



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos factores associados à
mortalidade infanto-juvenil na África
Austral: uma aplicação dos modelos de
sobrevivência de Cox com fragilidade**

Autora: Shelsea Luís Damião

Maputo, Setembro de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos factores associados à
mortalidade infanto-juvenil na África
Austral: uma aplicação dos modelos de
sobrevivência de Cox com fragilidade**

Autora: Shelsea Luís Damião

Supervisor: Rachid Joel Guidion Muleia, PhD, UHasselt

Maputo, Setembro de 2025

Declaração de Honra

Declaro por minha honra que este trabalho é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para obtenção do grau de Licenciada em **Estatística** na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Setembro de 2025



(Shelsea Luís Damião)

Dedicatória

Dedico este trabalho ao meu Pai, **Luís Damião** (*in memoriam*), por me ter inspirado, através da sua história de vida, a nunca desistir dos meus sonhos e objectivos.

Agradecimentos

Agradeço ao meu supervisor, Doutor Rachid Muleia, pelo seu constante apoio, incentivo, paciência e disponibilidade para me orientar durante o processo de elaboração deste trabalho.

Ao Prof. Doutor Alberto Mulenga, pelos valiosos ensinamentos transmitidos para a elaboração do protocolo durante a cadeira de Seminário.

Aos docentes do Departamento de Matemática e Informática da Faculdade de Ciências, pela didática e pelo profissionalismo na partilha de conhecimentos durante a minha formação.

À minha mãe, Marta Uate, minha eterna fonte de força e inspiração, a quem sou grata pelo apoio incondicional que sempre me ofereceu. Este trabalho é, em grande parte, um reflexo da sua dedicação incansável, dos seus sacrifícios diários e da confiança imensurável que sempre depositou em mim.

À todos aqueles que, de alguma forma, contribuíram para a realização deste trabalho, o meu sincero agradecimento.

Melhor é uma resposta aproximada à pergunta certa, que é frequentemente vaga, do que uma resposta exacta à pergunta errada, que pode sempre ser tornada precisa.”

(Tukey, 1962)

Resumo

Apesar dos avanços globais na redução da mortalidade infanto-juvenil, esta continua a ser um grave problema de saúde pública, especialmente nos países em desenvolvimento e, em particular, na África Austral. Este estudo analisa os factores associados à mortalidade infanto-juvenil nesta região, com o objectivo de identificar as variáveis que influenciam a probabilidade de sobrevivência das crianças. Para o efeito, foram utilizados dados provenientes dos Inquéritos Demográficos e de Saúde de 13 países da África Austral, que abrangem diferentes contextos socioeconómicos, demográficos e materno-infantis. De forma a explorar a dinâmica da mortalidade ao longo do tempo, foi utilizada a análise de sobrevivência, sendo o modelo de riscos proporcionais de Cox com fragilidade lognormal aplicado para avaliar o impacto das covariáveis no tempo de sobrevivência e capturar a influência de factores não observáveis que podem afectar a mortalidade. Os resultados revelaram que crianças do sexo feminino e aquelas com maior número de consultas pré-natais apresentaram um risco reduzido de morte, enquanto partos gemelares, peso à nascença excessivo e intervalos entre nascimento inferiores a 12 meses foram associados a um risco acrescido. Em suma, o estudo enfatiza a necessidade de reforçar os serviços de saúde materno-infantil, promover a educação para a saúde com abordagem familiar e comunitária, fortalecer o trabalho das equipas de saúde comunitária e desenvolver políticas públicas intersectoriais.

Palavras-chave: Mortalidade infanto-juvenil; Análise de Sobrevivência; Regressão de Cox; Modelos de fragilidade; África Austral.

Abstract

Despite global progress in reducing child and youth mortality, it remains a serious public health issue, particularly in developing countries and especially in Southern Africa. This study examines the factors associated with child and youth mortality in this region, aiming to identify the variables that influence children's survival probability. Data were drawn from Demographic and Health Surveys conducted in 13 Southern African countries, covering diverse socioeconomic, demographic, and maternal-child health contexts. To explore the dynamics of mortality over time, survival analysis was employed, with the Cox proportional hazards model with lognormal frailty used to assess the impact of covariates on survival time and to account for unobserved heterogeneity that may affect mortality. The findings indicate that female children and those whose mothers had more antenatal care visits showed a reduced risk of death, while twin births, excessive birth weight, and birth intervals shorter than 12 months were associated with increased risk. Overall, the study underscores the need to strengthen maternal and child health services, promote health education with a family and community-based approach, reinforce the work of community health teams, and develop intersectoral public policies.

Keywords: Under-five child mortality; Survival analysis; Cox regression; Frailty models; Southern Africa.

Lista de Figuras

2.1	<i>Ilustração de censuras</i>	14
4.1	<i>Equiplot comparativo da percentagem de crianças menores de cinco anos mortas por índice de riqueza em diversos países.</i>	42
4.2	<i>Equiplot comparativo da percentagem de crianças menores de cinco anos mortas por Local de residência.</i>	43
4.3	<i>Curva de sobrevivência estimada.</i>	46
4.4	<i>Curvas de Kaplan-Meier estimadas para os grupos da variável País.</i>	46
4.5	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Local de residência (I), Sexo da Criança (II) e Peso à nascença (III).</i>	47
4.6	<i>Representação gráfica dos Resíduos de Martingala e Deviance.</i>	50
5.1	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Duração da amamentação (IV) e Número de consultas pré-natais (V), Ordem de nascimento (VI).</i>	70
5.2	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Tipo de de nascimento(VII), Intervalo entre nascimentos (VIII) e Gravidez indesejada (IX).</i>	70
5.3	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Índice de Riqueza(X) e Número de membros do agregado familiar (XI), Nível de escolaridade do marido (XII).</i>	71
5.4	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Fonte de água (XIII), Sanitário (XIV), Sexo do chefe do AF (XV).</i>	71
5.5	<i>Curvas de Kaplan-Meier estimadas para as variáveis: Uso de rede mosquiteira (XIX), Estado marital (XX), Ocupação da mãe (XXI), Nível de escolaridade da mãe (XXII).</i>	72
5.6	<i>Representação gráfica dos resíduos de Schoenfeld.</i>	72

Lista de Tabelas

3.1	<i>Descrição das variáveis de estudo</i>	36
4.1	<i>Distribuição das variáveis segundo o estado de vida das crianças e Teste de Qui-quadrado</i>	43
4.2	<i>Teste log-rank para comparação das curvas de sobrevivência</i>	48
4.3	<i>Comparação das medidas de ajuste dos modelos de fragilidade</i>	48
4.4	<i>Teste de proporcionalidade dos riscos pelo método de Schoenfeld</i>	49
4.5	<i>Coeficientes de regressão do modelo de Cox com fragilidade lognormal</i>	51
5.1	<i>Seleção de covariáveis para o modelo</i>	73

Lista de Abreviaturas

AF	Agregado Familiar
AIC	Akaike's Information Criterion
DHS	<i>Demographic and Health Survey</i>
HIV	Vírus de Imunodeficiência Humana
IC	Intervalo de Confiança
IDS	Inquérito Demográfico e de Saúde
INE	Instituto Nacional de Estatística
KM	Kaplan-Meier
MISAU	Ministério da Saúde de Moçambique
OMS	Organização Mundial da Saúde
UEM	Universidade Eduardo Mondlane

Índice

1	INTRODUÇÃO	1
1.1	Problema de pesquisa	4
1.2	Objectivos	4
1.2.1	Objectivo geral	4
1.2.2	Objectivos específicos	4
1.3	Justificação	5
1.4	Estrutura do trabalho	5
2	REVISÃO DA LITERATURA	6
2.1	Conceitos de Mortalidade infanto-juvenil	6
2.2	Factores associados à mortalidade infanto-juvenil	6
2.2.1	Factores socioeconómicos e demográficos	7
2.2.2	Factores materno-infantis	9
2.2.3	Factores relacionados às características da criança	10
2.2.4	Outros estudos similares	11
2.3	Conceitos básicos de Análise de Sobrevivência	12
2.3.1	Censura	12
2.3.2	Truncagem	14
2.3.3	Função de sobrevivência	14
2.3.4	Função taxa de falha ou função de risco (<i>hazard</i>)	15
2.3.5	Função taxa de falha acumulada	15
2.3.6	Relações entre as funções de sobrevivência	16
2.3.7	Tempo médio e mediano de sobrevivência	16
2.4	Estimação não-paramétrica de sobrevivência	16
2.4.1	Estimador de Kaplan- Meier	17
2.4.2	Estimador de Nelson-Aalen	17
2.4.3	Intervalos de confiança da função de sobrevivência	18
2.4.4	Testes não paramétricos	18
2.5	Modelo de riscos proporcionais de Cox	19
2.6	Ajuste do modelo de Cox	20
2.7	Modelos probabilísticos de sobrevivência	22

2.7.1	Distribuição exponencial	22
2.7.2	Distribuição gama	22
2.7.3	Distribuição log-normal	22
2.8	Estimação paramétrica de sobrevivência	23
2.8.1	Teste de Wald	23
2.8.2	Métodos de estimação	23
2.9	Modelos de fragilidade	24
2.9.1	Teste de heterogeneidade	26
2.9.2	Distribuições de fragilidade	26
2.9.3	Estimação dos parâmetros do modelo de fragilidade	27
2.9.4	Adequação do modelo de fragilidade	28
2.9.5	Considerações sobre a estimação de modelos de fragilidade	29
2.10	Análise de Resíduos	29
2.10.1	Resíduos de Schoenfeld	30
2.10.2	Resíduos de martingala	30
2.10.3	Resíduos <i>Deviance</i>	31
2.10.4	Resíduos <i>Score</i>	31
2.11	Interpretação dos coeficientes	32
2.12	Análise de amostras complexas	33
3	MATERIAL E MÉTODOS	34
3.1	Material	34
3.2	Métodos	37
3.2.1	Análise exploratória	37
3.2.2	Teste de Qui-quadrado	37
3.2.3	Análise de sobrevivência	38
4	RESULTADOS E DISCUSSÃO	41
4.1	Análise descritiva	41
4.2	Curvas de Kaplan-Meier	45
4.3	Modelo de regressão de Cox com fragilidade	48
4.3.1	Escolha da distribuição de fragilidade	48
4.3.2	Seleção de variáveis	48
4.3.3	Análise residual	49
4.3.4	Interpretação dos coeficientes do modelo multinível	49
4.4	Discussão dos Resultados	52
5	CONCLUSÕES E RECOMENDAÇÕES	56
5.1	Conclusões	56
5.2	Recomendações	56

5.3 Limitações	57
REFERÊNCIAS	69
APÊNDICE	70

Capítulo 1

INTRODUÇÃO

A mortalidade infanto-juvenil é um indicador chave do desempenho do sistema de saúde para os países em todo o mundo (Gobebo, 2021). Entre 1990 e 2018, a taxa global de mortalidade infanto-juvenil reduziu em 52%, passando de 93 mortes por 1000 nascidos vivos para 39 mortes por 1000 nascidos vivos, com uma ligeira redução adicional para 37 mortes por 1000 nascidos vivos em 2022 (United Nations International Children’s Emergency Fund [UNICEF], 2024) . No entanto, apesar deste progresso, os países da África Subsaariana continuam a apresentar taxas notavelmente altas.

Apesar da África Subsaariana representar apenas 30% dos nascimentos em todo o mundo, em 2023, a região foi responsável por 57% das mortes de crianças menores de cinco anos (UNICEF, 2024). Em 2018, a taxa de mortalidade infanto-juvenil na África Subsaariana foi estimada em 78 mortes por 1000 nascidos vivos (Kazembe et al., 2012). Embora tenha havido uma redução para 71 mortes por 1000 nascidos vivos em 2023 , a região ainda enfrenta dificuldades para acelerar o progresso na redução dessa taxa (Organização Mundial da Saúde [OMS], 2022; UNICEF, 2024) .

Contudo, essa tendência de redução da taxa de mortalidade infanto-juvenil não ocorre de forma uniforme, o que evidencia disparidades entre as sub-regiões da África Subsaariana, onde as regiões leste e austral registam as taxas mais elevadas (Hill et al., 2012). Na África Austral, registou-se uma taxa de mortalidade infanto-juvenil de 125 mortes por 1000 nascidos vivos, que aumentou para 146 em 2006, antes de reduzir para 78 mortes por 1000 nascidos vivos em 2016 (Nyamuranga, 2016; OMS, 2013). E embora a taxa seja elevada, há muita heterogeneidade entre os países que compõem a região (Hill et al., 2012; Mahy, 2003).

Em Moçambique, a taxa de mortalidade infanto-juvenil é de 60 mortes por 1000 nascidos vivos. Alguns países da região apresentam taxas abaixo desse valor, como a África do Sul (42), a Tanzânia (43), Comores (50), Lesoto e Namíbia (ambos com 54 mortes por 1000 nascidos vivos). Por outro lado, várias nações registam taxas mais altas, como Angola (68),

Malawi (63), Zâmbia (61), Zimbabwe (69), e Madagáscar (75 mortes por 1000 nascidos vivos). As maiores taxas de mortalidade infanto-juvenil da região são observadas na República Democrática do Congo, com 104, e em Eswatini, com de 120 mortes por 1.000 nascidos vivos (Ministério da Saúde de Moçambique & Instituto Nacional de Estatística de Moçambique, 2023; Ministério da Saúde da África do Sul & Instituto Nacional de Estatística da África do Sul, 2016; Ministério da Saúde de Angola & Instituto Nacional de Estatística de Angola, 2015; Ministério da Saúde de Lesoto & Lesotho Bureau of Statistics, 2014; Ministério da Saúde de Madagascar & Instituto Nacional de Estatística de Madagascar, 2018; Ministério da Saúde de Eswatini & Central Statistical Office of Eswatini, 2014; Ministério da Saúde da Zâmbia & Central Statistical Office of Zâmbia, 2018; Ministério da Saúde e Bem-Estar da Família de Zimbabwe & Zimbabwe National Statistics Agency, 2015; Ministério da Saúde da República Democrática do Congo & Institut National de la Statistique de la République Démocratique du Congo, 2013; Ministério da Saúde das Comores & Institut National de la Statistique des Comores, 2012).

As altas taxas de mortalidade infanto-juvenil que se registam na África Austral e entre os diferentes países que a compõem podem estar associadas a problemas como as epidemias do Vírus de Imunodeficiência Humana (HIV), tuberculose, infecções respiratórias e doenças diarreicas, bem como a elevada prevalência de malária que se tem verificado na região. Neste contexto, as crianças e adolescentes, que representam quase 50% da população da Comunidade para o Desenvolvimento da África Austral (SADC), são particularmente vulneráveis a estas doenças, o que aumenta potencialmente o número de mortes de crianças menores de cinco anos nesta região (Comunidade para o Desenvolvimento da África Austral [SADC], 2017).

As características sociodemográficas, maternas e/ou infantis também são mencionadas na literatura como factores associados a mortalidade infanto-juvenil (Tesfa et al., 2021). Um estudo realizado em Moçambique evidenciou que o progresso na saúde infantil é condicionado por determinantes estruturais, nomeadamente o acesso a serviços básicos como educação, água potável e saneamento, e que, por outro lado, há factores que restringem a melhoria das condições de saúde pública, como a situação socioeconómico, o analfabetismo e as disparidades regionais no acesso a infraestruturas básicas (Taela et al., 2020). Outro estudo realizado na República Democrática do Congo revelou uma associação significativa entre a mortalidade infanto-juvenil e factores como nascimentos domiciliários e crianças que vivem com mães solteiras. Por outro lado, uma pesquisa conduzida na Namíbia destacou que muitas mortes são potencialmente evitáveis, apontando falhas nos cuidados pré-natais, na assistência ao parto e no acompanhamento hospitalar (Mdala & Mash, 2015).

Assim, o presente estudo visa analisar os factores associados à mortalidade infanto-juvenil

na África Austral com base na aplicação de técnicas de análise de sobrevivência de modo a compreender como estes factores afectam a sobrevivência de crianças nesta região.

1.1 Problema de pesquisa

Apesar dos avanços registados em outras partes do mundo, a África Austral continua a enfrentar condições que afectam a sobrevivência de crianças. Entre 1990 e 2006, a taxa de mortalidade infanto-juvenil na região aumentou de 125 para 146 mortes por 1000 nascidos vivos, com uma posterior redução para 78 mortes por 1000 nascidos vivos em 2016. No entanto, esta taxa continua alta, o que dificulta o alcance da meta dos Objectivos de Desenvolvimento Sustentável (ODS 3), que visa reduzir a mortalidade infantil para 25 mortes por 1.000 nascidos vivos até 2030.

Além de revelar a persistência de taxas elevadas, esse cenário mostra que a redução da mortalidade infanto-juvenil na região tem sido insuficiente para cumprir as metas estabelecidas pelos Objectivos de Desenvolvimento Sustentável. Diante do exposto, é fundamental compreender os factores associados à mortalidade infanto-juvenil na região, o que leva à seguinte questão de pesquisa: **Quais são os factores associados à mortalidade infanto-juvenil na África Austral?**

1.2 Objectivos

1.2.1 Objectivo geral

Analisar os factores associados à mortalidade infanto-juvenil na África Austral

1.2.2 Objectivos específicos

- Descrever o perfil socioeconómico e demográfico das crianças menores de cinco anos na África Austral;
- Verificar a associação entre a morte de uma criança menor de cinco anos e cada variável independente;
- Estimar as curvas de sobrevivência para o tempo de vida entre crianças menores de cinco anos na África Austral;
- Comparar o risco de morte de crianças em relação aos factores associados à mortalidade infanto-juvenil na África Austral;
- Estimar os coeficientes de um modelo de regressão que melhor descreve a relação entre variáveis explicativas e o tempo de sobrevivência na África Austral.

1.3 Justificação

A mortalidade infanto-juvenil é um indicador crucial do nível de saúde e desenvolvimento das populações, sendo, por isso, fundamental monitorar, quantificar e qualificar os óbitos infantis, especialmente devido à inevitabilidade da maioria destes. Na África Austral, onde as taxas de mortalidade infantil e infanto-juvenil são influenciadas por múltiplos factores socio-económicos, ambientais e de acesso aos cuidados de saúde, a análise dos factores associados à essas mortes permitirá identificar disparidades e desenvolver estratégias de intervenção mais eficazes, baseadas em evidências específicas da região.

Geralmente, as disparidades nas taxas de mortalidade infanto-juvenil entre países reflectem desigualdades que podem derivar de diferenças no acesso aos cuidados de saúde, à educação e aos recursos económicos. Neste contexto, o presente estudo permitirá não só orientar no desenvolvimento de políticas que promovem a equidade e a justiça social, mas também a concepção de estratégias eficazes para a redução da ocorrência de mortes evitáveis em crianças menores de cinco anos, contribuindo assim para o alcance dos Objectivos de Desenvolvimento Sustentável.

1.4 Estrutura do trabalho

O presente trabalho é composto por cinco capítulos. Sendo os capítulos subsequentes estruturados da seguinte forma:

No Capítulo 2, intitulado **Revisão de Literatura**, aborda-se, por um lado, a questão da mortalidade infanto-juvenil, incluindo os seus factores associados, e, por outro, apresentam-se os conceitos fundamentais que sustentam as técnicas estatísticas utilizadas para alcançar o objectivo geral do estudo.

No Capítulo 3, intitulado **Material e Métodos**, são descritas a fonte de obtenção dos dados, as variáveis consideradas no estudo e a metodologia usada.

No Capítulo 4, intitulado **Resultados e Discussão**, são apresentados os resultados obtidos através das técnicas estatísticas descritas no Capítulo 3. Posteriormente, os resultados são discutidos de forma a destacar semelhanças, diferenças e possíveis explicações.

Finalmente, no Capítulo 5, **Conclusões e Recomendações**, são apresentadas as conclusões do trabalho, suas limitações e as recomendações derivadas dos resultados obtidos. Adicionalmente, o trabalho inclui a apresentação das referências, o apêndice e o anexo.

Capítulo 2

REVISÃO DA LITERATURA

2.1 Conceitos de Mortalidade infanto-juvenil

A mortalidade infanto-juvenil mede a probabilidade de um indivíduo falecer antes de completar cinco anos de idade (0-59 meses) (INE & MISAU, 2024). A taxa de mortalidade em menores de cinco anos quantifica esta probabilidade, expressa por cada mil nascidos vivos, e descreve o risco de uma criança não atingir o quinto ano de vida (Bonita & Beaglehole, 2010).

A mortalidade infanto-juvenil pode ainda ser analisada sob o ponto de vista de diferentes segmentos, neonatal, pós-neonatal, infantil e pós-infantil. Segundo E. Oliveira et al. (2020), a mortalidade neonatal refere-se aos óbitos ocorridos em menores de 28 dias de idade (até 27 dias). A mortalidade pós-neonatal pode ser definida como sendo o número de mortes em crianças com 28 dias e menos de um ano, a mortalidade infantil é definida como o óbito de uma criança antes de completar um ano de vida e a mortalidade pós-infantil refere-se à probabilidade da criança morrer entre 12 e 59 meses de vida (Ely & Driscoll, 2022).

2.2 Factores associados à mortalidade infanto-juvenil

Estudos recentes, realizados em várias regiões de África, identificaram factores sociodemográficos, maternos e infantis como determinantes principais para a sobrevivência de crianças menores de cinco anos (Tesfa et al., 2021). De acordo com Yaya et al. (2018), a área de residência das famílias é um factor relevante, enquanto Sastry (2004) e Aly e Grabowski (1990) apontam a escolaridade materna e do pai, respectivamente, como factores significativos. Além disso, conforme destacado por Kanmiki et al. (2014), o estado civil materno é também determinante.

A literatura revela ainda que o sexo da criança, o local do parto e o intervalo entre nascimentos têm um papel importante. Outros factores, como gestações de gémeos, o número

de nascimentos no mesmo parto, a idade materna, a exposição a meios de comunicação, o nível socioeconómico familiar, o estado nutricional infantil e o tipo de parto, são também considerados como preditores significativos da mortalidade infantil (Costa & Victora, 2017; Ettarh, 2012; Rutstein, 2005; Amuka et al., 2024; Kozuki, 2013; Aheto, 2019; Sinha, 2016; Yaya et al., 2018) .

2.2.1 Factores socioeconómicos e demográficos

As características da área de residência têm sido enfatizadas como um dos determinantes da mortalidade infantil. A área de residência (urbana ou rural) afecta de forma indirecta os níveis de mortalidade ao condicionar o acesso das mulheres aos serviços de saúde. Em áreas urbanas, por exemplo, observa-se maior disponibilidade de transporte para as unidades sanitárias, maior facilidade de acesso a recursos financeiros para aquisição de medicamentos, além de melhores oportunidades de educação e maior capacidade de providenciar condições básicas no ambiente familiar (Mahy, 2003). Como resultado, as taxas de mortalidade infantil tendem a ser inferiores nos contextos urbanos do que rurais (Akinyemi et al., 2013).

Ainda associado ao local de residência, que geralmente compromete a distância das residências até as unidades sanitárias. As evidências disponíveis sugerem que as mães que têm partos em unidades sanitárias têm menores probabilidades de relatar a morte infantil, em comparação com aquelas que optam por ter o parto em casa, uma vez que as unidades oferecem um ambiente sanitário adequado e assistência médica especializada (Dhawan et al., 2023; Justice et al., 2012).

As condições socioeconómicas também intensificam as diferenças no acesso aos cuidados de saúde. Em contextos de maior vulnerabilidade económica, é menos provável que uma mãe procure serviços de saúde de qualidade para o seu filho, diferente do que pode ocorrer com mulheres em condições financeiras mais favoráveis. Além disso, pais com menos recursos tendem a não conseguir custear cuidados de saúde adequados, sendo também mais propensos a adiar a procura de tratamento ou a recorrer a medicamentos tradicionais para os filhos (Kunnuji et al., 2021). Em geral, a mortalidade entre crianças com menos de cinco anos é particularmente elevada nos países em desenvolvimento, sendo um indicador importante do nível de pobreza desses países (Wagstaff & Watanabe, 1999; Kyereme & Amo-Adjei, 2016).

