



**UNIVERSIDADE FEDERAL DE SERGIPE - UFS**

**CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS - CCET**

**DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS – DECAT**

**UTILIZAÇÃO DO MODELO DE REGRESSÃO DE POISSON PARA ANÁLISE DAS  
VÍTIMAS FATAIS ENVOLVENDO CICLISTAS EM ARACAJU**

**MARCEL BARBOSA DE OLIVEIRA**

**São Cristóvão – SE**

**07/03/2017**

# **MARCEL BARBOSA DE OLIVEIRA**

Monografia apresentada ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como um dos pré-requisitos para obtenção do grau de Bacharel em Estatística.

Orientadora: **Profa. Dra. Suzana Leitão Russo**

**São Cristóvão – SE**

**07/03/2016**

**MARCEL BARBOSA DE OLIVEIRA**

**UTILIZAÇÃO DO MODELO DE REGRESSÃO DE POISSON PARA ANÁLISE DAS  
VÍTIMAS FATAIS ENVOLVENDO CICLISTAS EM ARACAJU**

Apresentada em:

07/03/2017

Banca Examinadora:

---

Profª. Dra. Suzana Leitão Russo (UFS)  
Orientadora

---

Profª. MsC. Cristiane Toniolo Dias (UFS)  
1º Examinador

---

Profª. MsC. Marta Jeidjane Moura Borges (UFS)  
2º Examinador Externo

## AGRADECIMENTOS

Agradeço, primeiramente, a Deus por ter me iluminado até o fim dessa caminhada. Aos meus pais pelo investimento em minha formação e apoio incondicional.

Dedico esta conquista a minha orientadora, Profa. Dra. Suzana Leitão Russo pela orientação e auxílio na monografia e também por ter me concedido bolsas de Iniciação Científica que contribuíram para meu desenvolvimento acadêmico.

A todos os professores que me auxiliaram e incentivaram durante toda a minha formação. E também a todos os meus amigos e colegas de curso que contribuíram de maneira direta ou indireta para a conclusão dessa jornada.

*“Tenho espírito justiceiro e entendo que o amor deve seguir estes graus de preferência: Deus, humanidade, pátria, família e indivíduo” (D. Pedro II)*

## RESUMO

Estudos da Organização Mundial da Saúde apontam acidentes de trânsito como a nona causa de mortes externas no mundo. No Brasil, além dos homicídios, os acidentes de trânsito figuram entre as principais causas de mortes externas, na qual as categorias mais vulneráveis (pedestres, ciclistas e motociclistas) são as mais afetadas. Esse estudo busca encontrar um modelo de regressão de *Poisson* dos acidentes, que resultaram em vítimas fatais envolvendo ciclistas de Aracaju entre os períodos de 2006 até 2016. O presente trabalho é uma pesquisa básica quantitativa exploratória. Os dados foram obtidos no Núcleo de Estatística da Superintendência Municipal de Transportes e trânsito de Aracaju (SMTT). Para a elaboração das estatísticas básicas, histogramas e para encontrar o modelo de regressão de *Poisson* foi utilizado o *software* estatístico *STATISTICA*. O critério para a avaliação do modelo foi à observação da Escala de Deviance ( $G^2$ ) e da Escala de Pearson ( $X^2$ ), na qual os mesmos encontravam-se próximos de 1. O modelo apresentou um intercepto de -0,500775 e um erro padrão de apenas 0,111803. Além disso, foi realizada uma previsão em curto prazo para o primeiro trimestre de 2017.

**Palavras-chave:** acidentes de trânsito; vítimas fatais; ciclistas.

## ABSTRACT

Studies by the World Health Organization point to traffic accidents as a cause of no external deaths in the world. In Brazil, in addition to homicides, traffic accidents are among the main causes of external deaths, as the most vulnerable categories (pedestrians, cyclists and motorcyclists) are more affected. This study seeks to find a Poisson regression model of the accidents, which resulted in fatal victims involving cyclists from Aracaju between the periods 2006 to 2016. The present work is a basic exploratory quantitative research. The data were obtained without Statistical Nucleus of the Municipal Superintendence of Transports and Transit of Aracaju (SMTT). Statistical statistical software was used to compile the basic statistics, the studies and the Poisson regression model. The criterion for an evaluation of the model for the observation of the Deviance Scale (G2) and the Pearson Scale (X2), in which they approximate 1. The model presented for intercept of -0.500775 and a standard error of Only 0.111803. In addition, a short-term forecast was made for the first quarter of 2017.

**Keywords:** traffic-accidents; fatal victims; cyclists

## LISTA DE FIGURAS

FIGURA 1 – Distribuição Normal .....	21
FIGURA 2 – Histograma Conjunto de dados.....	29
FIGURA 3– Histograma sexo vítimas .....	30
FIGURA 4– Histograma idade vítimas .....	30
FIGURA 5– Teste de Aderência .....	32
FIGURA 6 – Gráfico Half-Normal.....	34
FIGURA 7 – Gráfico Resíduos Deviance.....	34
FIGURA 8 – Gráfico Resíduos Pearson.....	35

## LISTA DE TABELAS

TABELA 1 – Conjunto de Dados .....	27
TABELA - Distribuição de frequência conjunto de dados.....	28
TABELA 3 - Distribuição de frequência sexo vítimas.....	28
TABELA 4 – Distribuição de frequência idade vítimas.....	29
TABELA 5 – Estatísticas Descritivas.....	31
TABELA 6 – Parâmetros do modelo.....	32
TABELA 7 – Avaliação do modelo.....	33
TABELA 8 – Previsão.....	35



## LISTA DE ABREVIATURAS E SIGLAS

GL – Graus de Liberdade

MLG – Modelo Linear Generalizado

OMS – Organização Mundial da Saúde

ONU – Organização das Nações Unidas

SMTT – Superintendência Municipal de Transportes e Trânsito

$f_j$  – Frequência

$A_t$  – Amplitude total

$k$  – Número de classes

$\bar{X}$  – Média

$Md$  – Mediana

$S$  – Desvio-padrão

$S^2$  – Variância

$CV$  – Coeficiente de variação

$E(X)$  – Valor Esperado

$Var(X)$  – Variância

$\mu$  - Média Populacional

$\sigma$  – Desvio-padrão populacional

$H_0$  – Hipótese nula

$D$  – Desvio máximo

$\lambda$  - Parâmetro da distribuição *Poisson*

$L$  – Função de verossimilhança

$\text{Log } L$  – Função log-verossimilhança

$X^2_{n-p}(\alpha)$  – Distribuição Qui-Quadrado

$(X^T W X)^{-1}$  - Matriz Covariância Estimada

$\beta$  – Vetor parâmetros desconhecidos

$G^2$  - Função Desvio

## SUMÁRIO

RESUMO	v
ABSTRACT	vi
LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
LISTA DE SÍMBOLOS E ABREVIATURAS	ix
<b>1. INTRODUÇÃO</b>	<b>13</b>
<b>1.1. OBJETIVOS</b>	<b>15</b>
1.1.1. Geral	15
1.1.2. Específicos	15
<b>2. REVISÃO DA LITERATURA</b>	<b>16</b>
2.1. Análise Exploratória de Dados	16
2.1.1. Distribuição de Frequência	16
2.1.2. Histograma	17
2.1.3. Medidas de Tendência Central	18
2.1.4. Medidas de Dispersão	19
2.1.5. Erro Padrão	20
2.2. Distribuições de Probabilidade	20
2.2.1. Distribuição Normal	21
2.2.2. Distribuição de Poisson	22
2.3. Teste de Aderência	22
2.4. Modelos Lineares Generalizados	24
2.4.1. Modelo de Regressão de Poisson	24
<b>3. METODOLOGIA</b>	<b>27</b>
<b>4. RESULTADOS</b>	<b>28</b>
4.1. Análise Exploratória de Dados	28
4.1.1. Distribuição de Frequência	28
4.1.2. Histograma	28
4.1.3. Análise Descritiva	31
4.2. Teste de Aderência	31
4.3. Distribuição de Poisson	32
4.4. Modelo de regressão de Poisson	32
4.4.1. Parâmetros do Modelo	32
4.4.2. Avaliação do Modelo	33

4.4.3. Análise gráfica do modelo encontrado	33
4.4.4. Previsão	36
<b>5. CONCLUSÃO</b>	<b>36</b>
<b>6. REFERÊNCIAS</b>	<b>37</b>
ANEXO A – Valores críticos de Komolgorov	40

## 1. INTRODUÇÃO

O trânsito é definido pelo Código de trânsito Brasileiro (BRASIL, 1997), como sendo a utilização de vias terrestres por pessoas; animais e veículos para diversos fins. Atendo-se ao sentido semântico da palavra “acidente”, Oliveira (2011) caracteriza como sendo um evento inesperado ou uma tragédia. Logo, por silogismo, podem-se definir acidentes de trânsito como sendo qualquer evento inesperado ou uma tragédia ocorrida em vias terrestres, sejam por pessoas, veículos, animais, etc.

