



UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Rafaella Santos Vitorino

Modelo Dinâmico de Poisson Aplicado a Dados de Acidentes Automobilísticos nas Rodovias Federais do Estado da Paraíba

Campina Grande - PB

Agosto de 2019

Rafaella Santos Vitorino

**Modelo Dinâmico de Poisson Aplicado a Dados de
Acidentes Automobilísticos nas Rodovias Federais do
Estado da Paraíba**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Kleber Napoleão Nunes de Oliveira Barros

Campina Grande - PB

Agosto de 2019

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

V845m Vitorino, Rafaella Santos.
Modelo dinâmico de Poisson aplicado a dados de acidentes automobilísticos nas Rodovias Federais do Estado da Paraíba [manuscrito] / Rafaella Santos Vitorino. - 2019.
37 p. : il. colorido.
Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2019.
"Orientação : Prof. Dr. Kleber Napoleão Nunes de Oliveira Barros, Coordenação do Curso de Estatística - CCT."
1. Inferência Bayesiana. 2. Séries temporais. 3. Acidentes de trânsito. 4. Modelo de Poisson Dinâmico. I. Título
21. ed. CDD 519.5

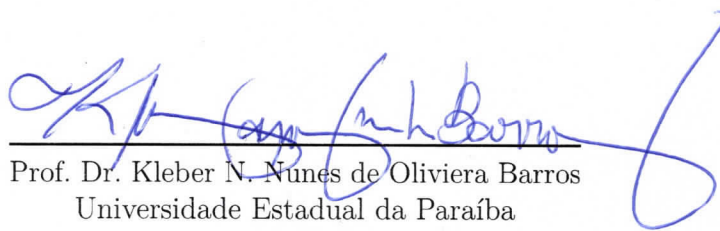
Rafaella Santos Vitorino

Modelo Dinâmico de Poisson Aplicado a Dados de Acidentes Automobilísticos nas Rodovias Federais do Estado da Paraíba

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 22 de Agosto de 2019.

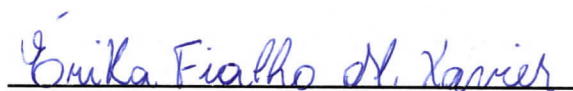
BANCA EXAMINADORA



Prof. Dr. Kleber N. Nunes de Oliveira Barros
Universidade Estadual da Paraíba
(Orientador)



Prof.ª Dra. Ana Patrícia Bastos Peixoto
Universidade Estadual da Paraíba



Prof.ª Dra. Érika Fialho Morais Xavier
Universidade Estadual da Paraíba

Agradecimentos

A Deus por ser amor, ser único, misericordioso e compassivo, o meu muito obrigado Meu Senhor e Meu Deus, obrigado apenas por ser A Tua vontade em minha vida e não a minha, e só Ti peço, seja tudo no Seu tempo; E que eu possa ser sinal de paz e amor ao próximo. Agradeço também a Virgem Maria por ser Tua filha, por me amar, me cuidar, me levar para uma vida de oração, pois é A Ti que peço verdadeiramente tudo e de coração; Obrigada também por me acolher, me abraçar, me embalar e proteger, obrigada Mãe por tudo, que eu possa dar sempre o meu sim a Deus como tu deste, pois és exemplo de esperança.

Aos meus pais pela enorme paciência, pelo amor, por sempre querer o meu melhor, por me fazer uma pessoa livre pela confiança, por mesmo no pouco eles nunca deixaram faltar nada, assim como nunca deixarão. Agradeço a eles por eles ser do jeitinho que são; Agradeço por Deus ter me dado os melhores, pois minha família é meu tudo e a gente se completa do jeito que somos; Agradeço também a minha tia Giselle por tudo, e até hoje sempre me cuidar, e a minha avó Denise pelo amor e ser única em nossas vidas, mesmo não estando mais aqui junto a nós.

Agradeço também a amiga que o curso me deu, Adriana por até hoje mesmo distante ela me escutar, ser uma pessoa a qual confio, por me aconselhar, por ter aguentado e aguenta todas as minhas loucuras, tristezas, felicidades e ansiedades, sou eternamente grata. E também a Andreza, Renata e Thuenne por serem minhas amigas, por me acolher em meio a alegrias e tristezas, por eu poder compartilhar tudo e toda a minha vida em todas as áreas, entre muitos outros momentos na qual sou grata pois elas mesmo longe, eu sei a importância delas não só na minha vida acadêmica e sim, principalmente na pessoal, pois se o nosso interior não está bem, nada funciona.

Grata também a todos que fizeram a minha vida acadêmica ter cor, sabor e vida, aos meus amigos e colegas que alcancei assim que entrei no curso e na UEPB, e a todos que fui conhecendo ao longo dele. A sr Antônio por ser um pai e amigo, grato a ele por todo carinho recebido, acolhimento e verdades. Agradeço também a todos os amigos e colegas que o espetinho me deu, sei a importância de cada um, e agradeço.

Agradeço também ao meu orientador Kleber Barros por sua amizade, paciência e conselhos. Assim, a todos os professores do departamento, a qual sou eternamente grata por além de serem professores, serem amigos, e me ajudar, seja na vida acadêmica, como pessoal, com verdades e conselhos, independente do curso; Agradeço a Deus por ter me apresentado a todos e que Deus só realize maravilhas a vocês e obrigado por todo ensinamento.

*"Não temerei se acaso
Tua vontade Senhor for maior que eu."*

Resumo

Os acidentes de trânsito são um grande problema de saúde pública no Brasil, partindo dessa ideia esse trabalho teve como objetivo utilizar o modelo dinâmico de Poisson para ajustar dados de acidentes automobilísticos nas Rodovias Federais (BRs) no Estado da Paraíba entre os anos de 2013 a 2018. Na metodologia a tabulação e sumarização dos dados foram realizadas com o auxílio do software Excel, as Análises de Séries Temporais obtidas com o software R empregando as simulações de Monte Carlo Via Cadeias de Markov (MCMC) para estimação de parâmetros e a Análise Bayesiana realizada pelo software OpenBUGS, usando a estatística de *Deviance Information Criterion* (DIC) para seleção de modelos. No final obteve-se uma modelagem útil, com um bom ajuste aos dados notificados que pode ser empregada pelo poder público para diminuir e prever o número de acidentes de trânsito no estado.

Palavras-chaves: Inferência Bayesiana; Séries Temporais; Rodovias Federais.

Abstract

Traffic accidents are a major public health problem in Brazil. Based on this idea, this study aimed to use the Poisson dynamic model to adjust data on automobile accidents on Federal Highways (BRs) in the state of Paraíba from 2013 to 2006. 2018. In the methodology the tabulation and summarization of the data were performed with the aid of Excel software, the Time Series Analyzes obtained with the R software employing Markov Chain Monte Carlo (MCMC) simulations for parameter estimation and the Bayesian Analysis. OpenBUGS software using *Deviance Information Criterion* (DIC) statistics for model selection. In the end a useful modeling was obtained, with a good adjustment to the notified data that can be used by the government to reduce and predict the number of traffic accidents in the state.

Key-words: Bayesian Inference; Time series; Federal highways.

Lista de ilustrações

Figura 1 – Mapa das Rodovias Federais da Paraíba	12
Figura 2 – Boxplot das rodovias com os maiores números de acidentes.	23
Figura 3 – Matriz de dispersão com correlações e histogramas.	24
Figura 4 – Gráfico de série nas rodovias com maiores números de acidentes.	25
Figura 5 – Gráfico <i>Monthplot</i> do número de acidentes para as rodovias BR-101, BR-104 e BR-230.	26
Figura 6 – Gráfico decompose para os números de acidentes na BR-101.	27
Figura 7 – Gráfico decompose para os números de acidentes na BR-104.	27
Figura 8 – Gráfico decompose para os números de acidentes na BR-230.	27
Figura 9 – Gráfico de autocorrelação das BRs.	28
Figura 10 – Boxplots para os números de acidentes previstos para a BR-101.	31
Figura 11 – Boxplots para os números de acidentes previstos para a BR-104.	31
Figura 12 – Boxplots para os números de acidentes previstos para a BR-230.	31
Figura 13 – Acidentes observados versus preditos para a BR-101.	32
Figura 14 – Acidentes observados versus preditos para a BR-104.	32
Figura 15 – Acidentes observados versus preditos para a BR-230.	32
Figura 16 – Acidentes observados versus preditos para a BR-110.	32

Lista de tabelas

Tabela 1 – Sumário das Estatísticas para as taxas de acidentes de trânsito nas rodovias federais da Paraíba.	22
Tabela 2 – Teste de Estacionariedade Phillips-Perron.	29
Tabela 3 – Critérios de comparação de modelos.	29
Tabela 4 – Estatísticas a posteriori para o modelo M2.	30

Sumário

1	INTRODUÇÃO	10
2	FUNDAMENTAÇÃO TEÓRICA	11
2.1	Acidentes de trânsito	11
2.2	Séries Temporais	12
2.2.1	Teste de Estacionaridade: Phillips-Perron	13
2.3	Inferência Bayesiana	14
2.3.1	Teorema de Bayes	14
2.3.2	Abordagem Bayesiana para previsão	15
2.3.3	Monte Carlo via Cadeias de Markov (MCMC)	15
2.3.3.1	Cadeias de Markov	15
2.3.3.2	Algoritmo de Metropolis-Hastings	16
2.4	Crítérios de Seleção de Modelos	16
2.4.1	Deviance Criterion Information	16
2.4.2	Raiz Quadrada do Erro Quadrático Médio	17
2.5	Modelos Lineares Dinâmicos (MLD)	17
2.6	Modelos Lineares Dinâmicos Generalizados (MLDG)	18
2.7	Modelo de Poisson Dinâmico	19
3	METODOLOGIA	21
4	APLICAÇÃO	22
4.1	Análise Exploratória	22
4.2	Análise Bayesiana	29
5	CONCLUSÃO	33
	REFERÊNCIAS	34
	APÊNDICES	36
	APÊNDICE A – CÓDIGO UTILIZADO	37

1 Introdução

Os dados do Instituto de Pesquisa Econômica Aplicada (IPEA)¹ juntamente com a Polícia Rodoviária Federal (PRF) apresentam os acidentes de trânsito como um grande problema de saúde pública no Brasil. Nos dados de pesquisa do instituto os acidentes de transportes terrestres no Brasil representam uma das principais causas de morte no país, matando cerca de 43 mil pessoas por ano e os acidentes nas rodovias correspondem 20% dessas mortes. Nos últimos dez anos no Brasil, segundo a Empresa Brasil de Comunicação (EBC)², entre 2009 e 2018 os acidentes de trânsito deixaram mais de 1,6 milhões de feridos com um custo de R\$ 2,9 bilhões para o Sistema Único de Saúde (SUS), trazendo um prejuízo enorme para saúde pública.