O nível de escolaridade dos pais é também um factor de destaque na literatura que é frequentemente indicado como um determinante na sobrevivência infantil. De facto, um estudo realizado por Adebowale et al. (2020) revelou que, na África Subsariana, a educação dos pais tem um impacto directo na sobrevivência de crianças menores de cinco anos. Os resultados indicaram que a mortalidade em crianças é menor em famílias onde ambos os cônjuges

possuem, pelo menos, o ensino secundário, e mais elevada em famílias onde os pais têm, no máximo, o ensino primário. Além disso, num estudo conduzido por Mirza et al. (2021), observou-se que a mortalidade infantil é inferior em famílias onde as mães têm um nível educacional superior ao dos pais, em comparação com aquelas em que o contrário ocorre. Nesse sentido, mães mais escolarizadas e com acesso aos meios de comunicação têm maior acesso a informações sobre cuidados infantis, incluindo nutrição, higiene e sinais de alerta para doenças, além de que mulheres com níveis mais altos de educação tendem a possuir maiores habilidades cognitivas, de compreensão e de comunicação, as quais se tornam determinantes de comportamentos mais saudáveis e, conseqüentemente, de uma menor mortalidade infantil e infanto-juvenil (Ayele et al., 2017; Muhuri, 1995). Por outro lado, estudos como o de Bicego e Boerma (1993), apontam que o efeito do nível de escolaridade dos pais pode ser observado no sentido inverso, já que pode estar associado à falta de tempo para cuidados directos da criança.

No que concerne ao sexo do chefe do agregado familiar, conforme estudado por Ekholuenetale et al. (2020), crianças menores de cinco anos em famílias lideradas por mulheres apresentam uma redução significativa de 10% no risco de mortalidade, em comparação com aquelas de famílias lideradas por homens. Este estudo constatou que, entre muitos outros factores, a chefia do lar é um determinante significativo da mortalidade infantil, o que sugere que a autonomia e o fortalecimento das mulheres, por meio do aumento da alfabetização materna e da capacidade de decidir de forma independente sobre a utilização dos serviços de saúde materna, incluindo o acesso a cuidados pediátricos, podem contribuir para a redução dos casos de morte infantil (Ekholuenetale et al., 2020). Por outro lado, há estudos que associam a liderança feminina no agregado familiar a um aumento na mortalidade infanto-juvenil, devido à maior probabilidade de insegurança alimentar e à sobrecarga de responsabilidades domésticas e sociais, o que dificulta a tomada de decisões sobre a saúde infantil (Girma et al., 2023).

Para além da chefia do agregado familiar, o estado civil da mãe também exerce influência na mortalidade infanto-juvenil. Segundo Mahy (2003), as mães solteiras geralmente são mais jovens, menos instruídas e possuem um estatuto socioeconómico mais baixo do que as mães casadas. Essa vulnerabilidade socioeconómica tem sido associada a piores desfechos de saúde materno-infantil, incluindo taxas mais elevadas de mortalidade neonatal e pós-neonatal. Estudos recentes estão alinhados a esses achados, por exemplo, Fonseca et al. (2017) também identificaram que mães adolescentes e aquelas com baixa escolaridade apresentam maiores casos de morte neonatal.

Outro aspecto importante a considerar é o número de membros no agregado familiar, em que segundo Ayele et al. (2017), o risco de uma criança não atingir os cinco anos de idade

aumenta à medida que o número de membros da família cresce, sobretudo para mulheres desempregadas. Os autores apontam ainda que a interação entre o tamanho da família e a situação laboral afecta de forma significativa a mortalidade infantil. De facto, as diferentes estruturas familiares proporcionam ambientes diversificados para as crianças, sendo um dos factores mais relevantes para o fortalecimento da saúde infantil (Ekholuenetale et al., 2020). Por outro lado, pesquisas, incluindo o estudo conduzido por Argawu e Hirko (2020), demonstram que crianças oriundas de famílias compostas por quatro a seis membros ou mais apresentam uma probabilidade reduzida de falecer antes dos cinco anos, em comparação às crianças provenientes de famílias menores, que possuem de um a três integrantes.

A literatura aponta igualmente o acesso a fontes melhoradas de água potável como um factor determinante para a redução da mortalidade infanto-juvenil. Tradicionalmente, a insegurança hídrica está associada à concentração de doenças diarreicas, o que representa um factor de risco associado a mortalidade infantil (Rhue et al., 2023). Outro aspecto relacionado à ocorrência de doenças em crianças é a falta de uso de rede mosquiteira nos agregados familiares, sendo a utilização deste tipo de meio de prevenção ainda desejável e recomendável (Fink et al., 2022).

2.2.2 Factores materno-infantis

A probabilidade de sobrevivência das crianças é parcialmente condicionada por factores demográficos e biológicos maternos. Um estudo conduzido na África Oriental demonstrou que crianças nascidas de mães com 35 anos ou mais enfrentam um risco maior de mortalidade antes dos cinco anos, em comparação com aquelas cujas mães têm entre 15 e 24 anos (Tessema et al., 2021). Esta relação pode dever-se ao aumento da incidência de complicações obstétricas em idades maternas mais avançadas, como hipertensão gestacional, diabetes e maior probabilidade de partos prematuros (Cleary Goldman et al., 2005). Além disso, mães mais velhas podem ter menos reservas biológicas para suportar uma gravidez saudável, o que pode impactar negativamente a saúde do recém-nascido (Jolly et al., 2000).

Estima-se que cerca de 55 a 75% dos óbitos infantis se devem a práticas inadequadas de amamentação entre as crianças afectadas, sendo que uma parte dessas mortes seria evitável caso a prevalência de práticas adequadas de amamentação aumentasse na África Subsaariana. Nesta região, as taxas de amamentação são inferiores em comparação com outros países de rendimentos baixo e médio, e a região apresenta as maiores taxas globais de mortalidade infantil em crianças com menos de cinco anos (Pretorius et al., 2020).

Além da amamentação, outra estratégia fundamental para reduzir a mortalidade infantil e

infanto-juvenil é o acesso a cuidados pré-natais adequados. A realização de pelo menos uma consulta pré-natal com um profissional de saúde qualificado durante a gravidez pode diminuir em aproximadamente 39% o risco de mortalidade neonatal, infantil e infanto-juvenil nos países da África Subsariana (Tekelab et al., 2019). Deste modo, para acelerar o avanço na diminuição da mortalidade entre recém-nascidos, é essencial que as gestantes tenham acesso a cuidados pré-natais ao longo da gravidez. A disponibilidade e qualidade desses serviços permitem prevenir complicações obstétricas, minimizar danos à saúde materno-infantil e reduzir a incidência de partos prematuros e do baixo peso ao nascer, além de contribuir para garantia de partos seguros e nascimentos mais favoráveis à sobrevivência neonatal (Olds et al., 2014).

A mortalidade infanto-juvenil também pode ser influenciada pela ordem e intervalo entre nascimentos. A investigação conduzida por Chikandiwa et al. (2014) evidenciou uma associação positiva entre a ordem de nascimento e nascimentos de alto risco. Em particular, os nascimentos primogénitos aumentam a susceptibilidade a complicações relacionadas com a gravidez, como hemorragia, pré-eclâmpsia, prematuridade e sofrimento fetal, factores que podem aumentar o risco de mortalidade antes do quinto aniversário. Por outro lado, Mahy (2003) identificou uma relação entre ordens de nascimentos elevadas e intervalos curtos entre gestações, dado que mulheres com um maior número de filhos tendem a tê-los com menor espaçamento temporal. O estudo demonstrou ainda que o risco de mortalidade infanto-juvenil é mais elevado tanto nos primeiros nascimentos como a partir do oitavo filho.

Outro estudo realizado por Akinyemi et al. (2013) na Zâmbia revelou uma forte associação entre múltiplos nascimentos e mortalidade infanto-juvenil, evidenciando que crianças nascidas de partos gemelares estão mais expostas ao risco de morte em comparação com as nascidas de partos únicos. Para além disso, outro factor importante destacado na literatura como preditor da mortalidade infanto-juvenil é a ocorrência de nascimentos indesejados. Um estudo realizado na Tanzânia revelou que os nascimentos indesejados frequentemente estão associados à uma recepção inadequada de nutrição e a um acesso limitado a serviços básicos de saúde infantil (Ogbo et al., 2019). A literatura aponta também para o tipo de parto como sendo determinante para a mortalidade infanto-juvenil, visto que a cesariana não indicada pelo médico está associada a uma maior mortalidade em comparação com os partos normais (Betrán et al., 2016).

2.2.3 Factores relacionados às características da criança

As crianças do sexo masculino apresentam um risco aumentado de mortalidade antes dos cinco anos. Este fenómeno é ainda mais acentuado nos primeiros dias após o nascimento,

com a idade gestacional a desempenhar um papel relevante nesse contexto, uma vez que recém-nascidos do sexo masculino têm maior tendência a nascer mais prematuramente, o que contribui para uma maior vulnerabilidade a complicações neonatais (Sahu & Nair, 2015; Remoaldo, 2002).

Outra característica da criança que afecta a sua sobrevivência é o seu tamanho ao nascer. Crianças com peso muito baixo ou muito alto ao nascer têm um risco maior de mortalidade antes de atingirem os cinco anos, em comparação com aquelas de peso médio. Os bebés com baixo peso ao nascer ou bebés macrossômicos, aqueles que apresentam peso ao nascer acima da média, podem apresentar imaturidade nos órgãos e condições médicas como doenças cardíacas congénitas, síndrome de *Down*, HIV/SIDA e Diabetes Mellitus, que aumentam o risco de mortalidade dentro dos primeiros cinco anos de vida (Croft et al., 2023; Dhawan et al., 2023).

2.2.4 Outros estudos similares

Vários são os estudos que investigam a relação entre mortalidade infanto-juvenil e diferentes factores, tanto a nível global como no contexto Africano. Na África Subsariana destaca-se o estudo realizado por Yaya et al. (2018) cujo objectivo era examinar os efeitos das variáveis explicativas na mortalidade infanto-juvenil. Nele, foram estimados coeficientes de modelos de riscos proporcionais de Cox, numa abordagem univariada, ajustados separadamente para cada base de dados dos Inquéritos Demográficos e de Saúde de distintos países. Posteriormente, a partir das variáveis que foram estatisticamente significativas, um conjunto de modelos de regressão multivariada foi obtido separadamente para cada país.

Por sua vez, Birhanie et al. (2025) realizaram um estudo na África Subsariana para identificar os factores associados à mortalidade infanto-juvenil. Para isso, foram compilados dados do IDS de 33 países da região, abrangendo o período de 2010 a 2020, e analisados simultaneamente com recurso ao modelo logístico multinível com abordagem bayesiana, pois foi considerado que crianças pertencentes ao mesmo *cluster* tendem a ser mais semelhantes entre si do que em relação a crianças de outros *clusters*.

Na Etiópia os factores associados à mortalidade infanto-juvenil foram avaliados através do estudo realizado por Yalew et al. (2022) que visava analisar o tempo até a morte de crianças menores de cinco anos e seus preditores na zona rural, assim, foi aplicado o modelo de fragilidade compartilhada Cox-Gamma. Nesse estudo, constatou-se uma diferença significativa na mortalidade em menores de cinco anos entre os domicílios cobertos por seguro de saúde e os não cobertos, conforme o teste Log-Rank. Constatou-se também que a relação entre a zona de residência e o seguro de saúde está associado a um risco acrescido de mortalidade

entre crianças com menos de cinco anos, o que pode estar relacionado com facto de que a maioria das pessoas nas zonas rurais pode temer o custo dos cuidados de saúde e ter um comportamento de busca de saúde inadequado para doenças infantis.

No estudo conduzido por Yeshaw et al. (2020) , cujo objectivo foi analisar os factores associados à mortalidade infanto-juvenil, recorreu-se à aplicação da regressão logística para analisar a relação entre diversas covariáveis e a mortalidade. Contudo, os autores reconhecem como uma limitação da sua abordagem o facto de que a técnica aplicada não explora a dinâmica de sobrevivência que pode ser aproveitada a partir das técnicas de análise de sobrevivência, uma vez que um modelo logístico assume que a probabilidade de eventos de morte é a mesma no tempo contínuo de zero a cinco anos, o que restringe a compreensão da evolução do risco ao longo do tempo.

2.3 Conceitos básicos de Análise de Sobrevivência

Conforme Kleinbaum e Klein (2012) , a análise de sobrevivência consiste em um conjunto de procedimentos estatísticos destinados a estudar situações em que a variável de interesse é o tempo até a ocorrência de um determinado evento. Esse tempo pode ser expresso em dias, semanas, meses ou anos, enquanto o evento em questão pode corresponder à morte, a incidência de uma doença ou qualquer outro evento relevante na vida de um indivíduo (Emmert-Streib & Dehmer, 2019).

No contexto da análise de sobrevivência, a variável temporal é o tempo de sobrevivência e o evento de interesse é a falha. A variável aleatória T , contínua e não negativa, representa o tempo de sobrevivência. Essa variável é habitualmente caracterizada por meio da função de sobrevivência ou da função taxa de falha (Colosimo & Giolo, 2024; Klein & Moeschberger, 2005). A análise de sobrevivência não se limita a problemas médicos, mas também pode ser aplicada a questões nas ciências sociais, engenharia ou *marketing* (Emmert-Streib & Dehmer, 2019).

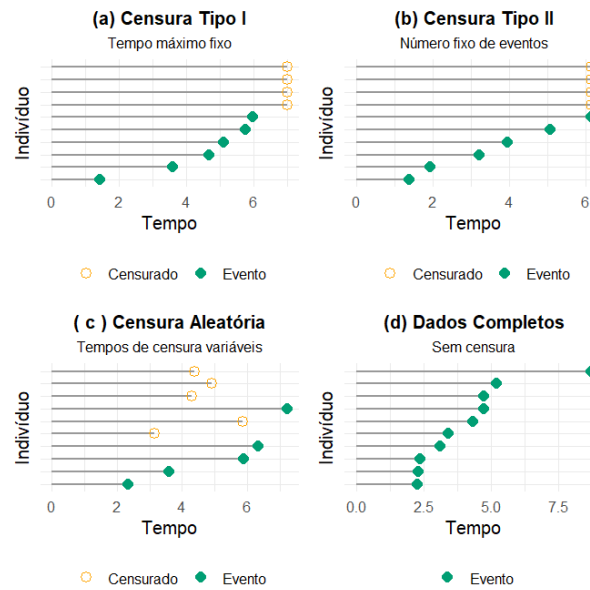
2.3.1 Censura

A maioria das análises de sobrevivência deve considerar um problema analítico fundamental denotado censura, que em essência, refere-se à presença de dados incompletos ou parcialmente observados, situação que pode ocorrer por diversas razões, entre elas, o término do período do estudo antes da ocorrência do evento, a perda de seguimento do indivíduo ou a ocorrência de evento concorrente que inviabiliza a observação do evento de interesse (Colosimo & Giolo, 2024; Schober & Vetter, 2024). Na ausência de censuras, as técnicas clássicas, como a análise de regressão e análise de variância, poderiam ser consideradas desde que

a variável resposta fosse transformada. Contudo, na presença de dados censurados, essas abordagens deixam de ser apropriadas. Assim, torna-se imprescindível recorrer a técnicas específicas para a análise sobrevivência, capazes de incorporar na análise a informação contida tanto nas observações completas quanto nas parciais (Colosimo & Giolo, 2024).

A censura pode ser classificada como:

- **Censura à direita:** designa-se, como tal, a observação incompleta do tempo até a ocorrência do evento de interesse, quando este ocorre para além do tempo registado. Conforme ilustrado na Figura 2.1, esse tipo de censura subdivide-se, por sua vez, em outros três tipos de censura:
 - **Censura do tipo I:** ocorre quando os estudos são encerrados após um período previamente estabelecido, sendo registados, no seu término, indivíduos que ainda não apresentaram o evento de interesse.
 - **Censura do tipo II:** verifica-se em estudos que são concluídos após a ocorrência do evento de interesse a um número de indivíduos previamente fixado.
 - **Censura aleatória:** ocorre quando o indivíduo é retirado do estudo antes da ocorrência do evento (Colosimo & Giolo, 2024; Carvalho et al., 2011).
- **Censura à esquerda:** este tipo de censura verifica-se quando o momento da ocorrência do evento de interesse antecede o instante em que o indivíduo passou a ser acompanhado (Colosimo & Giolo, 2024).
- **Censura intervalar:** este tipo de censura verifica-se em estudos em que os participantes são monitorados em visitas periódicas. Nesses casos, sabe-se apenas que o evento de interesse ocorreu em um determinado intervalo de tempo. Assim, o tempo até ao evento T é conhecido apenas por pertencer ao intervalo $T \in (L, U]$ (Colosimo & Giolo, 2024).

Figura 2.1: *Ilustração de censuras*

2.3.2 Truncagem

A truncagem ocorre quando o estudo apenas inclui indivíduos cujo tempo de vida se encontra dentro de um intervalo pré-estabelecido, denotado (Y_L, Y_R) . Se o evento de interesse tiver ocorrido fora desses limites, o indivíduo não é incluído no estudo (Carvalho et al., 2011). Diferentemente da censura, onde há pelo menos alguma informação parcial, na truncagem não se obtém qualquer dado sobre o indivíduo, nem se quer se reconhece a sua existência. (Colosimo & Giolo, 2024).

A truncagem pode ser à esquerda ou à direita:

- **Truncagem à esquerda:** é comum em estudos de análise de sobrevivência e ocorre quando $Y_R = +\infty$. Neste tipo de truncagem são incluídos no estudo os indivíduos cujo tempo de vida é superior ao tempo de truncagem Y_L . O instante Y_L é frequentemente designado tempo de entrada tardia, uma vez que o seguimento dos indivíduos tem início a partir deste instante (Anchisi, 2011).
- **Truncagem à direita:** ocorre quando o estudo inclui exclusivamente os indivíduos que apresentam o evento de interesse dentro do período de observação (Coelho, 2017).

2.3.3 Função de sobrevivência

De acordo com Lee (1980), citado por Licima (1992), o tempo de sobrevivência é geralmente caracterizado por três funções: a função de sobrevivência, a função densidade de probabilidade (f.d.p.) e a função de risco. Seja T uma variável aleatória contínua e não negativa, que

representa os tempos de vida dos indivíduos de uma dada população. Conforme Rocha e Papoila (2009), a função de sobrevivência associada a T é definida como a probabilidade de um indivíduo não falhar até ao tempo t . Matematicamente, esta função pode ser expressa como:

$$S(t) = P(T > t), \quad t \geq 0 \quad (2.1)$$

sendo que $S(0) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$. Consequentemente, a função de distribuição acumulada é dada por:

$$F(t) = P(T \leq t) = 1 - S(t), \quad (2.2)$$

onde $F(t)$ dada pela equação 2.2, representa a probabilidade do indivíduo falhar até o tempo t , ou seja, a probabilidade de não sobreviver além do tempo t (Anchisi, 2011).

2.3.4 Função taxa de falha ou função de risco (*hazard*)

A função de sobrevivência é frequentemente definida em termos da função de risco, conhecida como a taxa de falha instantânea. Esta representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t (Colosimo & Giolo, 2024). Matematicamente, a definição de risco pode ser expressa como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t}. \quad (2.3)$$

É importante ressaltar que, embora se utilize a designação de risco, $\lambda(t)$ é uma taxa, e não uma probabilidade, podendo assumir qualquer valor real maior ou igual a zero (Carvalho et al., 2011). Ao longo do tempo, podem ser identificados diferentes padrões de risco de mortalidade. Um padrão crescente é caracterizado pelo aumento progressivo do risco de morte à medida que o indivíduo envelhece, conforme indicado por J. Silva et al. (2015). Já o padrão decrescente, menos comum, verifica-se quando o risco de morte é inicialmente elevado, mas reduz-se com o passar do tempo, como ocorre em pacientes submetidos a transplantes (J. Silva et al., 2015).

2.3.5 Função taxa de falha acumulada

Outra função relevante é a função taxa de falha acumulada. Esta que fornece a soma acumulada das taxas de falhas $\lambda(u)$ dos indivíduos até o tempo t . Ela é útil na análise da função taxa de falha quando esta é difícil de ser estimada através de métodos não paramétricos (C. Oliveira, 2023). A expressão matemática é definida por:

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (2.4)$$

2.3.6 Relações entre as funções de sobrevivência

Uma característica relevante das funções de sobrevivência, taxa de falha, taxa de falha acumulada e densidade de probabilidade, é que estas são matematicamente relacionadas. Essa relação é útil nos processos de estimação, permitindo que, quando uma das funções é conhecida seja possível deduzir as restantes (Colosimo & Giolo, 2024). Essas relações são expressas por:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log(S(t)), \quad (2.5)$$

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log(S(t)) \quad (2.6)$$

e

$$S(t) = \exp \{-\Lambda(t)\} = \exp \left\{ -\int_0^t \lambda(u) du \right\}. \quad (2.7)$$

2.3.7 Tempo médio e mediano de sobrevivência

O tempo médio de sobrevivência é, segundo Aquino (2024), o valor esperado do tempo de sobrevivência,

$$E(T) = \int_0^\infty t f(t) dt \quad (2.8)$$

que, usando a integração por partes padrão, pode ser escrito como:

$$E(T) = \int_0^\infty S(t) dt \quad (2.9)$$

O tempo médio de sobrevivência só é definido se $S(\infty) = 0$, ou seja, se todos os sujeitos eventualmente falham (Colosimo & Giolo, 2024).

Por outro lado, o tempo mediano de sobrevivência fornece uma medida alternativa ao tempo médio. Este representa o tempo em que metade da população do estudo já experienciou o evento de interesse e a outra metade ainda não (Aquino, 2024). Formalmente, de acordo com Garcia (2013), o tempo mediano de sobrevivência pode ser definido como:

$$\hat{t}_{\text{med}} = \min\{t \mid \hat{S}(t) \leq 0.5\}, \quad (2.10)$$

ou seja, é o menor t tal que a função de sobrevivência é inferior ou igual a 0,5.

2.4 Estimação não-paramétrica de sobrevivência

Ao modelar a sobrevivência de humanos ou animais, é difícil saber qual família paramétrica escolher, e muitas vezes nenhuma das famílias disponíveis tem flexibilidade suficiente para

isto. Assim, em aplicações médicas e de saúde, métodos não paramétricos, que têm a flexibilidade para lidar com as peculiaridades da sobrevivência de seres vivos, têm consideráveis vantagens (Tiruneh et al., 2024).

2.4.1 Estimador de Kaplan- Meier

Quando há acompanhamento contínuo do estado de vida dos pacientes, a estimativa da função de sobrevivência deve ser feita, preferencialmente através do estimador de Kaplan-Meier, devido às suas propriedades teóricas (Bereta, 2002). Este método baseia-se nos princípios de independência de eventos e de probabilidade condicional para decompor a condição de sobreviver até o tempo t em uma sequência de eventos independentes que reflectem a sobrevivência em cada intervalo de tempo precedente a t e cuja probabilidade é condicionada aos indivíduos que estão em risco em cada período (Kaplan & Meier, 1958; Shimakura & Carvalho, 2005).

O estimador apresenta degraus característicos nos instantes de tempo t em que ocorrem as falhas. Formalmente expressa-se por :

$$\hat{S}(t) = \prod_{j: t_{(j)} \leq t} \left(\frac{n_j - d_j}{n_j} \right), \quad (2.11)$$

sendo $\hat{S}=1$ para $0 \leq t < t_{(1)}$. Em que os tempos de falhas $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ representam os tempos de falhas ordenados, d_j é o número de falhas ocorridas dentro do intervalo j e n_j é o número de indivíduos que não falharam até o instante anterior a $t_{(j)}$ (Colosimo & Giolo, 2024).

2.4.2 Estimador de Nelson-Aalen

Um estimador alternativo da função de sobrevivência é conhecido como estimador de Nelson-Aalen. Este estimador foi proposto inicialmente por Nelson (1972) e retomado por Aalen (1978). Ele é baseado na relação da função de sobrevivência com a função de risco. Uma estimativa da função de risco acumulado é a soma dos riscos estimados até um tempo t_i :

$$\Lambda(t) = \sum_{t_i \leq t} \frac{d_i}{n_i} \quad (2.12)$$

e a estimativa da função de sobrevivência é expressa por:

$$S(t) = e^{-\Lambda(t)}. \quad (2.13)$$

No entanto, o estimador em 2.12 requer um grande número de falhas em cada ponto no tempo para ser útil (Hosmer et al., 2008).

