



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise de factores associados à prevalência
de casamentos prematuros em mulheres
entre 15 a 18 anos em Moçambique**

Autor: Cremildo Atanásio Malate

Maputo, Janeiro de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise de factores associados à prevalência
de casamentos prematuros em mulheres
entre 15 a 18 anos em Moçambique**

Autor: Cremildo Atanásio Malate

Supervisor: Bonifácio José, Msc, KEU

Maputo, Janeiro de 2025

Didicatória

Dedico este trabalho aos meus amados pais, Atanasio Luis Malate e Maria Lote Nhantumbo Malate, e às minhas filhas, Daniela Cremildo Malate e Ashiley Cremildo Malate, na esperança de que se sintam incentivados a prosseguir com os estudos.

Declaração de Honra

Declaro que por minha honra que sou o autor deste trabalho e que o processo foi concebido para ser submetido apenas para a obtenção para o grau de **Licenciatura em Estatística**, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Janeiro de 2025

(Cremildo Atanásio Malate)

Agradecimentos

Agradeço primeiramente a Deus por me amar de forma incondicional e sem medida, e por estar presente em minhas atividades cotidianas. Agradeço profundamente aos meus Pais, Atanasio Luis Malate e Maria Lote Nhantumbo Malate, por terem cuidado de mim, pelos ensinamentos, aconselhamento de nunca desistir de lutar pelos meus sonhos. Agradeço também à minha família pelo aconchego que necessito, fornecendo-me as forças renovadas para persistir na luta pela vitória em diversas áreas da minha vida. Os meus agradecimentos estendem-se a todos os docentes do DMI, em particular ao Dr. Jonas Nassabe e ao Dr. Bonifácio, pela paciência e disponibilidade em transmitirem os ensinamentos. Aos meus colegas de curso, vai o meu agradecimento pela dedicação aos estudos, em especial ao Paulo Notiço, que esteve diretamente presente de maneira incansável, incentivando para a conclusão deste trabalho. E finalmente agradeço a todos que de forma direta ou indiretamente apoiaram na concretização deste trabalho.

MUITO OBRIGADO!

“No momento em que uma rapariga é [forçada a se casar], as suas asas são cortadas e os seus sonhos são contaminados pelo resto da sua vida”

(Graça Machel - ODS)

Resumo

O casamento prematuro é entendido como a união com o objetivo de formar uma família, na qual pelo menos um dos cônjuges tem menos de 18 anos de idade. Embora seja um fenômeno presente em várias partes do mundo, sua ocorrência é mais frequente nos continentes asiático e africano. Este estudo tem como objetivo analisar os fatores associados aos casamentos prematuros em Moçambique, com especial enfoque nas idades entre 15 e 17 anos. A investigação foi realizada através de amostragem complexa, assegurando a representatividade dos dados e a robustez dos resultados. A metodologia envolveu a aplicação de regressão logística para identificar e quantificar a influência das variáveis sobre o risco de casamento precoce. Os resultados indicam que a religião e a ocupação são fatores significativos associados ao aumento do risco de casamento prematuro. A análise revelou que a religião islâmica está associada a uma probabilidade significativamente maior de casamento antes dos 18 anos, com uma odds ratio (OR) de 2,87, enquanto a religião Zion também apresentou um risco elevado, com OR de 3,54. Adicionalmente, a presença de uma ocupação foi identificada como um fator de risco, com mulheres empregadas a apresentarem uma probabilidade três vezes superior de casar prematuramente em comparação com aquelas sem ocupação. A análise de multicolinearidade demonstrou que não existem relações lineares significativas entre as variáveis independentes, confirmando a adequação do modelo de regressão logística. A avaliação do ajuste dos modelos revelou que a amostragem complexa melhorou a representatividade e reduziu a variabilidade, conforme indicado pelos valores de AIC mais baixos para o modelo ajustado. Os resultados fornecem insights valiosos para a formulação de intervenções eficazes e políticas sociais voltadas para a proteção dos direitos das adolescentes.

Palavras-chave: casamentos prematuros, Moçambique, amostragem complexa, regressão logística, fatores de risco, religião, ocupação, educação.

Abstract

Premature marriage is understood as a union with the purpose of starting a family, in which at least one of the spouses is under 18 years of age. Although it is a phenomenon present in various parts of the world, it occurs more frequently in the Asian and African continents. This study aims to analyze the factors associated with early marriages in Mozambique, specifically focusing on individuals aged 15 to 17. The investigation utilized complex sampling methods to ensure data representativeness and robustness of results. Logistic regression was applied to identify and quantify the impact of various factors on the risk of early marriage. The results indicate that religion and occupation are significant factors associated with an increased risk of early marriage. The analysis revealed that Islamic religion is associated with a significantly higher likelihood of marrying before the age of 18, with an odds ratio (OR) of 2.87, while Zionist religion also showed a high risk with an OR of 3.54. Additionally, having an occupation was identified as a risk factor, with employed women being three times more likely to marry early compared to those without employment. Multicollinearity analysis demonstrated that there are no significant linear relationships among the independent variables, confirming the adequacy of the logistic regression model. Model fit assessment showed that complex sampling improved representativeness and reduced variability, as indicated by lower AIC values for the adjusted model. This study underscores the importance of targeted strategies to combat early marriages in Mozambique, highlighting the need for public policies that promote education and economic empowerment of young women. The research also suggests the implementation of awareness campaigns to reduce the incidence of this practice. The findings provide valuable insights for formulating effective interventions and social policies aimed at protecting adolescent rights.

Keywords: early marriage, Mozambique, complex sampling, logistic regression, risk factors, religion, occupation, education..

Lista de abreviaturas

AIC	Critério de Informação Akaike
BIC	Critério de Bayseano de Schwarz
MLG	Modelos Lineares Generalizados
OR	Razão de chances
OMS	Organização mundial da saúde
UNICEF	Fundo das Nações Unidas para a Infância
ROC	Receiver Operator characteristic curve
IDS	Inquérito Demográfico e de Saúde
INE	Instituto Nacional de Estatística
UNFRA	United Nations sexual and reproductive health agency

Índice

1	Introdução	1
1.1	Contextualização	1
1.2	Definição do problema	4
1.3	Objectivos	5
1.3.1	Objectivo geral	5
1.3.2	Objectivos específicos	5
1.4	Relevância do Estudo	5
1.5	Estrutura do Trabalho	6
2	Revisão da Literatura	7
2.1	Epidemiologia de casamentos prematuros	7
2.1.1	Erradicação dos casamentos prematuros	10
2.2	Modelos Lineares Generalizados	11
2.2.1	Escolha das variáveis independentes	12
2.2.2	Modelos Particulares do MLG	12
2.2.3	Formulação dos modelos	13
2.2.4	Ajuste do modelo	14
2.2.5	Inferência	14
2.2.6	Regressão logística	14
2.2.7	Modelo de Regressão logística simples	16
2.2.8	Estimação dos parâmetros do modelo de regressão logística simples	17
2.2.9	Modelo de Regressão logística múltipla	18
2.2.10	Teste da Razão de Verossimilhança	19
2.2.11	Teste de Wald	20
2.2.12	Pseudo R^2 de Cox e Snell	20

2.2.13	Critério de informação Bayesiano (BIC)	21
2.2.14	Teste Nagelkerke R2	21
2.2.15	Teste de Pearson	21
2.2.16	Função Logit	22
2.3	Amostragem Probabilística ou Aleatórios	23
2.3.1	Elementos dos desenhos complexos de amostragem	25
2.3.2	Questões relacionadas a análise estatísticas de dados	25
3	Material e Métodos	28
3.1	Material	28
3.1.1	Fonte de dados	28
3.1.2	Descrição das variáveis	29
3.2	Métodos	30
3.2.1	Amostragem Complexa	30
3.2.2	Teste de independência	30
3.2.3	Estimação dos parâmetros	31
3.2.4	Razão de chances	32
3.2.5	Multicolinearidade	33
3.2.6	Curva de ROC	34
4	Resultados e Discussão	37
4.1	Prevalência do Casamento prematuro	37
4.2	Avaliação da multicolinearidade	40
4.2.1	Risco de ocorrência de casamentos prematuros	41
4.2.2	Avaliação do modelo	43
4.3	Discussão dos resultados	45
5	Conclusões e recomendações	47
5.1	Conclusões	47
5.2	Recomendações	48
	Referências	49
	Apêndices	54
A	Tabelas	55

Lista de Figuras

2.1	Rank de casamento prematuro por países (top 4), FONTE: UNICEF	8
2.2	Estatísticas gerais de casamento prematuros, FONTE OMS	9
3.1	Curva de ROC	35
4.1	Prevalência de casamentos Prematuros	37
4.2	Mapeamento da prevalência de casamentos prematuros	38
4.3	Distribuição dos casamentos prematuros em relação ao nível de escolaridade	38
4.4	Distribuição dos casamentos prematuros em relação a religião	39
4.5	Distribuição dos casamentos prematuros em relação ao acesso à meios de comunicação	40
4.6	Avaliação do modelo	44

Lista de Tabelas

2.1	Funções de ligação canônica	13
3.1	Descrição das variáveis em estudo	29
4.1	Teste de independência e avaliação de multicolinearidade	41
4.2	Estimativas do OR	42
A.1	Avaliação do ajuste dos modelos	55
A.2	Avaliação do modelo	55
A.3	Estatística descritiva sem amostragem complexa	56
A.4	Sumário do modelo	57

Capítulo 1

Introdução

1.1 Contextualização

O casamento prematuro emerge como um fenómeno de alcance global, notável e inquietante, apesar dos esforços empreendidos por governos, organizações não governamentais e a sociedade civil para erradicá-la. Segundo o UNICEF (2018), os dados mais recentes indicam uma queda nos índices de casamentos prematuros em escala mundial. Em 2008, uma em cada quatro mulheres se casava antes dos 18 anos, mas nos registos de 2018, essa proporção reduziu para uma em cada cinco mulheres, representando uma diminuição de 25% para 21%, respectivamente. A UNICEF projecta que entre 120 e 150 milhões de casamentos de mulheres com menos de 18 anos ocorrerão até 2030. Além disso, estima-se que 12 milhões de casamentos prematuros aconteçam anualmente, equivalendo a aproximadamente 23 meninas casando a cada minuto (UNICEF, 2018, 2011).

O continente Asiático encara uma realidade alarmante quanto aos casamentos prematuros, impulsionada por tradições profundamente enraizadas em nações como Bangladesh, Paquistão e Indonésia. No caso específico do Bangladesh, onde a incidência de casamentos prematuros atinge 52%, aproximadamente 73% das mulheres contraem matrimónio enquanto ainda são crianças, evidenciando uma prática arraigada nesse tipo de casamento prematuro. Esse mesmo padrão é observado no Paquistão e na Indonésia, com taxas de 34% e 38%, respectivamente, ressaltando o aspecto mais sombrio de uma questão cultural que tolhe a liberdade de homens e mulheres, perpetuando um ciclo de pobreza e dependência cultural (Ajuda em acção, 2018).

Apesar do cenário desafiador, as políticas implementadas pela Índia têm desempenhado um papel crucial na redução do risco de casamentos prematuros, com destaque para a promoção da educação entre adolescentes, visando mantê-las na escola, e a clara divulgação da ilegalidade desse fenômeno, com o intuito de desencorajar sua prática. Essas iniciativas influenciaram significativamente na diminuição da incidência de casamentos prematuros, que passou de 50% para 30% em todo o continente Asiático ao longo de um período de 10 anos (UNICEF, 2018). Outras realidades preocupantes emergem em diversas partes do mundo como na América Latina e no Caribe onde as taxas de casamentos prematuros alcançam 29%, enquanto em países como Arábia Saudita (onde a idade mínima para o casamento é de 10 anos) e Iêmen (onde não há legislação específica), não existe uma idade mínima estabelecida para o casamento. Essas estatísticas evidenciam a urgência de medidas globais para combater esse fenômeno e proteger os direitos das crianças em todo o mundo (UNICEF, 2018).

Na América do sul, o Brasil figura como o 4º país com maior incidência de casamentos Prematuros globalmente em números absolutos, conforme o mais recente relatório do Banco Mundial, e o líder absoluto na América Latina nesse aspecto. Tais uniões estão associadas a 30% da evasão escolar, à perpetuação da pobreza, ao aumento da violência doméstica e à ocorrência de gravidez na adolescência, entre diversos outros malefícios (Schmidt, 2023).

A questão do casamento prematuro persiste como uma preocupação significativa na região da África Austral, sendo influenciada por uma variedade de elementos, tais como a prevalência da pobreza, disparidades de gênero, arraigadas tradições culturais, turbulências em tempos de conflito, limitações nos sistemas educacionais e a carência de estruturas legais sólidas nos países membros, muitas das quais revelam inconsistências (Fórum Parlamentar da SADC, 2016). De acordo com Fórum Parlamentar da SADC (2016), em alguns países da África Austral, aproximadamente 40% dos jovens se casam antes de completarem 18 anos de idade. É digno de nota que globalmente Malawi, Zâmbia e Madagascar apresentam maiores taxas de casamentos prematuros que ultrapassam os 40%, segundo Lei Modelo sobre erradicação dos casamento prematuros e proteção da criança em casamento (Fórum Parlamentar da SADC, 2016).

De acordo o relatório da do fundo internacional de emergência das nações unidas para infância, estima-se que cerca de 700 milhões de mulheres contraíram matrimónio antes dos 18 anos, sendo mais de 75% contraíram matrimónio antes dos 15 anos. No continente Africano, regista se uma redução fraca na taxa de mulheres que se casaram antes dos 18 anos onde estima-se um decréscimo de 10%, isto é, de 44% para 34% no período de 1990 à 2016 respectivamente, segundo um perfil do casamento prematuro em África (UNICEF, 2018).

Moçambique apresenta altos índices de casamentos prematuros, onde 48% das meninas se casam antes dos 18 anos. Na região da África Austral, Moçambique fica somente atrás do Malawi, que ocupa a segunda posição. A posição de Moçambique nesse cenário é sustentada pela falta de divulgação das leis pertinentes, pela forte influência das tradições e pela desigualdade de gênero no acesso ao emprego, aos serviços, à saúde e à educação, como apontado pelo IDS (2011). Globalmente, Moçambique ocupa o 10º lugar entre os países com altas taxas de casamento prematuro, de acordo com o IDS (2011) e o IMASIDA (2015). A erradicação desse problema contribuiria em 15,6% para o crescimento econômico do país.