Segundo a Organização Mundial da Saúde (OMS), em 2015 cerca de 1 milhão de pessoas no mundo foram mortas por acidentes de trânsito, assim representaram a nona maior causa de fatalidade no mundo, esses dados são preocupantes e apontam que entre 15 a 29 anos de idade são os primeiros responsáveis por tais mortes, em segundo, na faixa etária de 5 a 14 anos e em terceiro de 30 a 44 anos, o que representou um custo no orçamento público, de aproximadamente US\$ 518 bilhões por ano, ou um percentual entre 1% e 3% do produto interno bruto de cada país (OMS, 2016). Nos dados da EBC², no período de 2009 a 2018 informa que vítimas com idade entre 15 e 39 anos apontam 60%, nas faixas etárias que vão de zero a 14 anos (8,2%) e em maiores de 60 anos (8,4%). Do total o sexo masculino é quase 80% das vítimas.

Analisando o aumento das estatísticas dos acidentes de trânsito nas rodovias brasileiras este presente trabalho, realizou-se uma modelagem temporal nos dados observados ao longo do tempo nas Rodovias do Estado da Paraíba com o intuito de avaliar as características de tendência, sazonalidade e aleatoriedade mensalmente, juntamente com os Modelos Lineares Dinâmicos (MLD) e Modelos Lineares Dinâmicos Generalizados (MLGD), e assim fazendo a seleção do melhor modelo, e aplicando a previsão Bayesiana com a técnica de modelo dinâmico de Poisson para modelar, prevê e diminuir as observações dos números de acidentes. Tendo por objetivo geral utilizar o modelo dinâmico de Poisson para ajustar dados de acidentes automobilísticos nas Rodovias Federais (BRs) da Paraíba entre os anos de 2013 a 2018.

¹ http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=26292

² <http://agenciabrasil.ebc.com.br/saude/noticia/2019-05/acidentes-no-transito-deixaram-mais-de-16-milhao-feridos-em-10-anos>

² <http://agenciabrasil.ebc.com.br/saude/noticia/2019-05/acidentes-no-transito-deixaram-mais-de-16-milhao-feridos-em-10-anos>

2 Fundamentação Teórica

2.1 Acidentes de trânsito

Conforme a OMS o número de acidentes de trânsito tem atingido altas taxas de mortalidade no mundo (WHO, 2013). Devido à expansão da indústria, conduzido pela revolução industrial no século XX e a consequente melhora da economia. Com isso, cada vez mais pessoas acabaram adquirindo veículos para sua locomoção. No entanto, é nítido que este aumento da frota não teve uma melhoria no sistema viário e no planejamento urbano o que acabou gerando vários problemas, inclusive de saúde pública (MAURO et al., 2001). O delineamento das políticas públicas urbanas e viárias do Brasil não acompanhou a expansão de frota mundial. Assim, várias pessoas tiveram problemas de mobilidade em várias localidades. Por outro lado, foi possível identificar que os países mais desenvolvidos relataram maior número de níveis de motorização e maior número de vítimas e a começar a realizar ações a fim de evitá-lo (VASCONCELOS, 1998).

A expansão da frota de veículos automotores tende a agravar a situação dos acidentes de trânsito. Destacam-se, nessa expansão da frota, as vendas de motocicletas, que, apresentam baixas condições de proteção aos usuários em caso de colisão e queda. Esse tipo de veículo proporciona alto grau de gravidade aos acidentes, acrescentando as estatísticas de mortes e feridos graves. Desde 2003, houve um aumento na frota nacional de 136,5%; a de automóveis, 102,6% e a de motocicletas, 269,8%, principalmente no Nordeste, 414%, aumentando os transtornos nas ruas e rodovias e assim, gerando a quantidade de vítimas de trânsito (FEDERAL-PRF et al., 2015).

Os acidentes de trânsito no Brasil têm sido visados pelas políticas públicas devido ao seu prejuízo para a sociedade e para o estado. Os dados epidemiológicos disponíveis são fundamentais para a identificação desses problemas. Grande parte dos casos de acidentes ocorridos é provocado pela falta de atenção, bem como a mistura de álcool e direção, manuseamento de celular, passagem no sinal vermelho, enfim, são situações corriqueiras que colocam todos em riscos. Na Figura 1 é possível observar através do mapa da Paraíba, cada uma das rodovias federais em estudo.

Segundo a PRF³ em 2014 houve uma queda nas taxas de acidentes e vítimas na Paraíba, mesmo com o aumento de 6,2% da frota nacional de veículos, fator que tem influência direta no risco de acidentes. Em cerca de 4 anos, pôde-se observar uma redução de acidentes de 30,7%. A taxa de acidentes por milhão de veículos caiu de 51,8% para 41,9 acidentes/milhão. O número de vítimas seguiu a mesma tendência registrando queda

³ <https://www.pr.gov.br/portal/estados/paraiba/pb>



Figura 1 – Mapa das Rodovias Federais da Paraíba

de 32,78 para 27,32 feridos/milhão e de 2,19 para 2,08 mortos/milhão. Essas variações nas taxas de acidentes, feridos e mortos correspondem a quedas de 19,2%, 16,65% e 5,36%, respectivamente. Em números absolutos, foram registrados 3.633 acidentes, 2.369 feridos e 180 mortos.

No que se refere aos dados da Secretaria de Estado da Segurança e Defesa Social (SEDS)⁴ em 2018, na Paraíba foram contabilizadas 753 vítimas de acidentes letais de trânsito (ALT), embora a diferença seja apenas de 230 vítimas entre mortos no trânsito e em casos de violência, o número de acidentes na Paraíba caiu de 8,5% em relação ao ano de 2017, após serem pagos pelo Seguro de Danos Pessoais Causados por Veículos Automotores de Vias Terrestres (DPvat) 1.017 indenizações por morte. A terceira maior redução no Nordeste atrás apenas para Bahia (-9,8%) e Pernambuco (-9,3%).

Em função das estatísticas apresentadas, os dados podem ser compreendidas e descrever o comportamento temporal mensal, utilizando o modelo dinâmico de Poisson, baseado-se nas metodologias de Modelos Lineares Dinâmicos (MLD) e Modelos Lineares Dinâmicos Generalizados (MLDG), permitindo-se ser estudadas as técnicas da Inferência Bayesiana.

2.2 Séries Temporais

Uma série temporal pode ser definida como uma sequência de observações realizadas ao longo do tempo. Tempo pode ser: espaço, períodos, dias, semanas, etc. As observações próximas/vizinhas são dependentes. Exemplos de séries temporais podem ser vistos nas

⁴ <https://g1.globo.com/pb/paraiba/noticia/2019/05/05/transito-matou-quase-tres-pessoas-por-dia-em-2018-na-pb-aponta-estatistica-do-dpvat.ghtml>

mais diversas áreas seja na economia, medicina, epidemiologia, engenharias, meteorologia e ciências sociais. A análise de séries temporais tem por finalidade compreender e prever o futuro do comportamento da série e a previsão à partir dos valores presentes e passados, com o objetivo de reduzir o máximo de erros de previsão (BOX; JENKINS; REINSEL, 2008).

A série temporal é um processo estocástico formado pelo conjunto de observações $Y_t, t \in T$, onde Y_t é variável de interesse observada no tempo t e T é o conjunto de índices de tempo, podendo ser classificado como discreto ($T = t_1, t_2, \dots, t_n$) ou contínuo ($T = t : t_1 < t < t_2$). Caso se trate de uma série temporal multivariada as observações são definidas como $Y_t, t \in T$, onde Y_t' são vetores de variáveis aleatórias.

Existem diversos modelos utilizados para a previsão de séries temporais, dentre os quais se destaca a modelagem Box-Jenkins. Essa metodologia envolve o modelo Autorregressivo Integrado de Médias Móveis (ARIMA). Os modelos da classe ARIMA são famosos por serem flexíveis, pois permitem ajustes e adaptações nos seus parâmetros, podendo se adaptar a qualquer tipo de séries temporais.

Segundo Morettin e Tolo (2006), para compreender o comportamento de uma série temporal é preciso conhecer seus componentes característicos:

- a) Estacionaridade: é o comportamento da série.
- b) Tendência: é o efeito na média ao longo do prazo.
- c) Sazonalidade: é os efeitos ligados nas variações periódicas (semanal, mensal, anual, etc).
- d) Ciclos: variáveis periódicas, não associadas a nenhuma medida temporal;
- e) Erro aleatório.

Mais detalhes em Box, Jenkins e Reinsel (2008) Morettin e Tolo (2006).

