

Trabalho de Licenciatura em Estatística

**Análise dos Factores Associados ao Estado de
Vacinação em Crianças Menores de Dois Anos em
Moçambique**

Autora: Silvânia Victor Vilanculo

Maputo, Junho de 2025



Trabalho de Licenciatura em Estatística

**Análise dos Factores Associados ao Estado de
Vacinação em Crianças Menores de Dois Anos em
Moçambique**

Autora: Silvânia Victor Vilanculo

Supervisor: Miranda Albino Martins Muualo, PhD, UFRJ.

Maputo, Junho de 2025

Dedicatória

Dedico este trabalho à minha linda irmã **Shantel Victor Vilanculo**.

Declaração de Honra

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

Silvânia Victor Vilanculo

Maputo, Junho de 2025

Agradecimentos

Em primeiro lugar, agradeço à Deus pela dádiva da vida e por ter-me permitido chegar até aqui com saúde.

Expresso minha profunda gratidão aos meus pais, **Victor Jaime Vilanculo** (em memória) e **Alice Elvira Chachine**, por nunca terem desistido de mim, mesmo nos momentos mais difíceis da minha infância. Agradeço, com carinho, aos meus tios **Rosa Sambo** e **Timóteo Sambo**, pelo acompanhamento diário, apoio incondicional e pelas valiosas lições partilhadas ao longo da minha formação.

Ao meu supervisor, Doutor **Miranda Albino Martins Muaualo**, registro os meus mais sinceros agradecimentos pela disponibilidade, orientação e dedicação demonstradas durante todo o desenvolvimento deste trabalho.

Agradeço a todos os docentes do **Departamento de Matemática e Informática (DMI)**, com destaque especial aos docentes do Curso de Estatística, pelos conhecimentos transmitidos e conselhos que marcaram minha jornada académica.

Àqueles que caminharam ao meu lado, os meus colegas e eternos amigos **Alberto Gonçalves Vilanculos**, **Edvilce Jéssica Brás Combo** e **Shelsea Luís Damião**, deixo um agradecimento especial por tornarem esta caminhada mais leve, motivadora e repleta de boas memórias.

Por fim, deixo o meu sincero reconhecimento a toda a minha família, à minha eterna professora **Eng.^a Fátima Massicame**, aos funcionários da UEM, aos meus irmãos na fé, e a todos que, directa ou indirectamente, contribuíram para minha formação.

Muito obrigada!

Resumo

A vacinação constitui a primeira linha de defesa contra vários tipos de doenças infecciosas. Uma criança não vacinada pode enfrentar graves enfermidades, incluindo doenças potencialmente não tratáveis que podem ser mortais. O objectivo desta pesquisa é analisar os factores associados ao estado de vacinação em crianças menores de 2 anos em Moçambique. Foram usados dados do Inquérito Demográfico e de Saúde (2022-23). Para identificar associação das variáveis independentes e dependentes, aplicou-se a análise bivariada, por meio do teste Qui-quadrado e, para verificar a relação entre o estado de vacinação e as variáveis independentes, foi aplicada a regressão logística múltipla. O modelo apresentou bom ajuste segundo os testes de Wald e Hosmer-Lemeshow, e excelente capacidade preditiva ($AUC = 0,8225$), com acurácia de 80%. As variáveis que se mostraram estatisticamente significativas no modelo final foram: área de residência, número de visitas pré-natais, escolaridade da mãe (secundário ou mais), ocupação materna, quintil de riqueza (rico), local de parto (hospital público), religião do agregado familiar (Zion), idade da mãe e exposição à mídia. Estes resultados reforçam a importância de factores sociodemográficos, económicos e de acesso aos serviços de saúde na determinação da cobertura vacinal infantil em Moçambique.

Palavras-chave: Factores associados ao estado vacinal das crianças; Vacinação; Regressão logística.

Lista de Abreviaturas

- AIC - Critério de informação de Akaike
- AUC - Area Under the Curv
- BIC - Critério de informação Bayesiano
- DHS - Demographic and Health Survey
- I.C. - Intervalo de confiança
- IDS - Inquérito Demográfico de Saúde
- IFC - Inner City Fund
- INE - Instituto Nacional de Estatística
- MISAU - Ministério de saúde
- ONU - Organização das Nações Unidas
- OR - Odds Ratio
- OMS - Organização Mundial da Saúde
- ROC - Receiver Operating Characteristic
- UNICEF - Fundo das Nações Unidas para a Infância

Índice

Dedicatória	i
Declaração de Honra	ii
Agradecimentos	iii
Resumo	iv
Lista de Abreviaturas	v
Lista de Figuras	viii
Lista de Tabelas	ix
1 Introdução	1
1.1 Contextualização	1
1.2 Justificação	2
1.3 Problematização	3
1.4 Objectivos	3
1.4.1 Objectivo geral	3
1.4.2 Objectivos específicos	3
1.5 Hipóteses	4
1.6 Estrutura do trabalho	4
2 Revisão da literatura	5
2.1 Estado vacinal das crianças	5
2.2 Principais causas do estado vacinal da crianças	6
2.2.1 Factores associados às características do agregado familiar	7
2.2.2 Factores associados às características das mães das crianças	8
2.3 Estudos relacionados sobre vacinação de crianças	8
2.4 Regressão logística	11
2.4.1 Histórico da regressão logística	11
2.4.2 Definição da Regressão logística	11
2.4.3 Propriedades da regressão logística	12
2.5 Regressão logística simples	13
2.5.1 Transformação logit	14
2.5.2 Estimação dos parâmetros do modelo de regressão logística simples	15
2.5.3 Regressão logística múltipla	16
2.5.4 Estimação dos parâmetros na regressão logística múltipla	18
2.5.5 Seleção das variáveis independentes	24

2.5.6	Procedimento de seleção de variáveis	24
2.5.7	Critérios para seleção do melhor modelo	25
2.5.8	Curva Roc	27
3	Material e métodos	29
3.1	Classificação do estudo	29
3.2	Material	30
3.2.1	Fonte de dados	30
3.2.2	Descrição das variáveis independentes	31
3.3	Métodos	32
3.3.1	Análise descritiva	32
3.3.2	Análise bivariada	33
4	Resultados e discussão	36
4.1	Resultados	36
4.1.1	Tabelas cruzadas	38
4.1.2	Visualização gráfica	40
4.1.3	Teste de qui quadrado	43
4.1.4	Modelo logístico inicial	44
4.1.5	Modelo logístico com variáveis significativas no teste Qui-quadrado . .	45
4.1.6	Modelo final	46
4.1.7	Ajuste do modelo final	48
4.2	Discussão	51
5	Conclusão e recomendações	54
5.1	Conclusão	54
5.2	Recomendações	54

Lista de Figuras

2.1	Curva ROC.	28
4.1	Proporção do estado vacinal por sexo da criança	40
4.2	Proporção do estado vacinal por amamentação da criança	40
4.3	Proporção do estado vacinal pelo intervalo de nascimento	40
4.4	Proporção do estado vacinal pelo acesso aos documentos de vacinação	40
4.5	Proporção do estado vacinal por Estado civil	41
4.6	Proporção do estado vacinal por nível de escolaridade	41
4.7	Proporção do estado vacinal por sexo do agregado familiar	41
4.8	Proporção do estado vacinal pela ocupação do chefe do agregado	41
4.9	Proporção do estado vacinal pela idade da mãe	42
4.10	Proporção do estado vacinal pela ocupação da mãe	42
4.11	Proporção do estado vacinal pelo acesso à mídia	42
4.12	Proporção do estado vacinal pela religião	42
4.13	Curva ROC	49

Lista de Tabelas

3.1	Características da Criança	31
3.2	Características da mãe	31
3.3	Características do agregado familiar	32
3.4	Avaliação da curva roc	35
4.1	Descrição da amostra	37
4.2	Distribuição percentual da vacinação segundo variáveis sociodemográficas . . .	39
4.3	Associação entre estado vacinal e variáveis explicativas	43
4.4	Modelo logístico ponderado: factores associados à vacinação completa	44
4.5	Modelo logístico com variáveis significativas no teste Qui-quadrado	45
4.6	Modelo logístico ponderado com variáveis combinadas (teste Qui-quadrado e modelo anterior)	46
4.7	Razões de chances (OR) e intervalo de confiança (IC 95%) para o modelo logístico ponderado	47
4.8	Teste de multicolinearidade	48
4.9	Teste de Wald	48
4.10	Teste de Hosmer e Lemeshow para ajuste do modelo	49
4.11	Matriz de confusão do modelo logístico ponderado	50
4.12	Métricas de desempenho do modelo logístico ponderado	50
4.13	Comparação dos valores de Pseudo R ² entre os modelos logísticos	50

Capítulo 1

Introdução

1.1 Contextualização

O Fundo das Nações Unidas para a Infância (UNICEF, 2023a) aponta que a imunização desempenha um papel fundamental na prevenção de doenças infecciosas e na redução da mortalidade infantil. Segundo a Organização Mundial da Saúde (OMS, 2023) estima-se que, na década de 2010, mais de 20 milhões de bebês não completaram o esquema básico de vacinação no mundo a cada ano e, desses, mais de 13 milhões não receberam qualquer vacina por meio dos programas de imunização. Desse modo a OMS (2023) estabeleceu objectivos de fornecer serviços de vacinação acessíveis a todas as pessoas em todos os lugares com o intuito de salvar mais de 50 milhões de vidas até 2030.

Segundo a OMS (2020), a Assembleia Mundial da Saúde, em colaboração com regiões e parceiros, (África, Europa, América, Sudeste Asiático, Pacífico Oeste, Mediterrâneo Oriental), adaptou a Agenda de Imunização 2030 (IA2030), uma estratégia global para a próxima década, com o objectivo de enfrentar desafios na imunização e salvar mais de 50 milhões de vidas. A IA2030 visa garantir que pessoas de todas as idades, em qualquer lugar do mundo, tenham acesso pleno às vacinas, promovendo a saúde e o bem-estar global.

A implementação da IA2030 enfrenta desafios como desigualdade no acesso às vacinas, surtos de doenças evitáveis, hesitação vacinal e impactos das mudanças climáticas na saúde pública. Para superá-los, a estratégia enfatiza o uso de dados para orientar decisões, a colaboração entre governos, sector privado e organizações internacionais, além da integração da imunização com outros serviços de saúde para ampliar a cobertura e otimizar recursos. A IA2030 representa um compromisso global para garantir que ninguém fique para trás, reforçando a vacinação como um pilar essencial para a saúde pública e o desenvolvimento sustentável.

Globalmente, o número de crianças que receberam três doses da vacina contra difteria, tétano e coqueluche em 2023 manteve-se em 84% (108 milhões). No entanto, o número de crianças que não recebeu uma única dose da vacina aumentou de 13,9 milhões em 2022 para 14,5 milhões em 2023 (UNICEF,2023a).

A OMS (2022), afirmou que a região Africana está a enfrentar o reaparecimento de certas doenças evitáveis pela vacinação, como o sarampo e a poliomielite, devido a uma diminuição da cobertura vacinal. A cobertura de vacinas na África subsaariana manteve-se ao nível de 72% durante o período de 2019 à 2023. A situação expõe as populações da região a doenças e surtos que podem ser prevenidos pela imunização (ONU, 2019).

De acordo com o UNICEF (2023a) a África é a região com o maior número de crianças não vacinadas ou subvacinas. Entre 2019 e 2021, 12,7 milhões de crianças na região foram identificadas em situação de vacinação incompleta, com 8,7 milhões delas nunca tendo recebido qualquer vacina, sendo classificadas como "zero-dose". Além disso, a taxa de vacinação da primeira dose e terceira foi de 84% e 76% na África, enquanto na Europa foi de 97% e 94%, respectivamente.

Em Moçambique quase 40% de crianças menores de dois anos ainda carecem de serviços completos de vacinação (UNICEF, 2023a). A falta desses serviços é influenciada pelo acesso limitado aos serviços de saúde, especialmente em áreas rurais, pela falta de conhecimento da necessidade de vacinação, falta de vacinas nas unidades sanitárias (INE e ICF, 2023). Neste contexto o presente trabalho pretende estudar os factores associados ao estado de vacinação em crianças menores de 2 anos em Moçambique.

1.2 Justificação

A vacinação é uma das intervenções eficazes e custo-efetivas na saúde pública, desempenhando um papel central na redução da mortalidade infantil, especialmente em países em desenvolvimento como Moçambique. No entanto, a cobertura vacinal ainda permanece desigual em várias regiões do país, refletindo determinantes socioculturais, geográficos e de acesso aos serviços de saúde. Compreender os factores associados ao estado vacinal de crianças menores de dois anos é essencial não apenas para fortalecer as políticas públicas e programas de saúde, mas também para contribuir com o avanço do conhecimento científico nesta área.

Este estudo justifica-se, portanto, por sua relevância prática e académica. Para os investigadores e profissionais da área da saúde e das ciências sociais, os resultados oferecem evidências empíricas que podem sustentar intervenções mais direcionadas e eficazes. Para estudantes de Estatística e áreas afins, o trabalho representa uma aplicação real de métodos quantitativos, em especial a regressão logística na análise de dados demográficos e de saúde pública, contribuindo para sua formação técnica e crítica.

Além disso, ao abordar as desigualdades regionais, sociais e maternas que influenciam o acesso à vacinação, este estudo ajuda a revelar realidades que muitas vezes passam despercebidas. Quando uma criança não é vacinada, ela fica vulnerável a doenças evitáveis e torna sua infância menos saudável. Ao abordar este tema, o trabalho procura não só contribuir para melhorias nos

serviços de saúde, mas também despertar o olhar crítico da sociedade. É um convite para que todos os investigadores, estudantes, profissionais e cidadãos se envolvam na construção de um país onde o direito à saúde seja uma realidade para todos, sem distinção.

1.3 Problematização

Segundo o INE e ICF (2023), a imunização universal das crianças contra as doenças comuns preveníveis por vacinas é fundamental para reduzir a taxa de incidência de doenças e taxas de mortalidade neonatal e infantil. A nível mundial, Moçambique ocupa o 12º maior número absoluto de crianças com dose zero (UNICEF, 2023a). De acordo com INE e ICF (2023), a percentagem de crianças de 12–23 meses que receberam todas as vacinas básicas aumentou consideravelmente de 47% em 1997 para 66% em 2015, e posteriormente, diminuiu para 38% em 2023, que corresponde a uma redução de 28%. Neste contexto, levanta-se a seguinte questão de pesquisa: Quais são os factores associados ao estado de vacinação em crianças menores de 2 anos em Moçambique?

1.4 Objectivos

1.4.1 Objectivo geral

Analisar os factores associados ao estado de vacinação em crianças menores de 2 anos em Moçambique.

1.4.2 Objectivos específicos

- Descrever o perfil sociodemográfico das crianças menores de 2 anos em Moçambique;
- Identificar os factores associados ao estado de vacinação em crianças menores de 2 anos;
- Estimar os coeficientes de um modelo de regressão logística que descreve a relação entre o estado de vacinação e os factores associados.

1.5 Hipóteses

Como resposta antecipada ao trabalho, foi levantada a seguinte hipótese:

H_0 : Os factores socioeconômicos, culturais e geográficos não têm influência na adesão ao calendário vacinal de crianças menores de 2 anos em Moçambique.

H_1 : Os factores socioeconômicos, culturais e geográficos têm influência na adesão ao calendário vacinal de crianças menores de 2 anos em Moçambique.

1.6 Estrutura do trabalho

O presente trabalho está estruturado em cinco capítulos. O primeiro capítulo apresenta a introdução, onde se enquadra o problema do estudo, os objectivos do estudo, a justificativa e a relevância. O segundo capítulo apresenta a revisão da literatura, aborda os principais conceitos teóricos relacionados à vacinação infantil, bem como conceitos relacionados ao modelo de regressão logística. O terceiro capítulo descreve os materiais e métodos utilizados na pesquisa, incluindo a fonte de dados, e as variáveis consideradas. O quarto capítulo apresenta e discute os resultados obtidos a partir da análise dos dados. Por fim, o quinto capítulo traz as conclusões do estudo, recomendações para pesquisas futuras e as limitações do estudo.

Capítulo 2

Revisão da literatura

2.1 Estado vacinal das crianças

A vacinação constitui a primeira linha de defesa contra vários tipos de doenças infecciosas. Uma criança não vacinada pode enfrentar graves enfermidades, incluindo doenças potencialmente não tratáveis que podem ser mortais (UNICEF, 2009).

A vacinação é uma história de sucesso global de saúde e desenvolvimento UNICEF (2023b), pode ser considerada um sucesso porque, nos últimos dez anos, os programas de imunização salvaram a vida de mais de um milhão de crianças (Touray *et al.*, 2021). Actualmente, 3,5 a 5 milhões de mortes anualmente são evitadas por imunizações básicas contra crianças (BCG, pentavalente, poliomielite e sarampo) de doenças como difteria, tétano, coqueluche, gripe e sarampo (UNICEF, 2023a).

Segundo a OMS, as crianças devem receber durante o seu nascimento, uma dose de vacina contra a poliomielite e uma dose de vacina BCG (Bacillus Calmette Guérin) contra a tuberculose. Ao completar 12 meses, as crianças já devem ter recebido três doses de vacina contra poliomielite e de DTP (Difteria, Tétano e Pertússis) a pentavalente contra difteria, tétano, tosse convulsa, hepatite B e Haemophilus influenza e uma dose de sarampo (OMS, 2023).

