Adjusted Odds Ratio* ou Razão de Chances Ajustada (aOR).

As análises ponderadas foram feitas com auxílio do pacote *survey* [61, 62] do *software* R [113], onde as características do delineamento amostral, como estratificação, conglomeração e peso amostral, são incorporadas na análise usando a função `svydesign()`. Modelos de regressão logística foram ajustados com auxílio da função `svyglm()`, onde estimativas

3.4. APLICAÇÃO PARA OS FATORES QUE CONTRIBUEM PARA O USO DE REDES MOSQUITEIRAS POR MULHERES EM IDADE REPRODUTIVA EM MOÇAMBIQUE

pontuais e de variância considerando o delineamento foram extraídas. A qualidade do ajuste do modelo foi feita conforme [62].

3.4.3 Resultados

3.4.3.1 Análise Descritiva

A amostra do estudo foi constituída por 13,745 mulheres em idade reprodutiva (15 a 49 anos), em Moçambique no ano de 2011. As características sócio-económicas e demográficas da população de estudo levando ou não em consideração o plano amostral são sumarizadas na Tabela 3.2 que se segue. Os resultados, em geral, mostram que as estimativas pontuais de proporção são semelhantes comparando-se a análise ponderada e não ponderada.

Tabela 3.2: Características sócio-económicas e demográficas das mulheres e idade reprodutiva em Moçambique, usando os dados do MDHS 2011.

Variáveis	Não ponderado			Ponderado		
	n de mulheres			n de mulheres		
	frequência (%)	não (%)	sim (%)	frequência (%)	não (%)	sim (%)
Total	13,745 (100.0)	8,739 (63.6)	5,006 (36.4)	13,745 (100.0)	8,473 (61.6)	5,272 (38.4)
Idade						
15-19	3,065 (22.3)	2,141 (69.9)	924 (30.2)	3,061 (22.3)	2,087 (68.2)	973 (31.8)
20-24	2,468 (18.0)	1,492 (60.5)	976 (39.6)	2,454 (17.9)	1,378 (56.2)	1,076 (43.8)
25-29	2,340 (17.0)	1,396 (59.7)	944 (40.3)	2,275 (16.6)	1,323 (58.2)	952 (41.8)
30-34	1,975 (14.4)	1,210 (61.3)	765 (38.7)	1,997 (14.5)	1,207 (60.4)	790 (39.6)
35-39	1,691 (12.3)	1,031 (61.0)	660 (39.0)	1,698 (12.4)	999 (58.8)	699 (41.2)
40-44	1,156 (8.4)	758 (65.6)	398 (34.4)	1,159 (8.4)	737 (63.6)	422 (36.4)
45-49	1,050 (7.6)	711 (67.7)	339 (32.3)	1,101 (8.0)	742 (67.4)	359 (32.6)
Estado Civil						
solteira	2,852 (20.8)	2,036 (71.4)	816 (28.6)	2,514 (18.3)	1,771 (70.4)	743 (29.6)
casada	5,373 (39.1)	3,026 (56.3)	2,347 (43.7)	6,136 (44.6)	3,426 (55.8)	2,710 (44.2)
vive maritalmente	3,583 (26.1)	2,338 (65.3)	1,245 (34.8)	3,195 (23.2)	1,991 (62.3)	1,204 (37.7)
viúva	570 (4.2)	417 (73.2)	153 (26.8)	516 (3.8)	379 (73.4)	137 (26.6)
divorciada	236 (1.7)	153 (64.8)	83 (35.2)	307 (2.2)	200 (65.1)	107 (34.9)
separada	1,131 (8.2)	769 (68.0)	362 (32.01)	1,076 (7.8)	705 (65.5)	371 (34.5)
Província						
Niassa	927 (6.7)	493 (53.2)	434 (46.8)	681 (5.0)	366 (53.7)	315 (46.3)
Cabo Delgado	1,047 (7.6)	467 (44.6)	580 (55.4)	1,002 (7.3)	449 (44.9)	552 (55.1)
Nampula	980 (7.1)	443 (45.2)	537 (54.8)	1,926 (14.0)	887 (46.1)	1,039 (53.9)
Zambézia	1,330 (9.7)	876 (65.9)	454 (34.1)	2,532 (18.4)	1,670 (66.0)	862 (34.0)
Tete	1,126 (8.2)	739 (65.6)	387 (34.4)	1,608 (11.7)	1,060 (66.0)	547 (34.0)
Manica	1,174 (8.5)	743 (63.3)	431 (36.7)	951 (6.9)	580 (60.9)	371 (39.1)
Sofala	1,615 (11.8)	944 (58.5)	671 (41.6)	1,412 (10.3)	822 (58.2)	590 (41.8)
Inhambane	1,139 (8.3)	793 (69.6)	346 (30.4)	872 (6.3)	599 (68.7)	273 (31.3)
Gaza	1,259 (9.2)	1,126 (89.4)	133 (10.6)	813 (5.9)	732 (90.0)	81 (10.0)
Maputo Província	1,424 (10.4)	997 (70.0)	427 (30.0)	1,061 (7.7)	735 (69.3)	326 (30.7)
Maputo Cidade	1,724 (12.5)	1,118 (64.9)	606 (35.2)	888 (6.5)	573 (64.5)	315 (35.5)
(continua)						

CAPÍTULO 3. MODELOS DE REGRESSÃO LOGÍSTICA, AJUSTANDO PARA O DESENHO DA PESQUISA

Tabela 3.2 (continuação)

Nível de escolaridade						
Sem escolaridade	3,773 (27.5)	2,590 (68.7)	1,183 (31.4)	4,292 (31.2)	2,911 (67.8)	1,381 (32.2)
Primário	6,774 (49.3)	4,301 (63.5)	2,473 (36.5)	6,906 (50.2)	4,203 (60.9)	2,702 (39.1)
Secundário	2,943 (21.4)	1,676 (57.0)	1,267 (43.1)	2,362 (17.2)	1,242 (52.6)	1,120 (47.4)
Superior	255 (1.9)	172 (67.5)	83 (32.6)	185 (1.4)	117 (63.2)	68 (36.8)
Área de residência						
Urbana	5,804 (42.2)	3,336 (57.5)	2,468 (42.5)	4,773 (34.7)	2,587 (54.2)	2,186 (45.8)
Rural	7,941 (57.8)	5,403 (68.0)	2,538 (32.0)	8,972 (65.3)	5,886 (65.6)	3,086 (34.4)
Quinto de riqueza						
Mais baixo	1,833 (13.3)	1,207 (65.9)	626 (34.2)	2,597 (18.9)	1,738 (66.9)	859 (33.1)
Segundo	2,109 (15.3)	1,376 (65.2)	733 (34.8)	2,551 (18.6)	1,640 (64.3)	911 (35.7)
Médio	2,399 (17.5)	1,532 (63.9)	867 (36.1)	2,575 (18.7)	1,586 (61.8)	989 (38.4)
Quarto	2,946 (21.4)	1,955 (66.4)	991 (33.6)	2,783 (20.2)	1,733 (62.3)	1,050 (37.7)
Mais elevado	4,458 (32.4)	2,669 (59.9)	1,789 (40.1)	3,239 (23.6)	1,777 (54.9)	1,462 (45.1)
Se está a trabalhar						
não	8,454 (61.5)	5,45 (64.5)	3,000 (35.5)	8,335 (60.6)	5,268 (63.2)	3,067 (36.8)
sim	5,291 (38.5)	3,285 (62.1)	2,006 (38.0)	5,410 (39.4)	3,205 (59.3)	2,205 (40.7)
Se está grávida						
não / não tem certeza	12,336 (89.75)	7,842 (63.6)	4,494 (36.4)	12,229 (89.0)	7,511 (61.4)	4,718 (38.6)
sim	1,409 (10.25)	897 (63.7)	512 (36.3)	1,516 (11.0)	962 (63.5)	554 (36.5)

Analisando a Tabela 3.2, das características sócio-demográficas, constata-se que as mulheres são maioritariamente casadas ou vivem maritalmente (67.8%), seguindo-se as solteiras (18.3%); vivem predominante na área rural (65.3%); a grande maioria tem apenas o nível primário (50.2%), 31.2% são analfabetas e apenas 1.3% tem o nível superior; cerca de 11% das mulheres relataram estar grávidas; 61% das mulheres não estavam a trabalhar no momento. Em relação ao AF, constata-se que a maior parte das mulheres referiu que o seu agregado era chefiado por homens (64.7%).

Em relação ao grupo etário, pode notar-se que mulheres dos 20 aos 24 anos e dos 25 aos 29 anos apresentaram maior taxa do uso da rede mosquiteira (43.8% e 41.8% respetivamente). As mulheres com nível secundário apresentam maior taxa no uso da rede mosquiteira (47.4%). O estado civil das mulheres também é importante no uso da rede mosquiteira pois foi maior para as mulheres casadas e as mulheres solteiras apresentaram menor taxa; as mulheres trabalhadoras mostram maior uso de redes.

No que concerne a regiões geográficas, constatou-se que as diferenças geográficas também foram bastante substanciais: o uso da rede mosquiteira é menor nas zonas rurais (34.4%), o uso da rede mosquiteira é elevado em algumas áreas no norte do país (Nampula e Zambézia), enquanto que em Gaza é consideravelmente mais baixo (10.0%) (Tabela 3.2).

Quanto a posse de rede mosquiteira (Tabela 3.3) constata-se que a proporção de mulheres que possuíam a rede mosquiteira na noite anterior foi de 61.3%, contudo 38.4% das mulheres é que usaram a rede mosquiteira para dormir na noite anterior. Quanto à pulverização intradomiciliar, 22.6% das mulheres relataram que a sua casa foi pulverizada nos últimos 12 meses.

3.4. APLICAÇÃO PARA OS FATORES QUE CONTRIBUEM PARA O USO DE REDES MOSQUITEIRAS POR MULHERES EM IDADE REPRODUTIVA EM MOÇAMBIQUE

Tabela 3.3: Disponibilidade e uso da rede mosquiteira e pulverização intradomiciliar.

	Não ponderada frequência (%)	Ponderada frequência (%)
Posse de rede mosquiteira		
não tem rede	5,317 (38.7)	5,316 (38.7)
tem rede	8,428 (61.3)	8,429 (61.3)
Tipo de rede mosquiteira que usou para dormir na noite passada		
não usou rede	8,739 (63.6)	8,473 (61.4)
somente redes tratadas	4,333 (31.5)	3,518 (25.6)
redes tratadas e não tratadas	2 (0.0)	1 (0.0)
somente redes não tratadas	671 (4.9)	566 (4.1)
Pulverização intradomiciliar nos últimos 12 meses		
não	10,234 (74.5)	10,515 (76.5)
sim	3,352 (24.4)	3,109 (22.6)
não sei	159 (1.2)	121 (0.9)
Total	13,745 (100.0)	13,745 (100.0)

3.4.3.2 Ajuste do Modelo de Regressão Logística

No ajuste do modelo de regressão logística inicial foram incluídas todas variáveis citadas na seção 3.4.1. Em seguida reteve-se apenas as variáveis que foram estatisticamente significativas a nível de 5%. Esta estratégia foi confirmada usando testes de razão de verossimilhança (RV) para amostragem complexa. Do teste de independência Rao-Scott, para estudar a associação entre a variável de interesse (usar a rede mosquiteira) e as covariáveis, verifica-se que existe uma associação significativa entre as variáveis idade agrupada, província, área de residência e nível de escolaridade, ao nível de significância de 1%. Do teste RV ($p=0.36095$) verifica-se a não existência de evidências suficientes para manter as variáveis que foram excluídas do modelo, ou seja, estas variáveis em conjunto não afetam de forma significativa o *outcome* de interesse. Adicionalmente, foram feitos vários testes de RV do modelo inicial retirando cada uma das variáveis individualmente não significativas. Os resultados (valores- p) desses testes individuais para as variáveis sexo do chefe do agregado, se está a trabalhar, pulverização intradomiciliar, foram, respetivamente, 0.17907, 0.31101, 0.55253, comprovando-se mais uma vez que não há evidência de que as variáveis contribuem para o modelo.

As Tabelas 3.4 e 3.5 apresentam os resultados do ajuste do modelo final de regressão logística padrão e do modelo final de regressão logística levando em conta o delineamento amostral, respetivamente. É de realçar que as categorias que estão com um "Ref" na descrição dos parâmetros representam as categorias de referência das variáveis categóricas. Os resultados, em geral, mostram que as estimativas dos coeficientes da regressão logística tendo em conta o plano amostral são semelhantes às estimativas não ponderadas. Os erros padrões das estimativas ficam subestimados quando as características do plano amostral não são consideradas na estimação. Isto pode fazer com que os coeficientes estimados

CAPÍTULO 3. MODELOS DE REGRESSÃO LOGÍSTICA, AJUSTANDO PARA O DESENHO DA PESQUISA

para modelos de regressão, por exemplo, sejam considerados significativos quando na verdade não são.

O modelo final tanto para a regressão logística padrão quanto para a regressão logística tendo em conta o desenho da pesquisa inclui: idade agrupada, província, área de residência, nível de escolaridade, quinto de riqueza, se está grávida, número de membros do agregado e número de redes mosquiteiras no agregado.

Tabela 3.4: Modelo logit final Não Ponderado.

Variáveis	Coeficiente	Erro padrão	IC de 95%	valor-p	AOR
<i>intercept</i>	-0.832	0.150	[-1.125;-0.538]	0.000	0.44
Idade					
15-19 (Ref)	-	-	-	-	-
20-24	0.432	0.078	[0.279;0.585]	0.000	1.54
25-29	0.475	0.082	[0.314;0.637]	0.000	1.61
30-34	0.427	0.089	[0.252;0.601]	0.000	1.53
35-39	0.434	0.094	[0.259;0.627]	0.000	1.56
40-44	0.243	0.106	[-0.035;0.451]	0.022	1.28
45-49	0.279	0.110	[-0.063;0.496]	0.001	1.32
Estado Civil					
solteira (Ref)	-	-	-	-	-
casada	0.753	0.084	[0.588;0.918]	0.000	0.71
vive maritalmente	0.711	0.084	[0.546;0.875]	0.000	1.76
viúva	0.279	0.147	[-0.008;0.566]	0.057	1.42
divorciada	0.211	0.186	[-0.154;0.575]	0.258	1.34
separada	0.150	0.110	[-0.065;0.365]	0.172	1.31
Província					
Niassa (Ref)	-	-	-	-	-
Cabo Delgado	-0.339	0.114	[-0.561;-0.116]	0.003	
Nampula	-0.027	0.115	[-0.252;0.198]	0.815	0.97
Zambézia	-0.333	0.112	[-0.553;-0.113]	0.003	0.72
Tete	-0.658	0.117	[-0.887;-0.429]	0.000	0.52
Manica	-0.483	0.113	[-0.705;-0.262]	0.000	0.62
Sofala	-0.396	0.109	[-0.610;-0.182]	0.000	0.67
Inhambane	-1.92	0.130	[-2.175;-1.665]	0.000	0.15
Gaza	-2.888	0.150	[-3.181;-2.594]	0.000	0.06
Maputo Província	-1.546	0.130	[-1.800;-1.291]	0.000	0.21
Maputo Cidade	-1.090	0.128	[-1.342;-0.838]	0.000	0.34
Área de residência					
Urbana (Ref)	-	-	-	-	-
Rural	-0.512	0.068	[-0.645;0.378]	0.000	0.60
Nível de escolaridade					
Sem escolaridade (Ref)	-	-	-	-	-
Primário	0.278	0.059	[0.162;0.393]	0.000	1.32
Secundário	0.360	0.087	[0.190;0.530]	0.000	1.43

(continua)

3.4. APLICAÇÃO PARA OS FATORES QUE CONTRIBUEM PARA O USO DE REDES MOSQUITEIRAS POR MULHERES EM IDADE REPRODUTIVA EM MOÇAMBIQUE

Tabela 3.4 (continuação)

Variáveis	Coeficiente	Erro padrão	IC de 95%	valor-p	AOR
Superior	-0.680	0.200	[-1.072;-0.288]	0.001	0.51
Quinto de riqueza					
Pobre (Ref)	-	-	-	-	-
Segundo	0.0105	0.084	[-0.153;0.174]	0.900	1.01
Médio	0.202	0.084	[0.037;0.367]	0.016	1.22
Quarto	0.0341	0.092	[-0.146;0.214]	0.710	1.03
mais elevado	0.143	0.115	[-0.082;0.368]	0.213	1.15
Se está grávida					
não (Ref)	-	-	-	-	-
sim	-0.272	0.075	[-0.419;-0.125]	0.000	0.76
número de membros	-0.279	0.010	[-0.299;-0.259]	0.000	0.76
número de redes	1.300	0.025	[1.252;1.349]	0.000	3.67

Tabela 3.5: Modelo logit final Ponderado.

Variáveis	Coeficiente	Erro padrão	IC de 95%	valor-p	AOR
<i>intercept</i>	-1.108	0.196	[-1.494;-0.722]	0.000	0.33
Idade					
15-19 (R)	-	-	-	-	-
20-24	0.488	0.092	[0.308;0.668]	0.000	1.63
25-29	0.382	0.102	[0.181;0.582]	0.000	1.46
30-34	0.320	0.110	[0.104;0.536]	0.004	1.38
35-39	0.439	0.114	[0.215;0.664]	0.000	1.55
40-44	0.227	0.129	[-0.027;0.480]	0.080	1.25
45-49	0.248	0.29	[-0.005;0.500]	0.054	1.28
Estado Civil					
solteira (R)	-	-	-	-	-
casada	0.823	0.104	[0.619;1.027]	0.000	1.83
vive maritalmente	0.836	0.116	[0.608;1.064]	0.000	1.76
viúva	0.318	0.162	[-0.001;0.637]	0.051	1.42
divorciada	0.316	0.213	[-0.103;0.735]	0.139	1.34
separada	0.306	0.138	[0.035;0.578]	0.027	1.31
Província					
Niassa (R)	-	-	-	-	-
Cabo Delgado	-0.385	0.157	[-0.693;-0.076]	0.015	0.68
Nampula	-0.156	0.126	[-0.402;0.091]	0.216	0.86
Zambézia	-0.417	0.133	[-0.679;-0.155]	0.002	0.66
Tete	-0.715	0.165	[-1.039;-0.392]	0.000	0.49
Manica	-0.522	0.128	[-0.773;-0.271]	0.000	0.59
Sofala	-0.520	0.135	[-0.785;-0.256]	0.000	0.59

(continua)

CAPÍTULO 3. MODELOS DE REGRESSÃO LOGÍSTICA, AJUSTANDO PARA O DESENHO DA PESQUISA

Tabela 3.5 (continuação)

Variáveis	Coefficiente	Erro padrão	valor de t	P(> t)	OR
Inhambane	-2.108	0.253	[-2.605;-1.611]	0.000	0.12
Gaza	-3.211	0.306	[-3.812;-2.610]	0.000	0.04
Maputo Província	-1.700	0.177	[-2.047;-1.354]	0.000	0.18
Maputo Cidade	-1.171	0.176	[-1.516;-0.826]	0.000	0.31
Área de residência					
Urbana (R)	-	-	-	-	-
Rural	-0.356	0.095	[-0.542;0.170]	0.000	0.70
Nível de escolaridade					
Sem escolaridade (R)	-	-	-	-	-
Primário	0.310	0.075	[0.164;0.457]	0.000	1.36
Secundário	0.421	0.101	[0.222;0.620]	0.000	1.52
Superior	-0.628	0.202	[-1.025;-0.231]	0.002	0.53
Quinto de riqueza					
Pobre (R)	-	-	-	-	-
Segundo	0.052	0.104	[-0.152;0.257]	0.615	1.05
Médio	0.210	0.107	[0.001;0.420]	0.049	1.23
Quarto	0.070	0.130	[-0.185;0.326]	0.589	1.07
mais elevado	0.243	0.165	[-0.082;0.567]	0.142	1.27
Se está grávida					
não (R)	-	-	-	-	-
sim	-0.316	0.084	[-0.481;-0.151]	0.000	0.73
número de membros	-0.283	0.015	[-0.313;-0.254]	0.000	0.75
número de redes	1.440	0.051	[1.339;1.539]	0.000	4.22

Pela análise dos *odds ratios* ajustados estimados, ajustando para outras variáveis, mulheres mais velhas têm uma grande chance de usar a rede mosquiteira; por exemplo, mulheres de 20 a 24 anos têm mais 63% de chance de usar a rede mosquiteira (OR=1.63) quando comparadas com mulheres dos 15 aos 19 anos.

