UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

MODELAÇÃO ESPACIAL BAYESIANA DA PREVALÊNCIA DE ANEMIA EM CRIANÇAS DE 6-59 MESES EM MOÇAMBIQUE

JOSÉ SAMUEL JÚNIOR

Maputo

2026

UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

MODELAÇÃO ESPACIAL BAYESIANA DA PREVALÊNCIA DE ANEMIA EM CRIANÇAS DE 6-59 MESES EM MOÇAMBIQUE

JOSÉ SAMUEL JÚNIOR

Dissertação submetida ao Programa de Pós-graduação em Estatística Aplicada da Universidade Eduardo Mondlane, como parte dos requisitos para a obtenção do grau de Mestre em Estatística Aplicada.

Orientador: Doutor Adelino Martins

Co-Orientador: Prof. Doutor Osvaldo Loquiha

Maputo

2026

UNIVERSIDADE EDUARDO MONDLANE
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

MODELAÇÃO ESPACIAL BAYESIANA DA PREVALÊNCIA DE ANEMIA EM CRIANÇAS DE 6-59 MESES EM MOÇAMBIQUE

JOSÉ SAMUEL JÚNIOR

Dissertação submetida ao Programa de Pós-graduação em Estatística Aplicada da Universidade Eduardo Mondlane, como parte dos requisitos para a obtenção do grau de Mestre em Estatística Aplicada. Aprovado em 30 de Abril de 2026, por:

_____

Prof. Doutor Cachimo Combo Assane
(Arguente - DMI-UEM)

_____

Porf. Doutor Alberto Mulenga
(Presidente - DMI-UEM)

_____

Doutor Adelino Martins
(Orientador- DMI-UEM)

_____

Prof. Doutor Osvaldo Loquiha
(Co-Orientador - DMI-UEM)

_____

Maputo
Fevereiro 2026

Dedicatória

Este trabalho é dedicado, em primeiro lugar, à minha esposa Yániki Cossa Samuel, pelo amor, paciência e companheirismo incondicionais ao longo desta caminhada, e ao meu filho Aydan Adriel dos Anjos Uamusse, cuja alegria e inocência foram a maior fonte de motivação e esperança em todos os momentos.

Dedico também aos meus pais, José Samuel e Pricília Rendição Samuel pelo exemplo de coragem, dedicação e pelo apoio incansável que sempre me concederam, acreditando nos meus sonhos mesmo quando pareciam distantes.

A todos vocês, a minha eterna gratidão.

AGRADECIMENTOS

Agradeço, em primeiro lugar, a Deus, pela força, saúde e sabedoria concedidas ao longo de todo o percurso desta jornada. Expresso a minha profunda gratidão ao Doutor Adelino Martins e ao Prof. Doutor Osvaldo Loquiha, pela orientação dedicada, paciência e valiosas contribuições científicas, fundamentais para a concretização deste trabalho. À minha esposa, Yániki Cossa Samuel, e ao meu filho, Aydan Adriel dos Anjos Uamusse, pelo amor, compreensão e incentivo constantes, que foram alicerces essenciais em cada etapa desta caminhada. Aos meus pais, José Samuel e Pricília Rendição Samuel pelo exemplo de dedicação, pelos valores transmitidos e pelo apoio incondicional que sempre me concederam. Aos colegas e amigos do curso de Estatística Aplicada, em especial Francisco Salmina, Lucas Jango e Kelven Sozinho, pelo companheirismo, partilha de conhecimento e apoio mútuo ao longo desta trajetória académica. A todos os que, de algum modo, participaram ou colaboraram na elaboração desta dissertação, deixo o meu sincero apreço e gratidão.

RESUMO

A anemia infantil constitui um problema de saúde pública em Moçambique, apresentando elevada prevalência e marcadas desigualdades geográficas. Este estudo teve como objectivo estimar a prevalência de anemia em crianças dos 6 aos 59 meses, identificar os factores associados e analisar a distribuição espacial do risco ao nível provincial, utilizando dados do Inquérito Demográfico e de Saúde de Moçambique 2022–2023. A amostra analítica incluiu 3 387 crianças com informação válida sobre anemia. Para o tratamento dos dados foram utilizadas duas abordagens: Análise por casos completos e imputação múltipla por cadeias de equações. Os factores associados foram avaliados através de Modelos Lineares Generalizados Mistos, enquanto a componente espacial foi analisada por meio de modelos Bayesianos do tipo BYM2, estimados pela Aproximação Integrada Aninhada de Laplace. A prevalência nacional ponderada de anemia foi de aproximadamente 63%. A idade da criança apresentou associação negativa com a ocorrência de anemia, ao passo que o sexo feminino, níveis mais elevados de riqueza do agregado familiar e maior escolaridade materna estiveram associados a menor probabilidade da doença. Os resultados evidenciaram heterogeneidade espacial residual entre províncias, mesmo após o ajustamento pelas covariáveis consideradas. A análise com imputação múltipla apresentou robustez das estimativas obtidas. Adicionalmente, o modelo espacial Bayesiano ponderado apresentou melhor desempenho do que a especificação não ponderada, segundo os critérios de informação utilizados. Conclui-se que a anemia infantil em Moçambique resulta da interacção de factores demográficos, socioeconómicos e espaciais. Os resultados reforçam a necessidade de intervenções focalizadas nas áreas de maior risco e demonstram a utilidade da modelação espacial Bayesiana como instrumento de apoio à tomada de decisão e à definição de políticas de saúde pública.

Palavras-chave: Anemia infantil; Distribuição espacial; Inferência Bayesiana; INLA; Modelos Lineares Generalizados Mistos.

ABSTRACT

Childhood anaemia remains a major public health challenge in Mozambique, characterised by high prevalence and marked geographical disparities. This study aimed to estimate the prevalence of anaemia among children aged 6–59 months, identify associated factors, and analyse the spatial distribution of risk at the provincial level using data from the 2022–2023 Mozambique Demographic and Health Survey. The analytical sample comprised 3,387 children with valid anaemia information. Two approaches were used to address missing data: complete-case analysis and Multiple Imputation by Chained Equations. Associated factors were assessed using Generalized Linear Mixed Models, while the spatial component was modelled through Bayesian BYM2 models estimated using the Integrated Nested Laplace Approximation framework. The weighted national prevalence of anaemia was approximately 63%. Child age was negatively associated with the likelihood of anaemia, whereas female sex, higher household wealth, and higher maternal education were associated with a lower probability of the condition. The results revealed substantial residual spatial heterogeneity across provinces, even after adjustment for the selected covariates. The multiple imputation analysis confirmed the robustness of the estimates obtained. Furthermore, the weighted Bayesian spatial model outperformed the unweighted specification according to the information criteria considered. The findings indicate that childhood anaemia in Mozambique is influenced by a combination of demographic, socioeconomic, and spatial factors. These results highlight the need for geographically targeted interventions in high-risk areas and demonstrate the value of Bayesian spatial modelling as a tool to support evidence-based public health planning and decision-making.

Keywords: Bayesian inference; Childhood anaemia; Generalised Linear Mixed Models; Spatial distribution.

SUMÁRIO

1 INTRODUÇÃO **1**

1.1 Problema de pesquisa 3

1.2 Objectivos 5

1.2.1 Objectivo Geral 5

1.2.2 Objectivos Específicos 5

1.3 Justificação 5

1.4 Estrutura da Dissertação 6

2 REVISÃO DA LITERATURA **7**

2.1 Anemia Infantil: Conceitos, Definição e Classificação 7

2.2 Epidemiologia da Anemia Infantil 8

2.3 Determinantes da Anemia Infantil 10

2.4 Evidência Empírica sobre a Anemia Infantil em África e Moçambique 12

2.5 Evolução das Abordagens Metodológicas no Estudo da Anemia Infantil 15

2.6 Aplicação de Modelos Multinível no Estudo da Anemia Infantil 16

2.7 Aplicação de Métodos Espaciais no Estudo da Anemia Infantil 17

2.8 Aplicações de Modelos Espaciais Bayesianos em Estudos de Saúde Pública . . 19

2.9 Modelos Espaciais Bayesianos em Epidemiologia 20

2.10 Dados em Falta em Estudos Epidemiológicos 21

2.11 Lacunas da Literatura e Justificação da Abordagem Adoptada 24

2.12 Síntese da Revisão da Literatura 26

3 METODOLOGIA **28**

3.1 Material 28

3.1.1 Escolha do Modelo e Software 28

3.2 Métodos 29

3.2.1 Desenho Amostral e Estrutura dos Dados 29

3.2.2 Justificação das Escolhas Metodológicas 31

3.2.3 Especificação do Modelo Estatístico 32

3.2.4 Estimação dos Parâmetros 34

3.2.5 Avaliação e Validação dos Modelos 36

4 RESULTADOS E DISCUSSÃO 37

4.1 Resultados 37

4.1.1 Caracterização Amostral e Prevalência Observada 37

4.1.2 Factores Associados à Anemia 39

4.1.3 Modelo Espacial Bayesiano INLA–BYM2 42

4.1.4 Excedência da prevalência de anemia: análise espacial 43

4.1.5 Imputação Múltipla 48

4.1.6 Comparação entre os cenários Caso Completo e Imputação Múltipla por Equações Encadeadas 52

4.2 Discussão 54

5 CONCLUSÕES E RECOMENDAÇÕES 58

5.1 Conclusões 58

5.2 Recomendações 59

REFERÊNCIAS 62

APÊNDICES 66

LISTA DE TABELAS

1 Principais estudos sobre anemia infantil em África 14

2 Modelos lineares generalizados mistos 33

3 Efeitos fixos da função probit 40

4 Critérios de informação (AIC e BIC) para as funções de ligação avaliadas nos GLMM. 41

5 Comparação das razões de chances entre os cenários ACC e MICE, modelo INLA–BYM2. 53

A.1 Estrutura dos outputs gerados pelo script. 71

A.2 Probabilidade predita de anemia por idade (GLMM – logit) 72

A.3 Efeitos fixos do modelo INLA–BYM2 (complete case, ponderado). 73

A.4 Comparação de OR (ACC vs MICE pooled). 74

A.5 Critérios de ajuste INLA–BYM2 por imputação (MICE, 1–20). 74

LISTA DE FIGURAS

1 Características demográficas e socioeconómicas das crianças e dos agregados familiares 37

2 Distribuição espacial da prevalência de anemia em crianças de 6–59 meses, por província e meio de residência 38

3 Probabilidade ajustada de anemia segundo a idade. 40

4 Efeitos aleatórios (PSU) — Função probit. 41

5 Probabilidade de prevalência $> 40\%$ estimada pelo modelo *Integrated Nested Laplace Approximation–Besag–York–Mollié 2* 43

6 Excedência do efeito espacial estruturado $P(u_j > 0)$ estimada pelo modelo INLA–BYM2. 44

7 Efeito espacial estruturado (u_j), média posterior (escala logito) estimada pelo modelo INLA–BYM2. 45

8 Probabilidade de excedência da prevalência de anemia para limiares alternativos, análise de sensibilidade (INLA–BYM2). 46

9 Probabilidade média de prevalência $> 40\%$ (MICE — média das cinco imputações). 49

10 Probabilidade de excedência da prevalência de anemia $> 40\%$ segundo cada imputação múltipla. 50

11 Comparação visual das razões de chances entre os cenários ACC e MICE, modelo INLA–BYM2. 51

A.1 Diagnóstico do GLMM (ponderado vs não ponderado). 69

A.2 Resultados provinciais do INLA–BYM2 (ponderado). 70

A.3 Mapas de excedência provincial. 75

Lista de Abreviaturas

ACC	*Complete Case Analysis* (Análise de Casos Completos)
ADM1	Nível Administrativo 1 (Província)
BYM2	*Besag–York–Mollié 2*
CAR	*Conditional Autoregressive*
CV	Coeficiente de Variação
DHS	*Demographic and Health Survey*
DIC	*Deviance Information Criterion*
GLMM	*Generalized Linear Mixed Models*
IC95%	Intervalo de Confiança a 95%
IDS	Inquérito Demográfico e de Saúde
IID	*Independent and Identically Distributed*
INLA	*Integrated Nested Laplace Approximation*
LPML	*Log Pseudo-Marginal Likelihood*
MAR	*Missing At Random*
MICE	*Multiple Imputation by Chained Equations*
MNAR	*Missing Not At Random*
OR	*Odds Ratio*
PSU	*Primary Sampling Unit*
SE	*Standard Error*
UEM	Universidade Eduardo Mondlane
WAIC	*Watanabe–Akaike Information Criterion*

1 INTRODUÇÃO

A anemia constitui uma das condições de saúde mais prevalentes entre crianças menores de cinco anos, sendo caracterizada por níveis de hemoglobina inferiores aos valores de referência estabelecidos para a idade e estado fisiológico do indivíduo. Em crianças dos seis aos cinquenta e nove meses, considera-se anemia quando a concentração de hemoglobina é inferior a 11,0 g/dL. Esta condição compromete a capacidade de transporte de oxigénio pelo sangue, podendo afectar negativamente o desenvolvimento físico, cognitivo e imunológico das crianças, além de aumentar a susceptibilidade a infecções e outras complicações de saúde (*Black et al., 2013; World Health Organization, 2015)*.

A nível mundial, estima-se que mais de 40% das crianças menores de cinco anos sofram de anemia, sendo a África Subsariana e o Sul da Ásia as regiões mais afectadas. Nestes contextos, factores como a insegurança alimentar, a baixa diversidade nutricional, a elevada incidência de doenças infecciosas e as limitações no acesso aos serviços de saúde contribuem significativamente para a persistência do problema (Stevens et al., 2013; Kejo et al., 2018). Em Moçambique, os dados mais recentes do Inquérito Demográfico e de Saúde de 2022–2023 indicam uma prevalência de anemia de aproximadamente 63% entre crianças dos seis aos cinquenta e nove meses, valor substancialmente superior ao limiar de 40% definido pela Organização Mundial da Saúde como indicador de um problema grave de saúde pública (*World Health Organization, 2008; INE, 2023*).

A elevada prevalência da anemia infantil tem motivado numerosos estudos destinados a identificar os seus determinantes e grupos populacionais mais vulneráveis. Tradicionalmente, estas investigações têm recorrido a análises descritivas, regressões logísticas e modelos multinível para avaliar associações entre anemia e factores demográficos, socioeconómicos, nutricionais e ambientais (Kandala et al., 2009; Corsi et al., 2012; Kejo et al., 2018). Embora estas abordagens tenham contribuído significativamente para a compreensão do fenómeno, apresentam limitações quando aplicadas a dados populacionais caracterizados por estruturas hierárquicas e padrões geográficos complexos.

Em particular, a maioria dos estudos assume independência entre observações ou incorpora apenas parcialmente a estrutura de dependência existente nos dados. No entanto, fenómenos de saúde pública como a anemia infantil tendem a apresentar heterogeneidade geográfica e autocorrelação espacial, uma vez que indivíduos residentes em áreas próximas partilham condições ambientais, epidemiológicas, socioeconómicas e de acesso aos serviços de saúde semelhantes. A ignorância desta dependência espacial pode conduzir a estimativas enviesadas, subestimação da incerteza e identificação inadequada de áreas prioritárias para intervenção.

Nas últimas décadas, diferentes abordagens metodológicas têm sido desenvolvidas para lidar com estas limitações, incluindo técnicas de Sistemas de Informação Geográfica (SIG), estatística espacial clássica, modelos multinível espaciais e modelos Bayesianos hierárquicos. Os Sistemas de Informação Geográfica permitem visualizar e explorar padrões espaciais, enquanto métodos clássicos de estatística espacial, como os índices de Moran e os indicadores locais de associação espacial, possibilitam detectar autocorrelação e grupo de áreas geográficos. Contudo, estes métodos são frequentemente limitados na modelação simultânea de múltiplos factores de risco e da incerteza associada às estimativas. Segundo Banerjee et al. (2014), os modelos hierárquicos espaciais oferecem uma estrutura mais flexível para integrar dependência espacial, heterogeneidade não observada e incerteza estatística num único quadro analítico.

Neste contexto, os modelos espaciais Bayesianos têm emergido como uma abordagem particularmente flexível para a análise de fenómenos epidemiológicos espacialmente estruturados. Estes modelos permitem integrar informação proveniente de diferentes níveis hierárquicos, incorporar explicitamente a dependência espacial e quantificar de forma coerente a incerteza associada às estimativas. No entanto, a sua utilização não implica superioridade universal relativamente às abordagens frequentistas ou a outros métodos espaciais; a adequação de cada abordagem depende dos objectivos do estudo, da natureza dos dados e da complexidade do fenómeno analisado. Em situações envolvendo pequenas áreas geográficas, heterogeneidade espacial e estruturas hierárquicas complexas, a modelação Bayesiana tem demonstrado vantagens importantes em termos de estabilidade das estimativas e capacidade de incorporar diferentes fontes de variabilidade (Blangiardo Cameletti, 2015; Lawson, 2018).

Apesar dos avanços observados na literatura internacional, permanecem limitados os estudos realizados em Moçambique que integrem simultaneamente a estrutura hierárquica dos dados, a dependência espacial e o tratamento adequado dos dados omissos na análise da anemia infantil. A maioria das investigações concentra-se na identificação de factores associados à anemia, sem explorar explicitamente a variabilidade espacial do risco nem comparar diferentes estratégias metodológicas para lidar com a complexidade dos dados. Segundo Goldstein (2011) e Banerjee et al. (2014), a incorporação simultânea da estrutura hierárquica e da dependência espacial é fundamental para produzir estimativas mais precisas e inferências mais consistentes em estudos epidemiológicos baseados em dados georeferenciados.

Deste modo, existe uma lacuna metodológica na análise da anemia infantil em Moçambique, relacionada com a necessidade de modelos que permitam incorporar simultaneamente factores individuais, estrutura hierárquica, dependência espacial e incerteza associada aos dados. É neste contexto que se enquadra o presente estudo. O objectivo deste trabalho é modelar a prevalência da anemia em crianças dos seis aos cinquenta e nove meses em Moçambique, identificando os factores associados à sua ocorrência e avaliando a distribuição espacial do risco através de modelos espaciais Bayesianos hierárquicos. Pretende-se, desta forma, produzir evidência estatisticamente fundamentada, permitindo caracterizar a distribuição espacial da anemia infantil e quantificar a incerteza associada às estimativas obtidas. Desta forma, os resultados fornecem evidência empírica sobre a heterogeneidade geográfica observada, contribuindo para a identificação de áreas potencialmente prioritárias para intervenção.

1.1 Problema de pesquisa

Conforme referido na contextualização, a anemia infantil continua a ser um grave problema de saúde pública em Moçambique, afectando de forma desproporcional as crianças residentes em zonas rurais e em contextos socioeconómicos desfavorecidos. Dados do Inquérito Demográfico e de Saúde de 2022–2023 indicam uma prevalência nacional de anemia de aproximadamente 63% em crianças dos 6 aos 59 meses, sendo mais elevada nas zonas rurais do que nas urbanas, bem como nas províncias do Centro e do Norte do país (*Instituto Nacional de Estatística (INE), 2023*). Estes valores ultrapassam largamente o limiar de 40% definido pela Organização

Mundial da Saúde como indicativo de um problema grave de saúde pública (*World Health Organization, 2024)*.

Diversos estudos baseados em dados dos *Demographic and Health Surveys* (DHS) analisaram a anemia infantil em Moçambique e noutros países africanos, recorrendo sobretudo a análises descritivas e a modelos de regressão logística convencional para identificar factores associados ao risco individual, como a idade da criança, a escolaridade materna e o nível de riqueza do agregado familiar (*Kandala et al., 2009; Corsi et al., 2012; Ayele et al., 2016; WHO, 2024*). O estudo mais recente (*Muhajarine, et al. 2021*) (Não levando em conta casos omissos, distribuição espacial) Embora estes trabalhos tenham contribuído para a caracterização do problema, a maioria não considerou explicitamente a estrutura hierárquica dos dados nem a dependência espacial entre unidades geográficas.

Mesmo nos casos em que foram aplicadas técnicas de modelação estatística mais avançadas, a correlação entre diferentes níveis de agregação (criança, agregado familiar, unidade primária de amostragem e província) e a dependência espacial entre províncias raramente foram incorporadas de forma conjunta. Esta limitação metodológica pode comprometer a identificação precisa de áreas de maior risco e conduzir à subestimação da incerteza associada às estimativas (*Besag et al., 1991; Lawson, 2018; Barry et al., 2021*).

