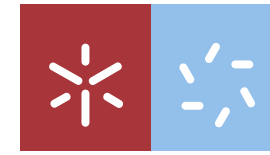




**Modelos de Regressão para Dados de Contagem:
Aplicação a Dados de Contagem Rodoviários**

Cátia Maria Oliveira de Sousa

UMinho | 2022

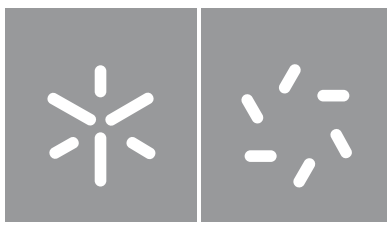


Universidade do Minho
Escola de Ciências

Cátia Maria Oliveira de Sousa

**Modelos de Regressão para Dados
de Contagem: Aplicação a Dados de
Contagem Rodoviários**

outubro de 2022



Universidade do Minho

Escola de Ciências

Cátia Maria Oliveira de Sousa

**Modelos de Regressão para Dados
de Contagem: Aplicação a Dados de
Contagem Rodoviários**

Dissertação de Mestrado

Mestrado em Estatística

Trabalho efetuado sob a orientação do(a)

Professora Doutora Susana Faria

Professora Doutora Elisabete Fraga

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



**Atribuição
CC BY**

<https://creativecommons.org/licenses/by/4.0/>

Agradecimentos

A realização deste trabalho decorreu enquanto trabalhador estudante e consegui-lo terminar foi uma árdua tarefa e só foi possível com a ajuda e apoio de várias pessoas às quais eu quero expressar os meus sinceros agradecimentos, em particular:

À professora Susana Faria, pela incansável disponibilidade, paciência e compreensão, bem como os incentivos dados, e pela partilha de conhecimentos.

À professora Elisabete Fraga, por todo o apoio, ajuda e dedicação.

À minha mãe, o meu maior pilar, que me incentivou e sempre me apoiou em todos os momentos.

Ao meu pai, por apoiar as minhas escolhas ao longo desta jornada.

Ao meu namorado, por todo o apoio e carinho incondicional.

Aos meus irmãos por estarem sempre do meu lado.

Um especial agradecimento ao Helder pelo apoio logístico na minha recolha de dados.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Resumo

Modelos de Regressão para Dados de Contagem: Aplicação a Dados de Contagem Rodoviários (em autoestradas)

Os acidentes de viação apresentam uma enorme carga social e económica que é imposta à sociedade. A modelação de frequência de acidentes é uma ferramenta importante para a gestão de segurança rodoviária. O aumento da procura dos meios de transporte, em especial, transporte rodoviário, tem resultado num claro aumento do número de acidentes nessa envolvente. Consequentemente, um dos grandes desafios para os investigadores e gestores rodoviários passa pela redução dos acidentes nas estradas.

Dessa forma, é apresentado um estudo dos Modelos Lineares Generalizados Mistos com distribuição de Poisson, sendo esta a mais utilizada para modelar dados de contagem.

Esta dissertação visa estudar o efeito das diferentes variáveis que definem o estado do pavimento, a geometria da estrada e o tráfego sobre o número de acidentes através da construção e análise de um modelo de regressão de Poisson com efeitos aleatórios. Os dados utilizados são referentes a segmentos das autoestradas A4, A44 e A41 em Portugal, disponibilizados pela Concessionária Ascendi. O modelo desenvolvido teve como variável resposta o número de acidentes, num período de quatro anos (2014 a 2017).

Palavras-chave: Acidentes; Distribuição de Poisson; Modelos Lineares Generalizados Mistos; Segurança Rodoviária.

Abstract

Regression Models for Count Data: Application to Road Count Data (on highways)

Road accidents present an enormous social and economic burden that is imposed on society. Accident frequency modeling is an important tool for road safety management. The increase in demand for means of transport, especially road transport, has resulted in a clear increase in the number of accidents in this environment. Consequently, one of the great challenges for road investigators and managers is the reduction of road accidents.

Thus, a study of mixed generalized linear models with Poisson distribution is presented, which is the most used to model counting data.

This dissertation aims to study the effect of the different variables that define the state of the pavement, the geometry of the road and the traffic on the number of accidents through the construction and analysis of a Poisson regression model with random effects. The data used refer to segments of the A4, A44 and A41 motorways in Portugal, made available by the Ascendi Concessionaire. The model developed had the number of accidents as a response variable, over a period of four years (2014 to 2017).

Keywords: Accidents; Mixed Generalized Linear Models; Poisson Distribution; Road safety.

Conteúdo

1	Introdução	1
1.1	Objetivos	2
1.2	Estrutura da Dissertação	3
2	Modelos Lineares Generalizados Mistos	4
2.1	Estrutura do modelo linear generalizado misto	4
2.2	Interpretação dos parâmetros do modelo	6
2.3	Inferência	7
2.3.1	Estimação dos parâmetros do modelo	7
2.3.2	Testes de hipótese e seleção do modelo	9
2.3.3	Análise de Resíduos	10
2.4	Modelos de Regressão de <i>Poisson</i> de Efeitos Aleatórios	10
3	Análise e Modelação dos Dados	12
3.1	Base de Dados	12
3.2	Análise Descritiva	21
3.3	Aplicação dos Modelos	40
4	Conclusão e Trabalho Futuro	47

Lista de Figuras

3.1	Mapa A44	13
3.2	Mapa A4	14
3.3	Mapa A41	14
3.4	<i>GripTester</i>	18
3.5	Número de Acidentes por Ano e Autoestrada	22
3.6	Número de Acidentes por Autoestrada e Sentido	23
3.7	<i>Box plot</i> da Variável <i>IRImd</i> e <i>IRImx</i> (A44)	27
3.8	<i>Box plot</i> da Variável <i>RDmd</i> e <i>RDmx</i> (A44)	28
3.9	<i>Box plot</i> da Variável <i>MPDmd</i> e <i>MPDmn</i> (A44)	28
3.10	<i>Box plot</i> da Variável <i>GRIPmd</i> e <i>GRIPmn</i> (A44)	29
3.11	Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A44)	29
3.12	<i>Box plot</i> da Variável <i>IRImd</i> e <i>IRImx</i> (A4)	32
3.13	<i>Box plot</i> da Variável <i>RDmd</i> e <i>RDmx</i> (A4)	33
3.14	<i>Box plot</i> da Variável <i>MPDmd</i> e <i>MPDmn</i> (A4)	33
3.15	<i>Box plot</i> da Variável <i>GRIPmd</i> e <i>GRIPmn</i> (A4)	34
3.16	Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A4)	34
3.17	<i>Box plot</i> da Variável <i>IRImd</i> e <i>IRImx</i> (A41)	37
3.18	<i>Box plot</i> da Variável <i>RDmd</i> e <i>RDmx</i> (A41)	37
3.19	<i>Box plot</i> da Variável <i>MPDmd</i> e <i>MPDmn</i> (A41)	38
3.20	<i>Box plot</i> da Variável <i>GRIPmd</i> e <i>GRIPmn</i> (A41)	38
3.21	Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A41)	39

3.22 Gráfico de Dispersão do Número de Acidentes por Ano e Autoestrada	40
3.23 Gráfico QQ dos Resíduos	45
3.24 Gráfico dos Resíduos <i>vs</i> Valores Preditos	46

Lista de Tabelas

3.1	Descrição das Variáveis da Base de Dados	16
3.2	Número Total de Acidentes por Autoestrada e Ano	22
3.3	Número de Acidentes por Autoestrada e Sentido	23
3.4	Análise Descritiva das Variáveis correspondentes às Características Geométricas.	24
3.5	Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A44)	26
3.6	Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário Anual (A44)	30
3.7	Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A4)	31
3.8	Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário (A4)	35
3.9	Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A41)	36
3.10	Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário (A41)	39
3.11	Reescala de variáveis	41
3.12	Resultados dos Modelos Simples das variáveis relativas ao estado do pavimento	42

3.13 Estimativas do modelo completo de regressão de Poisson com efeitos	
aleatórios	43
3.14 Estimativas do modelo final	44

Lista de abreviaturas

ANSR - Autoridade Nacional Segurança Rodoviária

C - Sentido Crescente

D - Sentido Decrescente

EmAR - Extensão(m) em Reta

EmCC - Extensão(m) em Curva Circular

EmCCV - Extensão(m) em Curva Concâva

EmCCX - Extensão(m) em Curva Convexa

EmCL - Extensão(m) em Clotóide

EmT - Extensão(m) em Trainel

GLM - *Generalized Linear Models* (em português, Modelos Lineares Generalizados)

GLMM - *Generalized Linear Mixed Models* (em português, Modelos Lineares Generalizados Mistos)

GRIP - Coeficiente de Atrito

IRI - *International Roughness Index* (em português, Índice Internacional de Rugosidade)

LMM - *Linear Mixed Models* (em português, Modelos Lineares Mistos)

MIN - Média da Inclinação Negativa

MIP - Média da Inclinação Positiva

MIR - Média do Inverso do Raio

MPA - Média do Parâmetro A

MPD - *Mean Profile Depth* (em português, Profundidade Média de Perfil)

MRCN - Média do Raio da Curva Negativa

MRCP - Média do Raio da Curva Positiva

RD - Rodeiras

TMDA - Tráfego Médio Diário Anual

VC - Via da Central

VD - Via da Direita

VE - Via da Esquerda

Capítulo 1

Introdução

Atualmente, a Sinistralidade Rodoviária é um problema muito grave, sendo considerada uma questão de Saúde Pública. Os acidentes rodoviários são a principal causa de morte a nível mundial na faixa etária mais jovem, e a oitava em todas as idades, apresentando uma elevada carga social e económica na população (ANSR, 2020).

A segurança dos utilizadores do sistema rodoviário constitui, cada vez mais, um dos principais objetivos no planeamento e operação deste meio de transporte (Sinay and Tamayo, 2005). Deste modo, as empresas e as instituições que atuam na segurança rodoviária, apresentam uma grande responsabilidade na implementação de medidas que incluem, entre outras, medidas que visam melhorar a concessão e operação das vias. Estas medidas podem representar uma importante redução da frequência de acidentes, assim como, na gravidade dos ferimentos sofridos pelos seres humanos envolvidos.

A nível mundial, quando se avaliam as causas que afetam a segurança rodoviária, conclui-se que esta depende da contribuição de múltiplos fatores, nomeadamente: (i) condutor; (ii) estrada e ambiente rodoviário; (iii) veículo. Estes fatores podem agir de forma isolada quer associados entre si. Durante décadas tem sido evidenciado o consenso sobre a responsabilidade do comportamento do condutor na maioria dos acidentes de viação. Porém, salienta-se a influência das características da estrada da infraestrutura rodoviária e, particularmente, das estratégias de gestão e controlo da velocidade ao longo de um determinado itinerário.

Em Portugal, a Autoridade Nacional de Segurança Rodoviária (ANSR) é a instituição responsável pela gestão da segurança rodoviária. Esta entidade tem

como missão o planeamento e a coordenação no território português, no que diz respeito à segurança rodoviária, bem como a aplicação do direito contraordenacional rodoviário.

Devido aos enormes custos dos acidentes de viação para a sociedade, o conhecimento dos fatores que afetam a ocorrência dum acidente tem sido uma área de investigação há muitas décadas. Estes estudos resultam, na sua grande maioria, na descrição da relação entre os acidentes de viação, o tráfego, número de interseções, características do piso, entre outras.

Em dados de contagem, é usual a mesma unidade experimental ser observada ao longo do tempo, pelo que se espera que haja uma correlação não nula entre as várias observações das unidades individuais, levando à violação do pressuposto de independência entre todas as observações. Uma forma de se ter em consideração os dados correlacionados passa por considerar os modelos lineares generalizados mistos (GLMM). Estes modelo são uma extensão dos Modelos Lineares Generalizados (GLM) dado que permitem a inclusão de efeitos aleatórios no preditor linear para que se tenha em conta a correlação existente entre as várias observações da mesma unidade experimental (Cabral e Gonçalves, 2011; Breslow e Clayton, 1993; Hedker, 2005).

1.1 Objetivos

Este trabalho tem como principal objetivo o desenvolvimento de modelos de regressão com efeitos aleatórios para dados de contagem relativamente ao número de acidentes de viação para segmentos das autoestradas A4, A44 e A41. Estes segmentos são Gulpilhares - Coimbrões (A44), Matosinhos - Águas Santas (A4) e Perafita - Alfena (A41). Para a modelação dos modelos são utilizados a frequência dos acidentes de viação, o tráfego médio diário anual (TMDA), as variáveis referentes ao estado do pavimento e às características geométricas tanto em perfil como em planta da via.

No âmbito da segurança rodoviária, este estudo permite, através de métodos estatísticos, relacionar o número de acidentes com outras variáveis, criando uma ferramenta robusta que permite por um lado, avaliar o efeito dessas variáveis na sinistralidade e por outro, estimar a sinistralidade em cenários específicos.

Todos os estudos computacionais realizados ao longo deste trabalho foram desenvolvidos recorrendo à linguagem de programação R (*R Development Core Team*, 2022)

1.2 Estrutura da Dissertação

A estrutura da dissertação encontra-se dividida por quatro capítulos. No Capítulo 1 é realizada uma introdução, uma breve explicação do tema e âmbito de estudo. No Capítulo 2 descrevem-se os conteúdos teóricos relacionados com os modelos lineares generalizados mistos, mais concretamente, a descrição dos modelos lineares generalizados mistos, a estimação dos parâmetros, testes de hipótese, seleção do modelo e análise dos resíduos. A aplicação das metodologias aos dados é apresentada no Capítulo 3. Inicialmente, é feita uma breve análise das autoestradas. Depois de apresentados os dados em estudo, realiza-se uma análise exploratória dos mesmos, seguindo-se a aplicação das metodologias. As principais conclusões do trabalho desenvolvido e sobre os resultados obtidos são descritas no Capítulo 4, assim como algumas sugestões de investigação para trabalho futuro.

Capítulo 2

Modelos Lineares Generalizados Mistos

Os Modelos Lineares Generalizados Mistos (GLMM) são uma extensão dos modelos lineares generalizados (GLM) que permitem a inclusão de efeitos aleatórios no preditor linear. O facto dos GLMM permitirem a introdução de efeitos aleatórios, permite que a estrutura de correlação entre observações de um mesmo indivíduo possa ser modelada.

Como mencionam Cabral e Gonçalves (2011), o objetivo dos Modelos Lineares Generalizados Mistos é descrever as alterações da resposta média de cada indivíduo e a relação destas alterações com as covariáveis de interesse. Estes modelos pretendem, assim, inferir sobre o indivíduo e não sobre a população (Cabral and Gonçalves, 2011).

A nível computacional e de interpretação dos parâmetros estimados, estes modelos apresentam algumas complicações provenientes da estrutura não linear agregada à inclusão de efeitos aleatórios.

2.1 Estrutura do modelo linear generalizado misto

Seja y_{it} o valor da resposta observada para o indivíduo i ($i = 1, \dots, n$) no instante $t = 1, \dots, T_i$, ou seja, y_{it} é a realização da variável resposta Y_{it} e

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})^T \quad (2.1)$$

representa o vetor das variáveis resposta para o indivíduo i . O perfil de cada indivíduo é assim dado pelo vetor $\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})^T$.

Seja $\mathbf{x}_{it}^T = (x_{1it}, \dots, x_{pit})$, o vetor das covariáveis de interesse associado a cada Y_{it} , de dimensão p , a matriz das covariáveis, \mathbf{X}_i , de ordem $(T_i \times p)$ e o vetor dos parâmetros de regressão desconhecidos $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, de dimensão $(p \times 1)$. Por fim,

$$\mathbf{z}_{it}^T = (z_{1it}, \dots, z_{qit}) \quad (2.2)$$

é o vetor $(q \times 1)$ de covariáveis (usualmente um subconjunto de \mathbf{x}_{it}) associado a um vetor $(q \times 1)$ de efeitos aleatórios \mathbf{b}_i (Cabral and Gonçalves, 2011).