As mulheres das províncias do sul do país (Inhambane, Gaza, Maputo Província e Maputo Cidade) são menos propensas a usar a rede mosquiteira quando comparadas com as da província do Niassa. Mulheres da área rural têm menos 30% (OR=0.70) de chance de usar a rede mosquiteira, quando comparadas com as mulheres da área urbana.

Mulheres casadas, que vivem maritalmente, as viúvas, as divorciadas e as separadas têm maior chance de usar a rede mosquiteira em relação às mulheres solteiras, OR= 1.83, 1.76, 1.42, 1.34, 1.31, respetivamente.

Mulheres com o nível superior têm menos 47% (OR=0.53) de chance de usar a rede mosquiteira quando comparadas com mulheres sem escolaridade, ajustando para outras variáveis. Mulheres com quinto de riqueza mais elevado apresentam maior chance de usar a rede mosquiteira (OR>1) quando comparadas com mulheres com quinto de riqueza mais baixo.

Mulheres grávidas são menos propensas a usar rede mosquiteira ($OR=0.73$). A Tabela 3.5 ainda revela que um aumento de 1 pessoa no tamanho do agregado familiar provoca uma diminuição de 25% ($OR=0.75$) na chance da mulher usar uma rede mosquiteira. Finalmente, por cada rede adicional no agregado familiar há um aumento ($OR=4.22$) na chance de usar rede mosquiteira por parte da mulher, ajustando para outras variáveis.

3.4.4 Conclusão

Neste capítulo, consideraram-se os modelos de regressão logística, ajustando para o desenho da pesquisa para identificar os fatores associados ao uso da rede mosquiteira pelas mulheres em idade reprodutiva em Moçambique. Ajustou-se modelos de regressão logística considerando-se ou não o plano amostral.

Os resultados do estudo mostraram que os fatores que estão associados ao uso da rede mosquiteira são: idade agrupada, estado civil, província, área de residência, nível de escolaridade, quinto de riqueza, se está grávida, número de membros do agregado e número de redes mosquiteiras no agregado. As variáveis se está a trabalhar no momento, pulverização intradomiciliar e sexo do chefe do agregado não afetam de forma significativa no uso da rede mosquiteira.

Desta análise constatou-se que mulheres grávidas tem menor chance de usar a rede mosquiteira, o que é preocupante visto que este é um grupo de risco para malária grave e complicações associadas tanto a mulher quanto ao bebé. Resultados similares foram observados no Mianmar [6]. Estudos adicionais devem ser feitos para identificar os fatores associados ao baixo uso de redes mosquiteiras por mulheres grávidas e melhorar a prevenção de malária neste grupo de risco. O baixo uso de redes mosquiteiras também foi mais observado na região sul, podendo dever-se ao facto de esta ser a região com menor prevalência de malária no país e em zonas rurais, onde há maior prevalência de malária. Semelhante ao que já foi observado anteriormente [24, 53] os resultados desta análise mostraram que o nível de escolaridade secundário está associado ao uso de redes mosquiteiras por mulheres. Isto pode ser justificado pelo facto destas mulheres poderem interpretar melhor as mensagens sobre a malária e entenderam os seus riscos.

Em relação à metodologia aplicada no estudo, e para enfatizar o facto de que foram escolhidos métodos que incorporam o desenho da amostra e o peso desigual das observações do estudo, obtendo assim resultados confiáveis fez-se uma comparação das estimativas ajustada e não ajustadas ao plano amostral. Os resultados, em geral, mostram que as estimativas pontuais de média e proporção foram semelhantes comparando-se amostragem complexa e amostragem aleatória simples. O mesmo foi observado nas estimativas dos coeficientes de regressão logística. Porém observou-se grande diferença nos erros padrão, eles ficam subestimados quando não se leva em consideração as características do desenho. Isto faz com que os coeficientes estimados para modelos de regressão sejam considerados

significativos quando na verdade não são.

Adicionalmente, verificou-se que existem recursos disponíveis que podem facilitar e melhorar a incorporação adequada dos diversos aspetos amostrais, tanto na estimação e precisão dos parâmetros quanto no ajuste e avaliação de modelos ajustados. E desta forma, encoraja-se aos investigadores em saúde que sempre que se tratar de uma AC, deve-se levar em consideração o efeito do plano amostral na análise descritiva e inferencial de modo a produzir resultados confiáveis, ou seja, é preciso selecionar e utilizar métodos adequados para compensar o efeito do delineamento amostral nas análises de interesse.

MODELOS LOGÍSTICOS MULTINÍVEL, AJUSTANDO PARA O DESENHO DA PESQUISA

No capítulo 3 usou-se o modelo de regressão logística ajustando para o desenho da pesquisa para identificar os fatores que influenciam no uso da rede mosquiteira por mulheres em idade reprodutiva em Moçambique. Neste capítulo, o objetivo principal é fazer uma extensão dos GLMs e utilizar o modelo logístico multinível ajustando para o delineamento da pesquisa que é uma abordagem alternativa para analisar dados de pesquisa complexa.

4.1 Introdução

Os modelos multiníveis são frequentemente usados para analisar dados de pesquisa complexa multietápica quando conglomeração, estratificação e probabilidade desigual de seleção estão envolvidas. As *outcomes* podem ser correlacionados devido à heterogeneidade não observada entre conglomerados, o que deveria ser considerado para fazer inferências estatísticas válidas. Nos dados de pesquisa complexa multietápica, as unidades da amostra podem não ser independentes dentro de um conglomerado ou entre conglomerados.

O modelo de regressão logística de efeitos mistos é uma escolha comum para a análise de dados dicotômicos multiníveis e é o mais usado no Modelos Lineares Generalizados Mistos (GLMMs).

Nesta tese, a *outcome* é binária. O modelo logístico de *intercept* aleatório ou o modelo logístico de coeficiente aleatório podem ser usados para analisar dados de pesquisa complexa com base na Máxima Pseudo-Verosimilhança Multinível (MPVM). Podemos considerar o peso da amostra para cada nível para analisar dados de pesquisa complexa multietápica. A probabilidade desigual de seleção é considerada usando os pesos amostrais dos indivíduos em cada nível dos dados de pesquisa complexa multietápica. As estimativas de parâmetros obtidas a partir da análise de dados de pesquisa complexa podem ser severamente enviesadas se os pesos de amostragem forem ignorados [5, 17, 57, 100].

De acordo com Chantala [17], a aplicação dos modelos multinível em dados de delineamento amostrais complexos é bastante recente. Estes modelos ainda são muito pouco

aplicados neste contexto de delineamento amostral complexo pois softwares estatísticos que incorporam essa metodologia possibilitando o uso dos pesos amostrais ainda é muito recente e requer pesos em diferentes níveis que muitas vezes não são disponibilizados. Assim, neste capítulo, modela-se a prevalência do uso da rede mosquiteira em Moçambique com base em dados coletados de uma pesquisa complexa. Introduzem-se os pesos de diferentes níveis no modelo usando o efeito do desenho, que é uma abordagem baseada no desenho. O principal interesse é comparar a variação na inferência quando modelamos os dados com esses pesos de pesquisa e quando ignoramos os pesos da pesquisa (assumindo que a amostragem aleatória simples foi usada na coleta dados), aplicados aos dados de prevalência do uso das redes mosquiteiras em Moçambique.

4.2 Modelos Logístico Multinível

Os Modelos Lineares Generalizados Mistos (GLMMs) (também conhecidos como modelos *cluster-specific*) são um caso particular dos modelos mistos e foram propostos por Breslow e Clayton (1993) [73]. Estes modelos, os GLMMs, são usados para dados agrupados ou em *cluster* onde as observações dentro de um *cluster* não podem ser assumidas como mutuamente independentes.

Os GLMMs são considerados como uma extensão dos Modelos Lineares Generalizados (GLMs). Ou seja, os GLMMs incluem para além de efeitos fixos, incluem igualmente uma componente de variabilidade devido a efeitos não observados no preditor linear [20, 73, 100]. A estrutura de dependência dos dados agrupados é modelada pela introdução do *intercept* aleatório no preditor linear.

Quando os dados possuem uma estrutura hierárquica, os GLMMs são chamados Modelos Multinível e os efeitos aleatórios entram no modelo em cada nível da hierarquia [33]. Desta forma, para além da inclusão dos efeitos aleatórios no preditor linear, os GLMMs permitem que a estrutura de dependência dos dados em vários níveis (variação dentro de cada *cluster* e a variação entre *clusters*) possa ser modelada. Estes modelos apresentam vantagens óbvias: em vez de tratar as observações incorretamente como não relacionadas, eles levam explicitamente em consideração o agrupamento das observações e permitem modelar como as características do nível mais alto (*macro* unidades) impactam unidades no nível mais baixo (*micro* unidades). Neste trabalho, a variável resposta é binária e portanto, o modelo logístico multinível será usado para análise e abordado a seguir.

Seja y_{ji} a variável resposta (*outcome*) de interesse binária para o indivíduo j aninhado dentro do *cluster* i ($i = 1, \dots, n_j; j = 1, \dots, J$), com distribuição de Bernoulli (π_{ji}), que é medida no nível mais baixo da estrutura hierárquica dos dados. Os modelos logísticos de dois níveis que são comumente usados são:

- **Modelo logístico multinível de *intercept* aleatório** - são modelos nos quais o nível geral de resposta é considerado como variável em relação aos *clusters* após o ajuste de potenciais covariáveis. O modelo logístico multinível de *intercept* aleatório (dois

níveis) para o indivíduo j (nível 1) do *cluster* i (nível 2) é dado por

$$\begin{aligned} \text{logit}[P(y_{ji} = 1 | x'_{ji}, u_i)] &= x'_{ji}\beta + u_i \\ &= \beta_0 + \beta_1 x_{1ji} + \dots + \beta_p x_{pji} + u_i \\ &= (\beta_0 + u_i) + \beta_1 x_{1ji} + \dots + \beta_p x_{pji}, \end{aligned}$$

onde u_i são os *intercept* aleatórios e são considerados como variáveis aleatórias. O efeito da heterogeneidade não observada pode ser representado por parâmetros aleatórios u_i , com $u_i | x_{ji} \sim N(0, \Omega_u)$. O modelo de *intercept* aleatório é assumido para capturar os efeitos combinados dos efeitos fixos β e os efeitos aleatórios u_j . Estes modelos são paralelos entre si por causa dos parâmetros de inclinação constantes β para cada modelo.

- **Modelo logístico multinível de coeficientes aleatórios** - são modelos em que o nível geral da resposta e os efeitos das covariáveis variam sobre os *clusters* depois de controlado para as covariáveis. O modelo logístico multinível (2 níveis) de coeficientes aleatórios para a unidade j (nível 1) dentro do *cluster* i (nível 2) é dado por

$$\text{logit}[P(y_{ji} = 1 | x'_{ji}, u_i)] = x'_{ji}\beta + z'_{ji}u_i$$

onde x_{ji} é o vetor de covariáveis com efeitos fixos β , z_{ji} é o vetor de covariáveis com efeitos aleatórios u_i ao nível do grupo. u_i são independentes entre as unidades i de nível 2 e o termo $u_i z_{ji}$ indica a interação entre os *clusters* e as covariáveis. O *intercept* aleatório e a inclinação aleatória tem uma distribuição normal bivariada com média zero e a matriz de variância-covariância Ω_u .

4.2.1 Estimação dos parâmetros do modelo

A estimação do vetor de parâmetros β e dos parâmetros Ω_u é feita pela maximização da função de verosimilhança marginal, obtida pela integração da função de verosimilhança em relação aos efeitos aleatórios u_i .

A contribuição da log-verosimilhança marginal para a unidade do nível 1, condicionada aos efeitos aleatórios no nível 2, i , pode ser escrita como:

$$l_i(\beta, \Omega_u) = \log \int \prod_{j=1}^{n_i} \pi_{ji}^{y_{ji}} (1 - \pi_{ji})^{1-y_{ji}} \phi(u_i; \Omega_u) du_i, \quad (4.1)$$

onde $\phi(u_i; \Omega_u)$ é função de densidade normal $N(0, \Omega_u)$. A função log-verosimilhança marginal é dada por

$$l(\beta, \Omega_u) = \sum_{i=1}^I l_i(\beta, \Omega_u). \quad (4.2)$$

Maximizando a função (4.2), obtêm-se os estimadores de máxima verosimilhança. Em delineamentos amostrais complexos de múltiplas etapas, as probabilidades de seleção

de unidades nos níveis são desiguais. As estimativas usuais de máxima verosimilhança são tendenciosas sem levar em consideração a probabilidade desigual de seleção [6]. O algoritmo da máxima pseudo-verosimilhança pode acomodar os pesos de probabilidade e reduzir o viés das estimativas dos parâmetros e será apresentado na secção que se segue.

4.2.2 Método de estimação de Máxima Pseudo-Verosimilhança

Considere-se um desenho amostral de duas etapas no qual π_i ($i = 1, \dots, I$) é a probabilidade de seleção de uma unidade de nível 2 e $\pi_{j|i}$ ($j = 1, \dots, n_i$) é a probabilidade de seleção da j -ésima unidade no nível 1 dentro do *cluster* i de nível 2. Seja também $w_i = 1/\pi_i$ e $w_{j|i} = 1/\pi_{j|i}$ o inverso da probabilidade de seleção da unidade i no nível 2 e a unidade j de nível 1 dentro da unidade i de nível 2, respetivamente.

As estimativas de MPVM são obtidas maximizando a log-verosimilhança ponderada. Os pesos de amostragem são incorporados na verosimilhança para a estimação dos parâmetros e os seus erros padrão. Maximiza-se a pseudo-verosimilhança ponderada em relação aos parâmetros de regressão. A contribuição da log-verosimilhança marginal para a unidade do nível 1, condicionada aos efeitos aleatórios no nível 2, i , pode ser escrita como:

$$l_i(\boldsymbol{\beta}, \boldsymbol{\Omega}_u) = \log \int \prod_{j=1}^{n_i} w_{j|i} \pi_{ji}^{y_{ji}} (1 - \pi_{ji})^{1-y_{ji}} \phi(u_i; \boldsymbol{\Omega}_u) du_i, \quad (4.3)$$

onde $\phi(u_i; \boldsymbol{\Omega}_u)$ é função de densidade normal $N(\mathbf{0}, \boldsymbol{\Omega}_u)$. A função log-verosimilhança marginal é dada por

$$l_w(\boldsymbol{\beta}, \boldsymbol{\Omega}_u) = \sum_{i=1}^I w_i l_i(\boldsymbol{\beta}, \boldsymbol{\Omega}_u). \quad (4.4)$$

Maximizando a função (4.4), obtêm-se os estimadores de máxima pseudo-verosimilhança. Contudo, geralmente o integral é intratável (não possui forma analítica), principalmente se houver muitos efeitos aleatórios a serem estimados [30, 32]. Consequentemente, faz-se necessário o uso de métodos de integrações numéricas. Os métodos de integração numérica, por sua vez, podem ser computacionalmente intensivos se o tamanho da amostra for grande e/ou o número de efeitos aleatórios no modelo for relativamente grande. Assim, para aliviar essa carga computacional, vários métodos de aproximações numéricas têm sido propostos e estes podem ser divididos em três grupos [9, 20, 73]:

- os que são baseadas na aproximação dos dados (consideram o desenvolvimento da série de Taylor e inserem-se na metodologia da quasi-verosimilhança penalizada (PQL) e o método de quasi-verosimilhança marginal);
- os que são baseados na aproximação da função a integrar (têm como base a aproximação de Laplace);
- os que são baseadas na aproximação do integral (o método da quadratura de Gauss-Hermite ou o método adaptativo da quadratura de Gauss-Hermite).

Outros métodos de integração numérica têm sido propostos para superar as limitações computacionais, dentre eles, os métodos baseados em Monte Carlo e a aproximação de Laplace aninhada e integrada que têm recebido grande atenção devido a desenvolvimentos paralelos em estatística bayesiana.

4.3 Aplicação para os fatores que contribuem para o uso de redes mosquiteiras em mulheres em idade reprodutiva em Moçambique

Em seguida, aplicou-se a metodologia dos modelos de regressão logística multinível para modelar o uso da rede mosquiteira em Moçambique, i.e., para identificar os fatores que contribuem para o uso de redes mosquiteiras pelas mulheres em idade reprodutiva (15–49 anos) em Moçambique, tendo em conta as diferentes *clusters*. Para tal, usou-se os dados do Inquérito Demográfico e de Saúde de Moçambique do ano 2011 [27] referentes as mulheres conforme descrito na Tabela 3.1 da secção 3.4.1.

4.3.1 Variáveis do estudo

A variável cujo comportamento se pretende explicar é o uso da rede mosquiteira (0 - não, 1 - sim). As possíveis covariáveis foram definidas ao nível individual e ao nível agregado são apresentadas na Tabela 4.1 que se segue.

Tabela 4.1: Variáveis do estudo

Fatores individuais	
Idade (V013)	(1) 15-19, (2) 20-24, (3) 25-29, (4) 30-34, (5) 35-39, (6) 40-44, (7) 45-49
Nível de escolaridade (V106)	(1) analfabeto, (2) primário, (3) secundário, (4) superior
Quinto de Riqueza (V190)	(1) mais baixo, (2) segundo, (3) médio, (4) quarto, (5) mais elevado
N. de membros do agregado familiar (V136)	-
Sexo do chefe do agregado (V151)	(1) masculino, (2) feminino
Estado civil (V501)	(1) solteiro, (2) casado, (3) vive maritalmente, (4) viúvo, (5) divorciado, (6) separada
Se está grávida (V213)	(1) não, (2) sim
Se está a trabalhar (V714)	(1) não, (2) sim
Número de redes mosquiteiras (HML1)	-
Posse de rede mosquiteira (V459)	(0) não, (1) sim
Pulverização intradomiciliar (HV253)	(1) não, (2) sim, (3) não sei
Fatores agregados	
Área de residência (V025)	(1) urbana, (2) rural
Província (V024)	(1) Niassa, (2) Cabo Delgado, (3) Nampula, (4) Zambézia, (5) Tete, (6) Manica, (7) Sofala, (8) Inhambane, (9) Gaza, (10) Maputo Província, (11) Maputo Cidade

4.3.2 Análise dos dados

Os Modelos de regressão logística multinível foram a técnica de análise usada para ter em conta a estrutura hierárquica dos dados do DHS e a variável de interesse binária.

Uma análise bivariada foi feita em primeiro lugar e as variáveis que se mostraram significativas foram consideradas na análise multivariada. Para analisar os fatores que influenciam no uso da rede mosquiteira a dois níveis, ou seja, considerou-se que as mulheres (nível 1) estão aninhadas dentro das áreas de enumeração (nível 2). Para tal, foram ajustados quatro modelos de regressão logística multinível para estimar os *odds ratio* ajustados e a variação aleatória entre as áreas de enumeração.

O modelo 1 (modelo nulo) foi ajustado sem covariáveis para testar a variabilidade aleatória no *intercept* e também para estimar o coeficiente de correlação intraclasse (ICC). O ICC indica a proporção não explicada da variância a nível agregado, ou seja, ele reflete

a magnitude da variância entre os grupos, calculado da seguinte forma:

$$ICC = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_\epsilon^2},$$

onde $\hat{\sigma}_u^2$ representa a variância estimada para o efeito aleatório, ou seja, a variância entre os grupos e $\hat{\sigma}_\epsilon^2$ a variância estimada dos erros do modelo. Para variável de interesse binária, $\hat{\sigma}_\epsilon^2 = \frac{\pi^2}{3}$.

O modelo 2 examinou os efeitos das características ao nível individual, o modelo 3 examinou o efeito das características do nível do *cluster*. E por fim, o modelo 4 examinou os efeitos das características individuais e agregadas (*cluster*) simultaneamente.