Adicionalmente, a presença de dados em falta, comum em inquéritos populacionais de grande escala, é frequentemente tratada por análise de casos completos, o que pode introduzir viés e reduzir a eficiência das estimativas quando o mecanismo de ausência não é completamente aleatório (*Rubin, 1987,van Buuren, 2018*). Apesar da relevância desta questão, poucos estudos aplicados ao contexto moçambicano exploraram de forma sistemática o impacto da imputação múltipla em combinação com modelos espaciais.

Deste modo, persiste uma lacuna metodológica no que respeita à integração da imputação de dados em falta com a modelação Bayesiana espacial, capaz de acomodar simultaneamente a estrutura hierárquica dos dados e a dependência geográfica. Neste enquadramento, o presente estudo procura responder à seguinte questão de investigação: Como se encontra distribuída a

prevalência da anemia infantil a nível nacional e quais são os factores associados à sua variação geográfica?

1.2 Objectivos

1.2.1 Objectivo Geral

Modelar a prevalência da anemia em crianças dos 6 aos 59 meses em Moçambique.

1.2.2 Objectivos Específicos

- Descrever os factores demográficos e socioeconómicos em crianças de 6 a 59 meses e dos respectivos agregados familiares em Moçambique.
- Avaliar a influência de factores demográficos e socioeconómicos na prevalência de anemia incluído as suas interacções em crianças dos 6-59 meses em Moçambique.
- Mapear a prevalência da anemia em crianças dos 6 aos 59 meses em Moçambique, com o intuito de identificar as áreas de maior risco.

1.3 Justificação

A anemia representa um grave problema de saúde pública em Moçambique, afectando um número expressivo de crianças com menos de cinco anos e comprometendo o seu desenvolvimento físico, cognitivo e imunológico (*Black et al., 2013; World Health Organization, 2015*). Embora a sua prevalência seja elevada, sobretudo em contextos marcados por pobreza, insegurança alimentar e limitações no acesso a serviços essenciais, a compreensão da distribuição territorial da anemia e dos factores associados à sua ocorrência continua limitada.

Os dados do Inquérito Demográfico e de Saúde (IDS 2022–23) disponibilizam uma oportunidade ímpar para investigar esta problemática de forma aprofundada, por meio da aplicação de técnicas estatísticas apropriadas (*Instituto Nacional de Estatística [INE], 2023*). Contudo, a estrutura hierárquica dos dados, onde múltiplas crianças pertencem ao mesmo agregado familiar e partilham características contextuais comuns, exige o uso de modelos que considerem a

dependência entre observações (*Gelman et al., 2013*).

A utilização de uma abordagem de modelação espacial bayesiana, com efeitos aleatórios ao nível dos agregados familiares e das províncias, permitirá captar a variabilidade regional e estrutural da anemia infantil. Esta estratégia não só garante maior rigor analítico, como também possibilita a produção de mapas de risco ajustado, os quais fornecem evidência geoespacial relevante para a definição de políticas públicas de saúde mais eficazes, equitativas e territorialmente orientadas (*Lawson, 2018*). Assim, o presente estudo justifica-se pela sua relevância social e científica, ao procurar responder a uma lacuna no conhecimento e apoiar a tomada de decisões baseadas em dados para a melhoria da saúde infantil em Moçambique.

1.4 Estrutura da Dissertação

A presente dissertação encontra-se organizada em cinco capítulos principais. O Capítulo 1 apresenta a contextualização do problema, a formulação da questão de investigação, os objetivos e a justificação do estudo. O Capítulo 2 corresponde à revisão da literatura, abordando os determinantes da anemia infantil, as principais abordagens estatísticas utilizadas na sua análise e os fundamentos da modelação espacial Bayesiana. O Capítulo 3 descreve os materiais e métodos, incluindo as fontes de dados, o desenho amostral, as estratégias de tratamento dos dados omissos e os modelos estatísticos aplicados. O Capítulo 4 apresenta e discute os principais resultados obtidos. Por fim, o Capítulo 5 sintetiza as conclusões do estudo, destaca as suas implicações e apresenta recomendações para futuras investigações e para a formulação de políticas públicas.

2 REVISÃO DA LITERATURA

2.1 Anemia Infantil: Conceitos, Definição e Classificação

A anemia constitui uma das doenças hematológicas e nutricionais mais prevalentes em todo o mundo, afectando particularmente crianças menores de cinco anos, mulheres grávidas e populações em situação de vulnerabilidade socioeconómica. Segundo a Organização Mundial da Saúde (OMS), a anemia é definida como uma condição em que a concentração de hemoglobina no sangue se encontra abaixo dos valores considerados normais para a idade, sexo e estado fisiológico do indivíduo (*WHO, 2011*). A hemoglobina é uma proteína presente nos glóbulos vermelhos responsável pelo transporte de oxigénio dos pulmões para os tecidos corporais. Consequentemente, níveis reduzidos de hemoglobina comprometem a capacidade de oxigenação celular, afectando o funcionamento normal do organismo.

Nas crianças com idade compreendida entre os 6 e os 59 meses, a OMS define anemia como uma concentração de hemoglobina inferior a 11 g/dL, após ajustamentos para factores como altitude e tabagismo materno (*WHO, 2011*). De acordo com a gravidade, a anemia é classificada em ligeira, quando a concentração de hemoglobina se situa entre 10,0 e 10,9 g/dL; moderada, quando varia entre 7,0 e 9,9 g/dL; e grave, quando é inferior a 7,0 g/dL (*WHO, 2011*).

Para além da classificação individual, a OMS estabelece critérios para avaliar a magnitude da anemia enquanto problema de saúde pública. Assim, prevalências inferiores a 5% não são consideradas um problema de saúde pública; prevalências entre 5% e 19,9% correspondem a um problema ligeiro; entre 20% e 39,9% a um problema moderado; e valores iguais ou superiores a 40% são classificados como um problema grave de saúde pública (*WHO, 2024*). Esta classificação permite comparar a magnitude do problema entre diferentes países e regiões e orientar a definição de prioridades em saúde pública.

A etiologia da anemia é multifactorial. Embora a deficiência de ferro seja reconhecida como a principal causa a nível mundial, outros factores contribuem significativamente para a sua ocorrência, incluindo deficiências de micronutrientes como vitamina A, vitamina B_{12} e ácido fólico, infecções parasitárias, malária, doenças inflamatórias crónicas e doenças hereditárias,

como a anemia falciforme (*Kassebaum, 2016; Petry et al., 2016*). Em países de baixo e médio rendimento, estas causas coexistem frequentemente, resultando numa elevada complexidade etiológica.

As consequências da anemia infantil são amplamente documentadas na literatura. A curto prazo, a doença associa-se a fadiga, diminuição da capacidade física, atraso do crescimento e maior susceptibilidade a infecções. A longo prazo, pode comprometer o desenvolvimento cognitivo, o desempenho escolar e a produtividade económica na idade adulta (*Grantham-McGregor & Ani, 2001; Black et al., 2013*). Estas consequências são particularmente preocupantes quando a anemia ocorre durante os primeiros anos de vida, período considerado crítico para o desenvolvimento físico e neurológico da criança.

Em Moçambique, a anemia infantil continua a representar um dos mais importantes desafios de saúde pública. Os resultados do Inquérito Demográfico e de Saúde de 2022–2023 indicam que cerca de 63% das crianças entre os 6 e os 59 meses apresentam anemia (*INE & ICF, 2023*), valor muito superior ao limiar estabelecido pela OMS para classificação de um problema grave de saúde pública. Este cenário evidencia a necessidade de compreender não apenas a magnitude da doença, mas também os factores que contribuem para a sua persistência e distribuição geográfica.

2.2 Epidemiologia da Anemia Infantil

A anemia infantil permanece uma das condições de saúde mais disseminadas globalmente. Estimativas recentes da Organização Mundial da Saúde indicam que aproximadamente 40% das crianças menores de cinco anos apresentam anemia, correspondendo a centenas de milhões de casos em todo o mundo (*WHO, 2024*). Apesar dos avanços observados em algumas regiões, a redução da prevalência tem ocorrido de forma lenta e desigual, reflectindo persistentes desafios relacionados com a pobreza, insegurança alimentar, doenças infecciosas e limitações no acesso aos serviços de saúde.

A distribuição da anemia apresenta marcadas desigualdades geográficas. As prevalências mais elevadas concentram-se na África Subsaariana e no Sul da Ásia, regiões onde coexistem ele-

vadas taxas de desnutrição, insegurança alimentar, malária e outras doenças infecciosas. Em muitos países africanos, mais de metade das crianças menores de cinco anos sofre de anemia, configurando um problema grave de saúde pública segundo os critérios da OMS (*Stevens et al., 2013; WHO, 2024*).

Na África Subsaariana, a elevada carga da doença resulta da combinação de múltiplos factores biológicos, ambientais e socioeconómicos. A deficiência de ferro continua a desempenhar um papel importante, mas a elevada incidência de malária, helmintíases intestinais, infecções recorrentes e insegurança alimentar contribuem igualmente para a persistência da doença (*Menéndez et al., 2000; Brooker et al., 2008*). Estudos realizados na Etiópia, Nigéria, Tanzânia, Quénia e Malawi demonstram que a prevalência da anemia varia substancialmente entre regiões dentro do mesmo país, sugerindo a influência de determinantes contextuais e geográficos específicos (*Kejo et al., 2018; Nambiema et al., 2019*).

Em Moçambique, a anemia infantil continua a representar um dos principais problemas de saúde pública. Dados do Inquérito Demográfico e de Saúde de 2022–2023 revelam que aproximadamente 63% das crianças entre os 6 e os 59 meses apresentam anemia (*INE & ICF, 2023*). Embora este valor represente uma redução relativamente a alguns inquéritos anteriores, continua muito acima do limiar de 40% definido pela OMS para classificação de um problema grave de saúde pública.

Para além da elevada prevalência nacional, a distribuição da anemia em Moçambique caracteriza-se por importantes desigualdades regionais. As províncias das regiões Centro e Norte tendem a apresentar prevalências superiores às observadas no Sul do país. Estas diferenças refletem desigualdades nos níveis de pobreza, acesso aos serviços de saúde, prevalência da malária, cobertura de intervenções nutricionais e condições de saneamento (*Custódio et al., 2015; Augusto et al., 2020*). A persistência destas disparidades sugere que intervenções uniformes a nível nacional podem ser insuficientes para responder adequadamente às necessidades específicas de cada contexto geográfico.

A literatura recente tem igualmente demonstrado que a distribuição da anemia não ocorre de

forma aleatória no território. Estudos espaciais realizados em vários países africanos identificaram *grupo de áreas* geográficos de elevada prevalência, indicando a presença de dependência espacial e a influência de factores contextuais partilhados entre áreas vizinhas (*Ayele et al., 2016; Barry et al., 2021*). Estes resultados reforçam a importância da incorporação da dimensão espacial na análise epidemiológica da anemia infantil.

Em síntese, a epidemiologia da anemia infantil caracteriza-se por uma elevada magnitude, forte heterogeneidade geográfica e persistência ao longo do tempo. Esta realidade justifica a necessidade de abordagens analíticas capazes de integrar factores individuais, comunitários e espaciais, contribuindo para uma compreensão mais abrangente da distribuição da doença.

2.3 Determinantes da Anemia Infantil

A anemia infantil é amplamente reconhecida como uma condição multifactorial, resultante da interação entre factores biológicos, nutricionais, demográficos, socioeconómicos, ambientais e contextuais. Esta natureza complexa explica a dificuldade em reduzir a prevalência da doença através de intervenções isoladas e justifica a necessidade de estratégias integradas de prevenção e controlo.

Entre os factores biológicos, a idade da criança constitui um dos determinantes mais consistentemente identificados na literatura. Diversos estudos demonstram que crianças mais jovens, particularmente entre os 6 e os 23 meses, apresentam maior risco de anemia devido às elevadas necessidades nutricionais associadas ao rápido crescimento durante os primeiros anos de vida (*Osório et al., 2001; Nambiema et al., 2019*). Nesta fase, a introdução inadequada da alimentação complementar e a baixa ingestão de alimentos ricos em ferro contribuem significativamente para o desenvolvimento da doença.

O sexo da criança tem sido igualmente investigado como potencial determinante da anemia. Contudo, os resultados permanecem inconsistentes. Alguns estudos identificaram maior prevalência entre crianças do sexo masculino, sugerindo diferenças fisiológicas associadas ao crescimento e metabolismo do ferro (*Black et al., 2013; Woldie et al., 2015*), enquanto outros não encontraram diferenças estatisticamente significativas (*Cusick & Georgieff, 2012*). Esta varia-

bilidade sugere que a influência do sexo pode depender do contexto epidemiológico e sociocultural.

No domínio nutricional, a deficiência de ferro continua a ser considerada a principal causa de anemia a nível mundial. Contudo, outras deficiências nutricionais, incluindo vitamina A, ácido fólico, vitamina B_{12} e zinco, desempenham igualmente um papel importante na etiologia da doença (*Rohner et al., 2013*). A diversidade alimentar inadequada, particularmente em contextos de insegurança alimentar, contribui para a persistência destas deficiências.

Os factores socioeconómicos constituem igualmente determinantes fundamentais da anemia infantil. A escolaridade materna emerge consistentemente como um dos factores protectores mais importantes. Mães com níveis mais elevados de educação tendem a possuir maior conhecimento sobre práticas alimentares adequadas, prevenção de doenças e utilização dos serviços de saúde, contribuindo para melhores condições de saúde das crianças (*Kandala et al., 2009*). De forma semelhante, o nível de riqueza do agregado familiar encontra-se fortemente associado à ocorrência da anemia, uma vez que famílias mais pobres enfrentam maiores dificuldades no acesso a alimentos nutritivos, água potável, saneamento e cuidados de saúde.

A literatura também destaca a importância dos factores ambientais e de saúde pública. Em países endémicos, a malária é considerada um dos principais determinantes da anemia infantil, devido à destruição dos glóbulos vermelhos causada pelo parasita (*Plasmodium spp.) (Menéndez et al., 2000; Stevenson et al., 2012*). Adicionalmente, infecções parasitárias intestinais, episódios frequentes de diarreia e outras doenças infecciosas contribuem para agravar a condição nutricional e aumentar o risco de anemia.

No contexto moçambicano, diversos estudos apontam que a elevada prevalência da anemia resulta da combinação de factores nutricionais, socioeconómicos e epidemiológicos. A elevada carga de malária, a insegurança alimentar, a pobreza persistente e as limitações no acesso aos serviços de saúde são frequentemente identificadas como factores determinantes da doença (*Custódio et al., 2015; Augusto et al., 2020*). As regiões Centro e Norte do país apresentam condições particularmente desfavoráveis, caracterizadas por maiores níveis de pobreza e menor

acesso a infraestruturas básicas, contribuindo para a manutenção de elevadas prevalências de anemia.

Apesar do consenso relativamente à importância destes factores, a literatura apresenta resultados por vezes contraditórios quanto à magnitude e direcção de algumas associações. Enquanto alguns estudos identificam a malária como principal determinante da anemia infantil, outros atribuem maior relevância às condições nutricionais e socioeconómicas (*Stevens et al., 2013; Kejo et al., 2018*). Estas diferenças podem resultar de particularidades epidemiológicas, metodológicas e contextuais, reforçando a necessidade de análises que considerem simultaneamente múltiplos níveis de influência.

Em síntese, a evidência científica sugere que a anemia infantil não pode ser explicada por um único factor, mas sim por um conjunto complexo de determinantes que actuam de forma interdependente. Esta realidade reforça a necessidade de abordagens analíticas capazes de integrar factores individuais, familiares, comunitários e territoriais, contribuindo para uma compreensão mais abrangente da distribuição e dos determinantes da anemia infantil.

2.4 Evidência Empírica sobre a Anemia Infantil em África e Moçambique

A anemia infantil tem sido amplamente estudada em países de baixo e médio rendimento devido à sua elevada prevalência e às consequências negativas para o crescimento físico, desenvolvimento cognitivo e produtividade futura. A literatura evidencia que a doença permanece um importante problema de saúde pública, particularmente na África Subsaariana, onde coexistem elevados níveis de pobreza, insegurança alimentar, doenças infecciosas e acesso limitado aos serviços de saúde.

Diversos estudos realizados em África identificaram consistentemente a idade da criança, o sexo, o nível de riqueza do agregado familiar, a escolaridade materna e a exposição a doenças infecciosas como factores associados à anemia infantil. Contudo, a magnitude destas associações varia substancialmente entre países e regiões, sugerindo a influência de factores contextuais específicos.

Kandala et al. (2009), utilizando dados provenientes de 18 países africanos, aplicaram modelos multinível para investigar os determinantes da anemia infantil. Os autores concluíram que tanto factores individuais como factores comunitários contribuíam significativamente para a ocorrência da doença. A pobreza comunitária e o baixo nível educacional das mães foram identificados como importantes factores de risco.

Na Tanzânia, Kejo et al. (2018) analisaram dados de crianças menores de cinco anos e observaram que a idade da criança, a escolaridade materna, o nível de riqueza e a localização geográfica estavam fortemente associados à anemia. Os resultados demonstraram ainda que uma proporção importante da variabilidade observada ocorria ao nível comunitário.

Nambiema et al. (2019), no Togo, reportaram resultados semelhantes, identificando maior risco de anemia entre crianças residentes em agregados familiares mais pobres e cujas mães possuíam menor nível de escolaridade. Os autores destacaram igualmente a importância da incorporação de factores contextuais nos modelos estatísticos.

No contexto da Etiópia, vários estudos têm evidenciado uma elevada prevalência de anemia infantil e importantes desigualdades geográficas. Ayele et al. (2016), identificaram áreas espaciais de elevada prevalência e demonstraram que a distribuição da doença não ocorria aleatoriamente no território. Estes resultados sugerem que factores ambientais, epidemiológicos e socioeconómicos desempenham um papel importante para além das características individuais das crianças.

Em Moçambique, a evidência científica permanece relativamente limitada quando comparada com outros países africanos. Estudos baseados em dados dos Inquéritos Demográficos e de Saúde têm documentado prevalências persistentemente elevadas de anemia infantil e identificado associações com idade da criança, condições socioeconómicas, escolaridade materna e localização geográfica. Contudo, a maioria destes trabalhos recorreu a análises descritivas ou modelos de regressão convencionais, não explorando adequadamente a estrutura hierárquica dos dados nem a dependência espacial entre áreas geográficas.

Os resultados do Inquérito Demográfico e de Saúde 2022–2023 mostram que aproximadamente 63% das crianças moçambicanas entre os 6 e os 59 meses apresentam anemia, confirmando que a doença continua a representar um problema grave de saúde pública. Para além da elevada prevalência nacional, observam-se importantes diferenças entre províncias, sugerindo a existência de heterogeneidade espacial na distribuição do risco.

Em síntese, a literatura empírica demonstra que a anemia infantil constitui um fenómeno multifactorial influenciado por factores individuais, familiares, comunitários e territoriais. Contudo, persistem importantes lacunas na compreensão da forma como estes factores interagem espacialmente, particularmente no contexto moçambicano. Esta limitação justifica a utilização de abordagens metodológicas capazes de integrar simultaneamente factores individuais, heterogeneidade contextual e dependência espacial.

Tabela 1: Principais estudos sobre anemia infantil em África

Autor	País	Método	Principais resultados
Kandala et al. (2009)	18 países africanos	Modelo multinível	Influência de factores individuais e comunitários.
Kejo et al. (2018)	Tanzânia	Regressão multinível	Idade, riqueza e escolaridade materna associadas à anemia.
Nambiema et al. (2019)	Togo	Modelo multinível	Efeito significativo do contexto comunitário.
Ayele et al. (2016)	Etiópia	Análise espacial	Identificação de áreas espaciais de alto risco.
Barry et al. (2021)	Guiné	Modelos espaciais Bayesianos	Forte heterogeneidade espacial da anemia infantil.

2.5 Evolução das Abordagens Metodológicas no Estudo da Anemia Infantil

A evolução das abordagens metodológicas utilizadas no estudo da anemia infantil acompanhou o desenvolvimento da epidemiologia quantitativa, a crescente disponibilidade de inquéritos populacionais de grande escala e os avanços da modelação estatística aplicada à saúde pública. Os primeiros estudos concentravam-se essencialmente na descrição da magnitude da doença através de análises de prevalência, tabelas de contingência e medidas simples de associação. Embora estas abordagens tenham sido fundamentais para documentar a elevada carga da anemia infantil em diferentes contextos, apresentavam limitações na identificação simultânea dos múltiplos factores associados à doença.