Os modelos lineares generalizados mistos podem ser especificados da seguinte forma:

1. A variável resposta $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})$, condicionada pelo vetor dos efeitos aleatórios \mathbf{b}_i , segue uma distribuição pertencente à família exponencial com função densidade

$$f(y_{it}|\mathbf{b}_i) = \exp \left\{ \frac{\omega_{it}}{\phi} (y_{it}\theta_{it} - b(\theta_{it})) + c(y_{it}, \phi) \right\} \quad (2.3)$$

O valor médio e a variância condicional são expressos, respetivamente, por:

$$E(Y_{it}|\mathbf{b}_i) = \mu_{it} = b'(\theta_{it}), \quad (2.4)$$

$$Var(Y_{it}|\mathbf{b}_i) = \sigma_{it} = b''(\theta_{it}) \frac{\phi}{\omega_{it}}. \quad (2.5)$$

Tal como nos modelos lineares generalizados, assume-se que:

$$g(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i, \quad (2.6)$$

$$\sigma_{it} = V(\mu_{it}) \frac{\phi}{\omega_{it}}, \quad (2.7)$$

onde $g(\cdot)$ é uma função de ligação e V é a função de variância, ambas conhecidas, ϕ é o parâmetro de dispersão e ω_{it} é uma constante conhecida atribuída às observações da variável resposta.

Uma vez fixados os efeitos aleatório \mathbf{b}_i , as variáveis resposta assumem-se mutuamente independentes.

2. A média condicional da variável resposta Y_{it} , depende dos efeitos fixos e dos efeitos aleatórios. Deste modo, o modelo linear generalizado misto pode ser escrito por

$$g \{E(Y_{it}|\mathbf{b}_i)\} = \eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i \quad (2.8)$$

3. Por fim, resta-nos especificar a distribuição dos efeitos aleatórios \mathbf{b}_i . À partida, qualquer distribuição multivariada poderia ser considerada para b_i . Contudo, na prática, assume-se que a distribuição comum aos efeitos aleatórios \mathbf{b}_i é a Gaussiana Multivariada com valor esperado zero e matriz de variância-covariância D . Destaca-se o facto dos efeitos aleatórios \mathbf{b}_i se assumirem independentes das covariáveis.

Modelo linear generalizado misto para contagens

Considere-se que Y_{it} , a variável resposta do indivíduo i no instante t , sob a forma de contagens. Nesta situação tem-se:

1. A variável resposta Y_{it} , condicional aos efeitos aleatórios \mathbf{b}_i são independentes e seguem uma distribuição de Poisson, com

$$E(\mathbf{Y}_{it}|\mathbf{b}_i) = Var(\mathbf{Y}_{it}|\mathbf{b}_i) \quad (\phi = 1). \quad (2.9)$$

2. A média μ_{it} está relacionada com a componente sistemática do modelo (que supomos incluir um termo constante) através da equação

$$\begin{aligned} \log(\mu_{it}) &= (1, \mathbf{x}_{it}^T)^T \boldsymbol{\beta} + (1, \mathbf{z}_{it}^T)^T \mathbf{b}_i \\ &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})x_{1it} + \dots + (\beta_p + b_{pi})x_{pit}, \end{aligned} \quad (2.10)$$

onde \log é a função de ligação canónica quando estamos na presença de dados de contagem.

3. Por fim, assume-se que o vector dos efeitos aleatórios, \mathbf{b}_i , segue uma distribuição Gaussiana de valor médio zero e matriz de variância-covariância D , de dimensão $(p + 2) \times (p + 2)$.

2.2 Interpretação dos parâmetros do modelo

A interpretação dos parâmetros fixos nos GLMM será assim afetada pelos efeitos aleatórios presentes no modelo, como explicam Cabral e Gonçalves (2011).

Nos GLMM

$$E(Y_{it}) = E \{E(Y_{it}|\mathbf{b}_i)\} = E [g^{-1}(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i)], \quad (2.11)$$

que, em geral, não pode ser simplificado devido à presença de funções não lineares em $g^{-1}(\cdot)$ e, portanto,

$$E[g(Y_{it}|\mathbf{b}_i)] \neq g[E(Y_{it}|\mathbf{b}_i)]. \quad (2.12)$$

Existe alteração na interpretação dos parâmetros dos efeitos fixos quando a função de ligação canônica dos GLMM não é linear.

De acordo com Cabral e Gonçalves (2011), nos GLMM podemos considerar as seguintes interpretações para os parâmetros fixos e para a componente aleatória do modelo:

- (i) **Parte Fixa:** As componentes β_j , $j = 0, 1, \dots, p$, do vetor β têm uma interpretação em termos específicos do indivíduo. Elas representam a influência das covariáveis de interesse na alteração da resposta média esperada de um determinado indivíduo. Estes coeficientes, β , são designados por *coeficientes de regressão específicos do indivíduo* dado que a interpretação de β_j depende dos efeitos aleatórios do i -ésimo indivíduo assumirem um valor fixo.
- (ii) **Parte Aleatória:** Uma forma de interpretar as estimativas das variâncias dos efeitos aleatórios passa por considerar percentis dos efeitos aleatórios baseados na hipótese Gaussiana e, com base nos percentis obtidos calcular os limites de variação dos valores esperados. Outra maneira é considerar a representação gráfica dos perfis de cada indivíduo.

2.3 Inferência

Para que se possa aplicar a metodologia dos Modelos Lineares Generalizados Mistos (GLMM) a um conjunto de dados, há necessidade, após a formulação do modelo que se pensa adequado, de proceder à realização de inferências sobre esse modelo. A inferência nos GLMM é baseada na verosimilhança. Com efeito, não só o Método da Máxima Verosimilhança é o método de eleição para estimar os parâmetros β e a matriz variância-covariância \mathbf{D} , como também os testes de hipóteses sobre os parâmetros do modelo são baseados na verosimilhança.

2.3.1 Estimação dos parâmetros do modelo

À semelhança do que acontece nos GLM, a estimação dos parâmetros nos Modelos Lineares Generalizados Mistos também recorre ao Método da Máxima Verosimilhança. Nesta situação, a estimação de β e dos parâmetros \mathbf{D} é feita através da

maximização da função de verosimilhança, obtida através da integração da função de verosimilhança sobre os efeitos aleatórios \mathbf{b}_i . A contribuição do i -ésimo indivíduo para a função de verosimilhança é

$$L^A(\boldsymbol{\beta}, \phi, \mathbf{D}) = \int \prod_{t=1}^{T_i} f(y_{it}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i, \quad (2.13)$$

onde $\prod_{t=1}^{T_i} f(y_{it}|\mathbf{b}_i, \boldsymbol{\beta}, \phi)$ diz respeito à função verosimilhança do modelo com efeitos aleatórios e $f(\mathbf{b}_i|\mathbf{D})$ representa a distribuição dos efeitos aleatórios, assumida como sendo uma distribuição normal multivariada. O expoente A na função de verosimilhança indica que se trata da função de verosimilhança do modelo com efeitos aleatórios (Cabral and Gonçalves, 2011).

Já a função de verosimilhança para a amostra, contendo os n indivíduos é expressa por

$$\begin{aligned} L^A(\boldsymbol{\beta}, \phi, \mathbf{D}) &= \prod_{i=1}^n L_i(\boldsymbol{\beta}, \phi, \mathbf{D}) \\ &= \prod_{i=1}^n \int L_i^F(\boldsymbol{\beta}, \phi, \mathbf{D}) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i, \end{aligned} \quad (2.14)$$

onde o expoente F na função de verosimilhança indica que se trata da função de verosimilhança do modelo com efeitos fixos. Tomando o logaritmo da função verosimilhança, tem-se, naturalmente,

$$l^A(\boldsymbol{\beta}, \phi, \mathbf{D}) = \sum_{i=1}^n \log L_i^A(\boldsymbol{\beta}, \phi, \mathbf{D}) \quad (2.15)$$

Maximizando a função (2.15), obtém-se os estimadores de máxima verosimilhança dos parâmetros de regressão $\boldsymbol{\beta}$ e da matriz de variância-covariância dos efeitos aleatórios \mathbf{b}_i . Contudo, ao contrário do que acontece nos GLM, onde estes parâmetros são facilmente estimados através do método iterativo de mínimos quadrados ponderados, nos modelos lineares generalizados mistos este processo não permite calcular estas estimativas. Neste tipo de modelos, a maximização da função log-verosimilhança é bastante complexa e o problema, geralmente, não tem uma solução analítica. Conseqüentemente, para que se possa inferir com base na verosimilhança, tem de se recorrer ao cálculo numérico. Como qualquer método, o cálculo numérico possui limitações, podendo tornar a obtenção das estimativas um processo computacionalmente intensivo, aumentando o seu tempo de resposta

quando se está na presença de grandes amostras e/ou quando a estrutura do modelo é complexa.

Como referem Cabral e Gonçalves (2011), de forma a resolver este problema têm sido propostas várias aproximações numéricas:

- (i) as que são baseadas na aproximação de dados (consideram o desenvolvimento da série de Taylor e insere-se na metodologia do método da quasi-verosimilhança penalizada (PQL), proposto por Breslow and Clayton (1993) e o método da quasi-verosimilhança marginal (MQL) apresentado por Goldenstein (1991));
- (ii) as que são baseadas na aproximação da função a integrar (têm como base a aproximação de Laplace);
- (iii) as que são baseadas na aproximação do integral (salienta-se, por exemplo, o método numérico da quadratura de Gauss-Hermite).

Outros métodos têm sido implementados na tentativa de evitar os problemas de integração numérica, como por exemplo, a utilização de Modelos de Cadeias de Markov via métodos de Monte Carlo (Clayton, 1996).

2.3.2 Testes de hipótese e seleção do modelo

Nos modelos lineares generalizados mistos, para comparar a significância estatística dos parâmetros de regressão β pode ser utilizado o teste de Wald. Já quando se pretende comparar a estrutura fixa de dois modelos encaixados que possuam os mesmos efeitos aleatório recorre-se, geralmente, a testes de razão de verosimilhança. O critério de seleção de modelos, quando estes não são encaixados, passa por considerar o Critério de Informação de *Akaike* (AIC) ou Critério de Informação Bayesiana (BIC).

Cabral e Gonçalves (2011) alertam para o facto de que, em qualquer modelo deve ter-se em atenção qual o método utilizado na maximização da função de verosimilhança, dado que esta tem de se basear nos dados e não na aproximação aos dados.

Caso se pretenda testar a presença de efeitos aleatórios, tem-se um problema fronteira (Cabral e Gonçalves, 2011) e, portanto, assume-se que a distribuição assintótica sob a hipótese nula para uma estatística de teste de razão de verosimilhança é uma mistura de Qui-quadrados.

2.3.3 Análise de Resíduos

A análise de resíduos é uma forma eficiente de verificar a qualidade de ajuste do modelo. Nesta pretende-se identificar valores da variável resposta que apresentem algum comportamento atípico. Isto é, que possuem comportamentos fora do padrão esperado em relação ao assumido para a variável y_i . Por meio desta análise é possível investigar possíveis falhas de suposições feitas a respeito do modelo, tais como a escolha da função de ligação ou da distribuição da variável resposta. Esta análise é feita recorrendo a métodos gráficos.

Resíduos Quantílicos Aleatorizados

Os resíduos quantílicos aleatorizados, que foram introduzidos por Dunn e Smyth (1996), apresentam distribuição normal, independentemente da distribuição da variável resposta e da sua dispersão. Assumindo que y_i é uma variável discreta, os resíduos quantílicos aleatorizados são definidos da seguinte forma.

$$\hat{r}_i^q = \Phi^{-1}\{u_i\}, i = 1, 2, \dots, n \quad (2.16)$$

em que Φ é a função da distribuição acumulada da Normal Padrão e u_i é uma variável aleatória uniformemente distribuída no intervalo $[a_i, b_i]$, com $a_i = F(y_i - 1; \hat{\mu}_i)$ e $b_i = F(y_i; \hat{\mu}_i)$. Assim, se os parâmetros do modelo são consistentemente estimados, então a distribuição \hat{r}_i^q converge para uma distribuição Normal padrão.

2.4 Modelos de Regressão de *Poisson* de Efeitos Aleatórios

O modelo de regressão de *Poisson* de efeitos mistos é proposto para analisar a frequência de uma variável resposta (dados de contagem). Seja $y_{it} = 0, 1, \dots$ o valor da resposta observada para o individuo i no instante t extraída de uma distribuição de *Poisson*.

O modelo de regressão de *Poisson* com efeitos mistos especifica o número esperado de contagens, dado por

$$\log(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i. \quad (2.17)$$

Aplicando a inversa da função logarítmica a ambos os lados da equação (2.17) obtém-se o número esperado de contagem, dado por:

$$\mu_{it} = \exp(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i). \quad (2.18)$$

Dado o processo de *Poisson* para a realização y_{it} , a probabilidade de $y_{it} = y$ condicional aos efeitos aleatórios \mathbf{b}_i , é dada por:

$$P(y_{it} = y | \mathbf{b}_i, \mathbf{x}_{it}, \mathbf{z}_{it}) = \exp(-\mu_{it}) \frac{(\mu_{it})^y}{y!}. \quad (2.19)$$

Capítulo 3

Análise e Modelação dos Dados

Como já referido anteriormente, o objetivo deste trabalho visa estudar o efeito de diferentes variáveis que definem o estado do pavimento, a geometria da estrada e o tráfego sobre a sinistralidade rodoviária, nos segmentos em estudo das autoestradas A4, A44 e A41. Para este estudo é considerado o modelo de regressão de *Poisson* com efeitos mistos.

Os valores necessários para a formação da base de dados foram cedidos pela Concessionária Ascendi e referem-se a três segmentos de autoestradas da região norte de Portugal concessionados pela mesma.

Este capítulo inicia-se com a descrição da base de dados, seguindo-se a análise exploratória dos dados. Por fim, os dados em estudo são modelados através do modelo de regressão de *Poisson* com efeitos aleatórios.

A implementação do modelo na análise será efetuada recorrendo à linguagem de programação R (*R Development Core Team* (2022)). Todas as bibliotecas necessárias para a implementação deste modelo serão devidamente mencionadas ao longo da análise efetuada.

3.1 Base de Dados

A base de dados em estudo refere-se a três autoestradas da região norte de Portugal, nomeadamente A44, A4 e A41. Nesta base de dados cada observação corresponde a um segmento de autoestrada, com uma extensão de 500m. Existem quinze segmentos distintos para a A4, oito para a A44 e vinte e nove para a A41. A cada segmento foram associadas variáveis que definem a sinistralidade, a geometria da estrada, o tráfego e o estado do pavimento num período compreendido entre 2014 e 2017.

- **A44: Autoestrada de Vila Nova de Gaia**

Esta autoestrada permite o acesso da zona litoral sul do Porto a esta cidade, representando uma alternativa ao tráfego que passa no centro de Gaia. O troço utilizado neste estudo será apenas o troço que liga Gulpilhares a Coimbrões com uma extensão de 3,926km. Este troço foi construído sobre a antiga EN109, fazendo inicialmente parte da autoestrada A29. Por este motivo, não possui um perfil adequado ao de uma autoestrada pois as faixas de rodagem são estreitas.

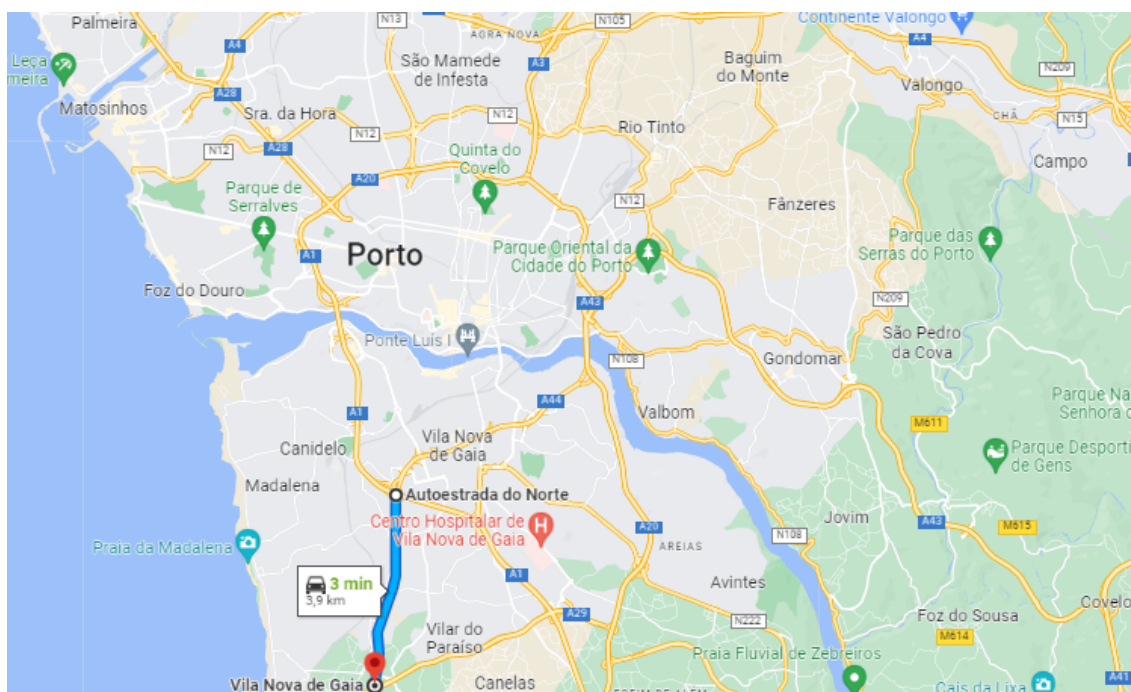


Figura 3.1: Mapa A44

- **A4: Autoestrada de Trás-os-Montes e Alto Douro**

Esta autoestrada liga Matosinhos a Bragança, com uma extensão total de 223km, sendo que apenas irá ser estudado o troço de Matosinhos a Águas Santas, inaugurado em 2006, com extensão de 7,504km.