Segundo Osório (2015), UNICEF (2018) e MISAU (2012), diversas causas fundamentam os casamentos prematuros em Moçambique. A predominância da cultura patriarcal na sociedade moçambicana é apontada como a primeira razão, enquanto a modalidade e certos conteúdos dos ritos de iniciação, em particular, promovem uma mentalidade de conformidade com a cultura patriarcal. Além disso, a pobreza multidimensional, abrangendo renda e consumo, também desempenha um papel significativo. Notavelmente, o recurso ao casamento prematuro como uma solução para a pobreza acaba por perpetuá-la, em vez de mitigá-la.

Globalmente, nota-se uma diminuição nos casamentos prematuros, contrastando com a tendência inversa observada em Moçambique. Tal cenário é alarmante, já que o país apresenta uma proporção consideravelmente acima da média global nesse aspecto. Diante dessa realidade, este estudo visa analisar os factores associados a casamentos prematuros em mulheres dentre 15 a 18 anos.

1.2 Definição do problema

Segundo UNICEF (2018), em Moçambique uma em duas raparigas casam-se antes dos 18 anos, representando uma grande violação dos direitos humanos das raparigas (IDS, 2011; IMASIDA, 2015). Esta situação influencia negativamente os esforços para a redução da pobreza e o alcance dos Objectivos de Desenvolvimento do Milénio (ODMs) - em particular influenciando para que as raparigas grávidas precocemente e deixem ter acesso a educação (Bassiano & de Lima, 2018).

Os casamentos prematuros em Moçambique têm um impacto duplo na educação. Por um lado, levam à desistência das adolescentes, enquanto, por outro lado, contribuem para o baixo desempenho das mesmas. À medida que as adolescentes avançam nos anos escolares, a taxa de desistência segue uma tendência crescente. Comparativamente à taxa de desistência entre rapazes e raparigas na 1ª e 2ª classes, há uma diferença de 2% a favor dos rapazes, mas na 7ª classe, essa diferença aumenta para 20% (Selemane, 2019). Além disso, esse fenómeno tem um impacto significativo na saúde, com 2.000 novos casos de fístula obstétrica surgindo a cada ano (MISAU, 2018). Desde 2010, tem-se observado um aumento nos casos de fístula obstétrica em adolescentes e mulheres jovens (MISAU, 2018).

Vários estudos limitam-se nas questões culturais e sociais, poucos são estudos que descrevem o padrão demográfico e socioeconômicos principalmente quando se trata de adolescentes dos 15 aos 18 anos, pois esta faixa etária encontra-se exposta ao risco de ocorrência da gravidez precoce, na qual observa-se que 40% das mulheres desta faixa etária são expostas de gravidez precoce (IDS, 2011; UNFPA, 2013), sendo que 51% das mortes neonatais são em mães adolescentes dos 15 a 18 anos (UNFPA, 2013), contudo fase as previsões feitas pelo Fundo das Nações Unidas para a Infância sobre os casamentos prematuros ao nível mundial, o padrão de ocorrência de casamentos prematuros poderá se intensificar em Moçambique e os indicadores que mensuram o bem estar da mãe e criança serão afectadas, entretanto diante deste cenário surge a seguinte pergunta de pesquisa:

- Quais os factores associados à prevalência de casamento prematuros em mulheres entre os 15 a 18 anos em Moçambique?

1.3 Objectivos

1.3.1 Objectivo geral

Analisar os factores associados à prevalência de casamentos prematuros em mulheres entre 15 à 18 anos em Moçambique

1.3.2 Objectivos específicos

- Descrever o perfil sócio-demográfico das vítimas de casamento prematuros;
- Mapear a distribuição dos casos de casamentos prematuros em Moçambique;
- Identificar os factores de risco associados à ocorrência de casamentos prematuros.
- Estimar o modelo para prever a probabilidade de ocorrência de casamentos prematuros com base na descrição socio-demográfica;

1.4 Relevância do Estudo

O presente estudo reveste de importância, pois conhecendo os factores associados à prevalência dos casamentos prematuros em Moçambique, poderá apoiar o governo e as organizações não governamentais na tomada de decisões e implementação de políticas específicas, em planos de ações de forma geral e, em particular, em grupos de risco, visando reduzir ou erradicar o crescimento dos casamentos prematuros, trazendo ganhos em várias áreas nomeadamente:

- **Saúde Materna Infantil:** Casamentos realizados no tempo adequado contribuem para planeamento familiar e gestações saudáveis, promovendo o bem-estar da mãe e do bebê. As crianças nascidas de casamentos oportunos podem contar com melhores condições de desenvolvimento, evitando dificuldades associadas à pobreza e à falta de experiência no cuidado parental.
- **Educação e Emprego:** Os casamentos realizados no momento certo fortalecem o compromisso com a educação das jovens, proporcionando maior acesso ao conhecimento e desenvolvimento profissional. Isso amplia as oportunidades de emprego e reduz a vulnerabilidade a salários baixos ao longo da vida.

- **Direitos Humanos:** Os casamentos realizados no tempo adequado promovem o respeito aos direitos das adolescentes, incluindo o direito à saúde, educação e liberdade de escolha.
- **Empoderamento Feminino:** Obter informações sobre os fatores associados aos casamentos prematuros permite que a sociedade desenvolva planos de ação para promover a igualdade de gênero e capacitar a jovem na tomada de decisões informadas sobre sua vida.
- **Contribuição para o Desenvolvimento Sustentável:** A erradicação dos casamentos prematuros está alinhada com os objetivos do desenvolvimento sustentável (ODS 5), que busca alcançar a igualdade de gênero e empoderar todas as mulheres e meninas.

1.5 Estrutura do Trabalho

A estrutura organizacional deste trabalho será composta pelos seguintes elementos:

- **Capítulo 2: Revisão de Literatura**

Neste capítulo, são discutidas sucintamente a prevalência de casamentos prematuros globalmente, com foco especial na África e em particular Moçambique, além das técnicas estatísticas que serão empregadas.

- **Capítulo 3: Material e Métodos**

Neste capítulo, são detalhados o material e a metodologia de análise utilizados para atingir os objetivos descritos na seção 1.3.

- **Capítulo 4: Resultados e Discussão**

Neste capítulo, são apresentados os resultados e uma breve análise dos achados obtidos por meio das técnicas e materiais descritos no capítulo 3.

- **Capítulo 5: Conclusões e Recomendações**

Neste capítulo, são discutidas as conclusões, as sugestões e as restrições identificadas ao longo da realização do estudo.

Capítulo 2

Revisão da Literatura

Neste capítulo serão abordadas questões relevantes aos casamentos prematuros e os seus factores assim como as técnicas que podem ser aplicadas neste contexto.

2.1 Epidemiologia de casamentos prematuros

No continente americano, o Brasil se destaca como o quarto país do mundo com o maior índice de casamentos de crianças e adolescentes do sexo feminino. Essa realidade afeta mais de 554 mil meninas com idades entre 10 e 17 anos no Brasil, sendo que mais de 65 mil delas têm entre 10 e 14 anos (PLAN INTERNACIONAL, 2018).

Segundo **Tavares (2017)**, No ranking mundial, o Brasil ocupa o quarto lugar em termos de número absoluto de casamentos prematuros e o primeiro lugar na região da América Latina. Não obstante ter uma das legislações mais avançadas do mundo. As Projeções indicam que 11% das mulheres se casam antes dos 15 anos e 36% antes dos 18 anos. Além disso, de acordo com o censo do IBGE de 2010, 88 mil crianças de até 14 anos estão envolvidas em casamentos, sendo que sete mil delas estão em uniões aceitas e amparadas pelo estado ou igreja.

Na América Latina e Caribe, o Brasil fica entre os 5 Países com índices mais altos, com 36%, em ordem decrescente Nicarágua 41%, república Dominicana 37%, Honduras e México 23%. Isto quer dizer que uma em cada quatro meninas dessa região se casam antes dos 18 anos, segundo Conforme GIRLS NOT BRIDES (n.d.), Na América Latina e no Caribe, cinco países se destacam pela alta prevalência de casamentos prematuros: Nicarágua (41%),

Brasil (36%), República Dominicana (37%), Honduras e México (23%). Assim, uma em cada quatro meninas nessa região se casa antes dos 18 anos de idade.

No continente Asiático, a Índia possui legislação que criminaliza os casamentos prematuros; no entanto, as projeções indicam que 27% das meninas indianas se casam antes de completar 18 anos, e 7% antes dos 15 anos. Apenas 4% dos meninos se casam antes de atingir os 18 anos. Essa prática é mais comum em áreas urbanas e nas regiões central e ocidental do país (GIRLS NOT BRIDES, n.d.).

O Bangladesh é o segundo país com maior prevalência de casamentos prematuros, onde 59% das meninas se casam antes de completar 18 anos de idade e 22% antes de atingir os 15 anos. Apenas 4% dos meninos se casam antes dos 18 anos. Registros indicam que a violência contra meninas e mulheres aumentou significativamente nos últimos 10 anos, com o número de casos de violência sexual dobrando. Muitas famílias veem o casamento como um meio de proteção nessas circunstâncias (GIRLS NOT BRIDES, n.d.).

Conforme GIRLS NOT BRIDES (n.d.), no continente Africano, a Nigéria se destaca por ser o terceiro país do mundo com maior prevalência de casamentos prematuros. Cerca de 43% das meninas se casam antes dos 18 anos e 16% antes dos 15 anos. No entanto, apenas 3% dos meninos se casam antes de completar 18 anos. Uma causa comum nesses países é a ocorrência de conflitos armados.



Figura 2.1: Rank de casamento prematuro por países (top 4), FONTE: UNICEF

Segundo a Fundo das Nações Unidas para a Infância em Moçambique (2016), A redução da prevalência dos casamentos prematuros na África não apresenta progressos encorajadores, tendo caído apenas 10% desde 1990. Naquele ano, a prevalência estava em 44%, enquanto atualmente está em 36%. Se não forem implementadas medidas que gerem resultados positivos, é esperado que até 2050 o continente africano tenha um número ainda maior de crianças e adolescentes em casamentos prematuros do que a região sul do continente asiático, que é reconhecida mundialmente por essa prática.

Para Moçambique, o MISAU (2012) aponta que quase metade (48%) das mulheres pesquisadas na faixa etária entre 20 e 24 anos se casaram prematuramente, ou seja, antes da idade legal. O MISSAU & INE (2013) também afirmam que 14% das meninas se casaram antes dos 15 anos. Esses dados são corroborados pelo relatório da Fundo das Nações Unidas para a Infância e Fundo das Nações Unidas para a População (2016a), que coloca Moçambique como um dos países com maior taxa de casamentos prematuros, ocupando o segundo lugar no ranking do continente africano e o décimo no contexto mundial.



Figura 2.2: Estatísticas gerais de casamento prematuros, FONTE OMS

2.1.1 Erradicação dos casamentos prematuros

Com o objetivo de garantir o bem-estar de todas as crianças no mundo, especialmente as meninas, o UNICEF e o UNFPA recentemente anunciaram o Programa Global, que visa impulsionar as ações para erradicação dos casamentos precoces Fundo das Nações Unidas para a Infância e Fundo das Nações Unidas para a População (2016b) . Em Moçambique, um dos países participantes, esse programa está sendo implementado por meio da Estratégia Nacional para a Prevenção e Combate aos Casamentos Prematuros para o período de 2016-2019 Bassiano & de Lima (2018).

O Diploma Ministerial n. 46 (2008), em Moçambique, estabelece três encontros anuais, no mínimo que pode podem ser reaproveitados no colegiado para procurar melhores mecanismos que visem à eliminação de casamentos prematuros, que forçam o abandono por meninas que ainda se encontra no ensino primário. Ainda segundo Ibraimo (2014), reconhece o papel de destaque que a escola desempenha, ao defenderem que o Conselho de Escola deveria ser um espaço de participação da comunidade.

O Diploma Ministerial n. 46 (2008) em Moçambique estipula a realização de três encontros anuais, no mínimo, que são utilizados para discutir no âmbito educacional as formas de erradicar os casamentos prematuros, os quais têm impacto na evasão escolar ainda no ensino primário.

Conforme a Resolução 66/170 (2011), incentiva-se a reflexão sobre a situação das adolescentes em todo o mundo em 11 de outubro de cada ano, visando a promoção e proteção de seus direitos, para alcançar o desenvolvimento sustentável até 2030, conforme as metas estabelecidas pela ONU em escala global.

De acordo com Bassiano & de Lima (2018), o governo de Moçambique tem envolvido a sociedade civil e a mídia jornalística, principalmente em ocasiões comemorativas ou festivas, nas grandes cidades, onde o fenômeno não é tão comum. No entanto, devido à sua prevalência, especialmente em áreas rurais, e à associação com a pobreza e fatores socioculturais, reduzir as taxas de ocorrência tem sido um desafio. É necessário estabelecer parcerias entre entidades governamentais e não governamentais para planejar iniciativas conjuntas e mecanismos que desencorajem essas práticas.

Os casamentos prematuros são predominantemente praticados de forma ritualizada na África subsaariana e no sudoeste da Ásia. Nas Américas, as uniões matrimoniais geralmente têm um caráter mais informal. Isso dificulta a fiscalização e tende a ser ignorado nas agendas nacionais dos governos, conforme apontado pelo UNICEF (2014).

2.2 Modelos Lineares Generalizados

Na década de 1970, o surgimento dos modelos lineares generalizados impulsionou significativamente o avanço da estatística aplicada. Inicialmente, seu uso era limitado a alguns pesquisadores devido à escassez de literatura. Os modelos lineares generalizados foram concebidos para enfrentar desafios em que a variável de interesse não é quantitativa, ou seja, não atende aos requisitos dos modelos clássicos de regressão linear.

Segundo Turkman & Silva (2000), Os modelos lineares generalizados oferecem uma vantagem substancial ao proporcionar uma unificação por meio de um único algoritmo para resolver equações de verossimilhança. Este algoritmo se baseia em uma sequência de problemas de mínimos quadrados e se beneficia de técnicas amplamente testadas. Segundo Nelder & Wedderburn (1972), este método iterativo de mínimos quadrados é aplicável a qualquer modelo linear generalizado, requerendo apenas pequenos ajustes conforme a distribuição de probabilidade e a função de ligação consideradas.

Segundo Nelder & Wedderburn (1972), o modelo linear generalizado possibilita a unificação de diferentes técnicas estatísticas, que tradicionalmente eram tratadas separadamente. Esta abordagem unificadora na modelagem estatística, uma expansão dos modelos de regressão convencionais, é denominada de modelos Lineares Generalizados (MLG). Esses modelos consistem em uma única variável dependente, variáveis explicativas e uma amostra aleatória de n observações independentes, onde:

- A variável resposta, componente aleatório do modelo, segue uma distribuição que pode ser da classe normal, gama ou normal inversa para dados contínuos, binomial para proporções, e poisson ou binomial negativa para contagens.
- As variáveis independentes são integradas na forma de uma estrutura linear, constituindo a componente sistemática do modelo.