Globalmente, estima-se que mais de 50% da mortalidade em crianças em idade pré-escolar se deva a doenças que são evitáveis e tratáveis por meio de intervenções simples e acessíveis, como uma vacina. Em 2016, a Organização Mundial da Saúde (OMS) estimou que 5,6 milhões de crianças menores de cinco anos morreram, traduzindo-se em 15.000 mortes todos os dias. A vacinação provou ser uma intervenção eficaz para a prevenção e eliminação de doenças infecciosas potencialmente fatais e estima-se que previna aproximadamente 2,5 milhões de mortes anualmente.

A África é a região com a maior contagem de crianças não vacinadas. Somente em 2021, 12,7 milhões de crianças se enquadraram nessa categoria, incluindo 8,7 milhões que não receberam nem mesmo uma única dose de vacinação, comumente chamadas de crianças de "dose zero".

Notavelmente, metade dos 20 principais países do mundo com o maior número de crianças com dose zero estão localizados na África. A Nigéria e a Etiópia têm precedência entre essas nações, respondendo por mais de 2,2 milhões e 1,1 milhão de crianças com dose zero, respectivamente, no continente (UNICEF, 2023a).

2.2 Principais causas do estado vacinal da crianças

O estado de imunização infantil é influenciado por diversos factores, como características da criança, condições obstétricas, factores socioeconômicos, demográficos e obstáculos geográficos.

De acordo com INE e MISAU (2017), entre 1997 e 2003 houve um aumento na cobertura vacinal para ambos os sexos, sendo a diferença entre eles mínima nesse período. No entanto, em 2015 observou-se um aumento na disparidade, com um rácio de 0,94, comparativamente aos anos anteriores, nos quais os rácios variavam entre 0,99 e 1,02 (de 1997 a 2011).

A cobertura vacinal completa também variou significativamente segundo a área de residência. Na zona rural, houve um aumento contínuo de 1997 a 2015, enquanto na zona urbana foi observada uma tendência de diminuição durante o mesmo período. Apesar disso, a zona urbana manteve as maiores taxas de cobertura vacinal. Em 1997, a diferença entre as zonas urbana e rural era bastante acentuada (rácio = 0,42), mas nos anos subsequentes essa disparidade foi reduzida, com rácios de 0,69 em 2003, 0,80 em 2011 e 0,79 em 2015, (INE e MISAU, 2017).

Em relação à escolaridade materna, constatou-se maior cobertura vacinal entre crianças cujas mães possuíam nível de escolaridade secundário, enquanto filhos de mães sem escolaridade apresentaram as menores coberturas ao longo dos anos. Em 1997, a diferença entre esses grupos era marcante (rácio = 0,34). Houve, contudo, uma melhora progressiva, com rácios de 0,50 em 2003, 0,76 em 2011 e 0,62 em 2015, (INE e MISAU, 2017).

A cobertura vacinal completa também se mostrou associada ao nível econômico materno. Crianças cujas mães pertencem ao quintil de riqueza mais elevado apresentaram, consistentemente, maiores coberturas vacinais em comparação àquelas pertencentes ao quintil mais pobre. Em 1997, a diferença entre os dois extremos era significativa (rácio = 0,23). Essa disparidade foi reduzida até 2011 (rácio = 0,71), mas entre 2011 e 2015 observou-se um leve aumento na diferença, sendo que em 2015 o quintil mais pobre continuava apresentando as menores taxas de cobertura vacinal, (INE e MISAU, 2017).

O estudo realizado sobre o estado de imunização infantil de acordo com o número de irmãos e a ordem de nascimento em 85 países de baixa e média renda mostra que a ordem de nascimento é o preditor mais significativo do status de imunização infantil (Santos e Silva, 2025). Um estudo realizado na região somali do leste da Etiópia examinou a cobertura vacinal completa e seus factores associados entre 612 crianças de 12 a 23 meses. Os pesquisadores utilizaram um

modelo de regressão logística binária para analisar os dados. Os resultados indicaram que vários factores foram significativamente associados à imunização completa. Dos quais o local de residência, estado civil, nível de escolaridade materna, renda familiar média, sexo da criança, local de parto e tempo de viagem para a unidade de saúde mais próxima (Yadita e Ayehubizu, 2021).

Na pesquisa realizada por Tesema *et al.* (2020), com base nos DHSs realizados de 2012 a 2018 de 12 países da África Oriental (Burundi, Etiópia, Comores, Uganda, Ruanda, Tanzânia, Moçambique, Madagascar, Zimbábue, Quênia, Zâmbia e Malawi), foi realizada análise de dados secundários. O estudo incluiu uma amostra total ponderada de 18.811 crianças com idades entre 12 e 23 meses. O modelo de regressão logística binária multinível revelou que a vacinação infantil foi significativamente associada à idade materna, escolaridade materna, educação do marido, exposição à mídia, número de visitas de pré-natais, visitas pós-natais, local de parto, tamanho da criança ao nascer, índice de riqueza, país e pobreza da comunidade.

Os factores associados ao estado de vacinação em crianças podem ser estudados sob o ponto de vista das características do agregado familiar em que a criança está inserida, características da mãe da criança, ou seja, factores materno-infantis, e também as características das crianças. Estes factores incluem questões socioeconômicas e demográficas. Neste trabalho, foram analisadas variáveis como escolaridade materna, renda familiar, ocupação dos pais e acesso aos serviços de saúde, entre outras.

2.2.1 Factores associados às características do agregado familiar

De acordo com a Tauil *et al.* (2016), a prevalência de crianças com vacinas incompletas está associada a escolaridade dos pais (alta ou baixa escolaridade), quintil de riqueza do agregado familiar, falta de informação a respeito do papel das vacinas na prevenção de doenças, medo das reações adversas, crença de que a vacinação não é benéfica ou causa danos, falta de motivação para vacinar os filhos, pressão social ou cultural contra a vacinação.

Para Bedford e Lansley (2007), o aumento do número de vacinas no calendário, juntamente com o uso de vacinas combinadas, leva alguns pais a questionar a segurança de dar no mesmo instante múltiplas vacinas às crianças.

Segundo Arter (2024), factores de nível comunitário, como nível de educação materna dentro das comunidades, índice de riqueza no nível comunitário, exposição à mídia da comunidade e sexo do chefe do agregado familiar são factores que estão significativamente associados ao estado de vacinação das crianças.

Alguns factores relacionados aos serviços de saúde abrangem o acesso e a distância da residência aos serviços de vacinação, oportunidades perdidas de vacinação, fornecimento inadequado das vacinas, conhecimento profissional de saúde a respeito das vacinas, o nascimento da criança em casa, uso de estabelecimentos privados para a vacinação das crianças, (Dummer *et*

al., 2012). Alguns estudos mostram que crianças apresentam baixa adesão aos programas de vacinação quando os pais vivem em zonas rurais, (Phukan *et al.*, 2008). Embora as zonas urbanas geralmente apresentem melhores níveis médios de cobertura vacinal, crianças residentes em áreas urbanas pobres podem ter cobertura tão baixa quanto as crianças em zonas rurais. Um estudo com dados de 97 países demonstrou que a prevalência de crianças com zero-dose foi de 12,6% entre os grupos urbanos pobres, enquanto nas zonas rurais foi de 14,7%; entre os urbanos não pobres, a taxa caiu para 6,5% (Restrepo-Mendez *et al.*, 2016).

2.2.2 Factores associados às características das mães das crianças

De acordo com Tertuliano e Stein (2025), a baixa escolaridade materna, falta de cuidados no pré-natal e o desmame precoce estão entre as causas mais comuns do estado vacinal da criança. Aliado à falta de conhecimentos básicos de higiene, a situação de desemprego ou subemprego e ganho insuficiente de peso na gestação.

Segundo Tauil *et al.* (2016), mães solteiras, jovens, e com intervalos de nascimento menores, pertencentes a grupos de minorias étnicas ou religiosas, desempregadas estão associadas à vacinação incompleta das crianças.

Outras pesquisas indicam que o nível de escolaridade das mães e o local de residência apresentam associações significativas com o estado de imunização infantil. Além disso, factores como ordem de nascimento, intervalo entre nascimentos e características do parto (por exemplo, local e tipo de assistência) também influenciam a cobertura vacinal. Estudos adicionais destacam que o acesso a cuidados pré-natais e pós-natais, a posse de documentação de vacinação e condições socioeconômicas, como o índice de riqueza familiar, são determinantes cruciais. No contexto comunitário, aspectos como o nível de educação das mães na comunidade, a exposição a mídias e factores económicos desempenham papéis relevantes. De acordo com Yokokura *et al.* (2013), crianças do sexo feminino, prematuras, e com o número de visitas pré-natais insuficientes tem mais chances de apresentar vacinação incompleta.

2.3 Estudos relacionados sobre vacinação de crianças

Um estudo conduzido no Quênia que utilizou dados do DHS de 2014 para analisar as proporções de crianças de 12 a 23 meses classificadas como não vacinadas, subvacinadas e totalmente vacinadas onde a análise empregou um modelo de regressão logística multinomial, mostrou que 61,9% das crianças estavam totalmente vacinadas, 28,7% subvacinadas e 9,4% não vacinadas. Essa baixa taxa de cobertura vacinal, associada a uma alta proporção de sub e não vacinação, foi atribuída a três factores principais: área de residência, índice de riqueza familiar e local de parto, (Shenton *et al.*, 2018).

Outro estudo, realizado na Etiópia por Sako *et al.* (2023), utilizou dados do DHS Etíope com uma amostra composta por 1.008 mães de crianças de 12 a 23 meses, cujo o objectivo era avaliar a prevalência de cobertura vacinal infantil e identificar seus principais determinantes. Os

resultados destacaram diversos factores que influenciam as taxas de imunização infantil. Entre os determinantes significativos, incluem-se:

- **Escolaridade Materna:** A educação das mães mostrou-se um fator preditivo importante para a cobertura vacinal.
- **Idade e Estado Civil Materno:** Mães mais jovens ou solteiras apresentaram menor probabilidade de assegurar a vacinação completa de seus filhos.
- **Documentação de Vacinação:** A presença de uma carteira de vacinação foi associada a melhores taxas de imunização.
- **Recebimento de Suplementação de Vitamina A:** Mães que receberam suplementação de vitamina A durante a gestação estavam mais propensas a vacinar seus filhos.

Ainda no mesmo estudo realizado na Etiópia, autores destacaram importantes limitações metodológicas. Embora tenham utilizado os dados secundários mais recentes disponíveis a nível nacional e aplicado um modelo de probabilidades proporcionais multinível para lidar com a estrutura hierárquica dos dados, o estudo não incluiu variáveis psicossociais dos pais, indicadores de hesitação vacinal e outras variáveis relevantes. Além disso, os autores reconhecem que, por se tratar de um estudo transversal, não foi possível estabelecer relações de causalidade entre os factores analisados e o estado vacinal das crianças. Outra limitação apontada foi o viés de recordação, considerado um dos principais desafios do estudo, já que parte das informações dependia da memória dos respondentes. Por sua vez, Massango (2018) apontou que vários indicadores apresentam diferenças consideráveis quanto ao local de residência (urbana e rural), estando quase todos em maior desvantagem para zonas rurais. Apesar do acesso aos serviços de saúde ter melhorado muito, o acesso à água potável e às condições de saneamento ainda são muito preocupantes, (Sequeira e Ndzava, 2010).

Em um estudo conduzido por Peretti-Watel *et al.* (2020), utilizando dados transversais da Pesquisa Demográfica e de Saúde do Senegal, foram analisados os factores que influenciam a imunização infantil em uma amostra de 1.143 crianças de 12 a 23 meses. A análise foi realizada por meio de modelos de regressão logística bivariada e multivariada. Os resultados mostraram que dois factores sociodemográficos, escolaridade materna e local de residência, estavam significativamente associados ao estado vacinal infantil. Após ajustes para factores de confusão, o modelo reduzido apontou que as variáveis explicativas da vacinação básica completa foram a idade da criança e a escolaridade materna.

No modelo completo, apenas as crianças cujas mães tinham até quatro anos de escolaridade apresentaram maior risco de não serem vacinadas. Embora o maior número de irmãos tenha sido associado ao risco de não receber vacinação básica completa na análise não ajustada, essa associação não permaneceu após o controle de variáveis de confusão. Crianças com três ou quatro anos de idade apresentaram um risco 1,29 vezes maior de não serem vacinadas em comparação às mais jovens. Além disso, o risco de uma criança cuja mãe cursou até quatro anos de

escola não ser vacinada foi 1,42 vezes maior do que o das crianças cujas mães possuíam maior nível educacional.

A baixa escolaridade materna manteve-se como um factor associado à baixa cobertura vacinal, mesmo após os ajustes para factores de confusão, corroborando os achados de outros estudos. A literatura aponta que a baixa escolaridade materna está relacionada a diversos desfechos negativos na saúde infantil, incluindo maiores taxas de mortalidade e morbidade, pior desenvolvimento físico das crianças e menor frequência de consultas preventivas de puericultura.

Entre os estudos realizados no contexto africano, destaca-se uma investigação conduzida em Bujumbura, capital do Burundi, que analisou os factores associados à vacinação completa entre crianças de 24 a 48 meses hospitalizadas em quatro hospitais nacionais. Embora as vacinas sejam ofertadas gratuitamente em unidades de saúde no país, muitas crianças continuam sem completar o esquema vacinal, mesmo diante de múltiplas estratégias para aumentar a cobertura. Após décadas de altas taxas vacinais com cobertura superior a 80% desde o lançamento do Programa Expandido de Vacinação (PEV) em 1980, o país passou a registrar um leve declínio nos indicadores entre 2015 e 2017, tendência confirmada pelos dados da OMS/UNICEF de 2018.

Os resultados mostraram que apenas 64,81% das crianças estavam corretamente vacinadas. Após a modelagem, os factores estatisticamente associados à vacinação correta incluíram: nível de escolaridade do marido (OR = 15,41; $p = 0,021$), renda familiar (OR = 10,23; $p = 0,021$), distância da residência até a unidade de vacinação (OR = 0,12; $p = 0,000$) e nível de conhecimento das mães sobre doenças evitáveis por vacinas (OR = 1,73; $p = 0,004$). O modelo apresentou bom ajuste (teste de Hosmer-Lemeshow, $p = 0,6697$) e classificou corretamente 90% das observações.

O estudo realizado por Begum e Moniruzzaman (2017), ao empregar a regressão logística binária para identificar os factores associados ao estado vacinal de crianças, observou que aquelas cujas mães possuíam escolaridade secundária tinham 2,07 vezes mais chance de estarem completamente vacinadas, em comparação com filhos de mães sem instrução formal. Além disso, crianças cujas mães estavam empregadas apresentaram um aumento de 53% na probabilidade de estarem com vacinação completa.

Neste trabalho, aplica-se a regressão logística binária como ferramenta estatística para investigar os factores associados ao estado vacinal de crianças moçambicanas com menos de dois anos de idade. Embora esse modelo já tenha sido amplamente utilizado em pesquisas similares como nos estudos conduzidos em Bangladesh e Burundi sua aplicação continua relevante. A regressão logística binária permite estimar a probabilidade de uma criança estar ou não completamente vacinada com base em variáveis explicativas como escolaridade materna, condições do agregado familiar, zona de residência e acesso aos serviços de saúde. Assim, sua utilização neste estudo visa não apenas confirmar tendências observadas em outras regiões, mas também oferecer resultados concretos que possam ajudar na construção de políticas públicas mais eficazes em Moçambique.

2.4 Regressão logística

2.4.1 Histórico da regressão logística

Segundo Mesquita (2014), a técnica de regressão logística, foi desenvolvida no século XIX, porém, obteve maior visibilidade após 1950 ficando então, mais conhecida. Conseguiu um grande avanço a partir dos trabalhos de (Cox, 1989) e (Hosmer e Lemeshow, 2000). Caracteriza-se por descrever a relação entre uma variável dependente qualitativa binária, associada a um conjunto de variáveis independentes qualitativas ou métricas.

Desde então vários estudos foram feitos utilizando esta metodologia, sendo um dos mais famosos e de grande relevância para a maior implementação desta técnica, o famoso *Framingham Heart Study*, realizado com a colaboração da Universidade de Boston (Corrar, 2011). Embora a regressão logística inicialmente fosse direcionada para área médica, a eficiência desta técnica viabilizou sua implementação nas mais diversas áreas do conhecimento, tendo grande aceitação entre os usuários dos modelos de regressão tradicionais. Desde então, a regressão logística tornou-se uma poderosa ferramenta para análise de variáveis dicotômicas (Cramer, 2023).

Segundo Hair (1998), o termo regressão logística tem sua origem na transformação usada com a variável dependente, que permite calcular diretamente a probabilidade da ocorrência do fenômeno em estudo.

2.4.2 Definição da Regressão logística

A regressão logística é uma técnica amplamente usada para modelar dados binários, ou seja, o modelo de regressão logística é usado quando a variável dependente Y é do tipo nominal dicotômica (0,1). A regressão logística é usada para prever a probabilidade de ocorrência de um evento. Essa probabilidade é o parâmetro da distribuição binomial, e assim sendo, o modelo logístico possui uma forma linear para o logit desta probabilidade (Agresti, 2007):

$$\text{logit}(p) = \log \left(\frac{p}{1-p} \right) \quad (2.1)$$

Segundo Gonzalez (2018), a regressão logística é uma técnica estatística usada para criar um modelo que prevê os valores de uma variável categórica, frequentemente binária, com base em uma ou mais variáveis independentes, contínuas ou binárias.