O site do Senado Federal divulgou um estudo realizado pela Organização Mundial da Saúde (OMS) apontando os acidentes de trânsito como a nona causa de mortes externas no mundo. Gonsaga et al. (2012), define mortes externas ou causas externas como qualquer tipo de traumatismo, lesão ou qualquer outro elemento que traga agravos à saúde, sejam intencionais ou não.

No Brasil, os homicídios e os acidentes de trânsito figuram entres os principais agentes causadores de mortes externas (MINAYO, 2009). Para Minayo (2009), a principal diferença entre essas duas variáveis é a de que os óbitos por homicídios estão distribuídas em pontos específicos e os óbitos provenientes de acidentes de trânsito estão distribuídos de forma dispersa por todo o território nacional.

A Organização das Nações Unidas afirma que os jovens são os mais afetados pela mortalidade no trânsito, jovens que possuem uma faixa etária entre 15 a 29 anos. Além do fato da maioria das vítimas serem compostas por: pedestres, ciclistas e motociclistas (ONU, 2015).

Em Aracaju essa realidade não seria diferente. As estatísticas em óbitos no trânsito não contrariam a tendência mundial, conforme o estudo apresentado pela ONU (2015). O Núcleo de Estatística da Superintendência Municipal de Transportes e Trânsito de Aracaju (SMTT, 2016) afirma que as maiores vítimas de acidentes de trânsito provêm das classes vulneráveis cujas mesmas são compostas por motociclistas, ciclistas e pedestres. As estatísticas da SMTT apontam que as maioria das vítimas, sejam fatais ou não, são pessoas com faixa etária entre os 19 e 44 anos, seguido por idosos com mais de 60 anos.

A Organização Mundial da Saúde (OMS) sugere que se considere como vítima fatal por acidente de trânsito qualquer óbito registrado até 30 dias após a ocorrência (EMDEC, SETRANSP).

Este trabalho busca analisar os acidentes de trânsito que resultaram em vítimas fatais envolvendo ciclistas ocorridos em Aracaju entre os anos de 2006 a 2016. Seguindo a sugestão

da Organização Mundial da Saúde (OMS), os óbitos analisados referem-se a vítimas de acidentes de trânsito que vieram a falecer em até 30 dias após as suas respectivas ocorrências. Utilizando-se da técnica de análise de regressão de Poisson busca-se verificar se haverá um acréscimo ou decréscimo nas vítimas fatais em Aracaju, analisando seu comportamento ao longo do ano de 2017.

Este estudo é de suma importância, pois poderá servir como um auxílio norteador a todos aqueles responsáveis e interessados em analisar os óbitos envolvendo ciclistas em Aracaju. Também poderá auxiliar se haverá ou não necessidades de políticas públicas, para o trânsito, mais eficientes.

## **1.1. OBJETIVOS**

### **1.1.1 Geral**

O objetivo deste trabalho é estudar o comportamento da série de vítimas fatais de ciclistas em Aracaju, por meio da análise dos modelos de regressão *Poisson*, verificando o a previsão à curto prazo

### **1.1.2 Específicos**

- Descrever os dados através de análises descritivas;
- Identificar a distribuição dos dados;
- Verificar os dados referentes às vítimas fatais de ciclistas na área delimitada de estudo;
- Verificar o comportamento da série;
- Modelar a série do número de vítimas fatais de ciclistas no período de janeiro de 2006 a dezembro de 2016, utilizando os modelos de regressão *Poisson*;
- Fazer uma previsão em curto prazo.

## **2. REVISÃO DA LITERATURA**

### **2.1. ANÁLISE EXPLORATÓRIA DE DADOS**

#### **2.1.1. Distribuição de Frequência**

Em um levantamento de dados, frequência é definida como o número de observações ou repetições de um valor. Uma tabela de frequências é uma tabela onde se procura fazer corresponder os valores observados da variável em estudo e as respectivas frequências. A tabela de frequências tem como vantagem, proporcionar uma apresentação mais estética dos dados (TOLEDO, OVALLE, 1992).

A distribuição de frequência permite uma leitura horizontal e vertical dos dados, na qual os dados são dispostos ordenadamente em linhas e colunas. (TOLEDO, OVALLE, 1992).

#### **Dados não agrupados em classes**

Toledo e Ovalle (1992) descrevem o método para se construir uma tabela de frequências para dados discretos ou descontínuos. Os autores determinam que a primeira coluna deva ser encabeçada pelo índice  $j$ , para corresponder à ordem dos valores das variáveis. Em seguida, os autores descrevem que a segunda coluna deve ser encabeçada por  $x_j$ , aonde serão alocados os valores da variável. Na terceira coluna, os autores a indicam como sendo uma coluna auxiliar a fim de computar a contagem dos valores repetidos. E por fim, a última coluna, Toledo e Ovalle (1992) determinam que essa deva ser encabeçada por  $f_j$ , aonde serão apresentados os resultados numéricos provenientes da contagem e a sua soma deve ser sempre igual ao número total de observações.

#### **Dados agrupados em classes**

Toledo e Ovalle (1992) descrevem um procedimento para a elaboração de dados agrupados em classes quando a variável do conjunto de dados é contínua ou se o número de valores representativos dessa variável for muito grande, evitando-se assim certos inconvenientes como, por exemplo, uma extensão desproporcional da tabela.

Para elaborar as classes, deve-se primeiro fazer o cálculo da amplitude total do número total das classes, com esse procedimento será possível elaborar os intervalos das classes da distribuição de frequência. O procedimento consiste no cálculo da amplitude total, no cálculo da quantidade de classes e no cálculo para a amplitude do intervalo da classe (TOLEDO, OVALLE, 1992).

Para o cálculo da amplitude total ( $A_t$ ), deve-se subtrair o maior valor do conjunto de dados com o menor valor (TOLEDO, OVALLE, 1992).

Para o cálculo do número de classes ( $k$ ) que a tabela de frequências irá conter Tolledo e Ovalle (1992), apresentam o Método de *Sturges* que consiste em:

$$k = 1 + 3,3 \log_{10} n$$

E por fim, Tolledo e Ovalle (1992), apresentam o método para se calcular a amplitude do intervalo da classe, como sendo:

$$\text{Amplitude do Intervalo da Classe} = \frac{A_t}{k}$$

### **2.1.2. Histograma**

Histogramas são gráficos, tipicamente de análises usados para representar uma distribuição de frequência simples. O histograma é um gráfico formado por um conjunto de retângulos justapostos, de forma que a área de cada retângulo seja proporcional à frequência da classe que ele representa. Assim sendo, a soma dos valores correspondentes às áreas dos retângulos será sempre igual à frequência total. (TOLEDO, OVALLE, 1992).

### **2.1.1. Medidas de Tendência Central**

Para Toledo e Ovalle (1992), as medidas de tendência central são as mais importantes das medidas de posição, dentre as quais se destacam a Média, a Mediana e a Moda.



## Média

A média é a medida de tendência central mais comumente usada para descrever resumidamente uma distribuição de frequências. A média aritmética simples é uma das mais conhecidas e frequentemente uma das mais utilizadas. Há outros tipos de média, além da média aritmética simples, como por exemplo: média aritmética ponderadas, geométrica ou harmônica, etc. (TOLEDO, OVALLE, 1992).

A **Média Aritmética Simples** é genericamente escrita como:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Onde  $X_i$  é o valor genérico das observações e  $n$  é o número de observações (TOLEDO, OVALLE, 1992).

## Mediana

A Mediana é uma separatriz que divide os dados pela metade (cinquenta por cento), em partes iguais (TOLEDO, OVALLE, 1992).