- Para conectar a componente aleatória e a componente sistemática, emprega-se uma função adequada, como a função logarítmica utilizada nos modelos log-lineares, também conhecida como função de ligação.

2.2.1 Escolha das variáveis independentes

Na prática, a seleção de um conjunto de variáveis explicativas para formar um modelo parcimonioso é uma tarefa complexa, dadas as combinações e questões estatísticas envolvidas. O desafio reside em determinar as combinações viáveis de variáveis explicativas a serem testadas para inclusão no preditor linear. Existem vários métodos de seleção de variáveis disponíveis, acompanhados de critérios para comparar a qualidade de ajuste dos modelos, como os métodos FORWARD, BACKWARD e STEPWISE, conforme, Abbad & Torres (2001). Segundo esses autores, o método BACKWARD começa com um modelo completo e, em seguida, remove as variáveis estatisticamente não significativas. Por outro lado, o método FORWARD parte do modelo inicial e adiciona variáveis de acordo com sua significância até que a inclusão da última não proporcione mais melhorias no ajuste do modelo. Por fim, o método STEPWISE combina os critérios anteriores de forma conjunta.

2.2.2 Modelos Particulares do MLG

Segundo Turkman & Silva (2000), É importante fazer referência a alguns modelos que representam casos particulares dentro do conjunto dos modelos lineares generalizados. Modelo de regressão linear clássico;

- Modelos de análise de variância e covariância;
- Modelo de regressão logística;
- Modelo de regressão de Poisson;
- Modelos log-lineares para tabelas de contingência multidimensionais;
- Modelo probit para estudos de proporções, etc.

Ainda segundo Turkman & Silva (2000), Considerando a função de ligação

$$g(\mu_i) = \theta_i = \eta_i \quad (2.1)$$

O preditor linear modela diretamente o parâmetro canônico θ_i e é denominado função de ligação canônica. Os modelos correspondentes são chamados de modelos canônicos. Isso geralmente resulta em uma escala adequada para a modelagem com interpretação prática dos parâmetros de regressão, além da importância teórica em termos da existência de um conjunto de estatísticas suficientes para o vetor de parâmetros β e alguma redução na complexidade do algoritmo de estimação.

$T = x^T Y$ com componentes $T_r = \sum_{i=1}^n x_{ir} Y_i, r = 1, \dots, p$. As funções de ligação canônicas para as principais distribuições são apresentadas na Tabela abaixo:

Tabela 2.1: Funções de ligação canônica

Distribuição	Função de ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \log \mu$
Binomial	Logística: $\eta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca do quadrado: $\eta = \frac{1}{\mu^2}$

O processo de adequação dos modelos generalizados pode ser dividido em três fases distintas: (i) formulação dos modelos, (ii) ajuste dos modelos e (iii) inferência. Esses modelos são relevantes devido à flexibilidade na etapa, (i) formulação do modelo, à computação simples no ajuste(ii) e a critérios razoáveis (iii).

2.2.3 Formulação dos modelos

Durante a formulação dos modelos, são feitas escolhas críticas relacionadas à distribuição de probabilidade da variável dependente, às covariáveis (matriz modelo) e à função de ligação. Essas escolhas desempenham um papel crucial na descrição precisa das características principais da variável dependente.

Para fazer uma escolha de distribuição que seja aceitável, é essencial examinar minuciosamente os dados, levando em consideração aspectos importantes, como assimetria, natureza

contínua ou discreta (contagens) e intervalo de variação, segundo Turkman & Silva (2000).

2.2.4 Ajuste do modelo

Durante a fase de ajustes, ocorre a estimativa dos parâmetros lineares dos modelos, assim como de certas funções relacionadas às estimativas desses parâmetros, refletindo a adequação dos valores estimados. Existem vários métodos disponíveis para estimar os parâmetros dos modelos lineares generalizados, sendo o método de máxima verossimilhança amplamente escolhido devido à sua eficácia e relativa simplicidade. De acordo com Cordeiro & Demétrio (2013).

2.2.5 Inferência

O objetivo principal da fase de inferência é avaliar a adequação global do modelo e examinar minuciosamente possíveis discrepâncias locais. Se essas discrepâncias forem significativas, podem indicar a necessidade de mudança de modelo ou identificação de outliers. Independentemente da escolha, isso implica em repetir todo o procedimento metodológico.

Ainda na fase de inferência, o analista deve investigar a precisão e a interdependência das estimativas, construir intervalos de confiança e realizar testes sobre os parâmetros de interesse, analisar estatisticamente os resíduos e fazer previsões. A precisão das previsões depende particularmente do modelo escolhido, para o qual temos critérios de avaliação de ajuste, verificando se a precisão de uma previsão específica é maximizada. Em muitas ocasiões, é possível otimizar a precisão modificando apenas o componente sistemático do modelo, segundo Turkman & Silva (2000)

2.2.6 Regressão logística

A análise de regressão logística tem suas raízes no século XIX, inicialmente direcionada à descrição do crescimento demográfico. Embora tenha começado a ganhar destaque nos anos 50, foi nos trabalhos de Cox & Snell (1989) e Hosmer & Lemeshow, que a regressão logística se destacou na literatura, especialmente entre os estatísticos. Aspectos teóricos desse modelo são extensivamente discutidos na literatura, com contribuições significativas de autores como

Kleinbaum & Klein (2002); Agresti (2002); Hosmer & Lemeshow (2000); Cox & Hinkley (1979).

A regressão logística é uma técnica estatística essencial para analisar variáveis qualitativas dicotômicas em um conjunto de variáveis independentes, permitindo a construção de modelos de previsão. Amplamente empregada em diversos contextos, mesmo quando a variável dependente não é originalmente binária, através da binarização da variável resposta para modelar a probabilidade de sucesso com a regressão logística. Essa abordagem é escolhida principalmente devido à clareza na interpretação dos parâmetros do modelo logístico e sua aplicabilidade em análises discriminatórias, Paula (2004).

Segundo Mesquita (2014), A regressão logística teve inicialmente sua aplicação e desenvolvimento restritos, principalmente na área da medicina. Devido à sua eficácia comprovada, ela se expandiu consideravelmente para outras áreas do conhecimento, desde as relacionadas à saúde até os campos de estudo de mercado e até mesmo na esfera política, abordando questões de intenção, e também foi adotada pelo setor bancário para avaliação de crédito, entre outros. Sua aplicação tem sido amplamente reconhecida, e essa diversidade de uso tem solidificado sua posição como uma ferramenta crucial na análise de dados categóricos. Isso tem levado à sua crescente notoriedade e aceitação entre os usuários de diversas técnicas de análise de dados.

A regressão logística é um dos principais modelos utilizados na análise de dados com variáveis de resposta binárias ou dicotômicas. Mesmo quando a variável de resposta não é inicialmente binária, é viável transformá-la em uma variável dicotômica para estimar a probabilidade de sucesso por meio de um modelo de regressão logística. Entre os modelos disponíveis para dados dicotômicos, a regressão logística se destaca pela sua flexibilidade matemática, facilidade de implementação e interpretação intuitiva de seus parâmetros, segundo Hosmer e Lemeshow (2000).

Na regressão logística, a variável dependente Y é geralmente binária, seguindo uma distribuição de Bernoulli com uma probabilidade desconhecida P . É importante destacar que a distribuição de Bernoulli é uma forma particular da distribuição binomial, em que $n=1$, representando a realização de um único experimento, segundo Belfiore (2015).

Na regressão logística, a probabilidade desconhecida P é estimada por meio de uma combinação linear das variáveis explicativas. É importante notar que a probabilidade de sucesso é restrita a valores entre 0 e 1 ($0 \leq p \leq 1$), enquanto o fracasso é representado por $q = 1 - p$.

2.2.7 Modelo de Regressão logística simples

Segundo Cabral (2013), O modelo de regressão logística simples ou univariada é Caracterizado por uma variável dependente Y , que é binária ou dicotômica, seguindo uma distribuição binomial e dependente de apenas uma variável explicativa. A variável Y pode variar entre 0 e 1, onde o valor 1 denota o evento de interesse, conhecido como "sucesso", enquanto o valor 0 representa o evento complementar, "fracasso", com suas respectivas probabilidades:

$$\pi_i = P(Y = 1|X = x_i) \quad (2.2)$$

e

$$1 - \pi_i = P(Y = 0|X = x_i) \quad (2.3)$$

De maneira a justar a resposta média ao modelo linear, é utilizada a função de ligação. na regressão linear, a variável respostas é expressa através de uma função $Y_x = E[Y|X = x] + \varepsilon_x$, onde o termo de erro denota o desvio de uma observação em relação á média. Partindo do pressuposto de que esse erro segue uma distribuição normal com média zero e variância constante. No entanto, quando lidamos com uma variável dicotômica, essa suposição não se mantém. o erro assume apenas dois valores: $Y = 1$ com probabilidade (x_i) e $Y = 0$ com probabilidade $1 - \pi(x_i)$.

Conforme Fávero et al. (2009), a probabilidade de sucesso no modelo logístico simples é expressa como:

$$\pi(x_i) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \quad (2.4)$$

E a probabilidade de fracasso é dada por:

$$1 - \pi_i = P(Y = 0|X = x_i) = 1 - \frac{1}{1 + e^{\beta_0 + \beta_1 X_i}} \quad (2.5)$$

O modelo logístico com a transformação logit é dado por:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_i \quad (2.6)$$

A interpretação dos parâmetros no modelo de regressão logística é fundamentada na comparação das probabilidades de sucesso e fracasso, através do uso da razão de chances (odds ratio), a qual é derivada da função de odds:

$$g(x) = \frac{\pi(x)}{1-\pi(x)} = \frac{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}{1-\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}} = \frac{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}{\frac{1}{1+e^{\beta_0+\beta_1 X_i}}} = e^{\beta_0+\beta_1 X_i} \quad (2.7)$$

porém, a regressão logística tem como objectivo a estimação do valor de P, aplicando o antilogaritmo:

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 X_i} \quad (2.8)$$

$$\hat{p} = \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}} \quad (2.9)$$

2.2.8 Estimação dos parâmetros do modelo de regressão logística simples

segundo Vaz (2020), Para ajustar o modelo, é necessário estimar os parâmetros β_0 e β_1 , utilizando métodos de estimação de máxima verossimilhança. A estimação por máxima verossimilhança permite encontrar os estimadores dos parâmetros do modelo que têm maior probabilidade de reproduzir o padrão das observações nos dados da amostra.

O modelo é desenvolvido através de um processo iterativo. No início, são atribuídos valores arbitrários aos coeficientes de regressão, formando um modelo inicial para prever os dados observados. Em seguida, os erros dessa previsão são avaliados e os coeficientes de regressão são ajustados para maximizar a probabilidade dos dados observados sob o novo modelo. Esse procedimento é repetido até que as discrepâncias entre o modelo mais recente e o anterior sejam minimizadas, (Mesquita, 2014) referencia citada.

Seja $\beta = (\beta_0, \beta_1)$ o vetor de coeficientes, e sejam as probabilidades $P(y_i = 1|x_i) = \pi(x_i)$ e $P(y_i = 0|x_i) = 1 - \pi(x_i)$. Então, para os pares (x_i, Y_i) onde $Y_i = 1$, a contribuição para a

função de verossimilhança é $\pi(x_i)$, e para os pares onde $Y_i = 0$, a contribuição para a função de verossimilhança é $1 - \pi(x_i)$, representando o valor de $\pi(x_i)$ avaliado em X_i . Segundo Figueira (2006).

A função de verossimilhança é dada pela seguinte equação:

2.2.9 Modelo de Regressão logística múltipla

De acordo com Vaz (2020), comparado ao modelo de regressão logística univariada, o modelo de regressão logística múltipla permite ajustar um modelo com uma variável resposta considerando mais de uma variável explicativa (covariável). Considera-se um conjunto de P covariáveis $\{x_1, x_2, x_3, \dots, x_p\}$, onde a probabilidade do acontecimento de interesse ocorrer é representada pela definição $P(Y = 1|x) = \pi(x)$ e a função de ligação ou logit é dada, nomeadamente por:

$$P(Y = 1|x) = \pi(x) \quad (2.10)$$

A função de ligação ou logit é dada por:

$$\text{logit}(\pi(x)) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) \quad (2.11)$$

Onde $E[Y]$ é designado por chance ou probabilidade de sucesso, e analogamente a probabilidade de fracasso vai ser dada pela expressão:

$$1 - E(Y) = 1 - \pi(x) = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} = \frac{1}{1 + e^{x_i^T \beta}} \quad (2.12)$$

Para o modelo múltiplo, a função de ligação logit é expressa pela seguinte equação:

$$f(x) = \ln \left[\frac{\pi_i}{1 - \pi_i} \right] = x_i^T \beta = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (2.13)$$

Portanto, para o logaritmo da função verossimilhança, pode-se escrever:

$$L(\beta) = \sum_{i=1}^n \left[y_i x_i^T \beta - \ln(1 + e^{(x_i^T \beta)}) \right] \quad (2.14)$$

2.2.10 Teste da Razão de Verossimilhança

Após verificar o ajuste do modelo, é crucial testar sua significância através do teste da razão de verossimilhança. Segundo Nelder e Wedderburn (1972), esse teste avalia se os coeficientes da regressão, excluindo o β_0 , são todos nulos de forma conjunta. A comparação entre os valores observados e esperados é realizada utilizando a função de verossimilhança, expressa da seguinte forma:

$$D = 2 \ln \left[\frac{\text{verossimilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right] \quad (2.15)$$

$$D = -2 \sum_{i=1}^n [y_i \ln(\pi_i/y_i) + (1 - y_i) \ln((1 - \pi_i)/(1 - y_i))] \quad (2.16)$$

Um modelo é considerado saturado quando incorpora todas as variáveis, enquanto o modelo ajustado inclui somente as variáveis relevantes para o estudo. A função D , também conhecida como deviance (desvio), é sempre positiva, sendo um valor menor indicativo de um ajuste mais eficaz do modelo.

Hipótese:

$$H_0 : \beta_1 = \dots = \beta_t = 0 \quad \text{e} \quad H_1 : \exists j = 1, \dots, p \beta_j \neq 0$$

Assim, na Hipótese nula H_0 a ser testada, os parâmetros do modelo serão igualados a 0 (zero). O modelo saturado que mantém o valor de seus coeficientes representa a hipótese alternativa H_1 .

Para avaliar a significância de uma variável independente, analisa-se o valor de D com e sem a variável na equação. A mudança prevista no valor de D ao adicionar a variável independente no modelo é determinada por:

$$G = D \left(\frac{\text{modelo sem a variável}}{\text{modelo com a variável}} \right) \quad (2.17)$$

Segundo Fahrmeir e Tutz (1994), como citado por Turkman & Silva (2000), recomenda-se o uso da estatística de razão de verossimilhança quando há poucas covariáveis e as amostras têm tamanho moderado. No entanto, em amostras grandes, essas estatísticas podem produzir resultados muito próximos. Nesses casos, é preferível optar por testes como W ou U , que são mais simples e rápidos de calcular.