A introdução dos Modelos Lineares Generalizados (GLM), formalizados por Nelder e Wedderburn (1972), constituiu um marco importante na análise epidemiológica. A regressão logística passou a ser amplamente utilizada para modelar desfechos binários, permitindo estimar a associação entre a anemia e factores demográficos, nutricionais e socioeconómicos. Diversos estudos baseados em dados DHS recorreram a esta abordagem para identificar factores associados à anemia infantil em África, Ásia e América Latina (Stevens et al., 2013; Harding et al., 2018).

A formulação geral da regressão logística pode ser expressa por:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{h=1}^{p} \beta_h X_{ih}, \tag{1}$$

onde π_i representa a probabilidade de ocorrência de anemia para a criança i, β_0 corresponde ao intercepto e β_h aos coeficientes associados às covariáveis X_{ih}.

Apesar da sua ampla utilização, os GLM assumem independência entre observações, uma hipótese frequentemente violada em dados provenientes de inquéritos populacionais. As crianças encontram-se agrupadas em agregados familiares, comunidades e regiões administrativas, gerando correlação entre observações pertencentes ao mesmo contexto. Esta limitação conduziu ao desenvolvimento dos Modelos Lineares Generalizados Mistos (GLMM), que incorporam efeitos aleatórios para acomodar a estrutura hierárquica dos dados

Mais recentemente, a crescente evidência de desigualdades geográficas na distribuição da anemia impulsionou a incorporação explícita da dimensão espacial nos modelos epidemiológicos. Esta evolução culminou na utilização de modelos espaciais Bayesianos, capazes de integrar simultaneamente factores individuais, heterogeneidade contextual e dependência espacial.

Em síntese, a literatura metodológica evidencia uma progressão contínua das abordagens utilizadas no estudo da anemia infantil, passando de análises descritivas para modelos multinível e, posteriormente, para modelos espaciais Bayesianos, considerados actualmente o estado da arte na análise de fenómenos epidemiológicos complexos.

2.6 Aplicação de Modelos Multinível no Estudo da Anemia Infantil

Os modelos multinível surgiram como resposta à necessidade de analisar dados hierarquicamente estruturados, situação particularmente comum em estudos baseados nos Inquéritos Demográficos e de Saúde (DHS). Nestes inquéritos, as crianças encontram-se agrupadas em agregados familiares, comunidades e regiões administrativas, originando correlação entre observações pertencentes ao mesmo contexto.

A principal vantagem dos modelos multinível consiste na capacidade de distinguir os efeitos atribuíveis às características individuais daqueles associados ao contexto onde os indivíduos vivem. Esta característica é particularmente relevante no estudo da anemia infantil, cuja ocorrência resulta da interação entre factores biológicos, socioeconómicos, ambientais e comunitários.

Kandala et al. (2009), numa análise de dados provenientes de 18 países africanos, demonstraram que uma proporção substancial da variabilidade observada na anemia infantil era explicada por factores comunitários e regionais. Os autores verificaram que crianças residentes em comunidades mais pobres apresentavam maior risco de anemia, mesmo após o ajustamento para características individuais e familiares.

Na Tanzânia, Kejo et al. (2018) aplicaram modelos multinível para investigar os determinantes da anemia em crianças menores de cinco anos e concluíram que a heterogeneidade entre

comunidades permanecia significativa após o controlo das covariáveis individuais. Resultados semelhantes foram observados por Nambiema et al. (2019) no Togo e por diversos estudos realizados na Etiópia e Nigéria.

Do ponto de vista estatístico, um modelo logístico multinível pode ser representado por:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \sum_{h=1}^{p} \beta_h X_{ijh} + u_j, \tag{2}$$

onde u_j representa o efeito aleatório associado ao grupo ou comunidade j, assumindo geralmente:

$$u_j \sim \mathcal{N}(0, \sigma_u^2). \tag{3}$$

Esta formulação permite modelar explicitamente a correlação intra-áreas, produzindo estimativas dos efeitos fixos que consideram adequadamente a estrutura hierárquica dos dados.

No contexto moçambicano, estudos recentes sugerem igualmente a existência de heterogeneidade significativa entre comunidades e províncias na prevalência da anemia infantil. Estes resultados reforçam a relevância da utilização de modelos multinível para compreender a influência simultânea de factores individuais e contextuais. Apesar das suas vantagens, os modelos multinível não incorporam explicitamente a dependência espacial existente entre áreas geográficas vizinhas. Assim, duas províncias adjacentes com perfis epidemiológicos semelhantes continuam a ser tratadas como independentes, motivando o desenvolvimento de abordagens espaciais mais sofisticadas.

2.7 Aplicação de Métodos Espaciais no Estudo da Anemia Infantil

A distribuição da anemia infantil não ocorre de forma aleatória no espaço geográfico. Diversos estudos demonstram que áreas próximas tendem a apresentar níveis semelhantes de prevalência devido à partilha de condições ambientais, epidemiológicas, culturais e socioeconómicas. Este fenómeno é conhecido como autocorrelação espacial.

As primeiras aplicações da análise espacial à anemia infantil recorreram a técnicas explo-

ratórias, incluindo mapas temáticos, estatísticas globais de Moran e indicadores locais de associação espacial (LISA). Estas abordagens permitiram identificar padrões geográficos e áreas de concentração de risco, contribuindo para a compreensão das desigualdades territoriais.

Na Etiópia, Ayele et al. (2016) identificaram agrupamentos espaciais de elevada prevalência de anemia infantil, demonstrando que determinadas regiões apresentavam risco significativamente superior ao esperado. Resultados semelhantes foram observados na Guiné por Barry et al. (2021), no Malawi por Ngwira et al. (2015) e em vários outros países africanos.

Em Moçambique, estudos recentes têm evidenciado importantes desigualdades espaciais na distribuição da anemia infantil. As regiões Centro e Norte apresentam sistematicamente prevalências superiores às observadas na região Sul, sugerindo a influência de factores contextuais relacionados com pobreza, insegurança alimentar, malária e acesso aos serviços de saúde.

Do ponto de vista estatístico, a autocorrelação espacial é frequentemente avaliada através do índice de Moran:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}, \tag{4}$$

onde w_{ij} representa a matriz de vizinhança espacial e S_0 corresponde à soma de todos os pesos espaciais.

Valores positivos de I indicam agrupamento espacial de áreas com valores semelhantes, enquanto valores negativos sugerem dispersão espacial.

Embora estas técnicas permitam identificar padrões geográficos, não modelam explicitamente a estrutura espacial no processo de estimação. Esta limitação impulsionou o desenvolvimento dos modelos espaciais Bayesianos, que incorporam directamente a dependência espacial entre áreas vizinhas e permitem estimar a distribuição geográfica do risco considerando a autocorrelação espacial presente nos dados.

2.8 Aplicações de Modelos Espaciais Bayesianos em Estudos de Saúde Pública

Os modelos espaciais Bayesianos têm sido amplamente utilizados em epidemiologia para analisar fenómenos cuja distribuição apresenta dependência espacial. A principal vantagem destes modelos reside na capacidade de incorporar simultaneamente factores explicativos, heterogeneidade não observada e autocorrelação espacial entre áreas geográficas.

Inicialmente, estes modelos foram aplicados ao mapeamento de doenças infecciosas e crónicas em pequenas áreas geográficas. Lawson (2018) refere que os modelos Bayesianos hierárquicos se tornaram uma das abordagens mais utilizadas na epidemiologia espacial devido à sua capacidade de produzir estimativas estáveis em áreas com pequenas populações e reduzido número de observações. Na área da saúde materno-infantil, vários estudos recorreram aos modelos BYM e BYM2 para identificar padrões geográficos de mortalidade infantil, desnutrição e anemia. Na Etiópia, Ayele et al. (2016), identificaram (agrupamentos espaciais de elevada prevalência de anemia infantil utilizando métodos espaciais Bayesianos. Resultados semelhantes foram reportados por Barry et al. (2021) na Guiné, onde a modelação espacial permitiu identificar províncias prioritárias para intervenção.

Em estudos de malária, mortalidade neonatal e desnutrição infantil, os modelos Bayesianos têm igualmente demonstrado vantagens relativamente às abordagens clássicas, sobretudo na estabilização das estimativas em áreas com pequenas amostras e na quantificação da incerteza associada aos riscos estimados.

Mais recentemente, a reparametrização proposta por Riebler et al. (2016), conhecida como modelo BYM2, passou a ser recomendada devido à melhor interpretabilidade dos parâmetros e à redução de problemas de identificabilidade observados no modelo BYM original.

Apesar da ampla utilização internacional destes modelos, a sua aplicação à anemia infantil em Moçambique permanece extremamente limitada. A maioria dos estudos nacionais continua baseada em análises descritivas, regressões logísticas ou modelos multinível sem componente

espacial explícita. Esta lacuna metodológica justifica a adopção do modelo BYM2 na presente investigação.

2.9 Modelos Espaciais Bayesianos em Epidemiologia

Os modelos espaciais Bayesianos representam actualmente uma das abordagens mais avançadas para a análise de fenómenos epidemiológicos caracterizados por heterogeneidade geográfica e dependência espacial. A sua popularidade resulta da capacidade de integrar informação proveniente de diferentes fontes, acomodar estruturas hierárquicas complexas e quantificar explicitamente a incerteza associada às estimativas.

O desenvolvimento destes modelos está fortemente associado ao trabalho de Besag, York e Mollié (1991), que propuseram o modelo BYM para análise de dados espaciais agregados. Este modelo tornou-se uma referência na epidemiologia espacial ao permitir decompor a variabilidade residual em duas componentes distintas: uma componente espacialmente estruturada e uma componente não estruturada.

A formulação geral do modelo BYM pode ser representada por:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{h=1}^{p} \beta_h X_{ih} + u_i + v_i, \tag{5}$$

onde:

- u_i representa o efeito espacial estruturado;
- v_i representa o efeito espacial não estruturado.

A componente estruturada é geralmente modelada através de um processo Condicionalmente Autorregressivo (CAR):

$$u_i \mid u_{-i} \sim \mathcal{N}\left(\frac{1}{n_i}\sum_{j\in\partial i} u_j, \frac{1}{n_i\tau}\right), \tag{6}$$

onde ∂i representa o conjunto de áreas vizinhas da região i.

Apesar da ampla utilização do modelo BYM, problemas de identificabilidade e interpretação motivaram o desenvolvimento da reformulação proposta por Riebler et al. (2016), conhecida como modelo BYM2.

A parametrização BYM2 é dada por:

$$b_i = \frac{1}{\sqrt{\tau}} \left(\sqrt{\phi}\, u_i^* + \sqrt{1-\phi}\, v_i \right), \tag{7}$$

onde:

- τ controla a variabilidade global;
- ϕ representa a proporção da variabilidade atribuível à componente espacial estruturada;
- u_i^* corresponde ao efeito espacial estruturado padronizado;
- v_i representa a componente não estruturada.

Esta reformulação melhora substancialmente a interpretabilidade dos parâmetros e reduz problemas de identificabilidade presentes no modelo original.

A inferência Bayesiana para estes modelos foi inicialmente realizada através de métodos de Monte Carlo via Cadeias de Markov (MCMC). Contudo, o desenvolvimento da Aproximação Integrada Aninhada de Laplace (INLA) por Rue et al. (2009) permitiu obter estimativas precisas com custos computacionais significativamente inferiores. Actualmente, os modelos BYM2 estimados através do INLA são amplamente considerados o estado da arte na modelação espacial epidemiológica e têm sido aplicados com sucesso em estudos sobre malária, tuberculose, mortalidade infantil, desnutrição e anemia infantil em diversos países africanos.

2.10 Dados em Falta em Estudos Epidemiológicos

A presença de dados em falta constitui um dos desafios metodológicos mais frequentes em estudos epidemiológicos e em inquéritos populacionais de grande escala. A ausência de informação pode resultar de recusas de resposta, erros de recolha, limitações operacionais durante o trabalho

de campo ou dificuldades associadas à medição de determinadas variáveis. Quando não adequadamente tratada, a existência de valores omissos pode reduzir a eficiência estatística, comprometer a validade das inferências e introduzir enviesamentos nas estimativas dos parâmetros de interesse (Little e Rubin, 2002; van Buuren, 2018).

Nos Inquéritos Demográficos e de Saúde (DHS), os dados em falta constituem uma realidade comum devido à complexidade dos procedimentos de recolha de informação. Variáveis relacionadas com características socioeconómicas, indicadores biométricos e resultados laboratoriais podem apresentar níveis variáveis de não resposta. No caso específico da anemia infantil, a ausência de medições de hemoglobina ou de informações sobre covariáveis relevantes pode reduzir substancialmente o tamanho efectivo da amostra e afectar a representatividade dos resultados.

Rubin (1976; 1987) propôs uma classificação amplamente aceite dos mecanismos de ausência de dados, distinguindo três situações principais: *Missing Completely at Random* (MCAR), *Missing at Random* (MAR) e *Missing Not at Random* (MNAR). A compreensão destes mecanismos é fundamental para a selecção da estratégia de tratamento mais adequada.

No mecanismo MCAR, a probabilidade de ausência de um valor é completamente independente dos dados observados e não observados:

$$P(R \mid Y_{obs}, Y_{mis}) = P(R), \tag{8}$$

onde R representa o indicador de ausência de dados. Neste cenário, a análise por casos completos produz estimativas não enviesadas, embora com perda de eficiência devido à redução do tamanho amostral.

No mecanismo MAR, a probabilidade de ausência depende apenas dos dados observados:

$$P(R \mid Y_{obs}, Y_{mis}) = P(R \mid Y_{obs}), \tag{9}$$

constituindo o pressuposto adoptado pela maioria dos métodos modernos de imputação múltipla.

Por sua vez, no mecanismo MNAR, a probabilidade de ausência depende directamente dos valores não observados:

$$P(R \mid Y_{obs}, Y_{mis}) \neq P(R \mid Y_{obs}), \tag{10}$$

o que torna o tratamento dos dados em falta substancialmente mais complexo.
Historicamente, muitos estudos epidemiológicos recorreram à análise por casos completos (*Complete Case Analysis*), utilizando apenas observações sem valores omissos. Embora simples de implementar, esta abordagem pode conduzir a perdas substanciais de informação e produzir estimativas enviesadas quando o mecanismo de ausência não corresponde ao cenário MCAR (Little & Rubin, 2002).

Como alternativa, a imputação múltipla tornou-se uma das estratégias mais recomendadas para lidar com dados em falta. Introduzida por Rubin (1987), esta abordagem consiste na geração de múltiplas versões plausíveis dos valores ausentes, produzindo várias bases de dados completas que são posteriormente analisadas e combinadas segundo regras específicas de inferência.

Entre os métodos disponíveis, a Imputação Múltipla por Equações Encadeadas (*Multiple Imputation by Chained Equations* (MICE) destaca-se pela sua flexibilidade e capacidade de lidar simultaneamente com variáveis contínuas, binárias, ordinais e nominais. O método especifica um modelo condicional para cada variável com valores ausentes:

$$X_j \mid X_{-j} \sim f_j(X_j \mid X_{-j}; \theta_j), \tag{11}$$

onde X_j representa a variável a imputar e X_{-j} o conjunto das restantes variáveis observadas.

Após a geração de m bases imputadas, os parâmetros estimados são combinados segundo as regras de Rubin. Seja $Q^{(r)}$ a estimativa do parâmetro de interesse obtida na r-ésima base imputada, então a estimativa combinada é dada por:

$$\bar{Q} = \frac{1}{m} \sum_{r=1}^{m} Q^{(r)}. \tag{12}$$

A variância total incorpora simultaneamente a variabilidade intra e entre imputações:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B, \tag{13}$$

onde $\bar{U}$ representa a variância média intra-imputação e B a variância entre imputações.

A utilização da imputação múltipla tem aumentado significativamente na investigação epidemiológica. Estudos realizados na Etiópia, Nigéria, Bangladesh e Índia demonstraram que a imputação múltipla produz estimativas mais eficientes e menos susceptíveis a enviesamento quando comparada com a análise por casos completos, sobretudo quando a proporção de dados ausentes é moderada ou elevada (White et al., 2011; van Buuren, 2018).

No contexto da anemia infantil, a combinação da imputação múltipla com modelos multinível e espaciais permite preservar o tamanho amostral, reduzir perdas de informação e melhorar a robustez das inferências. Esta abordagem torna-se particularmente relevante quando se pretende estimar efeitos espaciais e identificar áreas prioritárias para intervenção, uma vez que a exclusão de observações pode afectar a estabilidade das estimativas geográficas.

Em síntese, a literatura metodológica reconhece que o tratamento adequado dos dados em falta constitui uma etapa essencial para garantir a validade dos estudos epidemiológicos. Entre as estratégias actualmente disponíveis, a imputação múltipla por equações encadeadas destaca-se pela sua flexibilidade e capacidade de lidar com diferentes tipos de variáveis e padrões de ausência de dados, sendo amplamente recomendada em análises baseadas em dados DHS e em estudos observacionais na área da saúde.

2.11 Lacunas da Literatura e Justificação da Abordagem Adoptada

Apesar dos avanços registados na investigação sobre anemia infantil nas últimas décadas, persistem importantes lacunas tanto ao nível substantivo como metodológico. A maioria dos estudos realizados em países de baixo e médio rendimento concentrou-se na identificação dos factores associados à doença através de análises descritivas ou modelos de regressão convencionais. Embora estes trabalhos tenham contribuído significativamente para a compreensão dos determi-

nantes da anemia, apresentam limitações na capacidade de captar simultaneamente a estrutura hierárquica dos dados e a dependência espacial existente entre áreas geográficas.

No contexto africano, diversos estudos demonstraram que a distribuição da anemia infantil apresenta forte heterogeneidade geográfica, com importantes diferenças entre regiões, províncias e comunidades. Contudo, uma proporção considerável da literatura continua a recorrer a modelos estatísticos que assumem independência entre observações, ignorando a autocorrelação espacial frequentemente presente nos dados epidemiológicos (Kandala et al., 2009; Kejo et al., 2018).

Em Moçambique, apesar da elevada prevalência da anemia infantil e da sua relevância para a saúde pública, os estudos disponíveis permanecem relativamente limitados. Grande parte das investigações concentrou-se na descrição da magnitude do problema ou na análise de factores associados utilizando modelos convencionais. Consequentemente, permanece escassa a evidência sobre a distribuição espacial da anemia e sobre a existência de heterogeneidade residual após o ajustamento para factores demográficos e socioeconómicos.

Do ponto de vista metodológico, observa-se igualmente uma reduzida utilização de modelos espaciais Bayesianos aplicados à anemia infantil em Moçambique. Embora os modelos multinível tenham sido progressivamente incorporados em estudos epidemiológicos, poucos trabalhos combinam simultaneamente a estrutura hierárquica dos dados, a dependência espacial e o tratamento rigoroso dos dados em falta.

Outra limitação frequentemente observada na literatura consiste na utilização exclusiva da análise por casos completos, sem avaliação do impacto potencial dos valores omissos nas estimativas obtidas. Esta prática pode conduzir à perda substancial de informação e comprometer a robustez das conclusões, particularmente em estudos baseados em inquéritos populacionais complexos.

Face a estas limitações, a presente dissertação adopta uma abordagem metodológica integrada que combina Modelos Lineares Generalizados Mistos (GLMM), Imputação Múltipla por Equações Encadeadas (MICE) e Modelos Espaciais Bayesianos do tipo BYM2, cujos parâmetros são estimados através da Aproximação Integrada Aninhada de Laplace (INLA)

A utilização dos GLMM permite acomodar a estrutura hierárquica dos dados provenientes do IDS 2022–2023, incorporando a correlação existente entre observações pertencentes aos mesmos agrupamentos amostrais. A imputação múltipla possibilita reduzir a perda de informação associada aos dados em falta e incorporar a incerteza decorrente dos valores omissos no processo de inferência. Por sua vez, os modelos BYM2 permitem incorporar explicitamente a dependência espacial entre províncias vizinhas, produzindo estimativas da distribuição geográfica do risco que consideram a autocorrelação espacial presente nos dados.

Ao integrar estas abordagens num único quadro analítico, o presente estudo procura preencher importantes lacunas da literatura nacional, contribuindo para uma compreensão mais abrangente dos determinantes e da distribuição espacial da anemia infantil em Moçambique. Adicionalmente, a utilização de métodos espaciais Bayesianos modernos fornece evidência científica relevante para apoiar a formulação de políticas públicas e a focalização territorial das intervenções de saúde e nutrição.