Segundo Toledo e Ovalle (1992) caso o conjunto de dados apresente um número de observações resultem em um número par, o procedimento para o cálculo da mediana deve, segundo os autores:

$$E_{Md} = \frac{n}{2}$$

Se o número de observações no conjunto de dados resultar em um número ímpar, Toledo e Ovalle (1992) descrevem que o procedimento para o cálculo da mediana deve ser:

$$E_{Md} = \frac{n+1}{2}$$

## Moda

A moda ( $M_o$ ) é definida como o valor mais frequente que se repete em um conjunto de dados. Podendo haver conjuntos que são plurimodais ou amodais. Quando houver mais de um valor predominante, o conjunto de dados será classificado como plurimodal e quando houver ausência de um valor mais frequente, o conjunto de dados será classificado como amodal (TOLEDO, OVALLE, 1992).

### 2.1.2. Medidas de Dispersão

Toledo e Ovalle (1992) ressaltam que apenas uma análise das medidas de tendências centrais de um conjunto de dados apresentaria uma descrição inadequada dos dados, pois os dados diferem entre si. Logo, se torna importante avaliar o grau dessa dispersão dos dados. Dois tipos de medidas de dispersão a serem considerados são: as Medidas de Dispersão Absoluta na qual se destaca a variância e o desvio-padrão e as Medidas de Dispersão Relativa na qual se tem os coeficientes de variação de *Pearson* e de *Thorndike*.

### Desvio- Padrão

Segundo Toledo e Ovalle (1992) é a medida de dispersão mais utilizada, e a descrevem como:

$$s = \sqrt{\frac{\sum_i^N (X - \bar{X})^2}{N - 1}}$$

### Variância

Toledo e Ovalle (1992) definem a variância como sendo o quadrado do desvio-padrão, e a descrevem como:

$$s^2 = \frac{\sum_i^N (X - \bar{X})^2}{N - 1}$$

### 2.1.3. Erro Padrão

O erro padrão representa a variabilidade média entre as médias amostrais possíveis de serem coletadas. O erro padrão oferece, intuitivamente, uma noção da precisão da estimativa obtida para a média, sendo que aquela que apresentar maior erro padrão terá menos precisão (RUSSO, 2017).

De acordo com Russo (2017), o erro padrão é genericamente escrito como:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

## 2.2. DISTRIBUIÇÃO DE PROBABILIDADE

### 2.2.1. Distribuição Normal

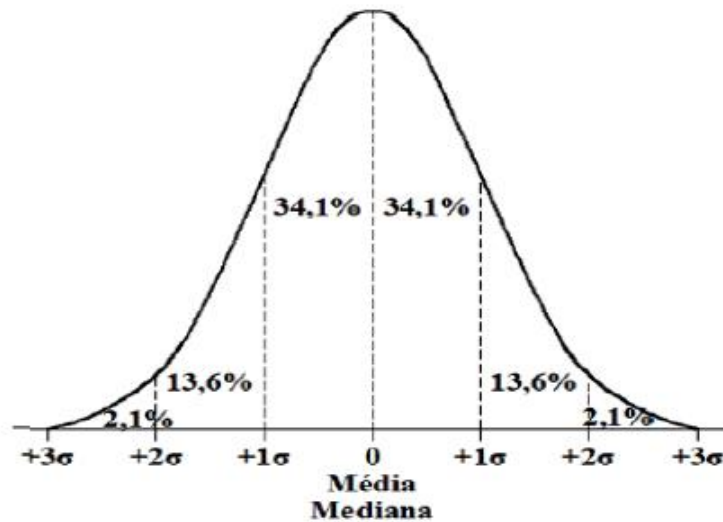
De acordo com Lopes et. al (2013), a distribuição Normal é umas das distribuições mais importantes distribuições de probabilidades, podendo ser conhecida como Distribuição de Gauss ou Gaussiana, cuja mesma é caracterizada pelos parâmetros de média populacional  $\mu$  e pelo desvio-padrão populacional  $\sigma$ . Lopes afirma ainda que quando o conjunto dos dados segue uma distribuição Normal, sua média irá se encontrar no centro da distribuição. Sendo que a média seguirá com os mesmos valores que a mediana e a moda, dando assim uma forma simétrica ao seu gráfico.

Lopes et al (2013) apresenta a formula da distribuição Normal, como sendo:

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}, x \in R$$

A Figura 1 representa o gráfico da distribuição Normal, aonde a proporção dos valores sob a curva é a seguinte: 68,26% dos valores estão dentro de um desvio padrão positivo e negativo de sua média ( $1\sigma$ ); 95,44% dos valores estão dentro de dois desvios padrão positivo e negativo ( $2\sigma$ ) de sua média e 99,72% estão dentro de três desvios padrão positivo e negativo ( $3\sigma$ ) de sua média (LOPES, M.; SOARES, J. B.; BRANCO, V. C.).

Figura 1 – Distribuição Normal



Fonte: Elaborado por Lopes et.al (2013)

Russo (2017) define o valor esperado da distribuição Normal ( $E(x)$ ) como sendo igual a 0. Em relação à variância ( $Var(X)$ ) da distribuição Normal, Russo (2017) define, como sendo igual a 1.

### 2.2.2. Distribuição Poisson

Uma variável aleatória de Poisson, é qualquer variável aleatória  $X$ , cujos valores podem ser de 0, 1, 2, ..... Tendo como parâmetro o  $\lambda$ , cujo parâmetro pode ser interpretado como sendo uma taxa média de ocorrência do evento. Para um  $\lambda > 0$ , tem-se:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

A Distribuição de *Poisson* foi introduzida por Siméon Denis Poisson em 1837 introduziu a distribuição de probabilidades de Poisson em 1837. A variável aleatória de Poisson pode ser aplicada em diversas áreas, pois a mesma pode ser utilizada como uma aproximação para a variável aleatória binomial com parâmetros  $(n, p)$ , no caso em que o  $n$  é grande e  $p$  é suficientemente pequeno para que  $np$  tenha tamanho moderado (ROSS, 2010).

Russo (2017) define o valor esperado ( $E(X)$ ) da distribuição de *Poisson*, como sendo igual a  $\lambda$ . Em relação à variância ( $Var(X)$ ) da distribuição de *Poisson* define-se, como sendo igual também a  $\lambda$ .

### 2.3. TESTE DE ADERÊNCIA

A estimação e os testes de hipóteses são uma dos exemplos da grande quantidade dos métodos estatísticos para supor que seus dados provêm de uma distribuição Normal. A verificação da normalidade do conjunto de dados pode ser feitas por vários testes como o de *Komolgorov-Sminorv* (LOPES, M.; SOARES, J. B.; BRANCO, V. C).

#### Teste de Komolgorov-Sminorv

O teste de *Komolgorov-Sminorv* é um teste de aderência na qual ele é concernente com o grau de concordância entre a distribuição de um conjunto de valores da amostra (escores observados) e alguma distribuição teórica observada. Em síntese, o teste busca especificar a distribuição de frequência acumulada que ocorreria dada a distribuição teórica e compará-la com a distribuição de frequência acumulada observada (SIEGEL, JUNIOR CASTELLAN, 2006).

A distribuição teórica representa o que seria esperado sobre a hipótese nula ( $H_0$ ). O ponto na qual essas duas distribuições, teórica e observada, mostram a maior divergência é determinado. A distribuição amostral indica se uma tão grande divergência é provável de ocorrer com base no acaso. Isto é, a distribuição amostral indica a possibilidade de que ocorresse uma divergência da magnitude observada se as observações fossem realmente uma amostra aleatória de uma distribuição teórica. O teste de *Komolgorov-Sminorv* admite que a distribuição da variável subjacente que está sendo testada é contínua, como especificado pela distribuição de frequências acumuladas. Assim, o teste é apropriado para testar a aderência para variáveis que são medidas pelo menos em uma escala ordinal (SIEGEL, JUNIOR CASTELLAN, 2006).

Seja  $F_0(X)$  uma função completamente especificada de distribuição de frequências relativas acumuladas – a distribuição teórica sob  $H_0$ . Isto é, para qualquer valor de  $X$ , o valor de  $F_0(X)$  é a proporção de casos esperados com escores iguais ou menores do que  $X$ . Seja  $S_N(X)$  a distribuição de frequências relativas acumuladas observadas de uma amostra aleatória

de  $N$  observações. Se  $X_i$  é um escore qualquer possível então  $S_N(X) = F_i/N$ , onde  $F_i$  é o número de observações menores ou iguais a  $X_i$  (SIEGEL, JUNIOR CASTELLAN, 2006).