Este método assim como as regressões lineares simples e múltiplas, estuda a relação entre uma variável resposta (dependente) e uma ou mais variáveis independentes. A diferença entre uma variável de regressão deve-se ao facto de que na regressão logística as variáveis dependentes estarem dispostas em categorias, enquanto na regressão linear estas variáveis são dados geralmente de natureza contínua. Outra diferença, é que na regressão logística a resposta é expressa por meio de probabilidade de ocorrência, enquanto na regressão simples se obtém um valor numérico qualquer (Castro e Ferreira, 2022).

Segundo Dobson e Barnett (2008), o modelo de regressão logística é uma generalização dos modelos lineares para variáveis de duas categorias, assumindo valores 0 para o insucesso e 1 para o sucesso. As variáveis respostas do modelo são caracterizadas pela distribuição de Bernoulli, onde designa-se de sucesso o resultado mais importante de resposta ou aquele que pretende-se relacionar com outras variáveis de interesse. A distribuição de Bernoulli para a variável aleatória binária Y com parâmetro π específica as probabilidades como:

$$P(Y = 1) = \pi \quad \text{e} \quad P(Y = 0) = 1 - \pi \quad (2.2)$$

Por definição:

$$\mathbb{E}(Y) = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi \quad (2.3)$$

Que é a proporção de respostas com $Y = 1$.

$$\text{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = 1^2 \cdot \pi + 0^2 \cdot (1 - \pi) - \pi^2 = \pi(1 - \pi). \quad (2.4)$$

A função de probabilidade de uma variável aleatória de Bernoulli é:

$$f(Y; \pi) = \pi^y (1 - \pi)^{1-y}. \quad (2.5)$$

2.4.3 Propriedades da regressão logística

A regressão logística além de explicar a relação entre as variáveis, possui algumas peculiaridades que em conjunto, fazem com que a técnica se torne diferente dos demais modelos de regressão tradicionais:

- A variável dependente é qualitativa dicotômica ou binária (falha ou sucesso, positivo ou negativo, sim ou não, aceitar ou rejeitar, certo ou errado, viver ou morrer, homem ou mulher, maior que ou menor que, etc...);
- A variável dependente segue uma distribuição de Bernoulli. Seja γ_i uma variável aleatória dummy definida como $\gamma_i \in \{0, 1\}$ tem distribuição de Bernoulli, cuja função de distribuição de probabilidade é dada por:

$$P(Y = y) = p^y (1 - p)^{1-y}, \quad y \in \{0, 1\}; \quad (2.6)$$

onde:

γ : indica a ocorrência do evento

p : indica a probabilidade de ocorrência do evento.

- Segundo Corrar (2011), necessita de amostras grandes, com aproximadamente 30 casos por variável independente;

- Mesmo quando a variável de interesse não é dicotômica, é possível torná-la binária, com a finalidade de aplicar a regressão logística.
- As variáveis independentes podem ser categóricas ou métricas
- Avalia a probabilidade de obtenção de uma das categorias da variável resposta

Trata-se, portanto, de uma técnica capaz de avaliar a probabilidade de ocorrência de determinado evento, e a influência de cada variável no resultado do fenômeno em estudo, sendo considerada como uma técnica de fácil interpretação pelos usuários dos modelos de regressão tradicionais (Mesquita, 2014).

Como se trata de uma sequência de eventos com distribuição de Bernoulli, a soma do número de sucessos ou fracassos nesta experiência terá distribuição Binomial de n parâmetros (número de observações) e p (probabilidade de sucesso). A função de distribuição de probabilidade da Binomial é dada por:

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (2.7)$$

2.5 Regressão logística simples

Segundo Hosmer Jr e Lemeshow (1989) os métodos de regressão são modelados com o objetivo de descrever as relações entre a variável resposta (Y) e a variável explicativa (X). Neste tipo de regressão a variável resposta é dicotômica, isto é, atribui-se o valor 1 para o evento de interesse (sucesso) e valor 0 para o acontecimento complementar. A probabilidade de sucesso é dada por:

$$\pi(x) = P(Y_i | x_i), \quad (2.8)$$

onde x_i é a variável explicativa associada à i -ésima resposta Y_i .

Considera-se uma amostra de respostas binárias, em que $(Y_1, Y_2, \dots, Y_n)$ são variáveis aleatórias independentes com distribuição de Bernoulli, com probabilidade de sucesso π_i , isto é: $Y_i \sim \text{Bernoulli}(\pi_i)$ e denota-se por $X_i^T = (1, x_i)$ a i -ésima linha da matriz X , em que $i = 1, 2, \dots, n$.

A probabilidade de sucesso do modelo logístico simples é dada por:

$$\pi = P(Y = 1 | X = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}. \quad (2.9)$$

E a probabilidade de fracasso:

$$1 - \pi = P(Y = 0 | X) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}. \quad (2.10)$$

Em que β_0 e β_1 são parâmetros desconhecidos. Nos modelos de regressão, modela-se o valor médio da variável dependente dado os valores das variáveis independentes. Esta quantidade é

chamada de *média condicional*, denotada por:

$$\mathbb{E}(Y_i \mid X = x_i) \quad (2.11)$$

Em que Y_i é a variável de resposta e x_i representa os valores das variáveis independentes. Devido à natureza da variável resposta, a amplitude da média condicional varia no intervalo $[0, 1]$, ou seja:

$$0 \leq \mathbb{E}(Y_i \mid X = x_i) \leq 1 \quad (2.12)$$

Usando a definição de variáveis aleatórias discretas, tem-se:

$$\mathbb{E}(Y_i \mid X = x_i) = 1 \cdot P(Y_i = 1 \mid X = x_i) + 0 \cdot P(Y_i = 0 \mid X = x_i) = P(Y_i = 1 \mid X = x_i) \quad (2.13)$$

A variável resposta Y_i dado x_i , é modelada por:

$$Y_i = \pi_i + \varepsilon_i. \quad (2.14)$$

O termo ε_i assume somente um dos dois valores possíveis, isto é:

$$\varepsilon_i = \begin{cases} 1 - \pi_i, & \text{se } Y_i = 1 \\ -\pi_i, & \text{se } Y_i = 0 \end{cases} \quad (2.15)$$

Segue que ε_i tem uma distribuição com média zero e variância dada por $\pi_i(1 - \pi_i)$, isto é, a distribuição binomial com probabilidade dada pela média condicional π_i .

2.5.1 Transformação logit

Para corrigir o problema de valores de probabilidade, ou seja, números que não pertençam ao intervalo $[0, 1]$, pode-se linearizar a função logística pela transformação designada logit.

A transformação logit para estudos de regressão logística é central e é definida como:

$$g(x_i) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right). \quad (2.16)$$

A partir das equações (2.9) e (2.10), tem-se:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad \text{e} \quad 1 - \pi_i = \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}; \quad (2.17)$$

Substituindo na função logit:

$$g(x_i) = \ln \left(\frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} \right) = \ln(e^{\beta_0 + \beta_1 x_i}) = \beta_0 + \beta_1 x_i. \quad (2.18)$$

$$g(x_i) = \ln(e^{\beta_0 + \beta_1 x_i}) = \beta_0 + \beta_1 x_i. \quad (2.19)$$

A função $g(x_i)$ apresenta as seguintes propriedades:

- É linear em seus parâmetros;
- É contínua;
- Varia no intervalo de $(-\infty, +\infty)$ com correspondência no intervalo de $[0,1]$, para i no contexto de modelos lineares generalizados.

2.5.2 Estimação dos parâmetros do modelo de regressão logística simples

Para ajustar coeficientes de um modelo de regressão, é necessário estimar parâmetros β_0 e β_1 do modelo.

Supondo que disponha de uma amostra independente com n pares de observações (x, y) , tal que y representa o valor da variável dicotômica e x o valor da variável independente da i -ésima observação em que $i = 1, 2, \dots, n$.

Para o ajuste do modelo de regressão logística simples, segundo a equação (2.10), é necessário estimar os parâmetros desconhecidos, isto é, valores de β_0 e β_1 . Na regressão linear clássica o método mais usado para estimar esses parâmetros é o de mínimos quadrados dos desvios para os valores observados de y_i em relação ao valor predito (y_i) baseado no modelo.

Na regressão logística simples os parâmetros β_0 e β_1 são estimados com base no método da máxima verossimilhança que, de uma forma geral, fornece valores para os parâmetros desconhecidos que maximizam a probabilidade de se obter determinado conjunto de valores.

Segundo Vaz (2020), na regressão não linear o método de máxima verossimilhança é utilizado de forma iterativa para que sejam encontradas as estimativas mais prováveis dos parâmetros.

A função de máxima verossimilhança é definida por:

$$\ell(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \quad \beta \in \mathbb{R}^2 \quad (2.20)$$

Os estimadores de máxima verossimilhança para os parâmetros β_0 e β_1 são os valores dos coeficientes estimados $\hat{\beta}_0$ e $\hat{\beta}_1$ que maximizam o logaritmo da função de verossimilhança.

Aplicando o logaritmo a função fica definida da seguinte forma:

$$\ell(\beta) = \ln \left(\prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right)$$

$$= \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \quad (2.21)$$

$$= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \quad (2.22)$$

$$= \sum_{i=1}^n \left[y_i(\beta_0 + \beta_1 x_i) + \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] \quad (2.23)$$

A expressão é definida por:

$$\ell(\beta) = \sum_{i=1}^n [Y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0 + \beta_1 x_i})] \quad (2.24)$$

Para obter os valores de β que maximizam $\ell(\beta)$, derivamos em relação a cada parâmetro (β_0, β_1) :

$$\frac{\partial \ell(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \quad (2.25)$$

$$\frac{\partial \ell(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(y_i x_i - \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \quad (2.26)$$

Igualando as derivadas a zero e substituindo β_0 e β_1 por suas estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$, temos:

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \left(\frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \right) \quad (2.27)$$

$$\sum_{i=1}^n y_i x_i = \sum_{i=1}^n \left(\frac{x_i e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \right) \quad (2.28)$$

Essas equações são resolvidas numericamente por métodos iterativos, como o método de Newton-Raphson. Com os parâmetros estimados, a probabilidade de sucesso estimada para o i -ésimo indivíduo é dada por:

$$\hat{\pi}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \quad (2.29)$$

Os coeficientes β_0 e β_1 têm interpretações análogas às da regressão linear, mas aqui baseiam-se na razão de chances (*odds ratio*).

2.5.3 Regressão logística múltipla

A regressão logística múltipla é uma generalização do modelo de regressão logístico simples, isto é, na regressão logística simples, trabalha-se com uma variável explicativa e na regressão logística múltipla com mais de uma variável explicativa.

A regressão logística múltipla representa um contexto da regressão logística, em que a variável dependente Y é binária ou dicotômica, ou seja, tem duas categorias e mais de duas variáveis independentes (Figueiro, 2006).

Considera-se a regressão logística múltipla como uma generalização da regressão logística binária, entretanto, nesta regressão, existe um conjunto de variáveis independentes denotadas por $X = (x_1, x_2, \dots, x_n)$. A combinação linear destas variáveis é definida como:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n. \quad (2.30)$$

De acordo com Kutner e Nachtsheim (2004), ao estender o modelo de regressão linear simples, simplesmente substituímos:

$$\beta_0 + \beta_1 x_i \quad \text{por} \quad \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1}. \quad (2.31)$$

Para simplificar as fórmulas, utiliza-se a notação matricial com os seguintes vectores:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_{p-1} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{i,p-1} \end{bmatrix} \quad (2.32)$$

Obtemos então:

$$\mathbf{X}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1} \quad (2.33)$$

$$\mathbf{X}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} \quad (2.34)$$

Com esta notação, a função de resposta logística simples estende-se para a regressão logística múltipla:

$$\mathbb{E}(Y) = \frac{\exp(\mathbf{X}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}^T \boldsymbol{\beta})}, \quad (2.35)$$

Logo, o logit igualando à $g(x)$, obtém-se:

$$\ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.36)$$

E, utilizando o antilogaritmo (função exponencial), obtemos a probabilidade estimada $\hat{p}$ no modelo de regressão logística múltipla:

$$\hat{p} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}. \quad (2.37)$$

Suposições da regressão logística múltipla

A regressão logística permite aos pesquisadores superar muitas das suposições de regressão linear que são muito restritivas. As seguintes suposições são relaxadas sob regressão logística:

- A relação linear entre as variáveis dependentes e independentes não são assumidas;
- A variável dependente não precisa ser normalmente distribuída;
- As variáveis dependentes não devem ter variância homocedástica (variância não precisa ser a mesma dentro das categorias);
- Não são assumidos termos de erro distribuídos normalmente;
- A variável resposta deve ser binária.

2.5.4 Estimação dos parâmetros na regressão logística múltipla

De acordo com Gonzalez (2018), para ajustar os parâmetros de um modelo de regressão linear múltipla, aplica-se o método de máxima verossimilhança a partir dos dados da amostra, este método procura os estimadores para o modelo de regressão logística, que são os valores do β que maximizam o logaritmo da função de máxima verossimilhança $\ell(\beta)$. Portanto a estimação por máxima verossimilhança permite encontrar os estimadores dos parâmetros do modelo que tem maior probabilidade de replicar o padrão de observações nos dados da amostra, o que exige um processo iterativo e que seja necessário derivar $\ell(\beta)$ em relação a cada parâmetro. A derivada da função de verossimilhança em relação ao parâmetro β_0 é dada por:

$$\frac{\partial \ell(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left[y_i x_{i0} - \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} x_{i0} \right] \quad (2.38)$$

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

$$\frac{\partial^2 \ln \ell(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (2.39)$$

$$\frac{\partial^2 \ln \ell(\beta)}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n x_{ij} x_{ik} \pi_i (1 - \pi_i) \quad (2.40)$$

$$\text{com } j, k = 0, 1, 2, \dots, p \quad (2.41)$$

Se for formada uma matriz quadrada de dimensão $(p + 1)$, constituída pelos valores simétricos médios referidos nas equações acima, obtém-se:

$$\mathcal{I}(\beta) = \text{Matriz de Informação de Fisher} \quad (2.42)$$

Matriz de Informação de Fisher

A matriz de informação de Fisher é dada por:

$$\mathcal{I}(\beta) = \mathbb{E} \left(-\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^T} \right) = X^T Q X, \quad (2.43)$$

onde:

- $Q = \text{diag}[\pi_i(1 - \pi_i)], \quad i = 1, 2, \dots, n;$
- X é a matriz dos dados (design matrix), com n linhas (observações) e $p + 1$ colunas (incluindo o intercepto)
- $\mathcal{I}^{-1}(\beta)$ é a matriz de variância e covariância das estimativas de máxima verossimilhança dos parâmetros.

A matriz X tem a forma:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \quad (2.44)$$

Cada linha de X representa o vetor de preditores para uma observação, incluindo a constante 1 para o intercepto β_0 .

Verificação da qualidade do ajuste do modelo logístico

Segundo Bender e Benner (2000), a avaliação do ajuste dos modelos logísticos pode detectar covariáveis importantes e interações omitidas. Embora muitos métodos tenham sido desenvolvidos para avaliar o ajuste de modelos de regressão logística binária, poucos desses métodos foram estendidos para dados de resposta ordinal.

Segundo Draper e Smith (1981), a qualidade do ajuste dos modelos ordinais é verificada usando o teste de Pearson. Esse teste envolve a criação de uma tabela de contingência na qual as linhas consistem em todas as possíveis configurações das covariáveis do modelo e as colunas são as categorias da resposta ordinal. As contagens esperadas dessa tabela são expressas por:

$$E_{ij} = \sum_{i=1}^{N_1} \hat{P}_{ij}, \quad (2.45)$$

onde: N_1 é o número total de indivíduos classificados na linha i ; $\hat{P}_{ij}$ representa a probabilidade estimada de um indivíduo na linha i ter a resposta j , calculada a partir do modelo adotado.

De acordo com Agresti (2013), o teste de Pearson para avaliar a adequação do ajuste do modelo compara as contagens observadas O_{ij} com as esperadas E_{ij} por meio da seguinte fórmula:

$$\chi^2 = \sum_{l=1}^L \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}. \quad (2.46)$$

A estatística deviance também compara contagens observadas e esperadas, Dobson e Barnett (2008), sendo dada por:

$$D^2 = 2 \sum_{l=1}^L \sum_{j=1}^k O_{ij} \log \left(\frac{O_{ij}}{E_{ij}} \right). \quad (2.47)$$

Os testes de qualidade do ajuste baseiam-se na aproximação das estatísticas χ^2 e D^2 para a distribuição Qui-quadrado, com:

$$(L - 1)(k - 1)p \text{ graus de liberdade;} \quad (2.48)$$

onde:

- L é o número de linhas da tabela de contingência;
- k é o número de colunas;
- p é o número de covariáveis do modelo.

Um valor-p significativo leva à conclusão de que o modelo apresenta falta de ajuste aos dados observados.

Em termos de ajuste geral do modelo, um dos principais testes utilizados é o de Hosmer e Lemeshow. Esse teste é considerado mais robusto do que o teste de Qui-quadrado comum, principalmente quando existem variáveis independentes contínuas ou quando o tamanho da amostra é pequeno (Garson, 2011).

