



UNIVERSIDADE
EDUARDO
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Análise dos factores associados ao número de vítimas
envolvidas em acidentes rodoviários
na Cidade de Maputo, entre 2017 à 2020**

Autor: Isménia Florinda Wilson

Maputo, Junho de 2024



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Análise dos factores associados ao número de vítimas
envolvidas em acidentes rodoviários
na Cidade de Maputo, entre 2017 à 2020**

Autor: Isménia Florinda Wilson

Supervisor: Adelino José Chingore Juga, PhD, UHasselt

Maputo, Junho de 2024

Dedicatória

Dedico este trabalho à minha família, em especial aos meus pais
José Leonardo Wilson e Alzira Mucavele

Agradecimentos

Em primeiro lugar agradecer a **DEUS** pelo dom da vida, pela graça e misericórdia que tem depositado em minha vida.

Agradecer ao meu pai **José Leonardo Wilson** - em memória (Acredito que se estivesse vivo, pai, estarias orgulhoso da filha que criaste), agradecer pelos ensinamentos, pelos puxões de orelha e incentivo que depositou em mim e a minha mãe **Alzira Mucavele** que tem estado comigo e me apoiado em absolutamente tudo.

Agradecer também a **minha avó** (em memória), **aos meus irmãos e a toda minha família** que directa ou indirectamente me apoiou e incentivou a alcançar os objectivos traçados no início desta jornada.

Agradecer ao meu supervisor **Doutor Adelino Juga**, que me deu todo apoio, orientação e também críticas construtivas e valiosas para que eu pudesse concluir o meu Trabalho de Licenciatura.

Os meus agradecimentos também se estendem aos docentes do Departamento de Matemática e Informática pelo profissionalismo e pela qualidade de ensino transmitida durante a minha formação.

À todos os meus amigos, em especial, a **Dádiva Manhiça, Dionilza Martins, Fanásia Elias, Jorge Matavele Júnior, Karina Cosmo, Keila Truzão, Inês Tembe, Ornela Lumbela e ao Bonifácio Nicola**, pela amizade, companheirismo, ensinamentos e apoio incondicional durante o curso.

MUITO OBRIGADA A TODOS!

Declaração de Honra

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

Maputo, Junho de 2024

*“Consagre ao Senhor tudo o que você faz,
e os seus planos serão bem sucedidos”
Provérbios 16:3*

Resumo

Os acidentes rodoviários são uma das principais causas de óbitos e lesões graves. Além do impacto humano, os acidentes acarretam implicações econômicas consideráveis, decorrentes de despesas médicas, perda de produtividade e danos materiais. Para este Trabalho, foram usados dados dos acidentes rodoviários ocorridos na Cidade de Maputo entre 2017 à 2020, fornecidos pelo Comando da Polícia da República de Moçambique. O objetivo principal do Trabalho é analisar os factores associados ao número de vítimas envolvidas em acidentes rodoviários na Cidade de Maputo entre 2017 à 2020. Os modelos de regressão de Poisson e Binomial Negativa, foram usados para investigar os factores associados ao número de vítimas em acidentes rodoviários. Os principais fatores associados ao número de vítimas em acidentes rodoviários são causados por: veículos pesados, acidentes ocorridos nos finais de semana, especialmente sextas-feiras e sábados, acidentes ocorridos entre 0-6 horas, a má travessia de peões, excesso de velocidade e consumo de álcool, o choque entre carros e choque entre carros e motos.

Palavras-chave: Acidente rodoviário, Número de vítimas, Regressão Binomial Negativa, Regressão de Poisson

Lista de Abreviaturas

AIC	Critério de Informação de Akaike
BIC	Critério de Informação Bayesiana
INE	Instituto Nacional de Estatística
INAV	Instituto Nacional de Viação
MLG	Modelos Linear Generalizados
OMS	Organização Mundial de Saúde
PRM	Polícia da República de Moçambique
PASR	Programa de Acção para a Segurança Rodoviária
WHO	World Health Organization

ÍNDICE

1	Introdução	1
1.1	Contextualização	1
1.1.1	Formulação do Problema	2
1.2	Objectivos	2
1.3	Justificativa	2
1.4	Estrutura do Trabalho	4
2	Revisão de Literatura	5
2.1	Principais Conceitos	5
2.2	Acidentes Rodoviários no Mundo	6
2.3	Acidentes Rodoviários em Moçambique	7
2.4	Causas associadas aos acidentes rodoviários	7
2.5	Modelos Lineares Generalizados	10
2.5.1	Família Exponencial	11
2.5.2	Características do Modelo Linear Generalizado	12
2.6	Seleção de Modelos	13
2.6.1	Critério de Informação	14
2.7	Modelos de Regressão para Dados de Contagem	15
2.7.1	Modelo de Regressão Poisson	16
2.7.2	Estimação dos Parâmetros pelo Método de Maxima Verossimilhança	17
2.7.3	Modelo de Regressão Binomial Negativa	18
2.7.4	Estimação dos Parâmetros	19
3	Material e Métodos	20
3.1	Classificação da Pesquisa	20
3.1.1	Natureza da Pesquisa	20
3.2	Fonte e Descrição dos Dados	21
3.2.1	Softwares usados	23
3.3	Métodos	23
3.3.1	Teste de Independência	23
3.3.2	Modelo de Regressão Poisson	24

3.3.3	Modelo de Regressão Binomial Negativa	26
3.3.4	Teste de Significância dos Parâmetros	26
3.3.5	Multicolinearidade	28
4	Resultados e Discussão	30
4.1	Análise Exploratória dos Dados	30
4.1.1	Distribuição dos condutores por sexo e faixa etária	30
4.1.2	Distribuição dos acidentes por tipo de veículo	31
4.1.3	Distribuição dos acidentes por dia da semana e hora de ocorrência	32
4.1.4	Distribuição dos tipos de acidentes	34
4.2	Comparação dos Modelos de Poisson e Binomial Negativa	35
4.3	Modelo de Regressão Binomial Negativa	35
4.3.1	Teste de Ajustamento a Distribuição Binomial Negativa	38
4.4	Multicolinearidade	38
4.4.1	Verificação da Multicolinearidade	38
4.5	Discussão dos Resultados	39
5	Conclusões e Recomendações	41
5.1	Conclusões	41
5.2	Recomendações	42
5.3	Limitações	42
	Referências	44

Lista de Figuras

2.1	Potenciais fatores de risco associados aos acidentes rodoviários	9
4.1	Distribuição dos condutores por sexo	30
4.2	Distribuição dos condutores por faixas etárias	31
4.3	Distribuição dos acidentes por tipo de veículo	32
4.4	Distribuição dos acidentes por dia de ocorrência	33
4.5	Distribuição dos acidentes por hora de ocorrência	33
4.6	Distribuição dos tipos de acidentes	34
4.7	Distribuição do número de vítimas observado vs. Previsto	38

Lista de tabelas

2.1	As 10 principais causas de doenças a nível mundial em 1990	7
2.2	Acidentes registrados pela polícia em todo o país entre 2015-2020	8
2.3	Funções de Ligação canônica por tipo de distribuição	13
3.1	Descrição das variáveis usadas neste Trabalho	22
4.1	Comparação dos modelos estimados incluindo o log-likelihood negativo duplo, o número de parâmetros, o AIC e o BIC	35
4.2	Parâmetros e erros padrão estimados, com os respectivos p-valores das análises de regressão simples e múltipla	37
5.1	Valores do fator de inflação da variância (VIF)	48

Capítulo 1

Introdução

1.1 Contextualização

Os acidentes rodoviários representam um problema de saúde pública em todo o mundo. De acordo com a World Health Organization (WHO, 2004), morrem cerca de 1,2 milhões de pessoas por ano vítimas de acidentes rodoviários e cerca de 50 milhões sofrem ferimentos graves ou ligeiros resultando assim em incapacidade temporária ou permanente das vítimas.

Ainda de acordo com a World Health Organization (WHO, 2009), em 2030 os acidentes rodoviários estarão entre as cinco principais causas de morte no mundo. Em 2008, a taxa de mortalidade nos países desenvolvidos foi de 10,3 mortes por cada 100,000 habitantes, enquanto que nos países subdesenvolvidos e em via de desenvolvimento a taxa de mortalidade foi de 21,5 e 19,5 por cada 100,000 habitantes respectivamente.

Segundo a World Health Organization (WHO, 2015) a taxa de mortalidade média a nível mundial, no ano de 2010, associada aos acidentes rodoviários, foi de aproximadamente 18 mortes por cada 100,000 habitantes, sendo 8,7 mortes por 100,000 habitantes nos países desenvolvidos e 18,3 e 20,1 por 100,000 habitantes nos países subdesenvolvidos e em via de desenvolvimento.

As consequências dos acidentes rodoviários não são só humanas mas também económicas resultando em elevados custos para a sociedade e para a economia mundial.

Ainda segundo a OMS (2015), o continente africano tem a mais elevada taxa de mortalidade por acidentes rodoviários com cerca de 26,6 por 100,000 habitantes. Em 2013 no continente africano morreram cerca de 246,718 pessoas vítimas de acidentes rodoviários, o que representa cerca de 675 pessoas mortas todos os dias nas estradas.

OMS (2015), Moçambique está entre os 25 países a nível mundial com altos índices de mortes por acidentes e entre os cinco países da Comunidade de Desenvolvimento da África Austral (SADC) com a mais elevada taxa de mortalidade nas estradas com 68%, seguida da Zâmbia

com 50%, Malawi com 45%, África do Sul com 39% e por último Zimbábwe com 26%.

Portanto, neste trabalho pretende-se analisar os factores associados à ocorrência de vítimas em acidentes rodoviários na Cidade de Maputo no período de 2017 à 2020.

1.1.1 Formulação do Problema

Segundo o Instituto Nacional de Viação (INAV, 2009) o número de vítimas que morrem nas estradas moçambicanas é de tal modo elevado que os acidentes rodoviários são entre as principais causas de morte em Moçambique, a par da malária ou de doenças associadas ao HIV/SIDA. Em 2020 foram registados 395 acidentes rodoviários na Cidade de Maputo. No primeiro semestre do ano 2019, foram registados 769 mortos e 1473 feridos graves, um acréscimo na ordem de 15% em relação a igual período do ano de 2018. O álcool, o estado deficiente das vias públicas, circulação de viaturas com deficiências mecânicas e/ ou o desrespeito pelas regras básicas do código influenciam a ocorrência dos acidentes rodoviários em Moçambique. Assim, este estudo pretende analisar os factores associados ao número de vítimas envolvidas em acidentes rodoviários na Cidade de Maputo entre 2017 à 2020.

1.2 Objectivos

Objectivo Geral

Analisar os factores associados ao número de vítimas envolvidas em acidentes rodoviários na Cidade de Maputo entre 2017 à 2020

Objectivos Específicos

- Descrever o perfil sócio-demográfico dos condutores envolvidos nos acidentes rodoviários;
- Descrever o perfil das viaturas envolvidas nos acidentes rodoviários ocorridos;
- Verificar a associação entre a variável número de vítimas e as demais variáveis do estudo;
- Identificar as variáveis explicativas associadas ao número de vítimas dos acidentes rodoviários.

1.3 Justificativa

Os acidentes rodoviários resultam em morte, traumas físicos e psicológicos nas vítimas, que convivem com sequelas para o resto de suas vidas. As famílias enlutadas e a economia

sofrem prejuízos devido aos danos humanos e materiais causados pelos acidentes. Além das consequências diretas mencionadas, os acidentes rodoviários têm um impacto profundo em diversos aspectos da sociedade. Os sistemas de saúde são sobrecarregados com o aumento na demanda por serviços médicos, desde atendimentos de emergência até reabilitação de longo prazo. Isso gera um custo significativo para os governos e contribuintes. Além disso, as vítimas de acidentes muitas vezes perdem a capacidade de trabalhar, seja temporariamente ou permanentemente, resultando em perda de renda e produtividade, tanto para os indivíduos quanto para a economia em geral.

Do ponto de vista psicológico, os sobreviventes de acidentes e suas famílias podem enfrentar transtornos como o estresse pós-traumático (TEPT), depressão e ansiedade, que exigem acompanhamento psicológico e podem afetar a qualidade de vida a longo prazo. As crianças que perdem pais ou responsáveis em acidentes rodoviários podem sofrer com a desestruturação familiar e as dificuldades econômicas subsequentes, impactando seu desenvolvimento e oportunidades futuras.

Em termos sociais, a perda de vidas jovens e economicamente ativas em acidentes rodoviários pode desestabilizar comunidades inteiras, reduzindo a força de trabalho e afetando negativamente o crescimento econômico. Além disso, a percepção de insegurança nas estradas pode afetar o comportamento dos cidadãos, levando a uma redução na mobilidade e na utilização de certos modos de transporte.

No contexto de Maputo, esses impactos são particularmente relevantes devido ao crescimento urbano acelerado e ao aumento no número de veículos nas estradas. A análise dos fatores associados ao número de vítimas de acidentes rodoviários na Cidade de Maputo, no período de 2017 a 2020, é essencial para entender melhor essas dinâmicas e desenvolver estratégias eficazes de prevenção e mitigação.