Para o ajuste dos modelos usou-se o *software* STATA versão 16 (Stata Corp LLC., 2017; College Station, TX, USA). As características do delineamento amostral, como estratificação, conglomeração e peso amostral foram incorporadas na análise usando a função *svyset*. Modelos de regressão logística ponderada foram ajustados com auxílio da função *svy: melogit* e os modelos de regressão logística não ponderada foram ajustados com auxílio da função *melogit*. Dado que os modelos multiníveis requerem pesos amostrais em cada nível de hierarquia e o DHS só fornece os pesos finais, os pesos de nível 1 e 2 foram aproximados conforme descrito em [26]. Ajustou-se 7 modelos com pesos aproximados usando vários níveis de importância ($\alpha=0, 0.1, 0.25, 0.5, 0.75, 0.9, 1$) e um oitavo modelo foi equipado com o peso de nível 1 declarado como 1 (uma prática comum usada quando os pesos de nível não estão disponíveis), conforme sugerido em [26].

4.3.3 Resultados

4.3.3.1 Ajuste do Modelo de Regressão Logística Multinível

A amostra do estudo foi constituída por 13,745 mulheres em idade reprodutiva (15 a 49 anos), em Moçambique no ano de 2011. O modelo de regressão logística multinível com *intercept* aleatório foi usado para determinar o impacto dos fatores individuais e agregados no uso da rede mosquiteira e os resultados do ajuste desses modelos são apresentados na Tabela 4.2 (análise não ponderada) e na Tabela 4.3 (análise ponderada). Para a análise multinível ponderada, ajustaram-se modelos com pesos aproximados usando vários níveis de importância, para os valores de $\alpha = 0, 0.1, 0.25$ e 0.5 não conseguimos obter resultados. Para os casos $\alpha = 0.75, 0.90$ e 1 , os coeficientes estimados e os seus erros padrão foram muito semelhantes. No entanto, em geral, os coeficientes do Modelo 8 são atenuados (mais próximos de 0) e os seus erros padrão são menores. Ao usar pesos de nível, a estimativa da variância do *intercept* no nível 2 varia entre 0,96 e 1,23, que é mais que o dobro da variância de 0,44 que é estimado quando o peso de nível 1 (no Modelo 8). Essa observação sugere que a variação entre os *clusters* no uso de da rede mosquiteira é subestimada se os pesos de nível não forem usados. Assim, dados que as estimativas são semelhantes, neste estudo consideraram-se os pesos com o $\alpha = 0.75$, os resultados para os restantes casos não foram

CAPÍTULO 4. MODELOS LOGÍSTICOS MULTINÍVEL, AJUSTANDO PARA O DESENHO DA PESQUISA

apresentados na tese. É de realçar que as categorias que estão com um "Ref" na descrição dos parâmetros representam as categorias de referência das variáveis categóricas.

Tabela 4.2: Modelo logístico multinível final não ponderado

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	AOR (M 4)
<i>intercept</i>	-0.644**	-2.030	-0.058	-0.587***	0.56
fatores de nível individual					
Idade				0.000	
15-19 (Ref)	-	-	-	-	-
20-24		0.440		0.296	1.34
25-29		0.452		0.437	1.55
30-34		0.380		0.432	1.54
35-39		0.376		0.453	1.57
40-44		0.181		0.270	1.31
45-49		0.218		0.247	1.28
Estado civil					
Solteira (Ref)	-	-	-	-	-
Casada		1.011		0.510	1.66
Vive maritalmente		0.705		0.418	1.52
Viúva		0.337		0.087	1.09
Divorciada		0.395		0.054	1.06
Separada		0.215		0.054	1.06
Nível de escolaridade					
Sem educação (Ref)	-	-	-	-	-
Primário		0.186		0.365	1.44
Secundário		0.292		0.692	2.00
Superior		-0.428		-0.204	0.82
Se está grávida					
Não (Ref)	-	-	-	-	-
Sim		-0.270		-0.143	0.87
N. de membros no agregado		-0.289		-0.066	0.94
N. de redes no agregado		1.367		-0.066	0.94
Quinto de riqueza				0.000	
Primeiro (Ref)	-	-	-	-	-
Segundo			0.025	0.012	1.01
Médio			0.293	0.252	1.29
Quarto			0.504	0.423	1.53
Mais elevado			0.908	0.783	2.19
fatores de nível agregado					
Província					
Niassa (Ref)	-	-	-	-	-
Cabo Delgado			0.369	0.364	1.44
Nampula			0.230	0.239	1.27
Zambézia			-0.310	-0.307	0.74
Tete			-0.468	-0.434	0.65
Manica			-0.616	-0.628	0.53
Sofala			-0.409	-0.226	0.80
Inhambane			-1.03	-0.926	0.40
Gaza			-2.37	-2.258	0.10
Maputo Província			-1.57	-1.537	0.21
Maputo Cidade			-1.45	-1.322	0.27
Área de residência					
Urbana (Ref)	-	-	-	-	-
Rural			-0.45	-0.420	0.66

(continua)

4.3. APLICAÇÃO PARA OS FATORES QUE CONTRIBUEM PARA O USO DE REDES MOSQUITEIRAS EM MULHERES EM IDADE REPRODUTIVA EM MOÇAMBIQUE

Tabela 4.2 (continuação)

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	AOR (M 4)
Efeitos aleatórios					
Var do nível das regiões (SE)	1.17 (0.092)	1.122 (0.090)	0.607 (0.054)	0.618 (0.0555)	
ICC (%)	26.2	25.4	15.82	15.82	
Ajustamento do modelo					
Log-Verosimilhança	-8215.8	-7992.0	-8019.5	-7819.6	
AIC	16435.58	16022.01	16073.1	15707.2	

CAPÍTULO 4. MODELOS LOGÍSTICOS MULTINÍVEL, AJUSTANDO PARA O DESENHO DA PESQUISA

Tabela 4.3: Modelo logístico multinível final ponderado

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	AOR (M 4)
<i>intercept</i>	-0.603	-2.458	-0.058	-0.383	0.29
fatores de nível individual					
Idade					
15-19 (Ref)	-	-	-	-	-
20-24		0.528		0.530	1.70
25-29		0.434		0.437	1.55
30-34		0.337		0.340	1.41
35-39		0.440		0.444	1.56
40-44		0.297		0.300	1.35
45-49		0.273		0.277	1.32
Estado civil					
Solteira (Ref)	-	-	-	-	
Casada		0.996		0.987	2.68
Vive maritalmente		0.894		0.904	2.44
Viúva		0.801		0.319	1.38
Divorciada		0.425		0.417	1.52
Separada		0.353		0.350	1.42
Nível de escolaridade					
Sem educação (Ref)	-	-	-	-	-
Primário		0.233		0.235	1.27
Secundário		0.348		0.349	1.42
Superior		-0.298		-0.306	0.74
Se está grávida					
Não (Ref)	-	-	-	-	
Sim		-0.343		-0.344	0.71
N. de membros no agregado		-0.301		-0.301	0.74
N. de redes no agregado		1.625		1.624	5.07
Quinto de riqueza					
Primeiro (Ref)	-	-	-	-	-
Segundo		0.038		0.039	1.04
Médio		0.221		0.226	1.25
Quarto		0.040		0.057	1.06
Mais elevado		0.019		0.044	1.04
fatores de nível agregado					
Província					
Niassa (Ref)	-	-	-	-	-
Cabo Delgado			0.482	-0.447	0.64
Nampula			0.265	-0.095	0.91
Zambézia			-0.404	-0.827	0.67
Tete			-0.580	-0.827	0.44
Manica			-0.419	-0.496	0.61
Sofala			-0.406	-0.591	0.55
Inhambane			-0.863	-2.150	0.12
Gaza			-2.997	-4.046	0.02
Maputo Província			-1.339	-1.76	0.17
Maputo Cidade			-1.285	-1.316	0.27
Área de residência					
Urbana (Ref)	-	-	-	-	-
Rural			-0.863	-1.241	0.59
Efeitos aleatórios					
Var do nível das regiões (SE)	2.023 (0.201)	2.252 (0.226)	1.241 (0.127)	1.186 (0.106)	
ICC (%)	38.08	40.6	27.39	26.50	

4.3. APLICAÇÃO PARA OS FATORES QUE CONTRIBUEM PARA O USO DE REDES MOSQUITEIRAS EM MULHERES EM IDADE REPRODUTIVA EM MOÇAMBIQUE

Da análise da Tabela 4.3, o Modelo 1 (modelo nulo), com o *intercept* aleatório apenas, mostra que existe uma variância significativa não explicada ao nível dos *clusters* em relação ao uso da rede mosquiteira. Da mesma forma, o ICC no modelo nulo mostra que 38.08% da variância total no uso das redes mosquiteiras pode ser atribuída às diferenças entre os *clusters*. A inclusão de fatores ao nível individual no modelo 2 explicou 40.6% da variação no nível dos *clusters* no uso da rede mosquiteira. A província, a área de residência e o quinto de riqueza incluídos no modelo 3 explicaram cerca de 27.39% da variação do uso das redes mosquiteiras que permaneceu não explicada no modelo 2. Da mesma forma, a inclusão dos fatores ao nível individual e agregado, em simultâneo, no modelo 4 explicou 26.50% da variação dos *clusters*. E também constata-se que a variância não explicada ao nível dos *clusters* no modelo 1 é de 2.023 e no modelo 4, com a inclusão dos fatores individuais e agregados, a variância diminuiu para 1.186.

Pela análise dos *odds ratios* ajustados do modelo 4 (Tabela 4.2), verifica-se que depois de controlado para as outras variáveis a nível individual e agregado, as mulheres que tinham educação superior eram 26% (OR = 0.74) tem menos chance de usar rede mosquiteira e mulheres que tinham ensino primário eram 27% (OR = 1.27) mais propensas de usar rede mosquiteira; mulheres com ensino secundário eram 42% (OR = 1.42) mais propensas a usar rede mosquiteira, em comparação com as mulheres que não tinham educação. Mulheres grávidas foram 29% menos propensas a usar rede mosquiteira em comparação com as mulheres que não estavam grávidas.

Mulheres no maior quinto de riqueza eram 1.04 vezes (OR = 1.04) mais propensas de usar rede mosquiteira em comparação com mulheres mais pobres. Em relação à província, Gaza, Maputo província e Maputo tinham menor chance de usar rede mosquiteira (OR= 0.02; 0.17; 0.27, respetivamente) quando comparadas com Niassa. As mulheres nas províncias de Cabo Delgado e Nampula tinham 36% e 9% mais chances de usar rede mosquiteira quando comparadas com as mulheres de Niassa, controlando para outros fatores. Mulheres da zona rural tinham menor chance de usar rede mosquiteira em comparação com as mulheres da zona urbana (OR = 0.59). Aumentando uma pessoa no tamanho do agregado familiar, a mulher tinha chance 26% menor chance de usar rede mosquiteira. Aumentando uma rede no agregado familiar, a chance da mulher usar a rede mosquiteira aumentava em 5 vezes.

4.3.4 Conclusão

Desenhos complexos de pesquisa muitas vezes envolvem probabilidades de seleção desiguais dos *clusters* e/ou de pessoas dentro dos *clusters*. A pesquisa inclui pesos do desenho (amostragem) para levar em conta as probabilidades de seleção desiguais. Ao estimar modelos multiníveis que são baseados em dados provenientes de amostragem complexa, os pesos de amostragem são incorporados na verosimilhança, produzindo uma pseudo-verosimilhança. Neste capítulo considerou-se a modelação multinível para dados

de pesquisa de 2 níveis com uma *outcome* binária, incorporando os pesos no processo de estimação para identificar os fatores que influenciam no uso da rede mosquiteira pelas mulheres em idade reprodutiva.

Para tal, consideraram-se modelos de regressão logística multinível não ajustados e ajustados para o plano amostral. Dos resultados verificaram-se diferenças nos erros padrão entre os métodos ajustados e não ajustados para o plano amostral, onde a abordagem padrão de modelação multinível pode estimar parâmetros e erros padrão adequadamente para dados agrupados que resultaram de uma amostragem com igual probabilidade de seleção. No entanto, com probabilidades de seleção desiguais, essa abordagem padrão não estima corretamente os parâmetros do modelo e os seus erros padrão.

Os resultados deste estudo mostraram que os *intercept* aleatório ao nível agregado (variância) foi grande e estatisticamente significativa, indicando diferenças entre os *clusters*. Ou seja, cerca de 38.08% das variações do uso da rede mosquiteira é atribuída as diferenças entre os *clusters*. O que suporta a aplicação dos modelos multinível para este estudo particular.

A idade, província, área de residência, nível de escolaridade, se está grávida, número de membros no agregado familiar, número de redes mosquiteiras no agregado, quinto de riqueza e o estado civil foram os fatores, ao nível individual e ao nível agregado, significativamente associados com o uso da rede mosquiteira.

Além disso, os resultados também mostram que ao nível agregado, a província e a área de residência onde a mulher reside desempenham um papel significativo no uso da rede mosquiteira.

MODELAÇÃO ESPACIAL E MAPEAMENTO DE DOENÇAS, AJUSTANDO PARA O DESENHO AMOSTRAL

Nos capítulos anteriores, o uso dos modelos de regressão logística e regressão logística multinível para dados binários incorporando o plano amostral foram ilustrados. Deste ponto em diante, será feita uma progressão, mais especificamente para o mapeamento de doenças, um campo científico relativamente novo onde os dados binários de doenças são modelados numa configuração de dados espaciais. Este capítulo é baseado na seguinte publicação [15].

5.1 Introdução

Em saúde pública e epidemiologia, os métodos de mapeamento de doenças têm sido usados há muito tempo para estimar padrões espaciais de risco de doenças. Um exemplo bem conhecido do seu uso são os estudos de John Snow sobre a epidemia de cólera na área da *Broad Street* de Londres em 1854 [110]. Graças aos avanços em *software*, *hardware* e tecnologia do computador nas últimas décadas, o mapeamento de doenças é agora usado para aplicações que vão desde o mapeamento do cancro [13, 65, 66, 72, 104] até a modelação da disseminação potencial de doenças infecciosas [8, 39, 54, 59, 70, 89, 90]. O mapeamento de doenças também é particularmente útil para a detecção de aglomerados espaciais que podem ser devido a fatores ambientais e demográficos comuns, ou efeitos culturais compartilhados por regiões vizinhas [1, 8, 13, 18, 19, 22, 34, 39, 43, 44, 45, 54, 59, 66, 70, 72, 75, 89, 90, 104, 114, 120, 121, 122]. Mapear a prevalência de doenças em locais com poucos recursos é uma abordagem cada vez mais importante para orientar a tomada de decisões, pois conseguimos visualizar tendências espaciais e temporais da carga de doenças [3, 31, 48, 112].

Os métodos de suavização espacial hierárquica são bem desenvolvidos e usados de maneira padrão no mapeamento de doenças. No entanto, especialmente na África Subsaariana, muitas análises de mapeamento de doenças que possuem dados de pesquisas

de saúde complexas não levaram em conta o desenho da pesquisa, como a amostragem desproporcional [69].

Em análises de pesquisa padrão, a amostragem desproporcional é corrigida na análise usando os pesos de amostragem de pesquisa que se ajustam à contribuição desproporcional de cada unidade de amostragem final. Ignorar o peso amostral nas análises de mapeamento de doenças pode levar a estimativas tendenciosas das distribuições espaciais do risco de doença, o que pode afetar negativamente as decisões políticas baseadas nelas. Assim, métodos de análise estatística apropriados que incorporem pesos amostrais na estimativa de padrões espaciais de doenças são imprescindíveis [18, 22, 75, 89, 120, 121].

O objetivo deste capítulo é apresentar os avanços estatísticos recentes nos métodos de mapeamento de doenças que incorporam os pesos amostrais na estimação da distribuição espacial do risco de doença. Os métodos foram então aplicados à estimação da distribuição geográfica da desnutrição infantil no Malawi ao nível distrital com base no MDHS 2015-16 [82], e da febre e da diarreia infantil ao nível provincial em Moçambique usando o IMASIDA 2015 [77].

5.2 Notação Geral

Suponha que uma população finita $U = \{1, 2, \dots, N\}$ seja distribuída em $i = 1, 2, \dots, I$ áreas, e uma amostra probabilística aleatória s da pesquisa com tamanho n é retirada desta população U de acordo com um determinado desenho. Sejam N_i e n_i o tamanho da população e da amostra para a área i , respetivamente tal que $U = \cup_{i=1}^I U_i$, $N = \sum_{i=1}^I N_i$, $s = \cup_{i=1}^I s_i$ e $n = \sum_{i=1}^I n_i$. Seja Y_{ij} um *outcome* de saúde de interesse binária do j -ésimo indivíduo da área i ($j = 1, \dots, N_i$; $i = 1, \dots, I$). Assume-se que N_i é conhecido para cada área i . Além disso, assumimos que o indivíduo ij tem uma probabilidade conhecida π_{ij} de ser incluído na amostra.

O interesse é estimar a verdadeira prevalência populacional específica da área P_i , definida como

$$P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}, \quad (5.1)$$

usando a amostra acumulada da área i . O estimador não ponderado específico da área da verdadeira prevalência P_i é dado por $\hat{P}_i^{UW}$ que é calculado como

$$\hat{P}_i^{UW} = \frac{1}{n_i} \sum_{j \in s_i} y_{ij}, \quad (5.2)$$

e a sua variância é estimada de forma ingênua usando

$$\widehat{var}(\hat{P}_i^{UW}) = \frac{\hat{P}_i^{UW}(1 - \hat{P}_i^{UW})}{n_i}. \quad (5.3)$$

No caso de um desenho de AAS sem reposição, o estimador (5.2) é não enviesado. No entanto, em AC seria inadequado, pois não considera o desenho, por exemplo, pesos amostrais [61].

5.2.1 Estimador de Horvitz-Thompson

O estimador não enviesado baseado no desenho amostral da pesquisa bem conhecido para a prevalência da população P_i é o estimador de Horvitz-Thompson (HT) [37], dado por

$$\hat{p}_i^{HT} = \frac{\sum_{j \in s_i} w_{ij}^d y_{ij}}{\sum_{j \in s_i} w_{ij}^d} = \frac{1}{n_i} \sum_{j \in s_i} \tilde{w}_{ij}^d y_{ij}, \quad (5.4)$$

onde s_i é o conjunto de indivíduos amostrados da área i , sendo y_{ij} o valor observado para $j \in s_i$ com $|s_i| = n_i$, $w_{ij}^d = \frac{1}{\pi_{ij}}$ é o peso do desenho, π_{ij} é a probabilidade de inclusão na amostra para a unidade j na área i (ou seja, são o inverso da probabilidade de inclusão na amostra ajustado para as não respostas [42]), e $\tilde{w}_{ij}^d$ dado por

$$\tilde{w}_{ij}^d = n_i \frac{w_{ij}^d}{\sum_{j \in s_i} w_{ij}^d}, \quad (5.5)$$

é o peso amostral normalizado (os pesos amostrais são o inverso da probabilidade de inclusão na amostra ajustado para as não respostas). A normalização envolve reponderar uma amostra para corresponder ao tamanho da amostra na área i , ou seja, $\sum_{j \in s_i} \tilde{w}_{ij}^d = n_i$ [18, 120, 121]. De acordo com [18], um estimador da variância de $\hat{p}_i^{HT}$ pode ser expresso da seguinte forma

$$\hat{v}ar(\hat{p}_j^{HT}) = \frac{1}{n_j} \left(1 - \frac{n_j}{N_j}\right) \frac{1}{n_j - 1} \sum_{i \in s_j} \tilde{w}_{ij}^{d^2} (y_{ij} - \hat{p}_j^{HT})^2. \quad (5.6)$$

O estimador HT é um estimador não enviesado do desenho de P_i [37, 102, 107]. Esse estimador é confiável para grandes amostras, mas para pequenas amostras terá uma grande variância [102]. As estimativas de uma área obtidas por uma abordagem baseada no desenho, ou seja, usando o estimador HT, são consideradas diretas, pois são baseadas apenas nos dados da área de interesse amostrada [98, 102].

5.3 Modelos Hierárquicos Bayesianos suavizados

Os modelos hierárquicos bayesianos suavizados recentemente ganharam atenção quanto ao seu uso na estimação de áreas pequenas ao invés de estimadores diretos [18, 75]. Esses métodos partem do pressuposto de que as estimativas de áreas específicas partilham informações de outras áreas, o que torna possível encontrar estimativas mais precisas. Além disso, isso cria a vantagem de que as estimativas podem ser obtidas em áreas que não foram amostradas. Os modelos hierárquicos bayesianos envolvem três etapas: (i) a verossimilhança da resposta definida condicionalmente em algumas variáveis latentes (efeitos aleatórios); (ii) é dada uma distribuição as variáveis latentes; e (iii) as distribuições a priori dos parâmetros desconhecidos são especificadas.