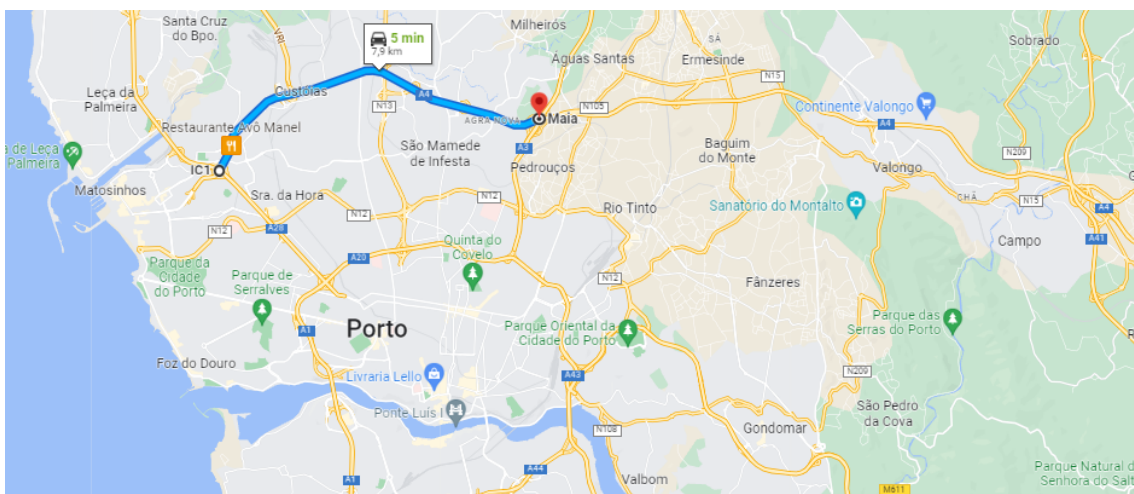


Figura 3.2: Mapa A4

- **A41: CREP- Circular Regional Exterior do Porto**

Esta autoestrada circunda a região do Grande Porto, contribuindo para o descongestionamento da Via de Cintura Interna e facilitando as ligações entre o sul e o leste da região. Esta autoestrada liga Perafita a Espinho tendo uma extensão total de 62km. Apenas estará em estudo o troço de Perafita a Alfena com uma extensão de 14,500km.

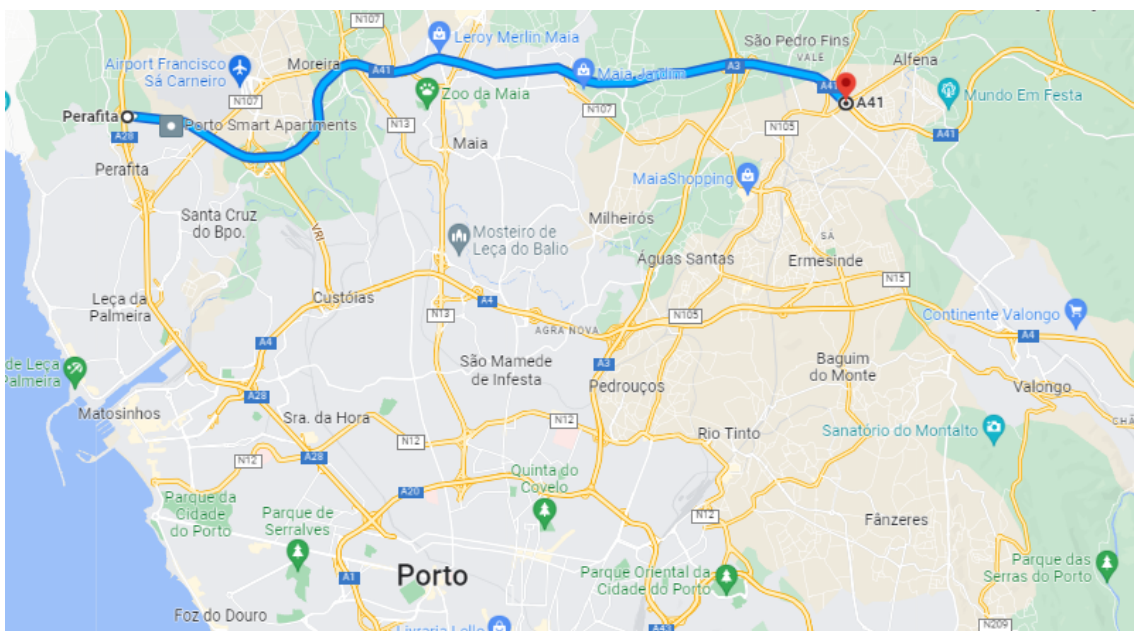


Figura 3.3: Mapa A41

Neste estudo a extensão analisada totaliza 25,930km correspondente a 7,504km da A4, 3,926km da A44 e 14,500km da A41. Na construção da base de dados, os

dados foram agrupados por segmentos com extensão de 500m, extensão já utilizada em outros trabalhos, calculando-se as médias aritméticas de todos os valores das variáveis daquele segmento. Estes valores foram obtidos das autoestradas compostas por três vias (via da esquerda- VE, via central- VC e via da direita- VD) em ambos os sentidos (sentido crescente e sentido decrescente). Posteriormente os dados foram agregados por sentido, calculando as médias aritméticas, os máximos e os mínimos dos valores das variáveis de todas as vias daquele segmento, obtendo assim 416 observações e 32 variáveis (Tabela 3.1). Destas observações 64 referem-se à autoestrada A44, 120 são referentes à A4 e por fim 232 observações são relativas à A29. Da mesma forma das 32 variáveis, 8 correspondem ao estado do pavimento, 5 às características geométricas em planta, 7 às características geométricas em perfil, 3 ao tráfego e as 9 restantes complementam a descrição da base de dados.

A base de dados refere-se aos anos de 2014, 2015, 2016 e 2017. Sendo que

- Para a variável do número de acidentes (*ACID*) existem valores para cada um dos anos. Esta variável será a variável resposta nos modelos construídos,
- As variáveis referentes às características geométricas, tanto em planta como em perfil, são características da infraestrutura da autoestrada, não se alteram com o tempo.
- Devido ao facto de apenas serem realizadas medições para o estado do pavimento de quatro em quatro anos, neste caso, realizadas no início de 2014 e início de 2018, foram associados os valores obtidos dessas características no ano de 2014 aos anos 2014 e 2015 e os resultados do ano de 2018 associados aos anos 2016 e 2017.
- Para as variáveis referentes ao tráfego (*TMDA*) foram associados os valores de 2014 a 2014 e 2015 e os valores de 2016 a 2016 e 2017.

Variáveis referentes ao estado do pavimento

Na superfície do pavimento é possível avaliar o estado do pavimento, que é definido pelos indicadores de estado de conservação. Os indicadores de estado de conservação dos pavimentos obtidos habitualmente numa rede de autoestradas são:

- *IRI*: Índice Internacional de Rugosidade (m/km)
Indica o quão irregular é a estrada. Este valor é obtido de 100m em 100m

Tabela 3.1: Descrição das Variáveis da Base de Dados

	VARIÁVEIS	DESCRIÇÃO
ESTADO DO PAVIMENTO	<i>IRImd</i>	Índice Internacional de Rugosidade - calculado através de uma média aritmética (m/km)
	<i>IRImx</i>	Índice Internacional de Rugosidade - calculado através do valor máximo (m/km)
	<i>RDmd</i>	Rodeiras - calculado através de uma média aritmética (mm)
	<i>RDmx</i>	Rodeiras - calculado através do valor máximo (mm)
	<i>MPDmd</i>	Profundidade Média de Perfil - calculado através de uma média aritmética (mm)
	<i>MPDmn</i>	Profundidade Média de Perfil - calculado através do valor mínimo (mm)
	<i>GRIPmd</i>	Coefficiente de Atrito - calculado através de uma média aritmética (mm)
	<i>GRIPmn</i>	Coefficiente de Atrito - calculado através do valor mínimo (GN- grip number)
CARACTERÍSTICAS GEOMÉTRICAS EM PLANTA	<i>EmAR</i>	Extensão e Alinhamentos Retos (m)
	<i>EmCL</i>	Extensão em Clotóide (m)
	<i>EmCC</i>	Extensão em Curva Circular (m)
	<i>MIR</i>	Média do Inverso do Raio (m-1)
	<i>MPA</i>	Média do Parâmetro A (m)
CARACTERÍSTICAS GEOMÉTRICAS EM PERFIL	<i>EmT</i>	Extensão em Trainel (m)
	<i>EmCCX</i>	Extensão em Curva Convexa (m)
	<i>EmCCV</i>	Extensão em Curva Concáva (m)
	<i>MIP</i>	Média da Inclinação Positiva (m/m)
	<i>MIN</i>	Média da Inclinação Negativa (m/m)
	<i>MRCP</i>	Média do Raio Positivo da Curva (m)
	<i>MRCN</i>	Média do Raio Negativo da Curva (m)
TRÁFEGO MÉDIO DIÁRIO ANUAL	<i>Ligeiros</i>	Número de Veículos Ligeiros
	<i>Pesados</i>	Número de Veículos Pesados
	<i>Total</i>	Número Total de Veículos (Ligeiros + Pesados)
VARIÁVEIS COMPLEMENTARES	<i>NÓ</i>	Número de nós presente no segmento
	<i>TROÇO</i>	Identificação de cada segmento
	<i>PKi</i>	Valor do Km inicial do Troço
	<i>PKf</i>	Valor do Km Final do Troço
	<i>ACID</i>	Número de Acidentes
	<i>NVIAS</i>	Número de Vias
	<i>ANO</i>	Ano do Acidente
	<i>SENTIDO</i>	Indica o Sentido (C- Crescente e D- Decrescente)
	<i>AUTO</i>	Indica a Autoestrada (A4, A44 e A41)

assim, como os segmentos em estudo têm uma extensão de 500m, o valor do índice é obtido calculando a média aritmética dos valores de cada segmento de 500m.

Em relação a este índice foram construídas duas variáveis:

1. *IRImd* : Média aritmética do IRI nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.
2. *IRImx* : Valor máximo do IRI nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.

- *RD*: Rodeiras (mm)

A profundidade das rodeiras é o indicador que melhor traduz a irregularidade transversal. A avaliação deste é importante pois, a presença de rodeiras para além de poder afetar as condições de conforto, é o fator de degradação da segurança quando a superfície do pavimento se encontra molhada. Sendo as medições realizadas de 10m em 10m, o valor para esta variável foi obtido da média aritmética dos valores de cada segmento de 500m.

Esta característica foi estudada em duas variáveis, sendo:

1. *RDmd* : Média aritmética da profundidade de RD nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.
2. *RDmx* : Valor máximo da profundidade de RD nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.

- *MPD*: Profundidade Média de Perfil (mm)

Este indicador caracteriza a textura da superfície. É uma medida de rugosidade. Neste caso, quanto maior for o valor deste indicador mais irregular e melhor serão as condições de drenabilidade e de atrito, espera-se maior segurança. Estes valores são determinados a cada 10m, portanto o valor da extensão de 500m é obtido do cálculo da média aritmética.

Com a agregação dos dados por sentido, obtemos duas variáveis:

1. *MPDmd* : Média aritmética do índice MPD nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.
2. *MPDmn* : Valor mínimo do índice MPD nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.

- *GRIP*: Coeficiente de Atrito (GN- grip number)

O atrito de um pavimento pode ser avaliado através de dois parâmetros. O coeficiente de atrito longitudinal (CAL), que é importante na avaliação da distância de paragem de um veículo, e o coeficiente de atrito transversal (CAT), que traduz a segurança de circulação dos veículos em curva. Neste estudo apenas se utiliza a medição realizada através do *GripTester* (Figura 3.4), que devolve o CAL.

Quanto maior for o valor do coeficiente de atrito, maior será a força de atrito gerada e conseqüentemente o movimento será feito com maior esforço, portanto a paragem de um veículo é realizada mais eficientemente. Da mesma forma que as outras características, existem duas variáveis em estudo:

1. *GRIP_{md}* : Média aritmética do índice GRIP nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.
2. *GRIP_{mn}* : Valor mínimo do índice GRIP nos 500m, considerando as observações efectuadas em todas as vias (VE, VC, VD) em cada sentido.



Figura 3.4: *GripTester*

Variáveis referentes às características geométricas em planta

O traçado em planta de uma estrada é composto por curvas circulares de raio constante, curvas de transição de raio variável e alinhamentos retos de raio infinito. As curvas são introduzidas no traçado para realizar a concordância entre alinhamentos retos consecutivos e abrangem dois tipos de curvas: as circulares e as de transição, habitualmente definidas por um troço de clotóide.

- *EmAR*: Extensão de Alinhamentos Retos (em metros)
Os alinhamentos retos são elementos com raio infinito, que permitem movimentos uniformes e facilitam as ultrapassagens entre veículos. Contudo, provocam o aumento da duração do encadeamento na condução noturna, podem tornar a condução monótona, dificultam a análise sobre as velocidades e distâncias em relação a outros veículos e, geralmente não se integram bem na morfologia do terreno natural.
- *EmCL*: Extensão da Clotóide (em metros)
As clotóides são curvas de transição que possibilitam a ligação entre alinhamentos retos e curvas circulares, garantindo uma variação gradual do raio de valor infinito, até ao valor do raio da curva circular ou vice-versa. A sua implementação no traçado assegura um maior nível de segurança e comodidade ótica que torna possível a percepção antecipada do traçado. São definidas pelo parâmetro da clotóide A .
- *EmCC*: Extensão de Curva Circular (em metros)
As curvas circulares são elementos que permitem efetuar a concordância entre alinhamentos retos consecutivos. Assim, sempre que a morfologia do terreno natural o permita, as curvas circulares devem apresentar os maiores valores de raio possíveis, com o intuito de facilitar a visibilidade e a percepção do traçado ao condutor. As curvas circulares são definidas pelo seu raio.
- *MIR*: Média do Inverso do Raio (m-1)
Nesta variável, é calculado o inverso da média do raio pois, como se trata de traçado em autoestrada, os raios das curvas são muito elevados dado que as velocidades praticadas são também elevadas.
- *MPA*: Média do Parâmetro A (m)
O parâmetro A representa um valor que permite avaliar a qualidade de percepção da estrada para o condutor, e define a clotóide.
A equação que define o parâmetro A é dada por

$$A^2 = R \times L \tag{3.1}$$

onde

A - Parâmetro A (clotóide)

R - Raio da Curva Circular

L - Extensão da Curva Circular

Variáveis referentes às características geométricas em perfil

A definição altimétrica de uma infraestrutura rodoviária é concretizada a partir de uma linha contínua localizada ao longo da respetiva plataforma, a qual se designa por rasante e corresponde ao perfil longitudinal da via.