2.2.11 Teste de Wald

De acordo com Araujo (2017), assim como o teste de verossimilhança, o teste de Wald é empregado na regressão logística para avaliar a significância dos coeficientes do modelo estimado. Este teste verifica se cada coeficiente é significativamente diferente de zero, permitindo avaliar se uma determinada variável independente possui uma relação estatisticamente significativa com a variável dependente.

Se os coeficientes logísticos forem estatisticamente significativos, podemos interpretá-los em termos de seu impacto na probabilidade estimada, ou seja, 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 do evento ($y = 0$).

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

$$W_j = \frac{\beta_j}{\text{var}(\beta_j)} \quad (2.18)$$

Hauck & Donner (1977) e Ferreira (2013) após examinar o desempenho do teste de Wald, constataram que, em alguns casos, ele se comporta de maneira inconsistente, falhando em rejeitar a hipótese nula mesmo quando o coeficiente é significativo. Portanto, é recomendável recorrer ao teste da razão de verossimilhança quando surgirem dúvidas quanto à eficácia do teste de Wald.

2.2.12 Pseudo R^2 de Cox e Snell

Segundo Ferreira (2013), o critério de informação de Akaike (AIC) permite a seleção de modelos derivados dos estimadores de verossimilhança. Ele é empregado para identificar o modelo mais apropriado ao comparar modelos aninhados. O processo de escolha do melhor modelo com base nesse critério é bastante simples: o modelo mais adequado é aquele que possui o menor valor de AIC. A Estatística AIC é definida como:

$$\text{AIC} = 2l(\beta) + 2(P + 1) \quad (2.19)$$

onde $L(\beta)$ é o logaritmo da função de verossimilhança do modelo e P é o número de parâmetros estimados.

2.2.13 Critério de informação Bayesiano (BIC)

O Critério de Informação Bayesiano, proposto por Schwarz (1978), assim como o AIC, o BIC oferece duas vantagens distintas sobre outros métodos de seleção. O BIC também se baseia na log-verossimilhança, semelhante ao AIC. Além disso, o BIC leva em consideração e penaliza a complexidade do modelo (ou seja, o número de parâmetros), aderindo ao princípio da parcimônia, fundamental na modelagem estatística. Além disso, os modelos comparados não precisam ser aninhados, o que é uma restrição significativa na comparação de modelos por meio de testes de hipóteses.

$$\text{BIC} = \frac{-2L}{N} + \frac{K \log n}{n} \quad (2.20)$$

onde: K é o número de restrições; n é o tamanho da amostra e L é a função de verossimilhança.

2.2.14 Teste Nagelkerke R^2

O indicador Cox-Snell R^2 leva em consideração o valor da verossimilhança e é situado em uma escala que varia de zero, mas não atinge o limite superior de 1. Por outro lado, o teste de Nagelkerke R^2 é um ajuste desenvolvido no indicador anterior para garantir que ele alcance o limite máximo de 1. Para Corrar et al. (2007), o teste de Nagelkerke R^2 situa-se numa escala que vai de 0 a 1 e tem a mesma finalidade do indicador Cox-Snell R^2 . Quando aplicado de modo prático, a principal diferença é que o teste Nagelkerke R^2 torna-se mais compreensível do que o Cox-Snell R^2 .

2.2.15 Teste de Pearson

Baseada nos resíduos de Pearson, a estatística de teste é obtida pela seguinte expressão:

$$\chi^2 = \sum_{j=0}^n [r_j(y_i \pi_i)]^2 \quad (2.21)$$

onde r_j representa o resíduo de Pearson para o j -ésimo elemento, conforme expresso pela seguinte equação:

$$r(y_i, \pi_i) = r_j = \frac{(y_i - \pi_i)}{\sqrt{\pi_i(1 - \pi_i)}} \quad (2.22)$$

Sob a suposição nula de adequação do modelo ajustado, a estatística de teste segue uma distribuição Qui-Quadrado com $n - (P + 1)$ graus de liberdade, onde P denota o número de covariáveis no modelo. No entanto, essa aproximação da distribuição Qui-Quadrado, utilizada no teste de Pearson, perde acurácia quando o número esperado de eventos por linha nos dados é reduzido. Assim, o teste de qualidade de ajuste de Pearson pode exibir imprecisões, especialmente em situações de dados no formato de resposta/frequência binária, segundo Freitas (2019).

Vantagens do modelo Logístico

- Simplificação na gestão de variáveis independentes categóricas;
- Resultados expressos em probabilidades;
- Facilidade na categorização de indivíduos;
- Requer poucas suposições;
- Altamente confiável;
- Uso quantitativas e qualitativas em simultâneo;
- De forma geral, permite um bom ajuste, com menos parâmetros do que modelos lineares.

2.2.16 Função Logit

Na regressão logística, a variável resposta é modelada conforme uma distribuição de Bernoulli. Para conectar as variáveis explicativas a essa distribuição presente na variável dependente, utilizamos o termo "Logit". Ao contrário dos problemas convencionais com distribuição de Bernoulli, na regressão logística, não dispomos de uma informação direta da probabilidade P . Portanto, o objetivo do modelo logístico é estimar essa probabilidade com base em uma combinação linear das variáveis independentes. O valor estimado de P é denotado por $\hat{P}$. De acordo com Powers & Xie (1999), é essencial empregar uma função que vincule a combinação linear de variáveis à distribuição de Bernoulli, garantindo que esta combinação retorne valores dentro de um intervalo entre 0 e 1, conforme a distribuição de probabilidade de Bernoulli. A razão de probabilidade é conhecida como "chance" ou "odds" em inglês, e o logaritmo natural dela é chamado de Logit.

De acordo com Goncalves (2012), a análise "Logit" ou regressão logística é uma técnica estatística que, por meio de ponderações, expressa a relação entre um conjunto de variáveis independentes e uma variável dependente. Essa variável resposta possibilita a conexão entre a probabilidade de ocorrência de um evento específico e a relevância das variáveis para esse evento. Em termos gerais, o objetivo da regressão logística é estabelecer uma expressão matemática ponderada que represente essa relação.

2.3 Amostragem Probabilística ou Aleatórios

Segundo Mattar (1996), A probabilidade de cada elemento da população fazer parte da amostra é igual, e todas as amostras selecionadas são igualmente prováveis. No entanto existe vários métodos de amostragem probabilísticas, sendo as mais utilizadas os seguintes:

1. **Amostragem aleatória simples:** Na abordagem clássica da inferência estatística, a amostragem aleatória simples é fundamental, garantindo que cada membro da população tenha uma chance igual e independente de ser selecionado. No entanto, a maioria dos inquéritos nacionais não a emprega devido a restrições orçamentárias e limitações de tempo relacionadas à coleta de uma grande quantidade de informações em um vasto território geográfico, Segundo Cochran (1977).
2. **Amostragem aleatória sistemática:** Segundo Barbetta (2006), esse método também é denominado de Pseudo-aleatório, pois não atribui a todas as amostras que podem ser selecionadas de uma população a mesma probabilidade de ocorrência. Para utilizá-lo, é necessário calcular o índice $K = N/n$ e a posterior, selecionar aleatoriamente um número dentro do intervalo $[1, K]$, que será utilizado como ponto de partida e primeiro elemento da amostra. Ao adicionar ao primeiro valor obtido o índice K (arredondando o resultado para baixo), obtemos o segundo elemento. o incremento sucessivo do mesmo razão permite encontrar os restantes elementos da amostra. Como se verifica, apenas o primeiro elemento é escolhido aleatoriamente, enquanto os restantes são determinados de modo sistemático pelo rácio.
3. **Amostragem aleatória estratificada:** A amostragem estratificada consiste em dividir a população em grupos mais homogêneos, chamados estratos, para identificar suas características e extrair uma amostra de cada um separadamente, sem sobreposição. Essa

abordagem abrange toda a população e é preferível quando se busca maior precisão nas estimativas dentro de cada estrato, segundo Barbetta (2006). Segundo Cochran, o uso de estratos homogêneos pode aumentar a precisão das estimativas da população total. Quando os estratos são homogêneos, é possível obter estimativas precisas com um número relativamente pequeno de observações em cada estrato. Essas estimativas são então ponderadas adequadamente para obter a estimativa populacional.

4. **Amostragem aleatória por conglomerado:** Segundo Cochran (1977), a amostragem por conglomerados, a unidade amostral é um conjunto de elementos da população, sendo recomendada quando há dificuldades financeiras na obtenção de listas de elementos da população. A aplicação da amostragem por conglomerados é comum em inquéritos populacionais, geralmente realizados em várias fases. Os conglomerados são unidades compostas por subunidades, que são selecionadas em cada etapa até alcançar o objeto de estudo. No entanto, o método amplamente utilizado para seleção dos conglomerados é a amostragem com probabilidade proporcional ao seu tamanho. Este método, além de sua simplicidade de aplicação, possui a vantagem de contribuir para a redução da variância entre as unidades de seleção.
5. **Amostragem complexa:** Os responsáveis do setor de saúde em alguns países têm adotado a prática de reservar tempo regularmente para avaliar o desempenho do setor, visando promover planos de ação destinados a melhorar a qualidade da assistência de acordo com as necessidades dos pacientes. A avaliação do desempenho do setor de saúde tem predominantemente empregado questionários em entrevistas que abordam temas como satisfação com o serviço, qualidade do atendimento e preocupações persistentes. A coleta dessas informações nos domicílios oferece vantagens na análise das inter-relações e na obtenção de dados que muitas vezes não estão disponíveis nos centros de saúde, como situação socioeconômica e características demográficas. Tais informações são essenciais para o monitoramento e avaliação, fundamentais para um planejamento de ações e investimentos eficaz, Szwarcwald & Damacena (2008). De acordo com Szwarcwald & Damacena (2008), a aplicação da amostragem aleatória simples nem sempre é viável em inquéritos nacionais devido a diversos fatores, como restrições financeiras, limitações de tempo em relação ao tamanho da amostra e à extensão geográfica, além da disponibilidade de recursos humanos qualificados. Como alternativa, outros métodos probabilísticos, como amostragem estratificada e

por conglomerados em múltiplas etapas, são empregados, envolvendo probabilidades de seleção desiguais, para obter uma amostra representativa da população. Esse tipo de amostragem, denominado desenho complexo, gera preocupações adicionais na análise dos dados com origem na amostragens complexas. A análise estatística de dados provenientes de amostras complexas tem sido objeto de debate em vários níveis do conhecimento acadêmico. Alguns autores destacam os problemas que podem surgir se a estrutura do plano amostral não for considerada nessas análises. Ignorar o desenho amostral e realizar análises estatísticas clássicas, presumindo uma amostragem aleatória simples, pode resultar em imprecisões nas estimativas médias e em suas respectivas variações, impactando os resultados, os testes de hipóteses e, por fim, as conclusões do estudo, Szwarcwald & Damacena (2008).

2.3.1 Elementos dos desenhos complexos de amostragem

De acordo com Szwarcwald & Damacena (2008), o desenho complexo de amostragem pode ser construído combinando os métodos de amostragem anteriormente referidos. É crucial salientar que as probabilidades de seleção de cada elemento são desiguais em cada etapa. Para equilibrar essas probabilidades desiguais, são atribuídas ponderações diferenciadas às unidades da amostra, denominadas de fatores naturais de expansão. Esses fatores correspondem ao produto das probabilidades de inclusão nas diferentes etapas de seleção. Portanto, é essencial realizar a calibração para totais conhecidos da população, ajustando os pesos naturais do desenho e/ou corrigindo os problemas decorrentes da ausência ou recusa de resposta. A calibração dos fatores naturais de expansão envolve a estimativa de novos pesos para cada unidade da amostra, através de ajustes dos pesos naturais do desenho com base nas informações das variáveis auxiliares da amostra. Um dos objetivos da calibração é obter distribuições amostrais das variáveis auxiliares muito próximas das populacionais obtidas pelo censo, ilustrando assim a importância desse procedimento, Szwarcwald & Damacena (2008).

2.3.2 Questões relacionadas a análise estatísticas de dados

Ao analisar os dados coletados por meio de inquéritos com amostras complexas, é fundamental realizar uma análise apropriada desses dados. Eles não podem ser tratados como observações independentes e identicamente distribuídas, como em uma amostragem alea-

tória simples. Portanto, é essencial considerar a estrutura complexa da amostra durante a análise para garantir a precisão e confiabilidade dos resultados, Szwarcwald & Damacena (2008).

Na análise deverá observar os seguintes aspectos:

1. Probabilidades distintas de seleção das unidades;
2. Efeitos de conglomeração das unidades;
3. Estratificação;
4. Calibração para ajuste de não resposta e outros ajustes para totais conhecidos da população.

O procedimento para extrair uma amostra do universo populacional, denominado plano de amostragem, desempenha um papel crucial na estimativa dos parâmetros. Isso ocorre porque a estimativa feita com base nesse plano amostral é fundamentada na ponderação de cada unidade da amostra pelo inverso de sua probabilidade de seleção. Essa abordagem é essencial para garantir que as estimativas sejam representativas da população e forneçam resultados confiáveis, segundo Szwarcwald & Damacena (2008).

Szwarcwald & Damacena (2008), também afirmam que quando a intenção por trás dos dados é descritiva, as estimativas específicas dos parâmetros populacionais, como totais, médias e proporções, são influenciadas unicamente pelos pesos amostrais determinados pelo plano de amostragem. No entanto, em um contexto analítico, quando o propósito é exploratório, medidas de variabilidade, como variância e desvio padrão, são necessárias para calcular intervalos de confiança e realizar testes de hipóteses. Essa diferenciação é crucial para uma análise estatística precisa e uma interpretação correta dos resultados.

Na análise analítica, as estimativas específicas estão intrinsecamente ligadas à precisão, a qual é determinada pelo tamanho do intervalo de confiança. Quanto mais amplo o intervalo de confiança, maior o erro padrão, o que resulta em uma estimativa menos precisa. A determinação do erro padrão é de suma importância para a inferência estatística, pois oferece uma avaliação da incerteza associada à estimativa específica, influenciando diretamente na interpretação dos resultados e na confiança das conclusões alcançadas, Szwarcwald & Damacena (2008).