2.2.1 Teste de Estacionaridade: Phillips-Perron

O teste de Phillips e Perron (1988) é conhecido como teste PP na literatura que é o resultado do teste de raiz unitária de Dickey e Fuller (1979) para o conjunto dados em que os erros são correlacionados. As hipóteses do teste PP são

$$\begin{cases} H_0 : \text{A série possui raiz unitária;} \\ H_1 : \text{A série não possui raiz unitária.} \end{cases}$$

Caso o valor-p do teste seja inferior 0,05, rejeitamos H_0 com 5% de significância, ou seja, a série não possui raiz unitária, portanto a série é estacionária.

2.3 Inferência Bayesiana

A inferência Bayesiana é uma área na estatística onde a estimação dos parâmetros do modelo pode ser quantificada através de uma distribuição de probabilidade. Os trabalhos da teoria Bayesiana desenvolvidos por Thomas Bayes em 1761, só foram publicados dois anos após sua morte, em 1763. A análise Bayesiana se baseia no conhecimento da distribuição a posteriori dos parâmetros e possibilita a construção de intervalos de confiança (melhor definido como intervalo de probabilidade ou intervalo de credibilidade) exatos para as estimativas dos parâmetros (GIANOLA; FERNANDO, 1986).

O objetivo da estatística Bayesiana é estimar uma quantidade de interesse desconhecida, denominada parâmetro, denotada geralmente por θ . Esse processo de estimação pode ser de duas maneiras: estimação pontual e estimação intervalar. Para esses processos de estimação, segundo Robertson (1955) considera-se que a informação a priori seja descrita, por uma distribuição de probabilidade. De modo geral, a inferência Bayesiana se baseia na noção do grau de confiança em determinada hipótese sobre as quantidades de interesse não observáveis. Ou seja, é uma probabilidade subjetiva. Assim, esta probabilidade subjetiva pode variar entre indivíduos ou em função dos acontecimentos, estabelecendo uma probabilidade a priori sobre a hipótese. A atualização das informações com dados e evidências sobre a hipótese é realizada via teorema de Bayes, modificando e refinando a probabilidade subjetiva continuamente.

A análise Bayesiana de dados é uma alternativa importante na estatística com relação aos procedimentos clássicos de estimação e testes de hipóteses que vêm tomando espaço crescente em aplicações nas mais diversas áreas. Segundo Paulino, Turkman e Murteira (2003), a inferência Bayesiana é o processo de encontrar um modelo de probabilidade para um conjunto de dados e resumir o resultado por uma distribuição de probabilidade sobre os parâmetros do modelo e sobre quantidades não observadas, tais como predição para novas observações. A metodologia Bayesiana consiste em agregar aos dados amostrais (função de verossimilhança) e do conhecimento prévio a respeito dos parâmetros (distribuição a priori) para obter a densidade a posteriori dos parâmetros por meio do teorema de Bayes. A informação a priori é incorporada através da especificação de uma distribuição para o parâmetro, devendo esta representar probabilisticamente o conhecimento que se tem sobre ele antes da realização do experimento.

2.3.1 Teorema de Bayes

Segundo Ntzoufras (2009), o vetor de parâmetros θ a ser estimado, sendo geralmente θ o espaço paramétrico e Y uma variável aleatória. A distribuição à posteriori para θ pelo Teorema de Bayes, é dada por

$$p(\theta | y) = \frac{p(\theta) p(y | \theta)}{\int_{\theta} p(\theta) p(y | \theta) d\theta}, \quad (2.1)$$

sendo θ contínuo, $p(\theta)$ a distribuição a priori e função de verossimilhança de θ é $p(y|\theta) = \prod_{i=1}^n p(y_i|\theta)$, onde se permite estimar o valor dos parâmetros a partir de um conjunto de dados. A equação (2.1) pode ser descrita da seguinte forma:

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (2.2)$$

Isto é, sendo $p(y)$ no denominador, a igualdade é substituída pela proporcionalidade, simplificando o teorema de Bayes. A partir da equação (2.2) a constante normalizadora $p(y)$ da posteriori em (2.1) é alcançada com

$$p(y) = \int p(y, \theta) d\theta = \int p(y|\theta)p(\theta) d\theta, \quad (2.3)$$

chamada de distribuição preditiva a priori. Essa é a distribuição esperada de y dado θ , antes da realização do experimento ou observação dos dados. A distribuição preditiva \tilde{Y} dado y é alcançada pela integração

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta, y) d\theta = \int p(\tilde{y}|\theta, y)p(\theta|y) d\theta. \quad (2.4)$$

2.3.2 Abordagem Bayesiana para previsão

Segundo Pole, West e Harrison (1994), a abordagem Bayesiana para previsão é sempre probabilística. A previsão é baseada em uma fonte de conhecimento que evoluiu para prática, portanto, as previsões são processos de probabilidade condicional, sendo o condicionamento o estado de conhecimento existente. O modelo Bayesiano fornece uma estrutura coerente e formal para combinar informações. As previsões Bayesianas são probabilísticas e são dadas através da eq. (2.4).

Em geral, a abordagem Bayesiana para previsão permite que informações não contidas nos dados sejam incorporadas ao modelo. A possibilidade de se obter informações externas é um ponto positivo para a abordagem Bayesiana.

2.3.3 Monte Carlo via Cadeias de Markov (MCMC)

O método de simulação de Monte Carlo via Cadeia de Markov (MCMC), pode ser utilizado para encontrar distribuições a posteriori intratáveis analiticamente. Assim, busca-se obter nessa metodologia uma amostra a posteriori e o cálculo das estimativas amostrais, baseado na utilização de técnicas iterativas nas cadeias de Markov. De acordo com Ntzoufras (2009), os métodos MCMC simulam um passeio aleatório no espaço θ , tendendo para uma distribuição estacionária, que é a distribuição de interesse do problema.

2.3.3.1 Cadeias de Markov

A cadeia de Markov é um caso particular de processo estocástico X_0, X_1, \dots , onde $X_i, i = 1, \dots, t$ com possíveis valores de probabilidade em um conjunto T . X_t é uma cadeia

de Markov se

$$P(X_t \in T | X_0, \dots, X_{t-1}) = P(X_t \in T | X_{t-1}),$$

ou seja, a distribuição X_t depende exclusivamente de X_{t-1} .

Segundo Ntzoufras (2009), o método MCMC requer que a cadeia de Markov seja homogênea (ou seja, é invariável as probabilidades na transição de um estado para o outro), irreduzível (é um número finito de interações entre a transição dos estados que é atingido) e aperiódica (onde não existem estados absorventes).

Supondo uma distribuição $\pi(x)$, $x \in \mathbb{R}^d$, com pelo menos uma constante multiplicativa conhecida, não possível obter uma amostra de forma direta. Dadas as realizações $X^{(t)}$, $t = 0, 1, \dots$ de uma cadeia de Markov onde π é a distribuição de equilíbrio, então,

$$X^{(t)} \xrightarrow{t \rightarrow \infty} \pi(x) \text{ e } \frac{1}{n} \sum_{t=1}^n g(X_i^{(t)}) \xrightarrow{n \rightarrow \infty} E_x[g(X_i)].$$

A média teórica é estimada pela média dos valores da cadeia. Assim, a cadeia excede a ter menos influência dos valores iniciais, com o aumento de número de interações e convergindo equilíbrio para distribuição, fazendo com que, em aplicações, as interações iniciais possam ser descartadas.

2.3.3.2 Algoritmo de Metropolis-Hastings

O algoritmo de Metropoli Hastings é caracterizado por um processo de reamostragem em dados não contínuos, que garante a convergência de cadeia para distribuição a posteriori, o algoritmo é um tipo de cadeia de Markov que gera valores de uma distribuição auxiliar. Mais detalhes em (HOFF, 2009).

2.4 Critérios de Seleção de Modelos

Quando existem diversos modelos competindo para o ajuste dos dados, faz-se necessária a escolha do melhor modelo. Diferentes critérios podem ser empregados. A seguir se discute alguns dos critérios utilizados na literatura Bayesiana.

2.4.1 Deviance Criterion Information

Na inferência Bayesiana o método mais utilizado para a escolha de modelos é o critério *Deviance Information Criterion* (DIC), em que sendo este uma generalização dos critérios *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC). De acordo com Moulder e Algina (2002), y é um conjunto de dados e θ é o parâmetro do modelo, a estatística Deviance é definida como

$$D(y, \theta) = -2 \ln(p(y | \theta)) \quad (2.5)$$

sendo também,

$$\begin{aligned} D_{\bar{\theta}}(y) &= D(y, \bar{\theta}) \\ \bar{D}(y, \theta) &= E(D(y, \theta) | y) \\ pD &= \bar{D}(y, \theta) - D_{\bar{\theta}}(y), \end{aligned}$$

em que $D_{\bar{\theta}}(y)$ é a deviance calculada na média a posteriori ($\bar{\theta}$), $\bar{D}(y, \theta)$ é a deviance média a posteriori e pD o número efetivo de parâmetros. Calculemos o DIC como

$$DIC = 2\bar{D}(y, \theta) - D_{\bar{\theta}}(y). \quad (2.6)$$

Assim como outros critérios, o melhor modelo, entre os modelos em estudo, é aquele que possuir menor valor de DIC.

2.4.2 Raiz Quadrada do Erro Quadrático Médio

Uma medida largamente utilizada para avaliar a avaliação global do modelo é a raiz do erro quadrático médio (REQM) dado por

$$REQM = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2}, \quad (2.7)$$

sendo y_i^* a mediana da densidade preditiva a posteriori $p(\tilde{y}_i | \mathbf{y})$. A REQM deve ser o menor possível para o melhor modelo.