Teste da razão de verossimilhança

Segundo Hosmer Jr e Lemeshow, o teste da razão de verossimilhança tem por objetivo comparar os valores observados da variável resposta (y) com os valores preditos obtidos no modelo. Esta medida testa simultaneamente se os coeficientes de regressão associados aos β são todos nulos com a exceção de β_0 , Cabral (2013). A comparação entre os valores observados e esperados com base na função de verossimilhança é expressa da seguinte forma:

$$D = -2 \ln \left(\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}} \right); \quad (2.49)$$

$$D = -2 \sum_{i=1}^n \left[\gamma_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) \right], \quad (2.50)$$

onde:

- $\gamma_i \in \{0, 1\}$ é a variável indicadora do evento;
- $\hat{\pi}_i$ é a probabilidade estimada pelo modelo;
- y_i é o valor observado da variável resposta.

Um valor de D elevado (com p-valor pequeno) indica que o modelo ajustado não fornece bom ajuste comparado ao modelo saturado, sugerindo falta de ajuste.

Um modelo saturado contém todas as variáveis, enquanto o modelo ajustado corresponde ao modelo com apenas as variáveis desejados no estudo. O teste da razão de verossimilhança é também conhecida como deviance (desvio) e avalia o valor ajustado na regressão logística, desempenhando o mesmo papel que a soma de quadrados residuais tem na regressão linear. A deviance sempre é positiva e quanto menor, melhor é o ajuste do modelo.

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 & (\text{modelo nulo}) \\ H_1 : \exists \beta_k \neq 0 & (\text{modelo ajustado}) \end{cases}$$

Teste de Wald

Segundo Hosmer jr e Lemeshow (1989), o teste de wald é usado na regressão logística para a determinação da significância dos coeficientes do modelo estimado, ele testa se cada coeficiente é significativamente diferente de zero. Deste modo, o teste de Wald verifica se uma determinada variável independente possui relação estatisticamente significativa com a variável dependente.

Se os coeficientes logísticos forem estatisticamente significativos pode-se interpretar em termos do impacto na probabilidade estimada, deste modo, na predição do objeto de estudo no grupo respectivo, isto é, no grupo do evento de interesse ($Y = 1$) ou no grupo da não ocorrência ($Y = 0$).

O teste de Wald é obtido comparando a estimativa de máxima verossimilhança de um coeficiente e a estimativa do erro padrão.

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}. \quad (2.51)$$

, examinaram o desempenho do teste de Wald e descobriram que, em alguns casos, ele se comporta de maneira inconsistente, falhando ao rejeitar a hipótese nula mesmo quando o coeficiente é significativo. Alternativamente, recomenda-se o uso do teste da razão de verossimilhança, quando há dúvidas de que teste de Wald tenha falhado.

As hipóteses do teste de Wald são:

$$\begin{cases} H_0 : \beta_i = 0; \\ H_1 : \beta_i \neq 0. \end{cases} \quad (2.52)$$

Pseudo R^2 de Cox e Shell

A estatística R^2 é uma medida que avalia em termos percentuais, a qualidade de um ajustamento de um modelo de regressão linear aos dados observados. Na regressão logística, não existe uma estatística equivalente ao R^2 . No modelo de regressão linear, as variáveis dependentes são contínuas, o que não é no caso da regressão logística, onde a variável dependente é categórica (Cox e Shell, 1989).

A denominação de pseudo R deve-se ao facto de que eles se parecem com um do modelo de regressão linear, pois estão em uma escala similar, variando de 0 a 1, cujos valores mais próximos de 1 indicam melhor ajuste do modelo, mas, apesar dessa similaridade, não podem ser interpretados da mesma forma como se interpreta um R^2 (Mesquita, 2014).

Existem várias formas de calcular o pseudo R^2 , alguns autores inclusive, citam mais de 10 formas, mas infelizmente, não há um consenso sobre o qual é a melhor no geral. As técnicas de pseudo R^2 não são muito precisas quanto a estatística R^2 e para a regressão linear. Cox (1989) apresentou o pseudo R^2 usado frequentemente, inclusive em alguns softwares estatísticos, com a seguinte definição:

$$R^2 = 1 - \left(\frac{L(\hat{\beta}_0)}{L(\hat{\beta}_M)} \right)^{\frac{2}{n}}, \quad (2.53)$$

onde:

- n é o tamanho da amostra;
- $L(\hat{\beta}_0)$ é o valor da função de verossimilhança do modelo sem preditores (modelo nulo);
- $L(\hat{\beta}_M)$ é o valor da função de verossimilhança do modelo estimado (modelo completo).

A racionalidade desta fórmula está no facto de que ela expressa a proporção de melhoria do ajuste do modelo completo em relação ao modelo nulo, considerando a probabilidade dos dados sob cada modelo. Assim, representa uma medida geral de qualidade do ajuste, embora não possa ser interpretada da mesma forma que o R^2 da regressão linear.

Teste Hosmer e Lemeshow

Conforme Hosmer e Lemeshow (2000), esta estatística corresponde a um teste Qui-quadrado que consiste em dividir o número de observações em aproximadamente dez classes e, em seguida comparar as frequências preditas com as observadas.

A finalidade deste teste é verificar se existem diferenças significativas entre as classificações realizadas pelo método e a realidade observada. Antes do cálculo da estatística teste, é necessário estimar a probabilidade de sucesso para cada observação e ordenar as probabilidades preditas em ordem crescente, e em seguida agrupar os dados de acordo com os decis de probabilidades preditas. Em cada decil, dividir os valores observados e os valores preditos para o sucesso e fracasso.

A um nível de significância estabelecido, busca-se não rejeitar a hipótese de que não existem diferenças entre os valores preditos e os observados. O critério de avaliação se distingue um pouco do convencional, pois geralmente o que se deseja é rejeitar a hipótese nula. Nesse caso se houver diferenças significativas entre as classificações preditas pelo modelo e as observadas, então o modelo não representa a realidade de forma satisfatória. Em tal situação, o modelo não seria capaz de produzir estimativas e classificações confiáveis.

Interpretação dos coeficientes de regressão logística

Os coeficientes estimados num modelo de regressão logística não apresentam uma interpretação directa da inclinação de um modelo de regressão linear. A razão é que o efeito de um aumento unitário em X varia para o modelo de regressão logística de acordo com a localização do ponto de partida na escala X . A interpretação dos coeficientes encontra-se na propriedade da função logística ajustada nas probabilidades estimadas:

$$\text{odds} = \frac{\hat{\pi}}{1 - \hat{\pi}}. \quad (2.54)$$

Em um modelo de regressão logística, a razão de chances é multiplicada por $\exp(b_1)$ para cada aumento unitário na variável X , mantendo as demais variáveis constantes. Isso significa que $\exp(b_1)$ representa a razão de chances ajustada associada a um incremento de uma unidade em X .

Considerando:

$$\hat{\pi}'(X_j) = \beta_0 + b_1 X_j; \quad (2.55)$$

A notação $\hat{\pi}'(X_j)$ indica especificamente o nível de X associado ao valor ajustado da função linear preditora.

Assim, a função logit completa pode ser expressa por:

$$\log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = \beta_0 + b_1 X; \quad (2.56)$$

A razão de chances ajustada para X é dada por:

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(\beta_0 + b_1 X). \quad (2.57)$$

Razão de Chances (odds ratio)

A razão de chances (OR) é uma ferramenta simples, amplamente usada em pesquisas clínicas. Kalra (2016), define a razão de chances (OR) como sendo uma medida de associação que compara as probabilidades de uma determinada doença ou de um evento expostos as probabilidades da doença ou do evento no qual esta exposto. Ele serve para determinar a relação entre a exposição e o resultado.

A razão de chances é uma expressão numérica da força de associação entre causa e efeito em situações em que as variáveis são categóricas. Como regra geral, quanto maior o OR, maior será o efeito no resultado. Segundo Kahn e Sempos (1989), a OR é utilizada como uma medida de associação entre a exposição e a doença, assumindo valores de zero a infinita (nunca menores que zero). Para o cálculo da OR, pode-se calcular com a seguinte fórmula:

$$OR = \frac{\text{Probabilidade da doença nos expostos}}{\text{Probabilidade da doença nos não expostos}}. \quad (2.58)$$

A interpretação da OR depende da sua magnitude e significância:

- Se a $OR = 1$, o fator não é considerado risco, ou seja, não há associação entre a exposição e a doença;
- Se a $OR > 1$, o fator é considerado de risco, ou seja, indica que a redução de chances da doença é maior quando exposto ao risco (associação positiva);
- Se a $OR < 1$, o fator é considerado de proteção, ou seja, indica a redução das chances para doença na exposição ao fator de risco (associação negativa);

2.5.5 Seleção das variáveis independentes

Segundo Tabakhi *et al.* (2014), a redução do número de variáveis contribui para a diminuição da dimensionalidade do conjunto de dados, de tal forma que os aspectos mais significativos dos dados sejam representados em um subconjunto de dados úteis a partir do conjunto de dados de entrada.

De acordo com Danasingh *et al.* (2016), a seleção de variáveis é um processo de remoção das características irrelevantes e redundantes de um conjunto de dados, a fim de melhorar o desempenho dos algoritmos em termos de precisão e tempo para construção do modelo. A seleção de variáveis é uma técnica para classificação de dados e redução de dimensionalidade, podendo ser usada para encontrar um subconjunto ótimo de características relevantes na qual a precisão global de classificação aumenta enquanto a dimensão dos dados é reduzida (Dash e Liu, 2003).

2.5.6 Procedimento de seleção de variáveis

Assim como na regressão linear múltipla, na regressão logística múltipla é também possível utilizar métodos de seleção de variáveis com poder preditivo. Segundo Pestana e Gageiro (2014),

existem três métodos nomeadamente backward delection, stepwise regression e forward selection.

Procedimento stepwise

Segundo Draper e Smith (1981), o procedimento stepwise começa escolhendo uma equação contendo apenas uma variável X , e posteriormente adiciona outras variáveis independentes com base nos valores parciais do teste F . O maior valor parcial de F é comparado com o valor do F padrão ou pré-selecionado. Ou seja, o procedimento consiste na seleção do z mais correlacionado com Y até que se observe melhoria na significância dos parâmetros e do R .

Procedimento backward

O método backward (eliminação para trás) é um procedimento que examina apenas as melhores regressões contendo um certo número de variáveis. Os passos básicos do procedimento são:

- Cálculo de uma equação de regressão contendo todas as variáveis;
- Cálculo do valor F parcial para cada variável preditora;
- Comparação dos F 's parciais calculados. Caso algum F parcial seja menor que o nível de significância pré-selecionado ou padrão, elimina-se essa variável.

Procedimento forward

De um modo geral, este tipo de método é similar ao método stepwise usado na regressão linear, que consiste na adição de uma variável no modelo se é significativa para 5%, (isto é, a variável entra no modelo se o seu p -value no teste de adição não exceder 0,05) e a remoção para o nível de significância de 10% (isto é, uma variável é removida se o seu p -value no teste de remoção exceder 0,10 parciais de máxima verossimilhança do modelo).

A variável X_i que melhor prevê Y , é a primeira escolhida a entrar. O critério seguido para selecionar a seguinte variável exógena é a que possua o maior coeficiente de correlação semi-parcial.

2.5.7 Critérios para seleção do melhor modelo

De acordo com Bello (2010), a seleção do melhor modelo, partindo de um conjunto de modelos candidatos, pode ser bastante complexa. Uma alternativa é considerar todas as combinações possíveis de termos do modelo completo e do número de parâmetros, utilizando critérios de seleção de modelos baseados na teoria da informação.

Critério de Informação de Akaike (AIC)

Akaike (1973), desenvolveu o critério de informação de Akaike (AIC), que baseia-se na minimização da quantidade de informação (ou distância) de Kullback-Leibler ($K - -L$) entre o

modelo verdadeiro e o modelo candidato.

A distância ($K - L$) é uma medida da discrepância entre esses dois modelos. Na prática, o modelo verdadeiro raramente é conhecido, sendo, na maioria das vezes, uma abstração. Por isso, é desejável obter um modelo suficientemente bom, que represente bem os dados observados.

Segundo Burnham e Anderson (2002) a escolha do modelo mais eficiente e interpretável é feita com base no Critério de Informação de Akaike (AIC). O critério de seleção de informação de Akaike é dado por:

$$AIC = -2 \sum_{i=1}^n \ln L(\hat{\mu}_i, y_i) + 2 \cdot (\text{número de parâmetros}). \quad (2.59)$$

O AIC torna-se útil quando são comparados diversos modelos. O melhor modelo será aquele que apresentar o menor valor de AIC. É recomendada apenas a utilização do AIC quando a razão entre o número de observações e variáveis preditoras (n/p) for maior ou igual a 40 (Burnham and Anderson, 2002). A expressão AIC pode ser simplificada, conforme a equação seguinte:

$$AIC = n \ln(\hat{\sigma}_p^2) + 2 \cdot (p + 1) \quad (2.60)$$

Critério de informação de Akaike corrigido (AICc)

Hurvish e Tsai (1989), desenvolveram um critério de seleção de modelos, o AICc, indicado para pequenas amostras e respostas com distribuição normal. Mcquarrie e Tsai (1998), afirmam que o critério AICc pode aumentar apreciavelmente a probabilidade de se escolher um modelo adequado, particularmente na seleção de modelos de regressão e de séries temporais. Entretanto, sua aplicação pode ser inadequada quando a suposição de normalidade dos erros é violada.

Para a seleção de modelos nos casos de respostas com distribuição normal e pequenas amostras ($n/p < 40$), Hurvich Tsai (1989) desenvolveram o critério AIC_c :

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2}$$

Critério de informação Bayesiano

O Critério de Informação Bayesiano (BIC), também conhecido como critério de Schwarz, é uma medida amplamente utilizada para avaliação e seleção de modelos estatísticos, (Schwarz, 1978). Diferente do AIC, o BIC é derivado de argumentos bayesianos e incorpora uma penalização mais forte para modelos com muitos parâmetros. Ele é definido pela expressão:

$$BIC = -2 \cdot \log(\hat{L}) + k \cdot \log(n), \quad (2.61)$$

onde:

- $\hat{L}$ é o valor da função de verossimilhança do modelo ajustado;

- k é o número de parâmetros estimados;
- n é o tamanho amostral.

O BIC favorece modelos que apresentam bom ajuste aos dados com menor complexidade, sendo particularmente útil em contextos de grandes amostras. Assim, quanto menor o valor de BIC, mais adequado é considerado o modelo (Emeliano, 2009).

2.5.8 Curva Roc

De acordo Hosmer e Lemeshow (2000), a curva ROC (Receiver Operating Characteristic) é uma ferramenta que permite avaliar o desempenho de um modelo de regressão binária (variável resposta é do tipo 0-1). Pode ser feita por meio de um gráfico simples e robusto, que permite estudar a variação da sensibilidade e especificidade, para diferentes pontos de quebra.

De acordo com Corrar (2011), a Curva ROC é uma ferramenta útil para avaliar a capacidade de um modelo de classificação entre duas classes. Ela traça a taxa de verdadeiros positivos (Sensibilidade) no eixo y e a taxa de falsos positivos (1-Especificidade) no eixo x , onde:

- Sensibilidade (Taxa de verdadeiros positivos) é a proporção de observações da classe positiva (1) que são correctamente classificadas pelo modelo.
- Especificidade (1 - Taxa de falsos positivos) é a proporção de observações da classe negativa (0) que são correctamente classificadas pelo modelo.

A Curva ROC-AUC é usada para medir a capacidade discriminativa do modelo, ou seja, quão bem ele consegue separar as duas classes. Quanto maior a área sob a Curva ROC melhor o modelo em classificar as observações. Assim, a representação da curva ROC permite evidenciar os valores para os quais existe otimização da Sensibilidade em função da Especificidade, correspondente ao ponto que se encontra mais próximo do canto superior esquerdo do diagrama, uma vez que o índice de verdadeiro positivo é 1 e o de falso positivo 0. A área abaixo da curva de ROC, fornece uma medida de discriminação.

Seja R , o valor que corresponde à área abaixo da curva de ROC, como regra geral tem-se as seguintes regras de orientação:

$$\left\{ \begin{array}{ll} R = 0.5 & \text{sem discriminação} \\ 0.5 < R < 0.7 & \text{discriminação fraca} \\ 0.7 \leq R < 0.8 & \text{discriminação aceitável} \\ 0.8 \leq R < 0.9 & \text{discriminação boa} \\ R \geq 0.9 & \text{discriminação excepcional} \end{array} \right.$$

A Curva ROC-AUC é usada para medir a capacidade discriminativa do modelo, ou seja, quão bem ele consegue separar as duas classes. Quanto maior a área sob a Curva ROC (ROC-AUC), melhor o modelo em classificar as observações:

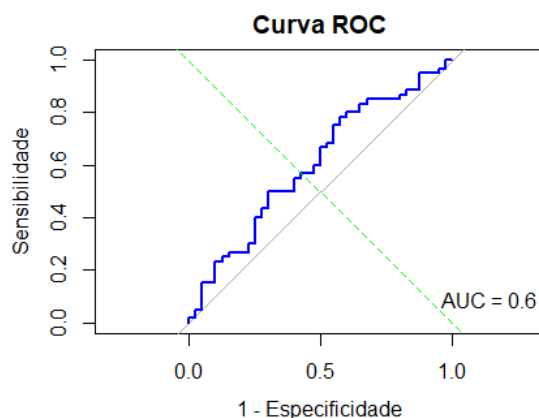


Figura 2.1: Curva ROC.

Capítulo 3

Material e métodos

3.1 Classificação do estudo

Quanto à natureza da pesquisa

Quanto à natureza, esta pesquisa pode-se classificar como aplicada, pois a mesma objetiva gerar conhecimentos para aplicação prática dirigidos à solução do problema específico (Gil, 2002).