Em um plano de amostragem complexa, a estimativa da variância é influenciada pelos pesos amostrais, pela estratificação e pelos conglomerados, de maneira combinada. No con-

texto da estratificação, quando esta é empregada de forma adequada, a variância da média estimada tende a ser menor do que se a amostragem fosse simplesmente aleatória. Isso implica em uma melhoria da precisão nas estimativas médias. Essa constatação destaca a importância de levar em conta a estrutura complexa da amostra ao calcular as estimativas de variância, o que pode resultar em resultados mais precisos e confiáveis em estudos de amostragem complexa, Szwarcwald & Damacena (2008).

Em geral, o efeito da conglomeração é oposto, levando a variâncias mais amplas em comparação com aquelas derivadas de observações independentes. Além disso, na amostragem por conglomerado, a variância da média é diretamente afetada pela variância total e pelo coeficiente de correlação intra-classe, que avalia a correlação entre as observações dentro dos conglomerados. Portanto, pode-se inferir que quanto maior a homogeneidade dentro dos conglomerados, maior será a variação entre os conglomerados e, por conseguinte, maior será a variância da média. Por outro lado, para avaliar o impacto do plano amostral (EPA) na variância das estimativas médias, é utilizado o conceito de "efeito de desenho", que é calculado como a razão entre a variância estimada pelo plano amostral e a variância estimada por uma amostra aleatória simples de igual tamanho. Além de fornecer insights sobre a perda de precisão na estimativa, o efeito de desenho também desempenha um papel crucial no planejamento futuro, especialmente no cálculo do tamanho da amostra, Szwarcwald & Damacena (2008).

Capítulo 3

Material e Métodos

Este capítulo tem como objectivo descrever os materiais e os métodos utilizados na elaboração deste estudo. Durante a pesquisa, foram conduzidas as seguintes abordagens: pesquisa exploratória para coleta de dados, pesquisa descritiva para análise da população de estudo e pesquisa explicativa para identificar os factores associados à prevalência de casamentos prematuros.

3.1 Material

3.1.1 Fonte de dados

Para a realização deste trabalho, serão utilizados dados secundários provenientes do Inquérito de Indicadores de Imunização, Malária e HIV/SIDA em Moçambique (IMASIDA) de 2015. Este inquérito teve como objetivo principal determinar a prevalência do HIV, SIDA, malária e outros indicadores de saúde das mulheres e crianças no país. Os dados são abrangentes, cobrindo informações a nível nacional, provincial e por áreas de residência (urbana e rural), além de outras características seleccionadas.

O IMASIDA 2015 utilizou uma amostra composta por 7.368 agregados familiares, seleccionados em 307 áreas de enumeração baseadas nos resultados do 3º Recenseamento Geral da População e Habitação (RGPH) de Moçambique. Dos 7.368 agregados familiares seleccionados, 7.342 estavam presentes durante a coleta de dados, dos quais 7.169 concordaram em participar na entrevista do agregado familiar, resultando em uma taxa de resposta de 98%. Entre os agregados entrevistados, foram identificadas 8.204 mulheres elegíveis entre 15 e 59

anos para a entrevista individual, das quais 7.749 foram entrevistadas.

3.1.2 Descrição das variáveis

Para a elaboração do trabalho, foram consideradas variáveis sociodemográficas, acesso a meios de comunicação, localização geográfica e o agregado familiar serão explorados visando identificar e analisar os factores associados à prevalência de casamentos prematuros.

Primeiro fez-se o filtro de mulheres menores de 18 anos, onde a posterior foi criada uma nova variável chamada casamento prematuro que recebeu o atributo (Sim=1) se tiver um estado civil casada, divorciada, viúva ou vive maritalmente e recebe (Não=0) se caso for solteira. Foi também criada uma variável chamada mídia que expressa a exposição de pelo menos um meio de comunicação descrita na Tabela 3.1, onde se for exposto a pelo menos um dos meios de comunicação é considerado (Sim=1) e caso contrário (Não=0).

Tabela 3.1: Descrição das variáveis em estudo

X_i	VARIÁVEL	CODIFICAÇÃO	Classificação	
			Unidade	Relação
X_1	Casamento Prematuro	0-Não; 1-Sim	Qualitativa nominal	Dependente
X_2	Tercil de riqueza	Baixo, médio e alto	Qualitativa nominal	Independente
X_3	Região	Sul, centro e norte	Qualitativa nominal	Independente
X_4	Nível de escolaridade	Nenhuma, primário, secundário e superior	Qualitativa nominal	Independente
X_5	Religião	Católica/anglicana, islâmica, evangélica/protestante, Zion, nenhuma e outras	Qualitativa nominal	Independente
X_6	Acesso á Mídia	0-Não; 1-Sim	Qualitativa nominal	Independente
X_7	Acesso ao Telefone	0-Não; 1-Sim	Qualitativa nominal	Independente
X_8	Acesso ao Jornal	0-Não; 1-Sim	Qualitativa nominal	Independente
X_9	Acesso a Televisão	0-Não; 1-Sim	Qualitativa nominal	Independente
X_{10}	Província	-	Qualitativa nominal	Independente
X_{11}	Zona	Rural e Urbano	Qualitativa nominal	Independente
X_{12}	Ocupação	0-Não Trabalha; 1-Trabalha	Qualitativa nominal	Independente

Para o processamento de dados será usado o Software R versão 4.4.0

3.2 Métodos

Para a elaboração deste trabalho, foram calculadas as frequências absolutas para descrever as variáveis em estudo, mapeando as frequências de casamentos prematuros a nível nacional e distribuídas por províncias. Foi aplicada amostragem complexa para melhorar a eficiência na estimativa do risco de ocorrência de casamentos prematuros. Utilizando regressão logística múltipla, com nível de confiança de 5%, foram determinadas as odds ratios associadas. Além disso, foi realizado o teste qui-quadrado para verificar a associação entre casamentos prematuros e variáveis sociodemográficas e culturais.

3.2.1 Amostragem Complexa

A amostragem é o processo de extrair amostras de uma população, sendo um dos métodos mais comuns a amostragem aleatória simples (Szwarcwald & Damacena, 2008). Neste método, todos os elementos da população têm a possibilidade de serem selecionados com uma probabilidade pré-definida. No entanto, surgem algumas dificuldades, como a obtenção de uma listagem completa da população a ser estudada, o que pode ser difícil e demorado, envolvendo custos elevados e nem sempre garantindo a fiabilidade desejada (Szwarcwald & Damacena, 2008). Dessa forma, muitas vezes é necessário recorrer à amostragem complexa, onde a probabilidade de seleção é desigual para prover uma amostra representativa da população, de acordo com orçamento previsto (Szwarcwald & Damacena, 2008).

Amostragem complexa envolve a utilização simultânea de diversos métodos de amostragem probabilística, como amostragem estratificada, por conglomerado e aleatória, com o objetivo de selecionar uma amostra representativa da população (Szwarcwald & Damacena, 2008). A utilização de estratos homogêneos pode conduzir a uma melhoria significativa na precisão das estimativas das características da população total, especialmente em estratos que contenham um número relativamente reduzido de observações (Szwarcwald & Damacena, 2008).

3.2.2 Teste de independência

De acordo com Ferreira (2009), recorre-se ao teste de independência quando os dados da pesquisa são representados em frequências e categorias discretas. Para o teste de hipóteses,

utiliza-se a estatística Qui-Quadrado para determinar a independência entre duas variáveis, conforme definido pela seguinte estatística:

$$Q^2 = \sum_{i=1}^r \sum_{j=1}^s \left(\frac{(O_{ij} - e_{ij})^2}{e_{ij}} \right) \quad (3.1)$$

Onde:

O_{ij} – Representa frequências observadas na linha i da coluna j .

e_{ij} – Representa frequências esperadas da linha i da coluna j .

r e s – Representam o número de linhas e o número de colunas, respectivamente.

Para calcular a frequência esperada e_{ij} em cada célula, multiplica-se os totais marginais associados a uma determinada célula e divide-se pelo total de casos. Os valores de χ^2 obtidos pela equação acima seguem uma distribuição aproximadamente Qui-Quadrado com graus de liberdade $gl = (L - 1)(C - 1)$, onde L é o número de linhas e C é o número de colunas, testando as seguintes hipóteses:

H_0 : as variáveis são independentes

H_1 : as variáveis são dependentes

Regra de decisão: Rejeita-se a hipótese nula caso o *P-value* associado à estatística do teste seja menor ou igual a 5% ($\alpha < 0.05$).

3.2.3 Estimação dos parâmetros

Hosmer & Lemeshow (1989) propõem estimar os parâmetros da regressão logística múltipla por máxima verossimilhança, encontrando o valor de β que maximiza $L(\beta)$. Esse processo é iterativo e requer a derivação de $L(\beta)$ em relação a cada parâmetro:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=0}^n \left[y_i x_{ij} - \frac{e^{(x_i^T \beta)}}{e^{(x_i^T \beta)}} x_{ij} \right] \quad (3.2)$$

A matriz de covariância dos coeficientes estimados é calculada a partir das segundas derivadas parciais do logaritmo da função de verossimilhança:

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j^2} = - \sum_{i=0}^n [x_{ij}^2 \pi_i (1 - \pi_i)] \quad (3.3)$$

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j \partial \beta_l} = - \sum_{i=0}^n [x_{ij} x_{il} \pi_i (1 - \pi_i)] \quad (3.4)$$

Onde $j, i = 0, 1, 2, \dots, p$. Ao formar uma matriz quadrada de dimensão $(p + 1)$ com os valores médios simétricos dos termos mencionados nas equações, obtém-se $L(\beta)$, conhecida como matriz de informação. A matriz de informação de Fisher é definida como:

$$L(\beta) = E \left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} \right) = X^T Q X \quad (3.5)$$

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

3.2.4 Razão de chances

De acordo com Goncalves (2012), a medida de associação Odds Ratio (OR), tradução em português, razão de chances, é frequentemente empregada na regressão logística para complementar o teste de significância da covariável x . Essa medida se destaca pela relação entre os parâmetros do modelo logístico e o Odds Ratio. Para representar essa relação, inicialmente assumimos que a covariável é binária. O Odds Ratio é o quociente entre a chance de sucesso ($Y = 1$) em indivíduos com $X = 1$ e a chance desse evento em indivíduos com $X = 0$. A probabilidade do evento de interesse ocorrer em indivíduos com $X = 1$ é definida por:

$$\frac{\pi(1)}{1 - \pi(1)} \quad (3.6)$$

De forma análoga, a chance de o evento de interesse ocorrer em indivíduos com $X = 0$ é definida como $\frac{\pi(0)}{1 - \pi(0)}$. Portanto, o Odds Ratio é uma maneira de comparar se a probabilidade de o evento de interesse ocorrer é a mesma para os indivíduos com $X = 1$ ou $X = 0$.

As probabilidades de o evento de interesse ocorrer para as duas categorias de X são dadas, respectivamente, por:

$$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \quad \text{e} \quad \pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \quad (3.7)$$

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

$$OR = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} = e^{\beta_1} \quad (3.8)$$

Evidenciando a relação entre o Odds Ratio e o coeficiente do modelo, o valor de OR depende da codificação adotada para a covariável binária X podendo ser definida por quaisquer valores. Ao considerar a codificação com os valores genéricos a e b , o cálculo do Odds Ratio é expresso como:

$$OR = \frac{\pi(a)[1 - \pi(b)]}{\pi(b)[1 - \pi(a)]} = e^{\beta_1(a-b)} \quad (3.9)$$

A interpretação do Odds Ratio é fundamentada na codificação de X , comumente definida em termos de 0 e 1, o que facilita uma interpretação direta dos parâmetros. O cálculo do Odds Ratio é realizado a partir de dados reais, frequentemente organizados em tabelas de contingência.

Valores de Odds Ratio maiores que 1 indicam aumento das chances, enquanto valores menores que 1 indicam redução das chances. Um Odds Ratio igual a 1 sugere que a variável não tem efeito sobre a variável dependente. A variação percentual na razão das chances é calculada pela expressão (Maroco, 2007):

$$OR(\%) = (OR - 1) \times 100\% \quad (3.10)$$

3.2.5 Multicolinearidade

Conforme Hair et al. (2005), a multicolinearidade é caracterizada pela forte correlação de uma variável independente com um conjunto de outras variáveis independentes. No contexto do modelo logístico, presume-se que as variáveis explicativas sejam linearmente independentes, ou seja, não correlacionadas. O método de estimação Stepwise permite identificar a multicolinearidade por meio de mudanças significativas nos coeficientes estimados. A intensidade da multicolinearidade é avaliada pela correlação entre as variáveis independentes, pela Tolerância, pelo VIF (Factor de Inflação da Variância) e outras técnicas.

A Tolerância mede o grau em que uma variável X_a é explicada por outras variáveis independentes e é dada por:

$$T_a = 1 - R_i^2 \quad (3.11)$$

Onde: T_a é a Tolerância da variável independente X_a ; R_i^2 corresponde ao coeficiente de determinação entre X_a e as restantes variáveis independentes.

Sendo assim, a Tolerância da variável X_a indica a proporção de sua variação não explicada pelas outras variáveis independentes restantes. Esta medida varia de 0 a 1, quanto mais próxima de zero, maior é a multicolinearidade, e quanto mais próxima de 1, menor é a multicolinearidade. O inverso da Tolerância é denominado Factor de Inflação da Variância (VIF - Variance Inflation Factor):

$$VIF = \frac{1}{1 - R_i^2} \quad (3.12)$$

Conforme Hair et al. (2005), quanto mais próximo de zero o VIF estiver, menor será a multicolinearidade. O valor comumente considerado para indicar multicolinearidade é 10. O índice de condição (Condition Index) é calculado como a raiz quadrada da razão entre cada valor próprio e o valor próprio total. Um índice de condição superior a 15 sugere a presença de multicolinearidade, enquanto um índice maior que 30 indica problemas significativos de multicolinearidade.

3.2.6 Curva de ROC

A curva ROC (Receiver Operating Characteristic) é uma ferramenta que permite avaliar o desempenho de um modelo de regressão binária, onde a variável resposta é do tipo 0-1. Geralmente, é empregado um gráfico simples e robusto para essa avaliação, possibilitando analisar a variação da sensibilidade e especificidade para diferentes pontos de corte. Ao escolher um ponto de corte, representado por C , cada probabilidade estimada é comparada com esse valor. O valor mais frequente para C é 0.5, segundo Hosmer & Lemeshow (2000).

Se a probabilidade estimada ultrapassar o valor de C , a variável dicotômica assumirá o valor 1; caso contrário, assumirá o valor 0.

Sensibilidade refere-se à probabilidade do teste fornecer um verdadeiro positivo, avaliando a capacidade do modelo em classificar corretamente um indivíduo como evento quando ele realmente é evento (ou seja, $P(\hat{Y} = 1|Y = 1)$). Já a especificidade, também chamada de verdadeiro negativo, mede a habilidade do modelo em prever corretamente um indivíduo como não evento quando ele realmente não é evento (isto é, $P(\hat{Y} = 0|Y = 0)$).