Agora sob, a hipótese nula de que a amostra tenha sido extraída de uma distribuição teórica especificada, é esperado que para qualquer valor de  $X_i$ ,  $S_N(X_i)$  esteja bastante próximo de  $F_0(X_i)$ . Isto é, quando a hipótese nula,  $H_0$ , é verdadeira, esperaríamos que as diferenças entre  $S_N(X_i)$  e  $F_0(X_i)$  fossem pequenas e dentro dos limites de erros aleatórios. O teste de *Komolgorov-Sminorv* focaliza sobre o maior dos desvios. O maior valor absoluto de  $F_0(X_i) - S_N(X_i)$  é chamado “desvio máximo D” (SIEGEL, JUNIOR CASTELLAN, 2006).

Siegel e Junior Castellan (2006) descrevem o desvio máximo D como sendo:

$$D = \max |F_0(X_i) - S_N(X_i)| \quad i=1,2,\dots,N$$

A distribuição amostral de D sob  $H_0$  é conhecida. A tabela A do Apêndice fornece certos valores críticos da distribuição amostral. A significância de um dado valor de D depende de N (SIEGEL, JUNIOR CASTELLAN, 2006).

## 2.4. MODELOS LINEARES GENERALIZADOS

Os GLM são apresentados por Nelder e Wedderburn (1972), na qual os mesmos são de grande importância na Estatística, pois generalizam o modelo de regressão normal, comumente já conhecido garantindo maiores possibilidades para a distribuição da variável resposta e dando maior flexibilidade entre a média e a parte sistemática do modelo (DOBSON, 2002; RUSSO, 2002)

De acordo com Cordeiro (2010), a seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística e envolve a procura de um modelo que seja o mais simples possível e que descreva bem o processo gerador dos valores observados que surgem em diversas áreas do conhecimento como agricultura, demografia, ecologia, economia, engenharia, geologia, medicina, ciência política, sociologia e zootecnia, entre outras. Esses modelos envolvem uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de  $n$  observações independentes, sendo que:

- a variável resposta, **componente aleatório** do modelo, tem uma distribuição pertencente a família de distribuições que engloba as distribuições normal, gama e

normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens.

- as variáveis explanatórias entram na forma de uma estrutura linear, constituindo o **componente sistemático** do modelo;
- a ligação entre os componentes aleatório e sistemático é feita por meio de uma função adequada como, por exemplo, logarítmica para os modelos log-lineares, denominada **função de ligação**.

#### 2.4.1. Modelo de Regressão de Poisson

Em Modelos Lineares Generalizados, o modelo de regressão de Poisson é o mais utilizado depois da regressão Logística. Os Modelos de Poisson são utilizados quando a variável resposta é uma contagem, tal como o número de eventos ocorridos num espaço de tempo (RUSSO, 2002).

Para Cordeiro (1992), o modelo de regressão de Poisson é um tipo específico de Modelos Lineares Generalizados e o método de máxima verossimilhança pode ser utilizado para estimar os parâmetros do modelo de regressão de Poisson. A função de verossimilhança é dada por:

$$L = \prod_{i=1}^n \Pr(Z_i / \lambda_i) = \prod_{i=1}^n \frac{e^{-\lambda_i} \cdot \lambda_i^{Z_i}}{Z_i!}$$

e a função log-verossimilhança, igual a  $\log L = \sum (Z_i \cdot \log(\lambda_i) - \lambda_i) - \sum \log(Z_i!)$

A função constante de  $Z$ , dada por  $\sum \log(Z_i!)$ , pode ser omitida, pois não envolve  $\lambda$  (FERRARI, 2002; FENGLER et. al., 2012).

A existência de uma função de ligação  $\log(\lambda_i)$  entre as médias das observações e a estrutura linear do modelo dada por  $\log(\lambda_i) = \beta x_i^T$ , pode ser admitida pela componente sistemática. Com isso, os valores ajustados de  $\lambda_i$ , assegurados pela função de ligação log, permanecem no intervalo  $[0, \infty)$ . O preditor linear  $\beta x_i^T$  será relacionado ao valor esperado

$\lambda_i$  do vetor  $Z_i$ , pela função de ligação log. O modelo Poisson com ligação log é algumas vezes chamado de modelo log-linear (SCHAFER, 1997; FENGLER et. al., 2012).

Transformando a função de ligação log obtêm-se a seguinte expressão para a variável resposta:  $\lambda = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}$

onde  $\beta x_i^T$  é o preditor linear,  $\beta = (\beta_1, \dots, \beta_k)^T$  é o vetor de parâmetros desconhecidos a serem estimados e  $x_i = (x_{i1}, \dots, x_{ik})^T$  representa os valores de  $k$  variáveis explicativas (PAULA, 1997; RUSSO 2002).

As distribuições normais, binomial, Poisson, gamma, normal inversa, e binomial negativa são membros da família exponencial (LEWIS, MONTGOMERY, MYERS, 2001; RUSSO, 2002).

**Estatísticas de interesse** – De acordo com Russo (2002), após a convergência (que pode ser feita através do algoritmo de Newton-Raphson), deve-se examinar as seguintes estatísticas:

- $(X^T W X)^{-1}$ , a matriz de covariância estimada para  $\hat{\beta}$ ;
- a função de log-verossimilhança  $l = l(Z, \lambda) = \sum_{i=1}^N \{Z_i \log \lambda_i - \lambda_i\}$

Para verificar se é satisfatória a relação funcional variância/média proposta para os dados, pode-se utilizar um gráfico de resíduos padronizados versus valores ajustados, sem nenhuma tendência. Gráficos dos resíduos versus covariáveis que não estão no modelo são bastante úteis. Se nenhuma covariável adicional é necessária, então não se deve encontrar qualquer tendência nesses gráficos. Dados com grande resíduos indicam que pode haver erros grosseiros ou o modelo ajustado necessita de mais covariáveis, como interações de ordem superior, por exemplo. A inspeção gráfica é um meio poderoso de inferência nos GLMs (PIEGORSCH, 1998; FENGLER et. al., 2012).



Um gráfico dos resíduos padronizados versus valores ajustados, sem nenhuma tendência, é um indicativo de que a relação funcional variância/média proposta para os dados é satisfatória. Gráficos dos resíduos versus covariáveis que não estão no modelo são bastante úteis. Se nenhuma covariável adicional é necessária, então não se deve encontrar qualquer tendência nesses gráficos. Dados com erros grosseiros podem ser detectados tendo grandes resíduos, ou o modelo ajustado deve requerer mais covariáveis, por exemplo, interações de ordem superior. A inspeção gráfica é um meio poderoso de inferência nos GLMs (PIEGORSCH, 1998; FENGLER et. al., 2012).

Os graus de liberdade associados ao desvio são definidos por  $\nu = n - p$ . Para o teste de um modelo compara-se o  $G^2$  e seus graus de liberdade  $\nu$  com uma distribuição teórica de probabilidade. Geralmente, adota-se a distribuição de qui-quadrado. Na prática, se contenta em testar um modelo (sem muito rigor) comparando o desvio com o valor crítico  $\chi^2_{n-p}(\alpha)$  da distribuição de qui-quadrado à um nível de significância igual a  $\alpha$ . Se este for maior que  $\chi^2_{n-p}(\alpha)$ , o modelo será rejeitado e, caso seja menor ou igual, aceito (SCHAFER, 1997; FENGLER et. al., 2012).

Deve-se esperar que um modelo bem ajustado aos dados tenha um desvio próximo dos seus graus de liberdade. Se o desvio  $G^2$  ou o  $\chi^2$  excedem o valor dos seus graus de liberdade, diz-se que o modelo é inadequado podendo tratar-se de um problema de superdispersão (WANG et al, 1996; FENGLER et. al., 2012).

Para se avaliar a existência de superdispersão deve-se empregar um critério uniforme. Adota-se que o desvio  $G^2$  excede o valor crítico  $\chi^2_{n-p}(\alpha)$  até 10%, caso seja maior o modelo será rejeitado. O efeito desta correção é mínimo nos estimadores pontuais. Este método está baseado na generalização de Breslow (1984) para modelos de Poisson.