Um modelo hierárquico bayesiano de suavização espacial de três etapas para o número total de indivíduos com a doença na área i dado por $y_i = \sum_{j \in s_i} y_{ij}$ usa uma distribuição binomial para a etapa 1 como

$$y_i | P_i \sim \text{Binomial}(n_i, P_i). \quad (5.7)$$

Na segunda etapa, modela-se a variação entre áreas em P_i usando o modelo de efeitos aleatórios de área. Nos últimos tempos, isso envolveu a incorporação de efeitos aleatórios não espaciais e espaciais usando o modelo de convolução de Besag, York e Mollié (BYM) [10]. Além do modelo espacial binomial padrão, descreve-se uma série de modelos baseados no modelo BYM que têm sido utilizados para realizar análises espaciais de dados de prevalência a partir de inquéritos de saúde.

Modelo 1: Modelo espacial binomial padrão. Usando a distribuição Binomial em (5.7), o primeiro modelo é uma abordagem de modelação espacial padrão para dados de contagem usando dados de pesquisa de saúde para estimar a distribuição espacial do risco de doença. Este modelo simplesmente liga a prevalência da doença com os dois tipos de termos aleatórios usando uma função logit como

$$\begin{aligned} \text{logit}(P_i) &= \beta_0 + u_i + v_i, \\ u_i | u_{i'}, i \neq i' &\sim N\left(\frac{1}{a_i} \sum_{i' \in ne(i)} u_{i'}, \frac{\sigma_u^2}{a_i}\right), \\ v_i | \sigma_v^2 &\sim_{iid} N(0, \sigma_v^2), \quad i = 1, \dots, I, \end{aligned} \quad (5.8)$$

onde β_0 é o *intercept*; v_i é a componente aleatória não estruturada; u_i é a componente aleatória espacial estruturada, $ne(i)$ indica o conjunto de vizinhos e a_i é o número de vizinhos para uma determinada área i . Aqui, adota-se a convenção comum de vizinhança, que considera duas áreas como vizinhas se elas compartilham uma fronteira comum. A especificação sobre os efeitos aleatórios estruturados é baseada numa priori autoregressiva condicional intrínseca (ICAR) [10, 105].

Na terceira etapa, atribui-se prioris para β_0 e as variâncias de efeitos aleatórios. O modelo da Equação (5.8) resulta na suavização de estimativas de áreas extremas em áreas com tamanhos amostrais pequenos, o que é um efeito desejável porque evita o sobreajuste. No entanto, sem ajustes para pesos amostrais, as estimativas de risco de doença de área podem ser tendenciosas.

5.3.1 Opções para ter em conta os pesos da pesquisa nos modelos hierárquicos

Seguindo [18, 75], várias abordagens foram propostas para explicar o desenho da pesquisa. Considerem-se agora os modelos 2 e 3 que, além de se basearem no estimador HT, também levam em conta os pesos da pesquisa. Uma maneira simples de permitir a suavização espacial do estimador HT é assumir um modelo de suavização espacial hierárquica para

$\hat{p}_i^{HT}$. Como a variância da distribuição binomial é uma função da média, a transformação da variância é usada para normalizar as estimativas da prevalência. As transformações logit e arco-seno são comumente usadas. Primeiro considera-se a transformação logit. Assim, o Modelo 2 é dado por

$$\begin{aligned} \text{logit}(\hat{p}_i^{HT}) | P_i &\sim N(\text{logit}(P_i), \sigma_i^2) \\ \text{logit}(P_i) &= \beta_0 + u_i + v_i, \end{aligned} \quad (5.9)$$

onde $\text{logit}(\hat{p}_i^{HT})$ tem variância $\sigma_i^2 = \widehat{\text{var}}(\hat{p}_i^{HT}) / (\hat{p}_i^{HT} (1 - \hat{p}_i^{HT}))^2$.

Chamaremos **Modelo 2** ao **modelo espacial Logit Normal (LN)**. Para a transformação da raiz quadrada do arco-seno, dada como $\arcsin(\sqrt{\hat{p}_i^{HT}})$ [75, 101], o modelo espacial transformação da raiz quadrada do arco-seno (AS), leva à seguinte especificação de modelo

$$\begin{aligned} \arcsin(\sqrt{\hat{p}_i^{HT}}) | P_i &\sim N(\arcsin(\sqrt{P_i}), \sigma_i^2) \\ \arcsin(\sqrt{P_i}) &= \beta_0 + u_i + v_i, \end{aligned} \quad (5.10)$$

e a variância da transformação do arco-seno é $\sigma_i^2 = \frac{1}{4n_i^E}$, onde $n_i^E = \hat{p}_i^{HT} (1 - \hat{p}_i^{HT}) / \widehat{\text{var}}(\hat{p}_i^{HT})$ é o tamanho efetivo da amostra na área i . Assim, o **Modelo 3** é o **modelo AS**.

Para o Modelo 4 considera-se a pseudo-verosimilhança (PL), que usa uma verosimilhança ponderada [75], onde os valores de resposta (y_{ik}) são ponderados usando os pesos do desenho normalizados. Assim, em vez de considerar os *outcomes* binomiais usados em (5.7) (Modelo 1), aqui usa-se $y_i^{PL} = \sum_{j \in s_i} \tilde{w}_{ij}^d y_{ij}$ como

$$\begin{aligned} y_i^{PL} | P_i &\sim \text{Binomial}(n_i, P_i) \\ \text{logit}(P_i) &= \beta_0 + u_i + v_i. \end{aligned} \quad (5.11)$$

Uma desvantagem da abordagem geral é que o erro padrão apropriado não é recuperado no caso de conglomeração. Rabe-Hesketh e Skrondal [100] usaram um método de pseudo-verosimilhança com pesos escalonados e usaram estimação sanduíche para fornecer estimativas de erro padrão válidas dentro de uma estrutura multinível, mas não consideraram a suavização espacial. Considere-se o **modelo espacial de pseudo-verosimilhança (PL)** como **Modelo 4**.

Os Modelos 5 e 6 também são variações do Modelo 1, mas agora dependem do tamanho efetivo da amostra (n_i^E) e do número de casos (y_i^E), e esses modelos podem ser descritos como

$$\begin{aligned} y_i^E | P_i &\sim \text{Binomial}(n_i^E, P_i) \\ \text{logit}(P_i) &= \beta_0 + u_i + v_i. \end{aligned} \quad (5.12)$$

Para o **Modelo 5**, o tamanho efetivo da amostra (n_i^E) é calculado conforme mostrado anteriormente e depende do estimador ponderado de prevalência [18]. O tamanho efetivo

da amostra é o tamanho da amostra que é necessário para tornar a variância sob o desenho complexo da pesquisa equivalente à de uma amostra aleatória simples [18]. Então o número efetivo de casos é facilmente encontrado como $y_i^E = n_i^E \hat{P}_i^{HT}$.

Por último, mas não menos importante, está o Modelo 6, que calcula os tamanho efetivo da amostra da área i através do efeito do desenho (Deff) na área i e é estimado como

$$n_i^E = \frac{n_i}{deff_i}, \quad (5.13)$$

onde

$$deff_i = \frac{s_i^2}{s_{ri}^2} \quad \text{para } i = 1, \dots, I, \quad (5.14)$$

onde s_i^2 é a estimativa direta não enviesada da variância da proporção com base no desenho complexo de amostragem e s_{ri}^2 é a estimativa direta não enviesada da variância da proporção com base no desenho de AAS [50]. Como antes, isso resultou no número efetivo de casos na área i dado por $y_i^E = n_i^E \hat{P}_i^{HT}$.

5.3.2 Inferência Bayesiana, Computação e Avaliação de Modelos

Assumiu-se uma priori uniforme imprópria para β_0 e prioris $Gamma(0.5, 0.008)$ para os parâmetros de precisão espacial e não espacial σ_u^{-2} e σ_v^{-2} como em [75]. A estimação dos parâmetros dos modelos foi feita usando aproximação de Laplace aninhada e integrada (INLA), que fornece uma alternativa rápida ao método de Monte Carlo via cadeias de Markov (MCMC) para aproximar as distribuições marginais a posteriori de modelos de campo aleatório de Markov (MRF) através da integração numérica [105].

A prevalência observada foi calculada considerando os pesos da pesquisa usando o pacote survey [60]. O INLA como implementado no pacote INLA [12, 108] pode ser baixado em <http://www.r-inla.org/>. Todas as análises estatísticas são realizadas dentro do ambiente computacional R [113].

A comparação e seleção do modelo foi realizada usando o critério de informação da deviance (DIC) [111]. O valor DIC é calculado como

$$DIC = \bar{D} + pD, \quad (5.15)$$

onde $\bar{D}$ é a média a posteriori da deviance que mede a qualidade do ajuste e pD é o número efetivo de parâmetros dado por

$$pD = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}),$$

onde $D(\bar{\theta})$ é a deviance calculada da média a posteriori dos parâmetros. Modelos com menores valores indicam um melhor ajuste.

5.4 Aplicação da modelação espacial à prevalência da malnutrição infantil no Malawi e da febre e da diarreia infantil em Moçambique

Neste capítulo, usaram-se os métodos apresentados nas secções anteriores para estimar a distribuição geográfica da prevalência da baixa estatura para a idade, do baixo peso para a estatura e do baixo peso entre crianças com menos de cinco anos a nível distrital no Malawi e a distribuição geográfica da prevalência da febre e da diarreia infantil a nível provincial em Moçambique.

5.4.1 Análise dos dados

Os dados analisados para o propósito do estudo foram do MDHS 2015-16 e do IMASIDA 2015 e encontram-se descritos no Capítulo 2 desta tese.

Os *outcomes* considerados neste estudo para a infância no Malawi é o estado nutricional das crianças (baixa estatura para a idade, baixo peso para a idade e baixo peso). Para tanto, foram incluídas no estudo crianças menores de cinco anos que possuíam dados válidos quanto aos seus indicadores antropométricos (medidas de altura e peso), resultando numa amostra de 5,149 crianças com a medição para a baixa estatura para a idade, 5,178 crianças com a medição para o baixo peso para a estatura e por fim, 5,223 com a medição do baixo peso.

Para os dados de Moçambique, os *outcomes* considerados são o *status* de febre e diarreia. Foram incluídas na análise as crianças menores de cinco anos cuja mãe respondeu se tiveram diarreia ou febre nas últimas 2 semanas. As demais crianças com valores ausentes para os desfechos foram excluídas da pesquisa. Assim, as análises incluem um total de 4,972 crianças menores de cinco anos para febre e 4,980 crianças para diarreia.

5.4.2 Resultados

5.4.2.1 Malawi: Variação distrital de saúde nas taxas de prevalência de malnutrição

A prevalência observada (ponderada) da baixa estatura para a idade, baixo peso para a idade e baixo peso no Malawi entre crianças menores de cinco anos é de 36.82% (95% IC 35.18-38.46), 2.79% (95% IC 2.24-3.33) e 11.58% (95% IC 10.49-12.67), respetivamente, com variação entre os distritos, que varia de 15.44% na cidade de Mzuzu a 45.88% em Mchinji para a baixa estatura para a idade; 1% em Balaka para 9.92% em Nsanje para o baixo peso para a idade; e 1.93% na cidade de Zomba para 18.85% em Nsanje para o baixo peso (Tabela 5.1).

CAPÍTULO 5. MODELAÇÃO ESPACIAL E MAPEAMENTO DE DOENÇAS,
AJUSTANDO PARA O DESENHO AMOSTRAL

Tabela 5.1: Resumo da prevalência observada (ponderada) da baixa estatura para a idade, baixo peso para a idade e baixo peso por distrito no Malawi.

Distrito	baixa estatura para a idade (N,%)	baixo peso para a idade (N,%)	baixo peso (N,%)
Chitipa	43 (33.08)	2 (1.39)	18 (13.89)
Karonga	38 (28.00)	2 (1.56)	14 (9.18)
Nkhata Bay	47 (31.33)	1 (0.17)	10 (5.89)
Rumphi	44 (31.87)	3 (1.74)	20 (13.75)
Mzimba	70 (44.79)	5 (3.19)	22 (13.45)
Likoma	33 (26.99)	5 (4.26)	11 (9.13)
Mzuzu City	7 (15.44)	1 (2.72)	1 (2.72)
Kasungu	73 (35.90)	5 (2.73)	14 (6.56)
Nkhotakota	69 (32.58)	7 (1.82)	32 (13.05)
Ntchisi	68 (40.55)	4 (1.80)	20 (11.53)
Dowa	74 (39.42)	2 (1.05)	17 (9.31)
Salima	78 (36.75)	4 (1.60)	28 (13.81)
Lilongwe Rural	63 (43.28)	1 (0.64)	14 (9.42)
Mchinji	95 (45.88)	8 (3.35)	26 (12.20)
Dedza	72 (41.18)	5 (2.85)	28 (15.41)
Ntcheu	78 (40.77)	8 (3.74)	26 (12.99)
Lilongwe City	15 (19.55)	3 (4.14)	6 (8.07)
Mangochi	107 (44.33)	2 (0.90)	32 (12.08)
Machinga	95 (38.50)	9 (3.71)	40 (15.58)
Zomba Rural	67 (36.90)	8 (4.50)	22 (11.93)
Chiradzulu	48 (33.62)	9 (6.53)	19 (12.81)
Blantyre Rural	28 (32.84)	5 (5.53)	7 (8.00)
Mwanza	46 (31.24)	12 (7.03)	23 (14.55)
Thyolo	54 (34.43)	6 (3.78)	22 (13.29)
Mulanje	66 (36.91)	6 (3.58)	29 (16.30)
Phalombe	68 (33.09)	4 (2.03)	21 (10.31)
Chikwawa	55 (30.18)	8 (4.80)	21 (11.27)
Nsanje	48 (31.70)	17 (9.92)	27 (18.85)
Balaka	69 (32.73)	0 (0.00)	26 (13.19)
Neno	72 (45.28)	7 (4.70)	30 (18.44)
Zomba City	8 (17.75)	3 (5.56)	1 (1.93)
Blantyre City	28 (30.47)	2 (2.10)	7 (7.57)
Total	1,826 (36.82)	164 (2.79)	634 (11.58)

A Figura 5.1 mostra a prevalência observada da baixa estatura para a idade, do baixo peso para a idade e do baixo peso por distrito no Malawi. A partir da figura observamos que houve uma grande variação tanto na prevalência da baixa estatura para a idade, do baixo peso para a idade e do baixo peso entre os distritos. Assim, para capturar e modelar a variabilidade espacial dos dados, as prevalências da baixa estatura para a idade, do baixo peso para a idade e do baixo peso em cada distrito são analisadas através de seis modelos hierárquicos bayesianos diferentes (separados para cada *outcome*).

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

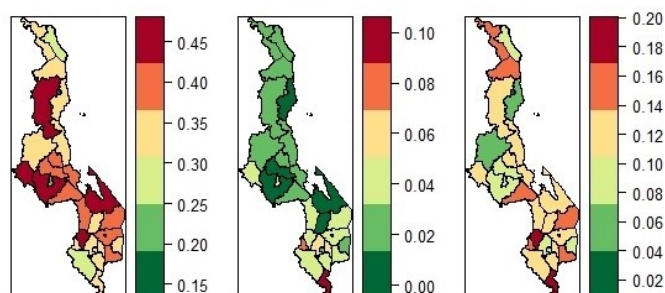


Figura 5.1: Mapas da prevalência observada de baixa estatura para a idade (esquerda), baixo peso para a idade (centro) e baixo peso (direita) por distrito gerados pelo estimador de HT no Malawi usando os dados do MDHS 2015-16.

Ajustaram-se os modelos 1-6 para produzir estimativas de nível distrital para esses indicadores. As estimativas dos parâmetros são apresentadas nas Tabelas 5.2 a 5.4. Em geral, os resultados mostram *intercept* semelhantes para todos os modelos e todos os *outcomes*, excepto para o Modelo 3 (AS) que está numa escala diferente. Os intervalos de credibilidade para cada modelo ajustado para os pesos são geralmente mais estreitos quando comparados com o modelo binomial não ajustado (UB).

Ao comparar os modelos com o DIC para ajuste e complexidade, verifica-se que o modelo 3 (AS) tem um desempenho ligeiramente melhor entre todos os modelos, para a baixa estatura para idade e baixo peso para idade (DIC=-86.69 para baixa estatura para idade; DIC=-94.76 para baixo peso para idade). Além disso, os resultados mostram que os modelos ajustados para os pesos parecem melhorar um pouco o desempenho quando comparados ao modelo 1 (UB).

Tabela 5.2: Comparação de modelos BYM para o mapeamento da baixa estatura para a idade no Malawi usando MDHS 2015-16.

Parâmetros	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
β_0 (CI)	-0.605 (-0.706;-0.547)	-0.585 (-0.67;-0.505)	0.634 (0.608;0.659)	-0.607 (-0.691;-0.526)	-0.594 (-0.68;-0.512)	-0.609 (-0.694;-0.528)
Sd	0.04	0.042	0.013	0.042	0.043	0.042
σ_u	0.210	0.170	0.069	0.219	0.188	0.218
σ_v	0.114	0.115	0.053	0.125	0.125	0.126
-2LL	-135.88	-24.45	12.88	-137.78	-133.44	-137.21
$p(D)$	18.54	16.47	23.56	19.55	17.64	19.32
DIC	226.85	6.24	-86.69	229.23	222.80	228.22

CAPÍTULO 5. MODELAÇÃO ESPACIAL E MAPEAMENTO DE DOENÇAS, AJUSTANDO PARA O DESENHO AMOSTRAL

Tabela 5.3: Comparação de modelos BYM para o mapeamento do baixo peso para idade no Malawi usando MDHS 2015-16.

Parâmetros	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
β_0 (CI)	-3.514 (-3.715;-3.325)	-3.458 (-3.661;-3.257)	0.166 (0.143;0.19)	-3.577 (-3.78;-3.386)	-3.658 (-3.87;-3.457)	-3.609 (-3.814;-3.416)
Sd	0.099	0.103	0.012	0.1	0.105	0.101
σ_u^2	0.493	0.353	0.061	0.486	0.519	0.467
σ_v^2	0.148	0.121	0.0481	0.144	0.152	0.143
$-2LL$	-95.22	-48.39	18.11	-94.04	-93.43	-92.31
$p(D)$	13.94	9.68	22.59	13.13	13.22	12.44
DIC	150.84	60.80	-94.76	148.73	147.06	145.98

Tabela 5.4: Comparação de modelos BYM para o mapeamento do baixo peso no Malawi usando MDHS 2015-16.

Parâmetros	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
β_0 (CI)	-2 (-2.109;-1.897)	-1.981 (-2.09;-1.877)	0.344 (0.32; 0.368)	-2.003 (-2.112;-1.901)	-2.022 (-2.133;-1.916)	-2.008 (-2.116;-1.905)
Sd	0.054	0.054	0.012	0.053	0.055	
σ_u	0.1390	0.1197	0.0634	0.1523	0.1400	0.1420
σ_v	0.1371	0.1139	0.0495	0.1286	0.1261	0.1321
$-2LL$	-117.62	-29.98	15.99	-117.38	-113.23	-115.35
$p(D)$	13.38	10.42	22.91	13.41	12.23	12.91
DIC	197.75	24.96	-89.77	197.28	190.55	193.68

A Figura 5.2 apresenta mapas das estimativas da prevalência da baixa estatura para idade em nível distrital. Embora se observe que o modelo UB apresente resultados bastante semelhantes às abordagens que levam em conta os pesos, observa-se algum impacto nos mapas de consideram os pesos no processo de modelação. Além disso, nota-se que a tendência espacial das estimativas dos modelos é mais suave em relação aos estimadores UW e HT, devido à inclusão do efeito aleatório espacial.