2.4.3 Intervalos de confiança da função de sobrevivência

Para que se possa construir intervalos de confiança e testar hipóteses para $S(t)$, é necessário avaliar a precisão do estimador. A variância assintótica desse estimador é dada pela fórmula de *Greenwood*:

$$\widehat{Var} [\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \quad (2.13)$$

Ao definir um nível de confiança $1 - \alpha$, espera-se que o verdadeiro valor da função de sobrevivência, num instante pré-determinado t_0 , se situe dentro do respectivo intervalo. Assim, o intervalo de confiança para a função de sobrevivência num instante t_0 é determinado pela expressão:

$$\hat{S}(t_0) - z_{1-\alpha/2} \sqrt{\widehat{var}\hat{S}(t_0)}, \quad \hat{S}(t_0) + z_{1-\alpha/2} \sqrt{\widehat{var}\hat{S}(t_0)}, \quad (2.14)$$

em que $z_{1-\alpha/2}$ representa o quantil de probabilidade $1 - \alpha/2$ da distribuição normal e $\widehat{var}\hat{S}(t_0)$ corresponde à estimativa da variância do estimador de Kaplan-Meier no ponto t_0 (Anchisi, 2011). Pode acontecer que os seus limites estejam fora do intervalo $(0, 1)$ quando a estimativa $\hat{S}(t_0)$ estiver próxima de zero ou de um. Uma alternativa é utilizar uma transformação para $\hat{S}(t_0)$, como, por exemplo, $\log [-\log \hat{S}(t_0)]$ (Hosmer et al., 2008).

2.4.4 Testes não paramétricos

Teste log-rank

O teste de Mantel & Haenszel (1959), ou log-rank, é amplamente utilizado para comparar valores observados e esperados do estrato sob a hipótese de que o risco é o mesmo para todos os grupos. Conforme referem Emmert-Streib e Dehmer (2019), a hipótese a ser testada é:

$$H_0 : \lambda_1(t) = \lambda_2(t)$$

Para a obtenção da estatística de teste, considera-se que existam k tempos de falha distintos $t_{(1)}, t_{(2)}, \dots, t_{(k)}$, na amostra formada pela combinação dos dois grupos e que no tempo $t_{(j)}$, m_{1j} indivíduo no grupo 1 e m_{2j} indivíduo no grupo 2 falham, $j=1,2,\dots,k$. Os indivíduos sob risco imediatamente inferior a $t_{(j)}$ são representados por n_{1j} e n_{2j} , nos grupos 1 e 2, com $n_j = n_{1j} + n_{2j}$ (Pereira, 2002).

A estatística do teste é dada por :

$$\frac{\sum_{j=1}^k (m_{ij} - e_{ij})^2}{Var(D)}, \quad (2.15)$$

em que

$$e_{ij} = \frac{n_{ij}}{n_j} (m_{1j} + m_{2j}) \quad (2.16)$$

e

$$Var(D) = \sum_{j=1}^k \frac{n_{1j} n_{2j} (m_{1j} + m_{2j}) (n_{1j} - m_{1j} - m_{2j})}{n_j^2 (n_j - 1)}. \quad (2.17)$$

Desta forma, o valor da estatística de teste é comparado com o valor crítico da distribuição χ^2 , com $k-1$ graus de liberdade. A hipótese nula (H_0) é rejeitada caso o valor calculado seja superior ao valor crítico tabelado (Rizopoulos, 2019).

O teste de log-rank é considerado o mais eficaz na detecção de afastamento da hipótese de igualdade das distribuições que sejam do tipo riscos proporcionais. Além disso, apresenta bom desempenho mesmo em situações em que as funções de risco não são proporcionais, desde que estas não se cruzem (Rocha & Papoila, 2009). Assim, a validade da hipótese de riscos proporcionais pode ser avaliada de forma preliminar por meio da análise gráfica das funções de sobrevivência estimadas. Quando estas não se cruzam, tal hipótese é sustentada, e o uso do teste de log-rank é considerado apropriado (Rocha & Papoila, 2009).

Teste Peto

O teste Peto & Peto (1972) é uma adaptação do teste de log-rank, na qual é atribuído um peso maior ao perfil de sobrevivência dos tempos menores, enfatizando as diferenças (ou semelhanças) no início da curva. Essa abordagem baseia-se no facto de que os tempos iniciais concentram a maior parte dos dados e, por isso são, mais informativos. Para conseguir esta ponderação maior nas fases iniciais, o teste incorpora um factor de ponderação correspondente ao valor estimado da sobrevivência $S(t)$ no estimador log rank (Carvalho et al., 2011).

2.5 Modelo de riscos proporcionais de Cox

Em muitos estudos, as covariáveis podem estar relacionadas com o tempo de sobrevivência, e para isso, alguns modelos de regressão apropriados para dados censurados têm sido propostos na literatura. O método estatístico tradicionalmente utilizado é o modelo de Cox (1972), caracterizado pela sua robustez e adaptabilidade às diferentes situações, tais como a estratificação, a inclusão de variáveis dependentes do tempo, a análise de eventos múltiplos, a consideração de riscos competitivos e a modelação de efeitos de fragilidade (Minguillo, 2016).

Para definição do modelo, considera-se T uma variável aleatória contínua que representa o tempo de sobrevivência, $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ um vector de covariáveis, e $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ o vector dos coeficientes de regressão ainda desconhecidos. O modelo de Cox assume que:

$$\lambda(t|x) = \lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}) = \lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p), \quad (2.18)$$

nesta formulação, $\lambda_0(t)$ representa uma função não negativa do tempo, conhecida como risco de base quando $\mathbf{x} = 0$. Os coeficientes de regressão $\beta_j, j = 1, 2, \dots, p$ reflectem o efeito das variáveis preditoras sobre o tempo até a ocorrência do evento em estudo (Hosmer et al., 2008). Este modelo assume que as covariáveis influenciam a função de risco de forma multiplicativa, de modo que, para dois indivíduos com preditores $\mathbf{x}_r$ e $\mathbf{x}_s$, respectivamente, a razão entre as suas funções de risco é dada por:

$$\frac{\lambda(t|\mathbf{x}_s)}{\lambda(t|\mathbf{x}_r)} = \frac{\lambda_0(t) \exp(\mathbf{x}_s'\boldsymbol{\beta})}{\lambda_0(t) \exp(\mathbf{x}_r'\boldsymbol{\beta})} = \exp\{(\mathbf{x}_s - \mathbf{x}_r)'\boldsymbol{\beta}\}. \quad (2.19)$$

Esta razão de riscos é independente do tempo t . Assim, por exemplo, se no início do estudo o indivíduo r apresenta um risco de morte que é o dobro do risco do indivíduo s , então essa razão de riscos manter-se-á constante ao longo do período de acompanhamento. Essa característica define o modelo como um modelo de riscos proporcionais (Anchisi, 2011).

Uma consideração essencial na estimação do modelo de Cox é o tamanho da amostra. Conforme indicado por Schober e Vetter (2024), a eficácia da análise de dados de sobrevivência está mais relacionada ao número de eventos do que ao total de observações, pelo que recomenda a inclusão de uma covariável no modelo, seja necessário observar pelo menos 10 eventos.

2.6 Ajuste do modelo de Cox

O vector de parâmetros $\boldsymbol{\beta}$ é estimado por meio da função de verossimilhança parcial, na qual elimina-se a função de risco de base, considerando-se apenas a informação dos indivíduos sob risco a cada instante t . Esta abordagem é análoga à dos modelos não paramétricos de Kaplan-Meier, mas permite a estimação dos efeitos das covariáveis sobre o tempo de sobrevivência (Carvalho et al., 2011). A verossimilhança parcial é obtida ao maximizar a probabilidade condicional de que um indivíduo sofra o evento, dado que ele está em risco no momento em que o evento ocorre, o que permite a estimação dos parâmetros $\boldsymbol{\beta}$ sem precisar conhecer a função de risco de base $\lambda_0(t)$ (Colosimo & Giolo, 2024). A função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [f(t_i | \mathbf{x}_i)]^{\delta_i} [S(t_i | \mathbf{x}_i)]^{1-\delta_i} = \prod_{i=1}^n [\lambda(t_i | \mathbf{x}_i)]^{\delta_i} [S(t_i | \mathbf{x}_i)], \quad (2.20)$$

onde δ_i indica se o evento foi observado (1) ou censurado (0). Para o contexto do modelo de Cox, a função de verossimilhança é definida por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [\lambda_0(t_i) \exp(\mathbf{x}_i' \boldsymbol{\beta})]^{\delta_i} [S_0(t_i)]^{\exp(\mathbf{x}_i' \boldsymbol{\beta})}. \quad (2.21)$$

De acordo com Rocha e Papoila (2009), em uma amostra de n indivíduos, com $k (\leq n)$ falhas observadas nos tempos $t_1, t_2, \dots, t_k$, a probabilidade condicional de um indivíduo falhar no tempo t_i , dado o seu historial, é expressa como:

$$P(\text{falhar em } t_i | \text{histórico até } t_i) = \frac{\lambda(t_i | \mathbf{x}_i)}{\sum_{j \in R(t_i)} \lambda(t_i | \mathbf{x}_j)}, \quad (2.22)$$

onde $R(t_i)$ é o conjunto de indivíduos em risco no tempo t_i . A componente não-paramétrica desaparece desta expressão. A função de verossimilhança parcial $L(\boldsymbol{\beta})$ é definida como o produto dos termos da equação 2.22 para todos os tempos de falha, sendo dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \left(\frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j' \boldsymbol{\beta})} \right)^{\delta_i}, \quad (2.23)$$

onde $\delta_i = 1$ se o evento foi observado e $\delta_i = 0$ se foi censurado.

Para obter os valores do vector de parâmetros $\boldsymbol{\beta}$ que maximizam a função de verossimilhança parcial, deve-se resolver o conjunto de equações escore definido por $U_q(\boldsymbol{\beta}) = 0$, para $q = 1, \dots, p$, em que $U_q(\boldsymbol{\beta})$ corresponde à derivada parcial de primeira ordem da log-verossimilhança parcial $\log L(\boldsymbol{\beta})$, este vector é dado por

$$U_q(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_q}.$$

Os autores, S.C.Oliveira e Almeida (2019), destacam que a verossimilhança parcial assume tempos contínuos de sobrevivência e não considera empates. Contudo, na prática estes empates podem surgir devido à precisão limitada da escala utilizada na medição dos eventos. Para lidar com esse cenário, foram propostas diferentes aproximações: o método de Breslow, para empates raros, o método de Efron, preferido quando há muitos empates, e a verossimilhança exacta, mais precisa mas computacionalmente dispendiosa, recomendada para amostras pequenas ou alta frequência de empates (S. C. Oliveira & Almeida, 2019).

2.7 Modelos probabilísticos de sobrevivência

2.7.1 Distribuição exponencial

Várias distribuições de sobrevivência estão disponíveis para modelar dados de sobrevivência. A distribuição exponencial tem um risco constante, $h(t) = \lambda$. A função de risco acumulado é derivada da seguinte forma:

$$\Lambda(t) = \int_0^t \lambda du = \lambda t \Big|_0^t = \lambda t. \quad (2.24)$$

A função de sobrevivência é

$$S(t) = e^{-\lambda t} \quad (2.25)$$

e a função de densidade de probabilidade é dada por

$$f(t) = \lambda e^{-\lambda t}. \quad (2.26)$$

2.7.2 Distribuição gama

De acordo com Sousa (2019), a distribuição gama não é tão utilizada como a distribuição Weibull, apesar de se ajustar adequadamente a vários dados de tempo de vida e incluir a distribuição exponencial como o seu caso particular ($k = 1$).

A função densidade de probabilidade é:

$$f(t; a, k) = \frac{1}{\Gamma(k)a^k} t^{k-1} \exp\left(-\frac{t}{a}\right), \quad t \geq 0,$$

em que $k > 0$ é o parâmetro de forma e $a > 0$ é o parâmetro de escala.

2.7.3 Distribuição log-normal

O modelo log-normal é adequado para situações em que o tempo de falha é bem descrito através de uma distribuição de probabilidades log-normal (S. J. Silva, 2022). O logaritmo de uma variável com distribuição log-normal, com parâmetros μ e σ , tem distribuição normal, com média μ e desvio-padrão σ . A função densidade de probabilidade para a variável aleatória tempo de falha T é:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right), \quad t > 0, \mu \in \mathbb{R}, \sigma > 0, \quad (2.27)$$

onde μ e σ representam a média e o desvio-padrão do logaritmo do tempo de falha, respectivamente. E a função de sobrevivência do tempo de falha é dada por:

$$S(t) = \Phi \left(\frac{-\log(t) + u}{\sigma} \right)$$

e a função taxa de falha expressa por:

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad t > 0. \quad (2.28)$$

onde Φ é a função de distribuição acumulada de uma distribuição normal padrão (Kleinbaum & Klein, 2012).

2.8 Estimação paramétrica de sobrevivência

2.8.1 Teste de Wald

A estatística de Wald testa a hipótese nula H_0 de que o parâmetro $\mathbf{B}_j$ da regressão é igual a zero (Carvalho et al., 2011). Este teste é talvez o teste mais comumente usado, e sua execução é directa a partir da saída da maioria dos *softwares* (Freitas, 2022). A hipótese nula $\beta = 0$ pode ser avaliada por meio da estatística

$$Z = \frac{\hat{\beta}}{\text{ep}(\hat{\beta})} \sim N(0, 1), \quad (2.29)$$

onde $\text{ep}(\hat{\beta})$ representa o erro padrão de $\hat{\beta}$. Sob H_0 , a estatística Z segue aproximadamente a distribuição normal padrão. De forma equivalente, pode-se considerar o quadrado dessa estatística:

$$W = Z^2, \quad (2.30)$$

a qual segue aproximadamente uma distribuição qui-quadrado com um grau de liberdade. Valores de W superiores ao valor crítico da distribuição qui-quadrado sugerem que a co-variável associada ao parâmetro β exerce influência significativa sobre a variável resposta.

2.8.2 Métodos de estimação

Método de máxima verossimilhança

Segundo Santos e Gomes (2023), ao construir uma função de verossimilhança tomando um produto de termos da distribuição exponencial, por exemplo, um para cada observação, se não houver censura, a função de verossimilhança assume a forma geral:

$$L(t_1, t_2, \dots, t_n) = \prod_{i=1}^n f(t_i; \lambda). \quad (2.31)$$

Se algumas observações forem censuradas, é necessário fazer um ajuste nesta expressão.

Para uma observação de uma morte, coloca-se a função de densidade de probabilidade como a equação 2.31. Mas para uma observação censurada à direita, coloca-se a função de sobrevivência, indicando que a observação é conhecida apenas por exceder um valor particular. A verossimilhança em geral assume a forma:

$$L(t_1, t_2, \dots, t_n) = \prod_{i=1}^n f(t_i; \lambda)^{\delta_i} S(t_i; \lambda)^{1-\delta_i} \quad (2.32)$$

Quando t_i é uma observação censurada, têm-se $\delta_i = 0$ e insere-se um factor de sobrevivência. Alternativamente, pode-se inserir um factor de risco para cada observação censurada e um factor de sobrevivência para cada observação, censurada ou não (Klein & Moeschberger, 2005).

2.9 Modelos de fragilidade

Os métodos tradicionais de análise de sobrevivência assumem que existe heterogeneidade não observada e que as observações são independentes (Minguillo, 2016). Esta limitação torna-se particularmente relevante em estudos epidemiológicos, onde os tempos de falha tendem a estar agrupados em unidades naturais, como famílias ou escolas. Nestes casos, algumas características não observadas, mas partilhadas pelos membros do mesmo grupo, podem influenciar o tempo até ao evento estudado (Gorfine et al., 2005).

Uma abordagem para lidar com essa dependência dentro dos grupos é o uso de modelos de fragilidade. Para Kleinbaum e Klein (2012), a fragilidade é uma variável aleatória que segue uma distribuição específica, comumente assumida como tendo média igual a 1. O parâmetro chave a ser estimado é a variância da distribuição da fragilidade, que quantifica o grau de heterogeneidade na população (Rocha & Papoila, 2009). Os modelos de fragilidade podem assumir diferentes formas, dependendo da maneira como a fragilidade é atribuída aos indivíduos. Segundo Kleinbaum e Klein (2012), os tipos de fragilidade podem ser:

- **Fragilidade não compartilhada ou univariada:** este modelo assume que cada indivíduo tem o seu próprio nível único de fragilidade, independente dos outros indivíduos na população. A fragilidade não compartilhada é apropriada quando não existe uma estrutura de grupo nos dados e assume-se a existência de variações biológicas não mensuráveis entre os indivíduos (Carvalho et al., 2011). Para considerar a heterogeneidade mencionada, um efeito aleatório, não negativo, é introduzido no modelo de Cox como factor multiplicativo, que fica então expresso por:

$$\lambda(t \mid x_i, z_i) = z_i \lambda_0(t) \exp(\beta' x_i), \quad i = 1, \dots, n, \quad (2.33)$$

onde, β é o vector de parâmetros associados às covariáveis x_i , $\lambda_0(t)$ é a função taxa de

falha de base e $z_1, \dots, z_n$ são os valores das fragilidades, que representam heterogeneidade não observada entre os indivíduos (Kleinbaum & Klein, 2012).

- **Fragilidade compartilhada ou multivariada:** neste modelo, assume-se que grupos de indivíduos, como membros da mesma família ou pacientes tratados na mesma clínica, compartilham o mesmo nível de fragilidade. A fragilidade representa, neste modelo, o risco comum compartilhado pelos indivíduos de um mesmo grupo. Assim como no contexto univariado, o modelo de fragilidade compartilhada é formulado pela introdução, no modelo de Cox, de um efeito aleatório para cada grupo. Formalmente, considere $T_j = \{T_{1j}, T_{2j}, \dots, T_{n_j}\}$ como o conjunto dos n_j tempos de sobrevivência dos indivíduos no j -ésimo grupo, e seja Z_j a variável que representa a fragilidade desse grupo, com $j = 1, \dots, m$.

Condicionais aos valores de Z_j , denotados por z_j , os componentes de T_j são assumidos independentes, e o modelo de fragilidade compartilhada é expresso por:

$$\lambda(t \mid \mathbf{x}_{ij}, z_j) = z_j \lambda_0(t) \exp(\beta' \mathbf{x}_{ij}), \quad i = 1, \dots, n_j, \quad j = 1, \dots, m, \quad (2.34)$$

onde β é o vector de parâmetros associados às covariáveis x_{ij} , $\lambda_0(t)$ é a função taxa de falha de base e z_j representa a fragilidade compartilhada dentro do grupo j (Colosimo & Giolo, 2024).

- **Fragilidade aninhada:** Os modelos de fragilidade aninhados consideram a estrutura hierárquica dos dados ao incluir dois efeitos aleatórios aninhados que actuam multiplicativamente sobre a função de risco. Estes modelos são apropriados quando as observações estão agrupadas em vários níveis hierárquicos, como em áreas geográficas Rondeau et al. (2012). Considerando T_{ijk} o tempo de sobrevivência em estudo para o k -ésimo indivíduo ($k = 1, \dots, K_{ij}$) do j -ésimo subgrupo ($j = 1, \dots, n_i$) do i -ésimo grupo ($i = 1, \dots, G$), C_{ijk} o tempo de censura à direita e L_{ijk} o tempo de truncamento à esquerda. As observações são dadas por $Y_{ijk} = \min(T_{ijk}, C_{ijk})$ e os indicadores de censura δ_{ijk} . Assumindo que T_{ijk} , L_{ijk} e C_{ijk} são independentes. Observam-se Y_{ijk} , L_{ijk} e δ_{ijk} . A função de risco para um modelo de fragilidade aninhado é dada por:

$$\lambda_{ijk}(t \mid v_i, z_{ij}) = v_i z_{ij} \lambda_0(t) \exp(\mathbf{x}_{ijk} \beta) = v_i z_{ij} \lambda_{ijk}(t), \quad (2.35)$$

onde $\lambda_0(t)$ é a função de risco base e X_{ijk} representa as covariáveis associadas ao indivíduo.

2.9.1 Teste de heterogeneidade

Conforme Balan e Putter (2019), o teste de razão de verossimilhança (LRT) é geralmente usado para testar a hipótese nula de ausência de fragilidade. Para os modelos de fragilidade gama e inversa gaussiana, isso é equivalente a testar $H_0 : \text{Var}[Z] = 0$ versus $H_a : \text{Var}[Z] > 0$.

2.9.2 Distribuições de fragilidade

A escolha da distribuição da fragilidade deve basear-se em fundamentos biologicamente plausíveis (Carvalho et al., 2011). Conforme os parâmetros definidos, as distribuições podem variar desde uma forma exponencial até uma curva em forma de sino que lembra uma distribuição normal (Bedia, 2022).

Distribuição de fragilidade gama

O modelo de fragilidade gama assume uma distribuição gama para as fragilidades. Este é o modelo de fragilidade mais popular devido à sua tratabilidade matemática (Munda et al., 2017). Suponha-se então que a fragilidade Z segue uma distribuição gama com função densidade de probabilidade:

$$f(z) = \frac{\theta^\delta z^{\delta-1} \exp(-\theta z)}{\Gamma(\delta)}, \quad z > 0, \quad (2.36)$$

onde $\theta > 0$ e $\delta > 0$. Sendo que a distribuição gama tem parâmetros de escala e forma α e γ tais que

$$\mathbb{E}[Z] = \gamma\alpha, \quad \text{Var}(Z) = \gamma\alpha^2.$$

A função de sobrevivência e a função de risco populacionais são dadas, respectivamente, por:

$$S(t) = \left(\frac{\theta}{\theta + \Lambda(t)} \right)^\delta \quad (2.37)$$

e

$$h(t) = \frac{\delta \lambda(t)}{\theta + \Lambda(t)}. \quad (2.38)$$

Distribuição de fragilidade lognormal

Os autores Rocha e Papoila (2009), propõem a utilização da distribuição lognormal como alternativa à distribuição gama. Neste modelo, assume-se que a população sobrevivente tende a torna-se mais homogênea ao longo do tempo, o que é coerente com presença de um mecanismo de selecção natural, em que os indivíduos mais frágeis são gradualmente eliminados. Admitindo que Z segue uma distribuição lognormal com parâmetros u e σ^2 , que

se relacionam com o valor esperado e variância, como segue:

$$\mathbb{E}[Z] = \exp\left(u + \frac{\sigma^2}{2}\right), \quad \text{Var}(Z) = \exp(\sigma^2) - 1.$$

2.9.3 Estimação dos parâmetros do modelo de fragilidade

A estimação de parâmetros em modelos de fragilidade pode ser realizada pelo algoritmo EM (Esperança-Maximização), pela verossimilhança penalizada ou pela verossimilhança restrita. O uso de cada abordagem depende do tipo de fragilidade considerada, da complexidade do modelo e da disponibilidade computacional (Carvalho et al., 2011).

Inferência baseada na Esperança-Maximização

Os procedimentos de estimação baseiam-se, em geral, na construção de uma função de verossimilhança e na sua maximização. Assim, se for assumida uma forma paramétrica para a função de risco de base $\lambda_0(t)$ e uma distribuição Gama para as variáveis de fragilidade Z_j , obtém-se uma expressão analítica para a função de verossimilhança. Nessa abordagem, a fragilidade Z_j é modelada como uma variável aleatória seguindo uma distribuição Gama com parâmetro de forma e escala iguais a θ , que quantifica o grau de heterogeneidade entre os grupos. A função de verossimilhança marginal é então obtida pela integração da fragilidade fora da expressão da verossimilhança condicional. Esse procedimento permite que as estimativas dos parâmetros do modelo, incluindo os coeficientes β associados às covariáveis e o parâmetro θ que mede a variabilidade entre os grupos, sejam obtidas directamente pela maximização da função de verossimilhança (Colosimo & Giolo, 2024).

No caso do modelo semiparamétrico de fragilidade gama, uma proposta para estimar os parâmetros baseia-se na utilização do algoritmo Esperança-Maximização (EM). A ideia por detrás do método é tratar as fragilidades como dados não observáveis, que são então estimados no passo E do algoritmo e no passo M são obtidos os valores dos coeficientes de regressão (Carvalho et al., 2011). No entanto, o algoritmo é lento e pode apresentar problemas de convergência em algumas situações (Colosimo & Giolo, 2024).