### 3. METODOLOGIA

O presente trabalho é uma pesquisa básica quantitativa e exploratória. Para a consecução dos objetivos propostos, utilizou-se uma abordagem qualitativa e quantitativa de exploração bibliográfica e documental com o intuito de analisar a série de vítimas fatais em Aracaju, e informações disponíveis acerca desses registros.

A fonte dos dados provém do Núcleo de Estatística da Superintendência Municipal de Transportes e Trânsito de Aracaju (SMTT), referentes ao período de janeiro de 2006 até dezembro de 2016, os quais correspondem as vítimas fatais (até 30 dias) dos ciclistas de Aracaju.

Foi empregada uma análise exploratória, seguido pelo Teste de Komolgorv-Sminorv e a verificação dos parâmetros para identificar a distribuição de *Poisson*, após para analisar com mais eficácia o estudo, aplicou-se o modelo de Regressão de *Poisson*.

O software utilizado, para todos os cálculos e análises estatísticas, foi o STATISTICA 11.

A Tabela 1 apresenta o conjunto de dados referente aos ciclistas que foram vítimas fatais ocorridos entre os anos de 2006 a 2016, em Aracaju.

Tabela 1 – Conjunto de dados

CICLISTAS	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Janeiro	4	1	0	1	1	1	0	0	0	0	0
Fevereiro	0	1	0	0	0	1	0	0	1	0	1
Março	0	2	0	1	0	0	1	0	0	0	0
Abril	1	1	0	0	1	0	1	0	1	1	1
Mai	3	0	1	1	1	0	0	0	0	1	0
Junho	0	0	2	0	0	1	2	0	1	1	0
Julho	0	0	0	2	1	0	0	0	0	0	0
Agosto	1	0	0	1	2	0	0	0	1	0	1
Setembro	3	2	2	0	1	0	1	1	0	1	1
Outubro	1	0	0	1	1	1	0	0	1	2	1
Novembro	2	0	1	0	1	0	0	0	0	1	0
Dezembro	0	0	0	1	1	2	1	0	1	0	1
<b>TOTAL</b>	<b>15</b>	<b>7</b>	<b>6</b>	<b>8</b>	<b>10</b>	<b>6</b>	<b>6</b>	<b>1</b>	<b>6</b>	<b>7</b>	<b>6</b>

Fonte: Elaborado pelo autor (2017)

## 4. RESULTADOS

### 4.1. ANÁLISE EXPLORATÓRIA DOS DADOS

#### 4.1.1. Distribuição de frequência

A Tabela 2 apresenta a distribuição de frequência do conjunto de dados. Analisando os 132 meses (2006 até 2016) em que houve ocorrência de vítimas fatais, tem-se: 69 meses não apresentaram vítimas fatais, 50 meses apresentaram uma vítima fatal, 10 meses apresentaram duas vítimas fatais, 2 meses apresentaram três vítimas fatais e 1 mês apresentou 4 vítimas fatais. O mês que apresentou mais vítimas fatais foi o mês de janeiro de 2006.

Tabela 2 – Distribuição de frequência do conjunto de dados

$j$	Quantidade de óbitos de ciclistas $x_j$	Quantidade de meses $f_j$
1	0	69
2	1	50
3	2	10
4	3	2
5	4	1

Fonte: Elaborado pelo autor (2017)

A Tabela 3 apresenta a distribuição de frequência do conjunto de dados, analisando o sexo das vítimas fatais. Dos 80 ciclistas que vieram a óbito, 75 eram do sexo masculino e apenas 5 eram do sexo feminino.

Tabela 3 – Distribuição de frequência do sexo das vítimas

$j$	Sexo das vítimas $x_j$	Quantidade de óbitos $f_j$
1	Masculino	75
2	Feminino	5

Fonte: Elaborado pelo autor (2017)

A Tabela 4 apresenta a distribuição de frequência do conjunto de dados, analisando a idade das vítimas.

Tabela 4 – Distribuição de frequência da idade das vítimas

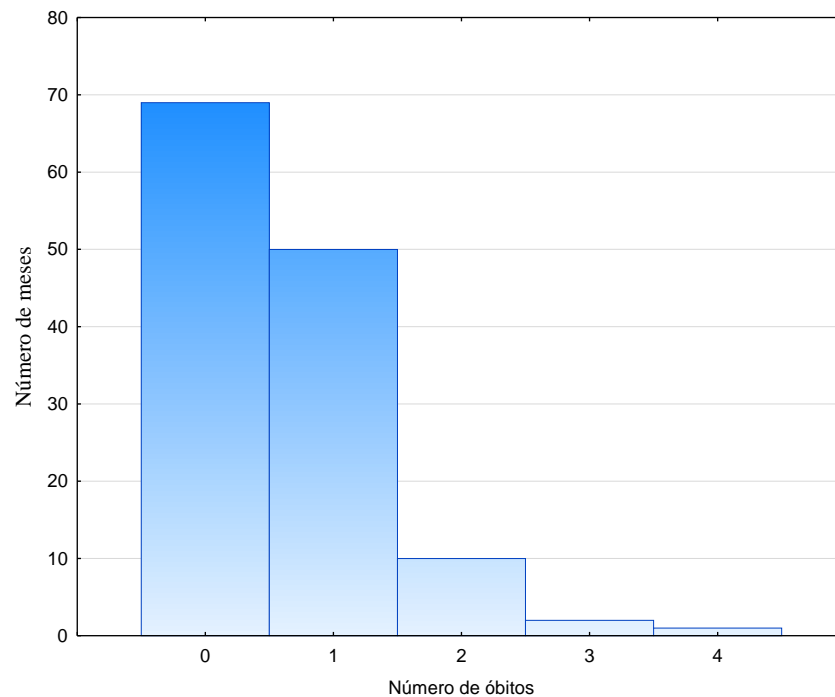
$k$	Idade das vítimas $x_k$	Quantidade de óbitos $f_k$
1	5   - 15	3
2	15   - 25	10
3	25   - 35	21
4	35   - 45	16
5	45   - 55	11
6	55   - 65	15
7	65   - 75	3
8	75   - 85	1

Fonte: Elaborado pelo autor (2017)

#### 4.1.2. Histograma

A Figura 2 apresenta o histograma do conjunto de dados.

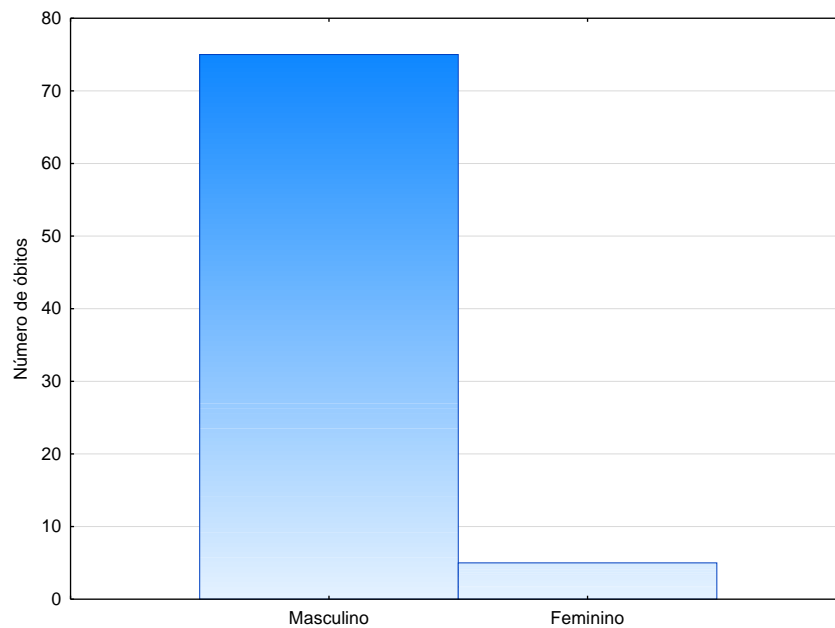
Figura 2 – Histograma do conjunto de dados



Fonte: Elaborado pelo autor (2017)

A Figura 3 apresenta o histograma do sexo das vítimas.