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

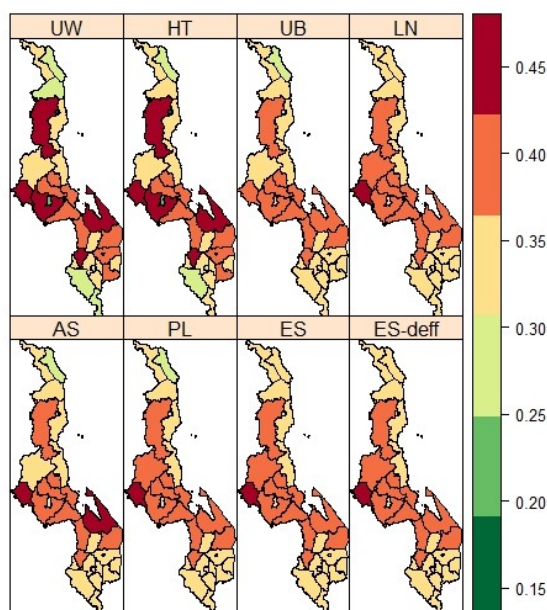


Figura 5.2: Mapas da prevalência observada (UW: Não ponderado, HT: Horvitz-Thompson) e espacialmente estimadas da baixa estatura para a idade (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Transformação de raiz quadrada de arco-seno estimador (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho efetivo de amostra usando efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.

Para comparar a prevalência estimada da baixa estatura para a idade por distrito a partir desses diferentes estimadores, apresentam-se os boxplots. Da Figura 5.3, observa-se que a média não ponderada e o estimador HT apresentam grande heterogeneidade entre os distritos, com alguns valores extremos. Essa variabilidade é substancialmente reduzida ao usar modelos hierárquicos Bayesianos. Além disso, no geral, a forma dos boxplots dos modelos são semelhantes, com os modelos ajustados apresentando uma localização ligeiramente inferior ao modelo UB.

Estes mapas de prevalência estimados revelam prevalência mais elevada nos distritos de Mzimba (na parte norte), Mchinji, Lilongwe (Rural), Dedza, Ntcheu e Dowa (na parte central), Mangochi e Neno (no sul).

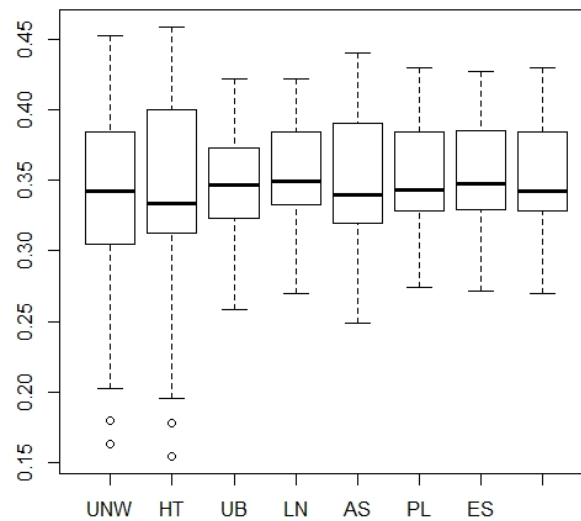


Figura 5.3: Boxplots da prevalência observada (UNW: não ponderado e HT: Horvitz Thompson) e prevalência estimada do baixo peso para a idade obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando efeito de desenho (Modelo 6)) em distritos no Malawi usando os dados do MDHS 2015-16.

A Figura 5.4 exibe mapas de estimativas de prevalência do baixo peso para a idade ao nível do distrito. Observa-se que o modelo UB apresenta resultados bastante semelhantes aos das demais abordagens. Em geral, as estimativas produzidas usando as diferentes abordagens baseadas em modelos são amplamente semelhantes às estimativas diretas, embora apareçam algumas diferenças entre estimativas diretas e as estimativas baseadas em modelos devido à suavização. Além disso, os mapas mostram *hotspots* claros para a prevalência do baixo peso para a idade nos distritos do Malawi, com alta prevalência na região sul do país, particularmente Nsange é a região mais afetada.

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

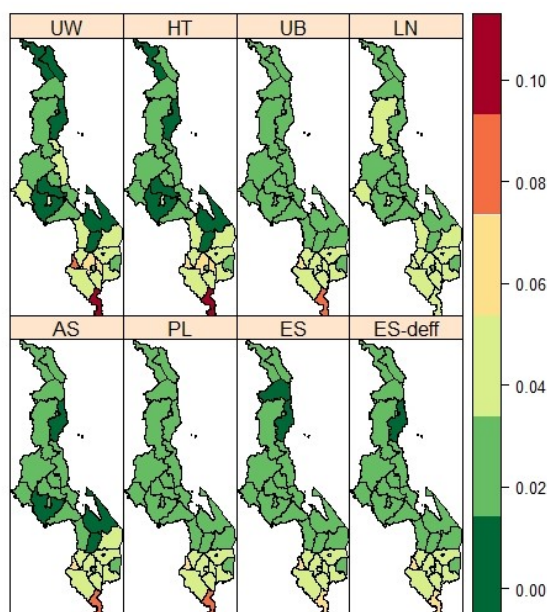


Figura 5.4: Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacial estimada da baixa estatura para a idade (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.

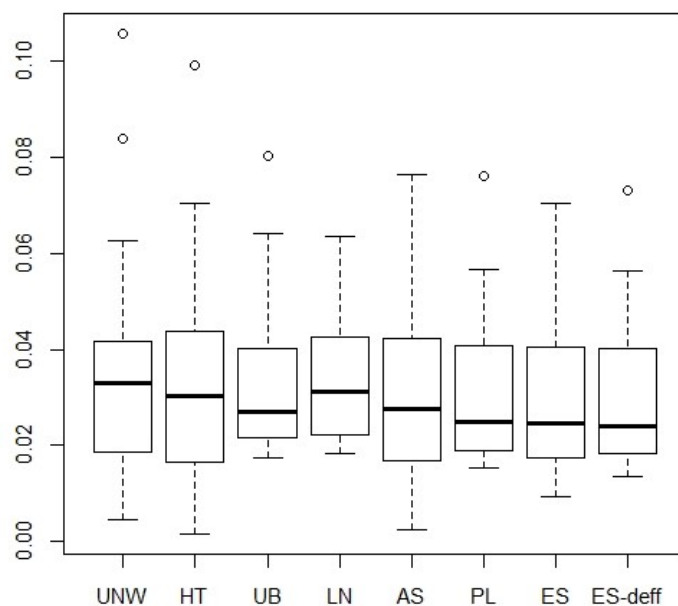


Figura 5.5: Boxplots da prevalência observada (UW: não ponderada e HT: Horvitz Thompson) e prevalência estimada da baixa estatura para a idade obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit-estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando efeito de desenho (Modelo 6)) nos distritos do Malawi usando os dados do MDHS 2015-16.

Dos boxplots da prevalência observada e estimada da baixa estatura para a idade por distrito obtidos usando diferentes estimadores, fica claro que o estimador UW e HT apresentam uma grande heterogeneidade entre os distritos (Figura 5.5). Essa variabilidade é reduzida ao usar os modelos hierárquicos.

A Figura 5.6 apresenta os mapas de prevalência do baixo peso para os estimadores UW, HT e indiretos. As tendências espaciais dos estimadores UW e HT são as mais variáveis. As tendências espaciais estimadas com base nos estimadores indiretos se comportam de maneira semelhante. Embora pareça haver uma distribuição do baixo peso semelhante em todo o país, uma tendência importante, com maior prevalência em Nsanje e Mwanza; e menor prevalência em Kasungu e Baía de Nkata.

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

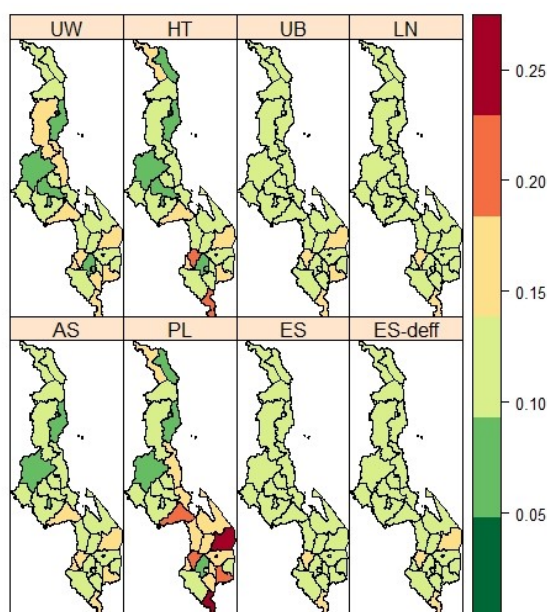


Figura 5.6: Mapas da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e espacialmente estimadas do baixo peso (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por distrito no Malawi usando os dados do MDHS 2015-16.

A partir dos boxplots da prevalência observada e estimada do baixo peso por distrito obtidos com diferentes estimadores, fica claro que os estimadores diretos (UW e HT) apresentam uma grande heterogeneidade entre os distritos (Figura 5.7). Essa variabilidade é substancialmente reduzida ao usar os modelos hierárquicos.

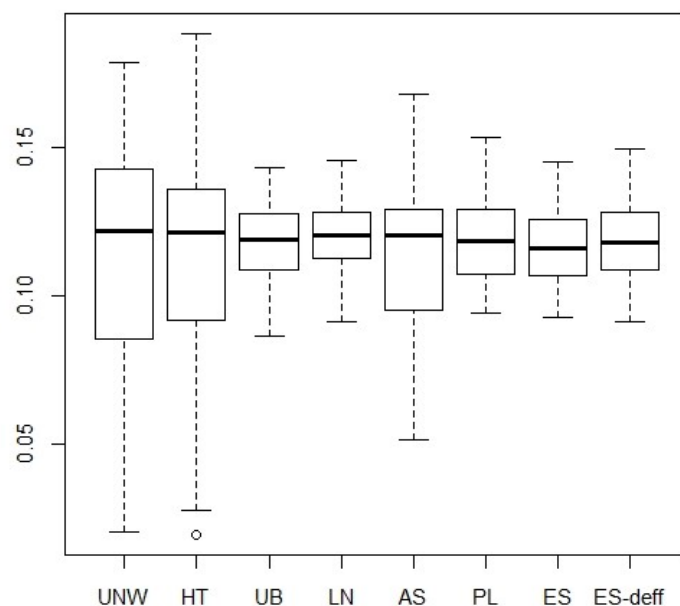


Figura 5.7: Boxplots da prevalência observada (UNW: não ponderado e HT: Horvitz Thompson) e prevalência estimada do baixo peso obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit-estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador do tamanho efetivo da amostra usando o efeito do desenho (Modelo 6)) nos distritos do Malawi estimado usando os dados do MDHS 2015-16.

5.4.2.2 Moçambique: Variações distritais de saúde nas taxas de prevalência de doenças na infância

Em geral, cerca de 29.37% (95% IC 26.99-31.87) das crianças menores de cinco anos tiveram febre e 11.11% (95% IC 9.93-12.41) tiveram diarreia, com variação entre as províncias de Moçambique, variando de 14.37% em Tete para 51.67% na Zambézia para febre e 6.8% em Tete para 17.19% no Niassa. Os resultados mostram uma variação na distribuição de diarreia e febre nas diferentes províncias em Moçambique.

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

Tabela 5.5: Resumo das taxas de prevalência observadas (ponderadas) de febre e diarreia por província em Moçambique.

Distrito	Febre (n,%)	Diarreia (n,%)
Niassa	146 (30.16)	94 (17.19)
Cabo Delgado	86 (21.94)	37 (9.90)
Nampula	228 (39.49)	64 (11.28)
Zambézia	264 (51.67)	90 (16.95)
Tete	65 (14.37)	39 (6.80)
Manica	81 (16.59)	43 (8.90)
Sofala	109 (21.58)	46 (8.43)
Inhambane	68 (18.23)	27 (7.24)
Gaza	141 (27.00)	61 (11.79)
Maputo Província	53 (15.86)	23 (8.25)
Maputo Cidade	70 (24.99)	28 (9.99)
Total	1,311 (29.37)	549 (11.11)

Ajustaram-se os modelos 1-6 detalhados na Subsecção 5.3 e 5.3.1, para produzir estimativas de febre a nível provincial e os modelos 1-5 para produzir estimativas de diarreia a nível provincial em Moçambique.

As estimativas dos parâmetros são apresentadas nas Tabelas 5.6 e 5.7, respetivamente para febre e para a diarreia. Os resultados mostram *intercept* semelhantes para os modelos hierárquicos Bayesianos não espaciais e espaciais, exceto para o Modelo 3 (AS) que está numa escala diferente, quando se considera a febre ou a diarreia. Os intervalos de credibilidade para cada um dos modelos ajustados aos pesos são geralmente mais estreitos quando comparados com os modelos não ajustados para os pesos (UB). O modelo 3 (AS) é o melhor modelo para febre (DIC=-37.64) e para diarreia (DIC=-39.66).

Tabela 5.6: Comparação do modelo BYM para mapeamento da febre em Moçambique.

Parâmetros	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
β_0 (CI)	-1.137 (-1.433;-0.844)	-1.123 (-1.453; -1.123)	0.523 (-0.45;0.597)	-1.129 (-1.496;-0.767)	-1.13 (-1.458;-0.805)	-1.13 (-1.458;-0.805)
Sd	0.146	0.162	0.036	0.183	0.161	
σ_u	0.1734	0.1863	0.1046	0.1854	0.1848	0.1848
σ_v	0.4660	0.5123	0.1094	0.5241	0.5172	0.5172
-2LL	-62.86	-16.20	-0.834	-63.98	-61.28	-61.28
$p(D)$	10.52	10.28	10.66	10.60	10.50	10.50
DIC	89.60	0.135	-37.64	89.54	86.86	86.86

Tabela 5.7: Comparação do modelo BYM para o mapeamento da diarreia em Moçambique.

Parâmetros	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
β_0 (CI)	-2.145 (-2.319;-1.979)	-2.14 (-2.335;-1.956)	0.329 (0.283;0.374)	-1.945 (-2.189;-1.708)	-2.152 (-2.34;-1.975)
Sd	0.085	0.095	0.023	0.118	0.091
σ_u^2	0.1651	0.1799	0.0748	0.2303	0.1724
σ_v^2	0.2257	0.2231	0.0640	0.3438	0.2442
$-2LL$	-50.27	-10.40	4.37	-53.70	-48.96
$p(D)$	8.75	7.90	10.15	9.83	8.64
DIC	81.04	4.30	-39.23	81.71	78.83

A Figura 5.8 apresenta a prevalência de febre obtida com base nos modelos não espaciais e espaciais entre as províncias. As tendências espaciais dos estimadores UW e HT são bastante semelhantes. Em geral, as estimativas produzidas usando as diferentes abordagens são amplamente semelhantes às estimativas diretas, embora algumas diferenças sejam aparentes devido à suavização.

Parece haver *hotspots* importantes bem evidenciados com maior prevalência de febre na parte norte de Moçambique (Nampula e Zambézia) e baixa prevalência em Maputo Cidade, Província de Maputo e Tete. Além disso, comparando o modelo UB com os modelos ajustados aos pesos, podemos observar alguma diferença nos mapas que levam em conta os pesos amostrais que podem ter impacto na intervenção das políticas.

A Figura 5.9 apresenta os boxplots da prevalência estimada de febre por província obtida usando diferentes estimadores, para os modelos espaciais. Observam-se diferenças nas estimativas de prevalência de febre usando diferentes estimadores e as estimativas ajustadas aos pesos apresentam semelhanças quase idênticas.

A prevalência estimada de diarreia em menores de cinco anos e a sua distribuição espacial estão mapeadas na Figura 5.11. Os mapas mostram uma distribuição desigual da prevalência da diarreia entre crianças menores de cinco anos, com maior prevalência nas províncias da Zambézia e Niassa e menor prevalência em Inhambane (sul do país).

5.4. APLICAÇÃO DA MODELAÇÃO ESPACIAL À PREVALÊNCIA DA MALNUTRIÇÃO INFANTIL NO MALAWI E DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

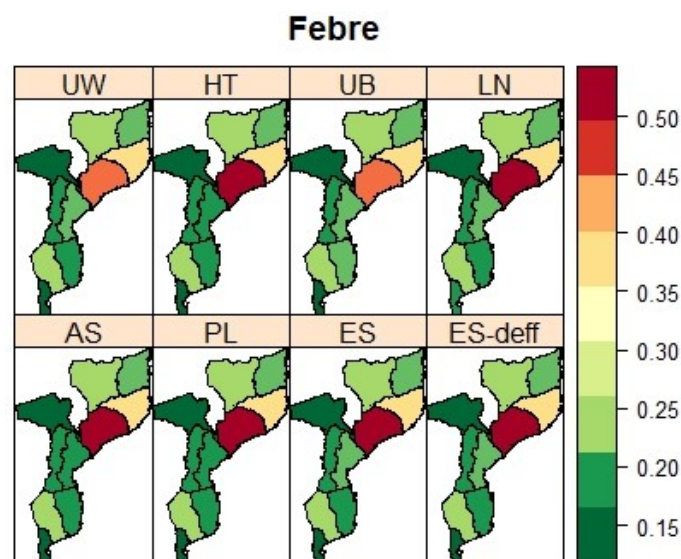


Figura 5.8: Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacial estimada de febre (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador de tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador de tamanho de amostra efetivo usando o efeito do desenho (Modelo 6)) por província em Moçambique usando o IMASIDA 2015.

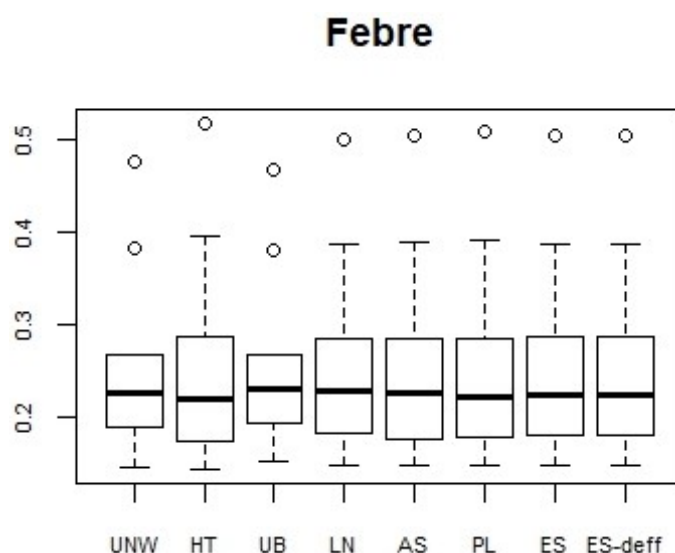


Figura 5.9: Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada de febre obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit-estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco-seno (Modelo 3), PL: Estimador de pseudo-verosimilhança (Modelo 4), ES: Estimador do tamanho de amostra efetivo (Modelo 5) e ES-deff: Estimador do tamanho efetivo da amostra usando efeito de desenho (Modelo 6)) em toda a província em Moçambique estimado a partir do IMASIDA 2015.

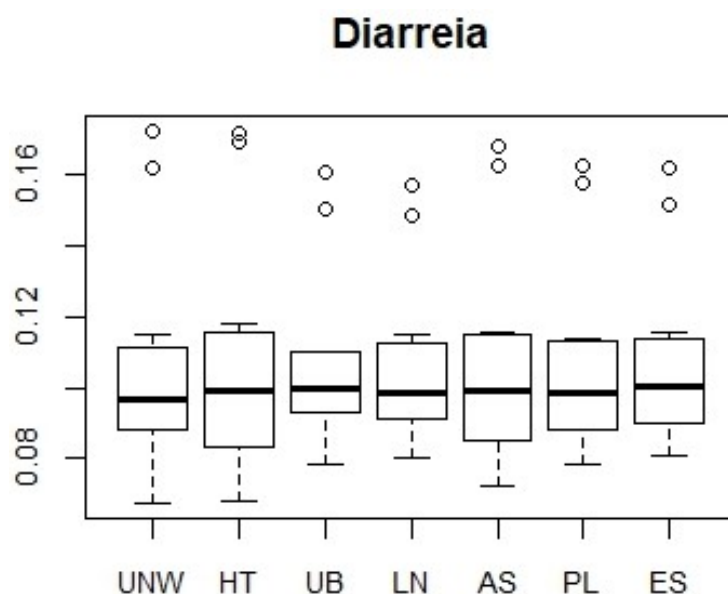


Figura 5.10: Boxplots da prevalência observada (UW: não ponderado e HT: Horvitz Thompson) e prevalência estimada de diarreia obtida usando várias abordagens (UB: Estimador binomial não ajustado (Modelo 1); LN: Logit- estimador normal (Modelo 2), AN: Estimador de transformação de raiz quadrada de arco (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4) e ES: Estimador de tamanho de amostra efetivo (Modelo 5)) em toda a província em Moçambique estimado a partir do IMASIDA 2015.