O perfil longitudinal é constituído por trainéis, que são elementos retos ascendentes (quando apresentam uma inclinação positiva) ou descendentes (quando apresentam uma inclinação negativa), e por concordâncias verticais, que se subdividem em curvas côncavas (de raio negativo) ou convexas (de raio positivo). As curvas verticais são elementos da rasante que permitem realizar a concordância entre dois trainéis de diferentes inclinações. As variáveis que resultam destes elementos são:

- *EmT*: Extensão em Trainel (em metros)
Os trainéis são os elementos mais simples da rasante, onde a inclinação é constante. No caso de ser positiva, é designada por rampa, e, caso seja negativa é designada por declive.
- *EmCCX*: Extensão em Curva Convexa (em metros)
As curvas convexas são utilizadas devido à necessidade de assegurar uma distância de visibilidade adequada, de forma a garantir a segurança na circulação da estrada. Estas são usualmente designadas como lombas.
- *EmCCV*: Extensão em Curva Concâva (em metros)
As curvas concâvas são introduzidas no traçado com o intuito de assegurar a visibilidade noturna e a comodidade na circulação da via. Usualmente designadas como depressões.
- *MIP*: Média da Inclinação Positiva (m/m)
Média aritmética dos valores positivos da inclinação.
- *MIN*: Média da Inclinação Negativa (m/m)
Média aritmética dos valores negativos da inclinação.
- *MRCP*: Média do Raio da Curva Positiva (m)
Média aritmética dos valores do raio positivo correspondentes às curvas concâvas.

- *MRCN*: Média do Raio da Curva Negativa (m)
Média aritmética dos valores do raio positivo correspondentes às curvas convexas.

Tráfego Médio Diário Anual (TMDA)

Volume total de tráfego que passa numa estrada durante um ano, dividido pelo número de dias do ano. O volume varia com a presença de nós e com o ano.

- *Ligeiros*: Número total de veículos ligeiros, num segmento de 500m.
- *Pesados*: Número total de veículos pesados, num segmento de 500m.
- *Total*: Número total de veículos ligeiros e pesados, num segmento de 500m.

Variáveis Complementares

- *Nó*: Indica a presença ou ausência de nós no segmento.
- *Troço*: Identificação de cada segmento de 500 metros.
- *PK_i*: Indica o km em que se inicia o segmento de 500 metros.
- *PK_f*: Indica o km em que acaba o segmento de 500 metros.
- *Acid*: Número de acidentes. Variável resposta nos modelos utilizados neste estudo.
- *Nvias*: Número de Vias.
- *Ano*: Indica o ano em que foram observados os dados. Apenas teremos em estudo os anos 2014, 2015, 2016, 2017.
- *Sentido*: Identifica o sentido. Está codificada como *C* - Crescente, *D* - Decrescente.
- *Auto*: Indica a autoestrada. Variável com as categorias: A4, A44 e A41.

3.2 Análise Descritiva

Inicialmente foi realizada uma análise exploratória do número de acidentes, consoante a autoestrada e sentido.

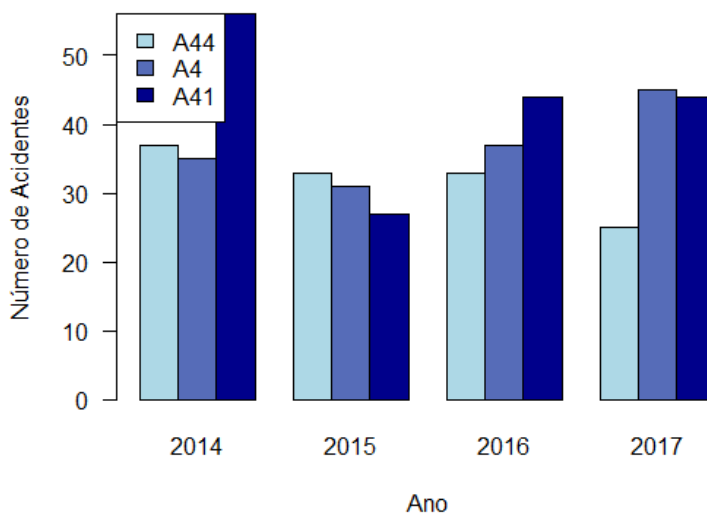


Figura 3.5: Número de Acidentes por Ano e Autoestrada

Da Figura 3.5 observa-se que a autoestrada com maior variação do número de acidentes é a A41, sendo 2014 o ano com mais acidentes e 2015 o ano com menos acidentes. Na autoestrada A44 observa-se um decréscimo do número de acidentes ao longo dos anos. Por fim, na A4 observa-se inicialmente um ligeiro decréscimo, mas a partir de 2016 observa-se um aumento do mesmo, sendo 2017 o ano com maior número de acidentes e 2015 o ano com menor número de acidentes.

Tabela 3.2: Número Total de Acidentes por Autoestrada e Ano

AUTOESTRADA	2014	2015	2016	2017	Nº TOTAL
A4	35	31	37	45	148
A41	56	27	44	44	171
A44	37	33	33	25	128
Nº TOTAL	128	91	114	114	447

Da observação da Tabela 3.2 reforça-se que 2015 é o ano com o menor número total de acidentes e 2014 é o ano com maior número total de acidentes. O número total de acidentes em 2016 e 2017 é igual, embora com variação entre as diversas autoestradas. Neste estudo o número total de acidentes é 447 para o período considerado.

Considerando a extensão em quilómetros para cada autoestrada no estudo

e o número total de acidentes das mesmas foi calculado o número médio de acidentes/km para cada uma delas. Assim obteve-se em média, 12 acidentes/km na A41, 20 acidentes/km na A4 e 33 acidentes/km na A44. Observando estes valores pode-se concluir que a autoestrada em estudo com maior incidência de acidentes é a A44 apesar de esta ser a de menor extensão. Em contrapartida, a A41 que é a mais extensa, apresenta menor sinistralidade média por km.

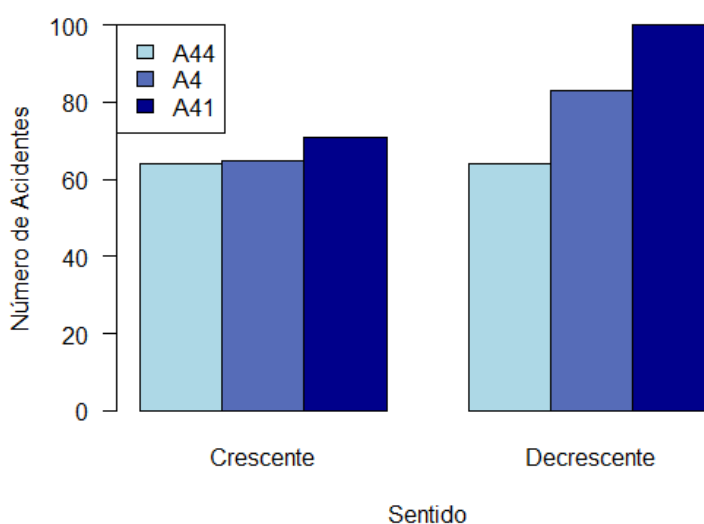


Figura 3.6: Número de Acidentes por Autoestrada e Sentido

Tabela 3.3: Número de Acidentes por Autoestrada e Sentido

AUTOESTRADA	CRESCENTE	DECRESCENTE	Nº TOTAL
A4	65	83	148
A44	64	64	128
A41	71	100	171
Nº TOTAL	200	247	447

Quanto ao estudo do número de acidentes por sentido, foi observado que o sentido com maior sinistralidade é o decrescente (Figura 3.6). Separando por autoestrada, apenas a A44 tem um valor igual de acidentes nos dois sentidos, as restantes têm mais acidentes no sentido decrescente (Tabela 3.3).

Tabela 3.4: Análise Descritiva das Variáveis correspondentes às Características Geométricas.

AUTOESTRADA	VARIÁVEL	NÚMERO DE OBSERVAÇÕES	MINÍMO	MÁXIMO	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
A4	<i>EmAR</i>	120	0	500	187,3	158,542	0,846
	<i>EmCL</i>	120	0	300	140	103,062	0,736
	<i>EmCC</i>	120	0	400	172	119,281	0,693
	<i>MIR</i>	120	0	0,002	0,001	0,001	0,775
	<i>MPA</i>	120	0	560	286,4	148,443	0,518
	<i>EmT</i>	120	0	500	264	180,338	0,683
	<i>EmCCX</i>	120	0	500	161,3	194,819	1,208
	<i>EmCCV</i>	120	0	390	74,67	117,559	1,574
	<i>MIP</i>	120	0	0,050	0,010	0,0144	1,414
	<i>MIN</i>	120	-0,045	0	-0,011	0,013	-1,208
	<i>MRCP</i>	120	0	27000	9513	10114,76	1,063
<i>MRCN</i>	120	-7000	0	-2233	2909,639	-1,303	
A44	<i>EmAR</i>	64	0	320	132,5	113,781	0,859
	<i>EmCL</i>	64	0	350	107,5	120,712	1,123
	<i>EmCC</i>	64	90	480	251,2	132,898	0,529
	<i>MIR</i>	64	0,00007	0,002	0,001	0,001	0,823
	<i>MPA</i>	64	0	351	183,8	117,337	0,638
	<i>EmT</i>	64	80	500	302,5	146,385	0,484
	<i>EmCCX</i>	64	0	230	110	92,513	0,841
	<i>EmCCV</i>	64	0	240	78,75	98,810	1,255
	<i>MIP</i>	64	0	0,037	0,023	0,015	0,671
	<i>MIN</i>	64	-0,046	0	-0,023	0,018	-0,804
	<i>MRCP</i>	64	0	50000	9276	15842,39	1,708
<i>MRCN</i>	64	-6000	0	-2270	2414,656	-1,064	
A41	<i>EmAR</i>	232	0	500	190	141,189	0,743
	<i>EmCL</i>	232	0	280	156,9	75,138	0,479
	<i>EmCC</i>	232	0	490	156,9	115,323	0,735
	<i>MIR</i>	232	0	0,002	0,001	0,001	0,633
	<i>MPA</i>	232	0	970	331,8	244,362	0,737
	<i>EmT</i>	232	10	500	192,1	113,546	0,591
	<i>EmCCX</i>	232	0	490	154,1	149,840	0,972
	<i>EmCCV</i>	232	0	410	154,1	133,615	0,867
	<i>MIP</i>	232	0	0,050	0,012	0,0180	1,518
	<i>MIN</i>	232	-0,046	0	-0,012	0,016	-1,307
	<i>MRCP</i>	232	0	260000	29426,37	55278,76	1,819
<i>MRCN</i>	232	-832000	0	-103786	210011,9	-2,024	

Na Tabela 3.4 apresenta-se uma análise descritiva das variáveis correspondentes

às características geométricas. Observa-se que:

- Para a variável $EmCC$, a A44 é a autoestrada com maior valor médio de extensão em curva circular, e ainda menor valor de $EmCL$. Pelo facto de uma clotóide apenas ser utilizada em curvas que apresentem raios menores, logo, curvas mais "apertadas", infere-se que, a A44 é uma autoestrada que apresentará curvas com raios maiores e por consequência menos perigosas.
- A variável MPA devolve o valor médio do parâmetro A . Assim, visto que apenas existe parâmetro A quando existe clotóide, o valor desta variável irá depender da variável $EmCL$, de forma que quanto maior for o valor de A^2 maior será a extensão média da clotóide ou maior o raio da curva subjacente. Das três autoestradas em estudo, a que tem maior extensão média de clotóide é a A41, sendo também a autoestrada que apresenta o maior valor médio do parâmetro A , logo maior qualidade de percepção da estrada para o condutor.
- Com exceção da A41, que contém exatamente o mesmo valor em média de extensão de curvas convexas e concâvas, a A4 e A44 têm mais curvas convexas (lombas) do que curvas concâvas (depressões).
- Considerando os valores em módulo das variáveis correspondentes à média do raio da curva tanto positivo como negativo, para as autoestradas A44 e A4, estes tomam valores superiores para o raio da curva positivo que corresponde às curvas concâvas. Contudo acontece o contrário para a A41.
- Pela análise do valor médio das variáveis MIP e MIN pode concluir-se que nas três autoestradas existe praticamente a mesma inclinação tanto positiva como negativa para cada uma delas.

Visto que os parâmetros de estado e do tráfego irão variar consoante a autoestrada, sentido e ano, será realizada uma análise das mesmas separadamente por autoestrada.

Tabela 3.5: Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A44)

ANO	VARIÁVEL	SENTIDO	NÚMERO		MÍNIMO	MÁXIMO	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIACÃO
			DE	OBSERVAÇÕES					
2014/ 2015	<i>IRImd</i>	C	16	16	1,040	2,055	1,462	0,297	0,203
		D	16	16	0,830	1,745	1,375	0,309	0,224
	<i>IRImx</i>	C	16	16	1,130	2,350	1,570	0,354	0,225
		D	16	16	0,900	1,960	1,462	0,329	0,225
	<i>RDmd</i>	C	16	16	1,255	2,380	1,654	0,405	0,245
		D	16	16	0,435	2,130	1,448	0,574	0,397
	<i>RDmx</i>	C	16	16	1,420	2,710	1,960	0,439	0,224
		D	16	16	0,640	3,140	1,905	0,693	0,364
	<i>MPDmd</i>	C	16	16	1,140	1,805	1,562	0,214	0,138
		D	16	16	1,245	1,770	1,529	0,176	0,115
	<i>MPDmn</i>	C	16	16	1,050	1,730	1,490	0,215	0,144
		D	16	16	1,120	1,700	1,442	0,177	0,123
	<i>GRIPmd</i>	C	16	16	0,370	0,475	0,414	0,039	0,093
		D	16	16	0,365	0,455	0,411	0,032	0,078
<i>GRIPmn</i>	C	16	16	0,270	0,460	0,376	0,071	0,186	
	D	16	16	0,300	0,440	0,375	0,057	0,151	
2016/ 2017	<i>IRImd</i>	C	16	16	1,085	1,735	1,506	0,206	0,137
		D	16	16	0,945	1,895	1,464	0,301	0,206
	<i>IRImx</i>	C	16	16	1,260	1,880	1,625	0,206	0,127
		D	16	16	1,080	2,020	1,570	0,293	0,187
	<i>RDmd</i>	C	16	16	0,885	1,755	1,304	0,265	0,203
		D	16	16	0,545	1,750	1,129	0,416	0,368
	<i>RDmx</i>	C	16	16	1,110	1,930	1,561	0,291	0,186
		D	16	16	0,710	2,680	1,350	0,641	0,474
	<i>MPDmd</i>	C	16	16	1,540	2,075	1,809	0,199	0,111
		D	16	16	1,370	2,110	1,799	0,244	0,135
	<i>MPDmn</i>	C	16	16	1,530	2,110	1,799	0,176	0,103
		D	16	16	1,340	2,010	1,722	0,203	0,118
	<i>GRIPmd</i>	C	16	16	0,460	0,525	0,494	0,021	0,041
		D	16	16	0,465	0,530	0,498	0,025	0,062
<i>GRIPmn</i>	C	16	16	0,450	0,500	0,475	0,019	0,041	
	D	16	16	0,460	0,510	0,483	0,017	0,035	

Pela observação da Tabela 3.5 pode concluir-se que há um aumento nos valores médios das variáveis ao longo do tempo. Tanto em 2014/2015 como em 2016/2017 o sentido com valores médios maiores para os indicadores de *IRI* e *RD* é o sentido crescente, o que se traduz num piso mais irregular e degradado comparativamente ao sentido decrescente. Comparando os valores médios das variáveis referentes ao atrito (*GRIP*), observa-se um aumento dos mesmos com o passar dos anos, onde as

condições de atrito são melhores no sentido crescente em 2014/2015 e no sentido decrescente em 2016/2017.

Na Figura 3.7 não se observam grandes diferenças entre valores relativos às variáveis $IRImd$ e $IRImx$ no sentido e ano. Já para as variáveis relacionadas as rodeiras (Figura 3.8) o gráfico sugere que os valores para estas variáveis são maiores para os anos de 2014/2015 e no sentido crescente. No caso da Figura 3.9, observa-se valores das variáveis $MPDmd$ e $MPDmn$ mais altos nos anos de 2016/2017. Para a variável relacionada com o atrito, $GRIP$, observa-se uma maior variabilidade desta nos anos 2014/2015 em relação a 2016/2017. Observa-se ainda que os valores em 2016/2017 são superiores aos valores de 2014/2015.