A razão para se aprofundar no estudo desse tema é múltipla. Primeiramente, ao identificar e compreender os fatores que contribuem para os acidentes rodoviários, é possível implementar políticas públicas e campanhas de conscientização mais eficazes, visando a redução do número de acidentes e suas consequências. Além disso, melhorar a segurança rodoviária é fundamental para o desenvolvimento social, político e econômico da sociedade moçambicana e do mundo em geral, pois promove a saúde pública, a estabilidade econômica e a coesão social.

Assim sendo, torna-se crucial analisar este tema não apenas pelos danos humanos e materiais envolvidos, mas também pelo potencial de transformação positiva que uma abordagem informada e proativa pode trazer para a sociedade

1.4 Estrutura do Trabalho

O presente Trabalho está organizado da seguinte forma:

1. **Capítulo 1: Introdução** faz-se a contextualização dos acidentes rodoviários, definição do problema, objectivos e relevância do estudo ;
2. **Capítulo 2: Revisão de Literatura** é apresentada uma breve revisão sobre os acidentes rodoviários no mundo e em Moçambique em particular, bem como sobre as técnicas estatísticas que serão usadas.
3. **Capítulo 3: Material e Métodos** é apresentado o material usado para a realização do trabalho e os métodos usados para obter os resultados.
4. **Capítulo 4: (Resultados e Discussão)**, são apresentados os resultados obtidos através dos métodos descritos no capítulo 3 de modo a alcançar com os objectivos descritos na secção 1.3.
5. **Capítulo 5: Conclusões e Recomendações são apresnetadas** confere os resultados alcançados, bem como as limitações encontradas no decorrer do trabalho.

Capítulo 2

Revisão de Literatura

Neste capítulo são apresentados os conceitos ligados ao tema do trabalho que foram desenvolvidos por diversos autores e também conceitos teóricos das principais técnicas estatísticas usadas.

2.1 Principais Conceitos

Os principais conceitos deste estudo são definidos segundo o Instituto Nacional de Estatística (INE, 2020).

1. **Acidente:** é um acontecimento inesperado e indesejável, que pode resultar em vítimas humanas, materiais ou financeiros, e pode ocorrer no meio rodoviário, marítimo, ferroviário ou aéreo.
2. **Acidente rodoviário:** é um acontecimento fortuito, súbito e anormal ocorrido na via pública em consequência da circulação rodoviária, de que resultem vítimas ou danos materiais, quer o veículo se encontre ou não em movimento.
3. **Ferido grave:** é a vítima de acidente cujos danos corporais obriguem a um período de hospitalização superior a 24 horas e que não venha a falecer nos 30 dias após o acidente.
4. **Ferido ligeiro:** é a vítima de acidente que não seja considerada ferido grave e que não venha a falecer nos 30 dias após o acidente.
5. **Vítima:** ser humano que, em consequência do acidente, sofra danos corporais.

6. **Óbito/Vítima mortal (no local):** vítima cujo óbito ocorra no local do acidente ou durante o percurso até à unidade de saúde.

7. **Condutor:** é o indivíduo que detém o comando de um veículo ou animal na via pública.

2.2 Acidentes Rodoviários no Mundo

Em muitos países desenvolvidos o número de vítimas em acidentes rodoviários tende a diminuir, devido à aplicação de medidas adequadas para melhorar a segurança rodoviária. A Organização para a Cooperação e Desenvolvimento Económico (OCDE, 2008), que inclui a maioria dos países desenvolvidos, estabeleceu um Sistema de Seguranças com metas desafiantes para seus membros em direção à filosofia a longo prazo, denominada visão zero, cujo objetivo é reduzir o número de mortos e ferimentos graves a zero. Nestes países, a fatalidade foi reduzida em 50%, desde 1970, apesar do aumento do número de veículos motorizados.

Segundo o Plano Nacional de Prevenção Rodoviária (PNPR, 2010), em 2009, mais de 35,000 pessoas morreram nas estradas da União Europeia, o equivalente à população de uma cidade de média dimensão. Ainda no mesmo PASR, refere-se que por cada morte nas estradas europeias estima-se que haja 4 acidentados com incapacidades permanentes, tais como lesões cerebrais ou na coluna vertebral, 10 feridos graves e 40 feridos ligeiros.

O relatório “Global burden of disease”, elaborado pela OMS em 2004, reportou os principais problemas de saúde mundiais e seus impactos sobre as populações, referindo que os acidentes rodoviários representam a perda de 2,5% de anos de vida ajustados por incapacidade, e classificando-os como a nona causa de morbilidade e mortalidade a nível mundial, ultrapassando doenças como a malária e a tuberculose.

Um estudo realizado pela Universidade de Harvard previu que em 2020, os traumatismos rodoviários ocupariam o 3º lugar da tabela, referente às dez principais causas de morte por doença a nível mundial, tal como mostra a Tabela 2.1.

Tabela 2.1: As 10 principais causas de doenças a nível mundial em 1990

Doença ou Traumatismo	
1	Infecções do sistema respiratório Doença isquémia do coração
2	Doenças diarréicas Depressão unipolar major
3	Problemas perinatais Traumatismos rodoviários
4	Doença unipolar major Doença isquémia do coração
5	Doença isquémia do coração Doença pulmonar obstrutiva crónica
6	Doença cerebrovascular Infecções do sistema respiratório
7	Tuberculose Tuberculose
8	Rubéola Guerra
9	Traumatismos rodoviários Doenças diarréicas
10	Anomalias congénitas SIDA

Fonte: OMS (2004)

2.3 Acidentes Rodoviários em Moçambique

No caso particular de Moçambique, os acidentes rodoviários foram das principais causas de solicitação dos serviços de urgências do Hospital Central em Maputo em 2009, só nos fins-de-semana do Natal e de transição de ano. Segundo estes serviços cerca de 120 moçambicanos vêm a falecer vítimas de acidentes rodoviários, em consequência de 117 acidentes de viação, a maioria provocado pelo excesso de velocidade, e dos quais resultaram ainda 195 feridos (100 graves e 95 ligeiros).

O crescimento acelerado do parque automóvel no país, pode estar a influenciar no elevado número de acidentes rodoviários. INE (2018). Segundo a Tabela 2.2, todas as províncias da região Sul, Sofala e Nampula registaram maior frequência de acidente Cidade de Maputo foi a província que mais acidentes registou.

2.4 Causas associadas aos acidentes rodoviários

A ocorrência de um acidente rodoviário não está associada apenas a uma causa específica e única, ou seja, um acidente rodoviário nunca ocorre só porque o condutor pratica velocidades excessivas ou conduz embriagado.

Tabela 2.2: Acidentes registrados pela polícia em todo o país entre 2015-2020

Província	2015	2016	2017	2018	2019	2020
Total	2 848	2 212	1 992	1 570	1 232	1 038
Niassa	80	65	44	54	35	84
Cabo Delgado	69	65	28	26	61	36
Nampula	283	246	170	162	121	119
Zambézia	97	116	65	81	71	86
Tete	83	64	65	28	18	6
Manica	190	179	131	80	96	54
Sofala	215	216	165	142	93	94
Inhambane	235	151	153	105	152	100
Gaza	162	189	178	200	151	166
Maputo Província	748	376	598	210	132	120
Maputo Cidade	686	545	395	482	302	173

Fonte: INE 2015-2020

Para Vasconcellos (2010), os factores que influenciam o envolvimento num acidente rodoviário são amplamente discutidos e de conhecimento público. O excesso de velocidade, condução sob efeito de álcool ou drogas, a fadiga, o uso de telemóvel e o não cumprimento das regras de trânsito.

Portanto, de acordo com OLIVEIRA (2006), há três factores associados a ocorrência de acidentes rodoviários como representa a figura 2.1.

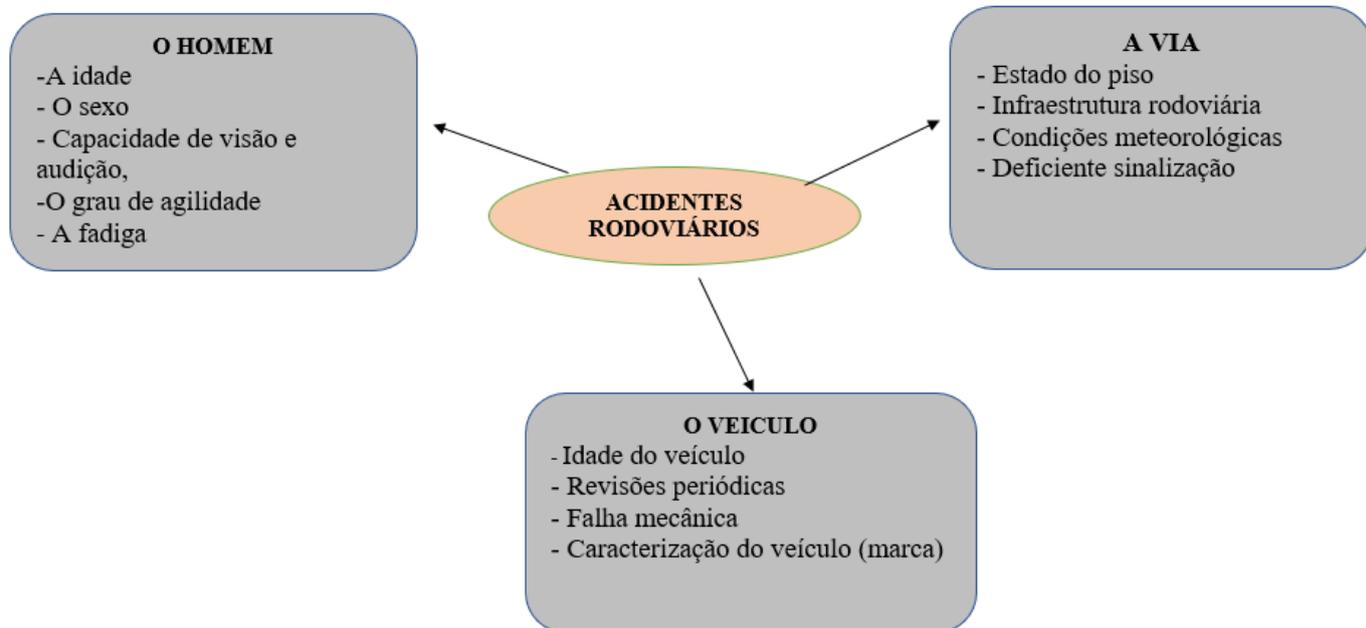


Figura 2.1: Potenciais fatores de risco associados aos acidentes rodoviários

Para Moçambique segundo o balanço feito da quadra festiva 2009-2010 no dia 18 de Janeiro de 2010 que envolveu as instituições responsáveis pela segurança rodoviária nomeadamente Ministério do Interior, Ministério dos Transportes e comunicações e Ministérios das obras públicas e Habitação consideram principais causas de acidentes

- Má travessia de pões;
- Excesso de velocidade;
- Álcool;
- Ultrapassagem irregular;
- Trânsito fora da mão;
- Mau estado técnico de alguns veículos;
- Transporte de passageiros em veículos não apropriado.

Dados publicados pelo INE (2018) indica que o atropelamento está associado a maiores chances de ocorrência tanto de óbitos como de feridos graves em acidentes rodoviários em todo o país, mas com maior enfoque para a capital do país, a Cidade de Maputo.

A condução sob efeito de álcool e a velocidade excessiva são factores com elevados prejuízos em termos de sinistralidade rodoviária, uma vez que influencia não só no risco de sofrer um acidente, como também aumenta a sua gravidade e a probabilidade de ocorrência de vítimas mortais.

2.5 Modelos Lineares Generalizados

Os modelos lineares generalizados (MLGs) foram propostos por Nelder e Weddeburn (1972) como uma extensão dos modelos lineares, onde a distribuição da variável resposta não tem de ser normal, mas sim uma outra distribuição da família exponencial.

O modelo linear normal, criado no início do século XIX por Legendre e Gauss, dominou a modelação estatística até meados do século XX, embora vários modelos não lineares ou não normais tenham entretanto sido desenvolvidos para fazer face a situações que não eram adequadamente explicadas pelo modelo linear normal.

São pois casos particulares dos modelos lineares generalizados, doravante referido como MLG, os seguintes modelos:

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

Segundo Cordeiro & Demétrio (2010), o processo de ajuste dos MLGs pode ser dividido em três etapas:

- i** Formulação dos modelos ;
- ii** Ajuste dos modelos;
- iii** Inferência.

Os MLGs formam uma ferramenta de grande utilidade prática, pois apresentam grande flexibilidade na etapa (i), computação simples em (ii) e critérios razoáveis em (iii). Essas etapas são realizadas se maneira sequencial.

Na análise de dados complexos, após a conclusão da etapa de inferência, pode-se voltar à etapa (i) e escolher outros modelos, a partir de informações mais detalhadas oriundas do estudo feito em (iii).