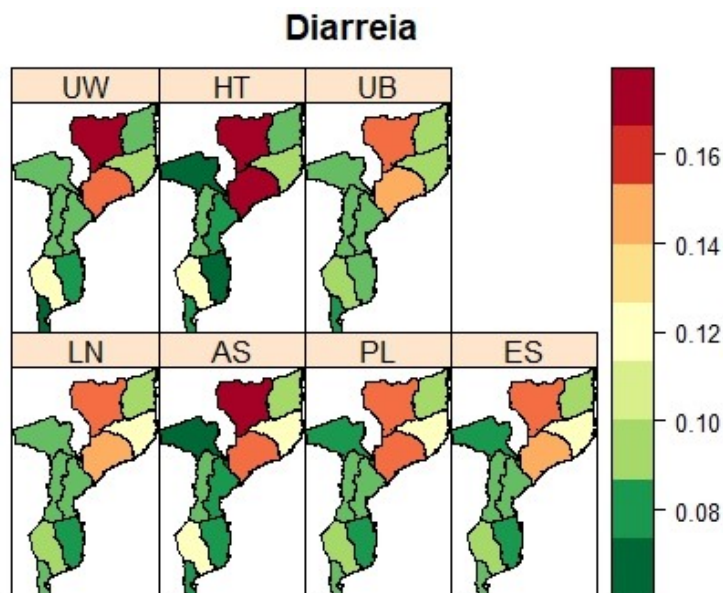


Figura 5.11: Mapas das prevalências observadas (UW: não ponderado e HT: Horvitz Thompson) e espacialmente estimadas de diarreia (UB: Estimador binomial não ajustado (Modelo 1); LN: Estimador Logit-normal (Modelo 2), AN: Estimador de transformação de raiz quadrada arcsine (Modelo 3), PL: Estimador de pseudo-verossimilhança (Modelo 4) e ES: Estimador de tamanho de amostra efetivo (Modelo 5)) por província em Moçambique usando o IMASIDA 2015.

5.4.3 Conclusão

Neste capítulo, compararam-se as estimativas de risco de doença entre áreas geográficas sob diferentes métodos de mapeamento espacial que levam em conta os pesos amostrais dos inquéritos de saúde. Em particular, usaram-se esses métodos para comparar a distribuição geográfica da baixa estatura para a idade, baixo peso para a idade e baixo peso entre crianças com menos de cinco anos ao nível distrital no Malawi usando o Inquérito Demográfico e de Saúde de 2015-16, e para a febre infantil e diarreia ao nível provincial em Moçambique usando o Inquérito de Indicadores de Imunização, Malária e HIV/SIDA de 2015.

Os métodos ajustados aos pesos foram os mais adequados, um resultado semelhante a estudos anteriores [1, 18, 75, 120, 121]. Embora o padrão espacial estimado tenha sido semelhante, os modelos que se ajustaram ao peso amostral produziram estimativas com menor variabilidade (intervalos de confiança reduzidos).

Assim, esta consideração dos pesos amostrais produziu estimativas de risco de doença que tiveram um nível de confiança maior. Existem algumas questões preocupantes decorrentes do nosso estudo. Em primeiro lugar, embora a transformação do arco-seno da prevalência ponderada tenha sido preferida neste estudo, ela não possui uma interpretação intuitiva da associação entre os *outcomes* binários e os preditores. Em segundo lugar, para o caso de Moçambique, mapeou-se a carga da doença utilizando a província com um nível muito mais grosseiro de agregação geográfica, no entanto, isso pode ter ocultado variações em alguma resolução espacial mais alta necessária para decisões de política local. Sugere-se para estudos futuros, realizar análises espaciais em resoluções espaciais mais altas, por exemplo, a nível distrital. Em terceiro lugar, estas análises usaram métodos espaciais univariados para as condições que poderiam ser correlacionadas em níveis ecológicos [49, 95]. Finalmente, não se realizou nenhum estudo de simulação para comparar o desempenho dos métodos estatísticos estudados para a estimativa de padrões espaciais de doenças usando dados complexos de pesquisas de saúde. No entanto, pensamos que isso não era necessário para este estudo, pois o objetivo era descrever os métodos de ajuste de pesos amostrais e aplicá-los para mapear a malnutrição, a diarreia e a febre. Outros trabalhos de pesquisa anteriores consideraram os seus desempenhos usando simulações [1, 18, 75, 120, 121].

Em conclusão, recomenda-se que os pesquisadores de epidemiologia espacial implementem métodos que considerem os pesos amostrais no mapeamento de doenças para estimar a distribuição espacial de riscos de doenças com base em dados complexos de pesquisas de saúde. As estimativas são mais precisas, fornecendo evidências de apoio confiáveis para orientar as políticas de saúde pública na focalização de recursos nas áreas de maior necessidade.

MODELAÇÃO ESPACIAL E MAPEAMENTO CONJUNTO DE DOENÇAS AJUSTANDO PARA O DESENHO AMOSTRAL

No capítulo anterior foram ajustados modelos de mapeamento de doenças para vários *outcomes* de saúde separadamente. No entanto, na prática, dados de múltiplas doenças estão disponíveis e o interesse pode estar na distribuição espacial de todas elas ou grupos populacionais, bem como na correlação entre as distribuições espaciais. Assim, neste capítulo esses modelos de mapeamento de doenças são estendidos para modelar conjuntamente doenças que parecem estar interligadas e produzir mapas suavizados da prevalência de doença. Além disso, ajustam-se esses modelos para o desenho da pesquisa.

6.1 Introdução

O peso da mortalidade infantil e da morbidade é o mais elevado nos países em desenvolvimento, com cerca de 12 milhões de mortes a ocorrerem todos os anos. Cerca de 70% dessas mortes são causadas por doenças comuns como diarreia, malária, malnutrição, tuberculose, entre outras. Crianças da SSA sofrem de múltiplas doenças que ocorrem frequentemente em simultâneo, em grande parte devido a fatores de risco partilhados ou sobrepostos, ou porque um distúrbio aumenta o risco do outro. As comorbidades na infância mais comumente observadas são a diarreia, a malária, o HIV e as doenças respiratórias agudas. Ultimamente, há um interesse crescente em compreender o grau de comorbidades nestas doenças, nomeadamente com a necessidade premente de uma gestão integrada das doenças infantis.

Os métodos de mapeamento de doenças são uma importante ferramenta para compreender as diferenças geográficas na saúde da população. Mapear a distribuição geográfica das taxas de incidência e prevalência de doença pode ajudar a compreender os padrões espaciais da doença e identificar diferenças na carga da doença numa área. A correta alocação geográfica dos recursos de saúde pode ser bastante aprimorada pela implantação de modelos estatísticos que permitem uma representação mais precisa das verdadeiras

taxas de doenças e os mapas resultantes podem ser usados por pessoas envolvidas no planejamento de serviços ou em programas de prevenção e controle de doença, tanto local quanto nacionalmente, e podem fornecer informações básicas úteis para tomada de decisões.

Grande parte do mapeamento de doenças concentra-se principalmente na modelação de uma única doença, mesmo quando várias doenças são analisadas. Isso pode ser devido ao fato de que essas doenças são impactadas pelo mesmo fator de risco [36, 51, 71]. Nesta linha, o mapeamento de doença tem aumentado ao longo dos anos, e estendeu-se desde o estudo de uma única doença de cada vez, para análise conjunta de várias doenças [36, 51, 71]. Atualmente, o mapeamento conjunto fornece informações úteis sobre a semelhança dos fatores de risco comuns e a abordagem é mais atraente do que a análise de resposta univariada. A modelação conjunta é apropriada quando diferentes doenças foram observadas nas mesmas unidades espaciais e quando as próprias doenças estão relacionadas entre si, diga-se, porque compartilham o mesmo conjunto de fatores de risco espacialmente distribuídos ou a presença de uma doença numa unidade espacial pode incentivar ou inibir a presença da outra doença na mesma unidade espacial. Os benefícios potenciais de um mapeamento multivariado de doenças incluem melhoria na precisão dos padrões de risco subjacentes de risco entre diferentes prevalências de doenças, facilidade de interpretação, capacidade de identificar padrões de risco compartilhados e específicos e melhoria na precisão do modelo [21, 36].

Knorr-Held sugeriram abordagens de modelação conjunta como uma melhoria das formulações de besag [10], compartilhando informações de diferentes doenças, as estimativas (taxas) podem ser melhoradas. Existem várias abordagens para a modelação conjunta espacial de doenças, e a sendo uma das abordagens mais usadas o modelo de componentes compartilhada [46, 51, 67]. No modelo de componente partilhado, a superfície de risco subjacente é decomposta em duas componentes, ou seja, a componente específica da doença, representando fatores espacialmente variáveis, e a componente compartilhada, que é um proxy de fatores espacialmente variáveis não observados que são comuns a ambas ou a todas as doenças [68]. Embora vários estudos ilustrem as vantagens da modelação espacial conjunta de múltiplas doenças ou de dados coletados de múltiplas fontes em comparação com a modelação univariada [21], essa abordagem ainda é pouco utilizada em estudos de contexto da SSA [69] e a maior parte dessas análises espaciais conjuntas, baseadas em dados obtidos de pesquisas com delineamento complexos, não levaram em conta os pesos da pesquisa. Recentemente, esforços foram feitos para estender os métodos de mapeamento de doenças espaciais e incorporar o desenho de pesquisa nos modelos espaciais [18, 75].

Assim, este capítulo descreve os modelos de componente espacial bivariada que incorpora os pesos da pesquisa no processo de modelação, levando em consideração a estrutura de correlação entre as doenças. Os métodos foram então aplicados aos dados do IMASIDA 2015, para investigar as distribuições geográficas conjuntas da febre e da diarreia infantil em Moçambique.

6.2 Modelo de Componente Compartilhada

Modelos de mapeamento conjunto de doenças são uma extensão direta de modelos espaciais univariados que usam tanto a estrutura de dependência espacial global e local para modelar o risco de doenças. As extensões permitem que os analistas façam uma avaliação sobre semelhanças, bem como diferenças entre fatores riscos para as doenças que compartilham perfis de risco comuns [21, 36, 46]. Um desses modelos de mapeamento conjunto de doenças é o modelo de componente compartilhada que foi introduzido por Held et al. [36]. A ideia foi emprestada por [51], no seu trabalho na detecção de *clusters* de doenças conjuntas. O caso de duas doenças é mais comumente aplicado na literatura, onde se descreve a associação na prevalência das duas doenças e a segunda componente introduzida é específica de uma das doenças. As duas componentes são formulados para incorporar os fatores espaciais não observados que afetam o risco da(s) doença(s) [36, 51]. Uma descrição detalhada do modelo de componente compartilhada é dada abaixo.

Seja y_{ijk} o *status* da doença k , para o indivíduo j na região i ($i = 1, 2, \dots, I$; $j = 1, \dots, n_i$; $k = 1$ (doença 1), 2 (doença 2)) e atribui-se um valor de 1 se o indivíduo tem a doença e 0 caso contrário, onde n_i é o número de indivíduos que foram amostradas na área i . O *outcome* y_{ik} do número total de indivíduos com a doença k segue uma distribuição de binomial com P_{ik} sendo a probabilidade da doença k ocorrer na área i , $y_{ik} = \sum_{j \in s_i} y_{ijk}$. O modelo de geração de dados é definido como

$$y_{ik}|P_{ik} \sim \text{Binomial}(n_{ik}, P_{ik}). \quad (6.1)$$

É modelada a variação entre áreas em P_{ik} usando o modelo de efeitos aleatórios de área, que em modelos separados, os efeitos aleatórios são introduzidos da seguinte forma

$$\text{logit}(P_{ij1}) = \beta_1 + u_{1i} + v_{1i}, \quad (6.2)$$

$$\text{logit}(P_{ij2}) = \beta_2 + u_{2i} + v_{2i}, \quad (6.3)$$

onde β_1 e β_2 são os *intercept* para a doença 1 e doença 2, respectivamente; u_1 e u_2 são os efeitos aleatórios espacialmente estruturados e são modelados usando prioris condicionais autoregressivas intrínsecas independentes enquanto v_1 e v_2 são efeitos de heterogeneidade aleatória e são modelados usando distribuições normais independentes. Esses modelos podem ser aplicados a diferentes doenças para produzir estimativas de risco espacial que podem ser mapeadas e analisadas para identificar padrões particulares de alto risco. Doenças com etiologias semelhantes podem apresentar padrões semelhantes, ou seja, variação espacial semelhante, e uma análise multivariada pode ser realizada para obter melhores estimativas desses padrões compartilhados. Ao mesmo tempo, o modelo deve permitir desvios específicos do padrão compartilhado em certas áreas. A fim de construir modelos de conjuntos para k doenças diferentes, será apresentado um dos modelos de conjunto comumente usado, o modelo de componente compartilhada. Num modelo conjunto, os riscos relativos incluem termos que são compartilhados por várias doenças, ou seja, a superfície de risco é dividida em componente compartilhada e específica da

doença, respetivamente, conforme [36, 51]. O efeito dos efeitos compartilhados pode ser ponderados, de modo que esses pesos meçam a dependência da distribuição geográfica da doença no padrão compartilhado. Assim, o modelo resultante é dado por

$$\text{logit}(P_{ij1}) = \beta_1 + u_{1i}\delta + u_{2i} + v_{1i}, \quad (6.4)$$

$$\text{logit}(P_{ij2}) = \beta_2 + \frac{u_{1i}}{\delta} + v_{2i}, \quad (6.5)$$

onde β_1 e β_2 são os *intercept* para a doença 1 e doença 2, respetivamente; $\mathbf{u}_1$ é a componente compartilhada enquanto $\mathbf{u}_2$ é a componente específica apenas da primeira doença. Essas duas componentes são modeladas usando prioris condicionais autoregressivas intrínsecas com parâmetros de precisão τ_1 e τ_2 respetivamente; $\mathbf{v}_1$ e $\mathbf{v}_2$ são efeitos de heterogeneidade aleatória e são modelados usando distribuição normal. O parâmetro extra $\delta(> 0)$ é incluído para permitir o gradiente diferencial da componente compartilhada para as duas doenças e é dado uma priori como $\log \delta \sim N(0, \sigma^2)$. A razão dos parâmetros de escala δ a $\frac{1}{\delta}$ compara o peso da doença 1 em relação à doença 2 associada a componente compartilhada. Neste modelo, as tendências espaciais comuns são investigadas usando a componente compartilhada, e as tendências divergentes são investigadas através da componente específica da doença. Os parâmetros de escala permitem que cada doença tenha uma associação única com o padrão geral da doença. Para interpretação da componente compartilhada, um valor de δ próximo de um indica que a doença 1 e a doença 2 têm uma magnitude similar de associação com o padrão geral da doença, enquanto um valor grande e positivo de δ indica que a doença 1 tem uma associação mais forte com o padrão geral da doença do que a doença 2.

Estes modelos têm sido amplamente aplicados na prática usando dados de pesquisa, no entanto não levam em conta o desenho da pesquisa (exemplo os pesos amostrais). Uma alternativa é usar modelo de pseudo-verosimilhança que levam em consideração o desenho da pesquisa, como o modelo de pseudo-verosimilhança que usa uma verosimilhança ponderada [75], onde o número total de indivíduos com a doença k na área i (y_{ik}) são ponderados usando os pesos do desenho normalizados. Assim, em vez de considerar os *outcomes* binomiais usados em 6.1, aqui usa-se $y_{ik}^{PL} = \sum_{j \in s_i} \tilde{w}_{ij}^d y_{ijk}$. O modelo resultante pode ser descrito como

$$y_{ik}^{PL} | P_{ik} \sim \text{Binomial}(n_{ik}, P_{ik})$$

$$\text{logit}(P_{i1}) = \beta_1 + u_{1i}\delta + u_{2i} + v_{1i}, \quad (6.6)$$

$$\text{logit}(P_{i2}) = \beta_2 + \frac{u_{1i}}{\delta} + v_{2i}, \quad (6.7)$$

onde β_1 e β_2 são os *intercept* para a doença 1 e doença 2, respetivamente. Os efeitos aleatórios (u_{1i}, u_{2i}) são assumidos como seguem uma distribuição ICAR, e os efeitos aleatórios (v_{1i}, v_{2i}) seguem uma distribuição normal.

6.2.1 Inferência Bayesiana, Computação e Avaliação de Modelos

A estimação do modelo foi realizada usando a abordagem Bayesiana e distribuições a priori apropriadas foram especificadas para todos os parâmetros dos modelos. Os interceptos específicos da doença β_k são atribuídas prioris uniformes impróprias e prioris $\text{Gamma}(0.5, 0.008)$ para os parâmetros de precisão espacial e não espacial. O logaritmo do parâmetro de escala (ou seja, $\log \delta$) foi atribuída uma distribuição normal com média zero e precisão de $1/5.9$. Adicionalmente, as especificações das prioris seguiram as ideias nas aplicações anteriores do modelo como em [96].

Modelos bivariados de componente compartilhada não ponderados (Modelo 1) e modelos bivariados de componente compartilhada usando a pseudo-verossimilhança (Modelo 2) foram ajustados. Os modelos foram ajustados no *software* R, os parâmetros dos modelos foram então estimados usando a abordagem INLA [105] e o pacote INLAMSM (<https://CRAN.R-project.org/package=INLAMSM>) recentemente desenvolvido por Palmí-Perales et al. (2019) [96] para implementar modelos espaciais multivariados para dados de áreas usando o R-INLA. A comparação dos modelos foi feita com base no DIC, onde o melhor modelo foi aquele com o menor valor de DIC [111].

6.3 Aplicação da modelação espacial conjunta à prevalência da febre e da diarreia infantil em Moçambique

6.3.1 Dados e variáveis do estudo

Os dados analisados para o propósito do estudo foram do IMASIDA 2015 e encontram-se descritos no Capítulo 2 desta tese. Os *outcomes* considerados neste estudo são os *status* da febre e da diarreia. Foram incluídas na análise as crianças menores de cinco anos cuja mãe respondeu se tiveram diarreia ou febre nas últimas 2 semanas. As demais crianças com valores ausentes para os desfechos foram excluídas deste estudo. Assim, as análises incluem um total de 4,969 crianças menores de cinco anos.

6.3.2 Resultados

Do total de 4,969 crianças incluídas no estudo, 26.34% apresentaram febre, 11.11% apresentaram diarreia, enquanto 5.74% apresentaram febre e diarreia em simultâneo (Tabela 6.1). Os resultados mapeados na Figura 6.1 mostram claramente alguns *hotspots* relativos à distribuição da febre e da diarreia nas diferentes províncias. Além disso, a febre foi mais prevalente do que a diarreia.

CAPÍTULO 6. MODELAÇÃO ESPACIAL E MAPEAMENTO CONJUNTO DE DOENÇAS AJUSTANDO PARA O DESENHO AMOSTRAL

Tabela 6.1: Tabela cruzada da prevalência da febre e da diarreia infantil em Moçambique, usando os dados do IMASIDA 2015.