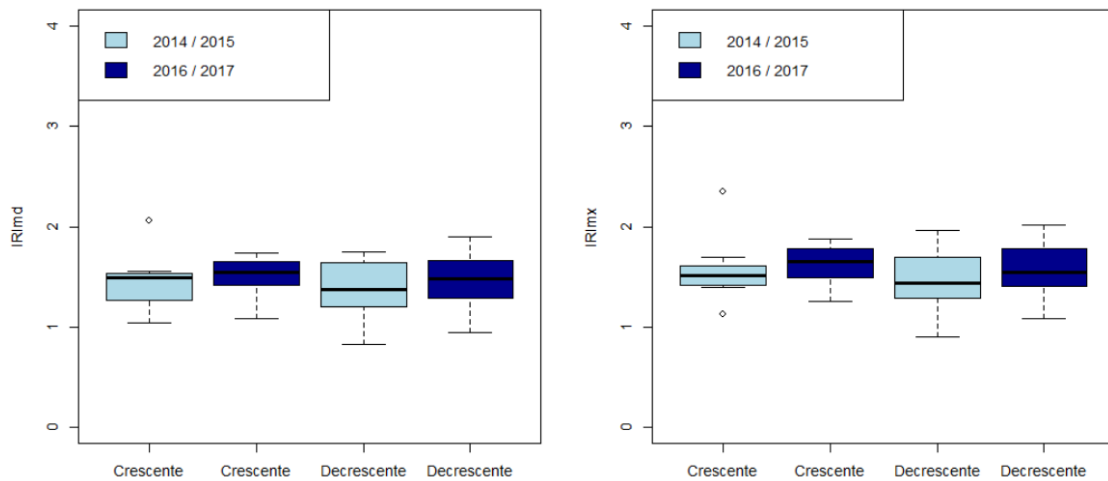


Figura 3.7: *Box plot* da Variável $IRImd$ e $IRImx$ (A44)

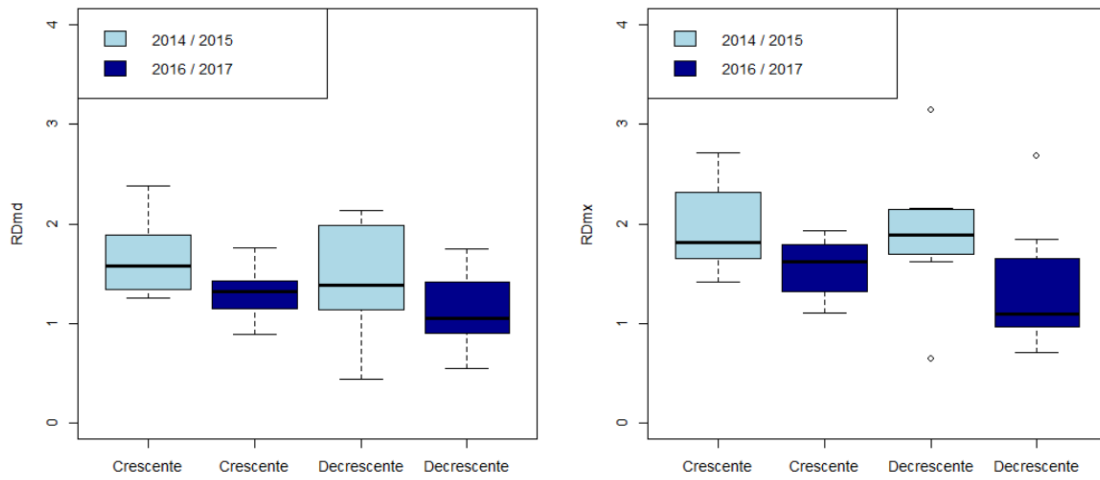


Figura 3.8: *Box plot* da Variável RD_{md} e RD_{mx} (A44)

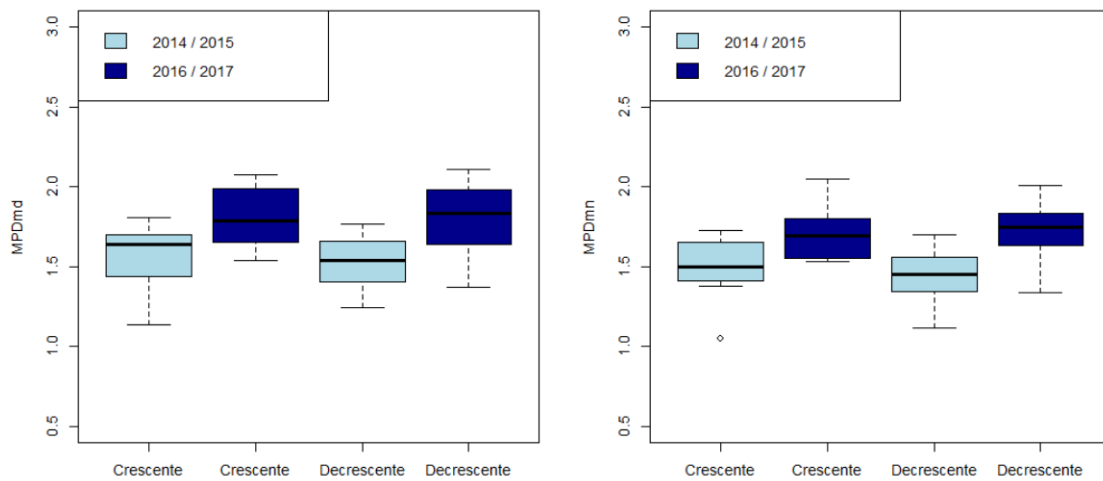


Figura 3.9: *Box plot* da Variável MPD_{md} e MPD_{mn} (A44)

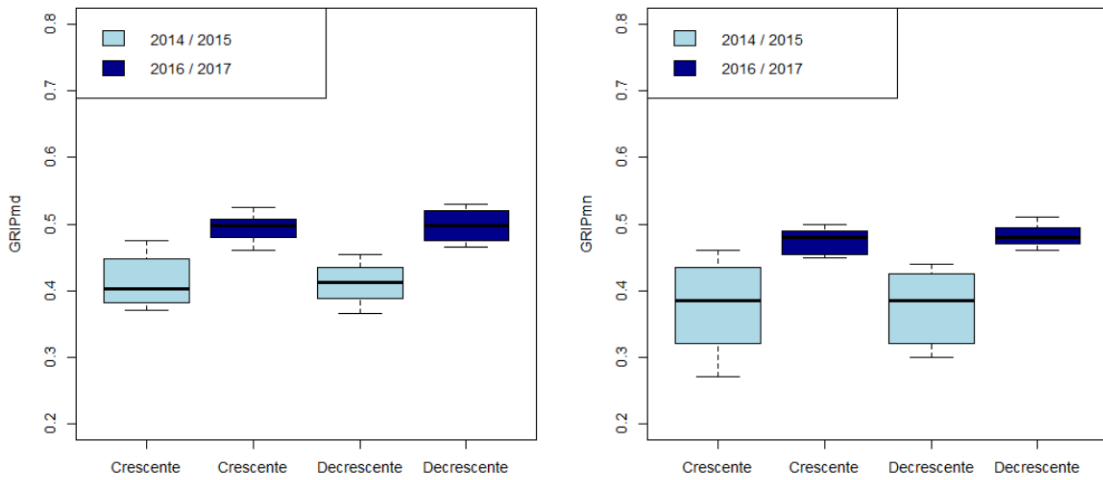


Figura 3.10: *Box plot* da Variável *GRIPmd* e *GRIPmn* (A44)

A Figura 3.11 descreve o número de acidentes em cada troço na autoestrada A44 no decorrer dos anos. Observa-se que o número de acidentes em cada troço ao longo dos anos é praticamente constante.

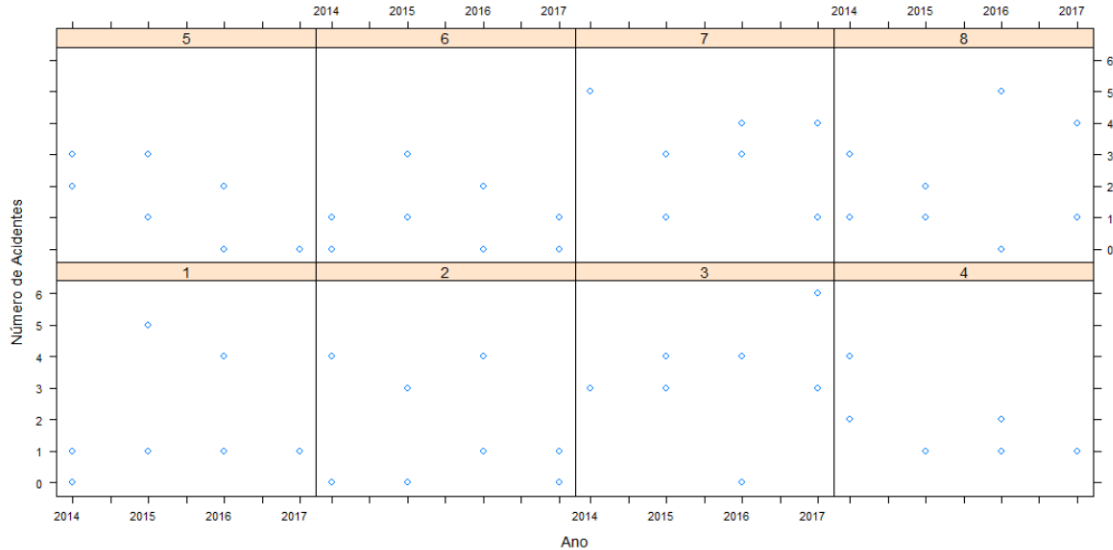


Figura 3.11: Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A44)

Pela Tabela 3.6, o tráfego médio diário anual (*TMDA*) na autoestrada A44 registou um decréscimo de 3% $\left(\frac{(34252+31608)-(30391+33461)}{34252+31608}\right)$ no valor médio do número total de veículos nos dois sentidos. Por sentido, o número médio de veículos em 2014/2015 é superior no sentido crescente, mas ao longo dos anos torna-se mais utilizado o sentido decrescente. Quanto aos veículos ligeiros, observa-se um aumento no sentido decrescente e um decréscimo no sentido crescente com o tempo. Em

Tabela 3.6: Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário Anual (A44)

ANO	VARIÁVEL	SENTIDO	NÚMERO DE OBSERVAÇÕES	MINÍMO	MÁXIMO	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
2014/ 2015	<i>Ligeiros</i>	C	16	20839	42360	32390	7291,401	0,225
		D	16	19974	37820	29888	6262,329	0,209
	<i>Pesados</i>	C	16	1765	1921	1862	49,545	0,027
		D	16	1610	1811	1721	65,009	0,038
	<i>Total</i>	C	16	22604	44281	34252	7333,9	0,214
		D	16	21584	39631	31608	6325,75	0,201
2016/ 2017	<i>Ligeiros</i>	C	16	20600	33731	28637	4552,561	0,159
		D	16	19742	43758	31803	7838,721	0,246
	<i>Pesados</i>	C	16	1439	2192	1755	281,669	0,161
		D	16	1585	1732	1658	58,947	0,036
	<i>Total</i>	C	16	22334	35520	30391	4681,305	0,154
		D	16	21328	45348	33461	7825,336	0,234

contrapartida, tanto em 2014/2015 como em 2016/2017 existe um número médio maior de veículos pesados no sentido crescente.

Ao estabelecer uma relação entre os valores das Tabela 3.5 e Tabela 3.6 observa-se que o sentido crescente é o piso que se encontra mais irregular e degradado (com valores para os indicadores do estado do pavimento *IRI* e *RD* maiores), o que se pode dever ao facto do tráfego total ser maior neste sentido em 2014/2015. Calculando o número total de veículos para cada sentido nos dois tempos em estudo obtém-se para o sentido crescente 64643 ($34252 + 30391$) de veículos e para o sentido decrescente 65069 ($31608 + 33461$) veículos, o que é praticamente o mesmo valor. Podendo ser uma das razões pelo qual existem o mesmo número de acidentes nos dois sentidos (Figura 3.6).

Tabela 3.7: Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A4)

ANO	VARIÁVEL	SENTIDO	NÚMERO		MÍNIMO	MÁXIMO	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
			DE OBSERVAÇÕES						
2014/ 2015	<i>IRImd</i>	C	30	1,067	1,833	1,329	0,0,273	0,206	
		D	30	0,633	2,000	1,258	0,339	0,269	
	<i>IRImx</i>	C	30	1,200	2,400	1,553	0,346	0,223	
		D	30	1,000	2,300	1,493	0,359	0,241	
	<i>RDmd</i>	C	30	1,100	1,600	1,298	0,158	0,121	
		D	30	0,633	1,867	1,207	0,317	0,263	
	<i>RDmx</i>	C	30	1,200	2,700	1,907	0,353	0,185	
		D	30	1,200	3,300	1,840	0,542	0,295	
	<i>MPDmd</i>	C	30	1,267	1,667	1,544	0,098	0,064	
		D	30	1,000	1,700	1,471	0,201	0,137	
	<i>MPDmn</i>	C	30	1,100	1,600	1,480	0,132	0,089	
		D	30	0,100	1,700	1,273	0,531	0,417	
	<i>GRIPmd</i>	C	30	0,407	0,513	0,448	0,029	0,064	
		D	30	0,280	0,486	0,418	0,047	0,113	
	<i>GRIPmn</i>	C	30	0,380	0,470	0,414	0,026	0,025	
		D	30	0,100	0,450	0,352	0,113	0,321	
2016/ 2017	<i>IRImd</i>	C	30	1,033	1,933	1,396	0,288	0,206	
		D	30	1,000	2,033	1,358	0,311	0,229	
	<i>IRImx</i>	C	30	1,200	2,200	1,593	0,333	0,209	
		D	30	1,100	2,300	1,627	0,333	0,205	
	<i>RDmd</i>	C	30	1,000	1,667	1,236	0,184	0,149	
		D	30	0,633	1,500	1,193	0,236	0,198	
	<i>RDmx</i>	C	30	1,100	2,900	1,807	0,496	0,275	
		D	30	1,200	2,500	1,793	0,445	0,248	
	<i>MPDmd</i>	C	30	1,333	1,767	1,587	0,111	0,071	
		D	30	1,000	1,833	1,547	0,193	0,125	
	<i>MPDmn</i>	C	30	1,200	1,700	1,520	0,115	0,075	
		D	30	0,100	1,800	1,440	0,422	0,293	
	<i>GRIPmd</i>	C	30	0,417	0,523	0,474	0,028	0,058	
		D	30	0,293	0,490	0,427	0,057	0,136	
	<i>GRIPmn</i>	C	30	0,390	0,480	0,440	0,027	0,061	
		D	30	0,100	0,440	0,349	0,148	0,424	

Na Tabela 3.7 é apresentada uma análise exploratória das características superficiais relativas à A4. Através da observação dos valores médios das variáveis presentes nesta tabela conclui-se que o pavimento encontra-se em pior estado no sentido crescente, pois os valores das variáveis relativas aos indicadores do estado do pavimento *IRI* e *RD* são maiores neste sentido, talvez por este ser o sentido com maior tráfego (Tabela 3.8). Existe também um agravamento do estado do pavimento com o tempo.

Pelas das Figuras 3.13 e 3.14 observa-se que não existem grandes diferenças nos valores referentes às rodeiras e à profundidade média de perfil entre sentido e ano. Pela Figura 3.12, observa-se uma maior variabilidade nos valores da variável $RDmx$ comparativamente à variável $RDmd$. Pela Figura 3.12 observa-se que os valores das variáveis $IRImd$ e $IRImx$ são mais elevados para os anos de 2016/2017. Por fim, na análise da Figura 3.15 os valores para o atrito são superiores para o ano 2016/2017.