Uma característica importante dos MLGs é que se supõe independência das variáveis resposta (ou pelo menos, não-correlação) e, portanto, dados exibindo auto-regressões como as séries temporais, em princípio, podem ser excluídos. Uma segunda característica é que a estrutura da variável resposta é supostamente única embora, usualmente, existam várias variáveis explanatórias na estrutura linear desses modelos.

2.5.1 Família Exponencial

Esta classe de distribuições foi proposta independentemente por Koopman, Pitman e Darmois através do estudo de propriedades de suficiência estatística. Posteriormente, muitos outros aspectos dessa família foram estudados e tornaram-se importantes na teoria moderna de Estatística.

O conceito de família exponencial foi introduzido na Estatística por Fisher mas os modelos da família exponencial apareceram na Mecânica Estatística no final do século XIX e foram desenvolvidos por Maxwell, Boltzmann e Gibbs.

Muitas das distribuições conhecidas podem ser reunidas em uma família denominada família exponencial de distribuições. Assim, por exemplo, pertencem a essa família as distribuições normal, binomial, binomial negativa, gama, Poisson, normal inversa, multinomial, beta, logarítmica, entre outras.

A importância da família exponencial de distribuições teve maior destaque, na área dos modelos de regressão, a partir do trabalho pioneiro de Nelder e Wedderburn(1972) que definiram os modelos lineares generalizados. Na década de 80, esses modelos popularizaram-se, inicialmente, no Reino Unido, e, posteriormente, nos Estados Unidos e na Europa.

Segundo Lima (2018) no âmbito dos MLGs, pressupõe-se que a variável resposta tem uma distribuição pertencente á família Exponencial, ou seja, que a sua função densidade de probabilidade (ou função massa de probabilidade) pode ser escrita na forma geral:

$$f(y|\theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \quad (2.1)$$

Onde y é a variável de interesse, θ é a forma canônica do parâmetro de localização (em geral conhecido), ϕ é o parâmetro de dispersão associado a variância, $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções reais conhecidas e também denominadas funções específicas.

Portanto, a média e a variância de uma v.a Y cuja distribuição pertence à família exponencial, na forma canônica usada por McCullagh & Nelder (1989), são dadas por:

$$\mu = E(Y) = b(\theta)\sigma^2 = Var(Y) = a(\phi)b''(\theta) \quad (2.2)$$

em que $a(\phi)$, em geral, pode ser escrita na forma $a(\phi) = \phi/w$, sendo ϕ chamado parâmetro de dispersão e w , peso *a priori*. Além disso, $b''(\theta) = d\mu/d\theta$ é uma função de μ e é representada por $V(\mu)$. Logo, a variância de Y pode ser escrita como:

$$Var(Y) = a(\phi)b''(\theta) = a(\phi)V(\mu) \quad (2.3)$$

2.5.2 Características do Modelo Linear Generalizado

O Modelo linear Generalizado segundo Alvarenga (2015) é caracterizado pela seguinte forma:

1. **Componente aleatória** Dado o vector de covariáveis X_i as variáveis aleatórias Y_i são condicionalmente independentes com distribuição pertencente à família Exponencial, e portanto o seu valor médio é dado por:

$$E(Y_i|X_i) = \mu_i = b''(\theta_i), i = 1, \dots, n \quad (2.4)$$

A variável resposta do modelo, tem uma distribuição pertencente à família de distribuições que engloba as distribuições normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens.

2. **O componente sistemática** é estabelecido durante o planejamento (fundamental para a obtenção de conclusões confiáveis) do experimento, resultando em modelos de regressão (linear simples, múltipla, etc.), de análise de variância (delineamentos inteiramente casualizados, casualizados em blocos, quadrados latinos com estrutura de tratamentos fatorial, parcelas subdivididas, etc.) e de análise de covariância.

Define-se o preditor linear n_i como combinação linear das variáveis explicativas, representado pela expressão dada por:

Tabela 2.3: Funções de Ligação canônica por tipo de distribuição

Distribuição	Ligação canônica
Normal	Identidade: $n = \mu$
Poisson	Logarítmica: $n = \ln(\mu)$
Binomial	Logística: $n = \ln\left(\frac{\phi}{1-\phi}\right) = \ln\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $n = \frac{1}{\mu}$
Normal Inversa	Recíproca ² : $\frac{1}{\mu^2}$

$$n_i = X_i^T \beta, \text{ ou, } n = X\beta \tag{2.5}$$

$X = (x_1, \dots, x_n)$ representa a matriz do modelo, $\beta = (\beta_1, \dots, \beta_r)$ o vector de parâmetros desconhecidos e n representa o vector de preditores lineares.

3. **Função de ligação** Relaciona uma ligação entre a componentes aleatório e sistemático relacionando a média ao preditor linear e é feita por meio de uma função adequada como, por exemplo, logarítmica para os modelos log-lineares.

$$n_i = g(.) \tag{2.6}$$

onde $g(.)$ é função real, monótona e diferencial.

A estatística do teste é definida por:

2.6 Seleção de Modelos

Em modelos de regressão, é necessário determinar um subconjunto de variáveis explicativas que melhor explique a variável resposta. Em outras palavras, entre todas as variáveis explicativas disponíveis, é crucial encontrar um subconjunto de variáveis importantes para o modelo. Qualquer procedimento para seleção ou exclusão de variáveis de um modelo é baseado em um algoritmo que avalia a importância das variáveis, incluindo ou excluindo-as do modelo com base em uma regra de decisão. A importância da variável é definida em termos de uma medida de significância estatística do coeficiente associado à variável para o modelo. Existem três procedimentos automáticos para selecionar modelos:

- Método Forward
- Método Stepwise
- Método Backward

Método Forward

Este procedimento considera o modelo inicial contendo apenas a constante. A ideia do método é adicionar uma variável de cada vez. A primeira variável selecionada é aquela com a maior correlação com a variável resposta.

Método Backward

Enquanto no método Forward o modelo inicial não possui nenhuma variável e adiciona variáveis a cada passo, o método Backward segue o caminho oposto. Ele incorpora inicialmente todas as variáveis e, em etapas subsequentes, retira sucessivamente as variáveis não significativas do modelo.

Método Stepwise

O método de Stepwise é uma combinação dos dois métodos, Forward e Backward.

2.6.1 Critério de Informação

Os critérios de informação são frequentemente utilizados para orientar a selecção dos modelos e objetivam prover uma medida de informação que equilibre uma boa medida de ajustamento e uma especificação reduzida do modelo (parcimonioso quanto ao número de parâmetros). Dada sua natureza de penalidade, os melhores modelos são aqueles que apresentam menores critérios de informação (Araújo, 2017).

Segundo Hosmer e Lemeshow (1989), o critério de informação de *Akaike*, conhecido simplesmente como AIC, e o critério de informação bayesiano, chamado de BIC, não são testes de hipótese, eles são ferramentas para selecção de modelos. O AIC é um índice de ajuste geral que pressupõe que são comparados modelos que são adequados ao mesmo conjunto de dados, usando o mesmo método de estimativa.

Os critérios de informação apresentam duas vantagens sobre outros métodos de selecção. O AIC e o BIC se baseiam na log-verosimilhança como o outro método e, além disso, consideram e penalizam a complexidade do modelo (número de parâmetros), o que busca o alinhamento

com o princípio da parcimônia, essencial na modelagem estatística. Além disso, os modelos que estão sendo comparados não necessitam ser aninhados, o que é a principal restrição na comparação de modelos via teste de hipótese (Araújo, 2017).

Nas equações subsequentes, são apresentados os critérios AIC e BIC, que são expressos respectivamente, por:

$$AIC = \frac{(-2l)}{n} + \frac{(-2k)}{n} \quad (2.7)$$

$$BIC = \frac{(-2l)}{n} + \frac{(k \log n)}{n} \quad (2.8)$$

Onde: k é o número de restrições; n é o tamanho da amostra e l é a função de verossimilhança.

2.7 Modelos de Regressão para Dados de Contagem

De acordo com Ramalho (1996), os modelos de regressão Poisson e binomial negativo fazem parte do que é conhecido por modelos de regressão para dados de contagem, e têm por objetivo analisar o comportamento, em relação a variáveis preditoras, de determinada variável dependente que se apresenta na forma quantitativa, porém com valores discretos e não negativos (dados de contagem).

Ainda segundo Ramalho (1996), é comum, quando estamos trabalhando com dados de contagem, iniciarmos a estimação dos parâmetros por meio de um modelo de regressão Poisson, devido à sua simplicidade. Neste caso, a variável dependente de um modelo de regressão Poisson deve seguir uma distribuição Poisson com média igual à variância. Entretanto, de acordo com Tadano et al., (2009), esta propriedade é frequentemente violada em estudos empíricos, já que é comum a existência de **superdispersão**, ou seja, é frequente que a variância da variável dependente seja maior do que a sua média. Nestes casos, trabalharemos com a estimação de um modelo de regressão binomial negativo.

Os modelos de regressão Poisson e binomial negativo, que também se inserem no contexto dos Modelos Lineares Generalizados (Generalized Linear Models), em que são utilizadas classes de modelos que oferecem alternativas para a transformação dos dados devido ao caráter não linear da variável dependente, tiveram sua origem na década de 1970, quando Wedderburn (1974) desenvolveu a teoria da quasi-verossimilhança.

Ao contrário da tradicional técnica de regressão estimada por meio de métodos de mínimos quadrados, os modelos de regressão para dados de contagem são estimados por máxima verossimilhança e a escolha da melhor estimação depende da distribuição da variável dependente, da relação entre sua média e variância e do objetivo do estudo, com base na teoria subjacente e na

experiência do pesquisador.

2.7.1 Modelo de Regressão Poisson

Os modelos de regressão para dados de contagem têm, por objetivo principal, estudar o comportamento de uma variável dependente, definida por Y , que se apresenta com valores discretos e não negativos, com base no comportamento de variáveis explicativas.

Segundo Cameron e Trivedi(2009), o ponto inicial para o estudo dos modelos de regressão para dados de contagem é a apresentação da distribuição Poisson que, para determinada observação i ($i = 1, 2, \dots, n$, em que n é o tamanho da amostra).

De acordo com Gujarati (2008), distribuição de Poisson é dada por:

$$f(Y) = \frac{\epsilon^{-\mu}}{Y!} Y_i = 0, 1, 2, \dots \quad (2.9)$$

em que $f(Y)$ denota a probabilidade de que a variável Y assuma valores inteiros não negativos e $Y!$ (lê-se factorial de Y) é representado por $Y! = Yx(Y-1)x(Y-2)x(Y-3)x\dots x2x1$.

Pode ser demonstrado por:

$$E(Y) = \mu \quad Var(Y) = \mu$$

um aspecto interessante da distribuição de Poisson é que a sua variância é a mesma que o seu valor médio. O modelo de Poisson pode ser escrito como:

$$Y_i = E(Y_i) + \mu_i = \mu_i + \mu_i \quad (2.10)$$

em que os Y são distribuídos independentemente como variáveis aleatórias de Poisson com média μ_i para cada indivíduo expresso como

$$\mu_i = E(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (2.11)$$

em que os X são algumas das variáveis que poderiam afectar o valor médio.

Para fins de estimação, escrevemos o modelo como:

$$f(Y) = \frac{\epsilon^{-\mu}}{Y!} + u_i \quad (2.12)$$

sendo μ substituído pela equação 2.11 como podemos ver, o modelo de regressão resultante terá parâmetros não lineares, necessitando da estimação de uma regressão não linear.

2.7.2 Estimação dos Parâmetros pelo Método de Máxima Verossimilhança

Os parâmetros de um modelo de regressão Poisson são estimados por máxima verossimilhança, em que a variável dependente segue uma distribuição Poisson.

A função de verossimilhança é dada por :

$$L(\beta) = \prod_{i=1}^n \frac{\exp(-\mu) \mu^{y_i}}{y_i!} \quad (2.13)$$

e a função de log-verossimilhança é dado por:

$$l(\beta) = \sum_{i=1}^n [-\mu_i + y_i \ln(\mu_i) - \ln(y_i!)] \quad (2.14)$$

Como $\mu_i = \exp(x_i^T \mu)$, vem

$$l(\beta) = \sum_{i=1}^n [-\exp(x_i^T \beta) + y_i(x_i^T \beta) - \ln(y_i!)] \quad (2.15)$$

Portanto, maximizando a função de log-verossimilhança $l(\beta)$ com respeito a β temos:

$$\frac{\partial L(\beta)}{\partial L(\beta_j)} = \sum_{i=1}^n [y_i - \exp(x_i^T \mu)] x_i = 0; j = 0, 1, 2, \dots, p \quad (2.16)$$

A função desvio no modelo de regressão de Poisson é definida pela seguinte expressão:

$$D(Y, \hat{\mu}) = 2 \sum_{i=1}^n \left\{ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\} \quad (2.17)$$

A função desvio reduz-se a

$$(Y, \hat{\mu}) = 2 \sum_{i=1}^n \left\{ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \right\} \quad (2.18)$$

para modelos com termo constante, β_0 , porque neste caso

$$\sum_{i=1}^n ((y_i - \hat{\mu}_i)) = 0 \quad (2.19)$$

A estatística de Pearson generalizada é dada por:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad (2.20)$$

2.7.3 Modelo de Regressão Binomial Negativa

O modelo de regressão Binomial Negativa é utilizado quando ocorre sobredispersão no modelo de regressão de Poisson. Considere então que $Y_i \sim \text{BN}(\mu_i, \alpha)$, com os parâmetros $\mu_i > 0$ e $\alpha > 0$, cuja função de probabilidade é dada por:

$$f(y; \mu, \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu_i}\right) \frac{1}{\alpha} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \text{ com, } y_i = 0, 1, 2, \dots \quad (2.21)$$

onde α é denominado parâmetro de dispersão.