Sensibilidade é a capacidade do teste em identificar corretamente um verdadeiro positivo, ou seja, avaliar a Probabilidade do modelo em classificar corretamente um indivíduo como evento quando realmente é evento, ou seja, $P(\hat{Y} = 1|Y = 1)$. Por outro lado, a especificidade é referente ao verdadeiro negativo, pois mede a capacidade do modelo em identificar corretamente um indivíduo como não evento quando na verdade não é evento, ou seja, $P(\hat{Y} = 0|Y = 0)$.

Segundo Hajian-Tilaki (2013), a curva ROC é um gráfico que relaciona a sensibilidade (ou taxa de verdadeiros positivos) com a taxa de falsos positivos. Isso é representado pela sensibilidade no eixo vertical (ordenadas) e $1 -$ especificidade no eixo horizontal (abscissas), demonstrando como essas métricas variam conforme o valor de corte ao longo do eixo de decisão.

A representação da curva ROC destaca os valores nos quais a sensibilidade é otimizada em relação à especificidade, evidenciando o ponto mais próximo do canto superior esquerdo do diagrama, onde o verdadeiro positivo é 1 e o falso positivo é 0. A área sob a curva ROC fornece uma medida de discriminação, indicando a habilidade de distinguir entre diferentes indivíduos.

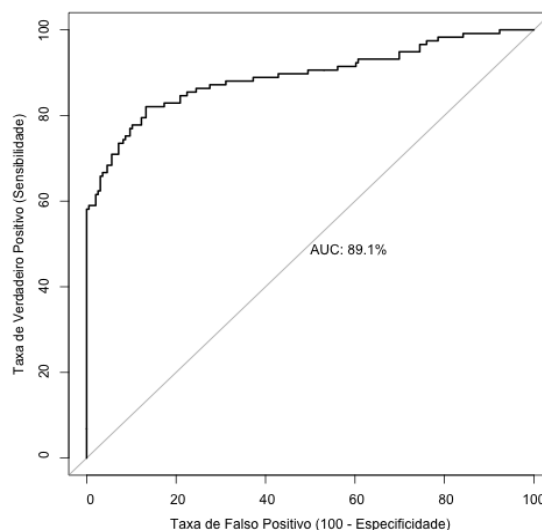


Figura 3.1: Curva de ROC. Fonte: Prates (2020)

CrITÉRIOS para classificar o poder discriminatório de um modelo de regressão logística:

- $ROC = 0,5$: O modelo não faz discriminação entre os indivíduos com e sem o endpoint;
- $0,6 \leq ROC < 0,7$: O modelo apresenta uma discriminação limitada;
- $0,7 \leq ROC < 0,8$: O modelo exibe uma discriminação aceitável;
- $0,8 \leq ROC < 0,9$: O modelo demonstra uma excelente discriminação;
- $ROC \geq 0,9$: O modelo revela uma discriminação quase perfeita.

Capítulo 4

Resultados e Discussão

Neste capítulo, apresentam-se os resultados conforme a metodologia descrita no capítulo anterior, incluindo as análises realizadas para responder às questões levantadas na seção 1.2.

4.1 Prevalência do Casamento prematuro

Da amostra (899) de mulheres com idades entre os 15 a 18 anos de idade, de acordo com a Figura 4.1, constatou-se que 30.57% (275/899) das mulheres casaram-se antes dos 18 anos e mais da metade (69, 43%; 624/899), nas idades mais avançadas.

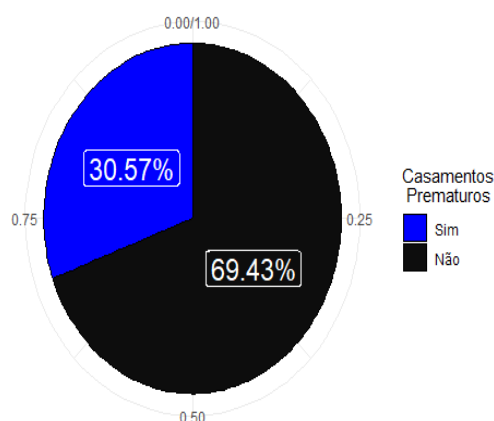


Figura 4.1: Prevalência de casamentos Prematuros

A Figura 4.2, ilustra uma grande variabilidade da prevalência de casamentos prematuros ao nível provincial, sendo notório que a Região norte tem prevalências superiores a 30%. Em termos proporcionais a Região sul teve maior proporção de casamentos prematuros e

ao se analisar a região observa-se que maior prevalência esta para a Zona rural com 9% de ocorrência de casamentos prematuros.

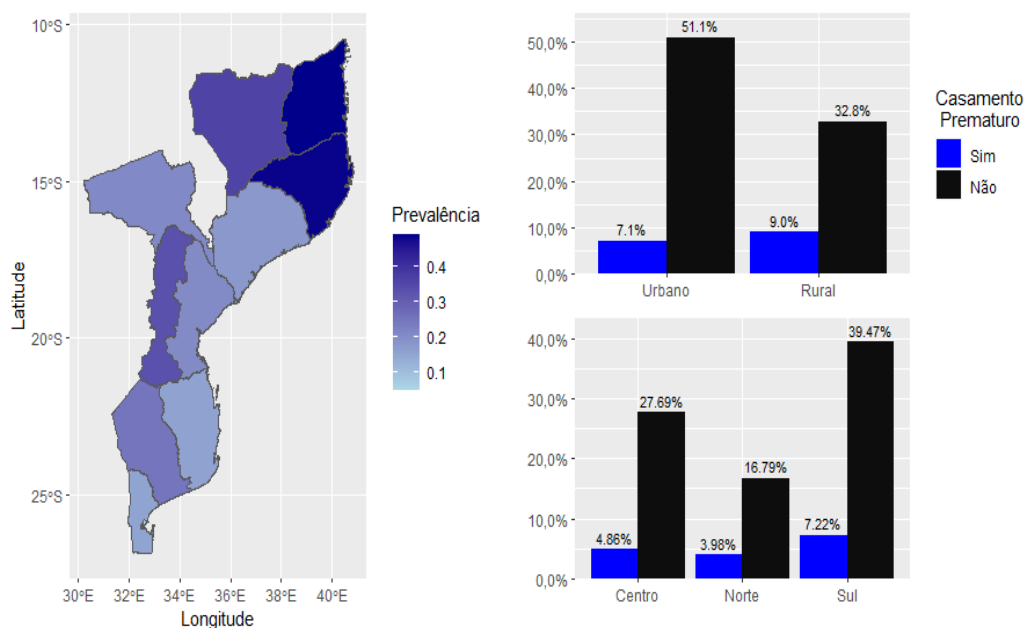


Figura 4.2: Mapeamento da prevalência de casamentos prematuros

De acordo com a Figura 4.3, o nível primário apresenta maior prevalência de casamentos prematuros com 21,72%, e nenhum (sem nível de escolaridade) 4,53% e para o nível superior 0,00%.

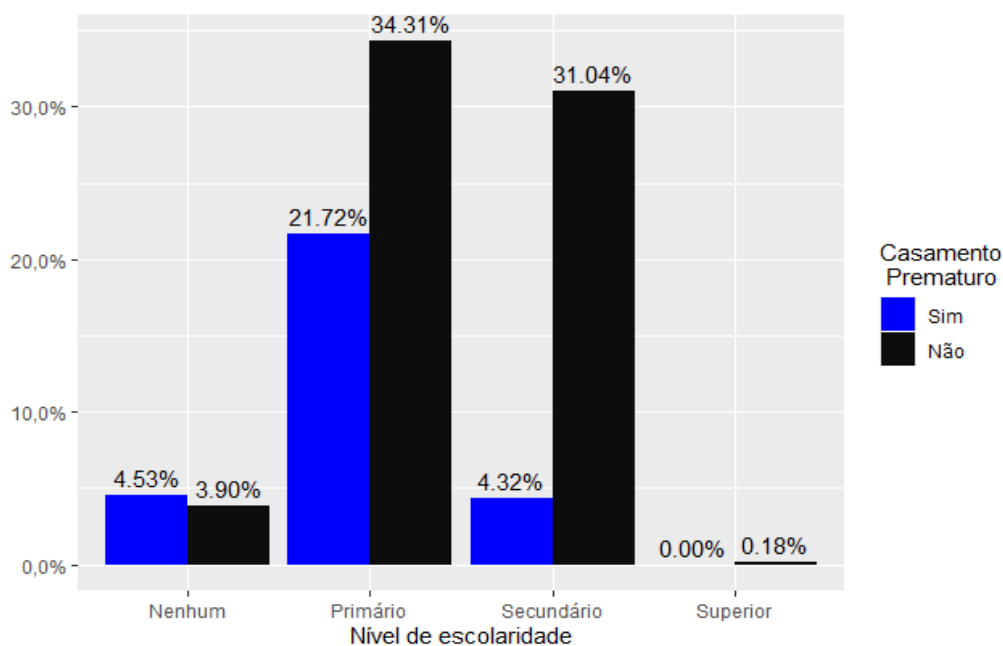


Figura 4.3: Distribuição dos casamentos prematuros em relação ao nível de escolaridade

Segundo a Figura 4.4, observa-se que a distribuição dos casamentos prematuros por religião, verifica se que, a religião católica/Anglicana apresentam maior de proporção de casamentos prematuros (10.53%), islâmica com 7.36%, Protestante/Evangélica 7,34% e Zion com 3,55%, com nenhuma (1.54%) e outras religiões com 0.24%.

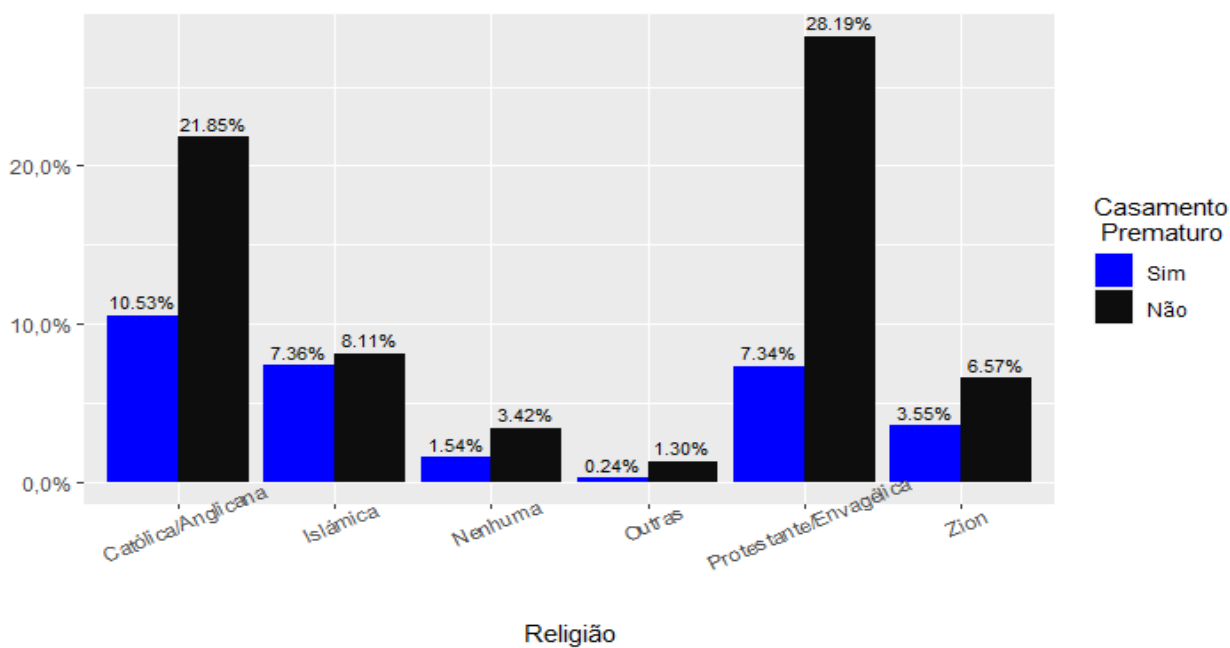


Figura 4.4: Distribuição dos casamentos prematuros em relação a religião

A Figura 4.5, mostra a distribuição dos casamentos prematuros em relação ao acesso a meios de comunicação social, nota-se que para mulheres com acesso a telefone tiveram maior proporção de casamentos prematuros 30.43%, enquanto que para mulheres com acesso a rádio tiveram a menor proporção de casamentos prematuros 19.02%.

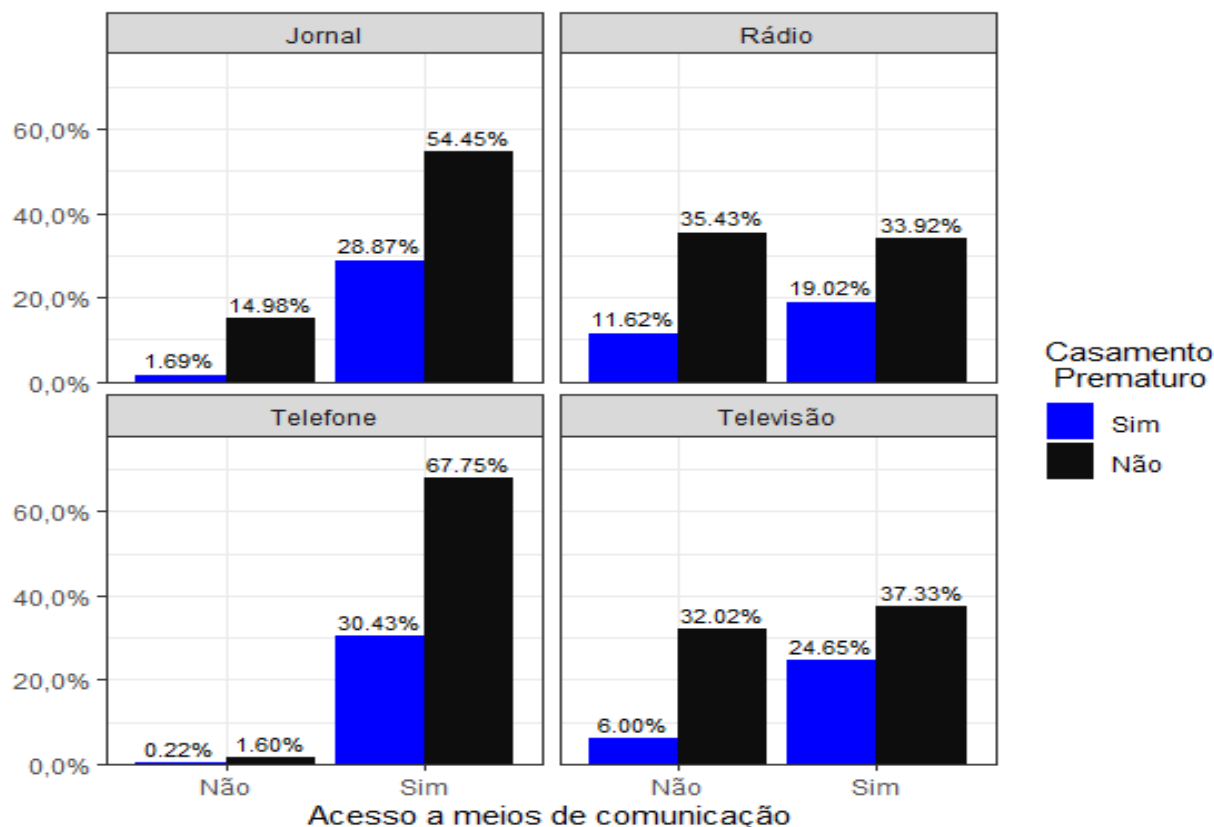


Figura 4.5: Distribuição dos casamentos prematuros em relação ao acesso à meios de comunicação

4.2 Avaliação da multicolinearidade

A Tabela 4.1 apresenta a avaliação de multicolinearidade entre as variáveis independentes, utilizando o Fator de Inflação da Variância (VIF) e a tolerância como métricas principais. Os valores de VIF variaram de 1.161 (telefone) a 6.529 (tercil de riqueza), permanecendo abaixo do limite crítico de 10, indicando ausência de multicolinearidade severa. A tolerância variou entre 0.153 e 0.861, sugerindo relações moderadas entre algumas variáveis, mas ainda dentro de níveis aceitáveis. Além disso, os testes de independência, baseados no qui-quadrado (χ^2), revelaram associações significativas para a maioria das variáveis, com exceção do telefone ($p = 0.107$), que não apresentou significância estatística.