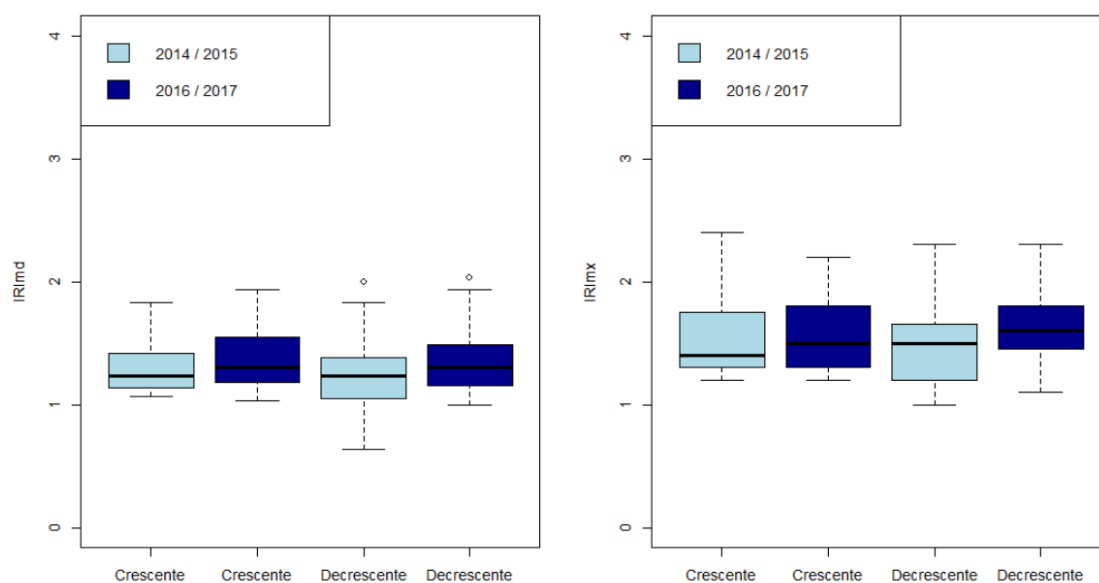


Figura 3.12: *Box plot* da Variável $IRImd$ e $IRImx$ (A4)

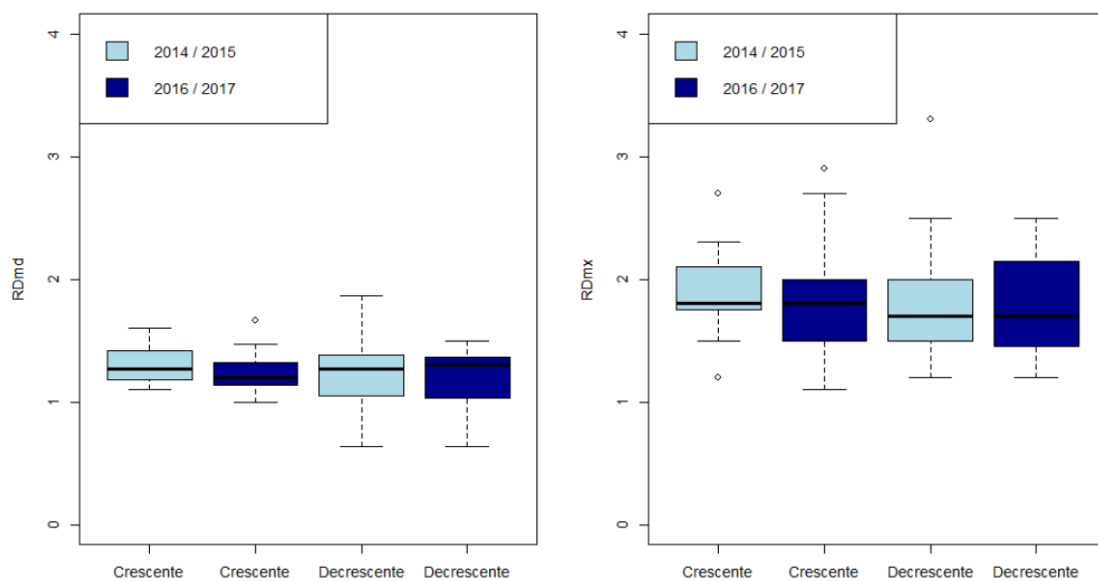


Figura 3.13: *Box plot* da Variável RD_{md} e RD_{mx} (A4)

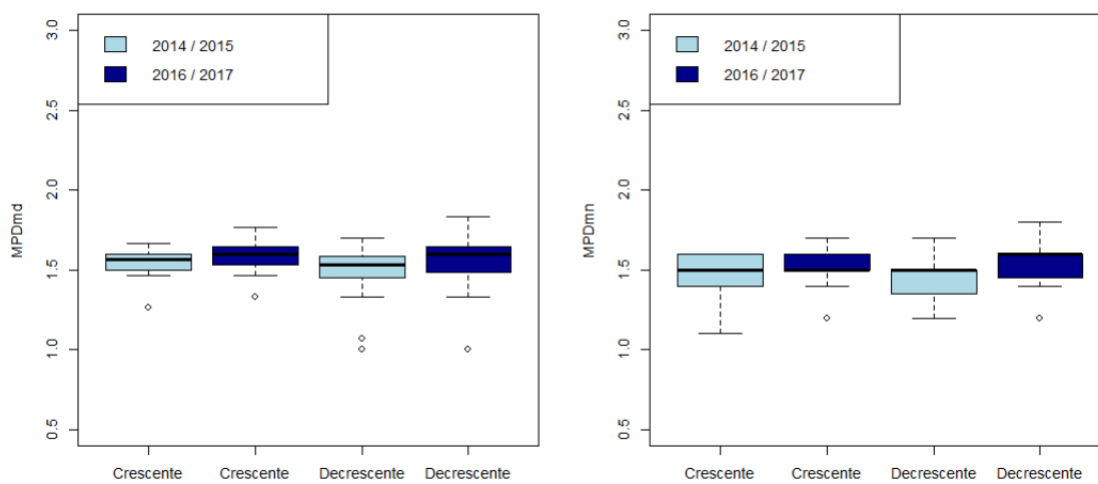


Figura 3.14: *Box plot* da Variável MPD_{md} e MPD_{mn} (A4)

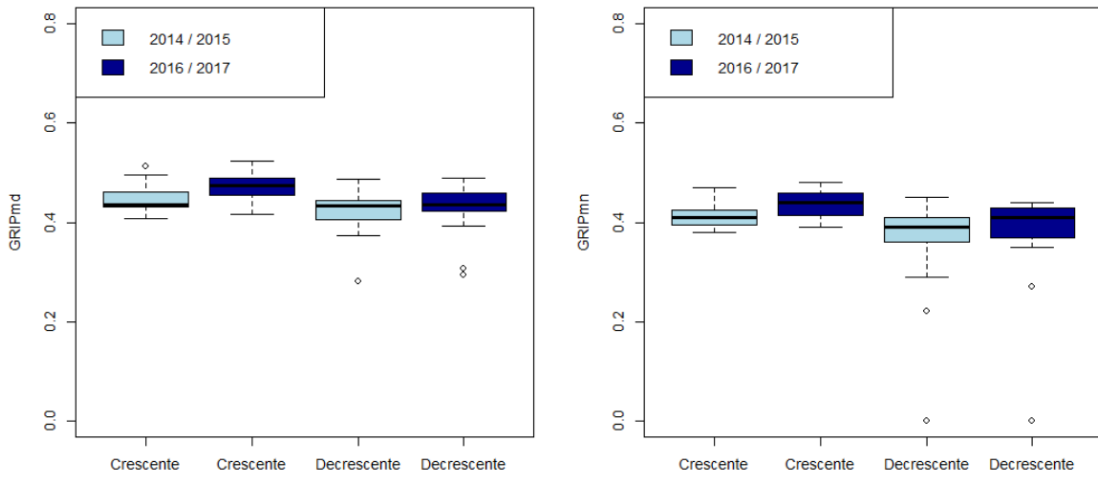


Figura 3.15: *Box plot* da Variável *GRIPmd* e *GRIPmn* (A4)

Observando o comportamento do número de acidentes por troço em relação ao anos, pode-se concluir que não existe uma grande variação entre 2014 e 2017 (Figura 3.16).

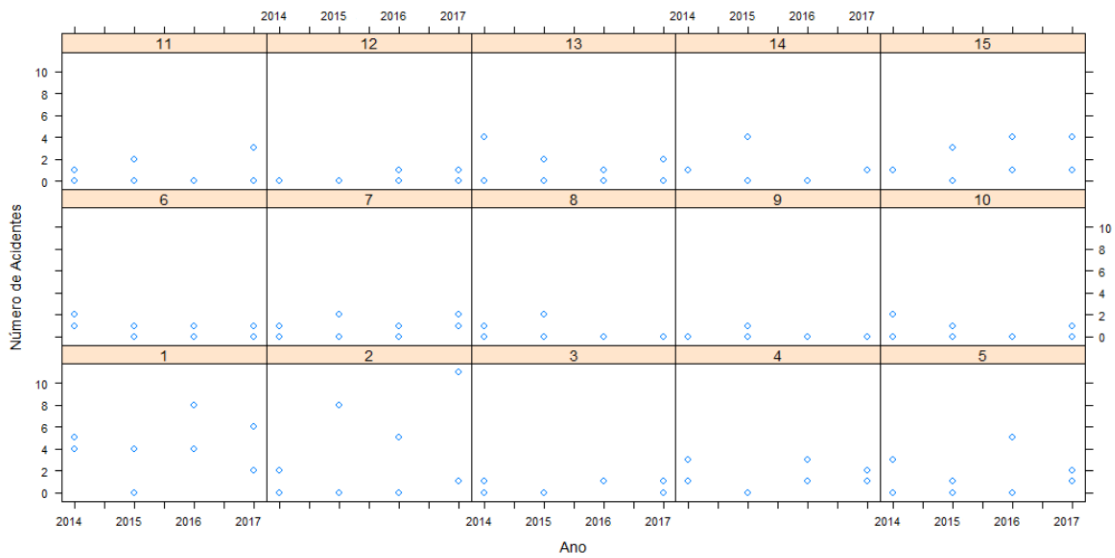


Figura 3.16: Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A4)

Na Tabela 3.8 apresenta-se uma análise exploratória das variáveis correspondentes ao tráfego médio diário anual. Observa-se que há um crescimento de aproximadamente 40% $\left(\frac{(21294+20844)-(28108+30881)}{21294+20844}\right)$ no valor total do tráfego médio diário anual de 2014/2015 para 2016/2017, verificando-se um aumento do tráfego em 30% $\left(\frac{(1000+938)-(1152+1346)}{1000+938}\right)$ de veículos pesados e de 40,5% $\left(\frac{(20294+19906)-(26956+29534)}{20294+19906}\right)$

nos veículos ligeiros. Nos anos de 2014/2015, observa-se um tráfego maior no sentido crescente, sendo que em 2016/2017 este é maior no sentido decrescente.

Tabela 3.8: Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário (A4)

ANO	VARIÁVEL	SENTIDO	NÚMERO			MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIACÃO
			DE OBSERVAÇÕES	MINÍMO	MÁXIMO			
2014/ 2015	<i>Ligeiros</i>	C	30	17092	22726	20294	1927,041	0,095
		D	30	16867	23574	19906	2096,850	0,105
	<i>Pesados</i>	C	30	616	1523	1000	388,168	0,388
		D	30	505	1479	937,9	426,830	0,455
	<i>Total</i>	C	30	17708	24249	21294	2298,899	0,109
		D	30	17427	25048	20844	2507,4	0,120
2016/ 2017	<i>Ligeiros</i>	C	30	23295	29655	26956	2099,6	0,078
		D	30	23395	33595	29534	3441,551	0,112
	<i>Pesados</i>	C	30	917	1384	1152	144,581	0,125
		D	30	1033	1541	1346	183,971	0,137
	<i>Total</i>	C	30	24212	3075	28108	2193,376	0,078
		D	30	24428	35109	30881	3553,676	0,115

Analisando os resultados obtidos relativamente ao número de acidentes por sentido (Figura 3.6) com o tráfego (Tabela 3.8) constata-se que apesar do sentido decrescente ser o que apresenta maior número de acidentes é o sentido crescente que tem o maior tráfego médio diário anual. Um dos motivos pelo qual o sentido decrescente ser o sentido com maior número de acidentes pode ser o facto deste ser o que apresenta menores valores para os indicadores de profundidade média de perfil (*MPD*) e de atrito (*GRIP*) o que sugere um piso com piores condições de drenabilidade e de atrito, não fornecendo segurança na imobilização de um veículo, podendo causar acidentes.

Tabela 3.9: Análise Descritiva das Variáveis correspondentes ao Estado do Pavimento (A41)

ANO	VARIÁVEL	SENTIDO	NÚMERO		MÍNIMO	MÁXIMO	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
			DE OBSERVAÇÕES						
2014/ 2015	<i>IRImd</i>	C	58	0,933	1,467	1,191	0,134	0,113	
		D	58	0,933	1,567	1,221	0,193	0,158	
	<i>IRImx</i>	C	58	1,000	1,600	1,303	0,166	0,127	
		D	58	1,000	1,800	1,362	0,229	0,168	
	<i>RDmd</i>	C	58	0,800	1,367	1,061	0,157	0,148	
		D	58	0,767	1,533	1,075	0,166	0,154	
	<i>RDmx</i>	C	58	1,200	2,600	1,859	0,376	0,202	
		D	58	1,100	3,000	1,841	0,407	0,221	
	<i>MPDmd</i>	C	58	0,833	1,567	1,355	0,162	0,119	
		D	58	0,933	1,733	1,359	0,178	0,131	
	<i>MPDmn</i>	C	58	0,300	1,500	1,224	0,327	0,267	
		D	58	0,100	1,700	1,217	0,385	0,316	
	<i>GRIPmd</i>	C	58	0,27	0,537	0,399	0,056	0,141	
		D	58	0,310	0,477	0,376	0,044	0,118	
	<i>GRIPmn</i>	C	58	0,100	0,51	0,361	0,086	0,237	
		D	58	0,280	0,460	0,349	0,041	0,118	
2016/ 2017	<i>IRImd</i>	C	58	1,000	1,633	1,256	0,155	0,123	
		D	58	0,967	1,7	1,279	0,212	0,166	
	<i>IRImx</i>	C	58	1,100	1,700	1,369	0,175	0,128	
		D	58	1,000	1,900	1,441	0,224	0,156	
	<i>RDmd</i>	C	58	0,767	2,567	1,381	0,399	0,289	
		D	58	0,767	1,833	1,271	0,242	0,191	
	<i>RDmx</i>	C	58	1,200	3,500	2,193	0,535	0,244	
		D	58	1,100	3,700	2,169	0,531	0,245	
	<i>MPDmd</i>	C	58	1,400	1,600	1,510	0,071	0,046	
		D	58	1,300	1,733	1,489	0,101	0,068	
	<i>MPDmn</i>	C	58	1,300	1,600	1,466	0,077	0,052	
		D	58	1,100	1,700	1,431	0,126	0,088	
	<i>GRIPmd</i>	C	58	0,463	0,567	0,514	0,025	0,048	
		D	58	0,450	0,547	0,504	0,025	0,067	
	<i>GRIPmn</i>	C	58	0,440	0,530	0,485	0,024	0,051	
		D	58	0,420	0,530	0,476	0,028	0,059	

Através da observação dos valores médios das variáveis correspondentes aos parâmetros de estado do pavimento da A41 (Tabela 3.9), apesar de existir um aumento dos valores com o passar dos anos, este não é assim tão elevado, sendo que os valores pouco diferem consoante o sentido. Contudo, os valores indicam que o piso que se encontra em melhor estado é o piso no sentido crescente, com um piso menos irregular (valores para *IRI* menores), menos degradado (valores de *RD* menores) e com

melhores condições de drenabilidade e de atrito (valores de *MPD* e *GRIP* maiores). O que pode ser uma das causas pelo qual o sentido crescente é o que apresenta o menor número de acidentes (Tabela 3.3).