2.5 Modelos Lineares Dinâmicos (MLD)

De acordo com West e Harrison (1997), o modelo linear dinâmico normal pode ser escrito da seguinte maneira por um conjunto de quádruplos

$$\{F, G, V, W\}_t = \{F_t, G_t, V_t\}, W_t, \quad (2.8)$$

onde, para cada t ,

- F_t é uma matriz de constantes conhecidas ($n \times r$), podendo ser um grupo de variáveis explicativas;
- G_t é uma matriz conhecida ($n \times n$), dos parâmetros de estado que mostram a evolução temporal, conhecida como matriz de evolução;
- V_t é uma matriz de variação conhecida ($r \times r$), que tem associação ao erro observacional;
- W_t é uma matriz de variação conhecida ($n \times n$), que tem associação ao erro de evolução de estado.

Em que r é o número de observações do vetor de constantes conhecidas Y_t ($r \times 1$) e n o número de parâmetros. Este modelo que relaciona Y_t ao vetor de parâmetro θ_t ($n \times 1$) no tempo t , e a sequência θ_t através do tempo, definem o quádruplo através das distribuições sequencialmente especificadas

$$(Y_t | \theta_t) \sim N [F_t' \theta_t, V_t], \quad (2.9)$$

e

$$(\theta_t | \theta_{t-1}) \sim N [G_t \theta_{t-1}, W_t]. \quad (2.10)$$

As equações (2.8) e (2.9) também são ocultamente condicionais em D_{t-1} , o conjunto de informações disponível antes do tempo t . Em particular, os valores das variâncias definidoras V_t e W_t são incluídos e as observações passadas Y_{t-1}, Y_{t-2}, \dots , tanto como o conjunto inicial de informações D_0 . Uma alternativa de representação dessas equações definidoras é

$$Y_t = F_t \theta_t + v_t, v_t \sim N [0, V_t], \quad (2.11)$$

e

$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N [0, W_t]. \quad (2.12)$$

- F_t é a matriz do design de valores conhecidos de variáveis independentes;
- θ_t é o estado, ou sistema, vetor;
- $\mu_t = F_t \theta_t$ é a resposta ao nível médio;
- v_t é o erro observacional,

são internamente e mutuamente independentes as sequências de erro v_t e w_t . A equação (2.10) é a equação de observação para o modelo, definindo a distribuição amostral para Y_t condicional à quantidade θ_t . Dado θ_t , é independente de todas as outras observações Y_t e valores de parâmetros; e, de modo geral, é independente do passado, o presente e o futuro. Através de uma regressão linear dinâmica esta equação relaciona o Y_t para θ_t com uma estrutura de erro normal multivariada tendo matriz de variância observacional conhecida.

2.6 Modelos Lineares Dinâmicos Generalizados (MLDG)

Pode ser definido um modelo linear dinâmico generalizado como uma extensão dos modelos lineares dinâmicos quando a variável resposta pertence a distribuições pertencentes à família exponencial, de acordo com West, Harrison e Migon (1985). Além disso, é

definido como uma extensão do Modelo linear Generalizado (MLG) o MLDG, proposto por (NELDER; WEDDERBURN, 1972), quando os parâmetros dependem do tempo.

$$f(y_t | \theta_t, \phi_t) = \exp \left\{ \frac{[y_t \theta_t - b(\theta_t)]}{a(\phi_t)} c(y_t, \phi_t) \right\}. \quad (2.13)$$

Com esperança $\mu_t = E[y_t | \theta_t, \phi_t]$ e com uma função de ligação monótona e diferenciável g e um vetor preditivo p -dimensional X_t incluindo um intercepto tal que

$$g(\mu_t) = X_t' \beta_t,$$

especificando um esquema de atualização linear de primeira ordem.

$$\beta_t = G_t \beta_{t-1} + w_t, t = 2, \dots, T,$$

onde w_t sua média 0 de covariância W , com sua condicional inicial w_1 . Em modelos de regressão no uso em séries temporais, suas variáveis no tempo podem ser apropriadas para aplicar em modelos lineares generalizados dinâmicos.

A abordagem Bayesiana linear é um dos métodos de estimação do MLDG, segundo West e Harrison (1997). O método é considerado como uma aproximação para estimar o primeiro e o segundo momento da distribuição a posteriori de maneira sequencial. Diversas são as aplicações com o MLDG via inferência Bayesiana, ver em (TRIANAFYLLOPOULOS, 2009);(CONGDON, 2005).

2.7 Modelo de Poisson Dinâmico

É definido o modelo dinâmico de Poisson como uma extensão de modelo de regressão de Poisson quando o seu parâmetro é variável no tempo. Assim, a série temporal para dados de contagem que seguem uma distribuição de Poisson assume $\{Y_t, t = 1, \dots, T\}$. Então os MLDGs, com um processo de observação e um processo de transição, podem ser expressos da seguinte maneira:

- Equações das observações do modelo:

$$y_t | \lambda_t \sim Pois(\lambda_t) = Pois(\exp(x_t)).$$

- A função de ligação:

$$\log(\lambda_t) = n_t = g(\lambda_t) = X_t' \beta_t.$$

- Equação para mudança de estados:

$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N[0, W],$$

onde w_t é um erro de evolução que segue a distribuição Normal com média zero e variância W . Existem extensões para o modelo dinâmico de Poisson que podem envolver efeitos de tendência, sazonalidade e superdispersão, como no modelo Ferreira e Gamerman (2000) para casos de meningite mensal no Rio de Janeiro.

$$y_t \sim \text{Pois}(\lambda_t).$$

$$\log(\lambda_t) = \alpha_t + s_t + n_t + \beta_t X_t,$$

com os efeitos mensais de tendência (α_t) e sazonalidade (s_t)

$$\alpha_t = \alpha_{t-1} + w_{1t}.$$

$$s_t = -(s_{t-1} + s_{t-2} + \dots + s_{t-11}) + w_{2t},$$

onde os erros

$$w_{1t} \sim N(0, W_1) w_{2t} \sim N(0, W_2),$$

com o controle de efeito de superdispersão definida como

$$n_t \sim N(0, W_3).$$

3 Metodologia

Neste trabalho, o intuito é avaliar casos de acidentes automobilísticos. Foram analisados números de acidentes nas principais rodovias do estado da Paraíba, entre os anos de 2013 a 2018. Os dados foram obtidos no site da PRF⁵. Foram utilizadas as 9 vias, sendo elas a BR-101, BR-104, BR-110, BR-116, BR-230, BR-361, BR-405, BR-412 e BR-427 do referido site. A tabulação e sumarização dos dados foram realizadas com o auxílio do software Excel. As análises de séries temporais foram obtidas com o software R. A análise Bayesiana foi realizada pelo software **OpenBUGS**. Utilizando a técnica do Modelo Dinâmico de Poisson para dados de contagem, foi considerado Y uma variável aleatória condicionada a um parâmetro λ , que é a média da distribuição. A verossimilhança do modelo é dada por

$$p(y | \lambda) = \prod_{t=1}^{72} \prod_{j=1}^9 \frac{e^{-\lambda_{tj}} \lambda_{tj}^{y_{tj}}}{y_{tj}!},$$

em que y a quantidade de acidentes, $t = 1, 2, \dots, 72$ é o índice referente ao mês, variando de janeiro de 2013 até dezembro de 2018 e $j = 1, 2, \dots, 9$ o número da rodovia. A média de acidentes λ_{tj} no tempo t e na BR j é dado por

$$\lambda_{tj} = \theta_{tj} \times \text{tam}_j \times e^{s_t},$$

em que θ_{tj} é a taxa de acidentes nas rodovias em meses, tam_j , é o tamanho das rodovias e e^{s_t} é um efeito que considera a sazonalidade de um período de 12 meses. Aplicando-se a escala logarítmica (Log), que fica,

$$\log(\lambda_{tj}) = \log(\theta_{tj}) + \log(\text{tam}_j) + s_t.$$

As distribuições a priori são definidas por:

$$\theta_{tj} \sim G(a, b)$$

$$a \sim G(0, 001; 0, 001)$$

$$b \sim G(0, 001; 0, 001)$$

$$w_t \sim U(-1, 1)$$

sendo G , a distribuição Gamma e U a distribuição Uniforme. O efeito sazonal foi definido por

$$s_t = \begin{cases} 0, & t = 1, 2, \dots, 11 \\ - \sum_{i=t-11}^{t-1} s_i + w_t, & t = 12, 13, \dots \end{cases}$$

As modelagens alternativas, sem e com sazonalidade foram comparados via DIC para decidir o modelo final.

⁵ <<https://www.prf.gov.br/portal/estados/paraiba/pb>>(último acesso em: 04/06/2019).

4 Aplicação

Primeiramente foi realizado, uma análise exploratória dos dados, com casos notificados de acidentes de trânsito por mês, de janeiro de 2013 a dezembro de 2018. Em seguida, uma análise de séries temporais para todas as rodovias do estado paraibano. Por fim, realizar-se uma análise Bayesiana considerando o modelo dinâmico de Poisson descrito na metodologia.

4.1 Análise Exploratória

A Tabela 1 apresenta o sumário das estatísticas referentes aos casos de acidentes de trânsito por mês. Note que as rodovias com maiores médias e medianas são BR101, BR104 e BR230, tendo a BR230 destacadamente mais acidentes que todas as outras rodovias combinadas em média (115,9 acidentes por mês). As quantidades mínimas e máximas de acidentes também são maiores para as três rodovias. Particularmente, as demais estradas registraram mínimo mensal de zero acidentes. Verificando o intervalo interquartilico ($\Delta Q = Q1 - Q3$), observa-se que as rodovias BR104 ($\Delta Q = 12$) e BR101 ($\Delta Q = 11$) têm menor dispersão comparadas com a BR230 ($\Delta Q = 38,5$). Porém, ao se examinar a variação levando em conta a magnitude as médias, pelo coeficiente de variação, chega-se a conclusão que as médias de acidentes das rodovias BR101 e BR230 são as mais representativas, com coeficientes 0,28 e 0,31, respectivamente. As autovias com menores médias de acidentes são também aquelas com maiores coeficientes de variação.