Na distribuição Binomial Negativa, a variância é superior à média, e elas são representadas por:

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \mu + \alpha\mu^2$$

Repara-se que, quando o valor do parâmetro de heterogeneidade (α) tende para zero, a distribuição Binomial Negativa tende para a distribuição de Poisson.

Considere-se a variável aleatória Y , com n observações, que representa o número de ocorrências de um determinado acontecimento num certo período de tempo ou espaço.

Dado um vetor de variáveis explicativas $X = (X_1, \dots, X_P)$ e uma observação $x_i^T = (1, x_{i1}, \dots, x_{ip})$ do indivíduo i , assume-se que:

$$Y | X = x_i \sim \text{NB}(\mu(x_i), \alpha)$$

onde

$$\mu_i = \mu(x_i)$$

representa o número de ocorrências de um determinado acontecimento dada a observação x_i . Naturalmente, tem-se que

$$\mu_i = E(Y | X = x_i) \quad \text{e} \quad \text{Var}(Y | X = x_i) = \mu_i + \alpha\mu_i^2.$$

O modelo de regressão Binomial Negativa é expresso por:

$$\log(\mu(x_i)) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (2.22)$$

onde o \log representa a função de ligação do modelo em questão.

2.7.4 Estimação dos Parâmetros

O método utilizado para estimar os parâmetros de regressão Binomial Negativa é a máxima verossimilhança. A função log-verossimilhança para n observações da distribuição Binomial Negativa é dada por Hilbe (2011),

$$l(\beta) = \sum_{i=1}^n \left[y_i \ln \left(\frac{\alpha \mu_i}{1 + \alpha \mu_i} \right) - \left(\frac{1}{\alpha} \right) \ln(1 + \alpha \mu_i) + \ln \left(\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1) \Gamma(\frac{1}{\alpha})} \right) \right] \quad (2.23)$$

onde $\mu(x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})$

O desvio do modelo de regressão binomial negativa é dado por:

$$D = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\mu_i} \right) - \left(\frac{1}{\alpha} + y_i \right) \ln \left(\frac{1 + \alpha y_i}{1 + \alpha \mu_i} \right) \right\} \quad (2.24)$$

Nesta situação, a estatística de Pearson generalizada é,

$$\chi_p^2 = \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i + \alpha \hat{\mu}_i^2} \quad (2.25)$$

Capítulo 3

Material e Métodos

No presente capítulo, são apresentados de forma detalhada a fonte de dados, a descrição das variáveis e metodologia aplicada nas análises dos dados.

3.1 Classificação da Pesquisa

3.1.1 Natureza da Pesquisa

Para este trabalho, será usada a pesquisa de campo, concretamente a pesquisa **ex-post-facto**, que tem por objetivo investigar possíveis relações de causa e efeito entre determinado facto identificado pelo pesquisador e um fenómeno identificado posteriormente. A principal característica deste tipo de pesquisa é o facto de os dados serem coletados após a ocorrência dos eventos.

Trata-se de uma pesquisa quantitativa, pois segundo Fonseca (2002), a pesquisa quantitativa recorre à linguagem matemática para descrever as causas de um fenómeno, as relações entre variáveis, etc.

Adicionalmente, esta pesquisa pode ser classificada como uma **pesquisa de transversal**, pois, apesar de abranger um período de quatro anos (2017-2020), os dados são analisados em conjunto, proporcionando uma visão geral das condições e fatores de risco associados aos acidentes rodoviários na Cidade de Maputo.

Este estudo é descritivo e exploratório, utiliza uma abordagem observacional para identificar fatores associados ao número de vítimas envolvidas em acidentes rodoviários, contribuindo assim para a formulação de políticas públicas eficazes e para a melhoria das medidas de segurança rodoviária.

3.2 Fonte e Descrição dos Dados

Para a realização deste Trabalho, foram utilizados dados fornecidos pelo Comando da Polícia da República de Moçambique (PRM) da Cidade de Maputo, referentes ao período de 2017 a 2020. Os dados incluem informações sobre acidentes rodoviários ocorridos na cidade durante esses anos. Cada registro de acidente contém informações sobre o condutor, sendo que cada condutor foi envolvido em apenas um acidente durante o período em análise.

Tabela 3.1: Descrição das variáveis usadas neste Trabalho

Variável	Contexto da Variável	Classificação
Vítimas**	Variável quantitativa que representa o número de indivíduos envolvidos em um acidente que tenham tido algum dano corporal	Número de vítimas
Ano de ocorrência do acidente*	Variável quantitativa que representa os anos de ocorrência dos acidentes	1. 2017 (referência) 2. 2018 3. 2019 4. 2020
Hora de ocorrência do acidente*	Variável quantitativa que representa a hora de ocorrência dos acidentes	1. 00h00-05h59 (referência) 2. 06h00-11h59 3. 12h00-17h59 4. 18h00-23h59
Idade do condutor*	Variável quantitativa que representa a idade do condutor causador do acidente	1. <18 anos (referência) 2. 18 aos 33 anos 3. 34 aos 49 anos 4. 50 aos 65 anos 5. > 66 anos
Sexo do condutor*	Variável qualitativa que representa o sexo do condutor causador do acidente	1. Feminino (referência) 2. Masculino
Tipo de veículo*	Variável qualitativa que representa o tipo de veículo causador do acidente	1. Ligeiro (referência) 2. Pesado 3. Outros
Dias de ocorrência do acidente*	Variável qualitativa que indica o dia de semana em que o acidente ocorreu	1. Domingo (referência) 2. Segunda-feira 3. Terça-feira 4. Quarta-feira 5. Quinta-feira 6. Sexta-feira 7. Sábado
Distrito urbano de ocorrência do acidente*	Variável qualitativa que representa o local de ocorrência do acidente	1. Kamavota (referência) 2. Kamaxaquene 3. Kampfumo 4. Kalhamankulo 5. Kamubukwana
Grau de Gravidade*	Variável quantitativa que representa a gravidade do acidente	1. Ligeiro (referência) 2. Grave
Causas do acidente*	Variável qualitativa que indica a causa de ocorrência do acidente	1. Álcool (referência) 2. Corte de prioridade 3. Excesso de Velocidade 4. Má travessia do peão
Tipo de acidente*	Variável qualitativa que representa o tipo de acidente ocorrido	1. Atropelamentos (referência) 2. Choques entre carros 3. Despistes e capotamentos 4. Choques contra obstáculos fixos 5. Queda de passageiros 6. Choques entre carros e motos 7. Choques entre carros e outros (veículos de manutenção de estrada) 8. Outros tipos (Colisão frontal, saída de pista, derrapagem)

Nota: * Variável independente. ** Variável dependente.

3.2.1 Softwares usados

No processo de análise de dados, foram usados os seguintes programas estatísticos SPSS, R Studio, Microsoft Excel. O relatório foi elaborado com recurso ao programa $\text{\LaTeX} 2_{\epsilon}$.

3.3 Métodos

3.3.1 Teste de Independência

Segundo Agresti e Finlay (2009), utiliza-se o teste de independência quando os dados da pesquisa se apresentam sob forma de frequências em categorias discretas. Para se aplicar este teste, recorre-se à estatística de Qui-quadrado (χ^2) para determinar se duas variáveis são independentes, usando a seguinte estatística:

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

Onde: O_{ij} - número de casos observados na linha "i" da coluna "j".

E_{ij} - número de casos esperados, na linha "i" da coluna "j".

Para obter a frequência esperada E_{ij} cada célula, multiplicam-se os totais marginais comuns a uma determinada célula e divide-se, multiplicam-se os totais marginais comuns a uma determinada célula e divide-se pelo total de casos.

Os valores de χ^2 obtidos pela equação 3.1, tem distribuição aproximadamente Qui-quadrado com $gl = (L-1)(C-1)$, onde L é o número de linhas e C é o número de colunas, cujas hipóteses a testar são:

$$\begin{cases} H_0 : \text{As variáveis são independentes} \\ H_1 : \text{As variáveis são dependentes} \end{cases} \quad (3.2)$$

Rejeita-se a hipótese nula ao nível de significância α se $\chi^2 > \chi_{(t-1)(c-1)}^2$

Rao e Scott (1984) propuseram um ajuste do teste Qui-quadrado, de modo a incorporar as características do plano amostral, que consiste em dividir o valor da sua estatística por um factor generalizado do efeito do plano amostral (FGEPA), $\chi_{R-S}^2 = \frac{\chi^2}{FGEPA}$, que representa o efeito do plano amostral. O FGEPA é tido como a média dos autovalores da matriz $D = V_{Design} V_{SRD}^{-1}$, em que V_{Design} é uma matriz de variância-covariância baseada no plano amostral para o vector de proporções usadas na construção da estatística Qui-Quadrado e V_{SRD}^{-1} é a matriz de variância-covariância para as proporções estimadas dada uma amostra aleatória simples.

Foi proposta uma correcção de segunda ordem que consiste em dividir a estatística de Rao e Scott pela quantidade $(1 + a^2)$ de modo a incorporar a variabilidade nos autovalores da matriz D, em que a representa o coeficiente de variação dos autovalores.

Rao e Scott (1984), propuseram uma versão F-ajustado da correcção de segunda ordem, que melhor controla o erro do tipo 1, dada pela expressão em

$$F_{T-R} = \frac{\chi_{R-S}^2}{(L-1)(C-1)} \quad (3.3)$$

3.3.2 Modelo de Regressão Poisson

Diversos estudos sobre os dados de contagem utilizam a metodologia de regressão de Poisson como modelo padrão para a análise deste tipo de dados. A razão pela qual o modelo de regressão de Poisson é considerado a metodologia base na modelação de dados de contagens é o facto de assumir apenas valores inteiros não negativos para o valor esperado da variável resposta (Lord e Mannering, 2010).

De acordo com Fávero e Belore (2017), poderemos estimar um modelo de regressão de Poisson, definido da seguinte forma:

$$\ln(\hat{Y}_i) = \ln(\mu_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k x_{kn} \quad (3.4)$$

que também é chamado de modelo log-linear. Sendo assim, o número esperado de ocorrências em dada exposição, para determinada observação i , pode ser escrito como:

$$\hat{\mu}_i = e^{\hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k x_{kn}} \quad (3.5)$$

em que $\hat{\beta}_i$ representa a constante, $\hat{\beta}_i (i = 1, 2, \dots, k)$ são os parâmetros estimados de cada variável explicativa, X são as variáveis explicativas e o subscrito i representa cada observação da amostra ($i = 1, 2, \dots, n$, em que n é o tamanho da amostra).

Sobredispersão

A sobredispersão é um problema que ocorre frequentemente na prática quando se aplica o modelo de regressão de Poisson, surgindo quando a variância da variável resposta é superior ao valor esperado.

Os critérios que indicam a sobredispersão aparente do modelo de regressão de Poisson são segundo (Hilbe, 2011):

- O modelo omite importantes variáveis explicativas;
- Os dados incluem outliers;
- O modelo não inclui um número suficiente de termos de interação;

- Uma variável explicativa precisa ser transformado em outra escala;
- Há situações em que os dados são muito escassos e mais dados precisam ser recolhidos e incluídos no modelo;
- Valores ausentes existem nos dados, mas não são distribuídos aleatoriamente nos dados.

Para identificar a sobredispersão nos dados, podemos utilizar o desvio, $D(y, \hat{\mu})$ também utilizado para testar a qualidade do ajustamento do modelo. O cálculo é baseado na aproximação χ^2 do desvio reduzido. Se existir sobredispersão, então $\frac{D}{\alpha}$ segue uma distribuição qui-quadrado com $n - p$ graus de liberdade, e isso leva ao seguinte estimador para α (Zuur et al., 2009)

$$\hat{\alpha} = \frac{D}{n - p}$$

Se a estimativa deste parâmetro for maior que um (1), é uma indicação da existência de sobredispersão. Caso seja menor que um (1), prossegue-se com o teste da dispersão definido em Cameron e Trivedi (2009).

Teste de dispersão - Um teste estatístico de sobredispersão é altamente desejável após a estimação de um modelo de regressão de Poisson. A maioria dos modelos de contagem que apresentam sobredispersão especificam a variância da seguinte forma:

$$Var[y_i|x_i] = \mu_i + \alpha * g(\mu_i), \tag{3.6}$$

onde α é um parâmetro desconhecido e $g(\cdot)$ é uma função conhecida. As especificações mais comuns da função $g(\mu_i)$ são $g(\mu) = \mu^2$ ou $g(\mu) = \mu$ (Cameron e Trivedi, 2009).