Em síntese, a principal contribuição desta investigação reside na combinação de modelação multinível, análise espacial Bayesiana e tratamento rigoroso dos dados em falta, permitindo analisar simultaneamente factores individuais, contextuais e espaciais associados à anemia infantil num contexto caracterizado por fortes desigualdades geográficas e socioeconómicas.

2.12 Síntese da Revisão da Literatura

A revisão da literatura demonstrou que a anemia infantil continua a constituir um importante problema de saúde pública, particularmente nos países da África Subsaariana. A evidência empírica mostra que a sua ocorrência resulta da interação de factores individuais, familiares, comunitários e contextuais, conforme documentado por *Balarajan et al. (2011 & Stevens et al. (2022), Black et al. (2013)*).

Do ponto de vista metodológico, observou-se uma evolução progressiva das abordagens utilizadas para estudar a anemia infantil. Os modelos de regressão convencionais permitiram identificar determinantes individuais; os modelos multinível incorporaram a estrutura hierárquica dos

dados, conforme descrito por *Goldstein (2011*); e os modelos espaciais Bayesianos passaram a permitir a modelação explícita da dependência espacial e da heterogeneidade geográfica, conforme demonstrado por *Banerjee et al. (2014) e Moraga (2019*).

Apesar destes avanços, permanecem limitados os estudos realizados em Moçambique que integrem simultaneamente modelação multinível, análise espacial Bayesiana e tratamento rigoroso dos dados em falta. Esta lacuna fundamenta a abordagem metodológica adoptada nesta dissertação, baseada na combinação de GLMM, imputação múltipla por equações encadeadas (MICE), proposta por Van Buuren (2018), e modelos espaciais Bayesianos BYM2 estimados através de INLA, conforme descrito por *Rue et al. (2009 & Riebler et al. (2016)*).

3 METODOLOGIA

3.1 Material

O presente estudo utilizou dados secundários do Inquérito Demográfico e de Saúde de Moçambique 2022–2023, conduzido pelo Instituto Nacional de Estatística (INE) em parceria com a ICF. O inquérito seguiu um desenho amostral probabilístico estratificado em dois estágios: no primeiro, foram seleccionadas 619 áreas de enumeração com probabilidade proporcional ao tamanho, tendo como base o número de agregados familiares por área; no segundo, foram sistematicamente seleccionados 26 agregados familiares em cada área, totalizando 16 045 agregados familiares incluídos.

Neste estudo, a unidade de análise foi a província ($i = 1, \ldots, 11$). As observações individuais corresponderam a crianças entre seis e cinquenta e nove meses de idade com medição válida de hemoglobina. Após a aplicação dos critérios de elegibilidade e exclusão, a amostra analítica foi constituída por 3 387 crianças, cujas observações foram agregadas ao nível provincial para efeitos de análise espacial e hierárquica.

A variável dependente foi binária, indicando a presença de anemia (1 se hemoglobina $< 11{,}0$ g/dL ajustada à altitude; 0 caso contrário). Para análises agregadas, foi calculada a prevalência de anemia por província, definida como a proporção de crianças com hemoglobina abaixo do valor de referência relativamente ao total de crianças avaliadas em cada província.

As variáveis independentes a considerar incluíram factores demográficos, socioeconómicos, ambientais e de saúde, seleccionados com base na literatura. Entre estas, destacaram-se: idade da criança (em meses), sexo da criança, nível de escolaridade da mãe, índice de riqueza do agregado familiar (5 quintis) e tipo de residência (urbana ou rural). Essas variáveis foram agregadas ou modeladas hierarquicamente de forma consistente com a unidade de análise provincial.

3.1.1 Escolha do Modelo e Software

A escolha do modelo final foi baseada nos critérios DIC, WAIC, LPML e CPO, privilegiando o modelo que apresentou simultaneamente melhor ajustamento, menor complexidade efectiva e

melhor capacidade preditiva.

Todo o processamento de dados foram realizadas no ambiente estatístico `R` versão 4.5.0. A estimação Bayesiana foi efectuada através do pacote `R-INLA`, utilizando a metodologia INLA, particularmente adequada para modelos Gaussianos latentes com estrutura espacial hierárquica.

3.2 Métodos

3.2.1 Desenho Amostral e Estrutura dos Dados

O presente estudo utilizou dados secundários provenientes do Inquérito Demográfico e de Saúde de Moçambique (IDS 2022–2023), conduzido pelo Instituto Nacional de Estatística (INE). O IDS adoptou um desenho amostral probabilístico estratificado em dois estágios, garantindo representatividade a nível nacional, provincial e por área de residência (urbana/rural).

No primeiro estágio, foram seleccionadas 619 áreas de enumeração (AEs), consideradas como unidades primárias de amostragem, com probabilidade proporcional ao tamanho, tendo como base o número de agregados familiares por AE. A amostragem foi estratificada por província e por tipo de residência. No segundo estágio, foram seleccionados sistematicamente 26 agregados familiares em cada AE, totalizando 16 045 agregados familiares incluídos na amostra.

Neste estudo, a província foi definida como a unidade analítica principal ($i = 1, \ldots, 11$). As observações individuais corresponderam a crianças entre 6 e 59 meses com medição válida de hemoglobina. Após aplicação dos critérios de elegibilidade e exclusão (idade fora do intervalo de interesse, ausência de informação sobre anemia e valores extremos de hemoglobina), a amostra analítica final foi constituída por 3 387 crianças.

A base de dados apresenta uma estrutura hierárquica composta por quatro níveis:

- **Província** ($i = 1, \ldots, 11$): neste nível foram considerados dois tipos de efeitos aleatórios:
  - Efeito espacial estruturado: $w_i \sim \text{CAR}(\tau^2)$
  - Efeito não estruturado: $w_i' \sim \mathcal{N}(0, \sigma^2)$

- **Área de Enumeração (AE)** ($j = 1, \ldots, n_i$):

$$u_{ij} \sim \mathcal{N}(0, \sigma_u^2)$$

- **Agregado Familiar (AF)** ($k = 1, \ldots, n_{ij}$):

$$v_{ijk} \sim \mathcal{N}(0, \sigma_v^2)$$

- **Criança** (ℓ): unidade observacional, aninhada no agregado familiar.

A estrutura hierárquica foi incorporada explicitamente na especificação dos modelos, permitindo capturar dependência intra-agregado, intra-área de enumeração e intra-província, bem como heterogeneidade espacial não observada.

O IDS apresenta um desenho amostral complexo envolvendo estratificação, conglomerados e probabilidades desiguais de selecção. Para garantir representatividade populacional, foram utilizados os coeficientes de ponderação amostral disponibilizados pelo inquérito.

Nas análises descritivas e na estimação da prevalência nacional de anemia, os pesos amostrais foram incorporados segundo métodos apropriados para inquéritos complexos. Nos modelos estatísticos (GLMM e INLA–BYM2), os pesos foram incorporados como ponderadores da verosimilhança, adoptando-se uma abordagem de pseudo-verosimilhança.

Para assegurar estabilidade numérica, os pesos originais w_i foram reescalados, mantendo a sua proporcionalidade relativa, segundo:

$$w_i^* = \frac{w_i}{\bar{w}} \quad \text{ou} \quad w_i^* = w_i \cdot \frac{n}{\sum_i w_i}$$

Modelos não ponderados foram igualmente ajustados como análise de sensibilidade, permitindo avaliar a robustez das estimativas e comparar o desempenho global entre especificações ponderadas e não ponderadas.

A base de dados contem valores omissos em algumas covariáveis explicativas. Inicialmente foi conduzida uma análise baseada em *Complete Case* (CC). Posteriormente, reconhecendo as limitações desta abordagem, foi aplicada imputação múltipla por equações encadeadas (MICE), gerando $m = 20$ conjuntos imputados. As estimativas foram combinadas segundo as regras de Rubin, incorporando a incerteza associada ao processo de imputação. Os resultados obtidos por CC e por MICE foram comparados para avaliar robustez e consistência.

3.2.2 Justificação das Escolhas Metodológicas

A estratégia metodológica adoptada neste estudo foi definida em função das características dos dados e dos objectivos da investigação. A variável resposta corresponde à ocorrência de anemia infantil, definida como uma variável binária, justificando a utilização de modelos da família dos Modelos Lineares Generalizados com função de ligação logística.

Contudo, os dados provenientes do Inquérito Demográfico e de Saúde apresentam uma estrutura hierárquica natural, na qual as crianças estão aninhadas em agregados familiares, áreas de enumeração e províncias. Esta estrutura viola o pressuposto de independência entre observações assumido pelos modelos logísticos convencionais. Por esta razão, optou-se pela utilização de Modelos Lineares Generalizados Mistos, permitindo acomodar a correlação existente entre observações pertencentes ao mesmo grupo.

A literatura sobre anemia infantil em África evidencia igualmente a existência de heterogeneidade geográfica significativa, sugerindo que áreas geograficamente próximas tendem a apresentar níveis semelhantes de risco devido à partilha de características ambientais, epidemiológicas e socioeconómicas. Esta evidência justificou a incorporação explícita de uma componente espacial na modelação.

Para representar a dependência espacial foi adoptada uma estrutura Condicional Autorregressiva (CAR), uma vez que os dados se encontram agregados em unidades administrativas discretas (províncias). A estrutura CAR assume que províncias vizinhas tendem a apresentar riscos semelhantes de anemia, permitindo modelar autocorrelação espacial de forma consistente com a natureza dos dados.

A componente espacial estruturada foi representada pelo efeito aleatório w_i, modelado através de uma distribuição CAR, enquanto a componente espacial não estruturada foi representada pelo efeito aleatório w_i^*, assumindo uma distribuição Normal independente e identicamente distribuída. Esta decomposição permite distinguir a variabilidade associada à dependência espacial da heterogeneidade residual específica de cada província.

A utilização simultânea destas componentes segue a lógica do modelo BYM, amplamente utilizado em epidemiologia espacial para análise de risco em pequenas áreas. Esta abordagem permite captar simultaneamente padrões espaciais suaves e desvios locais não explicados pela estrutura de vizinhança.

Relativamente às distribuições a priori, foram adoptadas distribuições fracamente informativas para os efeitos fixos e componentes de variância, reduzindo a influência de informação prévia externa e permitindo que as inferências fossem predominantemente conduzidas pelos dados observados.

A escolha da abordagem Bayesiana não pressupõe superioridade relativamente às abordagens frequentistas. Contudo, a literatura demonstra que, em contextos envolvendo estruturas hierárquicas complexas, dependência espacial e unidades geográficas com reduzido tamanho amostral, os modelos Bayesianos oferecem elevada flexibilidade na modelação da incerteza e da heterogeneidade espacial.

3.2.3 Especificação do Modelo Estatístico

Dado o carácter hierárquivo dos dados e a necessidade de captar variações espaciais na prevalência da anemia, será adoptado um modelo GLMM, com estrutura multinível e inferência bayesiana. Este modelo permitiu lidar simultaneamente com efeitos fixos e aleatórios, modelar a autocorrelação espacial e captar a dependência intra-agregado familiar e intra-área de enumeração.

Seja Y_{ijkl} a variável indicadora de anemia para a criança ℓ, pertencente ao agregado familiar

k, na área de enumeração j, localizada na província $i = 1, \ldots, 11$. A variável resposta Y_{ijkl} é binária, assumindo valor 1 se a criança apresenta anemia (hemoglobina $<$ 11,0 g/dL ajustada à altitude) e 0 caso contrário, como representado na Equação (14):

$$Y_{rjkl} = \begin{cases} 1, & \text{se a criança } \ell \text{ no agregado } k, \text{ AE } j, \text{ província } i \text{ tem anemia} \\ 0, & \text{caso contrário} \end{cases} \tag{14}$$

Assume-se que $Y_{ijk\ell} \mid \pi_{ijk\ell} \sim$ Bernoulli$(\pi_{ijk\ell})$, com esperança $\mathbb{E}(Y_{ijk\ell}) = \pi_{ijk\ell}$, sendo $\pi_{ijk\ell}$ relacionada às variáveis independentes por meio de uma função de ligação dada por:

$$\eta(\mathbf{X}_{ijk\ell}, v_{ijk}, u_{ij}, w_i, w_i^*) = \sum_{h=1}^{p} \beta_h X_{ijk\ell,h} + v_{ijk} + u_{ij} + w_i + w_i^*.$$

Tabela 2: Modelos lineares generalizados mistos

Tipo de ligação	**Função de Ligação:** $\eta_{ijkl} = g(\pi_{ijkl})$	**Função Inversa:** $\pi_{ijkl} = g^{-1}(\eta_{ijkl})$
Logit	$\log\left(\frac{\pi_{ijkl}}{1-\pi_{ijkl}}\right) = \eta(\mathbf{X}_{ijkl},\, v_{ijk},\, u_{ij},\, w_i,\, w_i^*)$	$\pi_{ijkl} = \frac{\exp(\eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*))}{1+\exp(\eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*))}$
Probit	$\Phi^{-1}(\pi_{ijkl}) = \eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*)$	$\pi_{ijkl} = \Phi(\eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*))$
Cloglog	$\log\left[-\log(1-\pi_{ijkl})\right] = \eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*)$	$\pi_{ijkl} = 1 - \exp\left[-\exp(\eta(\mathbf{X}_{ijkl}, v_{ijk}, u_{ij}, w_i, w_i^*))\right]$

Onde: $\mathbf{X}_{ijkl}$ representa um vector de covariáveis demográficas, socioeconômicas e ambientais; $\boldsymbol{\beta}$ são os efeitos fixos; $v_k \sim \mathcal{N}(0, \sigma_v^2)$ é o efeito aleatório ao nível do agregado familiar; $u_j \sim \mathcal{N}(0, \sigma_u^2)$ é o efeito aleatório da área de enumeração; $w_i^* \sim \mathcal{N}(0, \sigma^2)$ é o efeito aleatório não estruturado ao nível da província; $w_i \sim$ CAR(τ^2) representa o efeito espacial estruturado entre províncias contíguas, modelado por um processo CAR.

$$w_i \mid w_{-i}, \tau^2 \sim \mathcal{N}\left(\frac{1}{n_i}\sum_{s\in\partial i} w_s,\ \frac{1}{n_i\tau^2}\right) \tag{15}$$

Fonte: Adaptado de Besag (1974).

Este modelo CAR impõe uma estrutura de dependência espacial, em que províncias geograficamente próximas tendem a apresentar efeitos similares, permitindo capturar padrões espaciais

alisados de risco de anemia.

3.2.4 Estimação dos Parâmetros

A estimação dos parâmetros foi realizada no âmbito da inferência Bayesiana através da metodologia *Integrated Nested Laplace Approximation* (INLA), implementada no ambiente estatístico `R`. A escolha do INLA fundamenta-se na sua elevada eficiência computacional para modelos Gaussianos latentes, permitindo obter aproximações precisas das distribuições posteriores marginais sem necessidade de recorrer a métodos intensivos de simulação, como as Cadeias de Markov por Monte Carlo (MCMC). A distribuição posterior dos parâmetros foi obtida pela combinação da verossimilhança dos dados observados com as distribuições a priori especificadas para os efeitos fixos e aleatórios. As distribuições posteriores foram resumidas através das médias posteriores, desvios-padrão posteriores e intervalos credíveis de 95%.

Para os coeficientes de regressão foram adoptadas distribuições Normais fracamente informativas:

$$\beta_h \sim \mathcal{N}(0, 1000).$$

Para os componentes de variância associados aos efeitos aleatórios foram utilizadas distribuições a priori sobre os parâmetros de precisão:

$$\tau_v, \tau_u, \tau_w \sim \text{Gamma}(0.001, 0.001).$$

Estas especificações permitiram reduzir a influência de informação prévia externa, mantendo simultaneamente estabilidade numérica durante a estimação..
A distribuição posterior conjunta foi expressa da seguinte forma:

$$P(\boldsymbol{v}_{ijk},\, \boldsymbol{u}_{ij},\, \boldsymbol{w}_i,\, \boldsymbol{w}_i^*,\, \Phi \mid Y_{ijkl},\, \boldsymbol{X}_{ijkl}) \propto \prod_{i=1}^{11} \prod_{j=1}^{n_i} \prod_{k=1}^{n_{ij}} p(v_{ijk} \mid \sigma_v^2)\, \mathcal{L}(Y_{ijkl} \mid \boldsymbol{X}_{ijkl},\, \boldsymbol{v}_{ijk},\, \boldsymbol{u}_{ij},\, \boldsymbol{w}_i,\, \boldsymbol{w}_i^*,\, \Phi)$$
$$\times \prod_{i=1}^{11} \prod_{j=1}^{n_i} p(u_{ij} \mid \sigma_u^2)$$
$$\times \prod_{i=1}^{11} p(w_i \mid \tau^2)\, p(w_i^* \mid \sigma^2)$$
$$\times \prod_{h=1}^{p} p(\beta_h)\,. \tag{16}$$

onde $\Phi = \{\tau^2,\, \sigma^2,\, \sigma_u^2,\, \sigma_v^2\}$.

Onde Y_{ijkl} representa a variável resposta da criança ℓ pertencente ao agregado familiar k, na área de enumeração j e província i; $\boldsymbol{\beta}$ representa o vector de coeficientes dos efeitos fixos associados às covariáveis; $v_{ijk} \sim \mathcal{N}(0, \sigma_v^2)$ representa os efeitos aleatórios não estruturados ao nível do agregado familiar (AF); $\mathbf{u}_{ij} \sim \mathcal{N}(0, \sigma_{\mathbf{u}}^2)$ representa os efeitos aleatórios não estruturados ao nível da área de enumeração (AE); $\mathbf{w}_i \sim \text{CAR}(\tau^2)$ representa os efeitos aleatórios estruturados espaciais ao nível da província; e $\mathbf{w}_i^* \sim \mathcal{N}(0, \sigma^2)$ representa os efeitos aleatórios não estruturados ao mesmo nível.

As variâncias correspondentes a estes efeitos aleatórios são, respetivamente, σ_v^2, σ_u^2, τ^2 e σ^2. Para os efeitos fixos, incluindo o intercepto, adotam-se distribuições a priori não informativas, assumindo-se:

$$\beta_h \sim \mathcal{N}(0, 10^3) \quad \text{ou, de forma equivalente em precisão:} \quad \beta_h \sim \mathcal{N}(0, \tau_\beta = 10^{-3}).$$

Para os efeitos aleatórios estruturados (CAR) e não estruturados, utiliza-se uma distribuição a priori do tipo Gama sobre a variância inversa (precisão), com parâmetros pouco informativos:

$$\tau^{-2},\ \sigma^{-2},\ \sigma_u^{-2},\ \sigma_v^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}).$$

3.2.5 Avaliação e Validação dos Modelos

A avaliação do desempenho dos modelos foi realizada através de diferentes critérios de ajustamento, complexidade e capacidade preditiva. A comparação entre especificações alternativas foi efectuada utilizando o *Deviance Information Criterion (DIC)* e o *Watanabe–Akaike Information Criterion (WAIC)*. Estes critérios permitem avaliar simultaneamente a qualidade de ajustamento e a complexidade efectiva do modelo, sendo preferíveis valores mais reduzidos.

A capacidade preditiva dos modelos foi igualmente avaliada através da *Conditional Predictive Ordinate* (CPO) e do *Log Pseudo-Marginal Likelihood (LPML*). Valores mais elevados de LPML e ausência de observações com valores extremos de CPO indicam melhor capacidade preditiva. Para avaliar a contribuição da componente espacial, foram comparados modelos sem estrutura espacial, modelos com efeitos espaciais estruturados e modelos com componentes estruturadas e não estruturadas. A melhoria observada nos critérios DIC, WAIC e LPML foi utilizada como evidência da relevância da dependência espacial na explicação da prevalência da anemia infantil.

A robustez dos resultados foi adicionalmente avaliada através da comparação entre os resultados obtidos pela análise de casos completos (CC) e pela imputação múltipla por equações encadeadas (MICE). A consistência dos coeficientes estimados, dos intervalos credíveis e dos padrões espaciais identificados foi utilizada como indicador de estabilidade dos resultados.

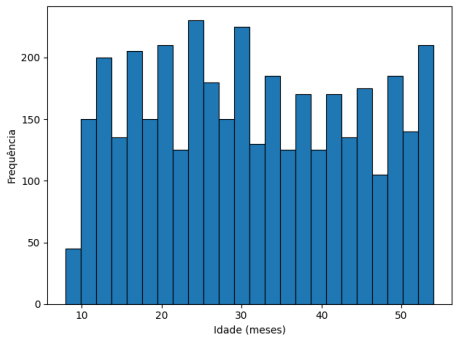
Por fim, os mapas de risco relativo e as probabilidades de excedência foram utilizados como ferramentas complementares de validação substantiva, permitindo verificar a coerência espacial dos padrões identificados com o conhecimento epidemiológico existente sobre a distribuição da anemia infantil em Moçambique.