Quanto à abordagem do problema

No que se refere à abordagem do problema, a pesquisa classifica-se em quantitativa, pois envolve a coleta e análise de dados numéricos e aplicação de testes estatísticos (Gil, 2002).

Quanto aos objectivos

Em relação aos seus objetivos, esta pesquisa classifica-se como explicativa. Para Nascimento (2016), a pesquisa explicativa tem como finalidade identificar os factores que influenciam ou determinam a ocorrência de determinados fenômenos, indo além da simples descrição de suas características. Esse tipo de estudo geralmente emprega métodos mais aprofundados e, por vezes, o método experimental, dada a complexidade envolvida.

Quanto aos procedimentos técnicos

Quanto aos procedimentos técnicos, a pesquisa é classificada em bibliográfica, pois utiliza material já publicado em livros, artigos de periódicos (Gil, 2002).

Quanto ao delineamento do estudo

Quanto ao delineamento, o presente estudo é do tipo transversal. Os estudos transversais têm como principal objectivo obter dados fidedignos que, ao final da pesquisa, permitam a elaboração de conclusões confiáveis e robustas, além de possibilitar a formulação de novas hipóteses a serem exploradas em investigações futuras (Raimundo *et al.*, 2018).

3.2 Material

3.2.1 Fonte de dados

Para o presente estudo, foi usada uma base de dados secundária referente ao IDS 2022-23, disponibilizada para a realização exclusiva deste trabalho.

A população alvo da pesquisa foi constituída por todas as mulheres de 15-49 anos de idade e por todas as crianças menores de 2 anos residentes no agregado familiar. A unidade amostral foi constituída por cada criança menor de 2 anos residente no agregado familiar.

Critério de inclusão

A população do estudo incluiu todos agregados familiares e visitantes que tenham passado na noite anterior no mesmo agregado.

Critério de exclusão

Foram excluídos os agregados familiares e respectivos membros moradores em residências colectivas, tais como hotéis, hospitais, quartéis militares, lares de estudantes e os sem-abrigo, que representam menos de 0,5% da população total do país.

Tipo de amostragem

A recolha de dados foi feita obedecendo duas etapas, na primeira etapa a seleção da amostra foi feita com base na amostragem por conglomerados (clusters), e no segundo estágio, os agregados familiares foram seleccionados sistematicamente com probabilidades iguais de cada área de enumeração.

Cálculo dos pesos amostrais

Segundo INE e ICF (2023) o cálculo dos pesos amostrais é um passo crucial para garantir que os resultados das pesquisas sejam representativos para toda a população-alvo, assim sendo, para esta pesquisa, foi considerada a variável v005 para o cálculo dos pesos amostrais.

Variável dependente

O estado de vacinação da criança foi considerado como variável dependente. Esta variável foi categorizada da seguinte forma:

- **Completamente vacinada:** para uma criança que tenha recebido as 8 doses de vacinas;
- **Parcialmente vacinada:** para uma criança que não recebeu todas as vacinas.

3.2.2 Descrição das variáveis independentes

A Tabela 3.1 apresenta as variáveis relacionadas às características das crianças incluídas na amostra.

Tabela 3.1: Características da Criança

Código	Nome da Variável	Categorias	Classificação
B4	Sexo da criança	Feminino	Categórica
		Masculino	
V437	Peso ao nascer	Baixo peso	Categórica
		Peso normal	
		Acima do peso	
HW57	Nível de hemoglobina	Nível Baixo	Categórica
		Nível normal	
		Acima do Nível	
M55	Amamentação da criança	Sim	Categórica
		Não	
B11	Intervalo entre nascimentos	Menor de 2 anos	Categórica.
		Entre 2 a 5 anos	
		Maior que 5 anos	

Características da mãe

Na Tabela 3.2 são descritas as variáveis associadas às características maternas.

Tabela 3.2: Características da mãe

Código	Nome da Variável	Categorias	Classificação
V012	Idade da mãe	Jovem	Categórica
		Adulto	
V106	Nível de escolaridade	Nenhum	Categórica
		Primário	
		Secundário ou +	
V714	Ocupação da mãe	Trabalha	Categórica
		Não trabalha	
M14	Número de consultas pré-natais	Não sabe	Categórica
		Nenhuma	
V501	Estado civil da mãe	Solteira	Categórica.
		Casada	
H1	Acesso aos documentos de vacinação	Teve	
		Não/Perdi	

Características do agregado familiar

A Tabela 3.3 resume as variáveis, relacionadas ao agregado familiar

Tabela 3.3: Características do agregado familiar

Código	Nome da Variável	Categorias	Classificação
V025	Local de residência	Urbano	Categórica
		Rural	
V190	Índice de riqueza	Pobre	Categórica
		Médio	
		Rico	
V151	Sexo do chefe do agregado	Masculino	Categórica
		Feminino	
V130	Religião do agregado	Católica	Categórica
		Islâmica	
		Zion	
		Outra	
V705	Ocupação do chefe do agregado	Desempregado	Categórica
		Empregado	
M15	Local do parto	Casa	Categórica
		Hospital público	
		Outro	
-	Exposição à mídia	Sim	Categórica.
		Não	

3.3 Métodos

3.3.1 Análise descritiva

A descrição das variáveis é um ramo da estatística que se preocupa com a organização, apresentação e sintetização de dados. Utilizam-se gráficos, tabelas e medidas descritivas como ferramentas. A descrição das variáveis é utilizada na etapa inicial da análise, destinada a obter informações que indicam possíveis modelos a serem utilizados numa fase final, (Medri, 2011).

De acordo com Medri (2011), a descrição tabelar numérica de dados é a representação das informações por intermédio de uma tabela. A tabela é a forma eficiente de mostrar os dados levantados, facilitando desta forma a compreensão e interpretação dos mesmos.

No estudo aplicou-se a descrição tabelar por meio de tabelas de frequências relativas percentuais e tabelas de frequências cruzadas.

Ajuste do plano amostral

Ajustou-se o plano amostral, usando ponderações, com pesos calculados com base nos critérios do DHS program, para garantir que as análises refletissem a representatividade nacional no processo de inferência com dados de amostras complexas (INE e ICF, 2011).

3.3.2 Análise bivariada

Teste de independência do Qui-quadrado

O teste Qui-quadrado aplica-se quando os dados se apresentam em categorias, para determinar se as duas variáveis são independentes (Silva, 1992). Neste trabalho, aplicou-se o teste de Qui-quadrado para testar a associação entre as variáveis independentes e o estado de vacinação.

As hipóteses a serem testadas são:

H_0 : As variáveis são independentes, ou seja, não há associação entre as variáveis;

H_1 : As variáveis não são independentes, ou seja, há associação entre as variáveis;

A estatística do teste é dada pela expressão:

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad (3.1)$$

onde:

- O_{ij} é a frequência observada na linha i e coluna j .
- E_{ij} é a frequência esperada na linha i e coluna j .
- l é o número de linhas.
- c é o número de colunas.

De modo a identificar os factores associados ao estado de vacinação das crianças, aplicou-se a regressão logística múltipla. Conforme Levin *et al.* (2012), esse método é adequado quando a variável dependente é dicotômica.

A regressão logística é uma técnica estatística amplamente recomendada para situações em que a variável dependente é dicotômica ou binária, permitindo que as variáveis independentes sejam tanto categóricas quanto não categóricas. Essa abordagem possibilita a estimativa da probabilidade associada à ocorrência de eventos específicos em relação a um conjunto de variáveis explicativas (Hosmer Jr e Lemeshow, 1989).

Segundo Mesquita (2014), os modelos de regressão podem ser descritos como técnicas que permitem explicar relação entre uma variável dependente e um conjunto de variáveis independentes. Diferentemente da regressão linear, a regressão logística não pressupõe os pressupostos

de homocedasticidade (homogeneidade das variâncias) e normalidade dos resíduos. A probabilidade de sucesso no modelo logístico múltiplo é dada por:

$$\pi_i = P(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \dots + \gamma_q D_{qi})}} \quad (3.2)$$

O logaritmo da razão entre as probabilidades da criança ser completamente e parcialmente vacinada é denominado logito e apresenta uma forma linear nos parâmetros (Hosmer e Lemeshow, 2000), sendo expresso da seguinte forma:

$$\ln \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \dots + \gamma_q D_{qi} \quad (3.3)$$

A interpretação dos parâmetros em regressão logística é feita a partir da odds (chances) (Hosmer and Lemeshow, 2000), que é a razão entre as probabilidades da criança estar completamente e parcialmente vacinada sendo expressa por:

$$\text{odds} = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \dots + \gamma_q D_{qi}) \quad (3.4)$$

Para fazer comparações entre dois indivíduos, usa-se a OR (odds ratio) que é a razão entre duas odds (chances) dada pela equação anterior (Hosmer e Lemeshow, 2000).

$$OR = \frac{\text{odds}_1}{\text{odds}_2}. \quad (3.5)$$

O processo de seleção de variáveis foi realizado com base no Teste de Qui-quadrado.

Para testar a significância individual de cada coeficiente no modelo, foi utilizado o teste de Wald (Agresti, 2007). O teste de Wald é uma abordagem comum para testar a significância de parâmetros individuais em um modelo de regressão (Maroco, 2007).

H_0 : O parâmetro de interesse é igual a um valor específico (usualmente zero, indicando nenhuma associação ou efeito).

H_1 : O parâmetro de interesse é diferente do valor específico.

O teste de Hosmer-Lemeshow foi aplicado para avaliar o ajuste dos dados no modelo logístico. Além disso, o coeficiente de Nagelkerke Pseudo R^2 foi utilizado para medir a proporção da variabilidade da variável dependente que é explicada pelo modelo, (Hosmer e Lemeshow, 2013).

Segundo Hair *et al.* (2009) o teste de Hosmer e Lemeshow é utilizado para demonstrar a qualidade do ajuste do modelo, ou seja, se o modelo pode explicar os dados observados.

$$HL = \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j(1 - \frac{E_j}{n})}. \quad (3.6)$$

Hipóteses do teste de Hosmer-Lemeshow:

H_0 : O modelo estimado se ajusta aos dados.

H_1 : O modelo estimado não se ajusta aos dados.

O teste de Nagelkerke R^2 é uma medida de ajuste para modelos de regressão logística, que estima a proporção de variância explicada pelo modelo em comparação com o modelo nulo (Hosmer e Lemeshow, 2000).

Hipóteses do teste de Nagelkerke R^2 :

H_0 : O modelo nulo (sem variáveis explicativas) é igualmente bom que o modelo proposto (com variáveis explicativas).

H_0 : O modelo proposto é significativamente melhor que o modelo nulo.

Por fim, aplicou-se a curva ROC para avaliar o desempenho dos coeficientes do modelo estimado, segundo (Hosmer e Lemeshow, 2013), a avaliação do modelo é feita com base nas seguintes regras:

Tabela 3.4: Avaliação da curva roc

Valor da AUC	Interpretação
0.5	Sem discriminação
0.5 - 0.7	Baixa capacidade de discriminação
0.7 - 0.9	Moderada capacidade de discriminação
> 0.9	Alta capacidade de discriminação.

No teste do Qui-quadrado foi adoptado um nível de significância de 10%, enquanto nos demais testes estatísticos foi utilizado o nível de significância de 5%.

Capítulo 4

Resultados e discussão

4.1 Resultados

De acordo com a Tabela 4.1, observa-se que 64,24% das crianças da amostra foram completamente vacinadas, enquanto 35,76% estavam apenas parcialmente vacinadas. Em relação ao sexo da criança, houve uma leve predominância do sexo feminino (55,25%) em comparação ao masculino (44,75%). No que diz respeito as visitas pré-natais, 75,59% das mães realizaram pelo menos uma visita durante a gestação.

Quanto aos níveis de hemoglobina, 71,73% apresentavam valores normais, enquanto 22,48% estavam em níveis baixos, sugerindo possíveis casos de anemia. Além disso, 5,78% apresentaram níveis acima do normal. A documentação vacinal esteve disponível para a maioria das crianças, com 91,22% das mães relatando posse do cartão de vacinação. Em relação ao local de parto, 50,10% dos partos ocorreram em hospitais públicos, 47,11% em casa, e uma pequena parcela 8,78% em outros locais.

Sobre a amamentação, 58,89% das crianças foram amamentadas, o que, embora majoritário, ainda deixa 41,11% sem esse cuidado essencial nos primeiros meses de vida. A maioria das famílias residia em áreas rurais 81,37%, o que pode influenciar o acesso aos serviços de saúde e, por consequência, afectar a adesão ao calendário vacinal completo.

Observa-se que a maior parte das mães possui baixo nível de escolaridade: 44,11% não possuem nenhuma instrução e 42,18% têm apenas o nível primário. Apenas 13,71% atingiram o nível secundário ou mais. Com relação à ocupação da mãe, 71,52% não exercem actividade laboral remunerada.

O quintil de riqueza mostra que 58,67% das famílias são pobres, enquanto apenas 21,63% estão na categoria rica. O chefe do agregado familiar é majoritariamente do sexo masculino (81,80%), enquanto as mulheres lideram apenas 18,20%. Quanto à religião, destaca-se a predominância de católicos (52,46%), seguidos por islâmicos (25,91%) e igreja Zion (10,92%). Outras religiões correspondem a 10,71% da amostra. No que diz respeito à idade da mãe, 96,79% são adultas, enquanto apenas 3,21% são jovens. A ocupação do chefe do agregado indica que 58,89% estão

empregados, enquanto 41,11% encontram-se desempregados. O estado marital das mães revela que 46,25% são casadas, e 53,75% estão solteiras. Por fim, no que se refere à exposição à mídia, 74,52% relataram ter acesso à leitura de jornais, escuta de rádio ou televisão, o que pode facilitar o acesso à informação sobre vacinação, enquanto 25,48% afirmaram não ter nenhuma exposição à mídia.

Tabela 4.1: Descrição da amostra

Variável	Categoria	F_i	F_r (%)
Estado vacinal	Parcial	167	35.76
	Completa	300	64.24
Sexo da criança	Masculino	209	44.75
	Feminino	258	55.25
Visitas pré-natais	Nenhuma	114	24.41
	Uma ou mais	353	75.59
Hemoglobina	Baixo	105	22.48
	Normal	335	71.73
	Acima do normal	27	5.78
Doc. de vacinação	Teve	426	91.22
	Não/perdi	41	8.78
Local de parto	Casa	220	47.11
	Hospital público	234	50.10
	Outros locais	13	2.78
Amamentação	Não	192	41.11
	Sim	275	58.89
Área de residência	Urbana	87	18.63
	Rural	380	81.37
Escolaridade da mãe	Nenhuma	206	44.11
	Primária	197	42.18
	Secundária ou mais	64	13.71
Ocupação da mãe	Não trabalha	334	71.52
	Trabalha	133	28.48
Quintil de riqueza	Pobre	274	58.67
	Médio	92	19.70
	Rico	101	21.63
Sexo do chefe do agregado	Masculino	382	81.80
	Feminino	85	18.20
Religião	Católica	245	52.46
	Islâmica	121	25.91
	Zion	51	10.92
	Outra	50	10.71
Idade da mãe	Jovem	15	3.21
	Adulta	452	96.79
Ocup. chefe do agregado	Desempregado	192	41.11
	Empregado	275	58.89
Estado civil da mãe	Casada	216	46.25
	Solteira	251	53.75
Exposição à mídia	Sim	348	74.52
	Não	119	25.48.

4.1.1 Tabelas cruzadas

De acordo com a Tabela 4.2, observa-se que a vacinação completa das crianças não ocorre de maneira uniforme entre os diferentes grupos sociodemográficos. Crianças do sexo feminino apresentam uma proporção ligeiramente maior de vacinação completa (36,8%) em comparação aos meninos (27,4%).

No que se refere à variável Amamentação da criança, observa-se que, entre aquelas que “Não” foram amamentadas, 36,40% estão completamente vacinadas, enquanto entre as que foram amamentadas, a proporção é de 27,84%. Quanto à variável Peso da criança, crianças classificadas como “Baixo Peso” registram 62,74% de cobertura completa, muito superior aos 1,49% observados em crianças “Acima do peso”, enquanto a variável Visitas pré-natais constitui um dos factores mais determinantes para a vacinação completa: crianças cujas mães tiveram “Uma ou mais” visitas apresentam 64,24% de cobertura completa, enquanto aquelas sem visitas têm apenas 17,99%.

Na variável Hemoglobina, crianças com nível “Normal” de hemoglobina têm 49,25% de vacinação completa, com nível “Baixo” (12,85%) e “Acima do nível” (2,14%). Quanto ao Intervalo de nascimento, os resultados indicam que crianças nascidas com intervalo entre 2 e 5 anos apresentam a maior taxa de vacinação completa (42,61%), seguidas por aquelas com intervalo “Maior que 5” (12,85%) e “2 anos” (8,77%).

A variável Acesso aos documentos de vacinação sugere forte associação, pois 56,96% das crianças que teve acesso ao documento estão completamente vacinadas, ao passo que apenas 7,28% daquelas que não tiveram, têm esquema completo de vacinação. Em relação à variável Idade da mãe, nota-se que mães “Adultas” têm 62,96% de crianças completamente vacinadas, ao passo que mães “Jovem” alcançam somente 1,28%. A variável Área de residência revela que em “Rural” a cobertura completa atinge 47,54%, ao passo que em “Urbana” é de apenas 16,70%.

A variável Quintil de riqueza sugere que famílias “Pobres” têm 32,12% de cobertura completa, enquanto “Médio” registram 12,85% e “Rico” 19,27%. Na variável Local de parto, observa-se que crianças nascidas em “Hospital público” apresentam 41,97% de esquema completo, superiores às de “Casa” (20,77%) e “Outros locais” (1,50%).