Se $\alpha = 0$, há equidispersão; se $\alpha < 0$ ocorre subdispersão e se $\alpha > 0$ indica sobredispersão. Para testar a sobredispersão, define-se:

$$H_0 : \alpha = 0 \text{ versus } H_1 : \alpha > 0$$

e depois de estimar o modelo de regressão de Poisson, calculam-se $\hat{\mu}_i = \exp(x_i^T \hat{\beta})$ e estima-se o modelo de regressão linear

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha \frac{g(\hat{\mu}_i)}{\hat{\mu}_i} + \epsilon_i$$

onde ϵ_i é um erro aleatório. Sob H_0 , a estatística de teste e a sua distribuição são

$$T = \frac{\hat{\alpha} - \alpha}{s(\hat{\alpha})} \sim N(0; 1)$$

3.3.3 Modelo de Regressão Binomial Negativa

O modelo de regressão Binomial Negativa também foi utilizado na modelagem, pois a variável dependente (Número de vítimas) é quantitativa discreta, ou seja, uma contagem. As variáveis independentes podem ser quantitativas e/ou qualitativas. Segundo Paula (2004), em certas situações práticas, o modelo Binomial Negativa apresenta superioridade em relação ao modelo de Poisson, em termos do desempenho dos estimadores e controle da variabilidade dos dados, pois não necessita que a variância seja igual a média.

Assim, supondo que a variável dependente assume valores inteiros não-negativos e, em particular, não existe um limite superior. A função de probabilidade Binomial Negativa é expressa por:

$$f(y; \mu, \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \text{ com, } y_i = 0, 1, 2, \dots \quad (3.7)$$

onde α é denominado parâmetro de dispersão.

3.3.4 Teste de Significância dos Parâmetros

Após a estimação dos parâmetros, deve-se proceder à verificação da significância estatística dos mesmos. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada a variável resposta. Para isso, existem os testes de hipóteses.

Teste de Wald

Segundo Silva (2016), o teste de Wald é uma das possíveis formas de testar se o parâmetro associado toma o valor zero, ou seja, este teste é utilizado para avaliar se o parâmetro é estatisticamente significativo para um determinado nível de significância. Logo, as hipóteses a serem testadas são:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0. \end{cases} \quad (3.8)$$

A estatística do teste é obtida por:

$$W_j = \frac{\hat{B}_j}{var(\hat{B}_j)} \bigcap N(0, 1) \quad (3.9)$$

Onde \hat{B}_j é o vector de estimativas para β e $var(\hat{B}_j)$ é a matriz de variâncias.

De modo a testar a significância dos parâmetros do modelo é usado um teste de Wald ajustado, baseado no plano amostral, dado pela equação 3.5:

$$F = \frac{s - p + 1}{sp} W \quad (3.10)$$

onde $s = (\sum_{k=1}^k m_k)$ - K é o número total de unidades amostrais primárias amostradas menos o número de estratos, p é o número de variáveis explicativas e W é a versão multivariada do teste de Wald, dado pela equação 3.4. O $p - value$ é gerado usando a distribuição F com p e (s-p+1) graus de liberdade como $Pr[F_{p,s-p+1} \geq F]$.

Por vezes, o teste de Wald é impreciso, e rejeita coeficientes que são estatisticamente significativos, assim sendo, é aconselhável que os coeficientes, identificados pelo teste de Wald como sendo estatisticamente não significativos, sejam testados novamente pelo teste da razão de verossimilhança.

Teste de Razão de Verossimilhança

Silva (2016) defendem que a razão de verossimilhança permite testar uma H_0 contra uma H_1 que contém um maior número de parâmetros. O teste da razão de verossimilhança é obtido através da comparação entre os modelos com e sem as variáveis testadas. Com este teste pretende-se testar simultaneamente se os coeficientes de regressão associados a β são todos nulos com exceção de β_0 .

As hipóteses à testar são:

$$\left\{ \begin{array}{l} H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1 : \text{Pelo menos um dos parâmetros é diferente de zero} \end{array} \right. \quad (3.11)$$

A estatística teste (D), definida abaixo, compara o valor obtido com o de uma distribuição Qui-quadrado com número de graus de liberdade igual à diferença no número de parâmetros.

$$D = -2 \log \left(\frac{\lambda_0}{\lambda_1} \right) \quad (3.12)$$

Onde λ_0 e λ_1 representam os valores de verossimilhança dos modelos correspondentes à hipótese nula e à hipótese alternativa.

ou

$$D = -2 \ln \left[\frac{\text{Verossimilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right] \quad (3.13)$$

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (3.14)$$

Onde: **Modelo ajustado** corresponde ao modelo com apenas as variáveis desejadas para o estudo e o **Modelo saturado** corresponde ao modelo com todas as variáveis e interações.

A estatística do teste de razão de verossimilhança é dada pela equação

$$G^2 = -2 \ln \left[\frac{L(R)}{L(F)} \right] \quad (3.15)$$

$L(R)$ é a função de verossimilhança do modelo ajustado, isto é, do modelo contendo q parâmetros e $L(f)$ é a verossimilhança do modelo completo.

G^2 segue aproximadamente uma distribuição χ^2_{p-q} , sendo a sua regra de decisão procedida de forma análoga ao teste G.

Lumley (2018), o teste de razão de verossimilhanças para amostras complexas é baseado em diferenças entre *Deviance* resultantes da ponderação das funções de pseudo-máxima verossimilhança.

3.3.5 Multicolinearidade

De acordo com Gujarati (2008), o termo Multicolinearidade deve-se a Ragnar Frisch. Originalmente, significava a existência de uma relação linear “perfeita” ou exata entre algumas ou todas as variáveis explanatórias do modelo de regressão.

Quanto maior for o valor de VIF, mais “problemática” ou colinear será a variável X_j . Como regra prática, se o VIF_j de uma variável for maior que 10 (o que acontecerá se R_j^2 for maior que 0,90), essa variável será tida como altamente colinear.

As consequências da multicolinearidade são as seguintes:

- Se existe colinearidade perfeita entre os X, seus coeficientes de regressão são indeterminados e seus erros padrão não são definidos;
- Se a colinearidade for alta, mas não perfeita, a estimação dos coeficientes de regressão será possível, mas seus erros padrão tendem a ser grandes. Como resultado, os valores populacionais dos coeficientes não poderão ser estimados com precisão. Entretanto, se o objetivo for estimar combinações lineares desses coeficientes, as funções estimáveis, pode-se fazer até mesmo na presença de multicolinearidade perfeita.

Embora não haja métodos seguros para detectar a colinearidade, há vários indicadores:

1. O sinal mais claro de multicolinearidade é quando R^2 é muito alto, mas nenhum dos coeficientes de regressão é estatisticamente significativo com base no teste t convencional. Este é um caso extremo, evidentemente.
2. Em modelos envolvendo apenas duas variáveis explanatórias, pode-se ter uma ideia boa da colinearidade examinando-se o coeficiente de correlação simples, ou de ordem zero, entre as duas variáveis. Se a correlação for alta, em geral esta é atribuída à multicolinearidade.
3. Entretanto, os coeficientes de correlação de ordem zero podem ser enganosos em modelos que envolvem mais de duas variáveis X , uma vez que é possível ter correlações baixas de ordem zero e, no entanto, encontrar alta multicolinearidade. Em situações como essas, pode ser preciso examinar os coeficientes de correlação parcial.
4. Se R^2 é alto, mas as correlações parciais são baixas, a multicolinearidade é uma possibilidade. Nesse caso, uma ou mais variáveis podem ser supérfluas. Mas, se R^2 for alto e as correlações parciais também, a multicolinearidade pode não ser detectável imediatamente.
5. Portanto, pode-se efetuar a regressão de cada uma das variáveis X_i contra as variáveis X remanescentes do modelo e descobrir os coeficientes de determinação correspondentes, R_i^2 . Um elevado R_i^2 sugere que X_i está estreitamente correlacionado com o restante dos X . Assim, podemos excluir esse X_i do modelo, contanto que ele não leve a um grave viés de especificação.

Capítulo 4

Resultados e Discussão

Neste capítulo, são apresentados os resultados obtidos de acordo com a metodologia apresentada no capítulo 3.

4.1 Análise Exploratória dos Dados

4.1.1 Distribuição dos condutores por sexo e faixa etária

O gráfico 4.1 ilustra a distribuição dos condutores por sexo envolvidos em acidentes rodoviários na cidade de Maputo entre 2017 à 2020.

Os condutores do sexo masculino representam a maioria dos envolvidos em acidentes rodoviários, representando 67% do total.

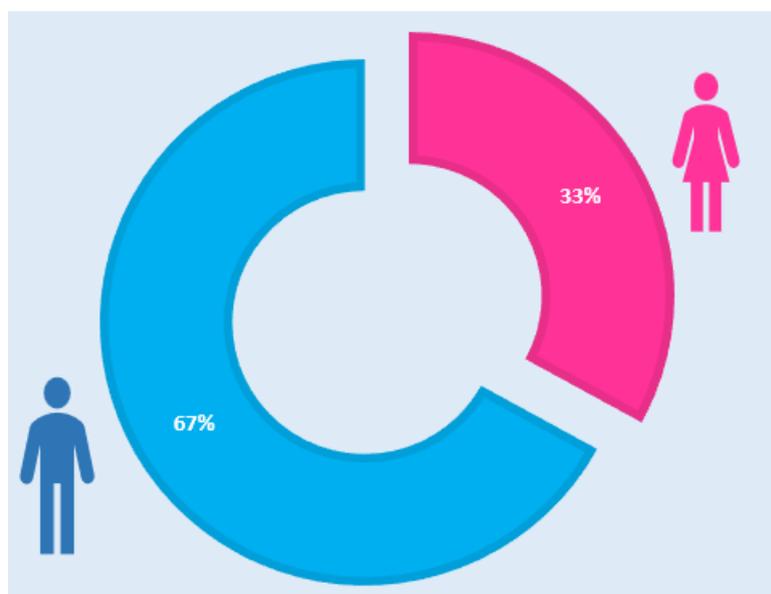


Figura 4.1: Distribuição dos condutores por sexo

Em relação as faixas etárias dos condutores, grande parte dos acidentes, 43% foram provocados por condutores dos 18 aos 33 anos e dos 34 à 49 anos (vide o gráfico 4.2).

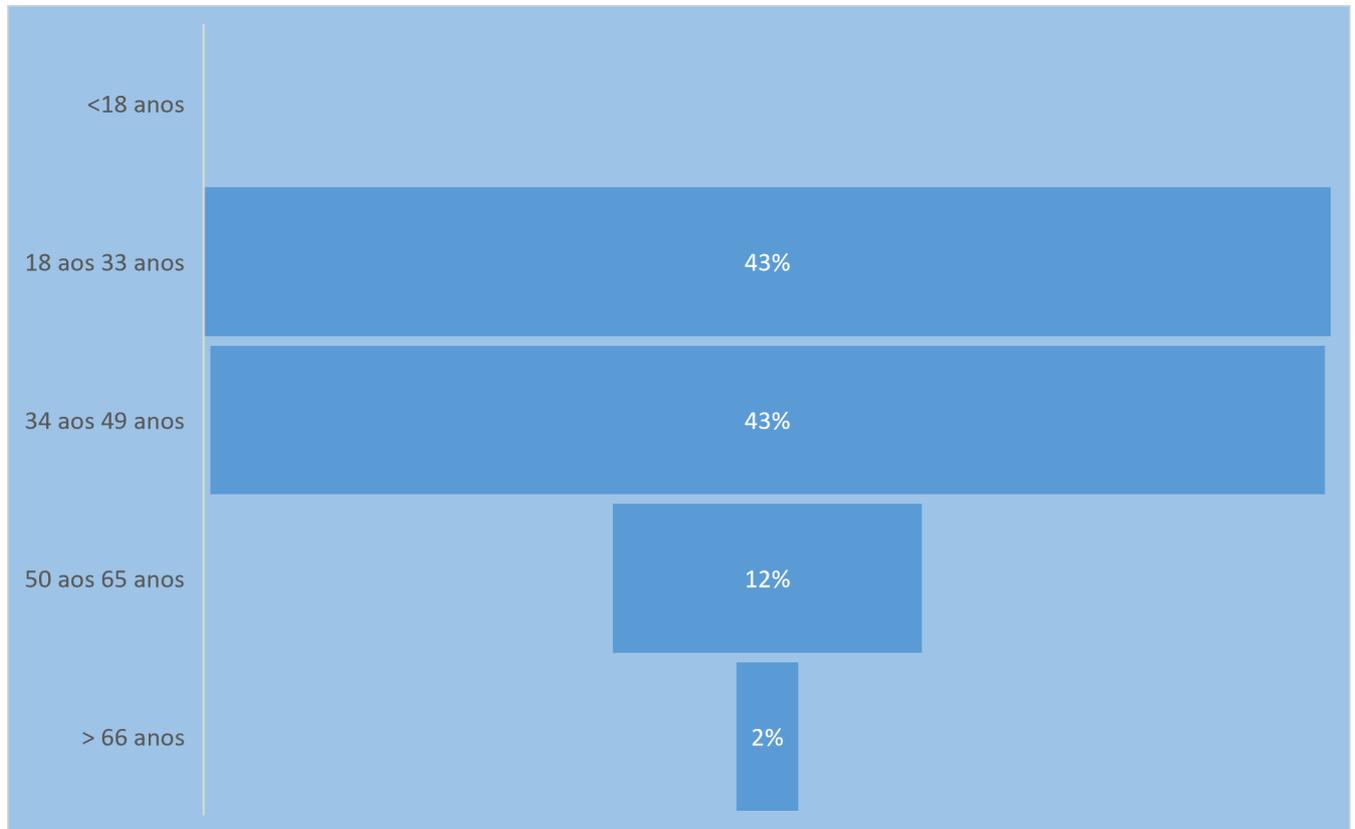


Figura 4.2: Distribuição dos condutores por faixas etárias

4.1.2 Distribuição dos acidentes por tipo de veículo

O gráfico 4.3 mostra a distribuição dos tipos de veículos envolvidos em acidentes rodoviários na Cidade de Maputo entre 2017 à 2020.