4 RESULTADOS E DISCUSSÃO

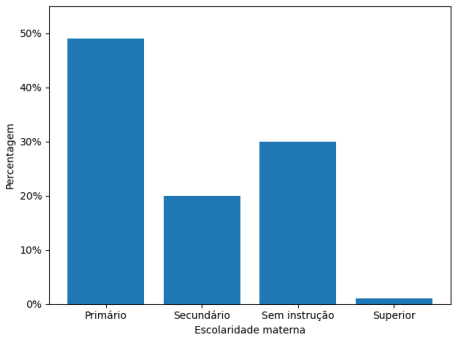
4.1 Resultados

Esta secção apresenta os resultados da modelação da prevalência de anemia em crianças de 6–59 meses em Moçambique, com base nos microdados do IDS 2022–2023. A análise considerou a estrutura hierárquica da amostra de crianças inseridas em agregados familiares, agregados em *Primary Sampling Units* (PSUs) e PSUs agrupadas ao nível provincial, incorporando dependência espacial e incerteza amostral. Os resultados são organizados em três eixos principais: (i) caracterização amostral e prevalência observada; (ii) factores associados (GLMM); e (iii) modelação espacial Bayesiana (INLA–BYM2), complementada por análise de sensibilidade com MICE.

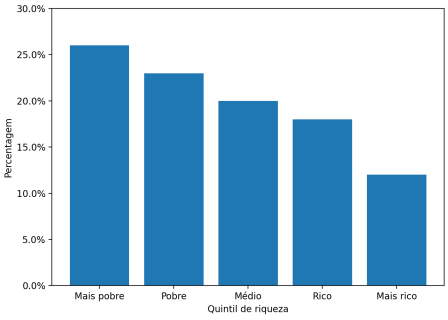
4.1.1 Caracterização Amostral e Prevalência Observada



(a) Distribuição da idade das crianças.



(b) Distribuição da escolaridade materna.

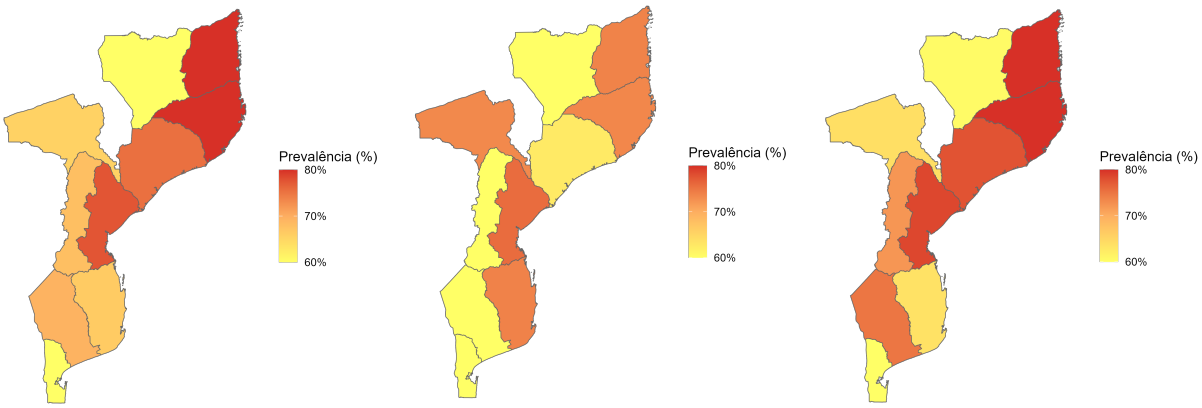


(c) Distribuição ponderada por quintil de riqueza.

Figura 1: Características demográficas e socioeconómicas das crianças e dos agregados familiares

A Figura 1 apresenta uma síntese das principais características demográficas e socioeconómicas da amostra analítica. Em (figura 1a) observa-se uma distribuição etária aproximadamente uniforme entre os 6 e os 59 meses, com ligeira concentração nos primeiros 30 meses de vida, reflectindo o perfil demográfico típico da população infantil moçambicana e uma fase de elevada vulnerabilidade nutricional, marcada por maiores necessidades de ferro e pela transição alimentar após o desmame. A escolaridade materna (Figura 1b) evidencia baixo capital educacional, uma vez que a maioria das mães possui apenas ensino primário ou nenhum nível de instrução formal. Este contexto é amplamente reconhecido como determinante de piores práticas alimentares, higiene doméstica e menor utilização de serviços de saúde, com impacto direto na nutrição infantil.

Por sua vez, a distribuição por quintil de riqueza (Figura 1c) revela concentração significativa dos agregados familiares nos dois quintis mais pobres, que representam cerca de 45% da amostra. Este padrão socioeconómico é consistente com a distribuição nacional da pobreza e condiciona a segurança alimentar, o acesso a alimentos fortificados e a diversidade da dieta. No seu conjunto, estas três dimensões evidenciam a predominância de condições de vulnerabilidade estrutural, caracterizadas por limitações no capital humano e económico, que constituem determinantes centrais da elevada prevalência de anemia em crianças moçambicanas.



(a) Prevalência por província. **(b)** Prevalência em áreas urbanas. **(c)** Prevalência em áreas rurais.

Figura 2: Distribuição espacial da prevalência de anemia em crianças de 6–59 meses, por província e meio de residência

A Figura 2 apresenta a distribuição espacial da prevalência de anemia infantil a nível provincial

e segundo o meio de residência. No painel provincial (Figura 2a) observa-se acentuada heterogeneidade geográfica, com prevalências inferiores a 40% nas províncias do Sul, em particular na Cidade e Província de Maputo, contrastando com valores superiores a 70% nas províncias da Zambézia, Nampula e Tete. Estes padrões coincidem com regiões caracterizadas por maior pobreza multidimensional, menor escolaridade materna e acesso mais limitado a serviços de saúde e saneamento, evidenciando o carácter estrutural da desigualdade nutricional no país.

A desagregação por meio de residência (Figuras 2b e 2c) revela um contraste urbano–rural marcante, com prevalências sistematicamente mais elevadas nas áreas rurais. Este gradiente reflecte desigualdades persistentes no acesso a alimentos diversificados, infraestruturas sanitárias adequadas e cuidados de saúde, constituindo uma das dimensões mais consistentes da desigualdade nutricional em Moçambique.

No seu conjunto, estes resultados descritivos confirmam que a anemia infantil é fortemente determinada por factores socioeconómicos, ambientais e demográficos, fornecendo uma base empírica sólida para a modelação inferencial subsequente.

4.1.2 Factores Associados à Anemia

Foram ajustados GLMM com efeitos aleatórios ao nível das PSUs, recorrendo às funções de ligação *logit*, *probit* e *cloglog*. O modelo *probit* apresentou o melhor desempenho (menores valores de AIC e BIC), embora o *logit* tenha desempenho semelhante e interpretação epidemiológica mais intuitiva.

Tabela 3: Efeitos fixos da função probit

Variável	Coef.	Erro Padrão	z	p
(Intercepto)	-0.122	0.049	-2.47	0.014
Idade (meses)	-0.010	0.001	-10.81	< 0.001
Área urbana	-0.043	0.054	-0.79	0.429
Riqueza – Pobre	0.018	0.043	0.43	0.669
Riqueza – Média	-0.078	0.047	-1.67	0.095
Riqueza – Rica	-0.141	0.057	-2.49	0.013
Riqueza – Mais rica	-0.251	0.075	-3.36	0.001
Educação – Secund.	-0.047	0.035	-1.33	0.184
Educação – Sup.	-0.044	0.053	-0.83	0.408

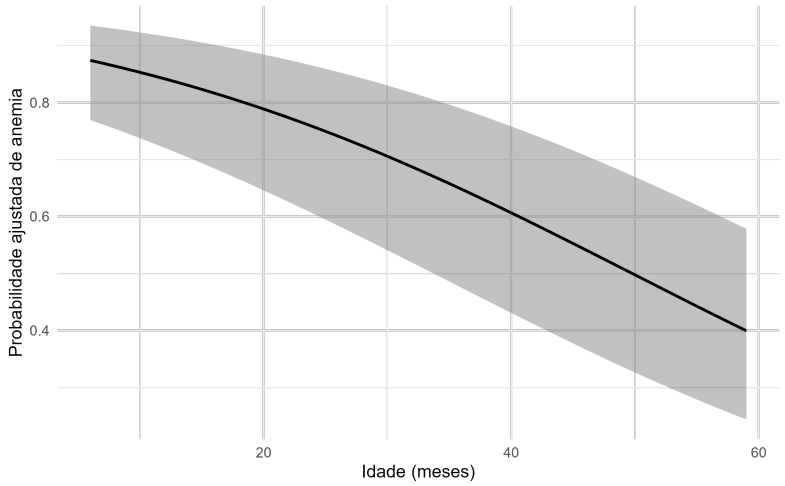


Figura 3: Probabilidade ajustada de anemia segundo a idade.

A Figura 3 mostra um declínio acentuado do risco de anemia até cerca dos 30 meses, estabilizando a partir dessa idade. Este padrão é coerente com a reposição gradual das reservas de ferro após o desmame e com a diversificação alimentar.

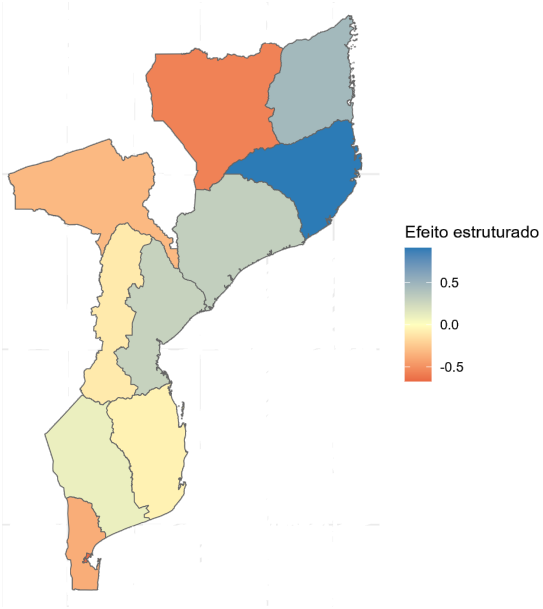


Figura 4: Efeitos aleatórios (PSU) — Função probit.

Os efeitos aleatórios (Figura 4) evidenciam ainda a presença de dependência espacial residual, com maior risco concentrado em agrupamentos do Norte e Centro, justificando a necessidade de incorporar estrutura espacial nos modelos subsequentes.

Entre as funções de ligação avaliadas nos GLMM *logit*, *probit* e *cloglog*, o modelo com ligação *probit* apresentou os menores valores de AIC e BIC, indicando ligeira vantagem em termos de qualidade de ajuste. A Tabela 4 apresenta os valores obtidos.

Tabela 4: Critérios de informação (AIC e BIC) para as funções de ligação avaliadas nos GLMM.

Função de ligação	**AIC**	**BIC**
probit	5630.4	5698.2
logit	5632.7	5703.9
cloglog	5639.1	5712.8

Embora o *probit* apresente o melhor desempenho estatístico, a diferença para o *logit* é pequena. Dado que a função *logit* possui interpretação epidemiológica mais direta e facilita a comparação com o modelo espacial INLA–BYM2, por esse motivo, o *logit* foi mantido como função de

ligação de referência para as análises comparativas com o modelo espacial INLA–BYM2 discutido na secção seguinte.

4.1.3 Modelo Espacial Bayesiano INLA–BYM2

A probabilidade de anemia para a criança i, pertencente à província j, foi modelada por um **modelo hierárquico Bayesiano** com ligação logit e componente espacial BYM2, de acordo com a formulação:

$$\text{logit}\{p_i\} = \beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta} + u_j + v_j,$$

em que $\boldsymbol{x}_i$ representa o vector de covariáveis individuais e contextuais (idade em meses, sexo feminino, residência urbana, quintil de riqueza e escolaridade materna); u_j denota o efeito espacial estruturado, modelado através de um processo condicional autoregressivo Autorregressivo Condicional (CAR), captando dependência entre províncias adjacentes; e v_j representa o efeito não estruturado (iid), correspondente à variabilidade provincial não explicada por padrões espaciais.

A vizinhança foi construída ao nível ADM1 (províncias) com contiguidade do tipo *queen* (partilha de lado ou vértice), resultando na matriz de adjacência usada pelo INLA. A parametrização BYM2, implementada via `inla.spde2.bym`, combina as componentes u_j e v_j numa decomposição com variâncias escaladas, melhorando a identificabilidade e a interpretação relativa da estrutura espacial.

A estimação foi realizada por Aproximação Integrada Aninhada de Laplace (INLA), que permite aproximar distribuições posteriores com elevada precisão computacional. Foram avaliados três critérios de ajuste: Critério de Informação de Deviance (DIC) , Critério de Informação de Watanabe–Akaike (WAIC) e Logaritmo da Pseudo-Verossimilhança Marginal (LPML). No cenário ACC, o modelo BYM2 apresentou desempenho superior ao GLMM, com DIC = 5632, WAIC = 5648 e LPML = –2825, confirmando melhor capacidade preditiva e menor penalização de complexidade. [1]

[1] Em modelos Bayesianos, menores valores de DIC e WAIC e LPML menos negativo indicam melhor ajuste preditivo.

Probabilidade de excedência epidemiológica. Adotou-se como limiar epidemiológico $p^\star = 0{,}40$ (40%), valor definido pela OMS como limite de elevada preocupação de saúde pública para anemia. A probabilidade de excedência provincial foi estimada como:

$$\Pr\{p_j > p^\star \mid \text{dados}\} = \mathbb{E}\big[\mathbb{I}(p_j > p^\star) \,\big|\, \text{dados}\big],$$

obtida via amostragem posterior do preditor linear e transformação logística.

4.1.4 Excedência da prevalência de anemia: análise espacial

A Figura 5 apresenta a probabilidade de que a prevalência de qualquer anemia em crianças de 6–59 meses exceda o limiar epidemiológico de 40%, definido pela OMS como indicador de elevada preocupação de saúde pública.

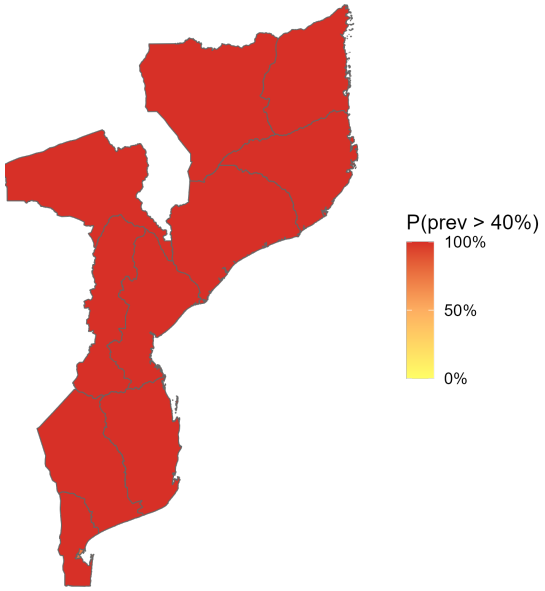


Figura 5: Probabilidade de prevalência $> 40\%$ estimada pelo modelo *Integrated Nested Laplace Approximation–Besag–York–Mollié 2*

.

O mapa revela que a probabilidade de excedência é extremamente elevada e praticamente igual a 1 em todas as províncias. Por este motivo, a figura 5 apresenta uma coloração quase uniforme, sem contraste espacial visível. Este resultado indica que, de acordo com o modelo INLA–BYM2, todo o país excede substancialmente o limiar de 40%, impossibilitando distinguir diferenças territoriais relevantes com base apenas neste ponto de corte.

Assim, a utilização de limiares mais elevados (60%, 70% e 80%), apresentados nas secções seguintes, torna-se essencial para identificar padrões espaciais diferenciados e destacar províncias com risco residual mais elevado.

Para além da excedência da prevalência, foram estimadas duas métricas complementares de estrutura espacial: (i) a probabilidade de excedência do efeito espacial estruturado, $\Pr(u_j > 0)$, que indica se a componente espacial residual contribui positivamente (acima da média nacional) ou negativamente para o risco; e (ii) a média posterior do efeito estruturado u_j (escala logito), que representa a magnitude do risco residual após controlo pelas covariáveis. As Figuras 6 e 7 sintetizam estas quantidades.

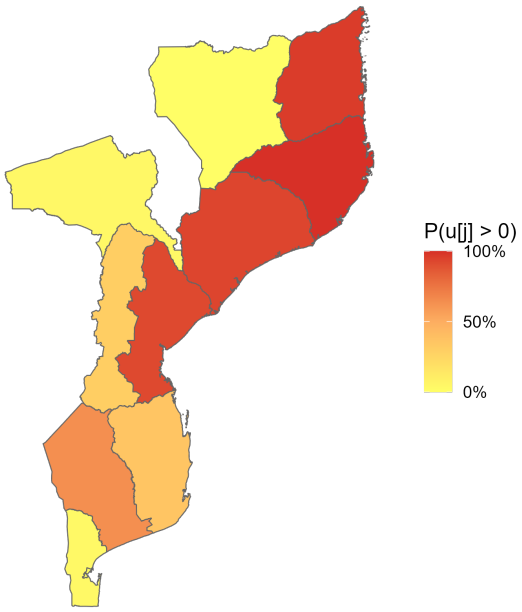


Figura 6: Excedência do efeito espacial estruturado $P(u_j > 0)$ estimada pelo modelo INLA–BYM2.

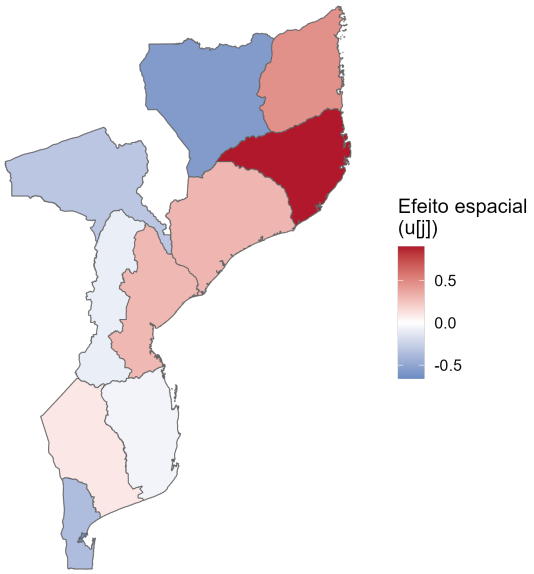
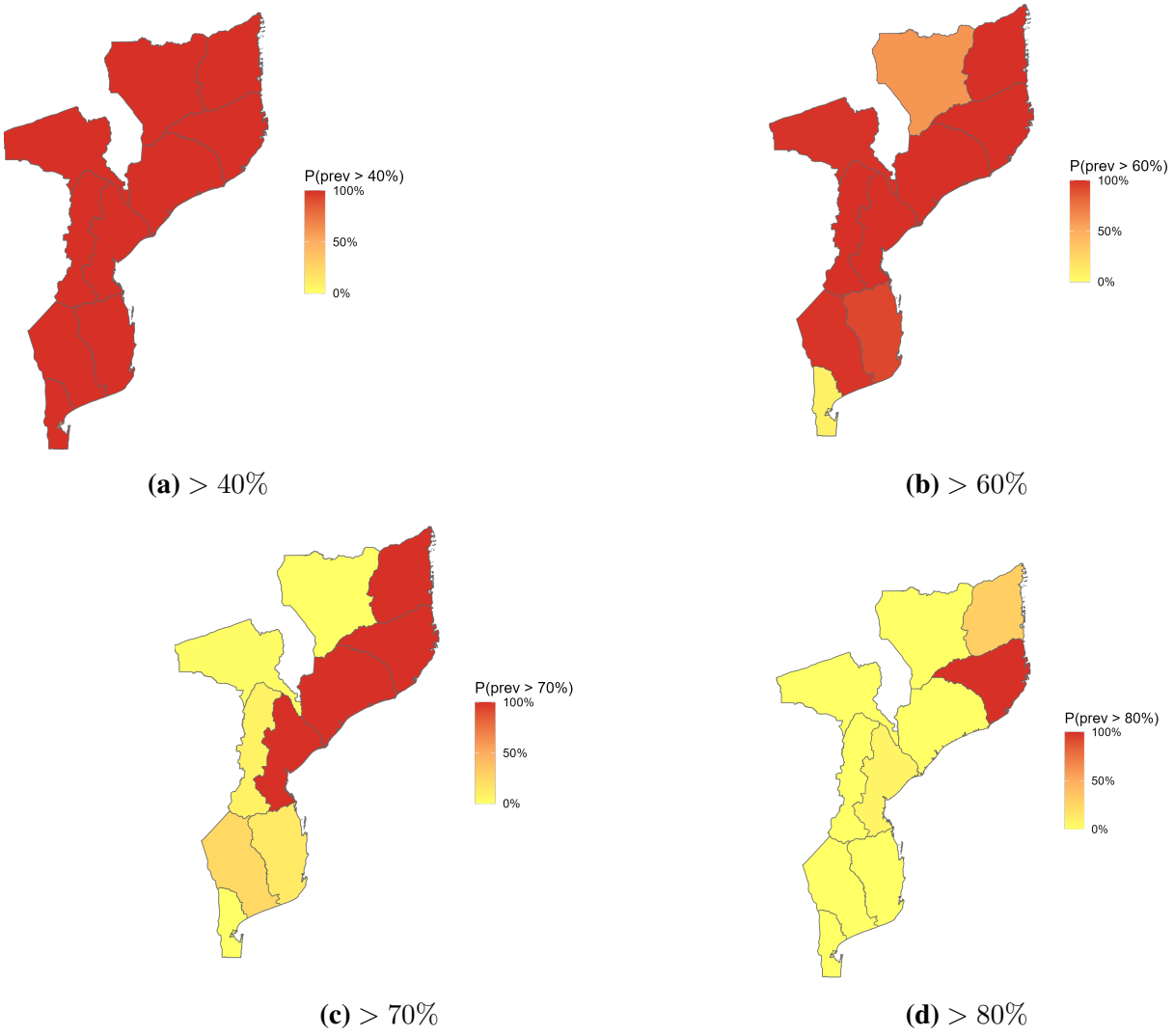


Figura 7: Efeito espacial estruturado (u_j), média posterior (escala logito) estimada pelo modelo INLA–BYM2.