O perfil da variável Escolaridade da mãe indica que aquelas com nível “Primário” têm 29,34% de cobertura completa, seguidas por “Nenhuma” (22,48%) e “Secundária +” (12,42%). Para a variável Ocupação da mãe, mães que “Não trabalha” têm 43,25% de filhos completamente vacinados, contra 20,99% das que “Trabalha”. A variável Sexo do chefe do agregado mostra cobertura completa de 51,61% quando o chefe é “Masculino” e 12,63% quando é “Feminino”. Em relação à variável Religião, crianças de famílias “Católica” e “Islâmica” têm cobertura completa próxima (19,91% e 17,99%, respectivamente), enquanto grupos como “Zion” (5,35%) e

“Outra” (7,07%) apresentam taxas variadas. Por fim, a variável Estado civil da mãe indica que mães “Solteiras” registram 36,40% de crianças completamente vacinadas, ligeiramente acima das mães “Casada” (27,84%).

Tabela 4.2: Distribuição percentual da vacinação segundo variáveis sociodemográficas

Variável	Cat.	Parciais (%)	Completas (%)
Sexo da criança	Masculino	17.34	27.41
	Feminino	18.42	36.83
Amamentação da criança	Sim	13.27	27.84
	Não	22.48	36.40
Peso da criança	Baixo Peso	32.54	62.74
	Acima do peso	3.21	1.49
Hemoglobina	Nível baixo	9.63	12.85
	Nível normal	22.48	49.25
	Acima do nível	3.64	2.14
Intervalo de nascimento	2 anos	6.42	8.77
	Entre 2 e 5 anos	24.20	42.61
	Maior que 5	5.14	12.85
Número de visitas pre-natais	Nenhuma	17.77	17.99
	Uma ou mais	35.76	64.24
Acesso os documentos de vacinação	Não tive	1.50	7.28
	Tive	34.26	56.96
Escolaridade da mãe	Nenhuma	21.63	22.48
	Primária	12.85	29.34
	Secundária +	1.28	12.42
Ocupação da mãe	Não trabalha	28.27	43.25
	Trabalha	7.49	20.99
Quintil de riqueza	Pobre	26.55	32.12
	Médio	6.85	12.85
	Rico	2.36	19.27
Sexo do chefe do agregado	Masculino	30.19	51.61
	Feminino	5.57	12.63
Local de parto	Casa	26.34	20.77
	Hospital público	8.78	41.97
	Outros locais	0.64	1.50
Religião	Católica	18.63	33.83
	Islâmica	7.92	17.99
	Zion	5.57	5.35
	Outra	3.64	7.07
Idade da mãe	Jovem	1.93	1.28
	Adulta	33.83	62.96
Ocup. do chefe do agregado	Desempregado	17.56	23.55
	Empregado	18.20	40.69
Área de residência	Urbana	1.93	16.70
	Rural	33.83	47.54
Estado civil da mãe	Casada	18.42	27.84
	Solteiro	17.34	36.40.

4.1.2 Visualização gráfica

Com base nas figuras 4.1 e 4.2, observa-se que tanto crianças do sexo masculino assim, como do sexo feminino apresentam maior proporção de casos de vacinação completa, todavia, crianças do sexo masculino apresentam maior proporção de crianças não vacinadas comparadas com as do sexo feminino.

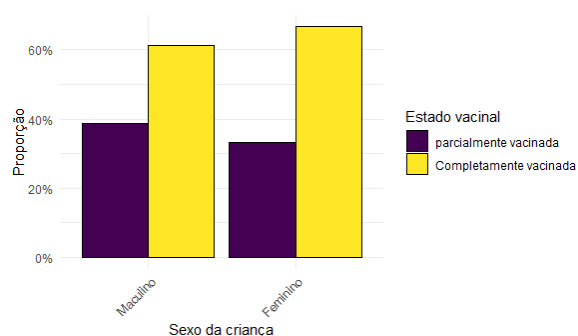


Figura 4.1: Proporção do estado vacinal por sexo da criança

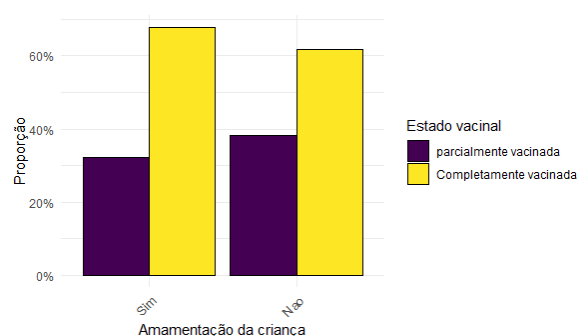


Figura 4.2: Proporção do estado vacinal por amamentação da criança

Observando as figuras 4.3 e 4.4, nota-se que crianças com intervalos de nascimento inferiores a 2 anos apresentam a maior proporção de vacinação parcial, seguidas daquelas com intervalos entre 2 e 5 anos. Já as crianças com intervalos superiores a 5 anos são as que apresentam, proporcionalmente, os menores níveis de vacinação incompleta.

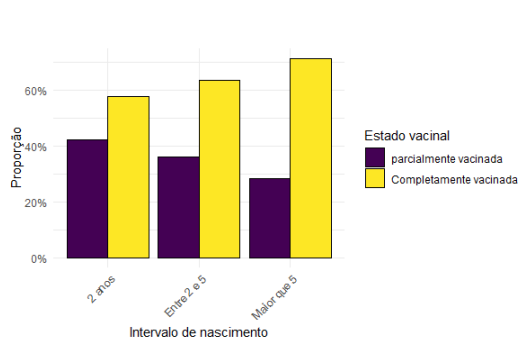


Figura 4.3: Proporção do estado vacinal pelo intervalo de nascimento

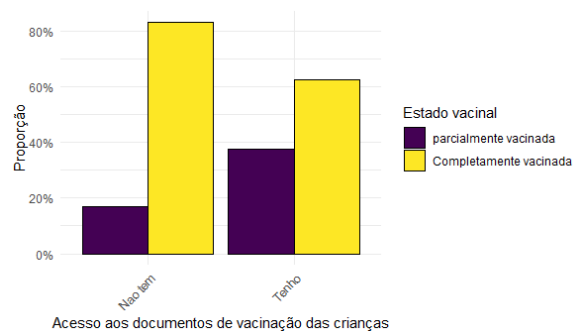


Figura 4.4: Proporção do estado vacinal pelo acesso aos documentos de vacinação

As figuras 4.5 e 4.6 mostram que o estado vacinal das crianças varia conforme o estado civil e o nível de escolaridade da mãe. Crianças de mães casadas apresentam maior proporção de vacinação incompleta comparadas com aquelas cujas mães são solteiras. Além disso, observa-se que, quanto menor o nível de escolaridade da mãe, menor é a proporção de crianças parcialmente vacinadas.

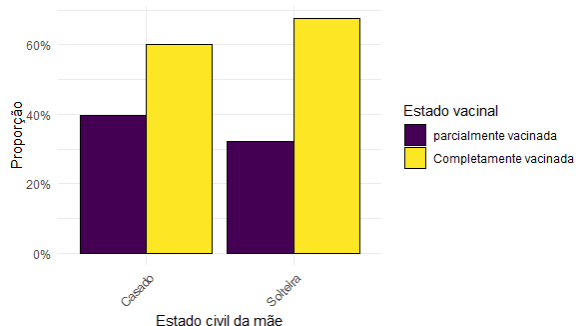


Figura 4.5: Proporção do estado vacinal por Estado civil

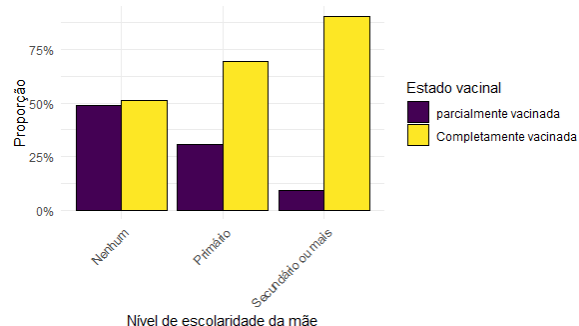


Figura 4.6: Proporção do estado vacinal por nível de escolaridade

As figuras 4.7 e 4.8 indicam que o sexo e a ocupação do chefe do agregado familiar influenciam a vacinação infantil. Quando o chefe do agregado é do sexo feminino, a proporção de vacinação completa é mais elevada. Da mesma forma, crianças de chefes empregados têm maiores chances de estarem completamente vacinadas em comparação com aquelas cujos chefes estão desempregados.

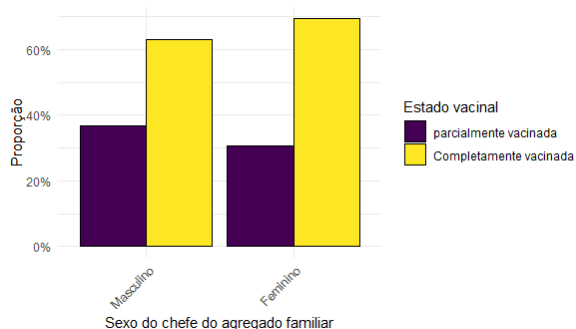


Figura 4.7: Proporção do estado vacinal por sexo do agregado familiar

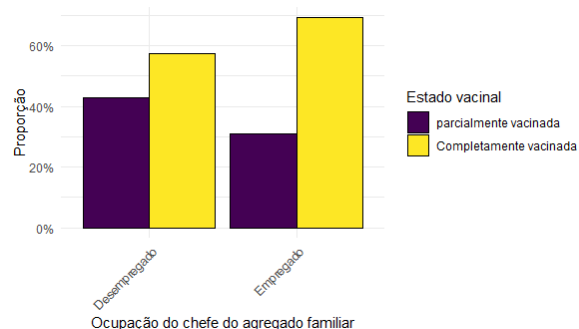


Figura 4.8: Proporção do estado vacinal pela ocupação do chefe do agregado

As figuras 4.9 e 4.10 mostram que tanto a idade quanto a ocupação da mãe influenciam o estado vacinal das crianças. Filhos de mães adultas têm maior proporção de vacinação completa do que os de mães jovens. De forma semelhante, quando a mãe está inserida no mercado de trabalho, a proporção de crianças completamente vacinadas é mais elevada em comparação às mães que não trabalham.

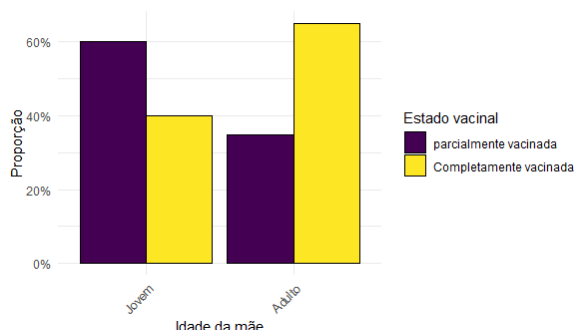


Figura 4.9: Proporção do estado vacinal pela idade da mãe

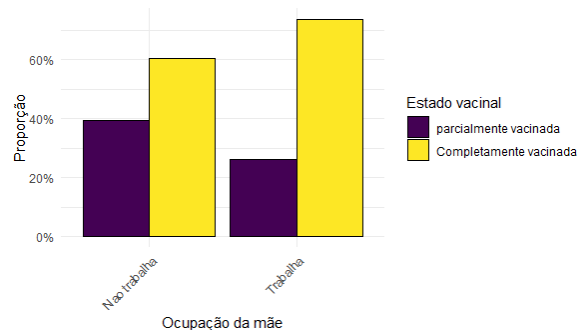


Figura 4.10: Proporção do estado vacinal pela ocupação da mãe

As figuras 4.11 e 4.12 indicam que a exposição à mídia e a religião do agregado familiar também estão associadas à vacinação. Crianças cujas mães têm acesso à mídia apresentam maiores taxas de vacinação completa do que aquelas sem acesso. Em relação à religião, observa-se que os grupos católico e islâmico apresentam maiores proporções de vacinação completa em comparação a outros grupos religiosos.

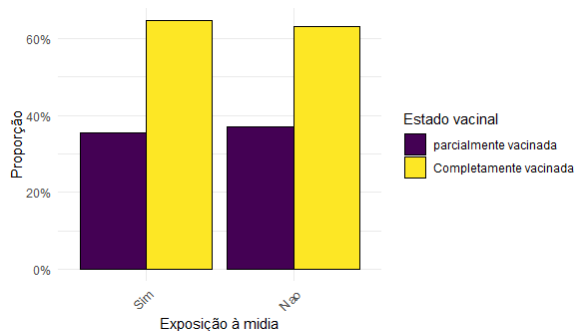


Figura 4.11: Proporção do estado vacinal pelo acesso à mídia

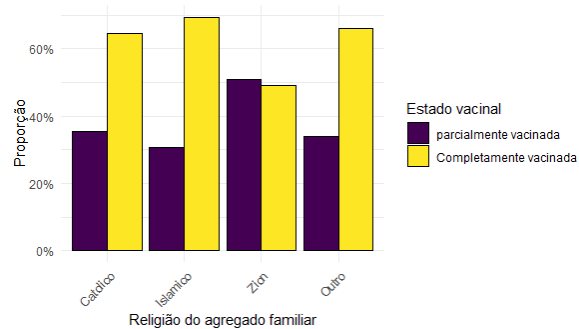


Figura 4.12: Proporção do estado vacinal pela religião

4.1.3 Teste de qui quadrado

Com base na Tabela 4.3, observa-se que algumas variáveis mostraram-se estatisticamente significativas a um nível de 10%, enquanto outras não apresentaram evidência suficiente de associação com o estado de vacinação.

Entre as variáveis que se destacaram como significativas estão a área de residência, o número de visitas pré-natais, o nível de escolaridade da mãe, a ocupação da mãe, o local de parto, o quintil de riqueza, a religião do agregado familiar, o acesso aos documentos de vacinação e o estado civil da mãe.

Por outro lado, as variáveis Sexo da criança, Peso da criança, Sexo do chefe do agregado familiar, Ocupação do chefe do agregado familiar, Amamentação da criança, Idade da mãe e Exposição à mídia não mostraram associação estatisticamente significativa.

Tabela 4.3: Associação entre estado vacinal e variáveis explicativas

Variável	F	p-valor	Significativo (10%)
Sexo da criança	1.352	0.2595	Não significativo
Peso da criança	1.2951	0.1372	Não significativa
Área de residência	14.672	1.822e-06	Significativa
Número de visitas pré-natais	5.3997	1.563e-08	Significativa
Nível de escolaridade da mãe	5.2905	5.905e-05	Significativa
Ocupação da mãe	6.657	0.0099	Significativa
Local de parto	7.4579	<0.001	Significativa
Quintil de riqueza	4.8883	9.122e-05	Significativa
Sexo do chefe do agregado familiar	0.10183	0.8720	Não significativa
Religião do agregado familiar	2.4721	0.007583	Significativo
Amamentação da criança	0.94251	0.3896	Não significativo
Idade da mãe	0.90444	0.5196	Não significativo
Acesso aos documentos de vacinação	22.475	<2.2e-16	Significativo
Ocupação do chefe do agregado familiar	1.3428	0.1891	Não significativo
Estado civil da mãe	0.21856	0.007802	Significativo
Exposição à mídia	0.33574	0.6843	Sim significativo.

4.1.4 Modelo logístico inicial

De acordo com a Tabela 4.4, as variáveis Escolaridade da mãe (superior), Número de visitas pré-natais, parto em hospital público, Religião (Zion), Área de residência (rural), Exposição a mídia (não), Idade da mãe (adulta) e Estado civil da mãe (casada) apresentaram associação estatisticamente significativa com o estado vacinal das crianças.

Tabela 4.4: Modelo logístico ponderado: factores associados à vacinação completa

Variável	Categoria	Coef. (log-OR)	EP	p-valor	Significância
(Intercepto)	–	0.085	1.162	0.942	
Área de residência	Rural	-1.001	0.510	0.051	.
Número de visitas pré-natais	Uma ou mais	1.521	0.426	<0.001	***
Escolaridade da mãe	Primário	0.347	0.314	0.271	
	Secundário +	1.267	0.712	<0.001	***
Ocupação da mãe	Trabalha	0.844	0.344	0.015	*
Quintil de riqueza	Médio	0.414	0.452	0.361	
	Rico	0.616	0.559	0.271	
Local de parto	Público	0.823	0.397	0.039	*
	Outro	1.226	0.770	0.113	
Religião do agregado familiar	Islâmico	0.332	0.367	0.367	
	Zion	-0.884	0.431	0.041	*
	Outro	-0.055	0.568	0.923	
Idade da mãe	Adulta	-0.393	0.782	0.015	
Acesso aos documentos	Tenho	0.026	0.587	0.965	
Ocupação do chefe do agregado	Empregado	-0.271	0.316	0.392	
Exposição à mídia	Sim	-0.834	0.429	0.053	.
Estado civil da mãe	casado	0.124	0.305	0.005	
Intervalo de nascimento	Entre 2 e 5	0.018	0.417	0.965	
	Maior que 5	0.357	0.505	0.480	
Hemoglobina	Nível normal	-0.032	0.384	0.933	
	Acima do nível	-0.713	0.696	0.307	
Sexo da criança	Feminino	0.175	0.309	0.572	
Amamentação da criança	Não	0.108	0.291	0.711	
Peso da criança	Acima do peso	-1.257	0.808	0.121	
Sexo do chefe do agregado	Feminino	0.155	0.352	0.660.	