A maior parte dos veículos envolvidos em acidentes são de categoria ligeira, representando 69% do total, representando 25% dos acidentes e por último 6% dos acidentes foram provocados por viaturas de outras categorias, conforme ilustra a figura 4.3.

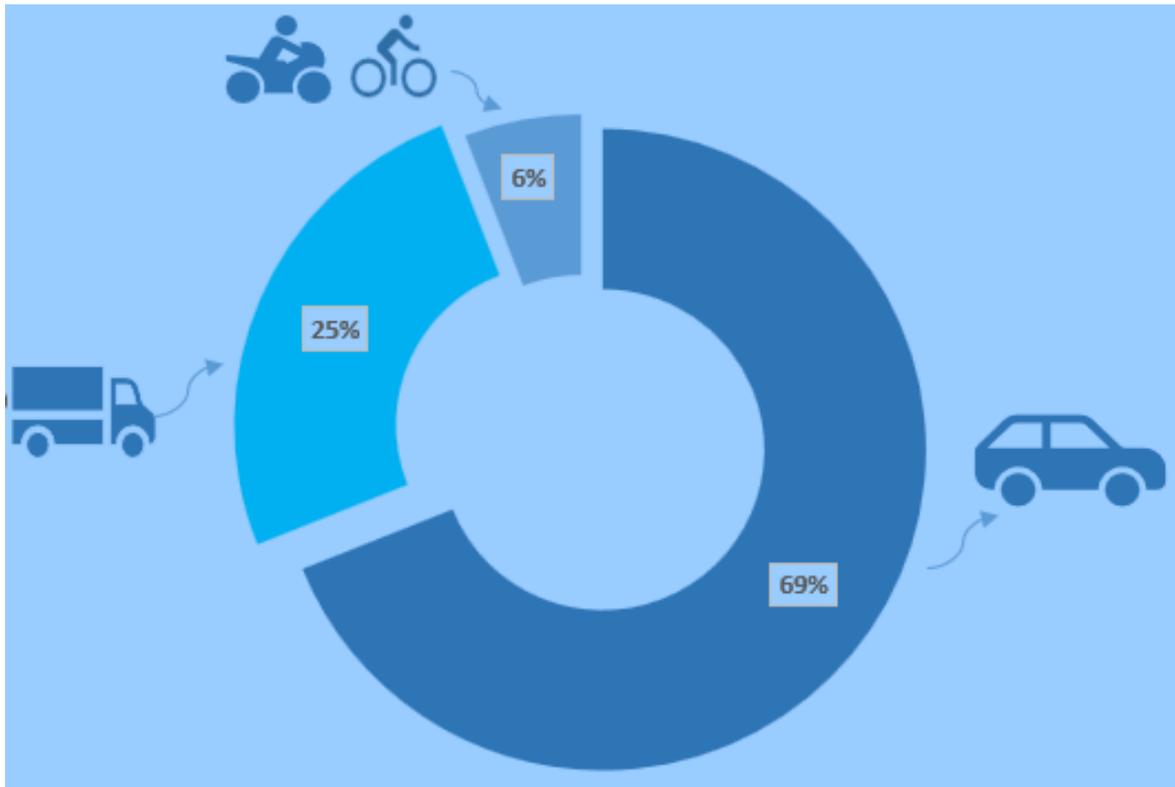


Figura 4.3: Distribuição dos acidentes por tipo de veículo

4.1.3 Distribuição dos acidentes por dia da semana e hora de ocorrência

De acordo com a Figura 4.4, constatou-se que 18% dos acidentes foram registrados aos Domingos, 20% aos sábados e 14% nas Sexta-feiras.

Por outro lado, do total de acidentes registrados entre 2017 e 2020, 33% ocorreram entre 18h00 às 24h, seguido por 31% entre as 12h00 e as 18h00, e 26% entre as 06h00 e as 12h00.

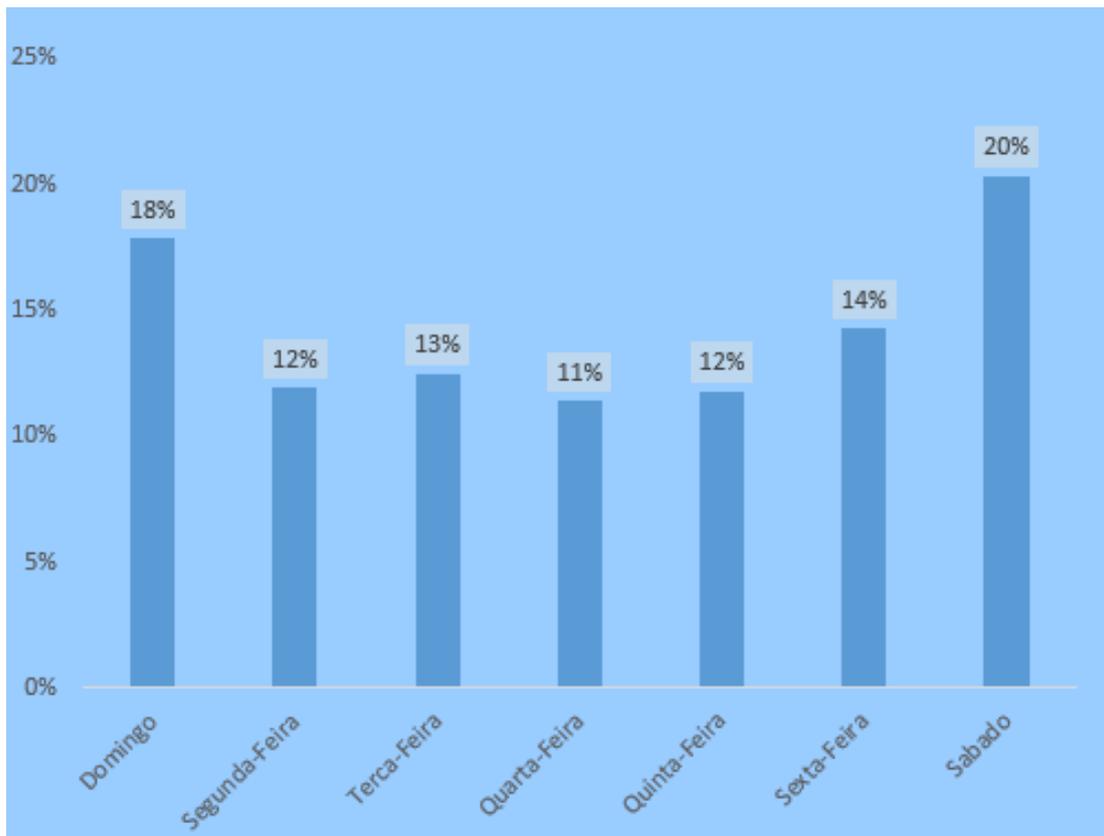


Figura 4.4: Distribuição dos acidentes por dia de ocorrência

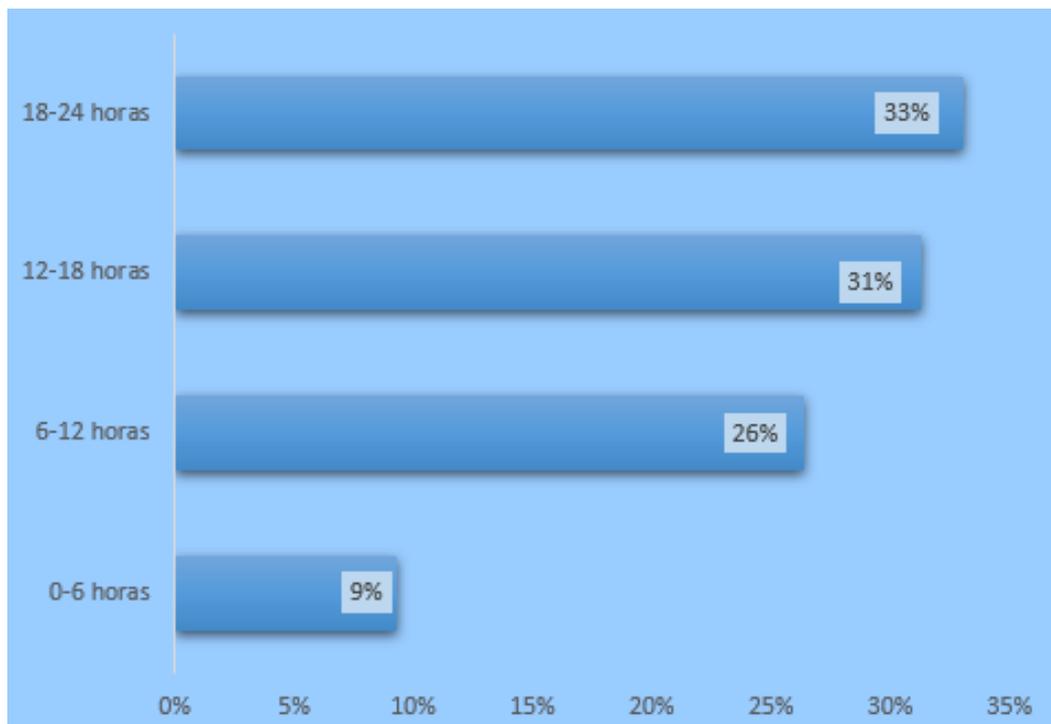


Figura 4.5: Distribuição dos acidentes por hora de ocorrência

4.1.4 Distribuição dos tipos de acidentes

De acordo com a Figura 4.6, constatou-se que 31% dos acidentes foram do tipo Atropelamentos, 23% Despiste e capotamento e 18% Choques entre carros.

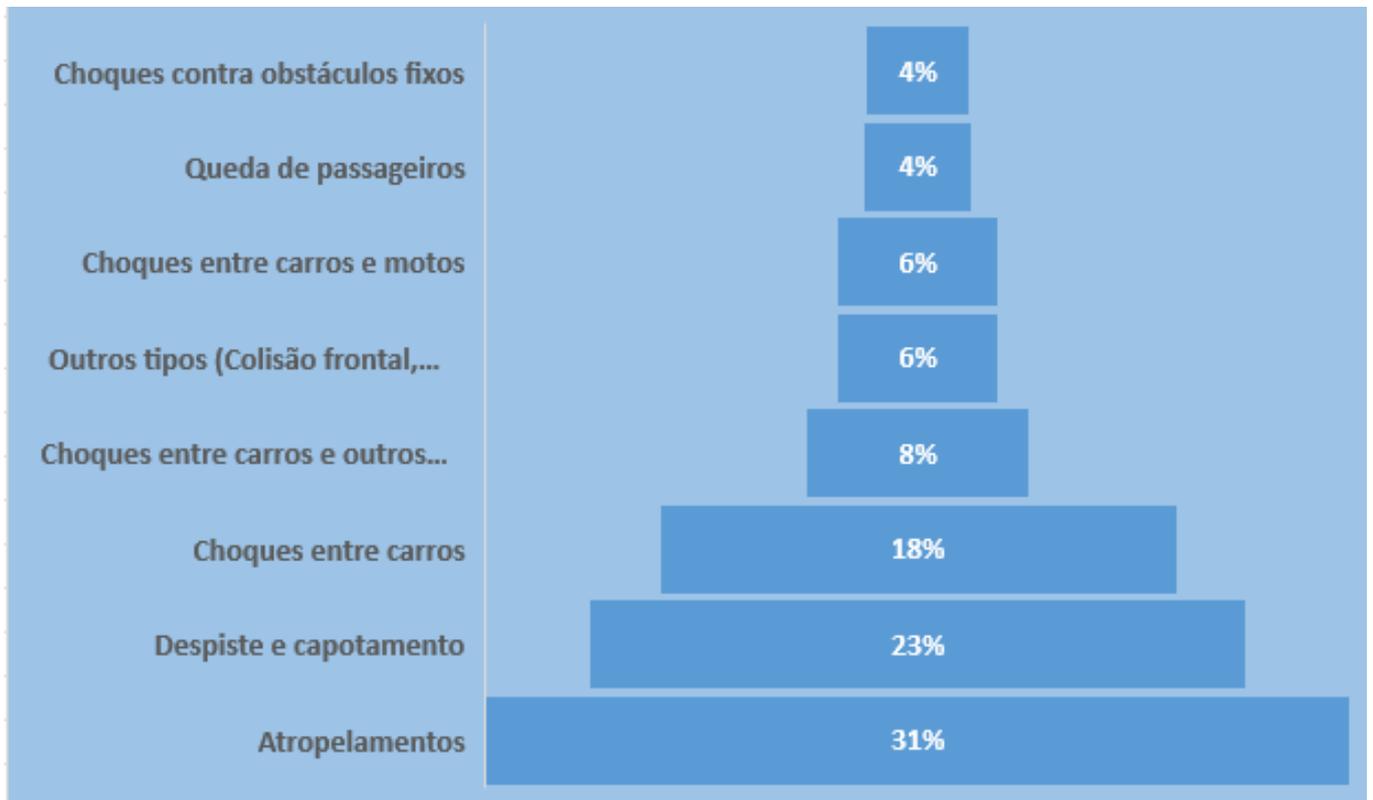


Figura 4.6: Distribuição dos tipos de acidentes

4.2 Comparação dos Modelos de Poisson e Binomial Negativa

A tabela 4.1 apresenta os valores dos critérios de informação de Akaike (AIC) e Bayesiano (BIC) para comparação dos modelos onde pode-se observar que o modelo de regressão Binomial Negativa é o melhor para descrever os dados, pois apresenta menores valores de AIC e BIC comparativamente ao modelo de regressão de Poisson.