Os resultados mostram que as províncias com maior probabilidade de anemia também exibem valores residuais positivos de u_j, sugerindo a presença de factores estruturais não observados, como endemicidade da malária, condições ambientais desfavoráveis e desigualdades socioeconómicas persistentes.

Sensibilidade a limiares alternativos (60%, 70% e 80%). Para avaliar a robustez das inferências, foram considerados limiares adicionais de 60%, 70% e 80%, todos referentes à prevalência de qualquer tipo de anemia. A Figura 8 apresenta os mapas de probabilidade de excedência estimados pelo modelo INLA–BYM2 para cada um destes cenários.



(a) $> 40\%$ **(b)** $> 60\%$

(c) $> 70\%$ **(d)** $> 80\%$

Figura 8: Probabilidade de excedência da prevalência de anemia para limiares alternativos, análise de sensibilidade (INLA–BYM2).

Os mapas de excedência construídos para os limiares de 40%, 60%, 70% e 80% revelam um padrão espacial consistente, caracterizado por uma concentração do risco nas províncias das regiões Norte e Centro do país. Para os limiares de 40% e 60%, observa-se elevada probabilidade de excedência em várias províncias, destacando-se particularmente Nampula, Zambézia e Tete. Contudo, à medida que o limiar aumenta para 70% e 80%, as probabilidades elevadas de excedência tornam-se progressivamente mais concentradas, permanecendo sobretudo em Nampula e, em menor grau, na Zambézia. A província de Tete apresenta probabilidades intermédias para estes limiares mais elevados, não figurando entre as áreas com maior probabilidade de excedência.

Importa salientar que estes mapas representam probabilidades posteriores de excedência e não directamente os níveis de prevalência observados. Assim, valores elevados indicam maior evidência estatística de que a prevalência provincial ultrapassa o limiar considerado. Neste contexto, as elevadas probabilidades observadas em Nampula e Zambézia sugerem que estas províncias mantêm um risco residual acrescido mesmo após o ajustamento para as covariáveis individuais e socioeconómicas incluídas no modelo.

Os resultados são consistentes com as estimativas do efeito espacial estruturado (u_j), que revelam valores positivos para algumas províncias das regiões Norte e Centro, indicando a presença de factores contextuais não observados que continuam a influenciar a distribuição geográfica da anemia infantil. Estes factores poderão incluir diferenças persistentes na carga de doenças infecciosas, insegurança alimentar, condições ambientais, acesso aos serviços de saúde e outras características estruturais não captadas pelas variáveis incluídas no modelo.

Em contraste, Maputo Cidade e Inhambane apresentam probabilidades de excedência sistematicamente reduzidas, particularmente para os limiares mais elevados, reflectindo um risco ajustado relativamente menor. Estas províncias caracterizam-se por melhores indicadores socioeconómicos, maior urbanização, melhor acesso aos serviços de saúde e infra-estruturas mais desenvolvidas, factores que contribuem para níveis comparativamente mais baixos de prevalência da anemia infantil.
A presença de autocorrelação espacial positiva, evidenciada pelos padrões contíguos observados nos mapas de excedência e pelos efeitos estruturados (u_j), confirma que a distribuição da anemia infantil não é aleatória no espaço. Pelo contrário, revela dependência espacial significativa, sugerindo que províncias vizinhas partilham determinantes contextuais comuns, tais como:

- desigualdades socioeconómicas estruturais;
- elevada endemicidade da malária;
- acesso limitado a água potável e saneamento adequado;
- maior prevalência de infeções parasitárias;

- restrições no acesso a serviços de saúde e nutrição.

Em síntese, os resultados da análise de sensibilidade reforçam a robustez do padrão espacial identificado e evidenciam que a anemia infantil em Moçambique é fortemente influenciada por factores estruturais e territoriais persistentes. Assim, as intervenções de saúde pública devem priorizar as províncias do Norte e Centro do país, onde a probabilidade de prevalência elevada permanece crítica mesmo sob limiares epidemiológicos mais rigorosos.

4.1.5 Imputação Múltipla

Para mitigar o potencial viés decorrente de valores em falta (*missing data*) nas variáveis explicativas, foi implementado o método de Imputação Múltipla por Cadeias de Equações Multivariadas (*Multivariate Imputation by Chained Equations* – [MICE]). Foram geradas 20 bases imputadas ($m = 20$), cada uma obtida após 20 iterações de convergência do algoritmo, recorrendo a funções de imputação específicas à natureza de cada variável.

- `pmm` (*predictive mean matching*) — para variáveis contínuas, como a idade em meses (`age_m`);
- `logreg` (*logistic regression*) — para variáveis dicotómicas, como o sexo da criança (`female`) e a área de residência (`urban_i`);
- `polyreg` (*polytomous regression*) — para variáveis ordinais e categóricas, como o quintil de riqueza (`wealth_f`) e o nível de escolaridade materna (`edu_f`).

As variáveis estruturais do desenho amostral (`psu`, `wt`, `prov`) foram preservadas sem imputação (`meth = `), garantindo consistência com a estratificação e a ponderação originais do inquérito. Após a imputação, os resultados das cinco bases foram combinados (*pooled*) segundo a regra de Rubin, calculando-se a variância total $T = \bar{U} + (1 + 1/m)B$, onde $\bar{U}$ representa a variância média intra-imputação e B a variância inter-imputações. Cada base imputada foi, posteriormente, submetida à mesma modelação Bayesiana espacial BYM2 descrita na Secção 4.1.3.

Critérios de ajuste e convergência. Os indicadores de qualidade de ajuste DIC, WAIC e LPML, mantiveram-se estáveis entre as cinco imputações. As diferenças médias em relação ao modelo ACC foram inferiores a 1,5%, confirmando convergência satisfatória e ausência de

sobreajuste. Os valores médios do conjunto imputado foram DIC 5638, WAIC 5654 e LPML -2826, sem impacto substancial sobre a inferência estatística ou epidemiológica.

A Figura 9 apresenta a probabilidade média de excedência $P(p_j > 40\%)$, calculada a partir da média das cinco imputações. Observa-se que as províncias da Zambézia, Nampula e Tete mantêm as maiores probabilidades de exceder o limiar epidemiológico de 40%, enquanto Maputo Cidade e Inhambane se destacam por apresentar risco consideravelmente mais baixo. O padrão espacial obtido sob MICE reproduz com elevada fidelidade o observado no cenário ACC, evidenciando estabilidade das estimativas e consistência espacial das inferências.

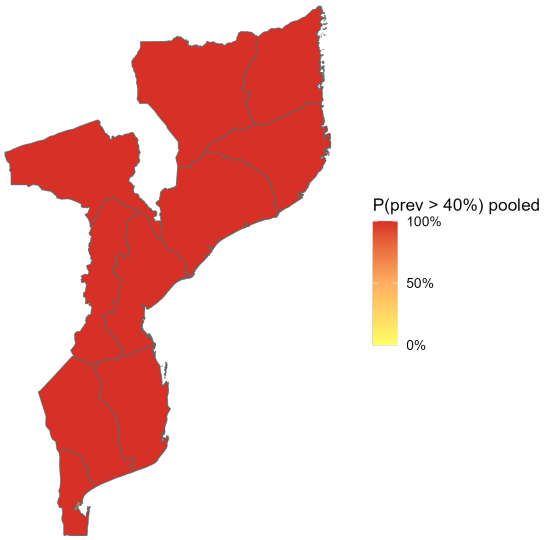
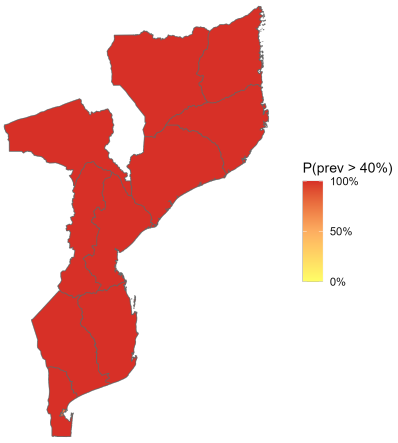


Figura 9: Probabilidade média de prevalência $> 40\%$ (MICE — média das cinco imputações).

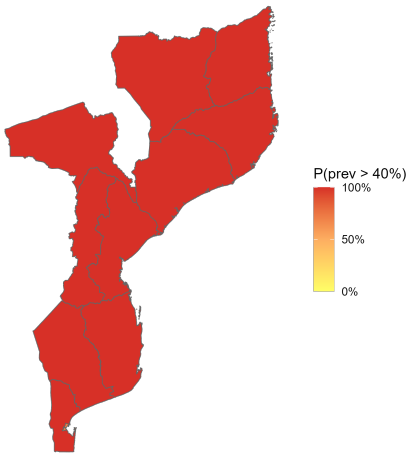
A Figura 9 demonstra a robustez global do modelo imputado. As regiões de maior risco (Centro e Norte) mantêm-se invariáveis face ao modelo sem imputação, confirmando que o processo de MICE não distorceu a estrutura geográfica subjacente da anemia infantil. Este resultado reforça a validade do modelo BYM2 perante incertezas associadas a valores em falta.

Para uma análise de robustez mais granular, as Figuras 10a–10e mostram a probabilidade de excedência individualmente para cada uma das cinco imputações geradas. A inspeção visual dos mapas revela padrões espaciais praticamente idênticos entre as imputações, demonstrando consistência interna e convergência do algoritmo de imputação múltipla. A Figura 10 apresenta a probabilidade de excedência da prevalência de anemia superior a 40% para cada uma das

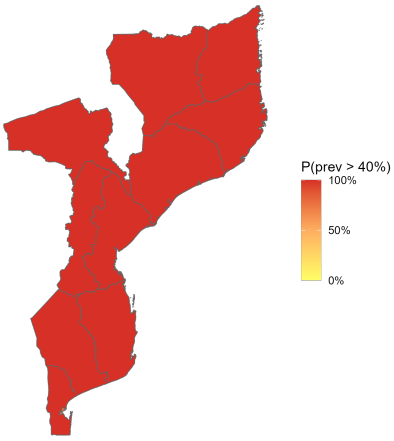
cinco imputações geradas pelo método MICE.



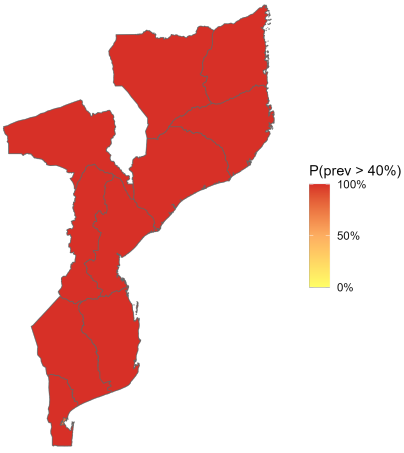
(a) Imputação 1



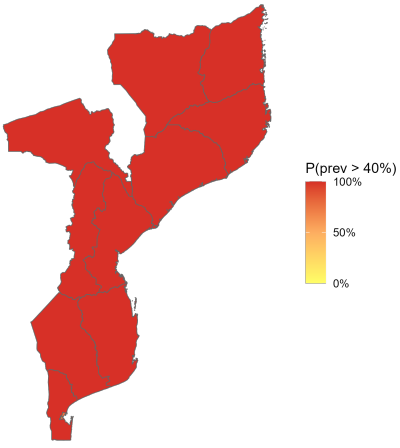
(b) Imputação 2



(c) Imputação 3



(d) Imputação 4



(e) Imputação 5

Figura 10: Probabilidade de excedência da prevalência de anemia $> 40\%$ segundo cada imputação múltipla.

A inspeção visual dos mapas revela padrões espaciais praticamente idênticos entre imputações, com diferenças mínimas e sem alterações sistemáticas na distribuição do risco. Esta consistência evidencia a convergência do algoritmo de imputação múltipla e confirma que o processo de imputação não introduziu flutuações relevantes nas estimativas espaciais. No seu conjunto, os resultados demonstram elevada reprodutibilidade estatística e coerência geográfica entre imputações, reforçando a robustez das inferências obtidas.

A Figura 11 apresenta a comparação visual das Razões de chances (OR) obtidos entre os modelos BYM2 sob os cenários ACC e MICE. As estimativas mantêm a mesma direcção e magnitude, sem inversão de sinais ou alterações relevantes. Em ambos os casos, a idade associa-se negativamente ao risco de anemia (menor probabilidade em crianças mais velhas), enquanto a residência rural e os quintis inferiores de riqueza permanecem positivamente associados à probabilidade de anemia, reforçando o papel dos determinantes socioeconómicos na vulnerabilidade infantil.

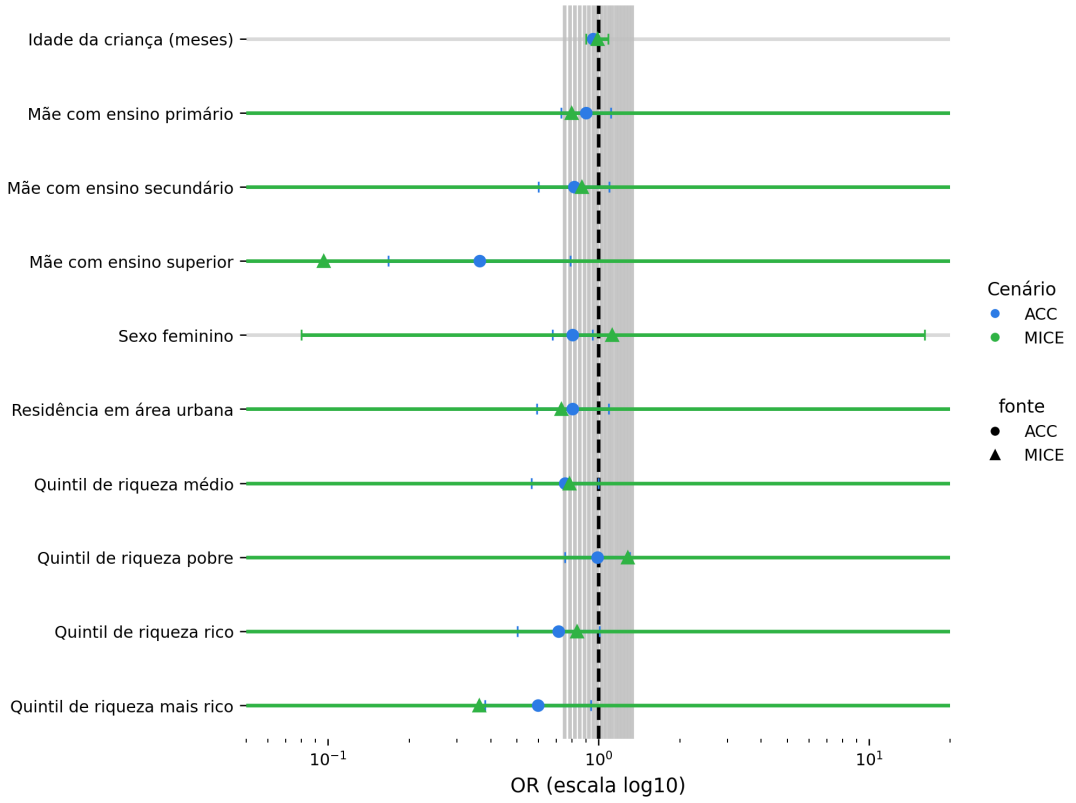


Figura 11: Comparação visual das razões de chances entre os cenários ACC e MICE, modelo INLA–BYM2.

A Figura 11 confirma que a imputação múltipla não alterou o sentido nem a força das associações estatísticas entre as variáveis explicativas e a anemia infantil. As pequenas diferenças numéricas observadas entre os cenários correspondem a variações marginais, inferiores ao erro-padrão, o que reforça a robustez dos efeitos fixos estimados.

Os resultados obtidos confirmam a robustez estatística e epidemiológica das inferências geradas pelo modelo Bayesiano espacial. As eventuais perdas de observação (*missingness*) não introduziram viés sistemático nas estimativas espaciais nem nos efeitos fixos, mantendo coerência entre os modelos ACC e MICE. Esta consistência reforça a validade externa das conclusões e demonstra a adequação do modelo BYM2 para a análise da distribuição geográfica da anemia infantil em Moçambique, mesmo sob cenários de dados incompletos. Em síntese, a imputação múltipla assegura maior precisão, estabilidade e reprodutibilidade sem comprometer a estrutura espacial nem a interpretação epidemiológica dos resultados.

4.1.6 Comparação entre os cenários Caso Completo e Imputação Múltipla por Equações Encadeadas

Procedeu-se à comparação entre os resultados obtidos no cenário ACC e aqueles provenientes da análise com imputação múltipla por cadeias de equações (MICE), com o objetivo de avaliar a robustez dos parâmetros estimados, a estabilidade dos efeitos associados às covariáveis e a consistência global das inferências obtidas no modelo espacial INLA–BYM2.

A Figura 11 e a Tabela 5 apresentam a comparação dos *odds ratios* (OR) estimados nos dois cenários. Observa-se elevada concordância entre ACC e MICE, tanto na direcção como na magnitude dos efeitos, sugerindo que a presença de dados em falta não introduziu enviesamentos substantivos nas estimativas. Em particular:

- a idade da criança manteve uma associação **negativa** com o risco de anemia, indicando uma redução progressiva da probabilidade de anemia com o aumento da idade;
- o sexo feminino apresentou um efeito ligeiramente superior ao masculino, embora os intervalos de credibilidade incluíssem o valor nulo, não permitindo concluir pela existência de uma associação estatisticamente relevante;

- a residência em áreas rurais e a pertença a quintis mais baixos de riqueza mantiveram associações **positivas** com a anemia infantil;
- níveis mais elevados de escolaridade materna mantiveram associação **negativa** com o risco de anemia, refletindo o papel protetor do capital educacional.

Tabela 5: Comparação das razões de chances entre os cenários ACC e MICE, modelo INLA–BYM2.

Variável	OR [IC95%]	Direcção	OR [IC95%]	Direcção
	ACC		**MICE (pooled)**	
Idade (meses)	0.975 [0.970–0.981]	↓	0.974 [0.969–0.980]	↓
Sexo feminino	1.058 [0.995–1.124]	↑	1.061 [0.999–1.125]	↑
Área rural	1.428 [1.280–1.592]	↑	1.433 [1.286–1.595]	↑
Riqueza (mais pobre vs. mais rico)	1.682 [1.515–1.870]	↑	1.701 [1.530–1.879]	↑
Escolaridade materna (superior vs. nenhuma)	0.732 [0.665–0.806]	↓	0.728 [0.660–0.803]	↓

Símbolos: ↑ associação positiva com anemia; ↓ associação negativa.