Inferência baseada na verossimilhança penalizada

A função de verossimilhança penalizada é composta por dois elementos principais: o primeiro corresponde à função de verossimilhança parcial que incorpora as fragilidades como parâmetros, enquanto o segundo representa um termo de penalização, cuja função é evitar diferenças grandes entre as fragilidades para os diferentes grupos (Colosimo & Giolo, 2024). Assim, a forma logaritmo da função de verossimilhança parcial penalizada pode ser descrita como:

$$PPL(\beta, w, \theta) = \log[L(\beta, w)] - g(w, \theta), \quad (2.39)$$

onde a função de verossimilhança logarítmica é dada por

$$\log[L(\beta, w)] = \sum_{j=1}^m \sum_{i=1}^{n_j} [\delta_{ij} \log \lambda(t_{ij} | x_{ij}, z_j) - \Lambda(t_{ij} | x_{ij}, z_j)], \quad (2.40)$$

sendo δ_{ij} o indicador de falha (1 se o evento foi observado, 0 caso contrário), $\lambda(t_{ij} | x_{ij}, z_j)$ a função de risco condicional e $\Lambda(t_{ij} | x_{ij}, z_j)$ a função de risco acumulada. O termo $g(w, \theta)$ representa a penalização que regula a complexidade do modelo (Colosimo & Giolo, 2024). A estimação do parâmetro θ pode ser feita minimizando o Critério de Informação de Akaike (AIC) (Carvalho et al., 2011).

A função de log-verossimilhança penalizada assume diferentes expressões consoante o modelo considerado. Para os modelos de Cox, de fragilidade partilhada, aninhada e aditiva, é definida por:

$$PPL(\theta) = \text{Log}(\theta) - \alpha \int_0^\infty \lambda_0(t)^2 dt, \quad (2.41)$$

em que $\alpha > 0$ é um parâmetro de suavização que controla o equilíbrio entre o ajustamento aos dados e a suavidade das funções Rondeau et al. (2012). O parâmetro pode ser fixado pelo utilizador ou estimado automaticamente. Valores baixos de α permitem maior flexibilidade no ajuste, enquanto valores elevados impõem maior suavidade às funções estimadas, podendo resultar em sub-ajuste. Esta abordagem funciona bem para os modelos de fragilidade gama e lognormal (Colosimo & Giolo, 2024).

Outras alternativas de estimação

É possível usar outros algoritmos de maximização para estimar a variância da fragilidade, como a máxima verossimilhança restrita (REML), que ajusta a verossimilhança considerando apenas componentes aleatórios. O algoritmo que minimiza o AIC, ao penalizar a complexidade, favorece modelos mais simples, mesmo que, por vezes, escolha modelos com um ajuste ligeiramente inferior em comparação à máxima verossimilhança (Carvalho et al., 2011).

2.9.4 Adequação do modelo de fragilidade

Actualmente, não há um método definitivo para determinar se a distribuição gama é superior à lognormal para ajustar o efeito aleatório, nem para identificar qual abordagem de estimação da variância da fragilidade é mais adequada (Carvalho et al., 2011). De acordo com Colosimo e Giolo (2024), a escolha de uma distribuição para a fragilidade é, na maioria das vezes, baseada na tratabilidade matemática ou disponibilidade do *software*, e não na forma como ela se ajusta aos dados. Entretanto, é importante testar a adequação da

distribuição com a finalidade de seleccionar a mais adequada, através do teste de razão de verossimilhança para a escolha da distribuição mais adequada. Os critérios de informação de Akaike (AIC) e o critério de informação bayesiano (BIC) também podem auxiliar no processo de escolha da distribuição mais apropriada para as variáveis de fragilidade.

Ainda segundo Colosimo e Giolo (2024), para avaliar se o modelo mantém a proporcionalidade (condicional) das taxas de falha, os métodos gráficos têm sido aqueles que envolvem as curvas das taxas de falha acumulada e os resíduos de Schoenfeld, descritos na Secção 2.10.

2.9.5 Considerações sobre a estimação de modelos de fragilidade

De acordo com Kleinbaum e Klein (2012), alguns aspectos devem ser levados em consideração na escolha de modelos de fragilidade:

- **Estrutura dos Dados:** a presença de grupos ou *clusters* nos dados sugere a utilização de um modelo de fragilidade compartilhada. Se não houver agrupamento, um modelo de fragilidade não compartilhada pode ser mais apropriado.
- **Distribuição da Fragilidade:** a distribuição da fragilidade, como gama ou lognormal, precisa ser especificada. A escolha da distribuição depende da plausibilidade biológica e das propriedades estatísticas de cada distribuição, a escolha entre gama e lognormal pode não ser crucial em muitos casos, desde que o objectivo seja corrigir a heterogeneidade (Kleinbaum & Klein, 2012).
- **Interpretação dos Resultados:** a inclusão da fragilidade no modelo afecta a interpretação dos coeficientes de regressão. As estimativas dos coeficientes representam os efeitos dos preditores na sobrevivência condicional ao nível de fragilidade. É fundamental ter em conta esta natureza condicional ao interpretar os resultados.
- **Software Estatístico:** nem todos os softwares estatísticos suportam a mesma variedade de modelos de fragilidade. É importante verificar as capacidades do *software* que se pretende usar antes de seleccionar um modelo.

2.10 Análise de Resíduos

A análise dos resíduos constitui uma etapa fundamental para verificar a adequação dos pressupostos do modelo. No contexto do modelo de Cox, os resíduos podem ser utilizados para examinar a suposição de riscos proporcionais, verificar a relação log-linear entre a variável resposta e as covariáveis, bem como identificar possíveis observações atípicas ou influentes. Contudo, a presença de dados censurados, aliada à natureza específica do modelo de Cox, torna a definição e interpretação dos resíduos mais complexa e menos imediata do

que nos modelos clássicos de regressão linear, dificultando a aplicação directa de conceitos análogos (Rocha & Papoila, 2009).

2.10.1 Resíduos de Schoenfeld

Segundo D. Silva (2016), os resíduos de Schoenfeld (1982) permitem averiguar a proporcionalidade de cada covariável z_i , definidos para cada i -ésimo indivíduo como:

$$r_{ji} = \delta_i(z_{ji} - a_{ji}), \quad (2.42)$$

em que,

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é uma observação não censurada;} \\ 0, & \text{se } t_i \text{ caso contrário} \end{cases}$$

e

$$a_{ji} = \frac{\sum_{l \in R_i} z_{jl} \exp(\hat{\beta}' z_l)}{\sum_{l \in R_i} \exp(\hat{\beta}' z_l)}. \quad (2.43)$$

O vector de resíduos é dado por $r_{ji} = (r_{j1}, r_{j2}, \dots, r_{jp})$, onde $i = 1, \dots, p$ representa os tempos em que os eventos foram observados. O conjunto R_i corresponde aos indivíduos em risco no instante t , enquanto z_{jl} designa o valor da covariável l para o indivíduo j .

A análise gráfica dos resíduos de Schoenfeld ao longo do tempo de vida permite verificar se a sua distribuição é homogénea, resultando numa nuvem de pontos aleatórios alinhados horizontalmente e centrados em zero. A ausência de padrões sistemáticos no gráfico de r_i padronizado em relação ao tempo de sobrevivência sugere que a suposição de riscos proporcionais é válida, indicando a adequação do modelo aos dados (Carvalho et al., 2011). Para avaliar possíveis desvios dessa suposição, adiciona-se uma curva suavizada acompanhada de bandas de confiança, o que possibilita a detecção de tendências ou falhas na proporcionalidade dos riscos (Colosimo & Giolo, 2024).

2.10.2 Resíduos de martingala

Os resíduos de martingala são úteis para avaliar se a especificação do modelo de Cox está adequada, isto é, se as covariáveis incluídas são apropriadas. Conforme explicado por Andrade e Cabete (2012), considera-se o acompanhamento de um indivíduo com vector de covariáveis $\mathbf{z}$ sendo $N(t)$ uma função que toma o valor zero até imediatamente antes do instante em que o acontecimento ocorre e toma o valor um a partir desse momento. Esta função caracteriza

um processo de contagem que indica a ocorrência ou não do acontecimento.

Para cada indivíduo j com tempo de vida t_j , o resíduo de Martingala corresponde a diferença entre o número de eventos ocorridos durante o tempo de observação e o número de eventos esperados sob o modelo de Cox (Andrade & Cabete, 2012). Estes resíduos são dados por:

$$M_j = N_j - \Lambda(t_j, z_j, \beta) = N_j - E_j. \quad (2.44)$$

Como o verdadeiro valor de β é desconhecido, então M_j é estimado por $\hat{M}_j$ em que se substituiu β por $\hat{\beta}$. Quando $\hat{M}_j < 0$, isso indica que o número de eventos observados é menor do que o estimado pelo modelo e a sobrevivência está sobrestimada (Colosimo & Giolo, 2024).

A análise dos resíduos de Martingala pode ser realizada através de dois tipo de representações gráficas: um que representa os valores $\hat{M}_j$ no eixo das ordenadas e os valores estimados $\beta'z_j$ no eixo das abcissas, destacando indivíduos mal ajustados, geralmente aqueles com tempos de vida muito longos ou curtos; e outro que exibe $\hat{M}_j$ estimados a partir de um modelo sem covariáveis ($\beta = 0$) em relação a uma covariável contínua, incluindo uma curva de suavização para sugerir a sua forma funcional (Rocha & Papoila, 2009).

2.10.3 Resíduos *Deviance*

Além da suposição de proporcionalidade das taxas de falha, há interesse em examinar outros aspectos do ajuste do modelo de Cox. Entre eles, verificar a presença de observações atípicas (*outliers*). Os resíduos *Deviance*, propostos por Therneau e Grambsch (2000), correspondem a uma transformada normalizada dos resíduos de Martingala, sendo simetricamente distribuídos em torno de zero, o que facilita a sua interpretação. Para cada indivíduo i o resíduo *Deviance* é dado por:

$$D_i = \text{sinal}(\hat{M}_i) \sqrt{-2 \times (l_{i(\text{modelo})} - l_{i(\text{saturado})})}, \quad (2.45)$$

onde $\text{sinal}(\hat{M}_i)$ é o sinal do resíduo martingala; $l_{i(\text{modelo})}$ e $l_{i(\text{saturado})}$ são as log-verossimilhanças parciais do modelo considerado e o modelo com ajuste perfeito, neste caso, o modelo saturado, respectivamente. Em contextos com baixa percentagem de censura, os resíduos D_i têm uma distribuição aproximadamente normal, o que faz com que, nesta situação, sejam mais úteis na detecção de *outliers* do que os resíduos de martingala (D. Silva, 2016).

2.10.4 Resíduos *Score*

Os resíduos *Score* possibilitam avaliar a diferença ocorrida na estimativa de β quando um determinado indivíduo é eliminado da análise (A. S. Silva et al., 2004). De forma semelhante aos resíduos de Schoenfeld, os resíduos *Score* são obtidos a partir da derivada parcial da

função log-verossimilhança em ordem a $\beta_k, k = 1, \dots, p$. Para uma amostra de dimensão n com m tempos de vida tem-se:

$$\frac{\partial(L\beta)}{\partial\beta_k} = \sum_{j=1}^n \left[\delta_j(z_{jk} - a_{jk}) + e^{\beta'z_j} \sum_{t_r \leq t_j} \frac{(a_{rk} - z_{jk})\delta_r}{\sum_{l \in R_r} e^{\beta'z_l}} \right], \text{ com } a_{jk} = \frac{\sum_l z_{lk} e^{\beta'z_l}}{\sum_l e^{\beta'z_l}}, \quad (2.46)$$

nessa formulação δ_j representa o estado do indivíduo j , em que o indivíduo j apenas contribui para a derivada da expressão (2.46) até ao instante t_j . Isto implica que, se a observação dos indivíduos terminasse neste instante, então a j -ésima componente da derivada não seria afectada. Assim, o resíduo *Score* para o indivíduo $j, j = 1, \dots, n$ e para a covariável $z_k, k = 1, \dots, p$ é:

$$\hat{r}s_{jk} = \delta_j(z_{jk} - a_{jk}) + e^{\beta'z_j} \sum_{t_r \leq t_j} \frac{(a_{rk} - z_{jk})\delta_r}{\sum_{l \in R_r} e^{\beta'z_l}}. \quad (2.47)$$

Assim como ocorre com os resíduos de Schoenfeld, a soma dos resíduos *Score* também é igual a zero; contudo, podem apresentar valores não nulos para indivíduos cujas observações foram censuradas. Para cada covariável z_k , os valores de $\hat{r}s_{jk}$ são calculados e representados num gráfico, onde o eixo das ordenadas corresponde aos resíduos e o eixo das abcissas aos valores de z_k . Observações com valores extremos tendem a ter um impacto significativo na estimativa de β_k . Para facilitar a interpretação, é comum apresentar os resíduos $\hat{r}s_{jk}$ ponderados pelo erro padrão estimado de $\hat{\beta}_k$ nos gráficos (Andrade & Cabete, 2012).

2.11 Interpretação dos coeficientes

Habitualmente, a interpretação não é realizada directamente em termos de β_j , mas sim através de $\exp(\beta_j)$, uma vez que esta possui um significado mais intuitivo no contexto do risco de morte (Aquino, 2024). Em que, $\exp(\beta_j)$ representa o risco relativo de ocorrência do evento para dois indivíduos que diferem em uma unidade nos valores da covariável z_j , sendo iguais os respectivos valores das demais covariáveis (Rocha & Papoila, 2009).

A avaliação do *Hazard Ratio* (HR) em modelos de sobrevivência pode ser realizada segundo A. S. Silva et al. (2004), da seguinte forma:

- Quando $HR < 1$, a covariável actua como um factor protector, reduzindo a probabilidade do desfecho.
- Para $HR = 1$, não há evidência de influência da covariável no risco de ocorrência do desfecho.

- Já para $HR > 1$, a covariável está associada a um aumento no risco do desfecho estudado.

2.12 Análise de amostras complexas

As ferramentas de análise convencionais disponíveis na maioria dos *softwares* estatísticos assumem, em geral, condições simplificadas que pressupõem a utilização de amostras aleatórias simples com reposição. No entanto, tais pressupostos mostram-se inadequados quando se trata de dados provenientes de amostras complexas de populações finitas, uma vez que ignoram elementos essenciais dos planos amostrais. Entre esses elementos incluem-se as diferentes probabilidades de seleção das unidades, a estrutura de conglomerados, a estratificação, bem como procedimentos de calibração e imputação aplicados em situações de não resposta, entre outros ajustamentos relevante (Pessoa & Silva, 2018).

A ideia estatística fundamental por trás de toda inferência baseada em desenho amostral é que um indivíduo amostrado com uma probabilidade de amostragem de n_i representa $\frac{1}{n_i}$ indivíduos na população. O valor $\frac{1}{n_i}$ é chamado de peso amostral (Lumley, 2010). Ao ignorar o plano amostral nas análises pode-se gerar estimativas enviesadas, erros padrão incorretos e, conseqüentemente, a conclusões estatísticas inadequadas, comprometendo a validade dos resultados e a generalização dos achados para a população-alvo (Pessoa & Silva, 2018).

Capítulo 3

MATERIAL E MÉTODOS

3.1 Material

Para o presente estudo, foram usados dados do IDS de 13 dos 16 países da África Austral. O IDS compreende uma amostra complexa, em que, primeiro, selecciona-se uma amostra estratificada de áreas de enumeração (EAs) com probabilidade proporcional ao tamanho. Em seguida, dentro dessas EAs seleccionadas, realiza-se uma listagem de todos os domicílios, depois selecciona-se um número fixo de domicílios, onde são colectadas informações de mulheres, homens e crianças por meio de questionários individuais, com a escolha das casas sendo feita por amostragem sistemática (DHS, 2012).

As bases de dados foram obtidas no *site* <https://www.dhsprogram.com/>, fornecidas após uma solicitação formal aprovada conforme a carta de autorização em anexo. Para África do Sul foi o IDS de 2016, em Angola o IDS de 2015-16, no Lesoto o IDS de 2014, em Madagáscar o IDS de 2021, em Malawi o IDS de 2015-16, em Moçambique o IDS de 2022-23, na Namíbia o IDS de 2013, em Eswatini o IDS de 2006-07, na Zâmbia o IDS de 2018, no Zimbabwe o IDS de 2015, na República Democrática do Congo o IDS de 2011-12, na Tanzânia o IDS de 2022, por fim, nas Comores o IDS de 2012. No total, a amostra bruta combinada para este estudo compreende 116679 observações de crianças abaixo de cinco anos. Para uma descrição detalhada do desenho amostral e dos procedimentos metodológicos utilizados, remete-se à referência DHS (2012).

Descrição da variável dependente

Neste estudo, a variável dependente foi o tempo até a ocorrência de óbito em crianças menores de cinco anos. Onde a morte ocorrida entre o nascimento e os 59 meses é o evento de interesse e crianças que sobreviveram aos 59 meses foram censuradas. Neste estudo, o óbito em crianças menores de cinco anos foi incluído na análise independentemente de qualquer

causa.

Descrição das covariáveis

Foram consideradas as variáveis apresentadas na Tabela 3.1 como potenciais variáveis independentes, estas são variáveis identificadas com base na literatura existente e são relativas a dados sociodemográficos, dados biomédicos da criança e características relacionadas à mãe e ao ambiente (Aly & Grabowski, 1990; Bonita & Beaglehole, 2010; Boettiger et al., 2021; Ettarh, 2012). As covariáveis quantitativas medidas numa escala no mínimo discreta, como a idade da mãe, peso da criança, duração da amamentação, foram categorizadas, pois segundo Eugênio (2017), o processo de categorização de variáveis quantitativas para inclusão em modelos de regressão em análise de sobrevivência, facilita a identificação de grupos de risco e a interpretação da relação da variável com a variável resposta. E que sobre o ponto de vista clínico e estatístico, as variáveis categóricas são preferidas. No caso da dicotomia, por exemplo, fornecem uma classificação simples em grupos de alto e baixo risco. Em geral, estabelecem um critério padrão para comparação com estudos anteriores e subsequentes.

Tabela 3.1: *Descrição das variáveis de estudo*

Descrição da Variável	Categorias
Estado de vida	1- Censura, 2- Evento
País	1- Moçambique; 2- Tanzânia; 3- África do Sul; 4- Zâmbia; 5- Zimbabwe; 6- Angola; 7- Congo; 8- Comores; 9- Lesoto; 10- Madagáscar; 11- Malawi; 12- Namíbia; 13- Eswatini
Local de residência	1- Rural; 2- Urbano
Sexo da criança	1- Masculino; 2- Feminino
Índice de riqueza	1- Baixo; 2- Médio; 3- Alto
Nível de escolaridade da mãe	1- Sem escolaridade; 2- Primário; 3- Secundário ou mais
Nível de escolaridade do marido	1- Sem escolaridade; 2- Primário; 3- Secundário ou mais
Idade da mãe	1- <18; 2- 18-29; 3- 30-39; 4- 40+
Sexo do chefe do agregado familiar	1- Masculino; 2- Feminino
Número de membros no agregado familiar	1- ≤5; 2- >5
Estado marital	1- Nunca se casou; 2- Casada/ Vive maritalmente; 3- Viúva/Divorciada/Separada
Ocupação da mãe	1- Desempregada; 2- Empregada
Peso à nascença	1- <2500g; 2- 2500-4500g; 3- >4500g
Duração da amamentação	1- Nunca amamentada; 2- Actualmente amamentada; 3- <6 meses; 4- 6-12 meses; 5- 12+ meses
Número de consultas pré-natais	1- Nenhuma; 2- 1 a 3; 3- 4 ou +
Ordem de nascimento	1- Primogénito; 2- 2-3; 3- 4+
Tipo de nascimento	1- Singular; 2- Gemelar
Intervalo entre nascimentos	1- <12 meses; 2- 12-48 meses; 3- 49+ meses
Gravidez indesejada	1- Não; 2- Sim
Tipo de parto	1- Cesariano; 2- Normal
Local de parto	1- Casa; 2- Hospital Público; 3- Hospital Privado; 4- Outro local
Fonte de água	1- Não melhorada; 2- Melhorada
Sanitário	1- Não melhorado; 2- Melhorado
Já vacinou	1- Não; 2- Sim
Incidência de diarreia	1- Não; 2- Sim
Incidência de infeccção respiratória	1- Sim; 2- Não
Uso de rede mosquiteira	1- Não; 2- Sim
Exposição à mídia	1- Sem exposição; 2- Uma vez por semana; 3- Pelo menos uma vez por semana 4- Todos os Dias

3.2 Métodos

3.2.1 Análise exploratória

Para compreensão da distribuição das variáveis, foram calculadas as frequências absolutas e estimadas as percentuais de cada variável em geral e entre eventos e censuras. Além disso, algumas variáveis foram apresentadas em forma de gráficos de modo a ilustrar a sua distribuição entre os grupos.

3.2.2 Teste de Qui-quadrado

Este teste foi usado para verificar se existe associação entre variáveis categóricas (G. Silva, 1992). Para tal, foram testadas as seguintes hipóteses:

$$\begin{cases} H_0 : & \text{Não há associação entre as variáveis} \\ H_1 : & \text{Há associação entre as variáveis} \end{cases}$$

Sendo a estatística do teste dada pela expressão:

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

Onde:

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

Regra de decisão:

- Se $\chi^2 \leq \chi_{\alpha, k-1}^2$, não rejeita-se a hipótese nula H_0 .
- Se $\chi^2 > \chi_{\alpha, k-1}^2$, rejeita-se a hipótese nula H_0 .
- Caso o p-valor associado à estatística do teste seja menor que o nível de significância fixado, rejeita-se a hipótese nula H_0 .

Em que, $\chi_{\alpha, k-1}^2$ é o valor crítico da distribuição qui-quadrado para um nível de significância α e $k - 1$ graus de liberdade. Rejeita-se a hipótese nula definida, caso o p-valor associado à estatística do teste seja menor que o nível de significância fixado (G. Silva, 1992). Entretanto, o teste de Qui-quadrado convencional não considera amostras complexas, mas

o ajustamento de Rao-Scott corrige a estatística para levar em conta a correlação dos dados, tornando o teste mais conservador e reduzindo a probabilidade de rejeição indevida da hipótese nula, sendo assim, foi aplicado o teste de Qui-quadrado com ajustamento de Rao-Scott, adequado para amostras complexas (Lee & Forthofer, 2006; Rao & Scott, 1987). A análise foi realizada em todas as variáveis da Tabela 3.1, excepto incidência de diarreia, incidência de infecção respiratória e vacinação. Estas foram excluídas porque possuem observações apenas para crianças censuradas.

3.2.3 Análise de sobrevivência

O tempo até a ocorrência do evento foi avaliado por meio das curvas de Kaplan-Meier, pois este permite a estimação da probabilidade de sobrevivência ao longo do tempo. Para comparar as curvas de sobrevivência entre os grupos, aplicou-se o teste de log-rank a um nível de significância de 5%. As hipóteses de igualdade entre as curvas descrevem-se por:

$$\begin{cases} H_0 : S_1(t) = S_2(t) \\ H_1 : S_1(t) \neq S_2(t). \end{cases}$$

Rejeita-se H_0 se o valor da estatística for maior que o valor tabelado da distribuição χ^2 ou se o p-valor associado for menor que $\alpha = 5\%$.

Modelo de Cox com fragilidade

Para estudar a relação entre o tempo até a morte e as covariáveis, considerou-se o modelo de Cox com riscos proporcionais. E uma vez que os dados do IDS têm uma estrutura hierárquica, o que viola a suposição de independência no modelo de Cox tradicional, incorporou-se o efeito de fragilidade de forma a considerar a variabilidade advinda dos *clusters* que são as áreas de enumeração. O modelo estimado é dado da seguinte forma:

$$\lambda(t \mid x_{ij}, z_j) = z_j \lambda_0(t) \exp(\beta' x_{ij}), \quad i = 1, \dots, n_j, \quad j = 1, \dots, m, \quad (3.2)$$

onde β é o vector de parâmetros associados às covariáveis x_{ij} , $\lambda_0(t)$ é a função taxa de falha de base e z_j representa a fragilidade compartilhada dentro dos *clusters*. A escolha da distribuição de fragilidade, bem como do método de estimação, foi feita por meio da comparação do AIC.

A selecção das covariáveis seguiu o procedimento proposto por Collet (2003), conforme descrito por J. Rocha (2015), e ocorreu em quatro passos:

- Passo 1: foram ajustados todos os modelos contendo uma única covariável, e seleccionadas para o próximo passo as covariáveis significativas ao nível $\alpha = 0,10$, com base na aplicação do teste de razão de verossimilhança.