Nota: EP = Erro Padrão; Significância: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, . $p < 0.1$

4.1.5 Modelo logístico com variáveis significativas no teste Qui-quadrado

Após identificar as variáveis associadas ao estado de vacinação por meio do teste do Qui-quadrado, foram estimados os coeficientes do modelo logístico considerando apenas aquelas que se mostraram estatisticamente significativas. Os resultados indicam que o **número de visitas pré-natais**, o **nível de escolaridade secundária ou mais da mãe**, a **ocupação materna (quando a mãe trabalha)**, **níveis adequados de hemoglobina**, o **nascimento em hospitais públicos** e a **afiliação religiosa à igreja Zion** estão significativamente associados ao estado de vacinação da criança.

Tabela 4.5: Modelo logístico com variáveis significativas no teste Qui-quadrado

Variável	Categoria	Coef. (log-OR)	EP	p-valor	Sig.
(Intercepto)	—	0.194	1.091	0.859	
Área de residência	Rural	-0.948	0.538	0.079	.
Número de visitas pré-natais	Uma ou mais	1.407	0.391	<0.001	***
Escolaridade da mãe	Primário	0.384	0.314	0.223	
	Secundário +	1.152	0.674	<0.001	***
Ocupação da mãe	Trabalha	0.766	0.329	0.021	*
Quintil de riqueza	Médio	0.434	0.442	0.328	
	Rico	0.280	0.547	0.609	
	Nível normal	-0.023	0.400	0.954	
Hemoglobina	Acima do nível	-1.726	0.569	0.003	**
	Público	0.879	0.396	0.027	*
	Outro	0.975	0.707	0.169	
Religião do agregado familiar	Islâmico	0.271	0.383	0.480	
	Zion	-0.993	0.398	0.013	*
	Outro	-0.099	0.552	0.858	
Idade da mãe	Adulta	-0.386	0.745	0.605	
Acesso aos documentos	Tenho	-0.035	0.572	0.952	
Ocupação do chefe do agregado	Empregado	-0.217	0.325	0.505	

Nota: EP = Erro Padrão. Significância: * p<0.05; ** p<0.01; *** p<0.001; . p<0.1

4.1.6 Modelo final

Com base nos resultados anteriores, foi estimado um modelo de regressão logística ponderado com variáveis combinadas, selecionadas a partir da análise bivariada (teste Qui-quadrado) e da significância observada no modelo inicial, com o objectivo de identificar os factores fortemente associados à vacinação completa de crianças. A Tabela 4.6 apresenta os coeficientes estimados do modelo final.

O modelo apresentou como variáveis estatisticamente significativas: Área de residência, Número de visitas pré-natais, Escolaridade da mãe (nível secundário ou mais), Ocupação da mãe, Quintil de riqueza (rico), Local de parto (hospital público), Religião do agregado familiar (Zion), Idade da mãe e Exposição à mídia.

Tabela 4.6: Modelo logístico ponderado com variáveis combinadas (teste Qui-quadrado e modelo anterior)

Variável	Categoria	Coef. (log-OR)	EP	p-valor	Sig.
(Intercepto)	—	0.308	1.086	0.777	
Área de residência	Rural	-1.035	0.553	0.063	.
Número de visitas pré-natais	Uma ou mais	1.486	0.416	<0.001	***
Escolaridade da mãe	Primário	0.347	0.317	0.274	
	Secundário +	1.293	0.698	<0.001	***
Ocupação da mãe	Trabalha	0.808	0.338	0.018	*
Quintil de riqueza	Médio	0.425	0.457	0.353	
	Rico	0.607	0.558	0.028	*
Hemoglobina	Nível normal	0.031	0.389	0.937	
	Acima do nível	-0.624	0.682	0.361	
Local de parto	Público	0.806	0.401	0.046	*
	Outro	1.102	0.765	0.151	
Religião do agregado familiar	Islâmico	0.323	0.368	0.381	
	Zion	-0.860	0.434	0.049	*
	Outro	-0.081	0.554	0.884	
Idade da mãe	Adulta	-0.339	0.743	0.049*	
Acesso aos documentos	Tenho	-0.026	0.585	0.964	
Ocupação do chefe do agregado	Empregado	-0.285	0.319	0.372	
Exposição à mídia	Sim	-0.804	0.411	0.052	.
Peso da criança	Acima do peso	-1.314	0.808	0.105.	

Nota: EP = Erro Padrão. Significância: * p<0.05; ** p<0.01; *** p<0.001; . p<0.1

Interpretação Razões de chances do modelo final

Com base na 4.7, as variáveis estatisticamente significativas revelam associações importantes com a vacinação completa, a seguir, as interpretações das razões de chances significativas:

- Crianças residentes em áreas urbanas apresentaram 3,36 vezes mais chances de estarem completamente vacinadas do que as que vivem em áreas rurais.

- Crianças cujas mães realizaram uma ou mais visitas pré-natais têm 4,42 vezes mais chances de estarem completamente vacinadas.
- Crianças de mães com escolaridade superior têm 4,30 vezes mais chances de estarem completamente vacinadas em comparação com aquelas cujas mães não possuem escolaridade.
- Filhos de mães que trabalham possuem 2,24 vezes mais chances de estarem vacinados do que os de mães que não trabalham.
- Crianças pertencentes ao quintil de riqueza mais elevado (rico) apresentam 3,84 vezes mais chances de estarem vacinadas do que aquelas do quintil mais pobre.
- Partos realizados em hospitais públicos estão associados a 2,24 vezes mais chances de vacinação completa das crianças em comparação aos partos domiciliares.
- Crianças cujas mães são adultas apresentam 4,71 vezes mais chances de estarem completamente vacinadas do que aquelas de mães jovens.
- Crianças expostas à mídia têm 2,45 vezes mais chances de estarem completamente vacinadas em comparação com as que não estão.

Tabela 4.7: Razões de chances (OR) e intervalo de confiança (IC 95%) para o modelo logístico ponderado

Variável	OR	IC 95% Inferior	IC 95% Superior
(Intercepto)	1.360	0.160	11.548
Área de residência (Urbano)	3.355	1.709	4.119
Número de visitas pré-natais (Uma ou mais)	4.420	1.949	10.023
Escolaridade da mãe (Primário)	1.415	0.758	2.642
Escolaridade da mãe (Secundário+)	3.644	2.107	8.7805
Ocupação da mãe (Trabalha)	2.243	1.153	4.363
Quintil de riqueza (Médio)	1.530	0.622	3.765
Quintil de riqueza (Rico)	3.835	2.611	5.507
Hemoglobina (Nível normal)	1.031	0.480	2.218
Hemoglobina (Acima do nível)	0.536	0.140	2.054
Local de parto (Público)	2.239	1.016	4.934
Local de parto (Outro)	3.011	0.667	13.590
Religião (Islâmico)	1.381	0.669	2.850
Religião (Zion)	2.423	1.996	3.180
Religião (Outro)	0.922	0.309	2.750
Idade da mãe (Adulta)	4.713	3.165	5.082
Documentos de vacinação (Tenho)	0.974	0.307	3.087
Ocupação do chefe do agregado (Empregado)	0.752	0.401	1.409
Exposição à mídia (Sim)	2.448	1.199	2.006
Peso da criança (Acima do peso)	0.269	0.055	1.321.

Nota: OR = razão de chances; IC = intervalo de confiança.

4.1.7 Ajuste do modelo final

Com base na Tabela 4.8, verifica-se que todos os VIFs estão abaixo de 5, o que indica que as variáveis independentes do modelo final não estão correlacionadas entre si.

Tabela 4.8: Teste de multicolinearidade

Variável	VIF
Área de residência	1.46
Nº de visitas pré-natais	1.36
Nível de escolaridade da mãe	1.46
Ocupação da mãe	1.58
Quintil de riqueza	1.90
Hemoglobina	2.32
Local de parto	1.30
Religião do agregado familiar	1.44
Idade da mãe	1.36
Acesso aos documentos vacinais	1.64
Ocupação do chefe do agregado	1.29
Exposição à mídia	1.57
Peso da criança	2.50.

Com base no Teste de Wald, observou-se que as variáveis: área de residência, número de visitas pré-natais, nível de escolaridade da mãe, ocupação da mãe, quintil de riqueza, local de parto, religião do agregado familiar, idade da mãe, acesso aos documentos vacinais da criança e exposição à mídia, apresentam um p-valor menor que o nível de significância de 5%, ou seja, rejeita-se a hipótese nula de que os coeficientes dessas variáveis sejam iguais a zero. Conclui-se, portanto, que essas variáveis incluídas no modelo são estatisticamente significativas.

Tabela 4.9: Teste de Wald

Variável	Wald	p-valor
Área de residência	3.16	0.035
Nº de visitas pré-natais	4.47	0.045
Nível de escolaridade da mãe	5.20	0.023
Ocupação da mãe	5.55	0.018
Quintil de riqueza	6.11	0.039
Hemoglobina	1.89	0.170
Local de parto	2.50	0.014
Religião do agregado familiar	2.85	0.037
Idade da mãe	6.01	0.006
Acesso aos documentos vacinais	4.14	< 0.001
Ocupação do chefe do agregado	0.62	0.432
Exposição à mídia	3.71	0.011
Peso da criança	1.85	0.602.

Ao nível de significância de 5%, não se rejeita a hipótese nula de que o modelo se ajusta bem aos dados. Conclui-se que o modelo apresenta um bom ajuste.

Tabela 4.10: Teste de Hosmer e Lemeshow para ajuste do modelo

Estatística	Valor
Qui-quadrado (X^2)	9.4854
Graus de liberdade (gl)	8
Valor-p	0.303.

A área sob a curva (AUC) foi de 0,8225, valor superior ao limiar de 0,8, indicando que o modelo apresenta excelente capacidade discriminatória. Portanto, o modelo distingue adequadamente as crianças parcialmente vacinadas e completamente vacinadas.

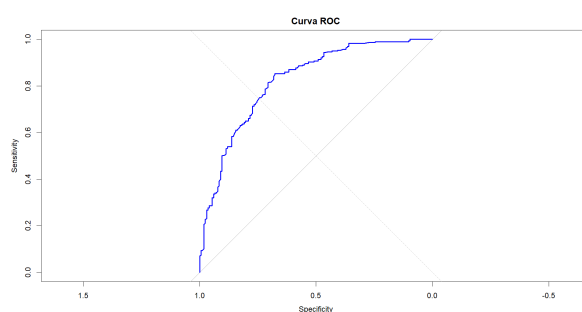


Figura 4.13: Curva ROC

A Tabela 4.11 apresenta a matriz de confusão do modelo. Foram corretamente classificadas 131 crianças parcialmente vacinadas e 241 completamente vacinadas. Houve 36 falsos positivos (parcialmente vacinadas preditas como completamente vacinadas) e 59 falsos negativos (completamente vacinadas preditas como parcialmente vacinadas).

Tabela 4.11: Matriz de confusão do modelo logístico ponderado

Classe Real / Predita	Parcialmente vacinada	Completamente vacinada
Parcialmente vacinada	131	36
Completamente vacinada	59	241.

A Tabela 4.12 apresenta as principais métricas de desempenho do modelo. A acurácia total foi de 79,66%, indicando que quase 80% das classificações foram corretas. A sensibilidade foi de 80,30%, mostrando que o modelo identificou corretamente a maioria das crianças completamente vacinadas. A especificidade foi de 78,45%, refletindo boa capacidade de identificar as crianças parcialmente vacinadas.

Tabela 4.12: Métricas de desempenho do modelo logístico ponderado

Métrica	Valor
Acurácia	79,66%
Sensibilidade (Recall)	80,30%
Especificidade	78,45%.

A Tabela 4.13 apresenta os valores do Pseudo R^2 para os três modelos logísticos avaliados. Observa-se uma melhoria progressiva na qualidade do ajuste, com o Modelo 1 apresentando um Pseudo R^2 de 0,329, seguido pelo Modelo 2 com 0,346. O Modelo 3, considerado final, obteve o melhor desempenho com um Pseudo R^2 de 0,436, indicando maior capacidade explicativa da combinação final de variáveis sobre o estado vacinal das crianças.

Tabela 4.13: Comparação dos valores de Pseudo R^2 entre os modelos logísticos

Modelo	Pseudo R^2
Modelo 1: Inicial com variáveis selecionadas	0.346
Modelo 2: Baseado no teste Qui-quadrado	0.329
Modelo 3: Modelo final ajustado	0.436.

4.2 Discussão

O presente trabalho teve como objectivo analisar os factores associados ao estado de vacinação das crianças menores de 2 anos em Moçambique, utilizando dados do Inquérito Demográfico e de Saúde (IDS) de 2022-23.

Numa primeira etapa, procedeu-se à limpeza da base de dados devido à presença de dados incompletos em algumas variáveis. Para lidar com esse problema, foi aplicada a técnica de exclusão de casos completos (listwise deletion), conforme sugerido por Schafer e Graham (2002), resultando em uma amostra final composta por 467 crianças com informações completas.

Após a limpeza dos dados, realizou-se uma análise exploratória, na qual se observou que 64,24% das crianças estavam completamente vacinadas, enquanto 35,76% apresentavam vacinação parcial. Esse resultado não se distancia significativamente do que é reportado na literatura. Segundo UNICEF (2023), cerca de 40% das crianças não têm acesso adequado aos serviços de vacinação. A diferença observada nos resultados deste estudo pode estar relacionada à redução da amostra após o processo de exclusão de dados incompletos.

A amostra analisada foi composta por 55,25% de crianças do sexo feminino e 44,75% do sexo masculino. Com relação aos níveis de hemoglobina, 71,73% apresentavam valores dentro da faixa considerada normal, 22,48% estavam abaixo do recomendado, e 5,78% apresentavam níveis elevados. Quanto ao acompanhamento pré-natal, observou-se que 75,59% das mães realizaram pelo menos uma visita, ao passo que 24,41% não participaram de nenhuma consulta.

A maioria dos partos ocorreu em hospitais públicos, seguidos por partos domiciliares. Verificou-se ainda que 58,89% das crianças foram amamentadas e 91,22% possuíam cartão de vacinação. Quanto à área de residência, 81,37% das crianças viviam em zonas rurais. Estes resultados estão em contraste com a literatura, pois em um estudo realizado por Araújo *et al.* (2020), 49,02% das crianças eram do sexo feminino, e 50,98% do sexo masculino, 85% das mães tiveram acesso às cadernetas de vacinação.

Ainda na análise exploratória dos dados, observou-se que 44,11% das mães não possuíam qualquer nível de escolaridade, 42,18% haviam concluído o ensino primário, e apenas 13,71% apresentavam nível secundário ou superior. Quanto à situação laboral, 71,52% das mães não exerciam actividade remunerada, ao passo que 28,49% estavam inseridas no mercado de trabalho. A maioria das mães era adulta, sendo que 3,21% eram jovens. Dentre estas jovens, 46,25% eram casadas e 53,75% solteiras.

Alguns desses resultados divergem da literatura. Por exemplo, no estudo realizado por Santos *et al.* (2020), a maioria das mães era jovem, com idades entre 19 e 30 anos. No que se refere à ocupação, a maior parte das mulheres era do lar, ou seja, não exercia nenhuma actividade remunerada. Em relação ao estado civil, maior parte das mulheres eram casadas. No que

diz respeito à escolaridade, a maioria das mães revelaram ter o primeiro grau incompleto e nenhuma das mães tinha formação secundária ou superior, convergindo, portanto, com os achados da presente pesquisa.

Em relação às características do agregado familiar, verificou-se que 81.80% dos agregados familiares eram liderados por homens e apenas 18.20% por mulheres. A maior parte dos chefes dos agregados familiares estava empregada, 58.67% das famílias eram pobres, e 52.46% frequentavam a igreja católica, 25.91% eram islâmicos, 10.92% eram da religião Zion e 10.72% eram de outras religiões.

Com o objectivo de identificar quais categorias apresentavam maiores proporções de crianças parcialmente vacinadas, foi elaborada uma tabela cruzada. Os resultados indicaram que crianças do sexo feminino, não amamentadas, com baixo peso ao nascer, com nível normal de hemoglobina e intervalo de nascimento entre dois e cinco anos apresentaram menor adesão ao esquema vacinal completo.

Observou-se ainda que crianças cujas mães tiveram partos em casa, casadas, que realizaram uma ou mais consultas pré-natais, que possuíam cartão de vacinação, e que residiam em áreas rurais também apresentaram baixas taxas de vacinação completa.

Verificou-se ainda que crianças cujas mães não possuem escolaridade, não exercem actividade laboral, são adultas, e vivem em agregados familiares chefiados por homens empregados, pertencentes aos quintis de menor renda e de religião católica, apresentaram menores taxas de vacinação.

Esses achados estão em consonância com a literatura. O INE e MISAU (2017), do recém-nascido, da criança e do adolescente indica que a cobertura vacinal tende a ser menor entre filhos de mães sem escolaridade e pertencentes a famílias em situação de pobreza. De acordo com Tertuliano e Stein (2025), factores como o desmame precoce, o desemprego e o ganho insuficiente de peso durante a gestação estão associados à baixa adesão ao calendário vacinal.

Para estimar os coeficientes do modelo logístico, aplicou-se o teste do Qui-quadrado com o intuito de identificar as variáveis associadas ao estado vacinal das crianças. Os resultados revelaram que as variáveis área de residência, número de visitas pré-natais, nível de escolaridade da mãe, ocupação materna, local do parto, quintil de riqueza, religião do agregado familiar, acesso ao cartão de vacinação e estado civil da mãe apresentaram p-valores inferiores ao nível de significância de 5%, sendo, portanto, estatisticamente significativas.