Tabela 4.1: Comparação dos modelos estimados incluindo o log-likelihood negativo duplo, o número de parâmetros, o AIC e o BIC

Modelos	-2ll	Parâmetros	AIC	BIC
Regressão de Poisson	-2975.803	12	5975.607	6038.11
Regressão Binomial Negativa	-2677.829	12	5381.659	5449.37

4.3 Modelo de Regressão Binomial Negativa

A tabela 4.2 apresenta os parâmetros estimados e os respectivos erros padrões dos modelos de regressão Binomial Negativa simples e múltipla.

Os resultados mostraram que o tipo de veículo, dias da semana, horas de ocorrência, causas dos acidentes, distrito urbano, tipo de acidente, sexo do condutor e grau de gravidade são as variáveis significativamente associadas ao elevado número de vítimas de acidentes.

A diferença do número de vítimas causadas por veículos pesados e viaturas ligeiras é positiva e estatisticamente significativa ($\beta = 0.040$, p-valor= 0.005), isso indica que o número de vítimas de acidentes causados por veículos pesados tem sido ligeiramente superior ao número de vítimas causadas por veículos ligeiros. Ademais, o número de vítimas dos acidentes que acontecem nas sextas-feiras e aos sábados é maior quando comparado com o número de vítimas dos acidentes que ocorrem no dia de referência e a diferença é estatisticamente significativa ($\beta = 0.062$, p-valor= 0.006) e ($\beta = 0.062$, p-valor= 0.003).

Para as horas de ocorrência, os coeficientes negativos e estatisticamente significativos indicam que os acidentes ocorridos entre 6 e 12 horas, 12 e 18 horas, e 18 e 24 horas tiveram um menor número de vítimas em relação ao período de 0-6 horas. por outro lado, o número de

vítimas causados pela má travessia de peões ($\beta = 0.219$, p-valor = 0.018) é maior que o número de vítimas causadas por outro tipo de acidente. Similarmente, acidentes causados por excesso de velocidade e o consumo de álcool apresentam maior número de vítimas quando comparados com o número de vítimas causados por outros tipos de acidentes ($\beta = 0.052$, p-valor = 0.013; $\beta = 0.270$, p-valor = 0.007, respectivamente).

A diferença do número de vítimas nos acidentes que ocorreram no distrito Kamaxaquene e no Kampfumo é negativo e estatisticamente significativa ($\beta = -0.150$, p-valor = 0.012) e ($\beta = 0.057$, p-valor = 0.514) respectivamente, sugerindo que estes distritos têm menos vítimas em comparação com o distrito de Kalhamankulo. O número de vítimas de acidentes envolvendo choques entre carros ($\beta = 0.219$, p-valor = 0.018), choques entre carros e motos ($\beta = 0.219$, p-valor = 0.018), choques entre carros e outros ($\beta = 0.219$, p-valor = 0.018), é maior que o número de vítimas de acidentes causados pelos atropelamentos.

A diferença entre o número de vítimas de acidentes graves e ligeiros é positiva e estatisticamente significativa ($\beta = 0.342$, p-valor = 0.002), sugerindo que acidentes graves tendem a resultar em um número maior de vítimas em comparação com acidentes ligeiros. Um coeficiente positivo e estatisticamente significativo ($\beta = 0.170$, p-valor = 0.001) indica que o número de acidentes provocados por condutores do sexo masculino tem sido ligeiramente maior que o número de vítimas de acidentes envolvendo os condutores do sexo feminino.

Tabela 4.2: Parâmetros e erros padrão estimados, com os respectivos p-valores das análises de regressão simples e múltipla

Variável	Categorias	Simples		Múltipla	
		Coefficiente (Erro padrão)	P-valor	Coefficiente (Erro padrão)	P-valor
Intercepto	—	—	—	1.638(0.126)	0.001*
Tipo de veículo	Ligeiro (referência)	—	—	—	—
	Pesado	0.119 (0.028)*	0.026	0.040 (0.054)*	0.005
	Outros (Motocicletas)	-0.016(0.054)	0.876	-0.009(0.105)	0.933
Dias da Semana	Domingo (referência)	—	—	—	—
	Segunda-feira	-0.085(0.088)	0.331	0.037(0.086)	0.669
	Terça-feira	0.040(0.085)	0.639	0.060(0.084)	0.468
	Quarta-feira	0.082(0.092)	0.317	-0.155(0.091)	0.088
	Quinta-feira	-0.071(0.090)	0.4294	0.153(0.088)	0.057
	Sexta-feira	-0.202(0.082)*	0.029	0.062(0.080)*	0.006
	Sábado	-0.036(0.075)	0.634	0.062(0.074)*	0.003
Horas de ocorrência	0-6 horas (referência)	—	—	—	—
	6-12 horas	-0.250(0.083)*	0.003	-0.259(0.084)*	0.002
	12-18 horas	-0.280(0.081)*	0.0006	-0.247(0.083)*	0.003
	18-24 horas	-0.378(0.081)*	<0.001	-0.334(0.083)*	<0.001
Má travessia do peão	Não ocorrência (referência)	—	—	—	—
	Ocorrência	-0.347(0.066)*	<0.001	0.219(0.092)*	0.018
Corte de prioridade	Não ocorrência (referência)	—	—	—	—
	Ocorrência	-0.005(0.109)	0.646	-0.303(0.122)	0.464
Excesso de velocidade	Não ocorrência (referência)	—	—	—	—
	Ocorrência	0.098(0.047)*	0.038	0.052(0.071)*	0.013
Álcool	Não ocorrência (referência)	—	—	—	—
	Ocorrência	-0.090(0.081)	0.263	0.270(0.090)*	0.007
Distrito urbano	Kalhamankulo (referência)	—	—	—	—
	Kamaxaquene	-0.133(0.060)*	0.026	-0.150(0.060)*	0.012
	Kampfumo	-0.183(0.089)*	0.041	-0.057(0.088)*	0.514
	Kamavota	-0.118(0.083)	0.155	-0.051(0.081)	0.529
	Kamubukwana	0.024(0.068)	0.721	0.087(0.067)	0.190
Tipo de acidente	Atropelamentos (referência)	—	—	—	—
	Choque entre carros	0.494(0.032)*	<0.001	0.416(0.069)	<0.001*
	Despistes e capotamento	-0.015(0.056)	0.919	-0.093(0.147)	0.526
	Choques contra obstáculos fixos	-0.076(0.143)	0.655	-0.113(0.172)	0.509
	Choques entre carros e motos	0.334(0.171)*	<0.001	0.181(0.078)*	0.020
	Choques entre carros e outros	0.516(0.154)*	<0.001	0.452(0.157)*	0.004
	Queda de passageiros	0.047(0.158)	0.767	-0.029(0.166)	0.863
	Outros tipos	0.045(0.196)	0.818	0.140(0.199)	0.481
Grau de Gravidade	Ligeiro (referência)	—	—	—	—
	Grave	-0.153(0.049)*	0.008	0.342(0.058)*	0.002
Sexo do condutor	Feminino (referência)	—	—	—	—
	Masculino	0.134(0.049)*	0.006	0.170(0.048)*	0.001

Nota: * indica significância estatística a um nível de 5%.

4.3.1 Teste de Ajustamento a Distribuição Binomial Negativa

A figura 4.7 mostra o teste de ajustamento para o modelo de regressão Binomial Negativa, é possível observar que as densidades das distribuições dos valores observados e previstos, representadas pelas linhas azul e vermelha, respectivamente, estão próximas em toda a faixa de valores. Isso sugere um bom ajuste do modelo aos dados observados. No entanto, em certas regiões do histograma, nota-se discrepâncias onde a densidade prevista parece subestimar a densidade observada.



Figura 4.7: Distribuição do número de vítimas observado vs. Previsto

4.4 Multicolinearidade

4.4.1 Verificação da Multicolinearidade

A tabela 5.1 em anexo, mostra os resultados do cálculo do VIF (Fator de Inflação da Variância) para cada variável independente no modelo de regressão Binomial negativa e os valores dos mesmos indicam que não há multicolinearidade entre as variáveis independentes e o número de vítimas de acidentes rodoviários.

4.5 Discussão dos Resultados

O presente Trabalho teve como objetivo principal analisar os fatores associados ao número de vítimas envolvidas em acidentes rodoviários na Cidade de Maputo entre 2017 e 2020.

A análise revelou que a maioria dos condutores envolvidos em acidentes rodoviários são do sexo masculino e têm idades entre 18 e 49 anos. Esta observação é consistente com várias pesquisas anteriores. De acordo com o relatório da OMS(2018) sobre segurança rodoviária, homens jovens são mais propensos a se envolver em acidentes rodoviários devido a uma combinação de fatores, como maior exposição ao risco, comportamento impulsivo e maior disposição para adotar comportamentos de condução perigosos OMS(2018). Estudos adicionais, como o de Dobson et al. (2013), corroboram que os jovens do sexo masculino tendem a subestimar os riscos associados à velocidade e ao consumo de álcool, resultando em maior incidência de acidentes.

O gráfico da distribuição dos acidentes por tipo mostrou que os atropelamentos, choques entre veículos e despistes/capotamentos foram as três categorias com maior número de acidentes registados. Este padrão é corroborado pelo relatório "Estatísticas de Acidentes de Viação, 2020" do INE. Os atropelamentos indicam uma necessidade urgente de melhorar a segurança dos pedestres, possivelmente através da implementação de passarelas, semáforos para pedestres e campanhas de conscientização sobre a importância de usar as passagens apropriadas.

Os choques entre veículos e os despistes/capotamentos sugerem que a infraestrutura rodoviária pode não estar adequadamente projetada para lidar com o volume e a velocidade do tráfego. Segundo uma análise da European Transport Safety Council (ETSC, 2015), a melhoria da infraestrutura, como a criação de faixas adicionais, melhorias na sinalização e a manutenção regular das estradas, pode reduzir significativamente o número de acidentes desse tipo.

A distribuição dos acidentes por dia da semana indicou que sextas-feiras, sábados e domingos têm uma maior frequência de acidentes rodoviários. Este fenômeno pode ser explicado pelo aumento do fluxo de condutores e pedestres durante esses dias devido ao fim de semana e atividades sociais mais intensas. A maior movimentação nas vias nesses dias aumenta a probabilidade de ocorrências de acidentes. Dados da National Highway Traffic Safety Administration (NHTSA, 2019) dos Estados Unidos também mostram um aumento nos acidentes durante os fins de semana, especialmente à noite, correlacionado com o aumento do consumo de álcool e a maior velocidade.

Em relação às causas dos acidentes, constatou-se que o excesso de velocidade, o consumo de álcool e a má travessia de pedestres estão fortemente associadas a um maior número de víti-

mas. O excesso de velocidade é um fator crítico que diminui o tempo de reação do condutor e aumenta a gravidade das colisões. Segundo Elvik et al. (2004), a redução de 1% na velocidade média pode resultar em uma redução de aproximadamente 4% nos acidentes fatais.

O consumo de álcool compromete significativamente a capacidade de condução segura. De acordo com a OMS (2018), condutores com níveis de álcool no sangue superiores ao limite legal têm uma probabilidade muito maior de se envolverem em acidentes graves. Políticas rigorosas de fiscalização, como a implementação de pontos de verificação de sobriedade e campanhas de conscientização pública, têm se mostrado eficazes em reduzir o número de acidentes relacionados ao álcool.