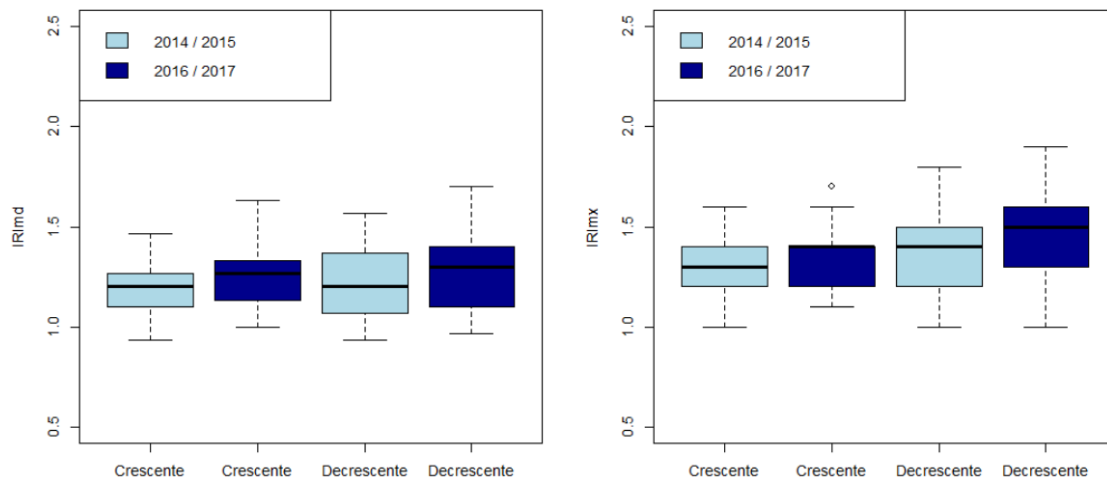


Figura 3.17: *Box plot* da Variável *IRImd* e *IRImx* (A41)

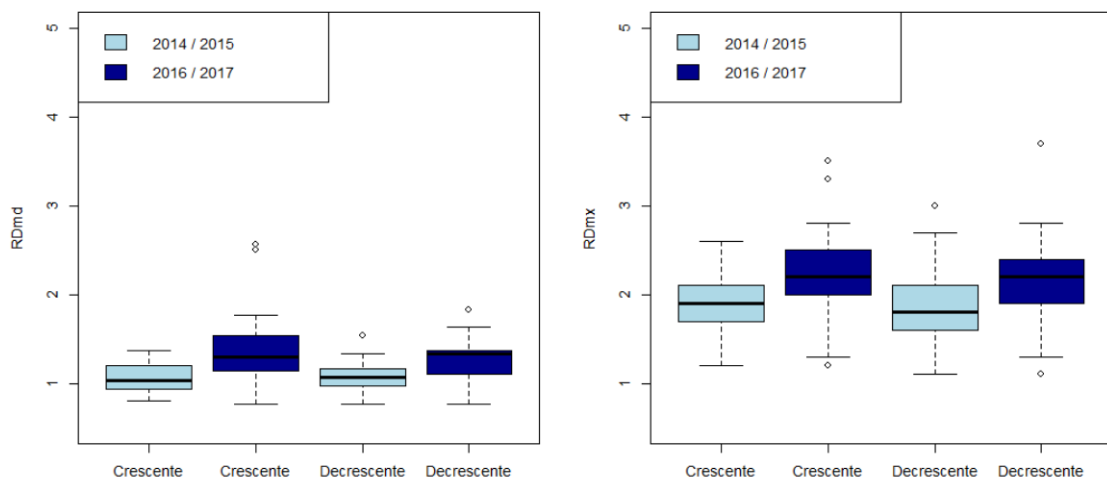


Figura 3.18: *Box plot* da Variável *RDmd* e *RDmx* (A41)

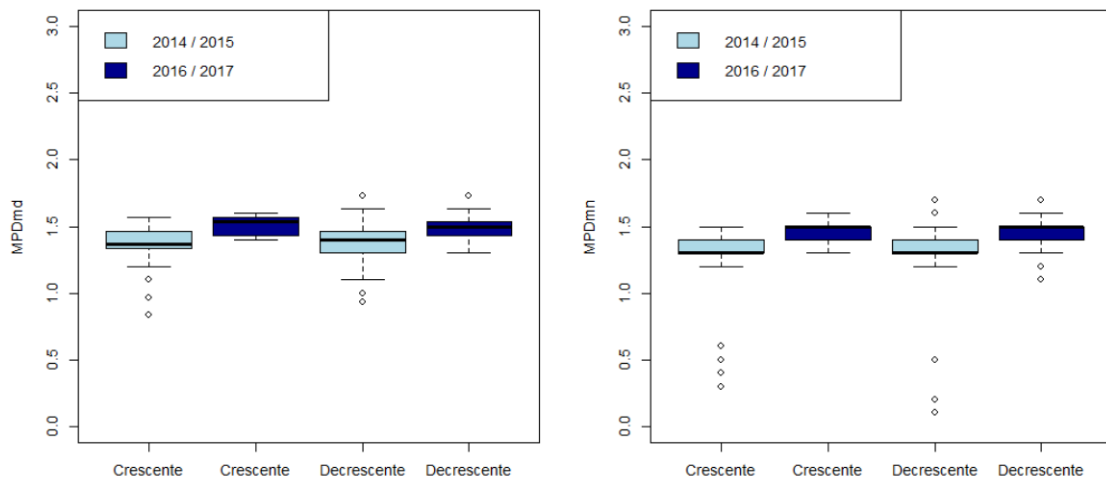


Figura 3.19: *Box plot* da Variável MPD_{md} e MPD_{mn} (A41)

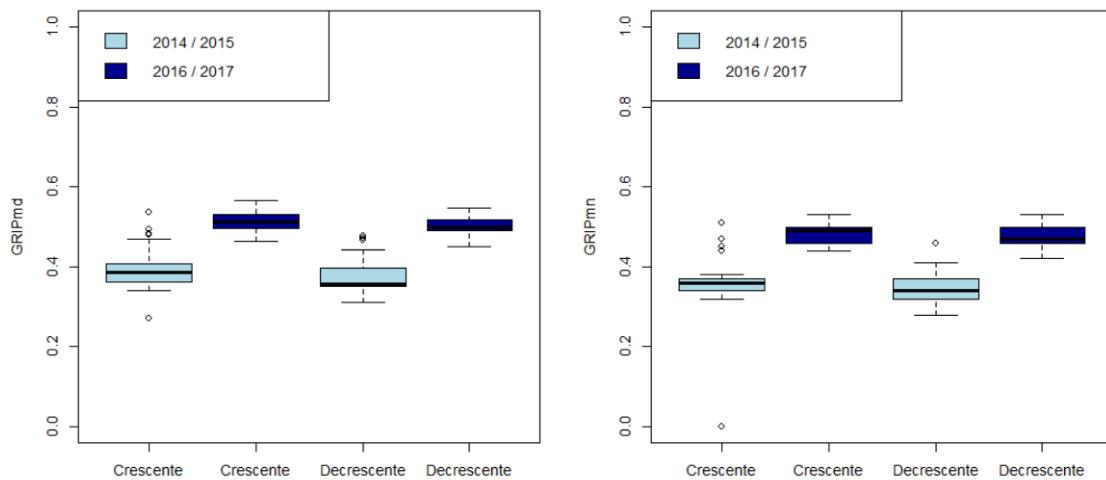


Figura 3.20: *Box plot* da Variável $GRIP_{md}$ e $GRIP_{mn}$ (A41)

Através da Figura 3.17 é observado que a variabilidade dos valores relativos ao índice internacional de rugosidade é superior no sentido decrescente. Pelas Figuras 3.18 e 3.19 verifica-se que os valores relativos às variáveis RD_{md} , RD_{mx} , MPD_{md} e MPD_{mn} nos anos 2016/2017 são superiores aos valores de 2014/2015. A Figura 3.20 sugere que os valores do atrito de 2016/2017 são muito superiores aos valores de 2014/2015.

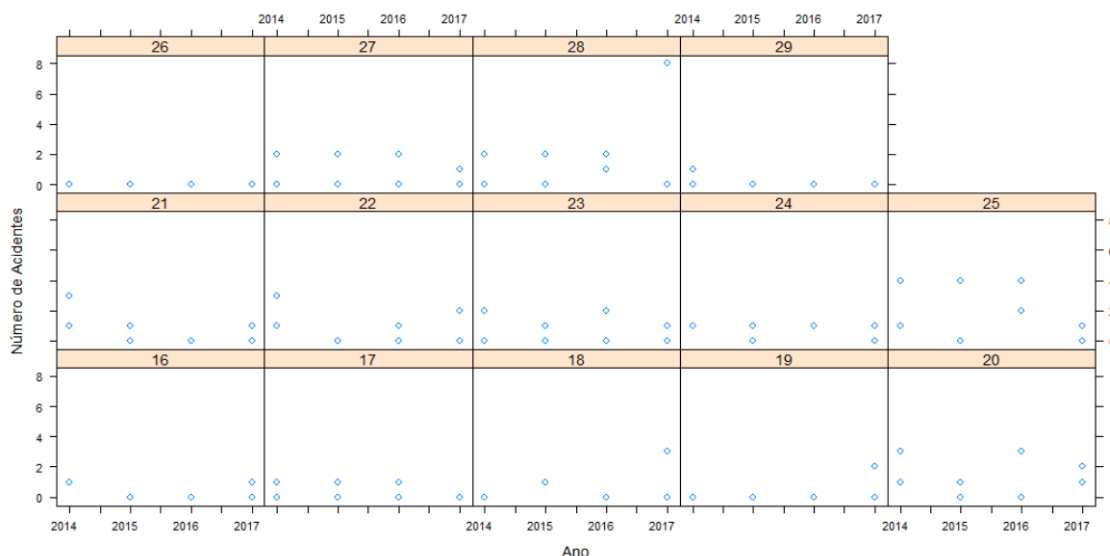


Figura 3.21: Gráfico de Dispersão do Número de Acidentes por Ano e Troço (A41)

A Figura 3.21 relativa ao número de acidentes por ano e troço sugere que existe um aumento, mesmo que pouco acentuado, do número de acidentes, em alguns troços, com o passar do tempo.

Tabela 3.10: Análise Descritiva das Variáveis correspondentes ao Tráfego Médio Diário (A41)

ANO	VARIÁVEL	SENTIDO	NÚMERO				DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
			DE OBSERVAÇÕES	MINÍMO	MÁXIMO	MÉDIA		
2014/ 2015	<i>Ligeiros</i>	C	58	10745	15830	12839	1957,524	0,152
		D	58	9992	16209	13049	2087,145	0,159
	<i>Pesados</i>	C	58	544	1227	840	287,919	0,343
		D	58	578	1249	805	261,036	0,324
	<i>Total</i>	C	58	11493	16374	13659	1996,563	0,146
		D	58	10990	16788	13854	2119,499	0,153
2016/ 2017	<i>Ligeiros</i>	C	58	14272	20719	17847	1996,717	0,112
		D	58	13528	19659	16561	2466,441	0,149
	<i>Pesados</i>	C	58	605	1621	1181	326,529	0,277
		D	58	1060	1650	1268	207,654	0,164
	<i>Total</i>	C	58	15711	21986	19027	2185,009	0,115
		D	58	14866	21309	17829	2606,149	0,146

O TMDA total na A41 teve um aumento de 34% $\frac{(13659+13854)-(19027+17829)}{13659+13854}$ com o tempo, sendo que existe um aumento de veículos quer ligeiros quer pesados. No entanto, comparando este valor entre sentido, não existe uma grande diferença, o que poderá vir a justificar o facto de as variáveis relativas ao estado do pavimento (Tabela 3.9) não variarem muito consoante o sentido.

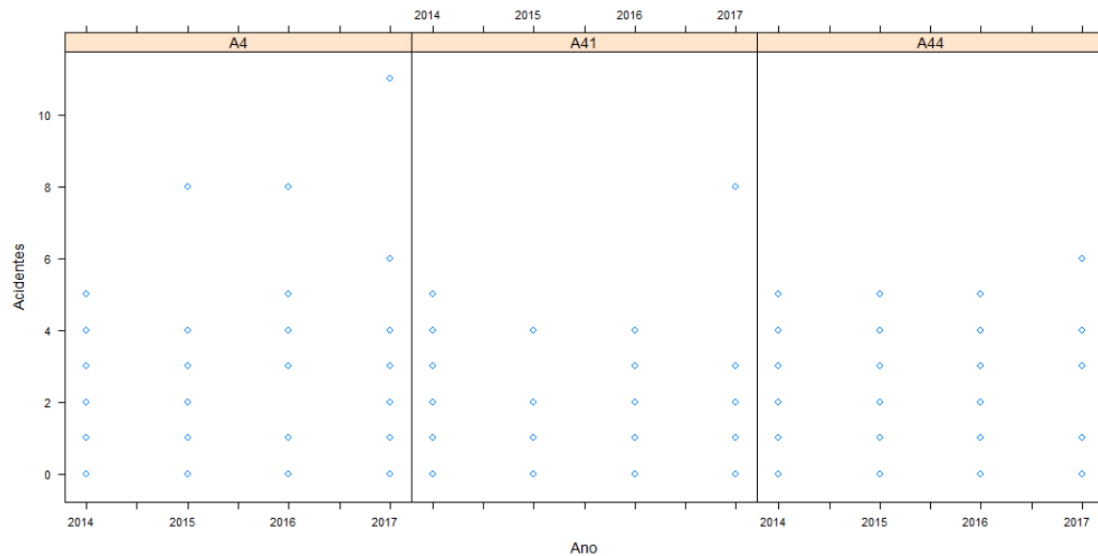


Figura 3.22: Gráfico de Dispersão do Número de Acidentes por Ano e Autoestrada

Na observação do gráfico de dispersão da Figura 3.22 que descreve o número de acidentes em cada autoestrada no período de quatro anos, pode-se observar que não existe um aumento relevante do número de acidentes ao longo dos anos.

3.3 Aplicação dos Modelos

A aplicação destes modelos no R foi executada através da função *glmer()* da biblioteca *lme4*. Esta função ajusta os dados a um modelo linear generalizado misto aplicando o método de estimação de máxima verosimilhança e utilizando o método numérico de Laplace. Em R existem outras funções que podem ser utilizadas no ajustamento a um modelo linear generalizado misto, como por exemplo a função *glmm()* da biblioteca *glmmML*, a função *glmmPQL()* da biblioteca *MASS* e a função *glmmadmb()* da biblioteca *glmmADMB*. A escolha da função utilizada recaiu sobre a função *glmer()* por ser a mais utilizada nos diversos estudos revistos.

Inicialmente foi necessária proceder-se a uma reescala de algumas variáveis de forma a facilitar a estimação do modelo. Assim, as variáveis iniciais resultam da multiplicação das variáveis novas com a sua respetiva reescala. (Tabela 3.11)

Tabela 3.11: Reescala de variáveis

VARIÁVEL INICIAL	Ligeiros	Pesados	Total	EmAR	EmCC	EmCL	MIR	MPA
VARIÁVEL NOVA	Ligeiros2	Pesados2	Total2	EmAR2	EmCC2	EmCL2	MIR2	MPA2
REESCALA	10 000	1 000	10 000	100	100	100	0,01	1 000
VARIÁVEL INICIAL	EmT	EmCCX	EmCCV	MIN	MIP	MRCN	MRCP	
VARIÁVEL NOVA	EmT2	EmCCX2	EmCCV2	MIN2	MIP2	MRCN2	MRCP2	
REESCALA	100	100	100	0,01	0,01	10 000	10 000	

Antes de se ajustar um modelo linear generalizado misto com resposta que segue uma distribuição de Poisson, dado que os dados são de contagem, é importante estudar quais os efeitos aleatórios que devem ser considerados. O efeito aleatório foi escolhido através do Critério de Informação de *Akaike*.

Desta forma, analisando os modelos

- Modelo Nulo com efeito aleatório Troço: AIC=1182,723
- Modelo Nulo com efeito aleatório Ano: AIC=1316,327
- Modelo Nulo com efeito aleatório Autoestrada: AIC=1258,487

O modelo nulo com menor valor AIC é o modelo com efeito aleatório Troço.

Na formação da base de dados foram construídas duas variáveis correspondente ao estado do pavimento, nomeadamente *IRImd* e *IRImx*, *RDmd* e *RDmx*, *MPDmd* e *MPDmn* e por fim *GRIPmd* e *GRIPmn*. De seguida, procede-se uma análise de qual a variável a usar no modelo.

Tabela 3.12: Resultados dos Modelos Simples das variáveis relativas ao estado do pavimento

VARIÁVEL	MODELO	AIC	VALOR P
<i>IRImd</i> vs <i>IRImx</i>	Regressão Simples: <i>IRImd</i>	1184,681	0,835
	Regressão Simples: <i>IRImx</i>	1181,329	0,0634
<i>RDmd</i> vs <i>RDmx</i>	Regressão Simples: <i>RDmd</i>	1184,694	0,862
	Regressão Simples: <i>RDmx</i>	1181,783	0,0834
<i>MPDmd</i> vs <i>MPDmn</i>	Regressão Simples: <i>MPDmd</i>	1184,502	0,634
	Regressão Simples: <i>MPDmn</i>	1183,127	0,194
<i>GRIPmd</i> vs <i>GRIPmn</i>	Regressão Simples: <i>GRIPmd</i>	1179,575	0,0210
	Regressão Simples: <i>GRIPmn</i>	1175,156	0,00137

Foram selecionadas as variáveis *IRImx*, *RDmx*, *MPDmn* e *GRIPmn* (modelos a negrito na Tabela 3.12), pois são os modelos com menor valor AIC. Observa-se ainda que essas variáveis correspondem ao Valor P mais pequeno no modelo simples.