Após a identificação das variáveis estatisticamente significativas, foram estimados três modelos logísticos. O primeiro modelo incluiu todas as variáveis disponíveis na amostra; o segundo considerou apenas aquelas que se mostraram significativas no teste do Qui-quadrado; e o terceiro modelo foi construído com as variáveis que se mantiveram significativas nos dois modelos

anteriores. Para identificar o melhor modelo, aplicou-se a estatística Pseudo, e verificou-se que o modelo que apresentava melhor desempenho e o modelo final.

No terceiro modelo, as variáveis que apresentaram significância estatística foram: área de residência, número de visitas pré-natais, nível de escolaridade da mãe (secundário ou mais), ocupação materna, quintil de riqueza (rico), local do parto (hospital público), religião do agregado familiar (Zion), idade da mãe e exposição à mídia.

De acordo com Shenton *et al.* (2018), em estudo conduzido no Quênia com dados do DHS, os principais factores associados à vacinação infantil foram a área de residência, o índice de riqueza familiar e o local de parto, resultados que convergem com os achados do presente estudo.

Outro estudo, realizado por Sako *et al.* (2023), na Etiópia, também utilizando dados do DHS, identificou como variáveis significativamente associadas à vacinação: o nível de escolaridade da mãe, a idade e o estado civil materno, a posse de documentação vacinal e a suplementação com vitamina A.

A suplementação com vitamina A não foi incluída no presente estudo, uma vez que não é amplamente abordada na literatura voltada à cobertura vacinal como fator determinante principal.

O ajuste do modelo final foi avaliado por meio dos testes de Wald, Hosmer-Lemeshow e da curva ROC. O teste de Wald indicou que as variáveis incluídas no modelo foram estatisticamente significativas. Já o teste de Hosmer-Lemeshow revelou que o modelo apresenta um bom ajuste aos dados observados. Para complementar a avaliação, foi traçada a curva ROC, que demonstrou elevada capacidade discriminatória do modelo, com uma área sob a curva (AUC) de 0,8225, valor considerado excelente.

Adicionalmente, foi apresentada a matriz de confusão do modelo, juntamente com as métricas de desempenho. Os resultados indicaram que aproximadamente 80% das classificações realizadas pelo modelo foram corretas, evidenciando sua boa acurácia na predição do estado vacinal.

Em suma, os resultados desta pesquisa evidenciam que a adesão ao calendário vacinal infantil está fortemente associada a factores sociodemográficos, econômicos e de acesso aos serviços de saúde. Variáveis como Área de residência, Número de visitas pré-natais, Escolaridade e Ocupação da mãe, Quintil de riqueza, Local do parto, religião do agregado familiar, Idade da mãe e Exposição à mídia demonstraram influência significativa sobre o estado vacinal das crianças.

Tais achados estão em consonância com estudos conduzidos em outros países africanos, como Quênia e Etiópia, reforçando a importância do contexto familiar e das condições de vida na determinação da cobertura vacinal. Além disso, o modelo final apresentou bom desempenho estatístico e preditivo, com elevada capacidade discriminatória (AUC = 0,8225) e acurácia de 80%, indicando sua adequação para análise e interpretação dos factores associados.

Capítulo 5

Conclusão e recomendações

5.1 Conclusão

Conforme os objectivos traçados neste estudo, foram achados os seguintes resultados: a amostra foi composta por 55.25% de crianças do sexo feminino e 44.75% do sexo masculino, das quais 64.24% estavam completamente vacinadas e 35.76% parcialmente vacinadas. Maior parte das crianças estava em zonas rurais (81,37%), 75,59% das mães realizaram pelo menos uma consulta pré-natal, maior parte dos partos ocorreu em casa.

A análise bivariada, conduzida com o teste de Qui-quadrado, permitiu identificar associações estatisticamente significativas entre o estado vacinal e variáveis como área de residência, número de visitas pré-natais, escolaridade e ocupação materna, local do parto, quintil de riqueza, religião do agregado, acesso ao cartão de vacinação e estado civil da mãe.

Na etapa final, a regressão logística revelou que as variáveis fortemente associadas ao estado vacinal completo foram: área de residência, visitas pré-natais, escolaridade materna (nível secundário ou superior), ocupação da mãe, quintil de riqueza (rico), parto em hospital público, religião (Zion), idade da mãe e exposição à mídia.

5.2 Recomendações

- Recomenda-se a realização de um inquérito comunitário nacional específico, com foco na vacinação das crianças;
- Recomenda-se a intensificação de campanhas de sensibilização sobre a importância da imunização, especialmente direccionadas à zonas rurais;
- Recomenda-se o fortalecimento da capacitação de profissionais de saúde em zonas rurais para ampliar o acesso e a qualidade das consultas pré-natais e programas de educação vacinal.
- Reforçar o investimento em infraestrutura e serviços de saúde em áreas de difícil acesso;

- Desenvolver políticas públicas que valorizem o uso e o acompanhamento do cartão de vacinação;
- Implementar campanhas de comunicação e sensibilização por meio de mídia local e comunitária, enfatizando a importância da vacinação completa e combatendo desinformação.

Limitações

A principal limitação deste estudo foi o elevado número de valores omissos em variáveis fundamentais para a condução da análise, especialmente aquelas relacionadas ao registro das vacinas. Observou-se que grande parte desses dados ausentes correspondia a crianças residentes em zonas rurais, o que pode indicar dificuldades específicas no acesso aos serviços de vacinação nessas áreas. Essa limitação não apenas reduziu o tamanho efetivo da amostra após a aplicação da técnica de exclusão de casos completos, como também pode estar relacionada a qualidade dos inquéritos.

Outra limitação está relacionada à qualidade dos dados, uma vez que as informações sobre a imunização infantil foram obtidas por meio de cartões de vacinação, documentos semelhantes ou, na ausência destes, com base na lembrança das mães. Esse tipo de colecta pode estar sujeito a falhas de memória ou à tendência de fornecer respostas socialmente desejáveis.

Referências Bibliográficas

- [1] Agresti, A. (2007). *Categorical Data analysis*. University of Florida.
- [2] Agresti, A. (2013). *Categorical Data Analysis*. Wiley, 3rd edition.
- [3] Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle 2nd International Symposium on Information Theory*. Akademiai Ki à do, Budapest.
- [4] Araújo, M., Silva, L., Balduino, L., Porto, T. N., De Sousa Martins, V., Carvalho, D. P., e De Sousa Neto, B. P. (2020). Fatores que interferem no cumprimento do calendário vacinal na infância. *Revista Eletrônica Acervo Saúde*.
- [5] Arter, A. (2024). Exploring factors influencing childhood immunization status in east africa using spatial and multilevel ordinal logistic regression. Master's thesis.
- [6] Bedford, H. e Lansley, M. (2007). More vaccines for children? parents' views. *vaccine. Epub 2007*.
- [7] Begum, S. and Moniruzzaman, M. (2017). Determinants of full vaccination coverage among children aged 12–23 months in bangladesh. *PLOS ONE*.
- [8] Bello, L. H. (2010). *Modelagens em experimentos Mistura processo par otimização de processos industriais*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- [9] Bender, R. e Benner, A. (2000). Calculating ordinal regression models in sas and s-plus. *Biometrical Journal*.
- [10] Burnham, K. P. e Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*.
- [11] Cabral, C. I. S. (2013). Aplicação do modelo de regressão logística num estudo de mercado. Master's thesis, Universidade de Lisboa, Faculdade de Ciências, Departamento de Estatística e Investigação Operacional.
- [12] Castro, H. M. e Ferreira, J. C. (2022). Linear and logistic regression models: when to use and how to interpret them. *Jornal Brasileiro de Pneumologia*.
- [13] Castro, H. M. and Ferreira, J. C. (2022). Linear and logistic regression models: when to use and how to interpret them. *Jornal Brasileiro de Pneumologia*.

- [14] Corrar, L. (2011). *Análise multivariada: para cursos de administração, ciências contábeis e economia*. Master's thesis, Fundação Instituto de Pesquisas Contábeis, Atuárias e financeiras.
- [15] Cox, D. R. and Snell, E. J. (1989). *Analysis of binary data*. Chapman and Hall/CRC.
- [16] Cox, D.R.; Snell, E. (1989). *Analysis of Binary Data*. London: Chapman and Hall. 2ª Edição.
- [17] Cramer, J. S. (2003). *Logit models from economics and other fields*. PhD thesis, Cambridge: Cambridge University.
- [18] Danasingh, A. A., Balamurugan, S., and Epiphany, J. (2016). Literature review on feature selection methods for high-dimensional data. *International Journal of Computer Applications*, 136.
- [19] Dash, M. and Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence*.
- [20] Dobson, A. J. and Barnett, A. (2008). *An introduction to generalized linear models*. Chemical Rubber Company press.
- [21] Draper, N. R. and Smith, H. (1981). *Applied regressions analysis*. 3rd ed. Wiley series in probability and statistics text and references section.
- [22] Dummer, L., Cui, Y., Strang, R., and Parker, L. (2012). Immunization completeness of children under two years of age in Nova Scotia, Canada. *Public Health*.
- [23] Emiliano, P. C. (2009). *Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano*. PhD thesis, Universidade Federal de Lavras.
- [24] Figueiro, C. V. (2006). *Modelos de Regressão logística*. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- [25] Garson, G. (2011). Multiple regression: Overview. statnotes: Topics in multivariate analysis. PR.
- [26] Gil, A. C. (2002). *Como classificar as pesquisas. como elaborar projetos de pesquisa. Projectos de Pesquisa*.
- [27] Gonzalez, A. L. (2018). *Regressão logística e suas aplicações*. Technical report, Universidade Federal de Maranhão.
- [28] Hair, J. (1998). *Multivariate analyses data*. New Jersey: Princeton University Press.
- [29] Hauck, W. W. e Donner, A. (1977). Wald test as applied to hypotheses in logit analysis. *American Statistical Association*.
- [30] Hosmer, D. W. e Lemeshow, S. Sturdivant, R. X. (2013). *Applied Logistic Regression (3rd ed.)*.

- [31] Hosmer, J. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley e Sons.
- [32] Hosmer Jr, D. W. e Lemeshow, S.e Sturdivant, R. X. (1989). *Applied logistic regression 1st ed.* John Wiley and Sons.
- [33] Hurvich, C. M. and Tsai, C. L. (1989). *Regression and Time Series Model Selection in Small Samples*. Biometrika.
- [34] INE and ICF (2011). *Inquérito Demográfico e de Saúde de Moçambique 2011*. INE e ICF International, Calverton, Maryland, EUA.
- [35] INE and ICF (2023). Inquérito demográfico e de saúde em Moçambique 2022-23. Technical report, Maputo, Moçambique e Rickvile, Maryland, EUA.
- [36] INE e MISAU (2017). Relatório sobre saúde e bem-estar da mulher, do recém-nascido, da criança e do adolescente. IIP-JHU.
- [37] Kahn, H. and Sempos, C. (1989). Statistical methods in epidemiology (monographs in epidemiology and biostatistics). Oxford University Press, Oxford.
- [38] Kalra, A. (2016). The odds ratio: Principles and applications. *Journal of the Practice of Cardiovascular Sciences*, 2:49.
- [39] Kutner, C. j. e Nachtsheim, J. N. (2004). *Applied linear regression models*. 4th ed.
- [40] Levin, J., James, A. F., e Forde, D. R. (2012). *Estatística para Ciências humanas*.
- [41] Maroco, J. (2007). *Analise Estatística com Utilizacao do SPSS*. 5a edicao. Lisboa. Edições Silabo.
- [42] Massango, E. T. (2018). Analise dos factores associados a desnutrição infantil.
- [43] Mcquarrie, A. D. R. and Tsai, C. L. (1998). Regression and time series model selection. *World Scientific*.
- [44] Medri, W. (2011). *Analise exploratória dos dados*. PhD thesis, Universidade Estadual de Londrina.
- [45] Mesquita, P. S. (2014). um modelo de regressão logística para avaliação dos programas de pos-graduação no brasil. *Rio de Janeiro, Brasil*.
- [46] Nascimento, F. P. (2016). Classificacao da pesquisa. natureza, método ou abordagem metodologica, objetivos e procedimentos. Brazil arquivos.
- [47] OMS (2020). Immunization agenda 2030: a global strategy to leave no one behind [internet]. *Geneva: World Health Organization*.

- [48] OMS (2022). Atlas das estatísticas da saúde em África: Análise da situação sanitária da região africana da oms, 2022 — relatório de síntese. Technical report, Escritório Regional da OMS para a África.
- [49] OMS (2023). A global strategy to leave no one behind. Technical report, Immunization Agenda 2030.
- [50] ONU (2019). Cobertura de vacinas em África “estagnada há cinco anos”. Technical report, Nações Unidas.
- [51] Peretti-Watel, P. and Cortaredona, S., Ly, E. Y., Seror, V. and Ndiaye, S., Gaye, I., e Fall, M. (2020). Determinants of childhood immunizations in senegal: Adding previous shots to sociodemographic background. *Human vaccines and immunotherapeutics*.
- [52] Pestana, H. e Gageiro, J. (2014). *Análise de dados para ciências sociais: A Complementaridade do SPSS 6ª Edição Revista, Atualizada e Aumentada* MARIA HELENA PESTANA JOÃO NUNES GAGEIRO. Research Gate.
- [53] Phukan, R. K., Barman, M. P., e Mahanta, J. (2008). Factors associated with immunization coverage of children in assam, india: Over the first year of life. *J Trop Pediatr*.
- [54] Raimundo, J., Echeimberg, J. O., e Leonel, C. (2018). Tópicos de metodologia de pesquisa: Estudos de corte transversal. *Journal of Human Growth and Development*.
- [55] Restrepo-Mendez, M. C., Barros, A. J. D., Wong, K. L. M., Hosseinpoor, A. R., Victora, C. G., e Boerma, T. (2016). Inequalities in full immunization coverage: trends in low and middle-income countries. *Bulletin of the World Health Organization*, 94(11):794–805B.
- [56] Sako, S., Gilano, G., e Hailegebreal, S. (2023). Determinants of childhood vaccination among children aged 12–23 months in ethiopia: a community-based cross-sectional study. *BMJ Open*, 13(3). Santos, C. e Silva, M. (2025). Redução na taxa de imunização infantil e cobertura vacinal nos Últimos anos no brasil e análise revisional das suas principais causas. *Revista Multidisciplinar do Nordeste Mineiro*, 5:1–11.
- [57] Santos, L., Barreto, C. C., e Silva, F. L., a. O. s. K. C. (2011). Percepção das mães quanto à importância da imunização infantil. *Rev Rene, Fortaleza*.
- [58] Schafer, J. L. e Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*.
- [59] Schwarz, G. (1978). *Estimating the dimension of a model. The annals of statistics*. Institute of Mathematical Statistics.
- [60] Sequeira, A. e Ndzava, A. (2010). Os serviços de abastecimento de água, saneamento e higiene em Moçambique. Technical report, WASHCost Mozambique. Revista Ndzava, Número 1.

- [61] Shenton, L. M., Wagner, A. L., Bettampadi, D. Masters, N. B., Carlson, B. F., e Boulton, M. L. (2018). Factors associated with vaccination status of children aged 12–48 months in India. *Matern Child Health J.*
- [62] Silva, G. L. (1992). *Modelo logístico para dados binários*. Mestrado em estatística. Instituto de Matemática e Estatística, USP.
- [63] Tabakhi, S., Moradi, P., and Akhlaghian, F. (2014). An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence*, 32:112–123.
- [64] Tauil, M. C., Sato, A. P., and Waldman, E. A. (2016). factors associated with incomplete or delayed vaccination across countries: A systematic review. *vaccine. Epub* 2016.
- [65] Tertuliano, G. C. e Stein, A. T. (2025). Atraso vacinal e seus determinantes: um estudo em localidade atendida pela estratégia saúde da família. *Associação Brasileira de Saúde Coletiva*.
- [66] Tesema, G. A., Tessema, Z. T., Koku, K. S., e Teshale, A. B. (2020). Complete basic childhood vaccination and associated factors among children aged 12–23months in east africa: a multilevel analysis of recent demographic and health surveys. *BMC Public Health*.
- [67] Touray, E., Barrow, A., Kinteh, B., Badjie, M. e Nget, M., Touray, J., e Ceesay, L. (2021). Childhood vaccination uptake and associated factors among children 12–23 months in rural settings of the gambia: a community-based cross-sectional study. *BMC Public Health*.
- [68] UNICEF (2009). *The State of the World's Children 2009: Maternal and Newborn Health*. UNICEF, New York, NY.
- [69] UNICEF (2023a). The state of the world's children 2023: For every child, vaccination. Technical report, Florence: UNICEF.
- [70] UNICEF (2023b). Unicef's the state of the world's children 2023. *For Every Child*.
- [71] Vaz, F. E. (2020). Aplicação da regressão logística no estudo de câncer da mama. SP.
- [72] Yadita, Z. S. e Ayehubizu, L. M. (2021). Full immunization coverage and associated factors.
- [73] Yokokura, A., Silva, A., Bernades, F., Filho, F. Alves, M., Cabra, N., e
- [74] Alves, R. (2013). Cobertura vacinal e fatores associados ao esquema vacinal básico incompleto aos 12 meses de idade. *São Luís Maranhão, Brasil, Saúde Publica*.