Os resultados destacam a relevância de variáveis como tercil de riqueza, escolaridade e média, que apresentaram associações estatisticamente significativas com a variável dependente. Apesar de alguns valores moderados de multicolinearidade, os indicadores gerais sugerem que o modelo é robusto e adequado para análises subsequentes. Esses achados reforçam

a necessidade de considerar fatores econômicos, educacionais e de acesso à informação ao analisar o fenômeno em estudo.

Tabela 4.1: Teste de independência e avaliação de multicolinearidade

Variável	VIF	Tolerância	χ^2	p-valor
Tercil de riqueza	6.529	0.153	86.581	<0.001
Escolaridade	3.395	0.294	80.966	<0.001
Zona	2.257	0.442	53.355	<0.001
Religião	3.778	0.264	34.326	0.0001
Radio	4.487	0.223	12.448	0.003
Televisão	4.361	0.229	53.725	<0.001
Telefone	1.161	0.861	2.541	0.107
Jornal	1.715	0.5828	33.119	<0.001
Trabalho	1.236	0.808	52.816	<0.001
Região	3.681	0.271	75.023	<0.001
Mídia	6.272	0.159	54.412	<0.001
Província	-	-	87.442	<0.001

4.2.1 Risco de ocorrência de casamentos prematuros

Entre os factores analisados, a religião destacou-se como uma variável de grande influência. Indivíduos identificados com a religião Islâmica apresentaram um aumento significativo na probabilidade (OR = 2,87; IC 95%: 1,20–6,86; p = 0,018), assim como os pertencentes à religião Zion (OR = 3,54; IC 95%: 1,58–7,97; p = 0,002), em comparação aos católicos/anglicanos. Por outro lado, a categoria "Outras religiões" apresentou um OR de 0,00 (p < 0,001), indicando uma associação estatisticamente muito forte, possivelmente devido à ausência de eventos em um dos grupos.

Outro resultado notável foi relacionado ao trabalho. Indivíduos que trabalham demonstraram uma probabilidade quase três vezes maior de serem afetados (OR = 3,34; IC 95%: 1,95–5,72; p < 0,001) em comparação aos que não trabalham. Esse fator sugere que ocupações podem estar associadas a maior exposição ou vulnerabilidade ao evento analisado. Além disso, o acesso a jornais também mostrou uma relação significativa, com indivíduos leitores apresentando uma menor probabilidade de eventos negativos (OR = 0,43; IC 95%: 0,19–0,97; p = 0,043), sugerindo que o acesso à informação pode ser protetor. Por fim, o nível de escolaridade e o acesso a mídia, apesar de não alcançarem significância estatística em todas as categorias, mostraram tendências interessantes. Indivíduos com nível superior apresentaram valores de OR extremamente baixos (0,00; p < 0,001), indicando uma provável

proteção associada ao maior nível educacional.

Tabela 4.2: Estimativas do OR

Variável	Modelo simples			Modelo múltiplo ajustado			
	OR	95% CI	p-value	OR	95% CI	β	p-value
TERCÍL DE RIQUEZA							
Alto	—	—	—	—	—	—	Ref
Baixo	0.94	0.26, 3.49	>0.9	1.67	0.56, 4.94	0.51	0.4
Médio	1.48	0.59, 3.68	0.4	2.00	0.89, 4.53	0.69	0.094
NÍVEL DE ESCOLARIDADE							
Nenhum	—	—	Ref				
Primário	0.64	0.25, 1.65	0.4				
Secundário	0.42	0.14, 1.25	0.12				
Superior	0.00	0.00, 0.00	<0.001				
ZONA							
Urbano	—	—	Ref				
Rural	1.37	0.69, 2.75	0.4				
RELIGIÃO							
Católica/Anglicana	—	—	—	—	—	—	Ref
Islâmica	2.67	1.10, 6.46	0.030	2.87	1.20, 6.86	1.05	0.018
Nenhuma	2.30	0.78, 6.82	0.13	2.38	0.87, 6.50	0.87	0.090
Outras	0.00	0.00, 0.00	<0.001	0.00	0.00, 0.00	0.00	<0.001
Protestante/Envagélica	1.68	0.79, 3.55	0.2	1.42	0.72, 2.80	0.35	0.3
Zion	3.80	1.58, 9.15	0.003	3.54	1.58, 7.97	1.26	0.002
RADIO							
Não	—	—	Ref				
Sim	1.42	0.56, 3.61	0.5				
TELEVISÃO							
Não	—	—	—	—	—	—	Ref
Sim	2.31	0.89, 6.04	0.087	2.01	0.82, 4.94	0.69	0.13
TELEFONE							
Não	—	—	Ref				
Sim	0.69	0.08, 6.06	0.7				
JORNAL							
Não	—	—	—	—	—	—	Ref
Sim	0.54	0.23, 1.28	0.2	0.43	0.19, 0.97	-0.84	0.043
TRABALHO							
Não	—	—	—	—	—	—	Ref
Sim	2.93	1.67, 5.13	<0.001	3.34	1.95, 5.72	1.21	<0.001
REGIÃO							
Centro	—	—	Ref				
Norte	1.29	0.55, 3.02	0.6				
Sul	0.85	0.45, 1.62	0.6				
ACESSO A MÍDIA							
Não	—	—	—	—	—	—	Ref
Sim	0.46	0.14, 1.52	0.2	0.67	0.34, 1.32	-0.40	0.3
						$\beta_0 = 2.57$	0.7

4.2.2 Avaliação do modelo

No primeiro gráfico, a Curva ROC (Receiver Operating Characteristic), observa-se a relação entre sensibilidade e especificidade. A curva está acima da linha diagonal, indicando que o modelo possui capacidade preditiva, pois é melhor do que uma classificação aleatória. Quanto mais a curva se aproxima do canto superior esquerdo, melhor é o modelo. A capacidade discriminatória do modelo é de 72.72%, o que indica uma capacidade satisfatória para diferenciar entre as classes. O modelo alcança 97.87% de sensibilidade, mas com uma especificidade de 3.17%, o que significa que ele identifica maior parte dos casos positivos, porém classifica maior parte dos negativos como positivos, resultando em uma alta taxa de falsos positivos.

O segundo gráfico, a Curva de Precisão-Recall, mostra a relação entre precisão (a proporção de verdadeiros positivos entre os positivos preditos) e recall (sensibilidade). Esse gráfico é útil para conjuntos de dados desbalanceados, onde a classe positiva (casamentos prematuros) é menos frequente. A curva indica que a precisão do modelo é relativamente baixa em níveis menores de recall, o que significa que, embora o modelo consiga identificar alguns casos de casamentos prematuros, ele o faz com uma precisão limitada, o que pode afetar a confiabilidade da previsão.

Por fim, o Gráfico de Resíduos vs. Ajustados ajuda a avaliar a adequação do modelo, mostrando os resíduos em função dos valores ajustados. Embora os resíduos estejam distribuídos ao redor da linha horizontal, há algumas tendências para valores extremos, o que indica que o modelo pode não estar capturando toda a variabilidade dos dados. Pontos como 950 e 74, destacados no gráfico, podem ser outliers ou observações influentes que impactam o ajuste do modelo.

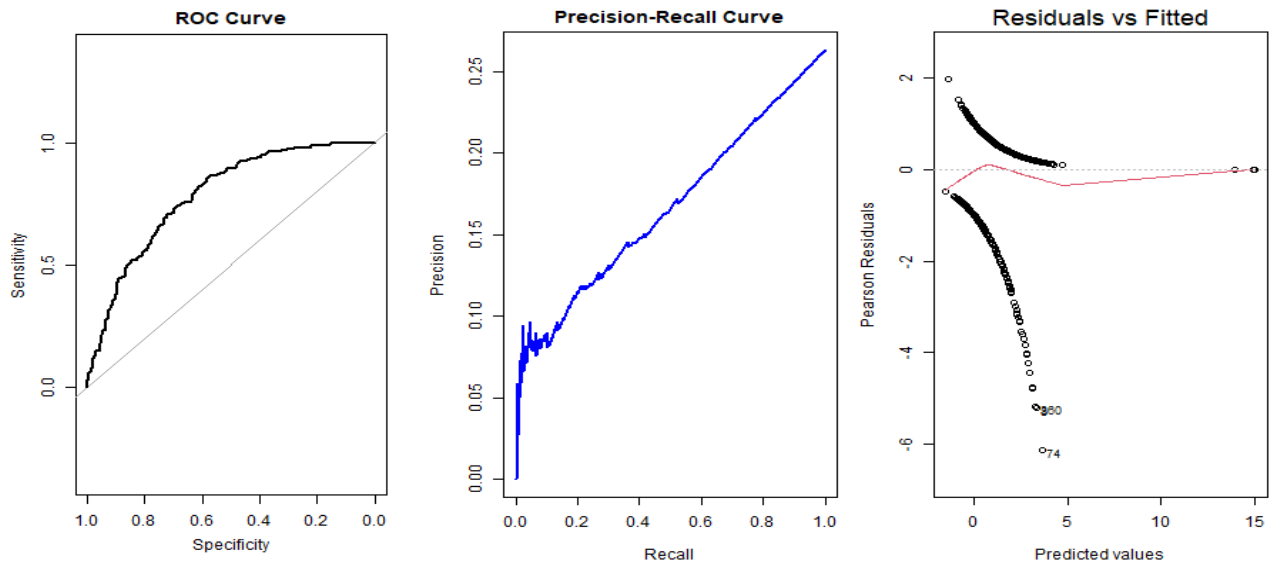


Figura 4.6: Avaliação do modelo

O teste de Wald ($F = 10,796$, $P\text{-valor} < 0,001$) sugere que as variáveis independentes incluídas no modelo têm uma contribuição estatisticamente significativa para explicar a variável dependente. Além disso, o teste de Hosmer-Lemeshow ($\chi^2 = 832$, $P\text{-valor} < 0,001$) indica que o modelo apresenta um bom ajuste global, uma vez que não há evidências significativas de discrepâncias entre os valores observados e os valores previstos pelo modelo.

4.3 Discussão dos resultados

Nesta seção, analisamos os fatores associados à ocorrência de casamentos prematuros (CP), conforme os resultados apresentados, à luz da revisão da literatura discutida na seção 2. Os resultados obtidos fornecem resultados valiosos sobre os determinantes sociais, econômicos, culturais e demográficos que influenciam a prevalência de CP, ao mesmo tempo que apontam algumas divergências e concordâncias com estudos prévios.

Os resultados indicam que o tercil de riqueza não foi significativo como determinante social para os casamentos prematuros. Embora a literatura existente, como o relatório da UNICEF (2018), aponte maior prevalência de casamentos prematuros entre grupos de baixa renda, os achados deste estudo sugerem uma realidade distinta. A pobreza é frequentemente identificada como um factor catalisador para o casamento prematuro, uma vez que famílias em situações econômicas precárias frequentemente utilizam o casamento como estratégia para reduzir encargos financeiros. No entanto, a ausência de significância do tercil de riqueza nos dados analisados pode refletir especificidades locais, como diferenças culturais, sociais ou econômicas, que requerem análises mais detalhadas e contextualizadas para melhor compreender as nuances entre riqueza e casamentos prematuros.

Os dados sugerem que as zonas rurais e urbanas não foram fatores significativos, assim como o nível de escolaridade da criança. Esses achados dialogam com Schmidt (2023), que associa o abandono escolar como uma consequência do casamento prematuro, e não como um fator preditivo. Adicionalmente, a identificação de três crianças no nível superior que não vivenciaram casamento prematuro pode refletir barreiras culturais ou econômicas em populações com maior nível educacional, alinhando-se a estudos que relacionam educação com empoderamento e adiamento do casamento. No entanto, os resultados divergem de Bassiano e Lima (2018), que destacam maior prevalência de casamento prematuro em áreas rurais devido a fatores culturais e econômicos. Essa discrepância pode indicar que, nas regiões analisadas, áreas urbanas enfrentam pressões sociais e econômicas distintas, sugerindo a necessidade de intervenções específicas e adaptadas ao contexto local.

A religião revelou-se um fator social relevante, mostrando que, ao comparar a religião Católica/Anglicana, os indivíduos de religião Islâmica têm duas vezes mais chances de viven-

ciar casamentos prematuros, enquanto para a religião Zion as chances são três vezes maiores. Embora a literatura, como a UNICEF (2019), indique que a religião geralmente não é um fator determinante, os dados deste estudo sugerem que doutrinas específicas podem influenciar tradições culturais que promovem casamentos precoces. Esses achados reforçam a importância de considerar a diversidade cultural e religiosa ao avaliar fatores de risco, destacando a necessidade de abordagens contextuais em intervenções e políticas públicas.