Febre	Diarreia		Total
	não	sim	
não	3,393 (68.28)	267 (5.37)	3,660 (5.73)
sim	1,024 (20.61)	285 (5.74)	1,309 (26.34)
Total	4,417 (88.89)	552 (11.11)	4,969 (100.0)

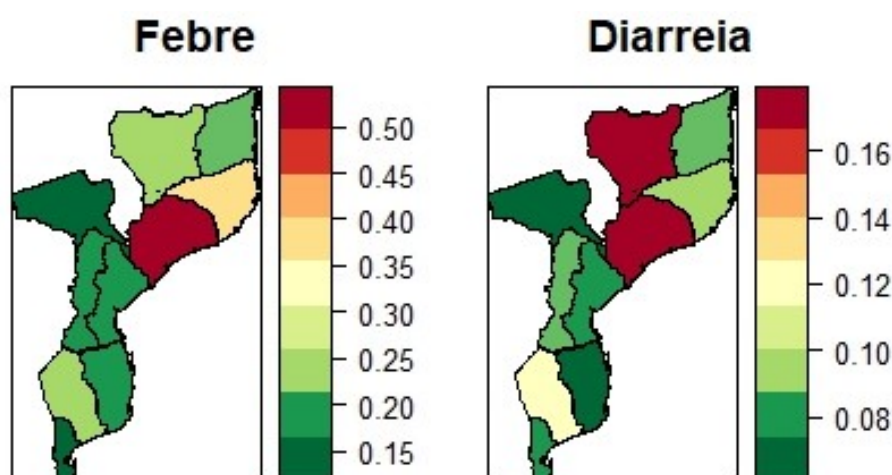


Figura 6.1: Mapas da prevalência observada (ponderada) da febre e diarreia por províncias em Moçambique usando IMASIDA 2015.

Aplicaram-se os modelos de componente compartilhada ajustando e não ajustando aos pesos da pesquisa para produzir estimativas de nível provincial da prevalência da febre e da diarreia. Ao comparar os modelos com o DIC para ajuste e complexidade, verifica-se que embora o valor do DIC para o modelo 1 e modelo 2 sejam bastante semelhantes, o modelo 2 ajustado aos pesos tem um ajuste melhor em relação ao modelo 1 que não leva em consideração os pesos da pesquisa (Tabela 6.2). O efeito relativo das variações compartilhadas devido à prevalência de febre em comparação com a prevalência de diarreia é medido por δ . Isso é estimado como $\delta = 1.12$, e sugere que os fatores de risco não observados têm efeito relativamente maior sobre a febre do que a diarreia (Modelo 2). De fato, $\delta^2 = 1.25$ demonstra que o gradiente da prevalência da febre em cada área foi superior ao da diarreia.

6.3. APLICAÇÃO DA MODELAÇÃO ESPACIAL CONJUNTA À PREVALÊNCIA DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

Tabela 6.2: Medidas de comparação dos modelos.

	Modelo 1	Modelo 2
DIC	168.08	167.23
pd	19.92	19.91
-2LL	-135.98	-136.31

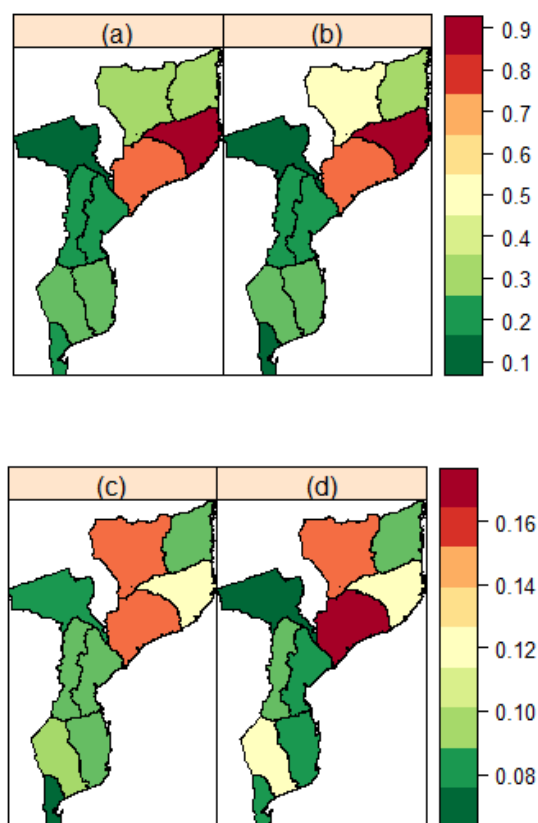


Figura 6.2: Mapas da prevalência estimada da febre (não ponderada (a), ponderada (b)) e diarreia (não ponderada (c), ponderada (d)) por províncias em Moçambique usando IMASIDA 2015.

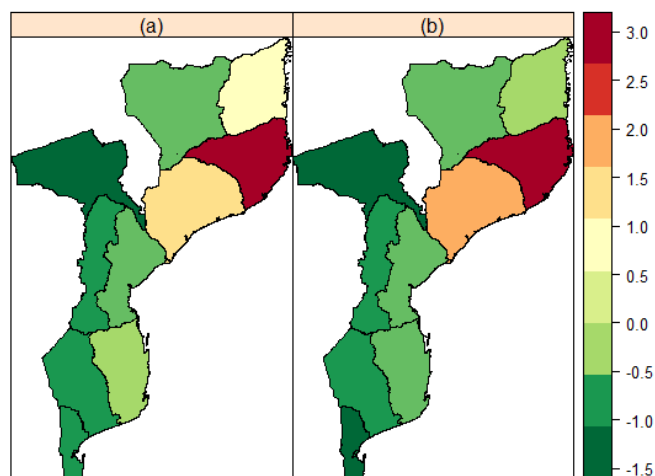


Figura 6.3: Mapas das estimativas a posteriori dos efeitos espaciais compartilhados da febre e da diarreia por províncias em Moçambique usando IMASIDA 2015 ((a) não ponderadas, (b) ponderadas).

As estimativas específicas da área para diarreia e febre em Moçambique são apresentadas na Figura 6.2. A coluna da esquerda fornece as estimativas não ponderadas, enquanto a coluna da direita as estimativas ponderadas. Os efeitos espaciais específicos da doença mostram áreas onde a febre e a diarreia diferiram (Figura 6.2). Parece haver *hotspots* da febre na parte norte de Moçambique (Nampula e Zambézia) e baixa prevalência na Cidade de Maputo, Província de Maputo e em Tete (Figura 6.2). Adicionalmente, observa-se que as províncias da Zambézia e Niassa são as mais afetadas pela diarreia e a área menos afetada do país é a província de Inhambane. Os efeitos aleatórios espacialmente estruturados compartilhados são apresentados na Figura 6.3. Os efeitos compartilhados foram altos principalmente nas províncias da Zambézia e Nampula. Além disso, ao comparar as estimativas ponderadas e não ponderadas observa-se alguma diferença nos mapas, o que pode ter impacto na intervenção das políticas.

6.3.3 Conclusão

Este estudo examinou simultaneamente os componentes espaciais comuns e específicas da prevalência da febre e diarreia infantil em Moçambique. Empregou-se o modelo de componente espacial compartilhada, o que permitiu dividir o risco geográfico de doenças em padrões espaciais comuns e incomuns. Este estudo pode ser considerado uma extensão para a análise espacial univariada, num contexto de aplicação estatística sobre doenças infantis na SSA, a maioria das quais se baseou no uso de modelos espaciais univariados [69]. Este estudo foi baseado numa análise anterior de Cassy *et al* [15] usando os mesmos

6.3. APLICAÇÃO DA MODELAÇÃO ESPACIAL CONJUNTA À PREVALÊNCIA DA FEBRE E DA DIARREIA INFANTIL EM MOÇAMBIQUE

dados, que ajustaram modelos separados para febre e diarreia. Embora o modelo agrupado pareça confirmar os resultados obtidos em Cassy *et al* [15], a análise agrupada mediu adicionalmente o grau de correlação espacial entre a prevalência de febre e diarreia. Essa é uma das características atraentes da modelação conjunta, pois permite isolar *clusters* de áreas de variação comum e incomum para as duas doenças.

Os resultados deste estudo mostraram que os métodos ajustados ao peso foram os mais adequados e que a comorbidade a nível provincial da febre e da diarreia foram predominantes nas províncias da Zambézia e Nampula. Estes resultados são consistentes com estudos anteriores que encontraram maior prevalência da febre e diarreia entre crianças que vivem nas províncias da Zambézia e Nampula. O padrão de prevalência partilhada também pode ser explicado por condições perenes de transmissão da malária. Num estudo de [46], a prevalência de ambas as doenças aumentam em áreas altamente endémicas da malária. Não é muito claro porque a prevalência da febre é mais pronunciada do que a da diarreia. Este facto deve ser investigado com maior profundidade. O uso de modelos espaciais bivariados ou multivariados em estudos epidemiológicos focados em países em desenvolvimento tem sido relativamente pequeno. Além disso, ainda tem havido pouco trabalho sobre distribuição espacial da febre e da diarreia com foco em nível sub-regional em Moçambique ajustando para o desenho da pesquisa, e este trabalho ajuda a preencher esta lacuna de conhecimento. Esta análise pode ser uma ferramenta valiosa para a orientação baseada em evidências de recursos para melhorar a saúde infantil. Atualmente, em países como Moçambique, estão a ser implementadas várias intervenções eficazes separadamente. Para um maior impacto, é necessário aplicar os recursos em conjunto como parte da estratégia do para atingir a meta dos ODS até 2030. Além disso, para um controlo eficaz, as intervenções devem ser orientadas espacialmente, por exemplo, os mapas dos efeitos partilhados poderia ser usado para planear a gestão integrada das duas doenças. Este mapa também pode ser útil para testar hipóteses relacionadas com os perfis epidemiológicos das duas doenças. Este é o grande destaque deste estudo em contraste com modelos separados de regressão.

CONCLUSÕES E TRABALHO FUTURO

7.1 Conclusões

Este trabalho concentrou-se em métodos voltados à modelação de *outcomes* de saúde basendo nos dados de inquéritos demográficos e de saúde. Mais especificamente, com métodos estatísticos para dados de resultados binários, que vem ganhando cada vez mais atenção no campo da estatística aplicada especialmente na área da saúde. Na análise de dados de pesquisa complexa, como a amostra é estratificada e conglomerada em várias etapas, com probabilidade de seleção desigual e ignorando efeitos do desenho como conglomeração, a estratificação e os pesos, geralmente levam a estimativas tendenciosas de parâmetros, erro padrão, intervalos de confiança e testes estatísticos enganosos. O desafio de analisar dados complexos é como incorporar as características do desenho no procedimento da estimação.

A fim de analisar dados com desenho da pesquisa complexa, foram aplicadas três abordagens: regressão logística de pesquisa, modelos multinível e modelos espaciais. A regressão logística de pesquisa foi usada para modelar o uso da rede mosquiteira por mulheres em idade reprodutiva em Moçambique e, como mencionado, a regressão logística de pesquisa levou em conta o desenho da pesquisa, usando os dados do DHS 2011 de Moçambique. Os resultados, em geral, mostram que as estimativas dos coeficientes da regressão logística tendo em conta o plano amostral são semelhantes as estimativas não ponderadas. Contudo, os erros padrão das estimativas ficam subestimados quando as características do plano amostral não são consideradas na estimação. Isto faz com que os coeficientes estimados para modelos de regressão, por exemplo, sejam considerados significativos quando na verdade não são. Estes resultados são similares aos apresentados por Korn e Graubard [52] que descrevem que as estimativas ao considerar-se os pesos, quando estes têm pouca variabilidade são similares às estimativas obtidas ao considerar-se a AAS. Os modelos multinível foram outra abordagem usada para modelar a prevalência do uso da rede mosquiteira. Dos resultados verificaram-se diferenças nos erros padrão entre os métodos ajustados e não ajustados para o plano amostral. A abordagem padrão de modelação multinível pode estimar parâmetros e erros padrão adequadamente para

dados agrupados que resultaram de uma amostragem com igual probabilidade de seleção. No entanto, com probabilidades de seleção desiguais, essa abordagem padrão não estima corretamente os parâmetros do modelo e os seus erros padrão.

Os resultados deste estudo mostraram que os *intercept* aleatório ao nível agregado (variância) foi grande e estatisticamente significativo, indicando diferenças entre os *clusters*. Ou seja, cerca de 38.08% das variações do uso da rede mosquiteira é atribuída às diferenças entre os *clusters*. O que suporta a aplicação dos modelos multinível para este estudo particular.

A idade, nível de escolaridade, se está grávida, número de membro no agregado familiar, número de redes mosquiteiras no agregado, quinto de riqueza e o estado civil foram os fatores, ao nível individual, significativamente associado com o uso da rede mosquiteira. Além disso, os resultados também mostram que ao nível agregado, a província e a área de residência onde a mulher reside desempenha um papel significativo no uso da rede mosquiteira.

Concomitantemente, foram usados métodos da estatística espacial, onde compararam-se as estimativas de risco de doença entre áreas geográficas sob diferentes métodos e a variação obtida entre um conjunto de métodos de mapeamento espacial que levam em conta os pesos amostrais dos inquéritos de saúde. Em particular, usaram-se esses métodos para comparar a distribuição geográfica do baixo peso para a idade, baixo peso para a altura e baixo peso entre crianças com menos de cinco anos ao nível distrital no Malawi usando o Inquérito Demográfico e de Saúde de 2015-16, e para a febre e diarreia infantil ao nível provincial em Moçambique usando o Inquérito de Indicadores de Imunização, Malária e HIV/SIDA de 2015. Os métodos ajustados aos pesos foram os mais adequados. Embora o padrão espacial estimado tenha sido semelhante, esta consideração dos pesos amostrais produziu estimativas de risco de doença que tiveram um nível de confiança maior, tanto na análise univariada quanto na análise bivariada. O uso dos métodos que consideram os pesos amostrais no mapeamento de doenças para estimar a distribuição espacial de riscos de doenças com base em dados complexos de pesquisas de saúde forneceu estimativas mais precisas, e também evidências de apoio confiáveis para orientar as políticas de saúde pública na focalização de recursos nas áreas de maior necessidade.

7.2 Recomendações

Para o caso de Moçambique, mapeou-se a carga da doença utilizando a província que tem um nível muito mais grosseiro de agregação geográfica, no entanto, isso pode ter ocultado variações em alguma resolução espacial mais alta necessária para decisões de política local. Sugere-se para estudos futuros, realizar análises espaciais em resoluções espaciais mais altas, por exemplo, a nível distrital e também considerando-se covariáveis. Além disso, é previsto explorar os modelos logísticos multiníveis no contexto de amostras provenientes de delineamentos complexos, com recurso ao *software* R e usando um pacote recentemente desenvolvido para o efeito. Além disso, os modelos em consideração

atendem apenas a dados de áreas; dados geoestatísticos e dados de padrões pontuais não foram aprofundados neste trabalho e ficando como trabalho futuro. Adicionalmente, esta prevista uma extensão dos modelos espaciais de forma a acomodar a componente temporal.

Em conclusão, recomenda-se que os pesquisadores de epidemiologia espacial implementem métodos que considerem os pesos amostrais no mapeamento de doenças para estimar a distribuição espacial de riscos de doenças com base em dados complexos de pesquisas de saúde. As estimativas são mais precisas, fornecendo evidências de apoio confiáveis para orientar as políticas de saúde pública na focalização de recursos nas áreas de maior necessidade.

Os dados do DHS têm sido uma importante fonte de monitorização e avaliação das tendências de saúde na maioria dos países da SSA. Os resultados apresentados nesta tese podem ser uma ferramenta valiosa para a orientação baseada em evidências de recursos para melhorar a saúde da mulher e da criança e consequentemente melhorar a qualidade de vida da sociedade. No entanto, para um controlo eficaz, sugere-se que as intervenções sejam mais espacialmente orientadas. Pois pode ajudar a otimizar a utilização dos recursos a todos os níveis da prestação de cuidados de saúde.

BIBLIOGRAFIA

- [1] T. N. Achia. «Spatial modelling and mapping of female genital mutilation in Kenya». Em: *BMC public health* 14.1 (2014), p. 276 (ver pp. 47, 67).
- [2] A. Agresti et al. «Random-Effects Modeling of Categorical Response Data». Em: *Sociological Methodology* 30.1 (2000), pp. 27–80. URL: <https://doi.org/10.1111/0081-1750.t01-1-00075> (ver pp. 20, 21).
- [3] S. K. Appiah, U. Mueller e J. Cross. «Spatio-temporal modelling of malaria incidence for evaluation of public health policy interventions in Ghana, West Africa». Em: (2011) (ver p. 47).
- [4] K. J. Archer, S. Lemeshow e D. W. Hosmer. «Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design». Em: *Computational Statistics & Data Analysis* 51.9 (2007), pp. 4450–4464. URL: <https://dx.doi.org/10.1016/j.csda.2006.07.006> (ver pp. 23, 25).
- [5] T. Asparouhov e B. Muthén. «Multilevel modeling of complex survey data». Em: *Proceedings of the joint statistical meeting in seattle*. 2006, pp. 2718–2726 (ver p. 35).
- [6] P. L. Aung, K. M. Win e K. L. Show. «Utilization of insecticide-treated bed nets among pregnant women in Myanmar–analysis of the 2015–2016 Demographic and Health Survey». Em: *PloS one* 17.3 (2022), e0265262 (ver p. 33).
- [7] K.-m. Ban. «Sustainable Development Goals». Em: (2016) (ver p. 6).
- [8] E. Barankanira et al. «Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention». Em: *BMC public health* 16.1 (2015), p. 118 (ver p. 47).
- [9] D. M. Bates. *lme4: Mixed-effects modeling with R*. 2010 (ver p. 38).
- [10] J. Besag, J. York e A. Mollié. «Bayesian image restoration, with two applications in spatial statistics». Em: *Annals of the institute of statistical mathematics* 43.1 (1991), pp. 1–20 (ver pp. 50, 70).

- [11] D. A. Binder. «On the variances of asymptotically normal estimators from complex surveys». Em: *International Statistical Review/Revue Internationale de Statistique* (1983), pp. 279–292 (ver pp. 2, 3, 19).
- [12] R. Bivand, V. Gómez-Rubio e H. Rue. «Spatial data analysis with R-INLA with some extensions». Em: American Statistical Association. 2015 (ver p. 52).
- [13] R. Carroll et al. «Bayesian model selection methods in modeling small area colon cancer incidence». Em: *Annals of epidemiology* 26.1 (2016), pp. 43–49 (ver p. 47).
- [14] S. R. Cassy, I. Natário e M. R. Martins. «Logistic Regression Modelling for Complex Survey Data with an Application for Bed Net Use in Mozambique». Em: *Open Journal of Statistics* 6.05 (2016), p. 898. URL: <https://dx.doi.org/10.4236/ojs.2016.65074> (ver p. 19).
- [15] S. R. Cassy et al. «Accounting for Sampling Weights in the Analysis of Spatial Distributions of Disease Using Health Survey Data, with an Application to Mapping Child Health in Malawi and Mozambique». Em: *International Journal of Environmental Research and Public Health* 19.10 (2022). ISSN: 1660-4601. DOI: 10.3390/ijerph19106319. URL: <https://www.mdpi.com/1660-4601/19/10/6319> (ver pp. 47, 76, 77).
- [16] R. L. Chambers e C. J. Skinner. *Analysis of survey data*. John Wiley & Sons, 2003 (ver pp. 3, 19).
- [17] K. Chantala e C. Suchindran. «Adjusting for unequal selection probability in multilevel models: A comparison of software packages». Em: *Proceedings of the American Statistical Association, Seattle, WA: American Statistical Association* (2006) (ver p. 35).
- [18] C. Chen, J. Wakefield e T. Lumely. «The use of sampling weights in Bayesian hierarchical models for small area estimation». Em: *Spatial and spatio-temporal epidemiology* 11 (2014), pp. 33–43. URL: <https://doi.org/10.1016/j.sste.2014.07.002> (ver pp. 3, 47–52, 67, 70).
- [19] L. A. Chimoyi e E. Musenge. «Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011». Em: *BMC public health* 14.1 (2014), p. 555 (ver p. 47).
- [20] G. M. Cordeiro e C. G. Demétrio. «Modelos lineares generalizados e extensões». Em: *São Paulo* (2008) (ver pp. 21, 36, 38).
- [21] A. R. Dabney e J. C. Wakefield. «Issues in the mapping of two diseases». Em: *Statistical methods in medical research* 14.1 (2005), pp. 83–112 (ver pp. 70, 71).
- [22] S. Das, H. Chandra e U. R. Saha. «District level estimates and mapping of prevalence of diarrhoea among under-five children in Bangladesh by combining survey and census data». Em: *PloS one* 14.2 (2019), e0211062 (ver pp. 47, 48).