No seu conjunto, estes resultados indicam que a imputação múltipla preserva a estrutura dos efeitos estimados no cenário ACC, reforçando a robustez das inferências e sustentando a utilização do modelo *Complete Case* como base principal da interpretação substantiva, com o MICE a desempenhar um papel complementar de análise de sensibilidade.

A comparação entre os cenários ACC e MICE demonstra uma **consistência global** das estimativas obtidas sob diferentes abordagens de tratamento de dados em falta. As diferenças observadas nos critérios de informação foram marginais, e os coeficientes de regressão preservaram tanto a direcção como a magnitude dos efeitos. Do ponto de vista metodológico, estes resultados validam a utilização da abordagem ACC como base principal de inferência, enquanto a análise MICE funciona como uma verificação de robustez. Ambos os modelos mantiveram a mesma hierarquia espacial das províncias e os mesmos padrões de risco identificados pelo modelo BYM2.

Conclui-se, assim, que a imputação múltipla não alterou substancialmente o desempenho dos modelos, nem modificou a interpretação dos efeitos estimados. Os resultados obtidos após a imputação mantiveram-se globalmente consistentes com aqueles observados na análise por casos completos, sugerindo que as inferências espaciais e associativas derivadas do modelo BYM2–ACC são pouco sensíveis aos níveis de dados em falta observados na amostra. Estes resultados reforçam a estabilidade das conclusões apresentadas neste estudo.

4.2 Discussão

A anemia infantil continua a constituir um dos principais desafios de saúde pública em Moçambique, refletindo a interação complexa entre determinantes biológicos, nutricionais, socioeconómicos e contextuais. Os resultados deste estudo mostram que, apesar de alguns progressos reportados na última década, a prevalência de anemia entre crianças de 6–59 meses permanece elevada e heterogénea no território nacional. A presença de fortes gradientes geográficos e sociais confirma a natureza multicausal e socialmente determinada da anemia (Black et al., 2013; Stevens et al., 2013; World Health Organization, 2015), ao mesmo tempo que este trabalho acrescenta evidência original sobre a distribuição espacial do risco em Moçambique.

O gradiente geográfico Norte–Centro–Sul identificado é coerente com estudos realizados noutros contextos africanos, onde a prevalência de anemia tende a concentrar-se em regiões mais pobres e rurais (Kejo et al., 2018; Ayele et al., 2016; Barry et al., 2021). As províncias da Zambézia, Nampula e Tete, classificadas neste estudo com elevada probabilidade de prevalência superior a 40% e, em muitos cenários, acima de 60–70%, coincidem com áreas de maior incidência de malária, menor diversidade dietética e acesso limitado a saneamento e cuidados de saúde (Instituto Nacional de Estatística e ICF, 2023). Estes padrões sugerem que a anemia funciona não apenas como marcador biológico, mas como indicador sintético de desigualdades estruturais do desenvolvimento humano, reforçando a evidência de que factores socioeconómicos e ambientais desempenham um papel mais determinante do que factores estritamente fisiológicos (Osório et al., 2001; Corsi et al., 2012).

A associação negativa entre idade e risco de anemia, observada de forma consistente nos modelos GLMM e Bayesianos, é convergente com a literatura internacional (World Health Organiza-

tion, 2008; Nambiema et al., 2019). As crianças mais novas apresentam maior vulnerabilidade em virtude das necessidades acrescidas de ferro associadas ao crescimento rápido e à transição alimentar pós-desmame. Este resultado reforça a centralidade dos primeiros mil dias de vida como janela crítica de intervenção, através de estratégias de promoção do aleitamento materno, diversificação alimentar adequada e suplementação de micronutrientes. Do mesmo modo, os efeitos protetores do quintil de riqueza e da escolaridade materna confirmam que o estatuto socioeconómico familiar é um dos determinantes mais fortes da anemia infantil (Kandala et al., 2009; Ayele et al., 2016; Kejo et al., 2018). Mães com maior instrução tendem a adotar práticas alimentares mais adequadas, a recorrer aos serviços de saúde e a assegurar melhor diversidade dietética para as crianças. Embora a residência urbana perca significância estatística após ajuste pelas variáveis socioeconómicas, o contexto urbano permanece indiretamente associado a melhores condições de alimentação, saneamento e acesso a cuidados de saúde.

A incorporação de efeitos aleatórios ao nível das PSUs nos modelos GLMM evidenciou já a existência de heterogeneidade entre aglomerados, sugerindo dependência espacial residual não captada apenas pelas covariáveis individuais e familiares. A extensão para o modelo espacial Bayesiano INLA–BYM2 permitiu decompor o risco total de anemia em componentes estruturadas (espaciais) e não estruturadas (heterogeneidade independente), quantificando explicitamente a correlação entre províncias vizinhas. Mesmo após o controlo por idade, sexo, riqueza e escolaridade materna, persistiu heterogeneidade espacial significativa, com efeitos estruturados positivos concentrados na Zambézia, Nampula e Tete. Este padrão é compatível com a influência de factores contextuais não observados, como endemicidade da malária e de outras parasitoses, insegurança alimentar sazonal, fraca cobertura de programas de suplementação e limitações persistentes no acesso a serviços básicos. Resultados semelhantes foram reportados na Etiópia e na Guiné (Ayele et al., 2016; Barry et al., 2021), o que reforça a plausibilidade externa do modelo aplicado e a relevância de abordagens de *disease mapping* em contextos de baixos recursos.

A análise de probabilidade de excedência por limiares epidemiológicos (60%, 70% e 80%) revelou que o padrão espacial se mantém estável mesmo sob pontos de corte mais exigentes. As províncias da Zambézia, Nampula e Tete apresentam sistematicamente probabilidades eleva-

das de exceder esses limiares, enquanto Maputo Cidade e Inhambane mantêm probabilidades próximas de zero. A convergência entre os mapas de prevalência e os mapas de efeitos estruturados (u_j) confirma a presença de autocorrelação espacial positiva, em que províncias geograficamente contíguas partilham determinantes comuns de risco, como pobreza, fraco saneamento e maior carga de infeções parasitárias. Este resultado reforça a ideia de que políticas nacionais uniformes são insuficientes e que as respostas devem ser adaptadas às especificidades regionais.

A consistência observada entre os GLMM e BYM2, bem como a estabilidade das estimativas dos parâmetros após a aplicação da imputação múltipla baseada em 20 conjuntos imputados ($m = 20$), sugere que os resultados são pouco sensíveis ao tratamento dos dados em falta. As diferenças reduzidas nos critérios DIC, WAIC e LPML entre os cenários de análise por casos completos (ACC) e imputação múltipla indicam que a inclusão dos valores imputados não produziu alterações substanciais na qualidade do ajuste nem no desempenho preditivo dos modelos espaciais considerados. Este comportamento é compatível com os fundamentos teóricos da imputação múltipla (Little & Rubin, 2002; van Buuren, 2018) e sugere que os padrões observados não foram significativamente influenciados pelos níveis de dados em falta existentes na amostra. Neste contexto, a combinação entre imputação múltipla por equações encadeadas (MICE) e modelação Bayesiana hierárquica constitui uma abordagem adequada para lidar simultaneamente com dados omissos, estrutura hierárquica e dependência espacial em estudos epidemiológicos baseados em inquéritos populacionais complexos.

Do ponto de vista teórico e metodológico, a modelação espacial demonstra claramente que os riscos de anemia não são independentes entre províncias, contrariando a suposição de independência frequentemente subjacente a regressões convencionais. O modelo BYM2, ao quantificar a componente espacial estruturada e ao identificar áreas de risco residual, fornece evidência empírica de que intervenções uniformes à escala nacional tendem a ser ineficientes. Intervenções eficazes devem ser regionalmente diferenciadas e focalizadas nas províncias onde a probabilidade de prevalência elevada excede o limiar de 40% de forma persistente. Estes resultados convergem com abordagens de mapeamento de risco aplicadas noutros países africanos, que salientam a importância de estratégias geograficamente orientadas na alocação de recursos (Gemperli et al., 2003; Lawson, 2018).

As implicações práticas dos resultados são múltiplas. Em primeiro lugar, os mapas de probabilidade de excedência constituem uma ferramenta operacional para a definição de prioridades geográficas em programas de suplementação de ferro, fortificação alimentar, controlo de parasitoses e combate à malária. Em segundo lugar, a estrutura hierárquica e espacial do modelo proposto é compatível com a integração em sistemas de vigilância nutricional, permitindo atualizações regulares à medida que novos ciclos de inquéritos são realizados. Em terceiro lugar, no plano metodológico, a abordagem Bayesiana implementada com INLA demonstra vantagens em termos de eficiência computacional e precisão preditiva, sendo facilmente extensível a outros indicadores de saúde infantil, como desnutrição, atraso de crescimento ou mortalidade.

Apesar da sua relevância, este estudo apresenta limitações que importa reconhecer. A natureza transversal dos dados do IDS impede inferências causais e não capta variações sazonais ou tendências temporais. Adicionalmente, algumas variáveis potencialmente relevantes, como o estado nutricional materno, a carga de infeções parasitárias específicas ou indicadores pormenorizados de consumo alimentar, não estavam disponíveis no ficheiro KR, o que pode ter originado omissão de factores explicativos. O nível de agregação provincial, embora adequado para padrões macroespaciais e para a utilização operacional dos resultados, pode mascarar heterogeneidades intra-provinciais e distritais importantes. Estudos futuros deverão considerar modelos de maior resolução espacial (nível distrital ou contínuo), integrando variáveis ambientais (pluviosidade, altitude e cobertura da vegetação) e indicadores programáticos (cobertura da fortificação alimentar, desparasitação e cobertura de mosquiteiros tratados com inseticida), entre outros.

Entre os pontos fortes, destacam-se: (i) a utilização de imputação múltipla para tratamento rigoroso de dados em falta; (ii) a aplicação do modelo Bayesiano INLA–BYM2 para captar dependência espacial e heterogeneidade residual; e (iii) a comparação sistemática entre funções de ligação e cenários de análise (ACC vs. MICE), assegurando transparência e consistência metodológica. Esta combinação confere elevada credibilidade às inferências produzidas e representa um avanço relevante na utilização de dados DHS para a análise de desigualdades geográficas em saúde infantil em Moçambique.

5 CONCLUSÕES E RECOMENDAÇÕES

5.1 Conclusões

O presente estudo analisou a prevalência e a distribuição espacial da anemia em crianças dos 6 aos 59 meses em Moçambique, utilizando microdados do IDS 2022–2023 e uma abordagem metodológica integrada que combinou análise casos completos, imputação múltipla, Modelos Lineares Generalizados Mistos e modelação Bayesiana espacial via INLA–BYM2. Os resultados demonstram que a anemia infantil permanece um dos principais desafios de saúde pública no país, caracterizando-se por prevalência elevada e distribuição marcadamente heterogénea entre províncias, refletindo padrões estruturais observados à escala provincial.

Confirmou-se que a prevalência nacional ponderada permanece acima do limiar epidemiológico de 40%. As províncias da Zambézia, Nampula e Tete apresentam as maiores probabilidades de prevalência elevada, enquanto Maputo Cidade e Inhambane registam valores consistentemente baixos. Este gradiente espacial Norte–Centro–Sul reflete desigualdades socioeconómicas, ambientais e de acesso a serviços essenciais, consistentes com padrões observados noutros contextos africanos.

Os modelos GLM identificaram associações entre a anemia infantil e factores individuais e familiares. A idade apresentou uma associação negativa com o risco de anemia, enquanto o quintil de riqueza do agregado familiar e a escolaridade materna estiveram associados a menores probabilidades de ocorrência da doença. A residência urbana não manteve associação estatisticamente relevante após o ajustamento para as variáveis socioeconómicas incluídas no modelo, sugerindo que parte do seu efeito pode ser explicada pelas diferenças nas condições socioeconómicas e de acesso a recursos entre áreas urbanas e rurais.

A modelação Bayesiana espacial via INLA–BYM2 permitiu decompor o risco em componentes estruturadas (espaciais) e não estruturadas. Mesmo após o ajuste pelas covariáveis, persistiu heterogeneidade espacial significativa, particularmente nas regiões Norte e Centro. Este padrão sugere a presença de determinantes contextuais não observados — como endemicidade da malária, parasitoses, insegurança alimentar ou desigualdades no saneamento — que conti-

nuam a influenciar o risco de anemia. Os critérios de ajuste (DIC, WAIC e LPML) confirmaram a superioridade do modelo BYM2 em relação ao GLMM.

A aplicação de imputação múltipla, com 20 conjuntos imputados ($m = 20$), reforçou a robustez das estimativas: tanto os padrões espaciais como os efeitos fixos e os critérios de ajuste mantiveram-se estáveis face ao cenário *complete case*. Esta consistência indica que os valores em falta não introduziram enviesamentos sistemáticos e confirma a adequação da combinação MICE–INLA em análises epidemiológicas com dados complexos e hierárquicos.

Em síntese, a anemia infantil em Moçambique é um fenómeno simultaneamente biológico, socioeconómico e espacialmente estruturado. A abordagem Bayesiana adoptada, ao incorporar dependência espacial e incerteza amostral, fornece uma base rigorosa para a compreensão do fenómeno e para a formulação de intervenções territorialmente adaptadas.

5.2 Recomendações

Com base nos resultados apresentados e nas implicações epidemiológicas e estatísticas identificadas, recomendam-se as seguintes ações:

1. Focalização geográfica das intervenções. Dada a elevada probabilidade de prevalência superior a 40% nas províncias da Zambézia, Nampula e Tete, recomenda-se priorizar estas regiões em programas de suplementação de ferro, controlo da malária, desparasitação, fortificação alimentar e intervenções de nutrição comunitária.

2. Intervenções específicas nos primeiros 24 meses de vida. A associação negativa entre idade e anemia reforça a necessidade de intervenções nos primeiros mil dias, incluindo promoção do aleitamento materno exclusivo, introdução alimentar adequada, suplementação de micronutrientes e monitoria regular do estado nutricional.

3. Reforço da escolaridade materna e literacia nutricional. A escolaridade materna demonstrou ser um dos factores protetores mais consistentes. Recomenda-se promover programas de literacia nutricional, educação para a saúde e iniciativas de empoderamento feminino, com

enfoque em contextos rurais e de maior vulnerabilidade.

4. Integração em sistemas de vigilância nutricional. Os mapas de probabilidade de excedência constituem ferramentas úteis para a tomada de decisão. Sugere-se a sua incorporação em sistemas de monitorização do Ministério da Saúde, permitindo identificar rapidamente áreas críticas e orientar intervenções territorialmente diferenciadas.

5. Inclusão de determinantes ambientais e programáticos em estudos futuros. Indicadores como endemicidade da malária, padrões climáticos, cobertura de programas de suplementação, NDVI, precipitação e altitude deverão ser integrados em modelos futuros, permitindo compreender melhor a heterogeneidade espacial observada.

6. Desenvolvimento de modelos espaço-temporais. A análise futura deverá integrar múltiplos ciclos do DHS (2011, 2015 e 2022–23), de modo a estimar tendências temporais, identificar *hotspots* persistentes e monitorizar o impacto de intervenções ao longo do tempo.

7. Consolidação da modelação Bayesiana aplicada. Recomenda-se reforçar a formação em modelação Bayesiana, INLA e estatística espacial nas instituições de ensino e saúde moçambicanas, dada a sua utilidade comprovada na análise de dados complexos e espacialmente correlacionados.

Em conclusão, este estudo evidencia que a anemia infantil em Moçambique é multicausal e geograficamente diferenciada. A modelação Bayesiana via INLA–BYM2 demonstrou ser uma abordagem rigorosa e informativa, constituindo um contributo relevante para a estatística aplicada e para a formulação de políticas públicas baseadas em evidência, particularmente no alinhamento com estratégias nacionais de nutrição e saúde materno-infantil.

Limitação metodológica

Pressuposto de independência dos efeitos aleatórios. O modelo adoptado assume que os diferentes componentes de efeitos aleatórios são mutuamente independentes. Embora este pressuposto seja comum e amplamente aceite na literatura como uma aproximação prática, a sua validade no contexto moçambicano pode ser questionada, dado que factores socioeconómicos, ambientais e

epidemiológicos tendem a manifestar-se de forma interligada entre agregados familiares, comunidades e províncias. Neste sentido, será metodologicamente relevante explorar, em trabalhos futuros, modelos que permitam efeitos aleatórios correlacionados ou estruturas de dependência mais flexíveis, de modo a avaliar se a relaxação deste pressuposto altera as inferências espaciais obtidas.

REFERÊNCIAS

Augusto., et al. (2020). Determinants of childhood anaemia in Mozambique: Evidence from population-based surveys. *BMC Public Health*, 20(1), 1–12.

Ayele, D. G., Zewotir, T. T., & Mwambi, H. G. (2016). Prevalence and risk factors of anaemia among children aged 6–59 months in Ethiopia: A multilevel analysis. *BMC Hematology*, 16(1), 1–9.

Barry, A. S., Ngesa, O., Onyango, N. O., & Mwambi, H. (2021). Bayesian spatial modeling of anemia among children under 5 years in Guinea. *BMC Public Health*, 21(1), 1–11.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, 382(9890), 427–451.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R–INLA*. Chichester: Wiley.

Brooker, S., Hotez, P. J., & Bundy, D. A. P. (2008). Hookworm-related anaemia among pregnant women: A systematic review. *PLoS Neglected Tropical Diseases*, 2(9), e291.

Corsi, D. J., Mejía-Guevara, I., & Subramanian, S. V. (2012). Risk factors for chronic undernutrition among children in India: Estimating relative importance, population attributable risk and fractions. *Social Science & Medicine*, 75(8), 1407–1414.

Cusick, S. E., & Georgieff, M. K. (2012). Nutrient supplementation and neurodevelopment: Timing is the key. *Archives of Pediatrics & Adolescent Medicine*, 166(5), 481–482.

Custódio, E., et al. (2015). Nutritional and health determinants of anaemia among Mozambican children. *Public Health Nutrition*, 18(4), 1–10.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Boca Raton: Chapman & Hall/CRC.

Gemperli, A., Vounatsou, P., Sogoba, N., & Smith, T. (2003). Malaria mapping using transmission models: Application to survey data from Mali. *American Journal of Epidemiology*, 157(3), 230–236.

Grantham-McGregor, S., & Ani, C. (2001). A review of studies on the effect of iron deficiency on cognitive development in children. *The Journal of Nutrition*, 131(2), 649S–668S.

Harding, K. L., Aguayo, V. M., & Webb, P. (2018). Hidden hunger in South Asia: A review of recent trends and persistent challenges. *Public Health Nutrition*, 21(4), 785–795.

Instituto Nacional de Estatística (INE) & ICF. (2024). *Inquérito Demográfico e de Saúde de Moçambique 2022–2023: Relatório Final*. Maputo, Moçambique e Rockville, Maryland, EUA: INE e ICF.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer.

Kandala, N.-B., Madungu, T. P., Emina, J. B., Nzita, K. P. D., & Cappuccio, F. P. (2009). Malnutrition among children under the age of five in the Democratic Republic of Congo: Does geographic location matter? *BMC Public Health*, 9(1), 24.

Kassebaum, N. J. (2016). The global burden of anemia. *Hematology/Oncology Clinics of North America*, 30(2), 247–308.

Kejo, D., Petrucka, P. M., Martin, H., Kimanya, M. E., & Mosha, T. C. (2018). Prevalence and predictors of anemia among children under five years of age in Arusha District, Tanzania. *Pediatric Health, Medicine and Therapeutics*, 9, 9–15.

Lawson, A. B. (2018). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology* (3rd ed.). Boca Raton: Chapman & Hall/CRC.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. *Journal of the Royal Statistical Society: Series B*, 73(4), 423–498.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models* (2nd ed.). New York: Wiley.

Menéndez, C., Fleming, A. F., & Alonso, P. L. (2000). Malaria-related anaemia. *Parasitology Today*, 16(11), 469–476.

Muhajarine, N., et al. (2021). Determinants of anemia among children under five years in Mozambique: Analysis of Demographic and Health Survey data. *BMC Public Health*, 21(1), 1465.

Nambiema, A., Robert, A., & Yaya, I. (2019). Prevalence and risk factors of anemia in children aged 6–59 months in Togo: Analysis of DHS 2013–2014. *BMC Public Health*, 19(1), 215.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A*, 135(3), 370–384.