- Passo 2: As covariáveis seleccionadas no passo 1, foram incluídas conjuntamente no modelo. Na presença de certas covariáveis, outras podem deixar de ser significativas e, assim, foram ajustados modelos reduzidos, excluindo-se uma covariável de cada vez. Neste passo, permanecem no modelo somente as covariáveis que produzem um aumento significativo na estatística da razão de verossimilhanças.
- Passo 3: Neste passo, foi ajustado um novo modelo com as variáveis retidas no passo 2. A seguir, as covariáveis excluídas no passo 1 retornaram ao modelo (uma de cada vez) para confirmar que elas não são estatisticamente significativas.
- Passo 4 : No último passo, as covariáveis que não foram significativas no passo 2 retornaram ao modelo para confirmar se definitivamente as mesmas devem ser excluídas do modelo final. Além disso a inclusão de outras variáveis foram testadas com o objectivo de verificar algum fundamento teórico. Ainda nesta fase, uma interação foi testada: a interação entre o local de residência e local de parto, baseado nos resultados obtidos no estudo de Ayele et al. (2017), descritos na secção [2.2.1](#).

Teste de razão de verossimilhança

O teste da razão de verossimilhanças foi usado para a comparação de modelos no processo de selecção das variáveis. A hipótese nula testada foi:

$$\left\{ H_0 : \beta_{k+1} = \beta_{k+2} = \dots = \beta_p = 0, \right.$$

A estatística da razão de verossimilhanças, foi obtida pela diferença entre os logaritmos das funções de verossimilhança dos dois modelos (Colosimo & Giolo, 2024). Este teste foi aplicado ao nível de significância $\alpha = 5\%$.

Análise de resíduos

A análise dos resíduos foi feita de modo a identificar alguns aspectos do modelo de regressão. A suposição de proporcionalidade de riscos foi avaliada usando o teste residual global de Schoenfeld, os resíduos de martingala foram usados para investigar a forma funcional de uma covariável, bem como para identificar *outliers*, os resíduos *Deviance* foram usados para identificar os valores *outliers*.

Interpretação dos coeficientes

A significância dos coeficientes estimados do modelo de regressão de Cox multinível foi medida ao nível de 5% e os *Hazard Ratios* foram usados para interpretação dos coeficientes estimados.

Ajuste para amostra complexa

Para ajustar a complexidade do plano amostral foram incluídas ponderações, de modo a garantir que as análises reflectam a representatividade desejada no processo de inferência como argumentado por Bianchini e Albieri (1999) e descrito na Secção 2.12. Desta forma foi utilizado o peso amostral após uma desnormalização, que é um processo recomendado em casos de combinação de IDS de diferentes anos ou países (Croft et al., 2023; Ren, 2022). Também foi usada a unidade primária de amostragem e os estratos para construção do objecto que representa o plano amostral (R Core Team, 2019). Esta componente do plano amostral foi aplicada para a questão das análises exploratórias e teste de qui-quadrado com vista a avaliar a relação entre o evento morte e as variáveis explicativas.

Para o processamento dos dados, utilizou-se o *software* R, versão 4.4.1, com o suporte dos pacotes *survival* para análise de sobrevivência, e *survey* para ponderação (R Core Team, 2019; Therneau, 2023). As análises foram conduzidas a um nível de significância de 5%, utilizando o *p*-valor associado à estatística dos testes para avaliação da regra de decisão. O código completo utilizado nesta monografia está disponível no repositório em:

<https://github.com/ShelseaDamiao/monografia--mortalidade-scripts>

Capítulo 4

RESULTADOS E DISCUSSÃO

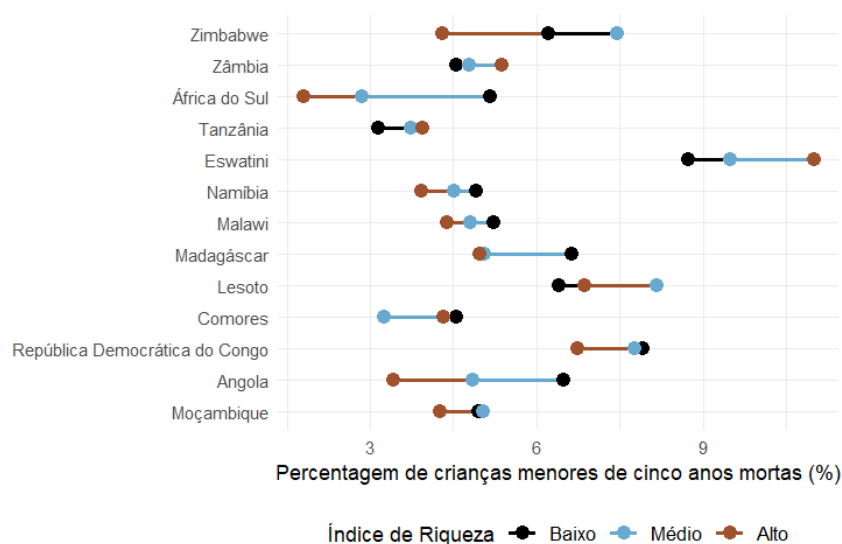
4.1 Análise descritiva

A Tabela 4.1 evidencia uma associação significativa entre o País e o Estado de vida, ressaltando que Eswatini e a República Democrática do Congo apresentam as mais elevadas taxas de mortalidade infanto-juvenil. Observa-se também que a mortalidade é maior em crianças do sexo masculino (5,8%) comparado ao feminino (4,8%). A maioria das mães era jovem, com 56,8% entre 18 e 29 anos, e 33% entre 30 e 39 anos. No caso do estudo, a idade materna está significativamente associada à condição de vida das crianças, reflectindo diferenças nas taxas de mortalidade infanto-juvenil conforme a idade das mães. Apesar de 79% das mães serem casadas ou viverem maritalmente, essa característica não mostrou associação significativa com o Estado de vida.

Foi observada uma diferença significativa na ocupação das mães: aquelas que estavam empregadas apresentaram 5,8% de eventos, enquanto as desempregadas registaram 4,9%. Ademais, a análise das condições ambientais e socioeconómicas destacou associações relevantes. Crianças residentes em áreas rurais exibiram eventos em 5,6% dos casos, em comparação com 4,8% naquelas oriundas de áreas urbanas. Entre as famílias de baixa renda, a percentagem de eventos foi de 5,7%, enquanto para renda média foi de 5,3% e para alta renda, de 4,7%. De modo semelhante, constatou-se que a utilização de fontes de água não melhoradas estava significativamente associada a uma maior mortalidade, com percentagem de 6,2% em comparação a 4,7% nas crianças que tinham acesso a fontes consideradas seguras. O mesmo se aplica ao saneamento, onde as condições não melhoradas estavam associadas a 6% de eventos, em contraste com 4,7% para as condições adequadas. Adicionalmente, a Exposição à mídia também apresentou diferenças significativas: 6% das crianças que não tinham exposição alguma à mídia registaram o evento, enquanto apenas 3% das crianças que tiveram contacto com mídias pelo menos uma vez por semana foram afetadas.

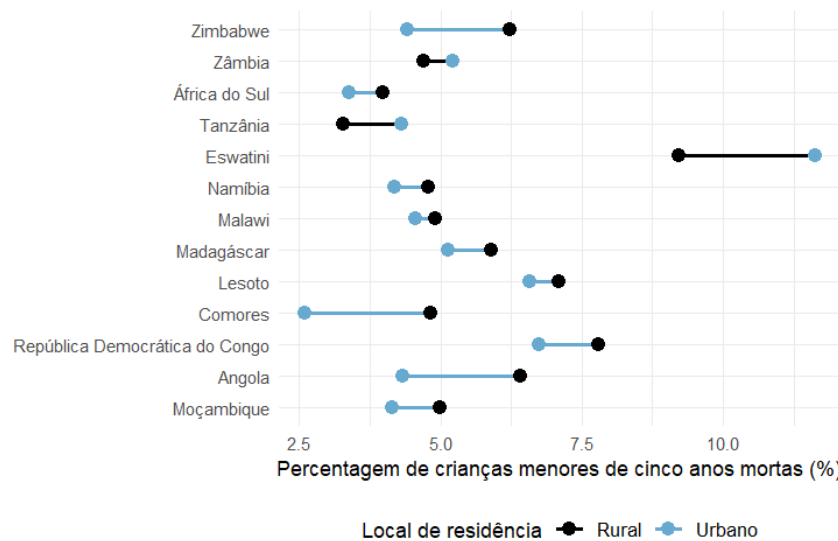
Quanto aos factores perinatais, o peso ao nascer foi um indicador chave, com 7,2% de mortalidade para crianças abaixo de 2500g em comparação a 3,6% para aquelas entre 2500 e 4500g e 8,5% para as acima de 4500g. A duração da amamentação teve forte impacto: 31,6% das crianças não amamentadas morreram, comparado a 6,5% entre as amamentadas por mais de 12 meses. O tipo de nascimento (singular vs. gemelar) também mostrou diferenças, com 18,2% de mortalidade entre gêmeos, superior aos 4,9% em nascimentos únicos. Por fim, factores como nível de escolaridade do marido, sexo do chefe do agregado familiar, estado marital da mãe e uso de rede mosquiteira não mostraram associações significativas.

Figura 4.1: *Equiplot comparativo da percentagem de crianças menores de cinco anos mortas por índice de riqueza em diversos países.*



A análise da Figura 4.1 indica que a mortalidade infantil em crianças abaixo de cinco anos é geralmente mais alta em contextos de pobreza, reduzindo-se nos níveis médio e alto. Eswatini, República Democrática do Congo revelam a maior disparidade, reflectindo desigualdades socioeconómicas. Por outro lado, Malawi e Namíbia apresentam diferenças menos evidentes, indicando menor desigualdade.

Figura 4.2: *Equiplot comparativo da percentagem de crianças menores de cinco anos mortas por Local de residência.*



A Figura 4.2 mostra que a mortalidade infantil em crianças abaixo de cinco anos é geralmente mais alta nas zonas rurais. Contudo, em Eswatini, Zâmbia e Tanzânia, a percentagem de mortes de crianças menores de cinco anos é maior nas áreas urbanas. .

Tabela 4.1: *Distribuição das variáveis segundo o estado de vida das crianças e Teste de Qui-quadrado*

Variáveis	n	Estado de vida		p-valor
		Censuradas (n=110322)	Evento (n=6357)	
País				
Moçambique	9289 (9,6)	9396 (95,3)	468 (4,7)	<0,001
Tanzânia	10783 (20,3)	10497 (96,4)	388 (3,6)	<0,001
África do Sul	3548 (10,7)	3444 (96,4)	128 (3,6)	<0,001
Zâmbia	9959 (5,2)	9361 (95,1)	480 (4,9)	<0,001
Zimbabwe	6132 (3,8)	6055 (94,3)	363 (5,7)	<0,001
Angola	14322 (9,9)	12669 (94,9)	688 (5,1)	<0,001
Congo	18716 (27,0)	17017 (92,5)	1373 (7,5)	<0,001
Comores	3149 (0,2)	3099 (95,8)	137 (4,2)	<0,001
Lesoto	3138 (0,4)	2896 (93,1)	216 (6,9)	<0,001
Madagáscar	12499 (7,6)	11632 (94,2)	713 (5,8)	<0,001
Malawi	17286 (4,8)	16548 (95,1)	847 (4,9)	<0,001
Namíbia	5046 (0,5)	4588 (95,5)	215 (4,5)	<0,001
Eswatini	2812 (0,1)	2553 (90,2)	276 (9,8)	<0,001
Sexo da criança				
Masculino	58449 (50,3)	54617 (94,2)	3439(5,8)	<0,001
Feminino	58230 (49,7)	55137 (95,2)	2853(4,8)	<0,001
Local de residência				
Rural	80179 (65,3)	75247 (94,4)	4512 (5,6)	0,004
Urbano	36500 (34,7)	34507 (95,2)	1779 (4,8)	0,004
Índice de riqueza				
Baixo	54397 (45,0)	49648 (94,3)	3057 (5,7)	<0,001
Médio	24127 (20,2)	21877 (94,7)	1262 (5,3)	<0,001
Alto	38155 (34,8)	38229 (95,3)	1973 (4,7)	<0,001
Nível de escolaridade da mãe				
Sem escolaridade	22397 (18,4)	19891 (94,6)	1139 (5,4)	<0,001

Tabela 4.1 (continuação)

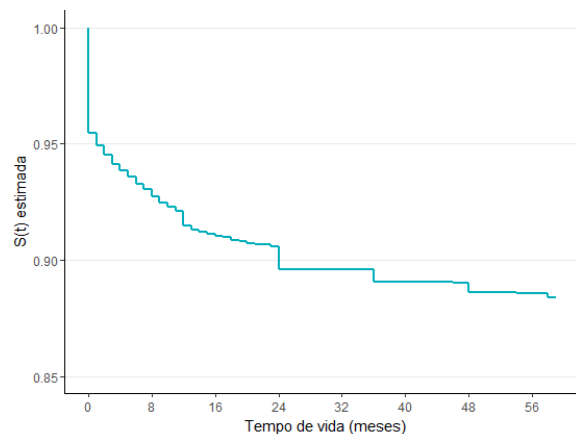
Variáveis	n	Estado de vida		p-valor
		Censuradas (n=110322)	Evento (n=6357)	
Primário	52718 (43,7)	49966 (94,2)	3133 (5,8)	<0,001
Secundário+	41553 (37,9)	39887 (95,4)	2020 (4,6)	<0,001
Nível de escolaridade do marido				
Sem escolaridade	12752 (12,3)	11727 (95,2)	624 (4,8)	0,37
Primário	35139 (37,1)	33253 (94,7)	2000 (5,3)	0,37
Secundário+	42899 (50,5)	40981 (94,2)	2280 (5,8)	0,37
Idade da mãe				
<18	2406 (1,8)	2189 (93,6)	167 (6,4)	0,002
18-29	66859 (56,8)	63088 (94,7)	3622 (5,3)	0,002
30-39	37699 (33,0)	35686 (94,9)	1885 (5,1)	0,002
40+	9715 (8,4)	8792 (93,7)	618 (6,3)	0,002
Sexo do chefe do agregado familiar				
Masculino	84273 (74,0)	80441 (94,7)	4580 (5,3)	0,685
Feminino	32406 (26,0)	29313 (94,8)	1712 (5,2)	0,685
Estado marital da mãe				
Nunca se casou	12854 (11,6)	11561 (94,6)	660 (5,4)	0,154
Casada/Vive marital-mente	92104 (79,0)	87354 (94,7)	4936 (5,3)	0,154
Viúva/Divorciada/Separada	11721 (9,4)	10840 (94,0)	696 (6,0)	0,154
Ocupação da mãe				
Desempregada	39607 (39,0)	37763 (95,1)	2058 (4,9)	0,001
Empregada	64530 (61,0)	60131 (94,2)	3578 (5,8)	0,001
Número de membros no agregado familiar				
≤5	55139 (46,4)	51725 (93,5)	3627 (6,5)	<0,001
>5	61540 (53,6)	58029 (95,3)	2665 (4,7)	<0,001
Tipo de nascimento				
Singular	112939 (96,8)	106738 (95,1)	5623 (4,9)	<0,001
Gemelar	3740 (3,2)	3016 (81,8)	669 (18,2)	<0,001
Tipo de parto				
Cesariano	6884 (8,0)	6629 (94,5)	432 (6,5)	0,877
Normal	101670 (92,0)	95073 (94,6)	5481 (5,4)	0,877
Peso à nascença				
<2500g	7949 (7,2)	7435 (92,8)	615 (7,2)	<0,001
2500-4500g	64760 (62,4)	63577 (96,4)	2444 (3,6)	<0,001
>4500g	35910 (30,5)	30787 (91,5)	2792 (8,5)	<0,001
Duração da amamentação				
Nunca amamentada	5107 (5,6)	3408 (68,4)	1574 (31,6)	<0,001
Actualmente amamentada	38829 (37,7)	38456 (100,0)	0(0,0)	<0,001
<6 meses	816 (2,1)	734 (87,7)	103 (12,3)	<0,001
6-12 meses	983 (1,9)	920 (93,5)	64 (6,5)	<0,001
12+ meses	62093 (52,6)	57607 (93,5)	4012 (6,5)	<0,001
Número de consultas pré-natais				
Nenhuma	6687 (10,1)	5808 (94,1)	385 (5,9)	<0,001
1 a 3	18410 (23,7)	17834 (96,3)	658 (3,7)	<0,001
4 ou +	49005 (66,3)	47470 (96,9)	1617 (3,1)	<0,001
Intervalo entre nascimentos				
<12 meses	1197 (1,4)	980 (82,5)	199 (16,5)	<0,001
12-48 meses	61849 (71,3)	57786 (94,7)	3271 (5,3)	<0,001
49+ meses	24261 (27,3)	23370 (96,1)	1022 (3,9)	<0,001
Ordem de nascimento				
Primogénito	29091 (24,2)	27424 (94,3)	1742 (5,7)	<0,001
2-3	43320 (37,8)	41287 (95,4)	2070 (4,6)	<0,001
4+	44268 (38,0)	41043 (94,2)	2479 (5,8)	<0,001
Local de parto				
Casa	30050 (25,0)	26237 (93,3)	1828 (6,7)	<0,001
Hospital público	66889 (65,9)	63718 (95,1)	3359 (4,9)	<0,001
Hospital privado	7188 (8,2)	7397 (94,9)	376 (5,1)	<0,001
Outro local	1177 (0,9)	1035 (92,6)	75 (7,4)	<0,001
Gravidez indesejada				
Não	73151 (68,3)	68083 (94,3)	4039 (5,7)	0,013
Sim	35519 (31,7)	33764 (95,3)	1828 (4,7)	0,013

Tabela 4.1 (continuação)

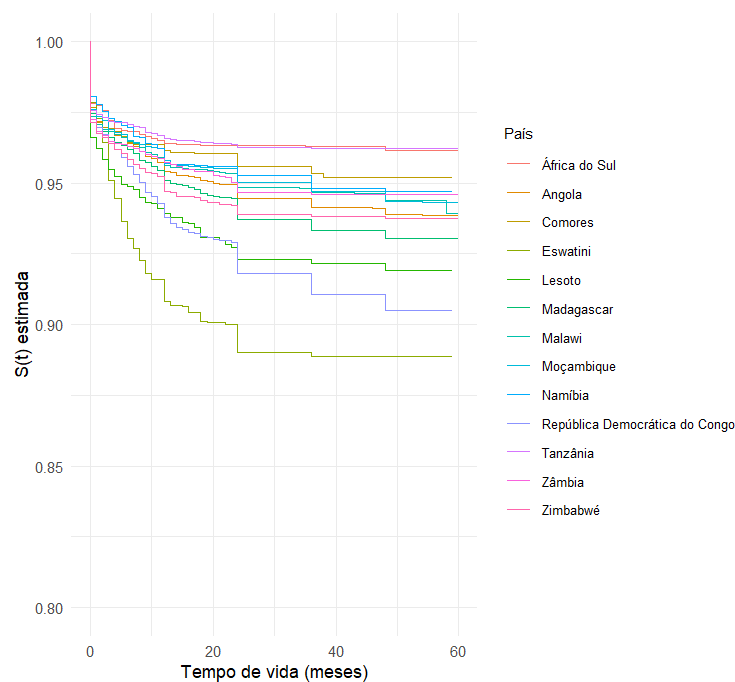
Variáveis	n	Estado de vida		p-valor
		Censuradas (n=110322)	Evento (n=6357)	
Exposição às mídias				
Sem exposição	49058 (87,5)	44673 (94,0)	2766 (6,0)	<0,001
Uma vez por semana	1286 (2,7)	1283 (96,9)	40 (3,1)	<0,001
Pelo menos uma vez por semana	3559 (9,3)	3461 (97,0)	122 (3,0)	<0,001
Todos os dias	400 (0,5)	468 (95,0)	28 (6,0)	<0,001
Já vacinou				
Não	7266 (20,0)	6643 (100,0)	0(0,0)	-
Sim	27080 (80,0)	26579 (100,0)	0(0,0)	-
Incidência de diarreia				
Não	92543 (86,7)	91622 (100,0)	0 (0,0)	-
Sim	15690 (13,3)	16020 (100,0)	0 (0,0)	-
Incidência de infecção respiratória				
Sim	2080 (1,5)	2075 (100,0)	0 (0,0)	-
Não	114599 (98,5)	107680 (94,5)	6292 (5,5)	-
Uso de rede mosquiteira				
Não	53124 (42,3)	50682 (94,4)	3048 (5,6)	0,303
Sim	56869 (57,7)	52733 (94,6)	2900 (5,4)	0,303
Fonte de água				
Não melhorada	42312 (39,0)	38376 (93,8)	2488 (6,2)	<0,001
Melhorada	72038 (61,0)	69287 (95,3)	3656 (4,7)	<0,001
Sanitário				
Não melhorado	53469 (50,5)	48664 (94,0)	3064(6,0)	<0,001
Melhorado	54359 (49,5)	53320 (95,3)	2795(4,7)	<0,001

4.2 Curvas de Kaplan-Meier

Para explorar melhor os dados, estimaram-se as curvas de sobrevivência de Kaplan-Meier para os grupos das covariáveis. Este procedimento proporciona uma visualização clara do comportamento dessas variáveis em relação às probabilidades de sobrevivência ao longo do tempo. A Figura 4.3 ilustra a estimativa de Kaplan-Meier para o tempo de sobrevivência geral, onde se verifica que a probabilidade de sobrevivência diminui progressivamente ao longo do tempo, embora não tenda a zero durante o período de observação. Observa-se, ainda, na Figura 4.3, um decréscimo mais acentuado das probabilidades no intervalo entre 0 a 24 meses de vida, o que sugere um risco acrescido de mortalidade neste período.

Figura 4.3: *Curva de sobrevivência estimada.*

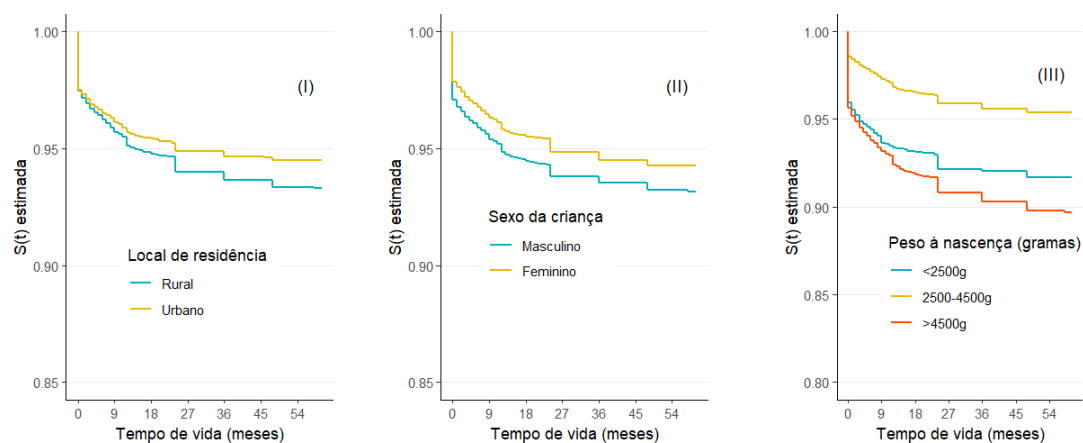
A Figura 4.4 apresenta as curvas de sobrevivência de Kaplan-Meier para diferentes países. Essas curvas revelam que a probabilidade de sobrevivência ao evento diminui de maneira mais gradual para crianças oriundas da Zâmbia, África do Sul e Tanzânia. Em contraste, para aquelas provenientes de Eswatini, essa probabilidade apresenta uma queda mais acentuada, indicando um risco mais elevado. Ademais, tanto as crianças que habitam em áreas urbanas quanto aquelas que residem em regiões rurais apresentam uma probabilidade de sobrevivência superior a 80%, no entanto, as que vivem em zonas rurais enfrentam um risco mais elevado (Figura 4.5).