O acesso à mídia, como televisão, rádio e telefone, não se mostrou um fator relevante para os casamentos prematuros. Contudo, o acesso a jornais revelou-se um fator protetivo, reduzindo em 46% as chances de ocorrência de casamentos prematuros. Por outro lado, crianças que possuem alguma ocupação apresentam um risco três vezes maior de vivenciar casamentos prematuros. Esses achados dialogam com a literatura em Moçambique, onde estudos têm apontado que o acesso limitado a fontes diversificadas de informação e a vulnerabilidade econômica de famílias frequentemente levam crianças a se envolverem em ocupações precárias, o que contribui para práticas como o casamento prematuro (UNICEF, 2018; Ministério do Género, Criança e Ação Social, 2020). Essa realidade evidencia a necessidade de políticas que promovam o acesso à educação e à informação, bem como a proteção infantil, como estratégias para mitigar os casamentos prematuros no contexto moçambicano.

Capítulo 5

Conclusões e recomendações

5.1 Conclusões

Este estudo analisou os factores associados à ocorrência de casamentos prematuros entre jovens de 15 a 17 anos em Moçambique, utilizando amostragem complexa e regressão logística para explorar essas relações. Os principais achados indicam que:

- Em um universo de 899 mulheres com idades entre 15 e 18 anos, constatou-se que 30,57% se casaram antes dos 18 anos.
- A análise de regressão logística identificou três (3) factores associados aos casamentos prematuros: a religião, acesso ao jornal e a ocupação.
- Os principais factores associados ao risco de casamentos prematuros foram a religião Islâmica, com duas vezes mais chances, a religião Zion, com três vezes mais chances, e a ocupação da mulher, que aumenta o risco em três vezes.

5.2 Recomendações

- **Educação e conscientização:** É fundamental implementar campanhas educativas que abordem os riscos e as consequências dos casamentos prematuros, com foco nas comunidades e nas escolas. Essas campanhas devem incluir temas relacionados aos direitos das meninas e a importância de retardar o casamento até a idade adulta.
- **Parcerias com líderes religiosos e comunitários:** Considerando que a religião foi identificada como um fator associado aos casamentos prematuros, é crucial envolver líderes religiosos e comunitários em iniciativas de conscientização. Esses líderes podem desempenhar um papel essencial na mudança de normas e atitudes que perpetuam a prática.
- **Empoderamento econômico das jovens:** A ocupação foi outro fator relevante associado aos casamentos prematuros. Portanto, recomenda-se promover programas de capacitação profissional e oportunidades de emprego para jovens mulheres, ajudando-as a alcançar independência financeira e a adiar o casamento.
- **Monitoramento e avaliação contínuos:** É importante que políticas e programas de prevenção de casamentos prematuros sejam acompanhados de sistemas de monitoramento e avaliação contínuos, de modo a medir seu impacto e ajustar estratégias conforme necessário.
- **Fortalecimento de leis e políticas:** O fortalecimento e a aplicação de leis que proíbam casamentos antes dos 18 anos, juntamente com políticas que promovam a proteção das jovens, são essenciais para reduzir a incidência de casamentos prematuros.

Referências

- Abbad, G., & Torres, C. V. (2001). Regressão múltipla stepwise e hierárquica em psicologia organizacional: aplicações, problemas e soluções. *SciELO*, 23.
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley Sons.
- Ajuda em acção. (2018). *Países em que se pratica o casamento infantil*. <https://ajudaemacao.org/blog/infancia/paises-em-que-se-pratica-o-casamento-infantil/>. (Acesso em: 23 de Fevereiro de 2024)
- Araujo, P. (2017). *Regressao logistica multinivel: um estudo sobre o consumo de alcool e inalantes entre crianças e adolescentes*.
- Barbetta, P. A. (2006). *Estatistica aplicada as ciencias sociais* (6th ed.). Florianopolis: Ed. Da UFSC.
- Bassiano, V., & de Lima, C. A. (2018). Casamentos prematuros em moçambique: causas e consequências de abandono escolar e meninas com futuros destruídos. *Imagens da Educação*, 8(2), e43085.
- Cabral, C. I. S. (2013). *Aplicacao do modelo de regressao logistica num estudo de mercado* (Unpublished master's thesis). Universidade de Lisboa, Lisboa.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley Sons.
- Cordeiro, G. M., & Demétrio, C. G. (2013). *Modelos lineares generalizados e extensões*. Piracicaba: ESALQ/USP. (Obtido em 09 de 07 de 2018, de http://www.ufjf.br/clecio_ferreira/files/2013/05/Livro-Gauss-e-Clarice.pdf)
- Corrar, S. R., Natario, I., & Martins, M. R. (2007). *Modelos de regressão logística com apoio computacional*. São Paulo: IME-USP.

- Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.
- Ferreira, M. (2013). *Modelos de regressão: uma aplicação em medicina dentaria*.
- Freitas, P. H. (2019). *Regressão logística na modelagem da probabilidade de vitória em jogos de futebol americano*. (Minas Gerais, Brasil)
- Fundo das Nações Unidas para a Infância e Fundo das Nações Unidas para a População. (2016a). *Comunicado de imprensa da unicef e unfpa contra os casamentos prematuros*. (Recuperado em 16 julho, 2017, de <http://www.wlsa.org.mz/comunicado-de-imprensa-da-unicef-e-unfpa-contra-os-casamentos-prematuros>)
- Fundo das Nações Unidas para a Infância e Fundo das Nações Unidas para a População. (2016b). *Comunicado de imprensa da unicef e unfpa contra os casamentos prematuros*. (Nova Iorque, 8 de Março de 2016)
- Fundo das Nações Unidas para a Infância em Moçambique. (2016). *Um perfil do casamento prematuro em África*. Maputo-Moçambique. (Recuperado em 16 abril, 2018, de <http://www.unicef.org.mz>)
- Fávero, P. L., Belfiore, P., Silva, F. L., & Chan, B. L. (2009). *Análise de dados: modelagem multivariada para tomada de decisões*. Rio de Janeiro: Elsevier.
- Fórum Parlamentar da SADC. (2016). *Lei do modelo da erradicação de casamentos prematuros e proteção da criança em casamento*. <https://www.girlsnotbrides.org/documents/486/PT-Model-Law-on-Eradicating-Child-Marriages-August-16-2016.pdf>. (Acesso em: 23 de Fevereiro de 2024)
- GIRLS NOT BRIDES. (n.d.). *About child marriage*.
- Goncalves, B. (2012). *Aplicacao de tecnicas estatisticas multivariadas a indicadores contaveis: uma contribuicao de evidenciacao das metricas contaveis*. (Brasil)
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análise multivariada de dados* (5a ed.). São Paulo: Prentice Hall.
- Hajian-Tilaki, K. (2013). Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627–635.

- Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, 72(360), 851–853.
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
- Ibraimo, M. N. (2014). *O conselho de escola como espaço de participação da comunidade* (Unpublished doctoral dissertation). Universidade Catolica Portuguesa (Portugal).
- IDS. (2011). *Inquérito demográfico de saúde*. MISAU e ORC Macro de Moçambique, Maputo.
- IMASIDA. (2015). *Malária e hiv/sida em moçambique 2015. Relatório Preliminar de Indicadores de HIV*.
- Kleinbaum, D. G., & Klein, M. (2002). *Logistic regression: A self-learning text* (2nd ed.). New York, NY: Springer.
- Maroco, P. (2007). *Análise estatística: com utilização do spss*. Lisboa: Edicoes Silabo.
- Mattar, F. (1996). *Pesquisa de marketing*. Atlas.
- Mesquita, P. S. B. (2014). Um modelo de regressão logística para avaliação de programas de pós-graduação no brasil.
- MISAU. (2012). *Impacto das uniões prematuras na educação, saúde e nutrição em moçambique*. chrome-extension://efaidnbmninnibpcapcglclefindmkaj/
<https://www.rosco.org.mz/index.php/documentos/relatorios/15-impacto-das-unioes-prematuras-na-educacao-saude-e-nutricao-em-mocambique/file>. (Acesso em: 9 de Março de 2024)
- MISAU. (2018). *Estratégia de saúde escolar e dos adolescentes e jovens 2018- 2024 [2029]*. Direcção Nacional de Saúde Pública.
- MISSAU, & INE. (2013). *Moçambique inquérito demográfico e de saúde 2011*. Maputo, Moçambique: MISAU, INE e ICFI.

- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A*, 135, 370–384.
- Osório, C. (2015). Os ritos de iniciação: Identidades femininas e masculinas e estruturas de poder. *Maputo Ed. Maria José Arthur*.
- Paula, G. A. (2004). *Modelos de regressão com apoio computacional*. São Paulo: IME/USP.
- PLAN INTERNACIONAL. (2018). *Plan internacional lança o documentário: Casamento infantil*. (Plan International, 27 fevereiro)
- Powers, D. A., & Xie, Y. (1999). *Statistical methods for categorical data analysis*. Austin: Academic Press.
- Prates. (2020). *Curva roc e auc em machine learning*. (<https://cienciaenegocios.com/curva-roc-e-auc-em-machine-learning/>)
- Schmidt, Í. C. B. (2023). Casamento infantil: Uma realidade brasileira. *SISTEMA DE JUSTIÇA, GÊNERO E DIVERSIDADES*.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistic*.
- Selemane. (2019). *O impacto das uniões prematuras na educação, saúde e nutrição em moçambique*. chrome-extension://efaidnbmnnnibpcajpcgiclfefindmkaj/
https://bettercarenetwork.org/sites/default/files/2021-04/O%20impacto%20das%20unio%CC%83es%20prematuras%20na%20educac%CC%A7a%CC%83o%2C%20sau%CC%81de%20e%20nutric%CC%A7a%CC%83o%20-%20FINAL_28Junho2019.pdf. (Acesso em: 9 de Março de 2024)
- Szwarcwald, C. L., & Damacena, G. N. (2008, May). Amostras complexas em inquéritos populacionais: planejamento e implicações na análise estatística dos dados. *Revista Brasileira de Epidemiologia*, 11, 38–45. Retrieved from <https://doi.org/10.1590/S1415-790X2008000500004> doi: 10.1590/S1415-790X2008000500004
- Turkman, A., & Silva, G. (2000). *Modelos lineares generalizados - da teoria à prática* -. Lisboa: Universidade Técnica de Lisboa. (Obtido em 22 de 07 de 2018, de <https://docs.ufpr.br/taconeli/CE225/tp.pdf>)

- UNFPA. (2013). Gravidez na adolescência moçambique desafios e respostas de moçambique. *Relatório Situação da População Mundial*.
- UNICEF. (2011). *Casamentos prematuros em moçambique*. <https://www.unicef.org/mozambique/casamento-prematuro-em-mo%C3%A7ambique>. (Acesso em: 23 de Fevereiro de 2024)
- UNICEF. (2014). *The state of the world's children 2014 in numbers: Every child counts*. New York: United Nations Children's Fund.
- UNICEF. (2018). *25 milhões de casamentos prematuros impedidos na última década*. <https://www.unicef.org/mozambique/comunicados-de-imprensa/25-milh%C3%B5es-de-casamentos-prematuros-impedidos-na-%C3%BAltima-d%C3%A9cada>. (Acesso em: 27 de janeiro de 2024)

Apêndices

Apêndice A

Tabelas

Tabela A.1: Avaliação do ajuste dos modelos (AIC)

Modelo	Não ajustado	Ajustado
Brutos	816.5321	808.8216
Amostragem complexa	594.4594	581.4876

Tabela A.2: Avaliação do modelo

Modelo ajustado	Amostragem complexa
LogLink	-276.9602
PseudoR	0.119646
Poder discriminatório	72.72%
Sensibilidade	97.87%
Especificidade	3.17%
AUC	0.5733

Tabela A.3: Estatística descritiva sem amostragem complexa

Característica	Sim, N = 219	Não, N = 613	p-value
Tercil de riqueza			<0.001
Alto	33 (15%)	288 (47%)	
Baixo	88 (40%)	107 (17%)	
Médio	98 (45%)	218 (36%)	
Número de agregado familiar	4.0 (3.0, 7.0)	7.0 (5.0, 8.0)	<0.001
Nível de escolaridade			<0.001
Nenhum	30 (14%)	24 (3.9%)	
Primário	146 (67%)	267 (44%)	
Secundário	43 (20%)	319 (52%)	
Superior	0 (0%)	3 (0.5%)	
Zona			<0.001
Urbano	79 (36%)	377 (62%)	
Rural	140 (64%)	236 (38%)	
Religião			<0.001
Católica/Anglicana	63 (29%)	179 (29%)	
Islâmica	51 (23%)	75 (12%)	
Nenhuma	14 (6.4%)	37 (6.0%)	
Outras	2 (0.9%)	12 (2.0%)	
Protestante/Evangélica	63 (29%)	252 (41%)	
Zion	26 (12%)	58 (9.5%)	
Acesso á Radio	0.011		<0.001
Não	124 (57%)	286 (47%)	
Sim	95 (43%)	327 (53%)	
Acesso a Televisão	<0.001		
Não	162 (74%)	274 (45%)	
Sim	57 (26%)	339 (55%)	
Acesso ao Telefone	0.581		<0.001
Não	216 (99%)	600 (98%)	
Sim	3 (1.4%)	13 (2.1%)	
Acesso ao Jornal			<0.001
Não	203 (93%)	466 (76%)	
Sim	16 (7.3%)	147 (24%)	
Ocupação	<0.001		<0.001
Não trabalha	137 (63%)	514 (84%)	
Trabalha	82 (37%)	99 (16%)	
Zona			<0.001
Centro	65 (30%)	210 (34%)	
Norte	100 (46%)	123 (20%)	
Sul	54 (25%)	280 (46%)	
Acesso a Mídia			<0.001
Não	108 (49%)	154 (25%)	
Sim	111 (51%)	459 (75%)	

Tabela A.4: Sumário do modelo

Variável	Categoria	Estimativa	Erro-padrão	z valor	p valor
(Intercept)	–	2.57	0.437	5.881	0.7
Tercil de Riqueza	Baixo	1.67	0.7015	-0.202	0.4
	Médio	2.00	0.5296	-0.879	0.094
Nível de escolaridade	Primário	-	-	-	-
	Secundário	-	-	-	-
	Superior	-	-	-	-
Zona	Rural	-	-	-	-
Religiao	Islamica	2.87	0.3870	-0.645	0.018
	Nenhuma	2.38	0.4539	-0.029	0.090
	Outras	0.00	0.7152	20.228	< 0.001
	Protestante/Evangélica	1.42	0.5078	-1.184	0.3
	Zion	3.54	0.4069	-3.694	0.002
Radio	Sim	-	-	-	-
Televisao	Sim	2.01	0.3602	-1.553	0.13
Telefone	Sim	-	-	-	-
Jornal	Sim	0.43	0.3960	-2.064	0.043
Trabalha	Sim	3.34	1.0147	0.657	0.001
Regiao	Norte	-	-	-	-
	Sul	-	-	-	-
Acesso a Mídia	Sim	0.67	0.4035	2.161	0.3