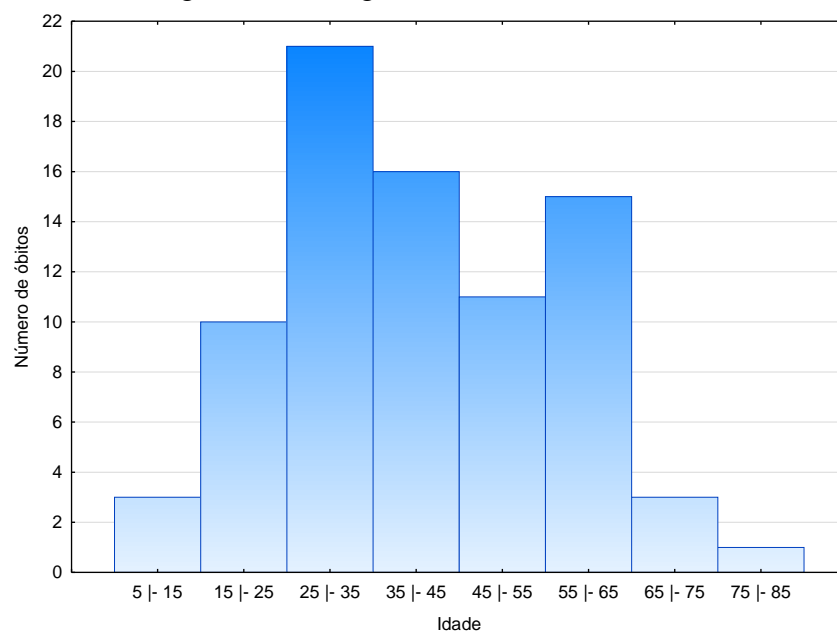
Figura 3 – Histograma do sexo das vítimas



Fonte: Elaborado pelo autor (2017)

A Figura 4 apresenta o histograma da idade das vítimas, aonde, percebe-se que a maiores vítimas são os ciclistas com idade entre 25 e 35 anos seguidos de ciclistas com idade entre 35 e 45 anos.

Figura 4 – Histograma da idade das vítimas



Fonte: Elaborado pelo autor (2017)

### 4.1.3. Análise Descritiva

A Tabela 5 apresenta as Estatísticas Descritivas. A Tabela 5 apresentou valores mínimos e máximos de 0 e 4, respectivamente, indicando o mínimo e máximo de vítimas fatais que podem ocorrer ao longo da série. Também, apresentou uma média de 0,6 vítimas fatais por mês e um desvio-padrão de 0,75.

Tabela 5 – Estatísticas Descritivas

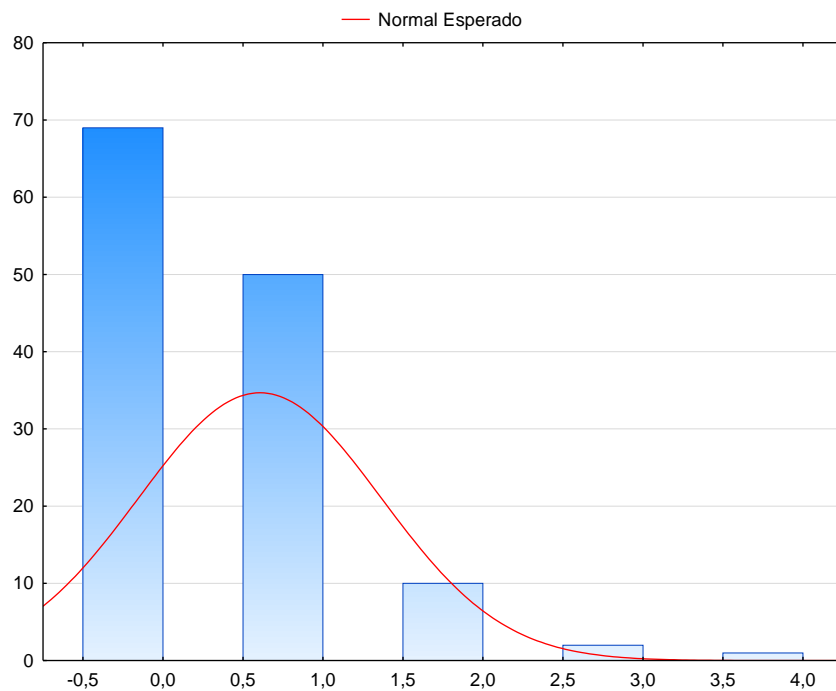
Dados	Mínimo	Máximo	Média	Desvio-Padrão	Mediana	Moda	CV
Total	0	4	0,6	0,75	0	0	125

Fonte: Elabora pelo autor (2017)

### 4.2. TESTE DE ADERÊNCIA

Aparentemente os dados não apresentam uma tendência sobre o tempo, o que se confirmará mais tarde. Foi realizado o teste de aderência, *Komolgorov-Sminorv*, para verificar se os dados seguem uma distribuição Normal. A Figura 5 apresenta o histograma da série e o gráfico esperado da distribuição Normal, apresentando um desvio máximo (D) de 0,3104. Considerando um nível de significância de 5% e fazendo-se o cálculo da regra de *Komolgorov-Sminorv*, o resultado obtido foi 0,11837. Pelo critério do Teste de *Komolgorov-Sminorv*, o resultado obtido foi menor que o desvio-máximo, logo a hipótese nula ( $H_0$ ) não é satisfeita, com isso pode-se concluir que o conjunto de dados não segue uma distribuição Normal.

Figura 5 – Teste de aderência



Fonte: Elaborado pelo autor (2017)

### 4.3. DISTRIBUIÇÃO DE POISSON

Pelos testes de *Komolgorov-Sminorv* e *Lilliefors* verificou-se que os dados não seguem uma distribuição Normal, logo precisa-se verificar se os dados seguem uma distribuição de *Poisson*. A média e desvio-padrão do conjunto de dados foram, respectivamente, 0,6 e 0,75. Pelo fato da média e desvio-padrão serem valores muito próximos, satisfaz a condição da distribuição de *Poisson*. Com todas as premissas verificadas e, determinando-se que os dados seguem uma distribuição *Poisson*, é possível aplicar o modelo de Regressão de *Poisson*.

### 4.4. MODELO DE REGRESSÃO DE POISSON

#### 4.4.1. Parâmetros do Modelo

A Tabela 6 apresenta os parâmetros do modelo de Regressão de *Poisson*.

Tabela 6 – Parâmetro do modelo

Efeito	Estimativa	Erro Padrão	Limite Inferior	Limite Superior	p
Intercepto	-0,5008	0,1118	-0,7199	-0,2816	0,0000
Escala	1	0	1	1	

Fonte: Elaborado pelo autor (2017)

O algoritmo de Newton-Raphson convergiu em 3 iterações. O modelo encontrado foi:

$$Y = -0,5008_{(0,1118)} + erro$$

#### 4.4.2. Avaliação do Modelo

Segundo Russo (2002) a média e a variância numa distribuição de Poisson são iguais, isso implica que a Estatística de Pearson ( $X^2$ ), dividido pelo grau de liberdade, deve ser aproximadamente 1 (um). A Estatística de Deviance ( $G^2$ ) e de Pearson ( $X^2$ ) dividido pelos respectivos graus de liberdade são usados para detectar se há superdispersão ou subdispersão. Valores maiores que um indica uma superdispersão, isto é, a variância real é maior do que a média, valores menores do que um, indicam subdispersão, a variância real é menor do que a média.

A Tabela 7 apresenta o critério de avaliação do modelo. Observa-se que a Escala de Deviance ( $G^2$ ) e Escala de Pearson ( $X^2$ ) estão adequados ao modelo, pois os seus respectivos graus de liberdade são iguais ou próximos a 1.