- [23] M. De Onis et al. «WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age». Em: (2006) (ver p. 6).
- [24] A. Diallo et al. «Factors associated with malaria preventive measures among pregnant women in Guinea». Em: *Infectious diseases in obstetrics and gynecology* 2021 (2021) (ver p. 33).
- [25] S. S. Dias et al. «Estimating risk factors for HIV infection among women in Mozambique using population-based survey data». Em: *African Journal of AIDS Research* 17.1 (2018), pp. 62–71 (ver p. 19).
- [26] M. Elkasabi, R. Ren e T. W. Pullum. «Multilevel modeling using DHS Surveys: a framework to approximate level-weights». Em: ICF. 2020 (ver p. 41).
- [27] I. N. de Estatística (Mozambique) e M. M. da Saúde. *Moçambique: Inquérito Demográfico e de Saúde 2011*. Calverton, Maryland, USA: MISAU, INE e ICFI, 2013 (ver pp. 2, 5, 14, 39).
- [28] I. N. de Estatística (Mozambique) e M. M. da Saúde. *Moçambique: inquérito demográfico e de saúde, 2011*. Instituto Nacional de Estatística, 2013 (ver p. 2).
- [29] E. Gaspar Manuel da Costa. *Moçambique, Inquérito Demográfico e de Saúde, 1997*. Calverton, Maryland, USA: Instituto Nacional de Estatística e Macro Internacional Inc, 1998 (ver p. 2).
- [30] A. Gelman e J. Hill. *Data analysis using regression and multilevel hierarchical models*. Vol. 1. Cambridge University Press New York, NY, USA, 2007 (ver pp. 3, 38).
- [31] E. Giorgi et al. «Geostatistical methods for disease mapping and visualisation using data from spatio-temporally referenced prevalence surveys». Em: *International Statistical Review* 86.3 (2018), pp. 571–597 (ver p. 47).
- [32] H. Goldstein. *Multilevel statistical models*. Vol. 922. John Wiley & Sons, 2011 (ver pp. 8, 38).
- [33] H. Goldstein. «Nonlinear multilevel models, with an application to discrete response data». Em: *Biometrika* (1991), pp. 45–51 (ver p. 36).
- [34] D. Haile et al. «Exploring spatial variations and factors associated with childhood stunting in Ethiopia: spatial and multilevel analysis». Em: *BMC pediatrics* 16.1 (2016), p. 49 (ver p. 47).
- [35] M. of Health - MoH/Malawi e I. International. *Malawi Service Provision Assessment 2013-14*. Malawi: MoH/Malawi e ICF International, 2014 (ver p. 2).
- [36] L. Held et al. «Towards joint disease mapping». Em: *Statistical methods in medical research* 14.1 (2005), pp. 61–82 (ver pp. 70–72).
- [37] B. HORVTTZ e D. Thompson. «A generalization of sampling without replacement from a finite universe». Em: *J. Am. Statist. Ass* 47 (1952), pp. 663–85 (ver p. 49).

- [38] D. W. Hosmer Jr, S. Lemeshow e R. X. Sturdivant. *Applied logistic regression*. Vol. 398. John Wiley & Sons, 2013 (ver pp. 7, 20, 22–24).
- [39] A.-K. Iddrisu e Y. A. Amoako. «Spatial modeling and mapping of tuberculosis using bayesian hierarchical approaches». Em: *Open Journal of Statistics* 6.3 (2016), pp. 482–513 (ver p. 47).
- [40] e. I. M. Instituto Nacional de Saúde (INS) Instituto Nacional de Estatística (INE). *Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique 2009*. Calverton, Maryland, EUA: INS, INE e ICF Macro, 2010 (ver p. 2).
- [41] Instituto Nacional de Saúde/INS et al. *Mozambique Inquérito Nacional sobre Indicadores de Malária (IIM)2018*. Rel. téc. INS/Mozambique, INE, PNCM, ICF, 2019. URL: <http://dhsprogram.com/pubs/pdf/MIS33/MIS33.pdf> (ver p. 2).
- [42] I. International. *Demographic and health survey sampling and household listing manual*. 2012 (ver pp. 17, 49).
- [43] N.-B. Kandala et al. «Geographic distribution of cardiovascular comorbidities in South Africa: a national cross-sectional analysis». Em: *Journal of Applied Statistics* 41.6 (2014), pp. 1203–1216 (ver p. 47).
- [44] N.-B. Kandala et al. «Geographic variation of hypertension in Sub-Saharan Africa: a case study of South Africa». Em: *American journal of hypertension* 26.3 (2013), pp. 382–391 (ver p. 47).
- [45] T. Kapwata e S. Manda. «Geographic assessment of access to health care in patients with cardiovascular disease in South Africa». Em: *BMC Health Services Research* 18.1 (2018), pp. 1–10. ISSN: 14726963. DOI: 10.1186/s12913-018-3006-0 (ver p. 47).
- [46] L. N. Kazembe, C. C. Appleton e I. Kleinschmidt. «Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi». Em: *Geospatial Health* 2.1 (2007), pp. 41–50 (ver pp. 70, 71, 77).
- [47] B. Keeley, C. Little e E. Zuehlke. *The state of the world's children 2019: children, food and nutrition: growing well in a changing world*. 2019 (ver p. 6).
- [48] S. P. Kigozi et al. «Spatial-temporal patterns of malaria incidence in Uganda using HMIS data from 2015 to 2019». Em: *BMC public health* 20.1 (2020), pp. 1–14 (ver p. 47).
- [49] D. K. Kinyoki et al. «Assessing comorbidity and correlates of wasting and stunting among children in Somalia using cross-sectional household surveys: 2007 to 2010». Em: *Bmj Open* 6.3 (2016), e009854 (ver p. 67).
- [50] L. Kish. *Survey sampling*. 04; HN29, K5. John Wiley & Sons, 1965 (ver pp. 2, 3, 52).
- [51] L. Knorr-Held e N. G. Best. «A shared component model for detecting joint and selective clustering of two diseases». Em: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164.1 (2001), pp. 73–85 (ver pp. 70–72).

-
- [52] E. L. Korn e B. I. Graubard. *Analysis of health surveys*. Vol. 323. John Wiley & Sons, 2011 (ver p. 79).
 - [53] R. Kumar et al. «Predictors of knowledge and use of long-lasting insecticidal nets for the prevention of malaria among the pregnant women in Pakistan». Em: *Malaria Journal* 20.1 (2021), pp. 1–7 (ver p. 33).
 - [54] J. Larmarange e V. Bendaud. «HIV estimates at second subnational level from national population-based surveys». Em: *AIDS (London, England)* 28.4 (2014), S469 (ver p. 47).
 - [55] E. S. Lee e R. N. Forthofer. *Analyzing complex survey data*. Vol. 71. Sage Publications, 2005 (ver pp. 22–24).
 - [56] A. H. Leyland e H. Goldstein. *Multilevel modelling of health statistics*. Wiley, 2001 (ver p. 8).
 - [57] B. Liu, P. Lahiri e G. Kalton. «Hierarchical Bayes modeling of survey-weighted small area proportions». Em: *Proceedings of the American Statistical Association, Survey Research Section*. 2007, pp. 3181–3186 (ver p. 35).
 - [58] J. M. Lourenço. *The NOVAthesis L^AT_EX Template User’s Manual*. NOVA University Lisbon. 2021. URL: <https://github.com/joaomlourenco/novathesis/raw/master/template.pdf> (ver p. iii).
 - [59] T. C. Lucas et al. «Mapping malaria by sharing spatial information between incidence and prevalence data sets». Em: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 70.3 (2021), pp. 733–749 (ver p. 47).
 - [60] T. Lumley. «Survey: Analysis of complex survey samples (R package version 3.30)». Em: *Vienna, Austria: R Foundation for Statistical Computing* (2014) (ver p. 52).
 - [61] T. Lumley et al. «Analysis of complex survey samples». Em: *Journal of Statistical Software* 9.1 (2004), pp. 1–19. URL: <https://dx.doi.org/10.18637/jss.v009.i08> (ver pp. 26, 48).
 - [62] T. Lumley. *Complex surveys: a guide to analysis using R*. Vol. 565. John Wiley & Sons, 2011 (ver pp. 23, 24, 26, 27).
 - [63] T. Lumley e A. Scott. «AIC and BIC for modeling with complex survey data». Em: *Journal of Survey Statistics and Methodology* 3.1 (2015), pp. 1–18. URL: <https://dx.doi.org/10.1093/jssam/smu021> (ver p. 25).
 - [64] T. Lumley e A. Scott. «Tests for regression models fitted to survey data». Em: *Australian & New Zealand Journal of Statistics* 56.1 (2014), pp. 1–14. URL: <https://dx.doi.org/10.1111/anzs.12065> (ver p. 24).
 - [65] T. Maiti. «Hierarchical Bayes estimation of mortality rates for disease mapping». Em: *Journal of Statistical Planning and Inference* 69.2 (1998), pp. 339–348 (ver p. 47).

- [66] C. B. Makura et al. «Choropleth mapping of cervical cancer screening in South Africa using healthcare facility-level data from the National Laboratory Network». Em: *AIMS public health* 3.4 (2016), p. 849 (ver p. 47).
- [67] S. Manda. «Joint mapping modelling for multiple health problems in South Africa». Em: *Retrieved January 23* (2011), p. 2013 (ver p. 70).
- [68] S. Manda, R. Feltbower e M. Gilthorpe. «Review and empirical comparison of joint mapping of multiple diseases». Em: *Southern African Journal of Epidemiology and Infection* 27.4 (2012), pp. 169–182 (ver p. 70).
- [69] S. Manda, N. Haushona e R. Bergquist. «A scoping review of spatial analysis approaches using health survey data in sub-Saharan Africa». Em: *International Journal of Environmental Research and Public Health* 17.9 (2020). ISSN: 16604601. DOI: 10.3390/ijerph17093070 (ver pp. 48, 70, 76).
- [70] S. Manda et al. «Mapping HIV prevalence using population and antenatal sentinel-based HIV surveys: a multi-stage approach». Em: *Population health metrics* 13.1 (2015), p. 22 (ver p. 47).
- [71] S. O. Manda, R. G. Feltbower e M. S. Gilthorpe. «Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes». Em: *European journal of epidemiology* 24.12 (2009), pp. 743–752 (ver p. 70).
- [72] K. Mansori et al. «Spatial inequalities in the incidence of colorectal cancer and associated factors in the neighborhoods of Tehran, Iran: Bayesian spatial models». Em: *Journal of preventive medicine and public health* 51.1 (2018), p. 33 (ver p. 47).
- [73] C. E. McCulloch e J. M. Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001 (ver pp. 8, 20, 36, 38).
- [74] C. McSweeney, M. New e G. Lizcano. «UNDP climate change country profiles Malawi». Em: (2014) (ver p. 12).
- [75] L. Mercer et al. «A comparison of spatial smoothing methods for small area estimation with sampling weights». Em: *Spatial Statistics* 8 (2014), pp. 69–85 (ver pp. 3, 47–52, 67, 70, 72).
- [76] I. N. d. E. (e. I. Ministério da Saúde (MISAU). *Inquérito de Indicadores de Imunização, Malária e HIV/SIDA em Moçambique 2015*. Maputo, Moçambique. Rockville, Maryland, EUA: INS, INE e ICF, 2015 (ver pp. 2, 15).
- [77] I. MISAU. *Inquérito de Indicadores de Imunização, Malária e HIV/SIDA em Moçambique (IMASIDA) 2015: Relatório de Indicadores Básicos*. 2016 (ver p. 48).
- [78] C. Mulher e C. Criança. «Estratégia Global para a Saúde das Mulheres, das Crianças e dos Adolescentes (2016-2030)[Internet]». Em: *Nova York* (2015) (ver p. 1).
- [79] J. A. Nelder e R. J. Baker. *Generalized linear models*. Wiley Online Library, 1972 (ver pp. 7, 20, 21).

- [80] N. M. C. P. -. NMCP/Malawi e ICF. *Malawi Malaria Indicator Survey 2017*. NMCP/Malawi e ICF International, 2018 (ver p. 2).
- [81] N. M. C. P. -. NMCP/Malawi e I. International. *Malawi Malaria Indicator Survey 2014*. NMCP/Malawi e ICF International, 2015 (ver p. 2).
- [82] NSO. «Malawi Demographic and Health Survey 2015-16». Em: *National Statistics Office The DHS Program* (2017), pp. 1–658. URL: <http://dhsprogram.com/pubs/pdf/FR319/FR319.pdf> (ver pp. 2, 16, 48).
- [83] N. S. O. -. NSO/Malawi e I. Macro. *Malawi Demographic and Health Survey 2010*. National Statistical Office e ICF Macro, 2011 (ver p. 2).
- [84] N. S. O. -. NSO/Malawi e O. Macro. *Malawi Demographic and Health Survey 2004*. National Statistical Office, 2005 (ver p. 2).
- [85] M. N. S. Office e M. International. *Malawi Knowledge, Attitudes, and Practices in Health Survey 1996*. National Statistical Office e ICF Macro, 1997 (ver p. 2).
- [86] M. N. S. Office e M. I. I. for Resource Development. *Malawi Demographic and Health Survey 2000*. National Statistical Office, 2001 (ver p. 2).
- [87] M. N. S. Office, M. I. I. for Resource Development. *Demographic e H. Surveys. Malawi Demographic and Health Survey: Summary report*. National Statistical Office, 1992 (ver p. 2).
- [88] N. S. Office. *2018 Malawi Population and Housing Census: Main Report*. 2018 (ver p. 12).
- [89] E. Okango, H. Mwambi e O. Ngesa. «Spatial modeling of HIV and HSV-2 among women in Kenya with spatially varying coefficients». Em: *BMC Public Health* 16.1 (2016), pp. 1–13. ISSN: 14712458. DOI: 10.1186/s12889-016-3022-0. URL: <http://dx.doi.org/10.1186/s12889-016-3022-0> (ver pp. 47, 48).
- [90] O. Oliveira et al. «Using Bayesian spatial models to map and to identify geographical hotspots of multidrug-resistant tuberculosis in Portugal between 2000 and 2016». Em: *Scientific reports* 10.1 (2020), pp. 1–9 (ver p. 47).
- [91] W. H. Organization et al. «Levels and trends in child mortality report 2017.» Em: *Levels and trends in child mortality report 2017*. (2017) (ver p. 1).
- [92] W. H. Organization et al. «The state of health in the WHO African Region: an analysis of the status of health, health services and health systems in the context of the Sustainable Development Goals». Em: (2018) (ver pp. 1, 4).
- [93] W. H. Organization. *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*. World Health Organization, 2018 (ver pp. 1, 4).
- [94] W. H. Organization et al. *World malaria report 2009*. World Health Organization, 2009 (ver p. 4).

- [95] O. S. Orunmoluyi, E. Gayawan e S. Manda. «Spatial Co-Morbidity of Childhood Acute Respiratory Infection, Diarrhoea and Stunting in Nigeria». Em: *International Journal of Environmental Research and Public Health* 19.3 (2022), p. 1838 (ver p. 67).
- [96] F. Palmí-Perales, V. Gómez-Rubio e M. A. Martínez-Beneito. «Bayesian multivariate spatial models for lattice data with INLA». Em: *arXiv preprint arXiv:1909.10804* (2019) (ver p. 73).
- [97] D. G. C. Pessoa e P. d. Silva. «Análise de dados amostrais complexos». Em: *São Paulo: Associação Brasileira de Estatística* 112 (1998) (ver pp. 2, 23).
- [98] D. Pfeffermann. «New important developments in small area estimation». Em: *Statistical Science* 28.1 (2013), pp. 40–68 (ver p. 49).
- [99] N. M. C. Programme-NMCP/Malawi e I. International. *Malawi Malaria Indicator Survey 2012*. National Statistical Office e ICF Macro, 2012 (ver p. 2).
- [100] S. Rabe-Hesketh e A. Skrondal. «Multilevel modelling of complex survey data». Em: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169.4 (2006), pp. 805–827. URL: <https://doi.org/10.1111/j.1467-985X.2006.00426.x> (ver pp. 35, 36, 51).
- [101] T. E. Raghunathan et al. «Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening». Em: *Journal of the American Statistical Association* 102.478 (2007), pp. 474–486 (ver p. 51).
- [102] J. N. Rao e I. Molina. *Small area estimation*. John Wiley & Sons, 2015 (ver p. 49).
- [103] G. Roberts, N. K. Rao e S. Kumar. «Logistic regression analysis of sample survey data». Em: *Biometrika* 74.1 (1987), pp. 1–12 (ver p. 23).
- [104] E. Rohner et al. «Mapping the cervical cancer screening cascade among women living with HIV in Johannesburg, South Africa». Em: *International Journal of Gynecology & Obstetrics* 152.1 (2021), pp. 53–59 (ver p. 47).
- [105] H. Rue, S. Martino e N. Chopin. «Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations». Em: *Journal of the royal statistical society: Series b (statistical methodology)* 71.2 (2009), pp. 319–392 (ver pp. 50, 52, 73).
- [106] L. G. Sambo et al. «Uma Década de Acção da OMS na Região Africana: Trabalhando juntos para alcançar os objectivos de saúde». Em: (2014) (ver p. 1).
- [107] C.-E. Särndal, B. Swensson e J. Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 2003 (ver p. 49).
- [108] B. Schrödle e L. Held. «Spatio-temporal disease mapping using INLA». Em: *Environmetrics* 22.6 (2011), pp. 725–734. URL: <https://dx.doi.org/10.1002/env.1065> (ver p. 52).

- [109] C. J. Skinner, D. Holt e T. F. Smith. *Analysis of complex surveys*. Wiley, 1989 (ver pp. 3, 19).
- [110] J. Snow. *On the mode of communication of cholera*. 1855 (ver p. 47).
- [111] D. J. Spiegelhalter et al. «Bayesian measures of model complexity and fit». Em: *Journal of the royal statistical society: Series b (statistical methodology)* 64.4 (2002), pp. 583–639 (ver pp. 52, 73).
- [112] D. Tafadzwa et al. «Spatiotemporal modelling and mapping of cervical cancer incidence among HIV positive women in South Africa: a nationwide study». Em: *International journal of health geographics* 20.1 (2021), pp. 1–12 (ver p. 47).
- [113] R. C. Team et al. «R: A language and environment for statistical computing». Em: (2013). URL: citeseer.ist.psu.edu/artho03highlevel.html (ver pp. 26, 52).
- [114] K. N. Turi, M. J. Christoph e D. S. Grigsby-Toussaint. «Spatial distribution of underweight, overweight and obesity among women and children: Results from the 2011 Uganda demographic and health survey». Em: *International Journal of Environmental Research and Public Health* 10.10 (2013), pp. 4967–4981. ISSN: 16604601. DOI: 10.3390/ijerph10104967 (ver p. 47).
- [115] M. A. A. Turkman e G. L. Silva. «Modelos Lineares Generalizados: da teoria à prática». Em: *VIII Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa*. 2000 (ver pp. 7, 21).
- [116] UNICEF et al. *WHO, World Bank. Joint child malnutrition estimates—levels and trends (2017 edition)*. Geneva: World Health Organization, 2017 (ver p. 6).
- [117] Unicef et al. *Committing to child survival: a promise renewed*. Rel. téc. eSocialSciences, 2015 (ver p. 1).
- [118] Unicef et al. «Levels and trends in child malnutrition: joint child malnutrition estimates». Em: (2018) (ver p. 6).
- [119] Unicef et al. «Progress for every child in the SDG era». Em: (2018) (ver p. 6).
- [120] Y. Vandendijck et al. «Model-based inference for small area estimation with sampling weights». Em: *Spatial Statistics* 18 (2016), pp. 455–473. ISSN: 22116753. DOI: 10.1016/j.spasta.2016.09.004. URL: <http://dx.doi.org/10.1016/j.spasta.2016.09.004> (ver pp. 47–49, 67).
- [121] K. Watjou et al. «Spatial small area smoothing models for handling survey data with nonresponse». Em: *Statistics in Medicine* 36.23 (2017), pp. 3708–3745. ISSN: 10970258. DOI: 10.1002/sim.7369 (ver pp. 47–49, 67).
- [122] A. Weimann, D. Dai e T. Oni. «A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa : A comparison between 2008 and 2012». Em: *Social Science & Medicine* 163 (2016), pp. 144–156. ISSN: 0277-9536. DOI: 10.1016/j.socscimed.2016.06.055. URL: <http://dx.doi.org/10.1016/j.socscimed.2016.06.055> (ver p. 47).

BIBLIOGRAFIA



1. **Identify the main components of the system.** What are the inputs, outputs, and internal processes?