Figura 4.4: *Curvas de Kaplan-Meier estimadas para os grupos da variável País.*

A Figura 4.5 também ilustra as curvas de sobrevivência, segmentadas conforme o sexo da

criança. É importante ressaltar que a probabilidade de sobrevivência para ambos os sexos permanece acima de 80% ao longo de todo o período de observação, no entanto, observa-se que as crianças do sexo masculino apresentam um risco superior de mortalidade em comparação às do sexo feminino. Na sequência, a Figura 4.5 traz as curvas correspondentes aos grupos definidos pela covariável Peso ao nascer. Nota-se que a curva relacionada ao grupo de crianças nascidas com peso entre 2500 e 4500 gramas permanece estabilizada acima das curvas dos grupos que apresentam peso inferior a 2500 gramas ou superior a 4500 gramas. Este padrão sugere que estes dois últimos grupos apresentam maior risco de morte em comparação com o grupo de crianças com peso entre 2500 e 4500 gramas à nascença.

Figura 4.5: *Curvas de Kaplan-Meier estimadas para as variáveis: Local de residência (I), Sexo da Criança (II) e Peso à nascença (III).*



As curvas de sobrevivência apresentadas nas Figuras 5.1 e 5.2 revelam diferenças entre os grupos das variáveis consideradas: Duração da amamentação, Número de consultas pré-natais, Ordem de nascimento, Tipo de nascimento, Intervalo entre nascimentos e Gravidez indesejada. Observa-se que a taxa de sobrevivência tende a estabilizar-se acima de 80% nas diversas ordens de nascimento. Em contraste, as crianças provenientes de gestações gême-lares exibem uma probabilidade de sobrevivência inferior a 70%, em comparação com as crianças de gestações singulares, cuja probabilidade se aproxima de 90%. De forma semelhante, observa-se (Figura 5.4) uma ligeira vantagem na sobrevivência entre crianças com acesso a sanitários e fontes de água melhorados. Tendências comparáveis são observadas nas demais variáveis analisadas.

Para complementar a comparação entre as curvas de sobrevivência estimadas, são apresentados os p -valores associados ao teste de log-rank na Tabela 4.2. Observou-se uma diferença significativa nas probabilidades de sobrevivência de variáveis como Local de residência, Sexo da criança, Peso da criança à nascença e Duração da amamentação. Em contrapartida, variáveis como Nível de escolaridade do marido, Sexo do chefe do agregado familiar, Número

de membros no agregado familiar, Tipo de parto e Uso de rede mosquiteira não mostraram diferenças significativas nas curvas de sobrevivência entre os grupos.

Tabela 4.2: *Teste log-rank para comparação das curvas de sobrevivência*

Covariáveis	Estatística	p-valor
País	383,56	<0,001
Local de residência	15,75	<0,001
Sexo da criança	2,38	<0,001
Peso à nascença	139,15	<0,001
Duração da amamentação	989,02	<0,001
Número de consultas pré-natais	24,27	<0,001
Ordem de nascimento	6,16	< 0,001
Tipo de nascimento	38,85	<0,001
Intervalo entre nascimentos	66,70	<0,001
Índice de riqueza	28,98	<0,001
Idade da mãe	21,24	0,005
Nível de escolaridade do marido	7,70	0,308
Sexo do chefe do agregado familiar	2,14	0,644
Nível de escolaridade da mãe	69,70	<0,001
Número de membros do agregado familiar	82,38	0,422
Estado civil da mãe	20,42	<0,001
Ocupação da mãe	14,32	<0,001
Gravidez indesejada	0,34	< 0,001
Tipo de parto	5,36	0,755
Local de parto	51,85	<0,001
Fonte de água	29,80	< 0,001
Sanitário	17,52	<0,001
Já vacinou	-	-
Incidência de diarreia	-	-
Incidência de infecção respiratória	-	-
Uso de rede mosquiteira	4,40	0,523
Exposição às mídias	6,16	<0,001

4.3 Modelo de regressão de Cox com fragilidade

4.3.1 Escolha da distribuição de fragilidade

A Tabela 4.3 apresenta medidas de ajuste para comparação de modelos de Cox com diferentes distribuições com vista a identificação da distribuição de fragilidade para os *clusters*. O modelo com fragilidade lognormal estimado pelo método de minimização do AIC foi o selecionado, uma vez que apresenta o menor AIC .

Tabela 4.3: *Comparação das medidas de ajuste dos modelos de fragilidade*

Medidas de ajuste	Gama (AIC)	Lognormal (AIC)	Gama (EM)	Lognormal (REML)
Variância do efeito aleatório	0,18	0,25	0,26	0,26
Concordance	0,79	0,79	0,79	0,79
AIC	145941,80	145771,40	145980,21	145774,12

4.3.2 Selecção de variáveis

A identificação das variáveis a incluir no modelo foi feita através da aplicação do método de seleção descrito por Collet (conforme exposto na secção 3.2.3 do capítulo 2), o modelo final

incluiu as variáveis que apresentaram associação estatisticamente significativa com o tempo até o evento, nomeadamente: Sexo da criança, Peso à nascença, Intervalo entre nascimentos, Idade da mãe, Tipo de parto, Exposição à mídia, Número de consultas pré-natal, Local de parto, Tipo de nascimento, Nível de escolaridade do marido e Número de membros do agregado familiar. Variáveis como País, Ordem de nascimento, Nível de escolaridade da mãe, Estado marital e Uso de rede mosquiteira foram excluídas por não apresentarem significância estatística. Ademais, algumas variáveis não foram consideradas no processo de selecção devido ao elevado número de valores omissos, o que poderia comprometer a robustez do modelo. A interação entre Local de residência e Local de parto também foi testada, não sendo significativa no contexto deste estudo.

4.3.3 Análise residual

Na análise exploratória inicial, parece plausível não duvidar da proporcionalidade das funções de risco na medida em que as estimativas das funções de sobrevivência correspondentes às categorias de cada covariável não se cruzam. Entretanto, esta suposição foi formalmente validada através da aplicação do teste residual de Schoenfeld apresentados na Tabela 4.4 e avaliação dos resíduos de Schoenfeld apresentados na Figura 5.6.

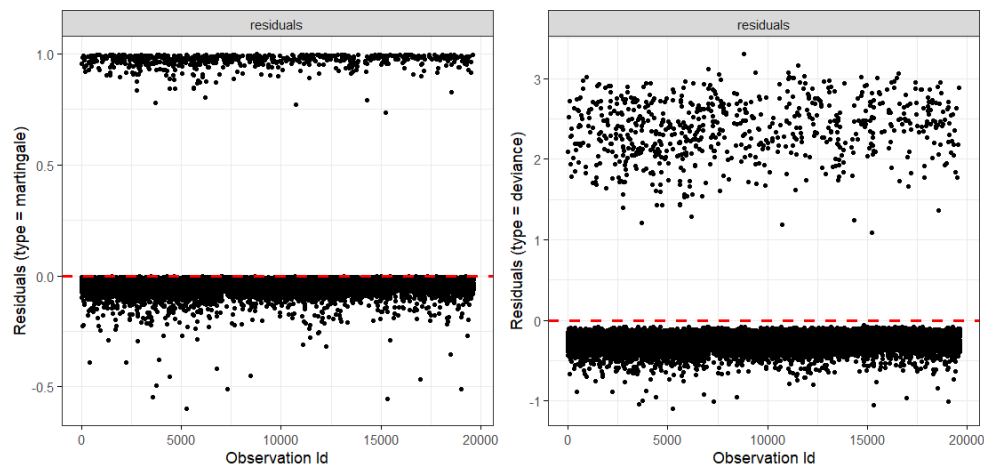
Tabela 4.4: *Teste de proporcionalidade dos riscos pelo método de Schoenfeld*

Covariável	Qui-quadrado	gl	p-valor
Sexo da criança	1,85	0,99	0,172
Peso à nascença	5,57	1,97	0,60
Intervalo entre nascimentos	2,54	1,98	0,276
Idade da mãe	0,27	2,98	0,964
Tipo de parto	5,86	0,99	0,15
Exposição à mídia	3,52	2,97	0,314
Número de consultas pré-natais	0,01	1,97	0,992
Local de parto	4,96	2,95	0,170
Tipo de nascimento	3,93	0,98	0,046
Nível de escolaridade do marido	0,50	1,97	0,772
Número de membros do agregado familiar	1,13	0,99	0,286
GLOBAL	29,34	116,96	0,9998

Com base na inspecção dos gráficos da Figura 4.6, entende-se que não há evidências de que pontos atípicos (*outliers*) estejam a impactar a adequação do modelo ajustado, visto que os gráficos mostram um comportamento aleatório na distribuição dos resíduos.

4.3.4 Interpretação dos coeficientes do modelo multinível

A Tabela 4.5 apresenta os coeficientes estimados do modelo de Cox com fragilidade lognormal. Com base nos efeitos ajustados das covariáveis no modelo e considerando a hipótese

Figura 4.6: *Representação gráfica dos Resíduos de Martingala e Deviance.*

de proporcionalidade dos riscos, afirma-se com 95% de confiança que:

As crianças do sexo feminino apresentam cerca de 14% menos risco de mortalidade em comparação às crianças do sexo masculino e aquelas com mães de idades entre 18 e 29 anos têm 48% menos risco de morte em relação àquelas cujas mães têm 40 anos ou mais. Para mães com idade entre 30 e 39 anos, espera-se que a redução seja de cerca de 39%, controlando os restantes factores. No contexto familiar, espera-se que o risco de morte em crianças menores de cinco anos seja aproximadamente 45% inferior para crianças inseridas em agregados familiares com acesso à mídia pelo menos uma vez por semana, e cerca de 56% inferior para aquelas com acesso menos frequente, em comparação com crianças sem qualquer exposição mediática. Além disso, a composição do agregado familiar mostra-se significativa, revelando que o risco de morte é cerca de duas vezes maior para crianças que pertencem a agregados familiares com menos de cinco membros. Quanto ao nível de escolaridade do marido, verificou-se uma redução do risco em torno de 16% para crianças cujas mães têm maridos com escolaridade primária, e 28% para os sem educação formal, face à escolaridade secundária ou superior, ajustando para os demais factores.

No que concerne às características relacionadas com o nascimento, espera-se que crianças com peso ao nascer entre 2500 e 4500 gramas apresentem uma redução do risco de morte na ordem dos 26%, relativamente às nascidas com peso inferior a 2500 gramas, ainda que esta diferença não tenha alcançado significância estatística no presente estudo. Em contrapartida, espera-se que o risco aumente em cerca de 57% para crianças com peso superior a 4500 gramas. O intervalo entre nascimentos revelou ser um factor protector, uma vez que para intervalos entre 12 e 48 meses, espera-se que o risco de morte seja aproximadamente 42% inferior, e para intervalos superiores a 49 meses a redução pode atingir cerca de 59%, em comparação com intervalos inferiores a 12 meses. Relativamente ao tipo de parto, espera-se

que os partos normais, apresentem um risco cerca de 47% inferior aos partos por cesariana. Já as crianças provenientes de gestações gemelares apresentam risco de morte superior a quatro vezes o observado nos nascimentos singulares.

No que concerne ao cuidado pré-natal e local de parto, espera-se que a realização de consultas pré-natais reduza substancialmente o risco de morte infantil: cerca de 28% para mães com 1 a 3 consultas e 29% para aquelas com 4 ou mais consultas, em comparação com mães que não realizam qualquer consulta. Adicionalmente, espera-se que os partos ocorridos em casa apresentem um risco cerca de 38% inferior face aos partos em outros locais que não sejam hospitais ou centros de saúde, ajustando para os restantes factores. Estes resultados reforçam a importância dos determinantes sociodemográficos, das condições do nascimento e do acesso a cuidados pré-natais na mortalidade infantil, destacando a complexidade e a inter-relação dos factores envolvidos.

Tabela 4.5: *Coeficientes de regressão do modelo de Cox com fragilidade lognormal*

Covariável	Coef.	HR	IC(95%)	Erro Padrão	P-valor
Sexo da criança					
Masculino (Ref.)	—	—	—	—	—
C ₁ : Feminino	-0,15	0,86	(0,74–1,00)	0,08	0,057
Peso da criança					
<2500g (Ref.)	—	—	—	—	—
C ₂ : 2500–4500g	-0,31	0,74	(0,53–1,02)	0,17	0,064
C ₃ : >4500g	0,45	1,57	(1,11–2,21)	0,18	0,010
Intervalo entre nascimentos					
<12 meses (Ref.)	—	—	—	—	—
C ₄ : 12–48 meses	-0,55	0,58	(0,35–0,95)	0,27	0,031
C ₅ : 49+ meses	-0,90	0,41	(0,24–0,68)	0,27	< 0,001
Idade da mãe					
40+ (Ref.)	—	—	—	—	—
C ₆ : <18	-1,40	0,25	(0,03–1,79)	1,008	0,17
C ₇ : 18–29	-0,66	0,52	(0,41–0,65)	0,12	< 0,001
C ₈ : 30–39	-0,49	0,61	(0,49–0,76)	0,11	< 0,001
Tipo de parto					
Cesariano (Ref.)	—	—	—	—	—
C ₉ : Normal	-0,63	0,53	(0,39–0,72)	0,15	< 0,001
Exposição à mídia					
Sem exposição (Ref.)	—	—	—	—	—
C ₁₀ : Pelo menos uma vez por semana	-0,60	0,55	(0,37–0,92)	0,20	0,003
C ₁₁ : Menos de uma vez por semana	-0,82	0,44	(0,21–0,93)	0,38	0,032
C ₁₂ : Todos os dias	0,46	1,59	(0,78–3,24)	0,36	0,204
Número de consultas pré-natal					
Nenhuma (Ref.)	—	—	—	—	—
C ₁₃ : 1–3 consultas	-0,32	0,72	(0,57–0,92)	0,12	0,008
C ₁₄ : 4+ consultas	-0,34	0,71	(0,57–0,89)	0,11	0,003
Local de parto					
Outro local (Ref.)	—	—	—	—	—

Continua na próxima página

Tabela 4.5 (continuação)

Covariável	Coef.	HR	IC(95%)	Erro Padrão	P-valor
C ₁₅ : Hospital público	-0,11	0,90	(0,50–1,58)	0,29	0,702
C ₁₆ : Hospital privado	-0,06	0,94	(0,50–1,78)	0,32	0,856
C ₁₇ : Casa	-0,48	0,62	(0,35–1,08)	0,29	0,093
Tipo de nascimento					
Singular (Ref.)	–	–	–	–	–
C ₁₈ : Gemelar	1,49	4,42	(3,23–6,05)	0,16	< 0,001
Nível de escolaridade do marido					
Secundário + (Ref.)	–	–	–	–	–
C ₁₉ : Sem escolaridade	-0,33	0,72	(0,58–0,90)	0,11	0,004
C ₂₀ : Primário	-0,17	0,84	(0,71–1,00)	0,08	0,052
Número de membros no agregado familiar					
>5 membros (Ref.)	–	–	–	–	–
C ₂₁ : ≤5 membros	0,73	2,08	(1,77–2,46)	0,09	< 0,001
Efeito de fragilidade					<0,001

4.4 Discussão dos Resultados

A mortalidade infanto-juvenil continua a ser um grave problema de saúde pública, especialmente nos países em desenvolvimento e, em particular, na África Austral (Gobebo, 2021; UNICEF, 2024). A sua complexidade é evidenciada pela variedade de factores que a influenciam (Ettarh, 2012; Yaya et al., 2018; Aheto, 2019). Este estudo aprofunda a compreensão da associação entre factores maternos, neonatais, contextuais e a mortalidade infanto-juvenil. A observação de que as crianças do sexo feminino apresentam um risco reduzido de mortalidade alinha-se aos resultados de pesquisas, como a conduzida por Costa e Victora (2017) que indicam um aumento nas taxas de mortalidade entre crianças do sexo masculino menores de cinco anos. Este fenómeno se revela, em muitas ocasiões, mais acentuado nos primeiros dias de vida, sendo também influenciado pela idade gestacional, conforme evidenciado por Sahu e Nair (2015).

Vários estudos destacam o papel do baixo peso ao nascer, do intervalo entre nascimentos reduzido, do acesso limitado a cuidados pré-natais qualificados e das desigualdades socio-económicas como determinantes críticos para a sobrevivência infantil Croft et al. (2023); Dhawan et al. (2023). No caso do presente estudo, o peso ao nascer revela-se como um factor determinante na mortalidade de crianças menores de cinco anos, visto que se observou que crianças nascidas com excesso de peso têm um risco relativo de morte mais elevado do que as nascidas com menos peso. Similarmente, para crianças cujas mães não realizaram consultas pré-natais, observou-se um risco aumentado. O que sugere que intervenções dirigidas à nutrição materna e à qualidade dos cuidados pré-natais continuam cruciais, tal como mencionado por Tekelab et al. (2019), que estimaram uma redução de quase 40% na mortali-

dade infantil para crianças cujas mães fizeram pelo menos uma consulta pré-natal qualificada.

A idade da mãe é um factor determinante na mortalidade infantil. No presente estudo, verificou-se que crianças nascidas de mães com idades entre 18 a 29 e 30 a 39 anos apresentam um risco inferior de morte quando comparadas às nascidas de mães com 40 anos ou mais. Este resultado corrobora os achados de Mahy (2003), que destacam que gestações em idades muito avançadas estão associadas a um maior risco de complicações obstétricas, podendo resultar em nascimentos com baixo peso e, consequentemente, maior vulnerabilidade à mortalidade precoce. Ademais, os pesquisadores D. Silva (2016) e Tesema et al. (2021) também apontam para o risco elevado de mortalidade entre crianças de mães mais velhas, frequentemente relacionado a uma diminuição na reserva biológica, factor que impacta negativamente a saúde do recém-nascido.

A importância do tipo de nascimento na sobrevivência infanto-juvenil foi evidente, revelando que os nascimentos gemelares representam um risco acrescido de mortalidade. Para intervalos entre 12 e 48 meses, o risco é cerca de 42% inferior, e para intervalos superiores a 49 meses, a redução pode atingir cerca de 59%. Essa relação é amplamente discutida por Akinyemi et al. (2013), que apontam, no contexto da Zâmbia, o aumento do risco associado aos partos múltiplos, frequentemente vinculados à prematuridade e ao baixo peso ao nascer. Além disso, os autores Rutstein (2005) e Kozuki (2013) ressaltam que intervalos de gestação curtos impõem uma sobrecarga fisiológica às mães, diminuindo o tempo disponível para os cuidados necessários, o que aumenta a vulnerabilidade das crianças. Entretanto, esse resultado também reflecte desigualdades nos sistemas de saúde. Em contextos onde o acompanhamento pré-natal e neonatal é de qualidade, os riscos associados a partos múltiplos podem ser significativamente reduzidos (Cleary Goldman et al., 2005).

Os resultados do presente estudo revelam um risco aproximadamente 47% menor de partos normais em comparação aos partos por cesariana, o que está em consonância com a literatura existente. Pesquisadores como Betrán et al. (2016) afirmam que cesarianas não indicadas clinicamente estão associadas a uma taxa de mortalidade significativamente mais alta em comparação aos partos normais. Essa observação ressalta a importância crucial das indicações médicas na decisão de realizar esse procedimento. Outro resultado relevante foi a influência do intervalo entre nascimentos. As crianças nascidas de gestações com intervalo intergestacionais maiores de 12 meses têm o risco de morte reduzido em relação aos intervalos menores, alinhando o que Mahy (2003) já havia evidenciado para contextos africanos, que os intervalos curtos sobrecarregam fisiologicamente as mães e reduzem o tempo disponível para cuidados, o que aumenta a vulnerabilidade infantil.

O número de membros no agregado familiar também destacou-se como um factor determi-

nante à mortalidade infanto-juvenil. Para o presente estudo, o risco de morte é maior para crianças inseridas em famílias compostas por menos de cinco comparativamente às famílias numerosas, o que pode estar associado ao número elevado de filhos e consequentemente mais experiência em cuidados pré e pós natais. Esta observação é consistente com evidências de pesquisas anteriores. Por exemplo, uma análise de sobrevivência multinível realizada na Etiópia demonstrou que crianças provenientes de famílias com quatro a seis membros ou mais apresentavam menor probabilidade de morrer antes dos cinco anos, em comparação com crianças de famílias menores, com um a três membros (Argawu & Hirko, 2020). Por outro lado, as conclusões de Ayele et al. (2017) e Ekholuenetale et al. (2020) indicam que, em contextos de pobreza, famílias numerosas podem enfrentar limitações económicas e logísticas que dificultam o fornecimento de cuidados adequados a todas as crianças. Assim, a divergência observada entre os estudos pode reflectir diferenças contextuais, metodológicas e temporais, incluindo variações no tamanho da amostra e nos modelos estatísticos aplicados.

O nível de escolaridade do marido é frequentemente associado à mortalidade de crianças. Autores como Mirza et al. (2021) e Aly e Grabowski (1990) argumentam que o aumento da escolaridade paterna tende a reduzir o risco de morte de crianças, por estar relacionado a um maior acesso à informação, melhores condições económicas e decisões mais favoráveis à saúde infantil. Entretanto, Bicego e Boerma (1993) enfatizam que, em certos contextos, essa relação positiva pode não ser observada ou até mesmo se inverter. No presente estudo, observou-se que níveis mais baixos de escolaridade do marido se associam a um menor risco de morte infantil. Entre as explicações possíveis para esse fenómeno, destaca-se o maior envolvimento dos pais com maior nível de escolaridade no mercado de trabalho, o que reduz o tempo disponível para o cuidado direto das crianças, assim como a migração para áreas urbanas, que frequentemente expõe as famílias a riscos ambientais e sociais maiores.

Em contraste, variáveis habitualmente associadas à mortalidade infanto-juvenil, como local de residência, uso de redes mosquiteiras e nível de escolaridade da mãe, não apresentaram efeitos estatisticamente significativos para inclusão no modelo final. Apesar do local de residência, não ter apresentado um efeito estatístico significativo no modelo final deste estudo, a literatura comumente reconhece sua relevância, uma vez que conforme evidenciado por Tesfa et al. (2021), este factor influencia o acesso a serviços de saúde e recursos, resultando, em geral, em taxas de mortalidade infantil mais baixas em ambientes urbanos. A falta de resultados observados pode reflectir uma heterogeneidade presente nos contextos de saúde, na cobertura vacinal e nas condições de saneamento. Além disso, sugere-se que avanços recentes em programas de saúde pública podem ter contribuído para a redução de algumas desigualdades entre áreas urbanas e rurais, como indicado por Fink et al. (2022) em relação à cobertura de redes mosquiteiras.

Na África Austral, as curvas de sobrevivência associadas ao resultado do teste de log-rank evidenciaram diferenças no risco de mortalidade infanto-juvenil entre países, especialmente na República Democrática do Congo e em Eswatini, em comparação com as Comores e a Namíbia. Embora essas discrepâncias não tenham alcançado significância estatística para inclusão no modelo final, estão em consonância com resultados de outros estudos, como os de Mahy (2003) e Hill et al. (2012), que destacam o impacto das desigualdades económicas, da instabilidade política e dos níveis variados de investimento em saúde pública na determinação da sobrevivência infantil. Alguns estudos modelaram a variável País como um termo de fragilidade compartilhada. Um exemplo notável é o estudo de Tesfa et al. (2021), que aplicaram este modelo em uma amostra que abrangeu 33 países da África Subsariana. Outros estudos adotaram abordagens com fragilidade aninhada, por exemplo, ao nível de *cluster* e região (Ayele et al., 2017). É importante salientar que algumas investigações alertaram para os desafios associados à implementação de múltiplos termos de fragilidade, que podem resultar em dificuldades de convergência. Além disso, a adoção de múltiplos termos de fragilidade pode restringir a capacidade de selecção entre as diferentes distribuições (Austin, 2017).

Em síntese, este estudo confirma a relevância de determinantes clássicos, como peso ao nascer e tipo de parto, idade da mãe mas também destaca a persistência de desigualdades estruturais e a importância de abordagens metodológicas robustas para informar políticas públicas mais eficazes e equitativas.

Capítulo 5

CONCLUSÕES E RECOMENDAÇÕES

5.1 Conclusões

A utilização de métodos estatísticos de análise de sobrevivência, incluindo modelos com fragilidade, permitiu identificar e compreender os factores de risco associados à mortalidade infanto-juvenil na África Austral. Com base nos resultados obtidos, conclui-se que:

1. Factores biológicos e demográficos da criança, como os partos múltiplos, peso ao nascer abaixo de 2500g ou acima de 4500g, intervalos curtos entre nascimentos e o menor número de membros no agregado familiar estão significativamente associados a um aumento do risco de morte de crianças menores de cinco anos na região.
2. Conclui-se, assim, que são necessárias medidas de intervenção integradas e contextualmente adaptadas, que combinem melhorias nos cuidados materno-infantis, planeamento familiar, nutrição e educação, com políticas de redução da pobreza e fortalecimento dos sistemas de saúde, para alcançar de forma equitativa a redução sustentável da mortalidade infantil na África Austral.