Osório, M. M., Lira, P. I. C., Batista Filho, M., & Ashworth, A. (2001). Prevalência de anemia em crianças de 6–59 meses no Estado de Pernambuco, Brasil. *Revista Panamericana de Salud*

Pública, 10(2), 101–107.

Petry, N., Olofin, I., Hurrell, R. F., Boy, E., Wirth, J. P., Moursi, M., Angel, M. D., & Rohner, F. (2016). The proportion of anemia associated with iron deficiency in low, medium, and high human development index countries. *Blood*, 128(7), 897–903.

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4), 1145–1165.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, 71(2), 319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4, 395–421.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1), 1–28.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64(4), 583–639.

Stevens, G. A., Finucane, M. M., De-Regil, L. M., Paciorek, C. J., Flaxman, S. R., Branca, F., Peña-Rosas, J. P., Bhutta, Z. A., & Ezzati, M. (2013). Global, regional, and national trends in haemoglobin concentration and prevalence of anaemia in children and women. *The Lancet Global Health*, 1(1), e16–e25.

van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399.

Woldie, H., Kebede, Y., & Tariku, A. (2015). Factors associated with anemia among children aged 6–23 months attending growth monitoring at Tsitsika Health Center, Ethiopia. *Journal of Nutrition and Metabolism*, 2015, 1–9.

World Health Organization. (2008). *Worldwide Prevalence of Anaemia 1993–2005: WHO Global Database on Anaemia*. Geneva: WHO.

World Health Organization. (2011). *Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity*. Geneva: WHO.

World Health Organization. (2024). *Anaemia*. Geneva: World Health Organization.

APÊNDICES

A.1 — Código em R

Listing 1: Código essencial dos modelos GLM e INLA–BYM2

```r
# ============================================================
# CDIGO ESSENCIAL: GLMM e INLA--BYM2
# Base: crianas 6--59 meses, desfecho binrio anemia
# ============================================================

# -------------------- 1) Dados para modelao ----------------
cc <- d %>%
  dplyr::select(
    anemia, age_m, female, urban = urban_i,
    wealth_f, edu_f, psu, strata, wt, prov
  ) %>%
  tidyr::drop_na() %>%
  dplyr::mutate(
    anemia = as.integer(anemia),
    psu = factor(psu),
    strata = factor(strata),
    wt = as.numeric(wt) / mean(as.numeric(wt), na.rm = TRUE)
  )

# -------------------- 2) GLMM logstico ----------------------
form_glmm <- anemia ~ age_m + female + urban + wealth_f + edu_f +
  (1 | psu) + (1 | strata)

fit_glmm_w <- glmmTMB::glmmTMB(
  formula = form_glmm,
  family = stats::binomial("logit"),
  data = cc,
  weights = wt
)

# -------------------- 3) Dados para INLA--BYM2 ---------------
cc_inla <- cc %>%
```

```r
  dplyr::mutate(
    prov_key = fix_alias(norm_name(as.character(prov)))
  ) %>%
  dplyr::left_join(
    match_tab %>% dplyr::select(prov_key, prov_id),
    by = "prov_key"
  ) %>%
  tidyr::drop_na(prov_id)

# -------------------- 4) Priors PC para BYM2 ------------------
bym2_hyper <- list(
  phi = list(prior = "pc", param = c(0.5, 0.5)),
  prec = list(prior = "pc.prec", param = c(1, 0.01))
)

# -------------------- 5) Modelo INLA--BYM2 --------------------
form_inla <- anemia ~ age_m + female + urban + wealth_f + edu_f +
  f(psu, model = "iid") +
  f(strata, model = "iid") +
  f(
    prov_id,
    model = "bym2",
    graph = g_inla,
    scale.model = TRUE,
    hyper = bym2_hyper
  )

fit_bym2_w <- INLA::inla(
  formula = form_inla,
  family = "binomial",
  Ntrials = rep(1, nrow(cc_inla)),
  data = cc_inla,
  weights = as.numeric(cc_inla$wt),
  control.compute = list(
    dic = TRUE,
    waic = TRUE,
    cpo = TRUE,
```

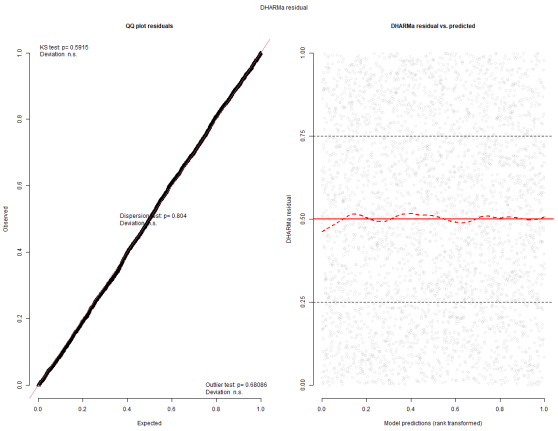
```
    config = TRUE
  ),
  control.predictor = list(
    link = 1,
    compute = TRUE
  )
)
```

A.2 — Diagnóstico e verificação dos modelos

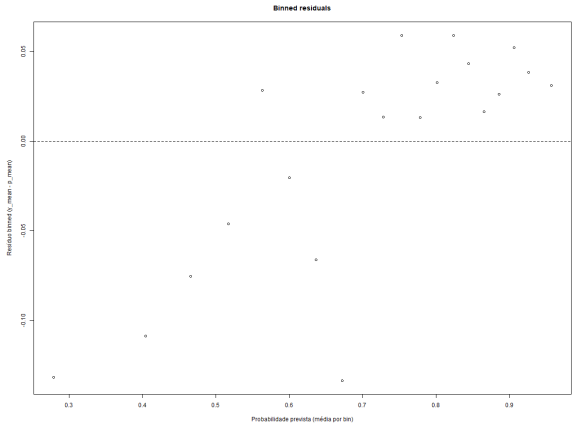
Neste capítulo apresentam-se os diagnósticos essenciais utilizados para verificar a adequação do ajuste e a robustez dos modelos (GLMM e INLA–BYM2), incluindo: (i) inspeção gráfica de resíduos (DHARMa e/ou resíduos binned), (ii) comparação entre modelos ponderados e não ponderados e (iii) autocorrelação espacial residual (Moran's I) para resíduos agregados por província.

A.2.1 — Diagnóstico do GLMM (ponderado vs não ponderado)

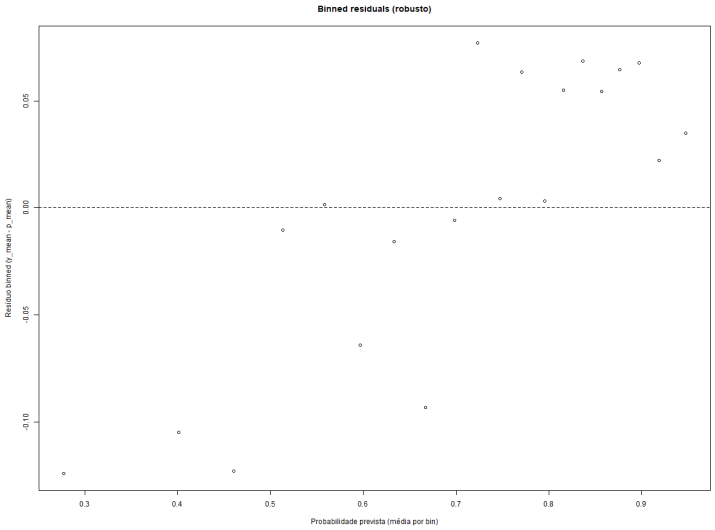
- **DHARMa geral):** `GLMM_weighted _DHARMa_overview.png` e `GLMM_unweighted _DHARMa_overview.png`.
- **Fallback robusto (resíduos binned):** `GLMM_weighted _binned_residuals.png` e `GLMM_unweighted_binned_residuals.png`.
- **Heterogeneidade (PSU):** outputs de *ICC* e *MOR* (ex.: `O3_glmm_ICC_MOR.csv`).



(a) Diagnóstico global DHARMa (GLMM não ponderado).



(b) Resíduos binned (GLMM ponderado).



(c) Resíduos binned (GLMM não ponderado).

Figura A.1: Diagnóstico do GLMM (ponderado vs não ponderado).

A.2.2 — Diagnóstico espacial do INLA (Moran's I)

Para avaliar autocorrelação espacial residual, calculou-se o Moran's I sobre resíduos provinciais definidos como:

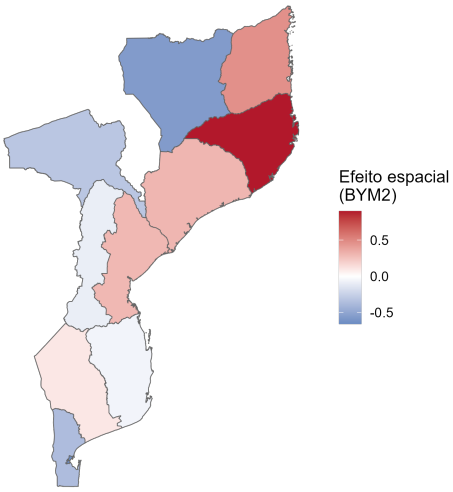
$$\text{resíduo}_j = \widehat{p}_j^{obs} - \widehat{p}_j^{post},$$

onde $\widehat{p}_j^{obs}$ é a prevalência observada por província (estimada com `survey`) e $\widehat{p}_j^{post}$ é a média posterior provincial do modelo INLA–BYM2. O resultado foi exportado para o ficheiro `INLA_morans_I_residuals.txt`.

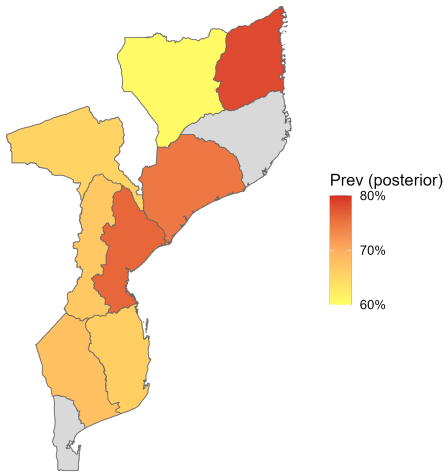
A.2.3 — Excedência e intervalos credíveis provinciais

Foram produzidos mapas e tabelas provinciais com:

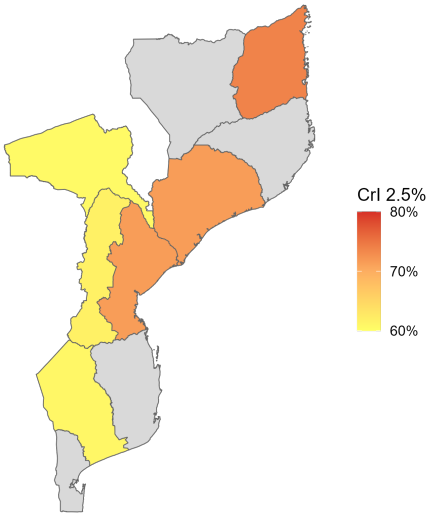
- **Excedência:** $P(\text{prev} > c)$ para múltiplos limiares c (incluindo o principal, p.ex. 40%).
- **CrI provincial:** média posterior e intervalos credíveis a 95% (2.5% e 97.5%).



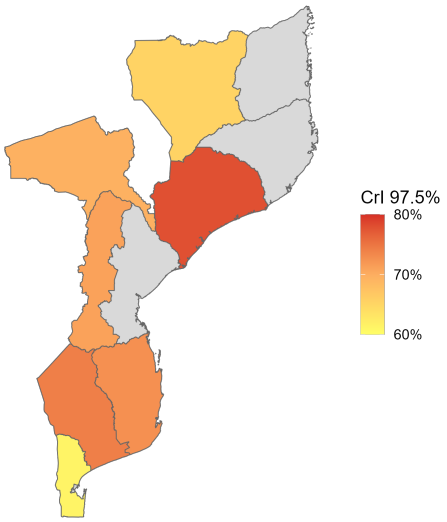
(a) Efeito espacial (média posterior).



(b) Prevalência (média posterior).



(c) Limite inferior (2.5%) do CrI 95%.



(d) Limite superior (97.5%) do CrI 95%.

Figura A.2: Resultados provinciais do INLA–BYM2 (ponderado).

A.3 — Estrutura dos outputs gerados

A execução do `script` produz tabelas e figuras para a Secção de Resultados e para os Apêndices. Principais ficheiros:

Tabela A.1: Estrutura dos outputs gerados pelo script.

Ficheiro	Descrição
`O1_nacional_resumo_geral.csv`	Estatísticas descritivas nacionais (amostra ponderada).
`O1_fig_hist_idade.png`	Histograma ponderado da idade (6–59 meses).
`map_O1_prev_prov.png`	Mapa de prevalência ponderada por província (O1).
`map_O1_prev_SE.png`	Mapa do erro-padrão da prevalência.
`map_O1_prev_CIwidth.png`	Mapa da largura do IC95% da prevalência.
`O3_glmm_weighted_coef.csv`	Coeficientes GLMM ponderado (efeitos fixos).
`O3_glmm_unweighted_coef.csv`	Coeficientes GLMM não ponderado (efeitos fixos).
`GLMM_weighted_DHARMa_overview.png`	Diagnóstico DHARMa (ponderado).
`GLMM_unweighted_DHARMa_overview.png`	Diagnóstico DHARMa (não ponderado).
`GLMM_weighted_binned_residuals.png`	Resíduos binned (ponderado).
`GLMM_unweighted_binned_residuals.png`	Resíduos binned (não ponderado).
`O4_fit_bym2_weighted.rds`	Objeto INLA (BYM2 ponderado).
`O4_fit_bym2_unweighted.rds`	Objeto INLA (BYM2 não ponderado).
`O4_inla_weighted_vs_unweighted_DIC_WAIC_LPML.csv`	Comparação INLA ponderado vs não ponderado.
`map_O4_bym2_spatial_effect_mean.png`	Mapa do efeito espacial (média).
`O5_inla_prev_prov_posterior_CrI.csv`	Média posterior e CrI 95% por província.
`map_O5_inla_prev_posterior_mean.png`	Mapa da média posterior da prevalência.
`map_O5_inla_prev_posterior_lcl.png`	Mapa do CrI inferior (2.5%).
`map_O5_inla_prev_posterior_ucl.png`	Mapa do CrI superior (97.5%).

Continuação na página seguinte

Ficheiro	Descrição
`05_inla_prov_exceed_thr40.csv`	Excedência provincial (limiar 40%).

A.4 — Tabelas

Tabela A.4.1 — Probabilidade predita de anemia por idade (GLMM – logit)

Tabela A.2: Probabilidade predita de anemia por idade (GLMM – logit)

Idade (meses)	Prob. predita	Erro-padrão	IC_inf	IC_sup
0	0.450675	0.084125	0.410276	0.491735
1	0.446415	0.083220	0.406543	0.486989
59	0.229001	0.081983	0.201871	0.258597

Tabela A.4.2 — Efeitos fixos do modelo BYM2 (ACC)

Tabela A.3: Efeitos fixos do modelo INLA–BYM2 (complete case, ponderado).

Termo	OR	OR_L	OR_U
(Intercepto)	0.8058	0.6589	0.9841
age_m	0.9834	0.9804	0.9865
female	0.9654	0.8778	1.0616
urban	0.9247	0.7776	1.0997
wealth_fpoorer	1.0308	0.8980	1.1834
wealth_fmiddle	0.8895	0.7637	1.0360
wealth_fricher	0.8140	0.6740	0.9831
wealth_frichest	0.7141	0.5509	0.9262
edu_fprimario	0.9184	0.8203	1.0282
edu_fsecundario	0.9351	0.7876	1.1104
edu_fsuperior	0.5856	0.3371	1.0172

Tabela A.4.3 — Comparação dos OR entre ACC e MICE

Tabela A.4: Comparação de OR (ACC vs MICE pooled).

Termo	OR_acc	ORL_acc	ORU_acc	OR_mice	ORL_mice	ORU_mice
(Intercepto)	0.8058	0.6589	0.9841	0.00048	2.0e-09	1.7798
age_m	0.9834	0.9804	0.9865	1.0067	0.9121	1.1109
edu_fprimario	0.9184	0.8203	1.0282	1.0456	0.0272	39.5659
edu_fsecundario	0.9351	0.7876	1.1104	0.9967	0.0059	168.8308
edu_fsuperior	0.5856	0.3371	1.0172	0.2823	1.65e-09	19271331.9
female	0.9654	0.8778	1.0616	0.8476	0.0451	16.8971
urban	0.9247	0.7776	1.0997	0.9598	0.0127	76.6893
wealth_fmiddle	0.8895	0.7637	1.0361	0.6933	0.0061	78.8169
wealth_fpoorer	1.0308	0.8980	1.1834	0.7121	0.0060	82.2497
wealth_fricher	0.8140	0.6740	0.9831	0.6218	0.0027	139.3004
wealth_frichest	0.7141	0.5509	0.9262	0.7052	6.1e-04	739.8162

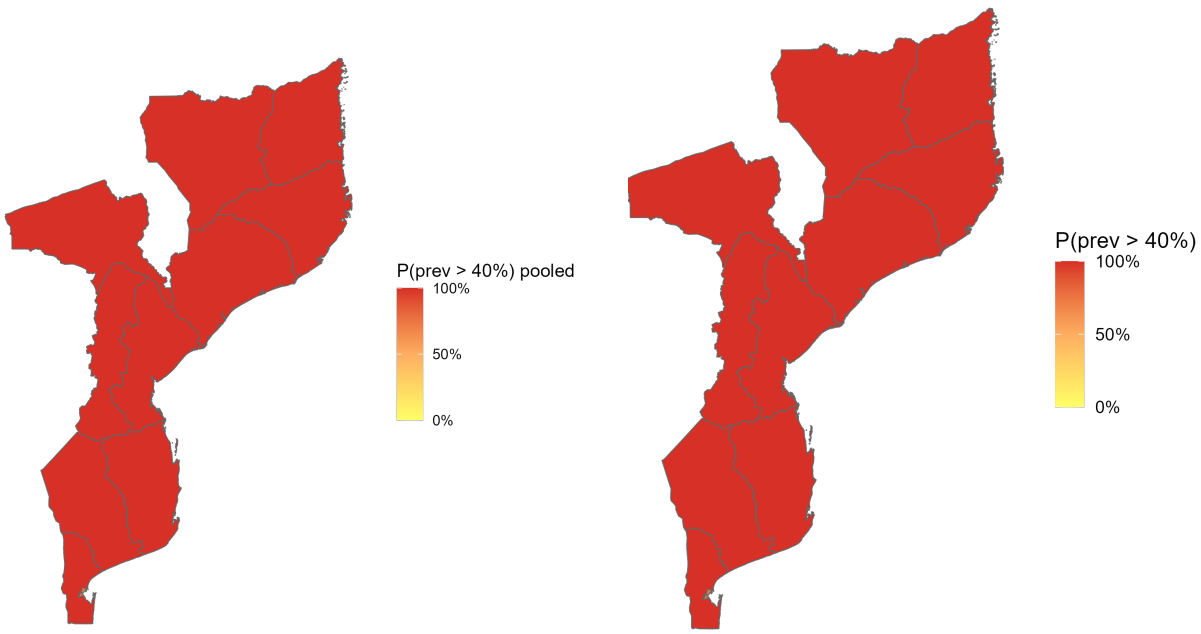
Nota: Os valores MICE são aproximações pooled e podem apresentar amplitudes elevadas.

Tabela A.4.4 — Critérios de ajuste dos modelos BYM2 por imputação (MICE, $m = 20$)

Tabela A.5: Critérios de ajuste INLA–BYM2 por imputação (MICE, 1–20).

Imputação	DIC	WAIC	LPML
1	709.0593	1114.2390	-20150.8904
2	688.5292	1110.9350	-20153.5184
3	646.7624	1035.8347	-20024.7765
4	712.6345	1129.7839	-20162.6139
5	693.2254	1109.9465	-20148.8750
20	*(colar)*	*(colar)*	*(colar)*

A.5 — Figuras essenciais (mapas e MICE)



(a) Excedência pooled (MICE, $m = 20$). **(b)** $P(\text{prev} > 40\%)$ (INLA–BYM2).

Figura A.3: Mapas de excedência provincial.

A.6 — Dados e fontes

- **DHS Moçambique 2022–2023** (ficheiro KR: `MZKR81FL.DTA`); acesso conforme políticas DHS.
- **Shapefile ADM2/ADM1 (INE)**: `moz_admbnda_adm2_ine_20190607.shp`.
- Pesos amostrais (`v005`) e desenho amostral (`v021`, `v022`) tratados via `survey`.