Tabela 1 – Sumário das Estatísticas para as taxas de acidentes de trânsito nas rodovias federais da Paraíba.

Rodovia	Média	Desvio Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
BR-101	32,35	9,21	15	26,0	32,0	37	62
BR-104	14,19	7,70	3	8,0	13,5	20	40
BR-110	0,67	1,27	0	0,0	0,0	1	6
BR-116	0,22	0,56	0	0,0	0,0	0	3
BR-230	115,93	35,96	67	91,5	111,5	130	257
BR-361	4,7	3,92	0	2,0	4,0	6	20
BR-405	1,60	1,87	0	0,0	1,0	3	8
BR-412	2,79	2,34	0	1,0	2,0	4	12
BR-427	1,68	1,85	0	0,0	1,0	2	9

Carmo et al. (2019) também afirma que as rodovias federais no estado da Paraíba com o maior número de vítimas de acidentes de transitos, são as vias BR-101, BR-104 e

BR-230, destacando-se a BR-230 com o maior número de vítimas, a rodovia concentrou-se na capital de João Pessoa dos quilômetros 10 a 40, e cidades próximas Cabedelo, Mamanguape, Bayeux e Santa Rita como os segmentos rodoviários críticos.

Na Figura 2, observa-se os boxplots para os acidentes ocorridos nas BR-101, BR-104 e BR-230. Nota-se que a mediana (linha central) dos acidentes na BR-230 é bem superior às demais medianas das outras vias. Observa-se ainda que as “caixas” são mais achatadas para as duas primeiras estradas, indicando menor variabilidade, tal como foi verificado pelos desvios-padrão da Tabela 1. Visualmente, as distribuições dos acidentes peracem simétricas (medianas aproximadamente nos centros das caixas) para as três autopistas. É possível ainda distinguir um mês com número de acidentes discrepante (ponto fora do boxplot) para BR-101 e para a BR-104. O BR-203 apresenta três meses com registros de acidentes muito superiores à mediana.

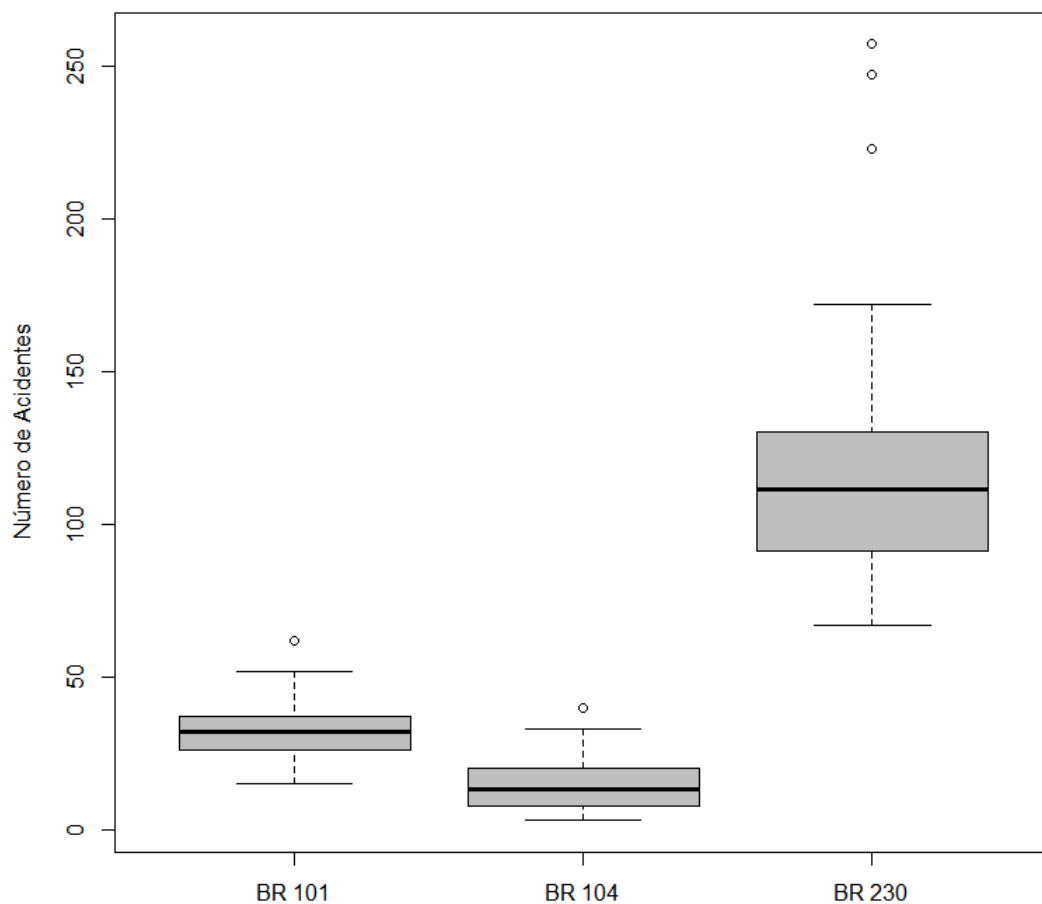


Figura 2 – Boxplot das rodovias com os maiores números de acidentes.

Na Figura 3, observa-se a matriz de dispersão (triângulo inferior), correlações

(triângulo superior) e histogramas (diagonal) para as nove rodovias federais estudadas na Paraíba. Boa parte dos gráficos de dispersão se encontram de forma aleatória, não aparentando haver tendências (linha central não-constante). As exceções são as dispersões entre BR-405 e BR-104; BR-405 e BR-116; BR-405 e BR-230; BR-104 e BR-230; BR-116 e BR-230; BR-427 e BR-230. Note ainda que as rodovias BR-110 e BR-116, por terem poucos acidentes, apresentam gráficos de dispersão com característica discreta. As demais rodovias apresentam maior espalhamento das coordenadas, exibindo as conhecidas nuvens de pontos. As correlações lineares significativas ocorreram entre as vias BR104 e BR-230; BR-116 e BR-230; BR-104 e BR-405; BR-230 e BR-405; BR-104 e BR-427; e BR-230 e BR-427, sendo esta a correlação maior (0,43). Note ainda que, embora pequenas, todas as correlações significativa são positivas, isto é, se aumenta o número de acidentes numa autoestrada a tendência é que ocorra o mesma na outra. Quanto aos histogramas, verifica-se que os mesmos são assimétricos para estradas com baixas notificações de acidentes e aparentemente simétricos para as rodovias com maiores registros. Os histogramas para as autopistas com baixas quantidades de acidentes também apresenta grande concentração de meses sem sinistros.

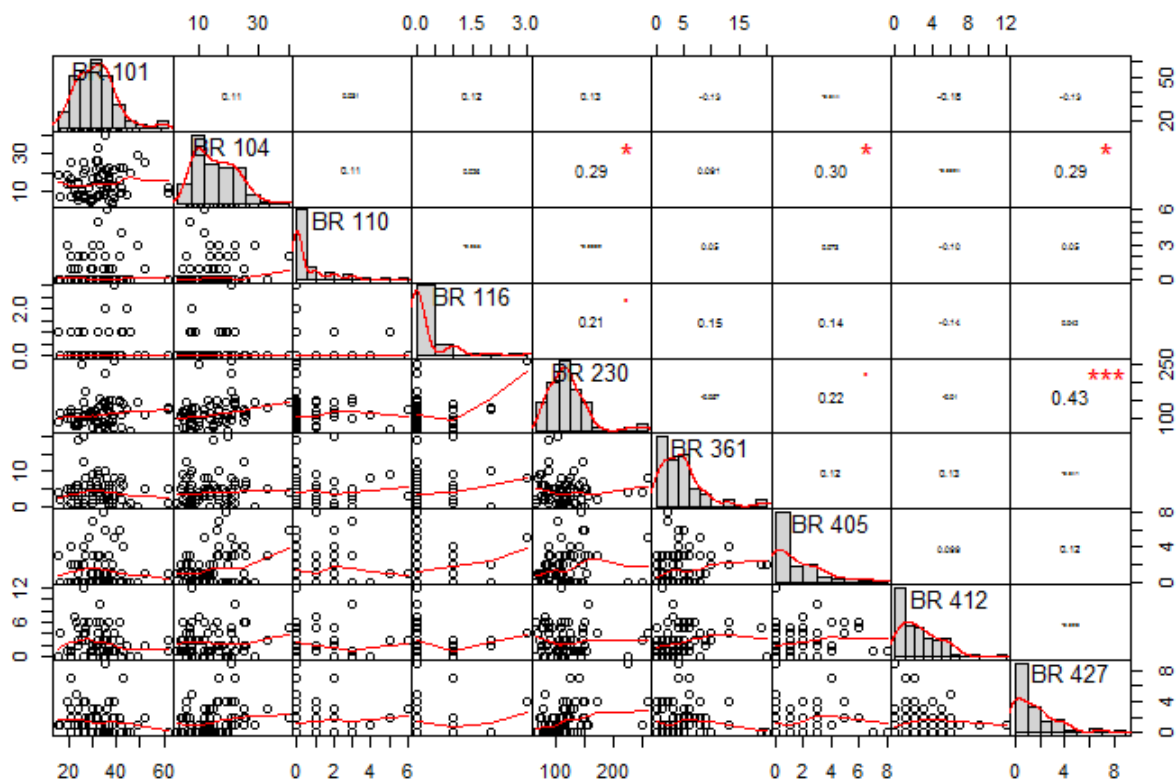


Figura 3 – Matriz de dispersão com correlações e histogramas.