As conclusões alcançadas neste estudo poderão também servir de base para reflexões em contextos comparáveis, auxiliando profissionais de saúde, responsáveis por políticas públicas e comunidades na adopção de estratégias orientadas para a melhoria da sobrevivência infantil e para o avanço no cumprimento dos objectivos globais de desenvolvimento sustentável.

5.2 Recomendações

Reconhecendo a complexidade dos determinantes da mortalidade infanto-juvenil na África Austral e a necessidade de soluções baseadas em evidências, recomenda-se:

1. Reforço dos serviços de saúde materno-infantil, garantindo acesso universal e de qualidade ao pré e pós-natal, com foco na prevenção do baixo peso ao nascer, no acompanhamento de gravidezes múltiplas e na promoção do planeamento familiar para espaçamento saudável entre nascimentos.
2. Promoção da educação para a saúde com abordagem familiar e comunitária, envolvendo mães, pais e encarregados de educação, para estimular práticas de cuidado infantil adequadas, valorizando o papel do parceiro, que se revelou um factor relevante nos resultados do estudo. Para reforçar esta estratégia, recomenda-se o desenvolvimento de campanhas de sensibilização em parceria com universidades, agentes comunitários e organizações locais.
3. Implementação de um sistema integrado de gestão de dados em saúde, capaz de integrar informações de diferentes fontes e fornecer dados completos sobre causas de morte e variáveis relevantes, incluindo crianças falecidas. Este sistema deve articular-se com programas de saúde materno-infantil permitindo a identificação precoce de riscos e a orientação de intervenções intersectoriais baseadas em evidências.

5.3 Limitações

Durante a pesquisa, algumas limitações foram identificadas:

- A disparidade temporal na recolha de dados (2013-2023), que compromete a actualidade dos resultados, acrescida da ausência de IDS em alguns países da África Austral;
- Algumas variáveis relevantes, como Incidência de diarreia, Vacinação e Incidência de infecção respiratória, são recolhidas apenas em crianças censuradas, o que impossibilita a inclusão na análise de sobrevivência, bem como a inclusão de variáveis recolhidas após a morte;
- A falta de dados sobre as causas de morte específicas que impede a análise de riscos competitivos;
- O carácter transversal do estudo, que não permite estabelecer uma relação de causa e efeito entre as variáveis analisadas;
- O excesso de observações omissas em algumas variáveis, que compromete a precisão das análises e pode resultar em erros do tipo II.

Embora existam limitações a serem consideradas, este estudo oferece contribuições significativas para a compreensão dos factores associados à mortalidade infanto-juvenil na África Austral. A utilização de dados representativos e metodologia adequada permitiu a obtenção de resultados que podem subsidiar políticas públicas e orientar futuras pesquisas na área.

Referências

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4). <https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-4/Nonparametric-Inference-for-a-Family-of-Counting-Processes/10.1214/aos/1176344247.full>
- Adebowale, A. S., Fagbamigbe, A. F., Morakinyo, O., & Obembe, T. (2020). Parental educational homogamy and under-five mortality in sub-Saharan Africa: Clarifying the association's intricacy. *Scientific African*, 8. <https://doi.org/10.1016/j.sciaf.2019.e00255>
- Aheto, J. (2019). Predictive model and determinants of under-five child mortality: evidence from the 2014 ghana demographic and health survey. *BMC Public Health*, 8, 175. <https://doi.org/10.1186/s12889-019-6390-4>
- Akinyemi, J. O., Bamgboye, E. A., & Ayeni, O. (2013). New trends in under-five mortality determinants and their effects on child survival in Nigeria: A review of childhood mortality data from 1990-2008. *African Population Studies*, 27(1), 1–10. <https://aps.journals.ac.za/pub/article/view/5> (Acesso em 11-09-2024)
- Aly, H., & Grabowski, R. (1990). Education and child mortality in egypt. *World Development*, 18(1), 733-745. [https://doi.org/10.1016/0305-750X\(90\)90021-0](https://doi.org/10.1016/0305-750X(90)90021-0)
- Amuka, E., Mitiku, A., & Zeru, M. (2024). Spatiotemporal modeling of under-five mortality and associated risk factors in Ethiopia using 2000-2016 EDHS data. *BMC Pediatrics*, 24(1), 201. <https://doi.org/10.1186/s12887-024-04676-4>
- Anchisi, M. (2011). *Métodos não paramétricos para análise de dados de sobrevivência* (Dissertação de Mestrado). Universidade de Lisboa, Lisboa, Portugal.
- Andrade, A. B., & Cabete, B. (2012). *Análise de sobrevivência com acontecimentos múltiplos: Aplicação ao estudo do tempo até a ocorrência de enfarte do miocárdio* (Dissertação de Mestrado). Universidade de Lisboa.
- Aquino, B. (2024). *Um estudo sobre a evasão no curso de gestão pública da ufcg-vdsa via análise de sobrevivência* (Trabalho de Conclusão de Curso, Campina Grande, Brasil).

- <https://dspace.bc.uepb.edu.br/jspui/handle/123456789/32530> (Acesso em 09-12-2024)
- Argawu, A. S., & Hirko, G. (2020). *Multilevel modelling of time to death for under-five children: Evidence from the 2016 ethiopia demographic and health survey* (Tech. Rep.). Ethiopia Demographic and Health Survey / Research Report.
- Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, 85(2). <https://doi.org/10.1111/insr.12214>
- Ayele, D., Zewotir, T., & Mwambi, H. (2017). Survival analysis of under-five mortality using Cox and frailty models in Ethiopia. *Journal of Health, Population and Nutrition*, 36(1), 1–10. <https://doi.org/10.1186/s41043-017-0103-3>
- Balan, T. A., & Putter, H. (2019). Nonproportional hazards and unobserved heterogeneity in clustered survival data: When can we tell the difference? *Statistics in Medicine*, 38(15), 2895–2910. <https://doi.org/10.1002/sim.8171>
- Bedia, E. (2022). *Modelos de sobrevivência bivariados induzidos por fragilidade* (Tese de Doutorado). Universidade de São Paulo, São Paulo, Brasil.
- Bereta. (2002). *Testes de significância repetidos em análise de sobrevivência* (Dissertação de Mestrado). Universidade Estadual de Campinas, São Paulo.
- Betrán, A. P., Ye, J., Moller, A.-B., Zhang, J., Gülmezoglu, A. M., & Torloni, M. R. (2016). The increasing trend in caesarean section rates: global, regional and national estimates: 1990–2014. *The Lancet*, 388(10052), 77–86. [https://doi.org/10.1016/S0140-6736\(16\)30672-8](https://doi.org/10.1016/S0140-6736(16)30672-8)
- Bianchini, Z., & Albieri, S. (1999). Uma revisão dos principais aspectos dos planos amostrais das pesquisas domiciliares realizadas pelo ibge. *Revista Brasileira de Estatística*, 60(210), 111–124.
- Bicego, G. T., & Boerma, J. T. (1993). Maternal education and child survival: A comparative study of survey data from 17 countries. *Social Science & Medicine*, 36(9), 1207–1227. [https://doi.org/10.1016/0277-9536\(93\)90241-D](https://doi.org/10.1016/0277-9536(93)90241-D)
- Birhanie, A. L., Tessema, Z. T., Endalew, B., & Tamirat, K. S. (2025). Under-five mortality and its associated factors in sub-saharan africa: a multilevel analysis of recent demographic and health surveys data based on bayesian approach. *BMC Pediatrics*. <https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-025-05454-6> (Acesso em 03-05-2025)

- Boettiger, D. C., Treleaven, E., Kayentao, K., & Guindo. (2021). Household factors and under-five mortality in Bankass, Mali: results from a cross-sectional survey. *BMC Public Health*, 1295. <https://doi.org/10.1186/s12889-021-11275-7>
- Bonita, R., & Beaglehole, R. (2010). *Epidemiologia básica* (2nd ed.). São Paulo, Brasil: Livraria Santos Editora Ltda.
- Carvalho, M., Andreozzi, V., Campos, D., & Barbosa, M. (2011). *Análise de sobrevivência: Teoria e aplicações em saúde*. Rio de Janeiro: Editora Fiocruz.
- Chikandiwa, A., Burges, E., Ottembe, K., & Chimoyi, L. (2014). *Use of contraceptives, high risk births and under-five mortality in Sub Saharan Africa: evidence from Kenyan (2014) and Zimbabwean (2011) demographic health surveys* (Vol. 18). <https://doi.org/10.1186/s12905-018-0666-1> (Acesso em 04-01-2025)
- Cleary Goldman, J., Malone, F. D., Vidaver, J., Ball, R. H., Nyberg, D. A., Comstock, C. H., ... D'Alton, M. E. (2005). Impact of maternal age on obstetric outcome. *Obstetrics & Gynecology*, 105(5), 983–990. <https://doi.org/10.1097/01.AOG.0000158118.75532.51>
- Coelho, L. (2017). *Modelos longitudinais e de sobrevivência para a recidiva do cancro da mama* (Dissertação de Mestrado). Universidade de Minho, Brasil.
- Colosimo, E. A., & Giolo, S. R. (2024). *Análise de sobrevivência aplicada*. São Paulo, Brasil: Edgard Blucher.
- Comunidade para o Desenvolvimento da África Austral [SADC]. (2017). *Minimum standards for child and adolescent hiv, tb and malaria continuum of care and support in the sadc region (2013–2017)*. https://www.sadc.int/sites/default/files/2021-08/SADC_MINIMUM_STANDARDS_FOR_HIV_AND_AIDS_TB_HEPATITIS_B_AND_C_AND_SEXUALLY_TRANSMITTED_INFECTIONS_PREVENTION_TREATMENT_CARE_AND_SUPPORT_IN_PRISONS_IN_THE_SADC_REGION_A4.pdf (Acesso em 2025-08-11)
- Costa, J. C., & Victora, C. G. (2017). Viés de gênero na mortalidade de menores de cinco anos em países de baixa/média renda. *BMJ Global Health*, 2(4), 549. <https://doi.org/10.1136/bmjgh-2017-000549>
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- Croft, T. N., Allen, C. K., Zachary, B. W., et al. (2023). Guide to dhs statistics [Computer software manual]. <https://dhsprogram.com/publications/publication-dhsg1-dhs-statistics-guide.cfm> (Acesso em 19-08-2024)

- Dhawan, D., Pinnamaneni, R., & Viswanath, K. (2023). Association between mass media exposure and infant and young child feeding practices in india: a cross-sectional study. *Scientific Reports*, 13, 19353.
- DHS. (2012). *Demographic and health survey's dhs toolkit of methodology for the measure dhs phase iii project, implemented from 2008-2013*. <https://www.dhsprogram.com>. (Acesso em 30-10-2024)
- Ekholuenetale, M., Wegbom, A. I., Tudeme, G., & Onikan, A. (2020). Household factors associated with infant and under-five mortality in sub-saharan africa countries. *International Journal of Child Care and Education Policy*, 14(1), 10. <https://doi.org/10.1186/s40723-020-00075-1>
- Ely, D. M., & Driscoll, A. K. (2022). *Infant mortality in the united states, 2020: Data from the period linked birth/infant death file* (Vol. 71; National Vital Statistics Reports No. 5). National Center for Health Statistics. <https://www.cdc.gov/nchs/data/nvsr/nvsr71/nvsr71-05.pdf> (Acesso em 28-11-2024)
- Emmert-Streib, F., & Dehmer, M. (2019). Introduction to survival analysis in practice. *Theoretical Biology and Medical Modelling*, 16(1), 1. <https://doi.org/10.1186/s12976-019-0102-1>
- Ettarh, J., R e Kimani. (2012). Determinants of under-five mortality in rural and urban kenya. *Rural and Remote Health*, 12(2), 1833. <https://doi.org/10.22605/RRH1833>
- Eugênio, G. C. (2017). *Métodos de selecção de pontos de corte em análise de sobrevivência* [Trabalho de Licenciatura]. São Paulo, Brasil.
- Fink, G., Mrema, S., Abdulla, S., Kachur, S. P., Khatib, R., Lengeler, C., ... Schellenberg, J. (2022). Mosquito net use in early childhood and survival to adulthood in tanzania. *New England Journal of Medicine*, 386(5), 428–436. <https://doi.org/10.1056/NEJMoa2112524>
- Fonseca, S., Flores, P., Camargo, K. J., Pinheiro, R., & Coeli, C. (2017). Escolaridade e idade materna: desigualdades no óbito neonatal. *Revista de Saúde Pública*, 51, 94. <https://www.scielo.org/article/rsp/2017.v51/94/pt/> (Acesso em 29-11-2024)
- Freitas, L. A. C. (2022). *Teste de wald para avaliação de parâmetros de regressão e dispersão em modelos multivariados de covariância linear generalizada* (Unpublished master's thesis). Universidade Federal do Paraná, Curitiba, PR.
- Garcia, P. N. A. (2013). *Aplicação de técnica de análise de sobrevivência em pacientes submetidos a intervenção coronária percutânea* [Trabalho de Licenciatura].

- Girma, M., Eshete, H., Asrat, R., Gebremichael, M., Getahun, D., & Awoke, T. (2023). Socio-demographic and environmental determinants of under-five child mortality in ethiopia: Using ethiopian demographic and health 2019 survey. *BMC Pediatrics*, 23, 294. <https://doi.org/10.1186/s12887-023-04026-w>
- Gobebo, G. (2021). Determinant factors of under-five mortality in southern nations, nationalities and people's region (snnpr), ethiopia. *Italian Journal of Pediatrics*, 47, 147. <https://doi.org/10.1186/s13052-021-01118-0>
- Gorfine, M., Zucker, D. M., & Hsu, L. (2005). *Prospective survival analysis with a general semiparametric shared frailty model: a pseudo full likelihood approach*. <https://arxiv.org/abs/math/0505387v1>
- Hill, K., You, D., Inoue, M., & Oestergaard, M. Z. (2012). Child mortality estimation: accelerated progress in reducing global child mortality, 1990–2010. *PLoS Medicine*, 9(8), e1001303. <https://doi.org/10.1371/journal.pmed.1001303>
- Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data*. Wiley.
- INE, & MISAU. (2024). *Moçambique inquérito demográfico e de saúde 2022–23: Relatório definitivo*. Maputo, Moçambique; Rockville, Maryland, USA: The DHS Program, ICF.
- Jolly, M., Sebire, N., Harris, J., Robinson, S., & Regan, L. (2000). The risks associated with pregnancy in women aged 35 years or older. *Human Reproduction*, 15(11), 2433–2437. <https://doi.org/10.1093/humrep/15.11.2433>
- Justice, A., Honorati, M., Renay, W., Shalom, A., & Seth, O.-A. (2012). Impact of place of delivery on neonatal mortality in rural tanzania. *International Journal of Maternal and Child Health and AIDS*. <https://doi.org/10.21106/ijma.10>
- Kanmiki, E. W., Bawah, A. A., Agorinya, I., Achana, F. S., Awoonor-Williams, J. K., Oduro, A. R., et al. (2014). Determinantes socioeconômicos e demográficos da mortalidade de menores de cinco anos no norte rural de gana. *BMC International Health and Human Rights*, 14, 24. <https://doi.org/10.1186/1472-698X-14-24>
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481.
- Kazembe, L., Clarke, A., & Kandala, N.-B. (2012). Childhood mortality in sub-saharan africa: Cross-sectional insight into small-scale geographical inequalities from census data. *BMJ Open*, 2(5), e001421. [10.1136/bmjopen-2012-001421](https://doi.org/10.1136/bmjopen-2012-001421)
- Klein, P. J., & Moeschberger, L. M. (2005). *Survival analysis- statistics for biology and health*. New York: Springer-Verlag.

- Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer.
- Kozuki, E. e. W. N., N e Sonneveldt. (2013). Confusão residual explica a associação entre alta paridade e mortalidade infantil. *BMC Public Health*, 13(3), 1–8. <https://doi.org/10.1186/1471-2458-13-S3-S5>
- Kunnuji, M., Eshiet, I., Ahinkorah, B. O., Omogbemi, T., & Yaya, S. (2021). Background predictors of time to death in infancy: Evidence from a survival analysis of the 2018 nigeria dhs data. *BMC Public Health*, 21, 12424. <https://doi.org/10.1186/s12889-021-12424-x>
- Kyereme, K., & Amo-Adjei, J. (2016). Household wealth, residential status and the incidence of diarrhoea among children under five years in ghana. *Journal of Epidemiology and Global Health*, 6(1), 15–23. <https://doi.org/10.1016/j.jegh.2015.05.001>
- Lee, E., & Forthofer. (2006). *Analyzing complex survey data* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Licima, V. d. A. S. C. (1992). *Inferência paramétrica e não-paramétrica em análise de sobrevivência* (Dissertação de Mestrado). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Estado de São Paulo, Brasil.
- Lumley, T. (2010). *Complex surveys: A guide to analysis using r*. Seattle, Washington: John Wiley & Sons, Inc.
- Mahy, M. (2003). *Childhood mortality in the developing world: a review of evidence from the demographic and health surveys*. Calverton, Maryland: ORC Macro.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4). <https://doi.org/10.1093/jnci/22.4.719>
- Mdala, J., & Mash, R. (2015). Causes of mortality and associated modifiable health care factors for children (<5 years) admitted at onandjokwe hospital, namibia. *African Journal of Primary Health Care & Family Medicine*, 7(1), 1–6. <https://doi.org/10.4102/phcfm.v7i1.840>
- Minguillo, M. (2016). *Introdução aos modelos de fragilidade: Uma maneira de analisar dados correlacionados de sobrevivência* (Trabalho de Licenciatura). Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.
- Ministério da Saúde da República Democrática do Congo, & Institut National de la Statistique de la République Démocratique du Congo. (2013). *Demographic and health survey:*

- Democratic republic of congo*. <https://dhsprogram.com/pubs/pdf/FR296/FR296.pdf> (Acesso em 06-08-2024)
- Ministério da Saúde da Zâmbia, & Central Statistical Office of Zâmbia. (2018). *Demographic and health survey: Zâmbia*. <https://dhsprogram.com/pubs/pdf/FR364/FR364.pdf> (Acesso em 11-09-2024)
- Ministério da Saúde da África do Sul, & Instituto Nacional de Estatística da África do Sul. (2016). *Demographic and health survey: South africa*. <https://dhsprogram.com/pubs/pdf/FR337/FR337.pdf> (Acesso em 06-08-2024)
- Ministério da Saúde das Comores, & Institut National de la Statistique des Comores. (2012). *Demographic and health survey: Comoros*. <https://dhsprogram.com/pubs/pdf/FR267/FR267.pdf> (Acesso em 07-08-2024)
- Ministério da Saúde de Angola, & Instituto Nacional de Estatística de Angola. (2015). *Demographic and health survey: Angola*. <https://dhsprogram.com/pubs/pdf/FR318/FR318.pdf> (Acesso em 06-08-2024)
- Ministério da Saúde de Eswatini, & Central Statistical Office of Eswatini. (2014). *Demographic and health survey: Eswatini*. <https://dhsprogram.com/pubs/pdf/FR290/FR290.pdf> (Acesso em 12-08-2024)
- Ministério da Saúde de Lesoto, & Lesotho Bureau of Statistics. (2014). *Demographic and health survey: Lesotho*. <https://dhsprogram.com/pubs/pdf/FR294/FR294.pdf> (Acesso em 08-08-2024)
- Ministério da Saúde de Madagascar, & Instituto Nacional de Estatística de Madagascar. (2018). *Demographic and health survey: Madagascar*. <https://dhsprogram.com/pubs/pdf/FR362/FR362.pdf> (Acesso em 08-08-2024)
- Ministério da Saúde de Moçambique, & Instituto Nacional de Estatística de Moçambique. (2023). *Demographic and health survey: Mozambique*. <https://dhsprogram.com/pubs/pdf/FR320/FR320.pdf> (Acesso em 06-08-2024)
- Ministério da Saúde e Bem-Estar da Família de Zimbabwe, & Zimbabwe National Statistics Agency. (2015). *Demographic and health survey: Zimbabwe*. <https://dhsprogram.com/pubs/pdf/FR313/FR313.pdf> (Acesso em 15-08-2024)
- Mirza, B., Hunter, W., Kam, S., Elodie, B., & Hanne, D. (2021). Parental education and inequalities in child mortality: a global systematic review and meta-analysis. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(21\)00534-1](https://doi.org/10.1016/S0140-6736(21)00534-1)

- Muhuri. (1995). Health programs, maternal education, and differential child mortality in matlab, bangladesh. *Population and Development Review*, 21(4), 813–834. <https://doi.org/10.2307/2137775>
- Munda, M., Arlenda, R., & Catherine. (2017). *parfm: Parametric frailty models in r*. <https://cran.r-project.org/web/packages/parfm/parfm.pdf> (Acesso em 13-10-2024)
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4), 945–966. <https://doi.org/10.1080/00401706.1972.10488991>
- Nyamuranga, C. (2016). *The effect of public health expenditure on childhood mortality in southern africa (sadc)* [Diseertacao de Mestrado].
- Ogbo, F. A., Ezech, O. K., Awosemo, A. O., Ifegwu, I. K., Tan, L., Jessa, E., et al. (2019). Determinants of trends in neonatal, post-neonatal, infant, child and under-five mortality in tanzania from 2004 to 2016. *BMC Public Health*, 19(1243), 1–12. <https://doi.org/10.1186/s12889-019-7547-x>
- Olds, D. L., Kitzman, H., Knudtson, M. D., Anson, E., Smith, J. A., & Cole, R. (2014, September). Effect of home visiting by nurses on maternal and child mortality: Results of a 2-decade follow-up of a randomized clinical trial. *JAMA Pediatrics*, 168(9), 800–806. <https://doi.org/10.1001/jamapediatrics.2014.472>
- Oliveira, C. (2023). *Análise de sobrevivência e random survival forest na predição do tempo até a morte de câncer de mama em mulheres do estado de pernambuco* (Dissertação de Mestrado). Centro de Ciências e Tecnologia, Campina Grande, Brasil.
- Oliveira, E., Lima, C. O., Cirino, P., Vera, P., Lima, L. H. d. O., & Conde, W. L. (2020). Mortalidade neonatal: causas e factores associados. *Universidade Federal do Piauí*.
- Oliveira, S. C., & Almeida, R. M. V. (2019). Modelo de cox para análise da sobrevida das empresas de plano de saúde no brasil. *Journal Brasileiro de Economia em Saúde*, 11(2), 123–135. <https://doi.org/10.1234/jbes.v11i2.2019>
- OMS. (2013). *Goal 4: Reduce child mortality*. World Health Organization. http://www.who.int/topics/millennium_development_goals/child_mortality/en (Acesso em 12-09-2024)
- OMS. (2022). *Análise da situação sanitária da região africana da oms* [Relatório Técnico]. Genebra, Suíça. <https://www.who.int/africa/publications/situation-report-2022> (Acesso em 13-09-2024)
- Pereira, J. (2002). *Métodos de comparação de curvas de sobrevivência* (Dissertação de Mestrado). Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