Tabela 7 – Avaliação do Modelo

<b>Crítérios</b>	<b>GL</b>	<b>Valores</b>	<b>Valores/GL</b>
Escala Deviance ( $G^2$ )	131	132,12	1,0855
Escala de Pearson ( $X^2$ )	131	124,60	0,9511
Verossimilhança	131	-133,75	

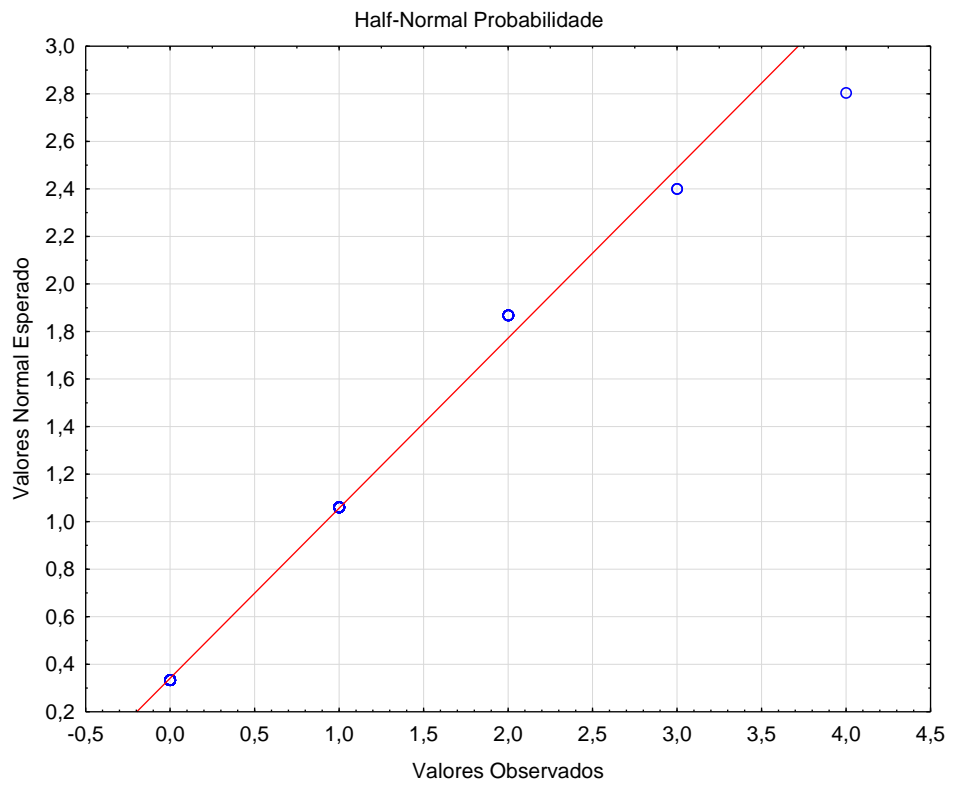
Fonte: Elaborado pelo autor (2017)

#### 4.4.3 Análise gráfica do Modelo Encontrado

Foi feito o gráfico da Half-Normal do modelo, a fim de verificar se há ou não presença de *outlier*. A Figura 6 apresenta o gráfico da Half-Normal, o gráfico aponta a ausência de *outliers*.

Figura 6 – Half-Normal

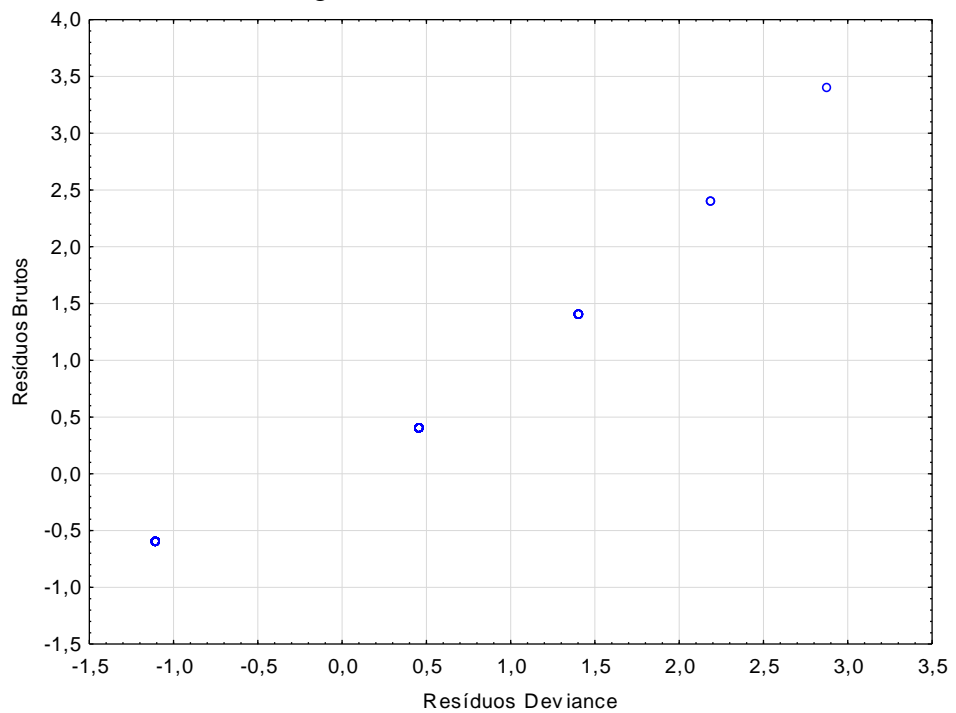




Fonte: Elaborado pelo autor (2017)

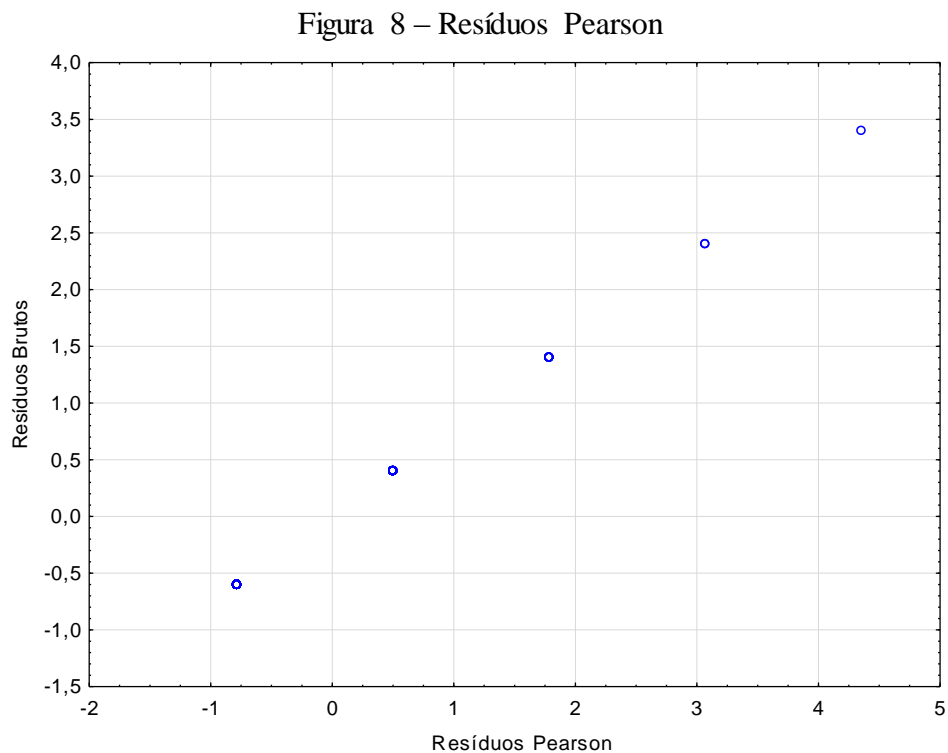
A Figura 7 apresenta o gráfico dos resíduos da Deviance a fim de verificar se há ou não presença de *outlier*. O gráfico aponta a ausência de *outlier*.

Figura 7 – Resíduos Deviance



Fonte: Elaborado pelo autor (2017)

A Figura 8 apresenta o gráfico dos resíduos de Pearson a fim de verificar se há ou não presença de *outlier*. O gráfico aponta a ausência de *outlier*.



Fonte: Elaborado pelo autor (2017)

#### 4.4.4. Previsão

Depois de cumpridas as etapas da avaliação e verificação do modelo. Pode-se fazer uma estimativa em curto prazo. A Tabela 8 apresenta a previsão para o primeiro trimestre de 2017. O modelo prevê 4 óbitos para o mês de janeiro e nenhum óbitos para os meses de fevereiro e março.

Tabela 8 – Previsão

<b>Mês/2017</b>	<b>Previsão</b>
Janeiro	4
Fevereiro	0
Março	0

Fonte: Elaborado pelo autor (2017)

## 5. CONCLUSÃO

Esse trabalho analisou o perfil dos ciclistas, vítimas de acidentes fatais (até 30 dias após a ocorrência do acidente) ocorridos em Aracaju entre o período de 2006 até 2016. Além das estatísticas básicas utilizou-se o modelo de regressão de *Poisson* para avaliar e estudar esses acidentes. Durante esse período foram computados 80 óbitos na qual 75 são ciclistas do sexo masculino e 5 são do sexo feminino. Além do sexo, computaram-se os acidentes por faixas etárias das vítimas na qual o maior índice de óbitos provém de ciclistas com faixa etária entre 25 a 35 anos, seguido de ciclistas com faixa etária entre 35 a 45 anos. O perfil da maioria das vítimas é composto, em sua grande maioria, por homens com faixa etária entre 25 a 45 anos.