Na Figura 4, pode-se observar os gráficos de séries temporais para as rodovias que apresentam as maiores quantidades de acidentes. Note que os dados da série da BR-230

estão sempre acima dos dados das outras duas estradas. Em geral, a série da BR-101 está acima da série da BR-104, com poucos cruzamentos, isto é, momentos em que o número de acidentes na BR-104 ultrapassam os mesmos números para a BR-101. Observe que apenas no fim do primeiro ano estudado (2013) há um aumento discrepante dos números de acidentes mensais, justamente os três pontos meses discrepantes (outubro, novembro e dezembro de 2013). Nota-se ainda uma tendência na diminuição de acidentes nesta rodovia. Nas demais rodovias não parece haver uma tendência de aumento ou diminuição nos casos de acidentes. Por este gráfico é possível comparar ainda a variabilidade da BR-230 com as demais rodovias, observando-se mais uma vez sua alta variabilidade em relação às demais rodovias.

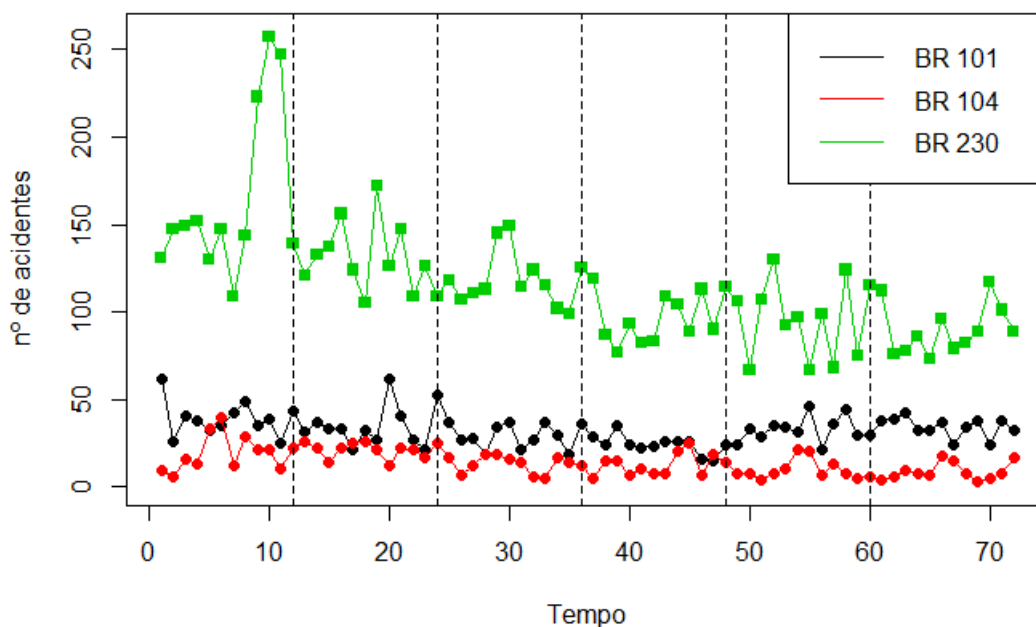


Figura 4 – Gráfico de série nas rodovias com maiores números de acidentes.

Lima (2019) relatou-se incidências e fatores associados aos acidentes de trânsito no Município de Mari/PB, uma distância aproximada de 64 km da capital de João Pessoa, e seu acesso principal se dá pela rodovia BR-230, mostrando ligação da via de entrada na cidade como influência, e afirmou-se que a partir dos anos de 2013 a 2014 ao longo dos anos começou-se a ter um decréscimo no número de óbitos e acidentes no trânsito.

Na Figura 5, é plotado o *Monthplot* que apresenta as médias mensais dos números de acidentes, que é indicada no gráfico por meio de uma reta horizontal, para a BR-101, BR-104 e BR-230, no período investigado. Para a BR-101 janeiro, agosto, setembro e dezembro são os meses com maiores números de acidentes. O mês com menores quantidades de acidentes, ao longo dos seis anos estudados, foi novembro. Em geral, os meses de janeiro,

agosto e dezembro iniciam com muitos acidentes nesta rodovia, enquanto que nos meses de julho e novembro ocorrem mais acidentes no fim do mês. Para os demais meses há uma oscilação com pequenos desvios em torno da média ao longo dos dias. Na BR-104 os meses de maio, setembro e dezembro apresentam as maiores ocorrências. Já os meses de Janeiro, fevereiro, março e novembro têm as menores médias. Os meses de janeiro, fevereiro, abril, maio, junho, agosto e dezembro têm, em geral, inícios mais violentos, enquanto os demais meses acontece uma oscilação em torno da média. Para a BR-230, os meses com maiores números de acidentes nos anos estudados foram abril, setembro, outubro e novembro, enquanto os meses de fevereiro, maio e julho registraram as menores médias. Note ainda que as concentrações de acidentes acontecem no início do mês, diminuindo drasticamente em torno do dia 10 até o fim de cada mês. Fenômeno este que é observado para quase todos os meses da série histórica.

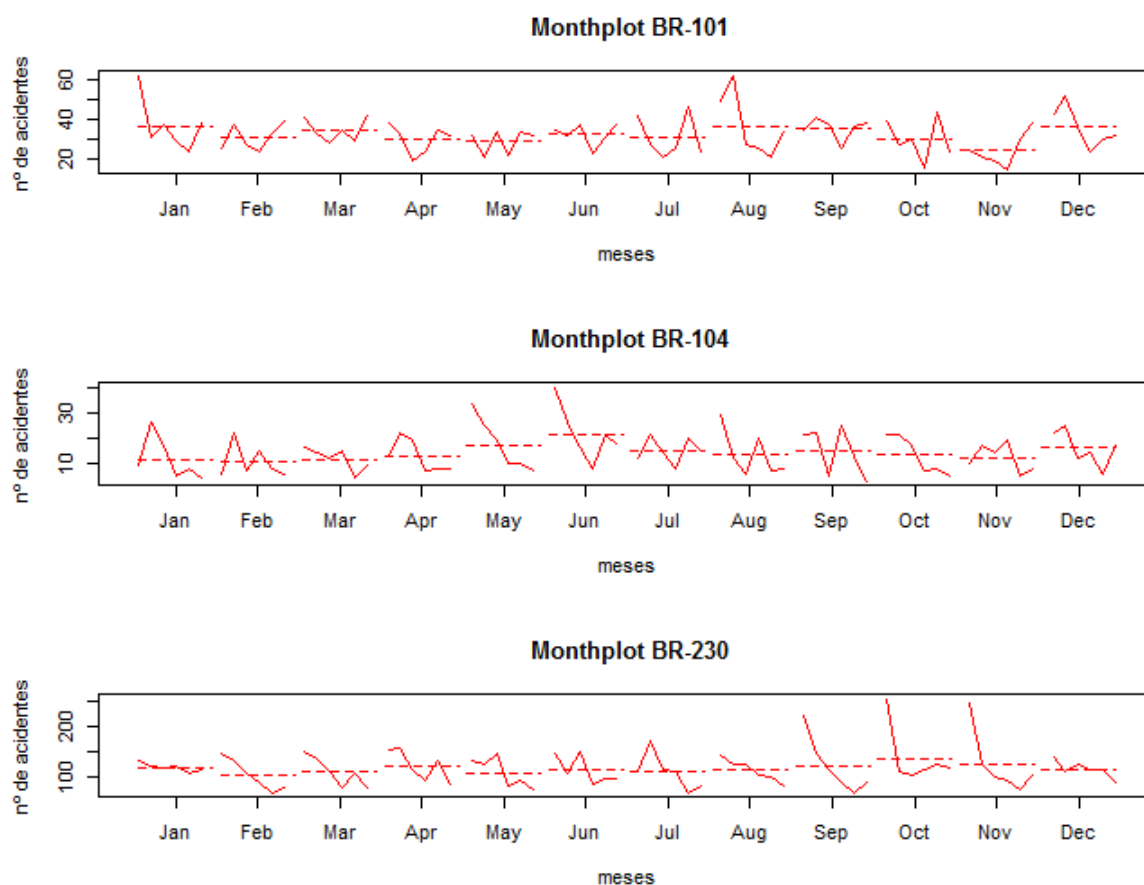


Figura 5 – Gráfico *Monthplot* do número de acidentes para as rodovias BR-101, BR-104 e BR-230.

Na Figura 6 é apresentada a decomposição da série da BR-101, podendo-se avaliar visualmente as características de tendência, sazonalidade e aleatoriedade dos dados. A componente de tendência apresenta um decréscimo, que é revertido a partir do final de 2017. A componente de sazonalidade apresenta uma periodicidade, indicando que realmente

deve existir uma certa sazonalidade na série. A componente aleatória, excetuando um breve período no fim de 2014 e entre 2016 e 2017, parece casual. Na Figura 7 é registrada a decomposição para os dados da BR-104. A componente de tendência indica que há um decréscimo na quantidade de acidentes para esta rodovia. A componente de sazonalidade indica que existe uma periodicidade na série. Na componente aleatória, observa-se não haver grandes discrepâncias em relação à aleatoriedade requerida. A decomposição dos dados para a BR-230 está plotada na Figura 8. Observando a componente tendência é possível perceber um decréscimo do número de acidentes na rodovia a partir do ano de 2014 ao longo do tempo. A componente sazonalidade indica que existe um movimento periódico persistente de decréscimo, oscilação e acréscimo no número de acidentes em cada um dos anos. Pela componente aleatória, nota-se não haver grandes disparidades em relação a um ruído branco, exceto para o final de 2013, coincidindo com os três meses discrepantes indicados pelo boxplot.