- Pessoa, D. G. C., & Silva, P. L. N. (2018). *Análise de dados amostrais complexos*. São Paulo: Associação Brasileira de Estatística.
- Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2). <https://doi.org/10.2307/2344317>
- Pretorius, C. E., Asare, H., Genuneit, J., Kruger, H. S., & Ricci, C. (2020, August). Impact of breastfeeding on mortality in sub-Saharan Africa: a systematic review, meta-analysis, and cost-evaluation. *European Journal of Pediatrics*, 179(8), 1213–1225. <https://doi.org/10.1007/s00431-020-03721-5>
- R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. <https://www.R-project.org/> (Acesso em 11-11-2024)
- Rao, K., & Scott, A. J. (1987). On simple adjustments to chi-square tests for sample survey data. *The Annuals of Statistics*, 15(1), 385–397. <https://doi.org/10.1214/aos/1176350365>
- Remoaldo, P. C. A. (2002). *Desigualdades territoriais e sociais subjacentes à mortalidade infantil em portugal*. Lisboa: Fundação Calouste Gulbenkian, Fundação para a Ciência e Tecnologia.
- Ren, R. (2022). *Note on dhs standard weight de-normalization*. <https://userforum.dhsprogram.com>. (Acesso em 10-06-2025)
- Rhue, S. J., Torrico, G., Amuzie, C., Collins, S. M., Lemaitre, A., Workman, C. L., ... Stole, J. (2023). The effects of household water insecurity on child health and well-being. *Journal of Water, Sanitation and Hygiene for Development*. <https://doi.org/10.2166/washdev.2024.042>
- Rizopoulos, D. (2019). *Biostatistical methods ii: Classical regression models (ep03) survival analysis*. (Material de curso)
- Rocha, & Papoila. (2009). Análise de sobrevivência. In *Xvii congresso da sociedade portuguesa de estatística (spe)*. Lisboa.
- Rocha, J. (2015). *Fatores que influenciam na proporção de indivíduos imunes às reações hansênicas em pacientes diagnosticados com hanseníase no rn* (Dissertação de Mestrado). Universidade Federal do Rio Grande do Norte, Brasil.
- Rondeau, V., Mazroui, Y., & Gonzalez, J. R. (2012). frailtypack: An r package for the analysis of correlated survival data with frailty models using penalized likelihood

- estimation or parametrical estimation. *Journal of Statistical Software*, 47(1), 1–28. <http://www.jstatsoft.org/v47/i01/> (Acesso em 12-09-2024)
- Rutstein, S. (2005). Efeitos de intervalos de nascimentos precedentes na mortalidade neonatal, infantil e de menores de cinco anos e no estado nutricional em países em desenvolvimento: evidências de pesquisas demográficas e de saúde. *International Journal of Gynecology & Obstetrics.*, 89, S7–S24. <https://doi.org/10.1016/j.ijgo.2004.11.012>
- Sahu, & Nair. (2015). Níveis, tendências e preditores de mortalidade infantil e infantil entre tribos programadas na Índia rural. *Indian Journal of Medical Research*, 144(6), 801–810. <https://doi.org/10.1007/s13312-016-0950-9>
- Santos, L. C. S. J., & Gomes, H. V. S. F. (2023). Análise de sobrevivência aplicada ao tempo de permanência de segurados em planos funerários: Um estudo de caso. *Revista Gestão e Organizações*, 8(2).
- Sastry, N. (2004). Urbanização, desenvolvimento e diferenciais de mortalidade de menores de cinco anos por local de residência em são paulo, brasil, 1970-1991. *Demographic Research*, S2, 89–106. <https://doi.org/10.4054/DemRes.S2.14>
- Schober, P., & Vetter, T. R. (2024). Survival analysis and interpretation of time-to-event data: The tortoise and the hare. *Anesthesia & Analgesia*.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1).
- Shimakura, & Carvalho. (2005). *Análise de sobrevida: Teoria e aplicações em saúde*. Rio de Janeiro: Editora Fiocruz.
- Silva, A. S., Menezes, M. J. F., & Fortunato. (2004). Resíduos generalizados de cox-snell na avaliação do ajuste de modelos. *Ciência e Agrotecnologia*, 28(4), 781–790. <https://doi.org/10.1590/S1413-70542004000400016>
- Silva, D. (2016). *Introdução ao modelo de cox com aplicação a dados de pneus 11.00r22* [Trabalho de Licenciatura]. Campina Grande.
- Silva, G. (1992). *Modelo logístico para dados binários* (Dissertação de Mestrado). Instituto de Matemática e Estatística, USP.
- Silva, J., nd Luiza Zambalde de Castro, L., Thais, & Campos, M. (2015). Análise da confiabilidade: Um estudo de caso. In *Xxv encontro nacional de engenharia de produção*. Fortaleza, CE, Brasil.

- Silva, S. J. (2022). *Predição dos tempos até a morte de mulheres com câncer de mama via random survival forest*. Centro de Ciências e Tecnologia, Departamento de Estatística, Curso de Bacharelado em Estatística, Campina Grande, PB, Brasil.
- Sinha, B. (2016). Maternal age at childbirth and perinatal and under-five mortality in a prospective birth cohort from delhi. *Indian Pediatr*, 53(12), 1055–1061. [10.1007/s13312-016-0950-9](https://doi.org/10.1007/s13312-016-0950-9)
- Sousa, G. C. V. (2019). *Modelos de regressão em análise de sobrevivência: uma aplicação na modelação do tempo de vida de Micrurus corallinus em cativeiro* (Dissertação de Mestrado). Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, São Paulo, Brasil.
- Taela, K., Boggs, M., Jodrey, P., Nhassico, D., & Calado, D. (2020). *Liderança e redes de saúde infantil em moçambique desde 2000 até à data: Perspectivas do país* (Relatório do Estudo de Caso). USAID. https://pdf.usaid.gov/pdf_docs/PA00WPHJ.pdf (Acesso em 11-09-2024)
- Tekelab, T., Chojenta, C., Smith, R., & Loxton, D. (2019). The impact of antenatal care on neonatal mortality in sub-saharan africa: A systematic review and meta-analysis. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0222566>
- Tesema, G. A., Teshale, A. B., & Tessem, Z. T. (2021). Incidence and predictors of under-five mortality in east africa using multilevel weibull regression modeling. *Archives of Public Health*, 79, 196. <https://doi.org/10.1186/s13690-021-00727-9>
- Tesfa, D., Tiruneh, S. A., Azanaw, M. M., Gebremariam, A. D., Engdaw, M. T., Ke-fale, B., ... Dessalegn, T. (2021). Time to death and its determinants among under-five children in sub-saharan africa using the recent (2010–2018) demographic and health survey data: country-based shared frailty analyses. *BMC Pediatrics*. <https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-021-02950-3> (Acesso em 23-10-2024)
- Therneau, T. M. (2023). *Survival: Survival analysis, including penalized likelihood*. <https://cran.r-project.org/web/packages/survival/index.html>. (Acesso em 25-10-2024)
- Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the cox model*. New York: Springer-Verlag.
- Tiruneh, M., Demissie, M., & Tafere, T. (2024). *Multilevel analysis of healthcare utilization for childhood diarrhea in high under five mortality countries* | *Scientific Reports*. <https://link.springer.com/article/10.1038/s41598-024-65860-1>

- Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*. <https://projecteuclid.org/euclid.aoms/1177704711> (Acesso em 11-10-2024)
- UNICEF. (2024). *Doenças infantis*. <https://www.unicef.org/health/childhood-diseases> (Acesso em 10-10-2024)
- Wagstaff, A., & Watanabe, N. (1999). *Desigualdades socioeconômicas na desnutrição infantil no mundo em desenvolvimento*. Banco Mundial. <https://doi.org/10.1596/1813-9450-2434>
- Yalew, M., Arefaynie, M., Bitew, G., Amsalu, E. T., Kefale, B., Muche, A., ... Dewau, R. (2022). Time to under-five mortality and its predictors in rural Ethiopia: Cox-gamma shared frailty model. *PLoS One*.
- Yaya, S., Bishwajit, G., Okonofua, F., & Uthman, O. A. (2018). Under five mortality patterns and associated maternal risk factors in sub-Saharan Africa: A multi-country analysis. *PLoS One*, e0205977.
- Yeshaw, Y., Worku, M. G., Tessema, Z. T., Teshale, A. B., & Tesema, G. A. (2020). Zinc utilization and associated factors among under-five children with diarrhea in East Africa: A generalized linear mixed modeling. *PLoS One*.

APÊNDICE

Apêndice A - Gráficos

Curvas de Kaplan-Meier

Figura 5.1: *Curvas de Kaplan-Meier estimadas para as variáveis: Duração da amamentação (IV) e Número de consultas pré-natais (V), Ordem de nascimento (VI).*

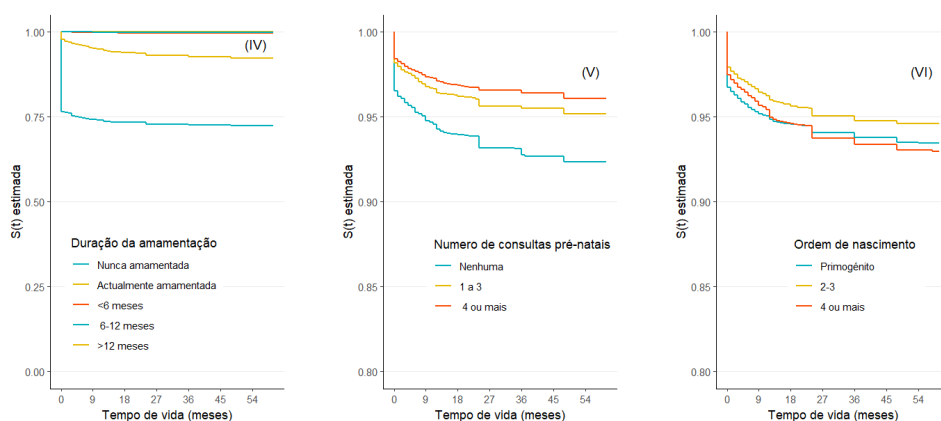


Figura 5.2: *Curvas de Kaplan-Meier estimadas para as variáveis: Tipo de de nascimento(VII), Intervalo entre nascimentos (VIII) e Gravidez indesejada (IX).*

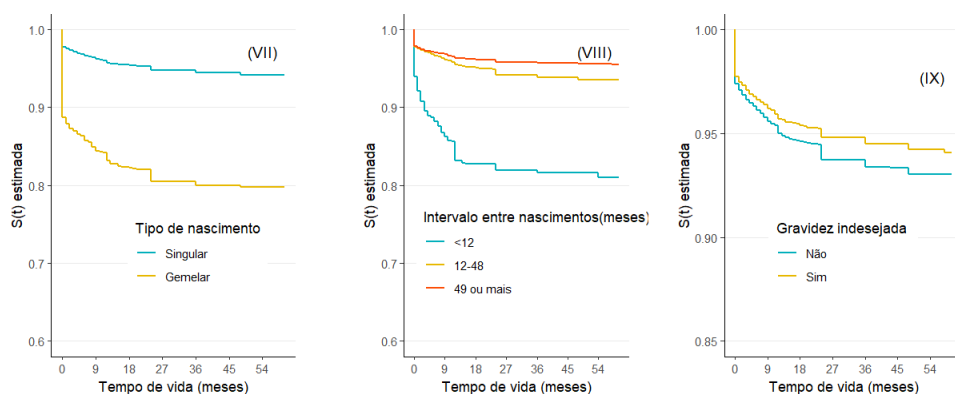


Figura 5.3: *Curvas de Kaplan-Meier estimadas para as variáveis: Índice de Riqueza(X) e Número de membros do agregado familiar (XI), Nível de escolaridade do marido (XII).*

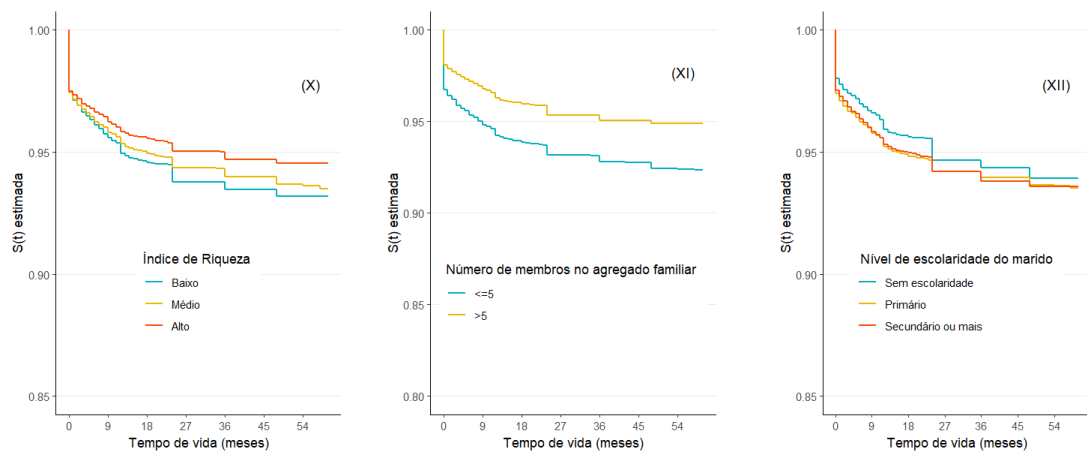


Figura 5.4: *Curvas de Kaplan-Meier estimadas para as variáveis: Fonte de água (XIII), Sanitário (XIV), Sexo do chefe do AF (XV).*

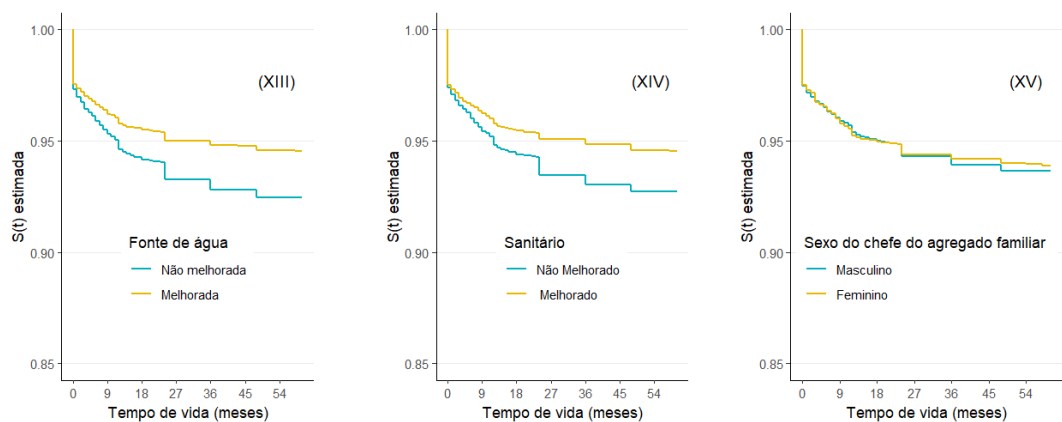


Figura 5.5: *Curvas de Kaplan-Meier estimadas para as variáveis: Uso de rede mosquiteira (XIX), Estado marital (XX), Ocupação da mãe (XXI), Nível de escolaridade da mãe (XXII).*

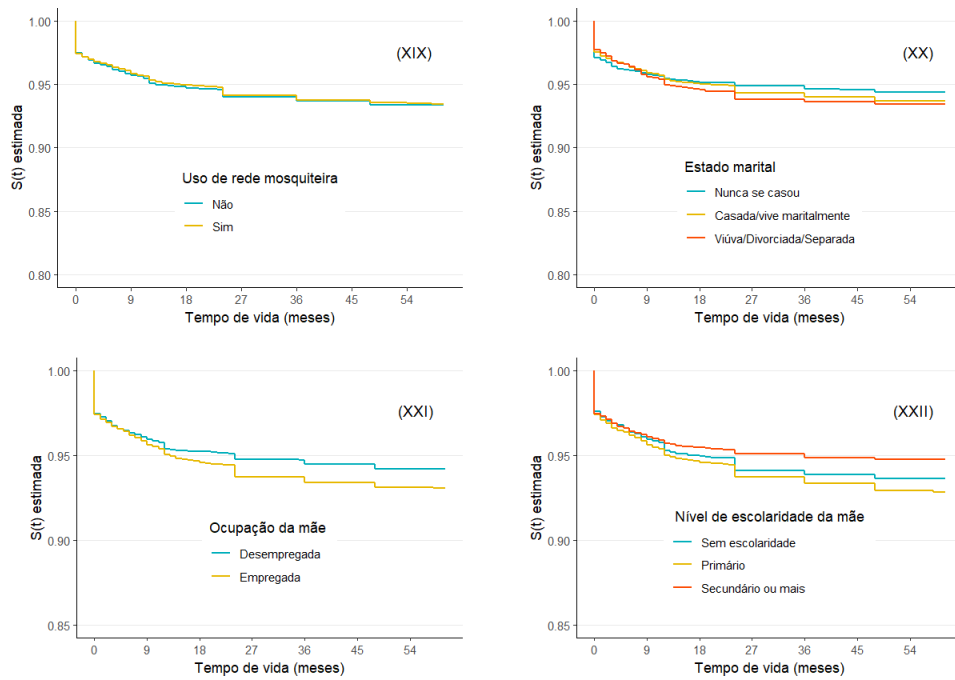
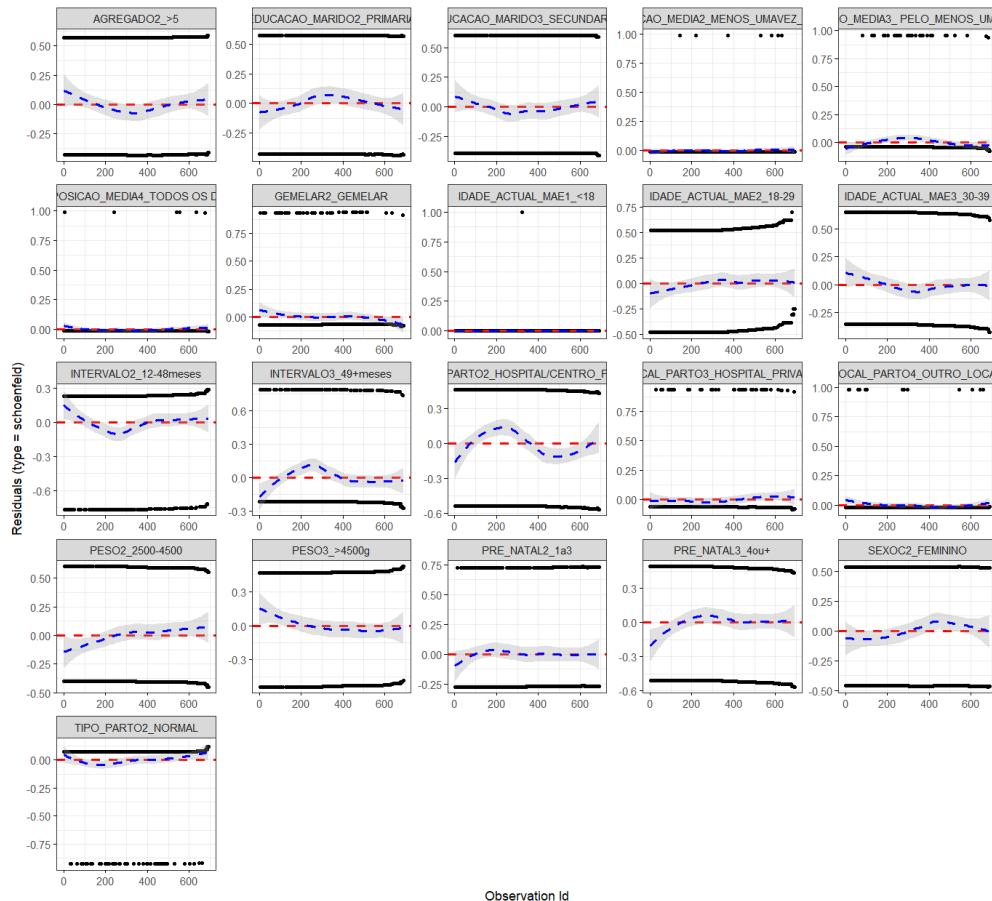


Figura 5.6: *Representação gráfica dos resíduos de Schoenfeld.*



Apêndice B- Tabelas auxiliares

Tabela 5.1: *Seleção de covariáveis para o modelo*

Passos	Modelo	Estatística	P-valor
Passo 1	Nulo		
	Sexo da criança	63,45	< 0,001
	País	57,29	0,99
	Local de residência	3,15	0,879
	Peso à nascença	1424,1	<0,001
	Número de consultas pré-natais	27,06	0,844
	Ordem de nascimento	70,01	< 0,001
	Tipo de nascimento	796,11	< 0,001
	Intervalo entre nascimentos	318,89	< 0,001
	Índice de riqueza	4,85	0,912
	Idade da mãe	56,44	< 0,001
	Nível de escolaridade do marido	17,02	0,001
	Sexo do chefe AF	0,071	0,753
	Gravidez indesejada	3,93	0,104
	Nível de escolaridade da mãe	25,05	0,001
	Número de membros do AF	312,65	< 0,001
	Estado marital da mãe	12,58	< 0,001
	Ocupação da mãe	15,51	0,104
	Tipo de parto	23,55	0,001
	Local do parto	8,56	0,998
	Fonte de água	36,77	0,302
	Sanitário	9,61	0,992
Passo 2	Uso de rede mosquiteira	6,08	0,001
	Exposição à mídia	38,06	<0,001
	Modelo completo passo 1		
	Sem Sexo da criança	9,04	< 0,001
	Sem Peso à nascença	124,92	< 0,001
	Sem Ordem de nascimento	42,61	0,265
	Sem Tipo de nascimento	31,17	0,99
	Sem Intervalo entre nascimentos	186,59	0,008
	Sem Idade da mãe	20,22	< 0,001
	Sem Nível de escolaridade do marido	13,14	0,176
	Sem Nível de escolaridade da mãe	4,32	0,188
	Sem Número de membros do AF	59,45	0,48
	Sem Estado marital	1,30	0,969
	Sem Tipo de parto	38,87	< 0,001
Passo 3	Sem Uso de rede mosquiteira	0,63	0,462
	Sem Exposição à mídia	19,26	< 0,001
	Modelo completo do passo 2		
	Com País	35,62	0,32
	Com Local de residência	1,23	0,284
	Com Número de consultas pré-natais	3,45	0,003
	Com Índice de riqueza	7,04	0,315
	Com Sexo do Chefe do AF	0,287	0,364
	Com Gravidez indesejada	0,978	0,192
	Com Ocupação da mãe	0,4	0,413
Passo 4	Com Local de parto	3,77	0,009
	Com Fonte de água	0,61	0,465
	Com Sanitário	0,914	0,456
	Modelo completo do passo 3		
	Com Ordem de nascimento	1,28	0,407
	Com Tipo de nascimento	19,4	0,009
	Com Nível de escolaridade do marido	24,15	0,076
	Com Nível de escolaridade da mãe	0,23	0,982
Passo 5	Com Número de membros do AF	65,41	< 0,001
	Com Estado marital	14,39	0,167
	Com Uso de rede mosquiteira	1,44	0,386
	Com Local de residência*Local de parto	4,32	0,273

ANEXOS



Jun 17, 2024

Shelsea Damião
Universidade Eduardo Mondlane
Mozambique
Request Date: 06/17/2024

Dear Shelsea Damião:

This is to confirm that you are approved to use the following Survey Datasets for your registered research paper titled: "Analysis of factors associated with under-five child mortality in Southern African":

Angola, Comoros, Congo Democratic Republic, Eswatini, Lesotho, Madagascar, Malawi, Mozambique, Namibia, South Africa, Tanzania, Zambia, Zimbabwe

To access the datasets, please login at: https://www.dhsprogram.com/data/dataset_admin/login_main.cfm. The user name is the registered email address, and the password is the one selected during registration.

The IRB-approved procedures for DHS public-use datasets do not in any way allow respondents, households, or sample communities to be identified. There are no names of individuals or household addresses in the data files. The geographic identifiers only go down to the regional level (where regions are typically very large geographical areas encompassing several states/provinces). Each enumeration area (Primary Sampling Unit) has a PSU number in the data file, but the PSU numbers do not have any labels to indicate their names or locations. In surveys that collect GIS coordinates in the field, the coordinates are only for the enumeration area (EA) as a whole, and not for individual households, and the measured coordinates are randomly displaced within a large geographic area so that specific enumeration areas cannot be identified.

The DHS Data may be used only for the purpose of statistical reporting and analysis, and only for your registered research. To use the data for another purpose, a new research project must be registered. All DHS data should be treated as confidential, and no effort should be made to identify any household or individual respondent interviewed in the survey. Also, be aware that re-distribution of any DHS micro-level data, either directly or within any tool/dashboard, is not permitted. Please reference the complete terms of use at: <https://dhsprogram.com/Data/terms-of-use.cfm>.

The data must not be passed on to other researchers without the written consent of DHS. However, if you have coresearchers registered in your account for this research paper, you are authorized to share the data with them. All data users are required to submit an electronic copy (pdf) of any reports/publications resulting from using the DHS data files to: references@dhsprogram.com.

Sincerely,

Bridgette Wellington

Bridgette Wellington
Data Archivist
The Demographic and Health Surveys (DHS) Program