Além disso, esse trabalho identificou que os acidentes fatais, envolvendo ciclistas, segue uma distribuição *Poisson* na qual foi possível, através do Modelo de Regressão de *Poisson*, elaborar um modelo que melhor descreve o comportamento desses acidentes. Além disso, foi realizada uma previsão, utilizando-se do modelo encontrado, aonde conseguiu-se prever os acidentes para o primeiro trimestre, resultando em 4 óbitos para janeiro de 2017 e nenhum óbito para os meses de fevereiro e março.

Faz-se necessário uma maior expansão das ciclovias em Aracaju, pois a principais vítimas eram homens jovens e adultos em plena idade de trabalho. Além disso, ressalta-se de que a bicicleta é considerado um dos meios de transporte mais saudáveis e não poluentes.

## 6. REFERÊNCIAS

BRASIL. *Código de Trânsito Brasileiro (CTB)*. Lei Nº 9.503, de 23 de Setembro de 1997 que institui o Código de Trânsito Brasileiro. 1997.

CORDEIRO, G. M. **Introdução à Teoria de verossimilhança**. Livro Texto do 10º Simpósio Nacional de Probabilidade e Estatística. UFRJ/ABE. Rio de Janeiro. 1992.

CORDEIRO, G. M. **Modelos Lineares Generalizados**. São Paulo, Campinas UNICAMP/UFPE. 2010.

DOBSON, A, J. **An introduction to generalized linear models**. 2 ed. CHAPMAN & HALL / CRC Press pp. 225.2002.

EMDEC, SETRANSP: Disponível em: <  
<http://www.emdec.com.br/eficiente/sites/portalemdec/pt-br/site.php?secao=Acidentalidade> >

Acesso: 10 de jan. 2017.

FERRARI, S. L. P.; DAVID, J.S. E.; ANDRÉ, P. A.; PEREIRA, L.A. A. **Use of overdispersed regression models in analyzing the association between air pollution and human health**. Relatório Técnico, RTMAE-2002-10, IME-USP. 2002.

GONSAGA, R. A. T.; RIMOLI, C. F.; PIRES, E. A.; ZOGHEIB, F. S.; FUJINO, V. T.; CUNHA, M. B. **Avaliação da mortalidade por causas externas**. Ver. Col. Bras. Cir. vol.39 no.4 Rio de Janeiro July/Aug. 2012.

LEWIS, S.; MONTGOMERY, D.C.; MYERS,R.H. **Examples of designed experiments with nonnormal responses**. Journal of Quality Technology V33 n.3 pp.265-278. 2001

LOPES, M. M.; BRANCO, V. T. F. C.; SOARES, J. B. **Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação**. TRANSPORTES v.21, n.1, 2013.

McCULLAGH, P. ;NELDER, J. A. **Generalized Linear Models**. Third Edition. New York: Chapman and Hall/CRC. Reprint. 1989.

MINAYO, M. C. S. **Seis características das mortes violentas no Brasil**. Rev. bras. estud. popul. vol. 26 no.1 São Paulo Jan/June 2009.

NELDER, J., A.; WEDDERBURN, R., W., M. **Generalized linear models**. Journal of Royal Statistical Society: v. 135, pp. 370-384. 1972.

OLIVEIRA, A. C. *Língua Portuguesa*. 1.ed. Blumenau: Vale das Letras, 2011.

ONU. Disponível em: < <https://nacoesunidas.org/onu-acidentes-no-transito-ainda-matam-125-milhao-por-ano-90-em-paises-de-renda-media-e-baixa/> > Acesso: 10 de jan. 2017.

PAULA, G. A. **Estimação e testes em modelos de regressão com parâmetros restritos**. Livro Texto da 5ª Escola de Modelos de Regressão. IME-USP/ABE. Campos do Jordão. 1997

PIEGORSCH, W., W. **An introduction to binary response regression and associated trend analyses**. Journal of Quality Technology. v.30, nº.3, pp. 269-281. 1998.

ROSS, S. **Probabilidade: um curso moderno com aplicações**. Tradução de Alberto Resende De Conti. 8ª ed. Porto Alegre: Bookman, 2010.

RUSSO, S. L. **Graficos de Controle Para Variáveis Não-Conformes Autocorrelacionadas**. UFSC. Florianopolis. 2002.

RUSSO, S. L.; FLENDER, D.; SILVA, G. F. **Poisson Regression Models for Count Data: Use in the Number of Deaths in the Santo Angelo (Brazil)**. Journal of Basic & Applied Sciences, 2012.

RUSSO, S. L. **Estatística Aplicada**. UFS. São Cristóvão. 2017.

SCHAFFER, J., L. **Analyses of incomplete multivariate data**. London Chapman & Hall. 1997.

SENADO FEDERAL. Disponível em: <  
<http://www.senado.gov.br/noticias/Jornal/emdiscussao/motos/saude/estudo-da-organizacao-mundial-da-saude-oms-sobre-mortes-por-acidentes-de-transito-em-178-paises-e-base-para-decada-de-aco-es-para-seguranca.aspx> > Acesso: 10 de jan. de 2017.

SIEGEL, S.; CASTELLAN JUNIOR, N. J. **Estatística Não-Paramétrica para Ciências do Comportamento**. Tradução de Sara Ianda Correa Carmona. 2ª ed. Porto Alegre: Artmed, 2006.

SMTT. Superintendência Municipal de Transportes e Trânsito da cidade de Aracaju. **Núcleo de Estatística**.

TOLEDO, G. L.; OVALLE, I. I. **Estatística Básica**. 2ª ed. São Paulo, SP: Atlas, 1992. 459 p.

WANG, P.; PUTERMAN, M., L.; COCKBURN,; LE, N. **Mixed Poisson regression models with covariate dependent rates**. Biometrics. v. 52, pp. 381-400. 1996.

**Tabela A – Valores Críticos de D no teste de uma amostra de *Komolgorov-Sminorv***

<b>Tamanho da amostra (N)</b>	<b>0,20</b>	<b>0,15</b>	<b>0,10</b>	<b>0,05</b>	<b>0,01</b>
<b>1</b>	0,900	0,925	0,950	0,975	0,995
<b>2</b>	0,684	0,726	0,776	0,842	0,929
<b>3</b>	0,565	0,597	0,642	0,708	0,828
<b>4</b>	0,494	0,525	0,564	0,624	0,733
<b>5</b>	0,446	0,474	0,510	0,565	0,669
<b>6</b>	0,410	0,463	0,470	0,521	0,618
<b>7</b>	0,381	0,405	0,438	0,486	0,577
<b>8</b>	0,358	0,381	0,411	0,457	0,543
<b>9</b>	0,339	0,360	0,388	0,432	0,514
<b>10</b>	0,322	0,342	0,368	0,410	0,490
<b>11</b>	0,307	0,326	0,352	0,391	0,468
<b>12</b>	0,295	0,313	0,338	0,375	0,450
<b>13</b>	0,284	0,302	0,325	0,361	0,433
<b>14</b>	0,274	0,292	0,314	0,349	0,418
<b>15</b>	0,266	0,283	0,304	0,338	0,404
<b>16</b>	0,258	0,274	0,295	0,328	0,392
<b>17</b>	0,250	0,266	0,286	0,318	0,381
<b>18</b>	0,244	0,259	0,278	0,309	0,371
<b>19</b>	0,237	0,252	0,272	0,301	0,363
<b>20</b>	0,231	0,246	0,246	0,294	0,356
<b>25</b>	0,21	0,22	0,24	0,27	0,32
<b>30</b>	0,19	0,20	0,22	0,24	0,29
<b>35</b>	0,18	0,19	0,21	0,23	0,27
<b>Acima de 50</b>	$1,07/\sqrt{N}$	$1,14/\sqrt{N}$	$1,22/\sqrt{N}$	$1,36/\sqrt{N}$	$1,63/\sqrt{N}$

Fonte: Siegel e Castellan Jr. (2006)