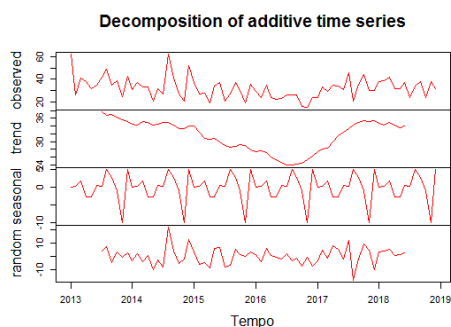


Figura 6 – Gráfico decompose para os números de acidentes na BR-101.

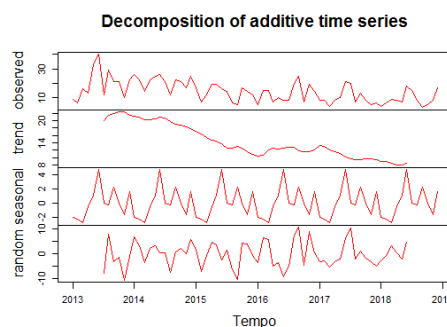


Figura 7 – Gráfico decompose para os números de acidentes na BR-104.

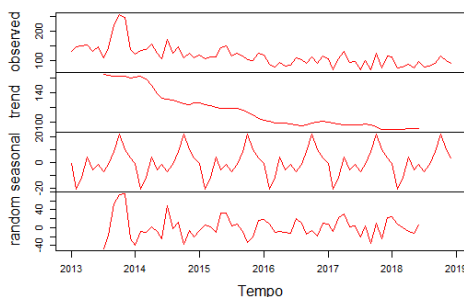


Figura 8 – Gráfico decompose para os números de acidentes na BR-230.

Afirma Santos et al. (2019), que destacam-se os anos de 2014 e 2017 com o maior número de registros ocorridos em acidentes de trânsito no estado de Alagoas, em que a BR-104, inicia-se no Rio Grande do Norte e vai até Alagoas, e passa pelos estados da Paraíba e Pernambuco, assim, pode-se afirmar uma grande influência nas demais vias BR-101 e BR-230 nos números de acidentes dentre estes anos.

Na Figura 9 são apresentados os gráficos da função de autocorrelação (ACF) para

as séries dos acidentes das rodovias BR-101, BR-104 e BR-230. As barras acima das linhas tracejadas em azul indicam defasagens (lag) significantes. Para a BR101 quase todas as barras estão dentro do limite da linha tracejada azul. Ou seja, as autocorrelações não diferem estatisticamente de zero, indicando que a série é aleatória. Na BR-104 os lags 1, 2, 6, 7, 8, 9, 10 e 11 são significativamente diferentes de zero, verificando que os acidentes do mês atual é correlacionado com os acidentes dos onze meses anteriores, indicando uma periodicidade na explicação da variável por ela mesma, isto é, apresentando possivelmente uma sazonalidade. Na BR-230 há uma correlação significativa nos nove primeiros lags. Note, entretanto que há um decaimento na função de autocorrelação seguida por um aumento, isto é, os primeiros meses tem maior influência na explicação dos acidentes do mês atual, seguidos por meses com menor influência na explicação, que por sua vez são seguidos por meses com maior poder de explicação, o que indica mais uma vez a presença de sazonalidade, corroborando a informação da Figura 8. A partir das informações da Figura 9, verifica-se que a modelagem para os dados dos acidentes das rodovias federais que atravessam o estado da Paraíba deve levar em conta um componente de sazonalidade.

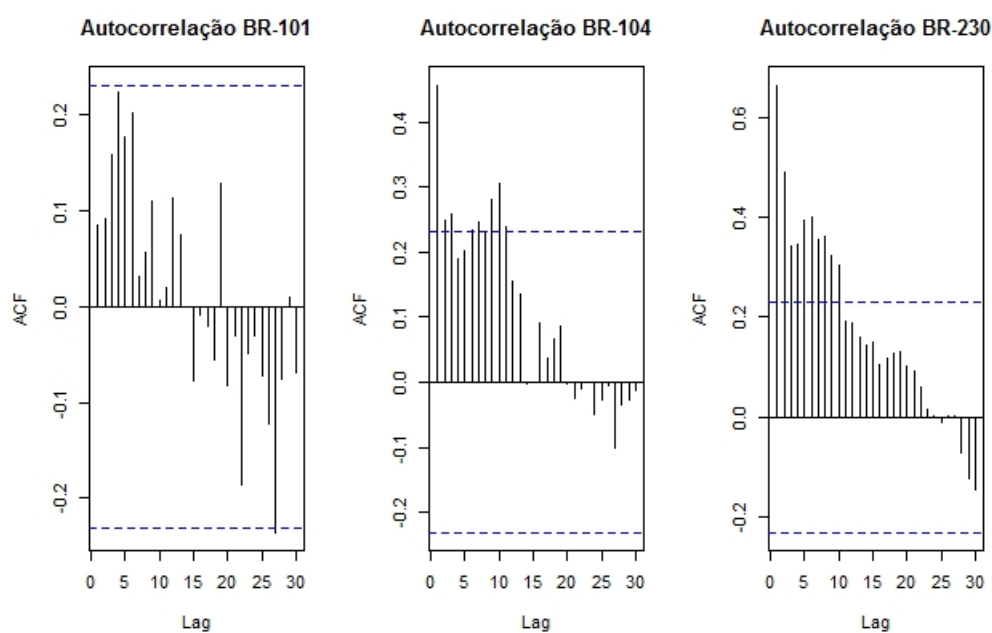


Figura 9 – Gráfico de autocorrelação das BRs.

Para verificação da presença de estacionariedade o teste de Phillips-Perron foi realizado. Na Tabela 2 são representados os resultados obtidos, onde a hipótese nula, de não-estacionariedade foi rejeitada para todas as séries das rodovias. Sendo assim, as séries dos números de acidentes no estado da Paraíba são estacionaria, isto é, os valores das séries se concentram, em torno da média com variabilidade aproximadamente constante.

Tabela 2 – Teste de Estacionariedade Phillips-Perron.

Teste Phillips-Perron	Estatística	P- valor
BR-101	-71,103	<0,01
BR-104	-47,409	<0,01
BR-110	-53,637	<0,01
BR-116	-85,273	<0,01
BR-230	-40,475	<0,01
BR-361	-67,458	<0,01
BR-405	-87,831	<0,01
BR-412	-79,428	<0,01
BR-427	-79,142	<0,01

4.2 Análise Bayesiana

A partir dos resultados apresentados até aqui, resolveu-se comparar o modelo comparado apresentado na metodologia sem sazonalidade (M1) e com sazonalidade (M2), isto é, sem e com o componente s_t no preditor, respectivamente. Para cada modelo foram simuladas 10000 iterações MCMC dispensando as 1000 primeiras simulações (*burn-in*), retendo-se 1 em cada 10 iterações (*thin*). O DIC e raiz quadrada do erro quadrático médio (REQM) dos modelos se encontram na Tabela 3. Percebe-se que, para os dois modelos, o REQM é praticamente o mesmo, isto é, os dois modelos tem a mesma capacidade preditiva. Porém, o DIC do modelo M2 (2158) é bem menor que o DIC do modelo M1 (2914), indicando que a presença de sazonalidade deve ser modelada. Portanto, escolhe-se o modelo M2.

Tabela 3 – Critérios de comparação de modelos.

Modelo	DIC	REQM
M1	2914	4,374
M2	2158	4,384

Na Tabela 4, encontram-se algumas estatísticas a posteriori do modelo escolhido. Como existem 649 parâmetros do tipo $\lambda_{t,j}$ foram escolhidos apenas os últimos correspondentes às taxas de acidentes no mês $t = 72$ (dezembro de 2018) para as rodovias $j = 1, \dots, 9$ (B-101, ..., BR-427), respectivamente. Utilizando os percentis 2, 5% e 97, 5%, verifica-se que o zero não está incluído em nenhum dos intervalos $[P_{2,5\%}; P_{97,5\%}]$, indicando que todos os parâmetros são significativos. A estimativa a posteriori para as taxas de acidentes nas rodovias BR-101, BR-104 e BR-230 são 30,95, 16,93 e 88,42 acidentes por mês, respectivamente. Todos os erros-padrão são, pelo menos, duas vezes menores que as médias a posteriori, indicando mais uma vez a significância das estimativas. Todas as medianas a estão próximas das médias a posteriori, o que indica que as distribuições a

posteriori são aproximadamente simétricas.

Tabela 4 – Estatísticas a posteriori para o modelo M2.