A má travessia dos peões indica a necessidade de melhor infraestrutura e educação para garantir a segurança nas vias. Implementações como faixas de pedestres elevadas, semáforos específicos para peões e campanhas educativas podem reduzir significativamente o risco de atropelamentos. Um estudo de Zegeer et al. (2005) mostra que tais medidas podem reduzir acidentes de pedestres em até 50%.

Os resultados deste estudo destacam a importância de intervenções multifacetadas para melhorar a segurança rodoviária na Cidade de Maputo. Uma abordagem abrangente deve incluir: educação e conscientização, com campanhas educativas focadas em jovens do sexo masculino para conscientizá-los sobre os riscos da velocidade e do consumo de álcool; fiscalização rigorosa, com aumento da fiscalização sobre o excesso de velocidade e o consumo de álcool, especialmente durante os fins de semana; melhoria da infraestrutura, com investimento em melhorias de infraestrutura rodoviária, como a construção de passarelas e a manutenção regular das estradas; e políticas de proteção ao pedestre, com a implementação de medidas de segurança para pedestres, como faixas elevadas e semáforos exclusivos para pedestres.

Em suma, os resultados deste estudo fornecem uma base sólida para a formulação de políticas públicas eficazes e para a melhoria das medidas de segurança rodoviária. A implementação dessas recomendações pode contribuir significativamente para a redução do número de vítimas em acidentes rodoviários na Cidade de Maputo.

Capítulo 5

Conclusões e Recomendações

5.1 Conclusões

A análise inicial foi a exploratória dos dados, pois ela pode sugerir se existe uma relação ou associação entre a variável resposta e as variáveis explicativas, relação essa que pode posteriormente ser melhor estudada e compreendida com a análise de regressão. A análise de regressão incide nos estudos dos modelos lineares generalizados, que neste trabalho abrangeu os modelos de regressão de Poisson e o de regressão Binomial Negativa.

As análises bivariadas assim como as multivariadas revelaram diversas associações significativas entre variáveis e o número de vítimas em acidentes de trânsito. Os resultados mostram que os acidentes causados por veículos pesados está associada a um aumento no número de vítimas, enquanto que acidentes que ocorrem as sextas-feiras e aos sábados estão associados a um aumento em comparação com outros dias da semana. Além disso, acidentes que ocorrem entre 6-12 horas, 12-18 horas e 18-24 horas têm menos vítimas em comparação com o período de 0-6 horas.

Acidentes causados por má travessia do peão, o excesso de velocidade e a presença de álcool estão todos relacionados a um aumento no número de vítimas.

Tipos específicos de acidentes, como choques entre carros, entre carros e motos, e entre carros e outros, também estão ligados a mais vítimas. Além disso, acidentes graves tendem a ter mais vítimas do que acidentes ligeiros, enquanto que acidentes causados por condutores do sexo masculino estão associados a um aumento no número de vítimas em comparação com os do sexo feminino.

5.2 Recomendações

O presente Trabalho fornece conclusões importantes para o desenvolvimento de políticas e estratégias eficazes para a redução do número de vítimas em acidentes rodoviários. Com base nos resultados obtidos, recomenda-se as seguintes medidas:

- Implementação de campanhas educativas e de conscientização para promover comportamentos seguros de condução, especialmente em relação ao excesso de velocidade, consumo de álcool e respeito às regras de trânsito;
- Reforço da fiscalização e aplicação das leis de trânsito, com ênfase em áreas e horários de maior risco.
- Investimento em pesquisas adicionais para uma compreensão mais abrangente dos fatores subjacentes aos acidentes rodoviários e a eficácia das medidas de prevenção implementadas.

Essas recomendações visam contribuir para a redução da morbidade e mortalidade associadas aos acidentes rodoviários, promovendo um ambiente rodoviário mais seguro e protegendo a vida e a saúde de todos os usuários das vias públicas.

5.3 Limitações

O presente Trabalho é limitado pela disponibilidade e qualidade dos dados utilizados. Embora os conjuntos de dados sobre acidentes rodoviários sejam valiosos, podem conter falhas e deficiências que afetam a interpretação dos resultados. A natureza heterogênea da coleta de dados, envolvendo relatórios policiais, registros médicos e seguradoras, pode resultar em uma representação incompleta dos acidentes, exigindo uma interpretação cuidadosa e consideração de possíveis vieses e lacunas na análise.

O registro incompleto de acidentes é comum, pois nem todos os acidentes são reportados, especialmente os menores, resultando em uma possível distorção dos dados. A qualidade e a consistência dos relatórios policiais podem variar devido à subjetividade, falta de treinamento e pressa na compilação dos dados, levando a inconsistências.

A coleta de dados ao longo de diferentes anos pode sofrer inconsistências temporais devido a mudanças nas práticas de registro ou critérios de classificação, afetando a comparabilidade dos dados. Muitas vezes, os dados não incluem informações contextuais importantes, como condições meteorológicas ou estado das vias, limitando a compreensão completa dos fatores de risco. O viés de informação é uma preocupação, pois a precisão das informações sobre as

causas dos acidentes pode ser comprometida pela dificuldade de determinar causas exatas, especialmente com múltiplos fatores envolvidos. A interpretação dos fatores de risco é complexa devido à inter-relação entre variáveis, como consumo de álcool e excesso de velocidade, dificultando a determinação do impacto isolado de cada fator. Mudanças em políticas de trânsito ou melhorias na infraestrutura durante o período de estudo podem afetar os padrões de acidentes e não serem refletidas adequadamente nos dados.

Referências

- [1] Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. John Wiley e Sons, Hoboken, New Jersey.
- [2] Agresti, A. e Finley, B.(2009). Statistical Methods for the Social Science. 4th ed. Prentice Hall.
- [3] Alvarenga, A. (2015). Dissertação: Modelos lineares generalizados: aplicação a dados de acidentes rodoviários. Lisboa, Portugal.
- [4] Araújo, P. (2017). Regressão Logística Multinível: um estudo sobre o consumo de álcool e inalantes entre crianças e adolescentes. Brasília, Brasil.
- [5] Balaniuk, R et (2018). Caracterização de acidentes rodoviários e as ações governamentais para a redução de mortes e lesões no trânsito: Um estudo de dados da rodovia BR-101 no período de 2014 a 2016. *Revista transporte y Território-Brasil*;
- [6] Cameron, a. e Trivedi, P. (2009) Microeconometrics using stata.
- [7] Cordeiro, G. (2008). Modelos Lineares Generalizados e Extensões. São Paulo, Brasil.
- [8] Cordeiro, G e Demétrio, C (2010) Modelos lineares generalizados. Universidade estadual de Campinas. Departamento de estatística. Campinas, Brasil.
- [9] Dobson, A., et al.(2013). Young Drivers: The Road Ahead. Journal of safety Research.
- [10] Elvik, R., et al. (2004). The Handbook of Road Safety Measures. Emerald Group Publishing.
- [11] European Transport Safety Council (ETSC) (2015). Reducing Road Risk for Pedestrians and Cyclists.
- [12] Fávero,P. e Belfiore, P. (2017). Manual de análise de dados. Estatística Modelagem Multivariada com excel, SPSS e Stata- Rio de Janeiro, Brasil. série. Lisboa, Portugal.
- [13] Fonseca, J.(2002) Metodologia da pesquisa científica. Fortaleza: UEC, Apostila

-
- [14] Gonçalves, A. M. (2013). Regressão Logística Aplicada à Preditores de Morte. Dissertação de Mestrado - Universidade de Coimbra.
- [15] Gujarati, D., e Porter, C. (2008). Basic Econometrics (5th ed.). New York, NY: McGraw-Hill Education.
- [16] Hill, M. (2005). Investigação por questionário. Lisboa, edições Silabo, Lda.
- [17] Hilbe, J.(2011) Modeling count data. Springer.
- [18] Hosmer, J. e Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley e Sons Inc.
- [19] Hosmer et al.,(2000) *Applied Logistic Regression* Edition Wiley Series in Probability and Sattistics texts and references section (2ª Edition), ISSN 0094306.
- [20] Instituto Nacional de Estatística (INE) *Anuário Estatístico*, Cidade de Maputo 2018.
- [21] Instituto Nacional de Estatística (INE) *Estatísticas de Acidentes de viação*, Cidade de Maputo 2020.
- [22] INAV- Instituto Nacional de viação, (2009), Análise da segurança rodoviária em Moçambique documento interno, Maputo.
- [23] INAV- in “STOP,” Revista Trimestral, (2011), do Instituto Nacional de Viação, Maputo.
- [24] Lima, E. (2018). Análise de Dados categóricos- Modelos Lineares Generalizados. Universidade Federal de Minas Gerais. Minas Gerais, Brasil
- [25] Lumley T. (2010). Complex Surveys: A Guide to Analysis Using R. John Wiley e Sons, Hoboken, New Jersey, USA.
- [26] Lord, D. and Mannering, F. (2010). The Statistical Analysis of Crash Frequency Data: A Review and Assessment of Methodological Alternatives. Accident Analysis and Prevention
- [27] Lumley T. (2018). Analysis of Complex Survey Samples: Package survey. UTC.
- [28] Mestre, P. (2009). Aplicação dos Modelos Lineares Generalizados às Telecomunicações Móveis: Caracterização dos clientes que desactivam os seus serviços. Lisboa, Portugal.
- [29] National Highway Traffic Safety Administration (NHTSA) (2019). Traffic Safety Facts Annual Report. Magis. Revista Internacional de Investigación Educativa
- [30] Nelder, J.A. and Wedderburn, R.W.M. (1972). *Generalized linear models*. Journal of the Royal Statistical Society, A 135, 370-384.

- [31] Oliveira, José Ferreira de, (2006), As políticas de Segurança e os Modelos de Policiamento, Coimbra, Almedina
- [32] Organização para a Cooperação e Desenvolvimento Económico (OCDE). (2008). *Improving Road Safety in Developing Countries: A Challenge for the OECD*.
- [33] Organização Mundial da Saúde (OMS) (2018). Global Status Report on Road Safety.
- [34] Organização Mundial da Saúde (OMS). (2015) Escritório regional para a África. *Segurança Rodoviária na região Africana* ISBN 978 929034088 1
- [35] Paula,G.(2004). Modelos de Regressão com apoio computacional. São Paulo.
- [36] Paulo,G.D. (2019). Modelos de Regressão de Poisson Generalizado: Análise de dados de ontagem com superdispersão e subdispersão. Dissertação (Mestrado em Estatística)- Universidade do Minho Escola de ciências.
- [37] PNPR, (2010), Plano Nacional de Prevenção Rodoviária, Lisboa.
- [38] Queiroz, N. (2004). Regressão Logística - Uma estimativa Bayesiana aplicada na identificação de Factores de Risco para HIV em doadores de Sangue. Recife, Brasil.
- [39] Ramalho, J. (1996).Modelos de regressão para dados de contagem. Tese de mestrado
- [40] Rao, J. e Scott, A. (1984). Chi-squared test for multiway contingency tables with cell proportions estimated from survey data, *Annals of Statistics*.
- [41] Silva, P. (2016). Modelos de Regressão Linear e Logística utilizando o software R. Porto, Portugal.
- [42] Turkman, A. e Silva, G. (2004). Modelos Lineares Generalizados - da teoria à prática - Lisboa, Portugal.
- [43] Tadano, Y. et al., (2009) Método de regressão de Poisson: metodologia para avaliação do impacto da poluição atmosférica na saúde populacional.
- [44] Vasconcellos, E.A (2010) Análise das políticas de transporte sustentável no Brasil -o caso da segurança de transito, Instituto Movimento e volvo Educacional foundations, São Paulo.
- [45] Wedderburn, R.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models and the Gauss Newton Method. *Biometrics*.
- [46] World Health Organization (WHO). (2004). *World Report on Road Trac Injury Prevention* . Geneva;
- [47] World Health Organization (WHO). (2009). *Global Plan for the Decade of Action for Road Safety 2009-2020*, Geneva, Switzerland;

- [48] World Health Organization (WHO) (2015) *Global status report on road safety 2015*, WHO Library Cataloguing-in Publication Data, Geneva, Switzerland.
- [49] Zegeer, C. V., et al. (2005). Safety Effects of Marked vs. Unmarked Crosswalks at Uncontrolled Locations. *Transportation Research Record*.
- [50] Zeileis, A. et al. (2008) .Regression models for count data in r.*Journal of statistical software*.
- [51] Zuur A. et al. (2009). *Mixed effects models and extensions in ecology with R*. Springer Science & Business Media,

Anexo

Tabela 5.1: Valores do fator de inflação da variância (VIF)

Variável	VIF(Fator de Inflação da Variância)	Df	VIF Ajustado
Tipo de veículo	1.217	2	1.050
Dias da semana	1.157	6	1.012
Excesso de velocidade	2.421	1	1.556
Corte de prioridade	1.359	1	1.166
Álcool	1.627	1	1.276
Distrito urbano	1.202	4	1.023
Tipo de acidente	2.375	7	1.064
Grau de gravidade	1.481	1	1.217
Má travessia do peão	2.079	1	1.442
Hora de ocorrência	1.201	3	1.031
Sexo do condutor	1.037	1	1.018