Parâmetro	Média	Erro-padrão	$P_{2,5\%}$	Mediana	$P_{97,5\%}$
a	0,4839	0,02999	0,4258	0,4832	0,545
b	4,288	2,047	1,507	4,114	8,157
$\lambda_{72;BR-101}$	30,950	5,484	21,250	30,650	42,570
$\lambda_{72;BR-104}$	16,930	4,083	9,817	16,610	25,670
$\lambda_{72;BR-110}$	0,475	0,693	<0,001	0,204	2,467
$\lambda_{72;BR-116}$	0,324	0,484	<0,001	0,140	1,665
$\lambda_{72;BR-230}$	88,420	9,490	70,850	88,120	107,800
$\lambda_{72;BR-361}$	10,110	3,078	5,110	9,815	16,920
$\lambda_{72;BR-405}$	3,081	1,673	0,729	2,763	7,096
$\lambda_{72;BR-412}$	4,261	2,017	1,273	3,971	8,996
$\lambda_{72;BR-427}$	0,420	0,606	<0,001	0,192	2,130

Na Figura 10 são observados os *boxplots* para as quantidades preditas de acidentes para cada um dos 72 meses na rodovia BR-101. A quantidade mediana, no período analisado, foi de 31,3 acidentes por mês. Nota-se que, em geral, esta quantidade foi superior a dez acidentes mensais. As maiores medianas se deram para os meses 1 (janeiro de 2013), 20 (agosto de 2014) e 24 (dezembro de 2014). As menores medianas acontecem para os meses 45 e 46, setembro e outubro de 2016, respectivamente. Na Figura 11 estão os *boxplots* para os acidentes preditos para a BR-104 nos 72 meses em estudo. O número de acidentes mediano, para o período, foi 14,3. Os meses com maiores acidentes previstos pelo modelo foram maio e junho de 2013. Na Figura 12 é possível se observar os *boxplots* para os números preditos de acidentes nos 72 meses para o rodovia BR-230. A quantidade mediana a posteriori de acidentes foi 115, considerando todos os 72 meses. As maiores medianas foram apresentadas para setembro, outubro e novembro de 2013. Observe que para os gráficos da BR-104 e BR-230 existe uma tendência de diminuição no número de acidentes no período analisado. Já na BR-101, a tendência de diminuição cessou por volta do mês 50, quando as previsões se estabilizam em torno dos 40 acidentes por mês.

acidentes mensais o modelo é capaz de prever corretamente a verdadeira quantidade.

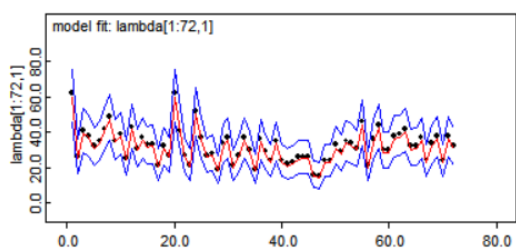


Figura 13 – Acidentes observados versus preditos para a BR-101.

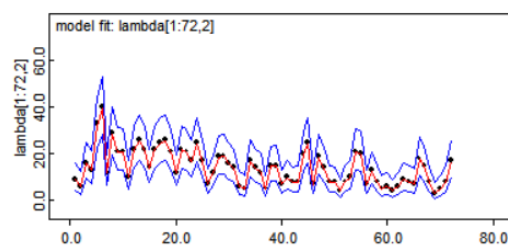


Figura 14 – Acidentes observados versus preditos para a BR-104.

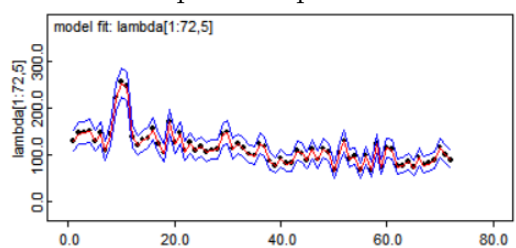


Figura 15 – Acidentes observados versus preditos para a BR-230.

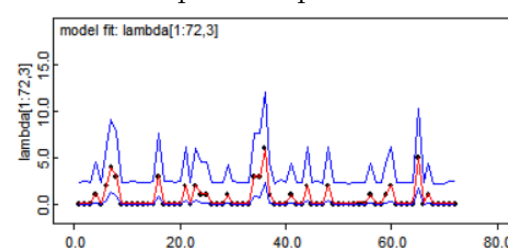


Figura 16 – Acidentes observados versus preditos para a BR-110.

Dos gráficos do modelo apresentados acima, verifica-se que a modelagem prevê corretamente o número observado de acidentes para as rodovias federais que percorrem o estado da Paraíba. Atendendo, de forma satisfatória, o objetivo do estudo.

5 Conclusão

Por meio da abordagem empregada neste trabalho foi possível constatar que das 9 rodovias federais que cruzam o Estado da Paraíba estudadas as que apresentaram maior número de acidentes de trânsito foram as rodovias BR-101, BR-104 e a BR-230. A BR-230 foi a que apontou maior número de acidentes quando comparadas as demais rodovias combinadas.

Através do Modelo Dinâmico de Poisson com sazonalidade foi possível ajustar a quantidade de acidentes nas rodovias e previsões foram realizadas de forma conjunta em função do tamanho das mesmas. O melhor modelo foi discriminado pelo DIC. Portanto, de acordo com esses resultados percebe-se que futuros trabalhos na área podem favorecer o ajuste além da distribuição de Poisson, como incluir covariáveis associadas as rodovias ou a modelagem de efeito de superdispersão ou ainda o ajuste por outras distribuições distribuições como a binomial negativa ou beta-binomial.

Referências

- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. Time series analysis: Forecasting and control. John Wiley and Sons, 2008. Citado na página 13.
- CARMO, C. L. d. et al. Segurança viária em trechos urbanos de rodovias federais. Universidade Federal de São Carlos, 2019. Citado na página 22.
- CONGDON, P. *Bayesian models for categorical data*. [S.l.]: John Wiley & Sons, 2005. Citado na página 19.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979. Citado na página 13.
- FEDERAL-PRF, P. R. et al. Acidentes de trânsito nas rodovias federais brasileiras. caracterização, tendências e custos para a sociedade. IPEA Brasil, 2015. Citado na página 11.
- FERREIRA, M. A.; GAMERMAN, D. Dynamic generalized linear models. *BIOSTATISTICS-BASEL-*, MARCEL DEKKER INC, v. 5, p. 57–72, 2000. Citado na página 20.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. *Journal of Animal Science*, Oxford University Press, v. 63, n. 1, p. 217–244, 1986. Citado na página 14.
- HOFF, P. D. *A first course in Bayesian statistical methods*. [S.l.]: Springer, 2009. v. 580. Citado na página 16.
- LIMA, A. S. d. Incidências e fatores associados aos acidentes de trânsito em mari/pb. 2019. Citado na página 25.
- MAURO, M. L. F. et al. Acidentes de trânsito: perfil epidemiológico de vítimas e caracterização de alguns traços de personalidade de motoristas infratores em campinas, são paulo. [sn], 2001. Citado na página 11.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2006. Citado na página 13.
- MOULDER, B. C.; ALGINA, J. Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, Taylor & Francis, v. 9, n. 1, p. 1–19, 2002. Citado na página 16.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972. Citado na página 19.
- NTZOUFRAS, I. Bayesian modeling using winbugs. Wiley, 2009. Citado 3 vezes nas páginas 14, 15 e 16.

- OMS, O. M. D. S. *Addressing the challenges of the United Nations Decade of Action for Road Safety (2011-2020): outcome of the Second Global High-level Conference on Road Safety – Time for Results: draft resolutions proposed by Brazil and the Dominican Republic*. 2016. Acesso de 18 Junho 2019. Disponível em: <https://apps.who.int/iris/bitstream/handle/10665/250762/B138_CONF4-en.pdf?sequence=1&isAllowed=y>. Citado na página 10.
- PAULINO, C.; TURKMAN, M.; MURTEIRA, B. estatística bayesiana fundação clouste gulbenkian lisboa. 2003. Citado na página 14.
- PHILLIPS, P. C.; PERRON, P. Testing for a unit root in time series regression. *Biometrika*, Oxford University Press, v. 75, n. 2, p. 335–346, 1988. Citado na página 13.
- POLE, A.; WEST, M.; HARRISON, J. *Applied Bayesian forecasting and time series analysis, 1994*. [S.l.]: Chapman & Hall, 1994. Citado na página 15.
- ROBERTSON, A. Prediction equations in quantitative genetics. *Biometrics*, JSTOR, v. 11, n. 1, p. 95–98, 1955. Citado na página 14.
- SANTOS, J. V. d. O. et al. Análise espacial dos acidentes de trânsito nas rodovias federais do estado de alagoas. Universidade Federal de Alagoas, 2019. Citado na página 27.
- TRIANTAFYLLOPOULOS, K. Inference of dynamic generalized linear models: on-line computation and appraisal. *International Statistical Review*, Wiley Online Library, v. 77, n. 3, p. 430–450, 2009. Citado na página 19.
- VASCONCELOS, E. A. *O que é trânsito*. [S.l.]: Brasiliense, 1998. Citado na página 11.
- WEST, M.; HARRISON, J. Bayesian forecasting and dynamic models. Springer, 1997. Citado 2 vezes nas páginas 17 e 19.
- WEST, M.; HARRISON, P. J.; MIGON, H. S. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 80, n. 389, p. 73–83, 1985. Citado na página 18.
- WHO, W. H. O. *Road traffic injuries*. 2013. Acesso em: 05 de Julho 2019. Disponível em: <<https://www.who.int/en/news-room/fact-sheets/detail/road-traffic-injuries>>. Citado na página 11.

Apêndices

APÊNDICE A – Código utilizado

Código BUGS utilizado para obter as estatísticas Bayesianas.

```

model{
  for(t in 1:T){
    for(j in 1:J){
      theta[t,j] ~ dgamma(a, b)
      log(lambda[t,j]) <- alpha[t] + log(theta[t,j]) + s[t] + log(tam[j])
      y[t,j] ~ dpois(lambda[t,j])
      SQ[t,j] <- pow(y[t,j]-lambda[t,j],2)
    }
    tempo[t] <- t
    w1[t] ~ dunif(-1,1)
    w2[t] ~ dunif(-1,1)
    #w[t] ~ dunif(wa,wb)
  }

  alpha[1] <- 0
  for(t in 1:11){s[t] <- 0}
  for(t in 2:T){ alpha[t] <- alpha[t-1] + w1[t] }
  for(t in 12:T){ s[t] <- -sum(s[(t-11):(t-1)]) + w2[t] }

  a ~ dgamma(0.01, 0.01)
  b ~ dgamma(0.01, 0.01)
  #wa ~ dunif(-10,0)
  #wb ~ dunif(0,10)

  REQM <- sqrt(sum(SQ[,,])/(J*T))
}

```