Antes da formulação do modelo final, foram realizadas comparações entre modelos para a escolha de combinações de duas variáveis entre três variáveis correlacionadas. Nesta situação existem três grupos de variáveis que estão relacionadas entre si. Estes três grupos serão tráfego (*Ligeiros2*, *Pesados2* e *Total2*), características geométricas em planta (*EmAR2*, *EmCC2* e *EmCL2*) e as características geométricas em perfil (*EmT2*, *EmCCX2* e *EmCCV2*).

No caso do tráfego, onde $Ligeiros2 + Pesados2 = Total2$, foram construídos três modelos com combinações de duas variáveis diferentes, sendo:

- *Ligeiros2* e *Pesados2* com AIC=1147,301
- *Ligeiros2* e *Total2* com AIC=1153,264
- *Pesados2* e *Total2* com AIC=1147,448

apesar da diferença mínima entre AIC dos modelos, a combinação escolhida será de *Ligeiros2* e *Pesados2*.

Em relação às características geométricas em planta, onde $EmAR2 + EmCC2 + EmCL2 = 500$:

- $EmAR2$ e $EmCL2$ com $AIC=1150,967$
- $EmAR2$ e $EmCC2$ com $AIC=1150,938$
- $EmCC2$ e $EmCL2$ com $AIC=1150,942$

mais uma vez, a diferença entre os valores de AIC foi mínima entre AIC, mas as variáveis escolhidas são $EmAR2$ e $EmCC2$. Nas características geométricas em perfil, onde $EmT2 + EmCCX2 + EmCCV2 = 500$:

- $EmT2$ e $EmCCX2$ com $AIC=1153,594$
- $EmT2$ e $EmCCV2$ com $AIC=1153,654$
- $EmCCX2$ e $EmCCV2$ com $AIC=1153,743$

assim, as variáveis escolhidas são $EmT2$ e $EmCCX2$.

Após a escolha das variáveis, é obtido o modelo completo (Tabela 3.13).

Tabela 3.13: Estimativas do modelo completo de regressão de Poisson com efeitos aleatórios

Variável Resposta: Número de Acidentes			
VARIÁVEIS EXPLICATIVAS	PARÂMETRO ESTIMADO(β)	ERRO PADRÃO	VALOR P
<i>intercept</i>	0,153	1,003	0,878988
<i>IRImx</i>	-0,173	0,277	0,531431
<i>RDmx</i>	-0,092	0,113	0,416459
<i>MPDmn</i>	0,079	0,223	0,724176
<i>GRIPmn</i>	-1,475	0,763	0,053323
<i>No</i>	0,685	0,171	6,02e-05
<i>NVias3</i>	-0,232	0,316	0,463613
<i>Ano</i>	-0,033	0,054	0,549113
<i>SentidoD</i>	0,144	0,098	0,143508
<i>Ligeiros2</i>	0,022	0,009	0,014210
<i>Pesados2</i>	0,487	0,229	0,033285
<i>EmAR2</i>	-0,103	0,149	0,490603
<i>EmCC2</i>	-0,103	0,129	0,428468
<i>MIR2</i>	0,038	0,018	0,035677
<i>MPA2</i>	-0,078	0,066	0,237915
<i>EmT2</i>	-0,103	0,079	0,193365
<i>EmCCX2</i>	-0,046	0,070	0,514312
<i>MIP2</i>	0,078	0,054	0,150614
<i>MIN2</i>	-0,239	0,072	0,000893
<i>MRCP2</i>	-0,006	0,009	0,529843
<i>MRCN2</i>	0,004	0,007	0,537448
Medidas de Avaliação do Ajuste			
AIC		1155,5	
Função de log-Verosimilhança		-555,7	

Retirando-se sucessivamente as variáveis estatisticamente não significativas, obtém-se o modelo:

$$Y_i | \mathbf{b}_i \sim P(\mu_{it}) \tag{3.2}$$

$$\begin{aligned} \log(\mu_{it}) = & \beta_0 + \beta_1 \times GRIPmn_{it} + \beta_2 \times No_{it} + \beta_3 \times Ligeiros2_{it} + \\ & + \beta_4 \times Pesados2_{it} + \beta_5 \times MIR2_{it} + \beta_6 \times MIN2_{it} + b_{0i} \end{aligned} \quad (3.3)$$

em que Y_{it} é o número de acidentes no troço $i = 1, \dots, 416$ no ano $t = 1, 2, 3, 4$. Com os resultados das estimativas apresentadas na Tabela 3.14.

Tabela 3.14: Estimativas do modelo final

Variável Resposta: Número de Acidentes				
VARIÁVEIS EXPLICATIVAS	PARÂMETRO ESTIMADO (β)	EXP(β)	ERRO PADRÃO	VALOR P
<i>intercept</i>	-1,099	0,330	0,340	0,001227
<i>GRIPmd</i>	-1,947	0,143	0,541	0,000318
<i>No</i>	0,689	1,991	0,141	1,05e-06
<i>Ligeiros2</i>	0,014	1,014	0,006	0,029955
<i>Pesados2</i>	0,707	2,027	0,160	1,04e-05
<i>MIR2</i>	0,037	1,037	0,012	0,002305
<i>MIN2</i>	-0,189	0,827	0,051	0,00149
Medidas de Avaliação do Ajuste				
AIC				1140,0
Função de log-Verosimilhança				-562,0

Na interpretação das estimativas obtidas do modelo final (Tabela 3.14) há:

- Com o aumento de uma unidade de GN no atrito (*GRIPmn*), o número de acidentes diminui cerca de 85,7% para o mesmo troço e permanecendo as outras variáveis explicativas constantes. O que já era o esperado, visto que quanto maior for o valor do atrito (*GRIPmn*) melhor é o estado do piso.
- Com o aumento do número de nós de uma unidade, o número de acidentes aumenta cerca de 2 vezes. Este resultado também era esperado uma vez que aumento do número de nós significa o aumento de uma entrada ou saída da autoestrada, o que pode provocar mais acidentes.
- Relativamente às variáveis do tráfego, para o mesmo troço, e com as restantes variáveis constantes, o aumento de 10 000 veículos ligeiros significa um aumento de cerca de 1,4% no número de acidentes. Já o aumento de 1 000 veículos pesados implica um aumento de cerca de 2 vezes o número de acidentes.

- O aumento de uma unidade da variável $MIR2$ significa um aumento de 0,01 da variável MIR , e assim um aumento de 100 metros no valor médio do raio, pois $MIR = \frac{1}{MR}$. Este aumento no valor médio do raio faz aumentar o número de acidentes em cerca de 3,7%. O aumento do valor médio do raio pode significar uma maior extensão em alinhamento reto ($EmAR$), o que pode originar um aumento de confiança, monotonicidade e causar mais acidentes. (MIR - média do inverso do raio; MR - média do raio)
- No caso do aumento de uma unidade de $MIN2$, existe um aumento de 0,01 da variável MIN (média inclinação negativa) significando um decréscimo no número de acidentes de 17,3%. Este aumento da variável MIN indica uma diminuição no declive da estrada, originado um declive da estrada menor.

Diagnóstico do Modelo

Neste estudo será aplicado o diagnóstico proposto por Florian Hartig (Hartig, F. (2022)) usando a package '*DHARMA*' no R baseado nos resíduos quantílicos. Esta abordagem é baseada em simulações de novos conjuntos de dados baseados no modelo final estimado e nos parâmetros ajustados.

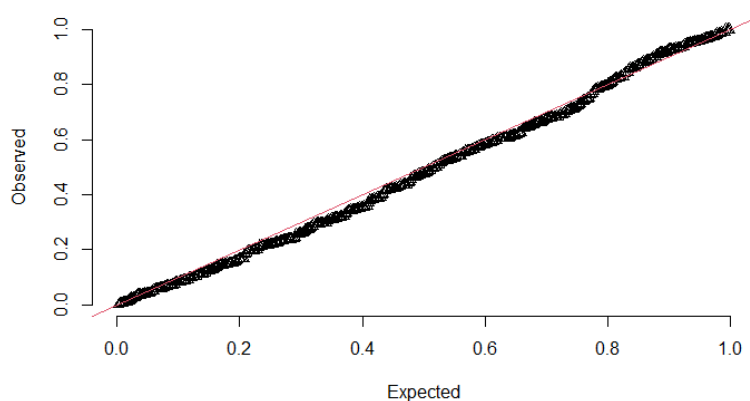


Figura 3.23: Gráfico QQ dos Resíduos Quantílicos

Na Figura 3.23 está representado o gráfico de Quantil-Quantil, que permite confirmar se os resíduos estão de acordo com o que é esperado pelo modelo. Neste gráfico, os pontos devem estar todos alinhados com a linha vermelha, caso contrário, existe problemas de especificação do modelo. No caso do modelo em estudo, não parece haver desvios relevantes dos valores esperados.

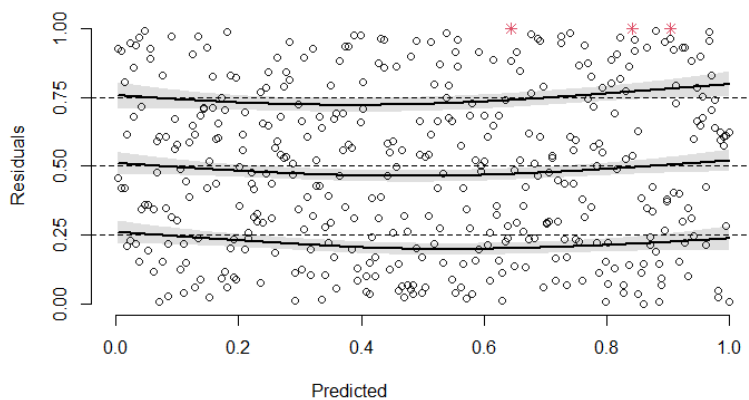


Figura 3.24: Gráfico dos Resíduos Quantílicos *vs* Valores Preditos

Na Figura 3.24 está representado um gráfico dos resíduos quantílicos *versus* os valores preditos pelo modelo. Este gráfico permite detetar desvios da uniformidade na direção y . As linhas presentes no gráfico são regressões quantílicas que mostram os quantis 0,25, 0,50 e 0,75. Numa situação ideal estas linhas deveriam ser retas horizontais para cada quantil. As linhas observadas no gráfico obtido para o modelo não são retas horizontais, mas muito próximas destas.

Capítulo 4

Conclusão e Trabalho Futuro

O principal objetivo deste trabalho foi identificar os fatores que influenciam a ocorrência de acidentes rodoviários. Fatores esses que podem ser características relativas ao estado do pavimento, características geométricas, ao tráfego e ao segmento de autoestrada. Com os valores obtidos nas três autoestradas em estudo (A4, A44 e A41) foi criada uma base de dados constituída por 416 observações e 32 variáveis.

A análise exploratória dos dados permite estudar uma possível relação ou associação entre a variável resposta e as variáveis explicativas, relação essa que pode posteriormente ser estudada e compreendida com a construção de um modelo.

Ao realizar a análise exploratória pode inferir-se que a autoestrada com maior tráfego médio diário anual total no tempo em estudo é a A44 (129 712 veículos), sendo também a autoestrada com menor extensão (3,926km), com maior incidência de acidentes por quilómetro (33 acidentes/km) e com menor número de acidentes em estudo (128). A autoestrada A41, pelo contrário, é a autoestrada com menor tráfego (64 369 veículos), menor incidência de acidentes por quilómetro (12 acidentes/km), maior extensão (14,500km) e com maior número de acidentes em estudo (171). No caso da A4 têm um tráfego médio diário anual de 101 127 veículos, com extensão de 7,504km, com incidência de 20 acidentes/km e com um total do número de acidentes em estudo de 148.

Neste trabalho estudou-se Modelos Lineares Generalizados Mistos para dados de contagem (número de acidentes) recorrendo à distribuição de *Poisson*. Foi tomado como efeito aleatório nos modelos a variável *troço* e como variável resposta

o número de acidentes.

Os resultados da análise da regressão com efeitos aleatórios com distribuição de *Poisson* sobre o número de acidentes ocorridos nas autoestradas A4, A44 e A41 do norte do país entre 2014 a 2017 demonstraram que as variáveis explicativas relativas ao atrito (*GRIP*), ao número de nós (*No*), o número total de ligeiros (*Ligeiros*), o número total de pesados (*Pesados*), a média do inverso do raio (*MIR*) e a média da inclinação negativa (*MIN*) influenciam o número de acidentes.

Os coeficientes estimados associados às variáveis *No*, *Ligeiros*, *Pesados* e *MIR* são positivos, o que significa o aumento do número de acidentes com o aumento do valor dessas variáveis. Relativamente às variáveis do atrito e da média da inclinação negativa, os coeficientes estimados são negativos o que leva a uma diminuição no número de acidentes.

O coeficiente associado à variável *GRIP_{mn}* relativa ao atrito sendo negativo significa que para o mesmo troço e com as restantes variáveis constantes há uma diminuição do número de acidentes em cerca de 85,7%. Já o coeficiente da variável *MIN* revela uma diminuição do número de acidentes em cerca de 17,3%. No caso da variável *Ligeiros* o seu coeficiente positivo significa um aumento de 1,4% no número de acidentes no mesmo troço e com as restantes variáveis constantes enquanto que na variável *Pesados*, o coeficiente revela um aumento de 2 vezes no número de acidentes. Relativamente à variável *No*, esta aumenta o número de acidentes cerca de 2 vezes. Por fim, com o coeficiente da variável *MIR* pode-se concluir que para o mesmo troço e as restantes variáveis constantes o aumento de uma unidade de medida significa um aumento de cerca de 3,7% no número de acidentes.

O aumento do atrito significa um estado do piso melhor e o aumento da média da inclinação negativa leva a uma diminuição do declive da estrada, o que proporciona uma diminuição do número de acidentes. No caso do aumento do número de veículos, quer ligeiros, quer pesados existe inevitavelmente um aumento de acidentes, bem como com o aumento de nós (entradas e saídas das autoestradas).

Para um trabalho futuro, seria interessante introduzir ainda o efeito aleatório da autoestrada. Seria ainda interessante estudar estes mesmos dados com Modelo Poisson Inflacionado de Zeros com efeito aleatório.

Bibliografia

- [1] ANSR (2020). Contributos para a visão zero 2030 – até 30 de Outubro. <http://www.ansr.pt/Noticias/Pages/Estrat%C3%A9gia-Nacional-de-Seguran%C3%A7a-Rodovi%C3%A1ria-2021-2030.aspx>
- [2] Antunes, M., Marecos, V. (2008). Indicadores de Desempenho para Pavimentos Rodoviários. *V Congresso Rodoviário Português*, Centro Rodoviário Português, Estoril.
- [3] Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- [4] Breslow, N., Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.
- [5] Cabral, M. S., Gonçalves, M. H. (2011). *Análise de Dados Longitudinais. Sociedade Portuguesa de Estatística*.
- [6] Cameron, A., Trivedi, P. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1, 29–53.
- [7] Costa, S. C. (2003). *Modelos Lineares Generalizados Mistos para Dados Longitudinais*. Tese de Doutoramento em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica., Escola Superior de Agricultura Luiz de Queiroz - Universidade de São Paulo, Brasil.
- [8] Couchino, R. P. (2011) *Medição do Coeficiente de Atrito da Superfície de Pavimentos*. Tese de Mestrado em Engenharia Civil na Área de Especialização de Vias de Comunicação e Transportes. Instituto Superior de Engenharia de Lisboa, Portugal.

- [9] Cunto, F. J. (2009). Projeto de Pesquisa: Modelos de Previsão de Acidentes de Trânsito em Vias Urbanas Brasileiras – Edital MCT/CNPq 18/2009 – UFC - Fortaleza - CE.
- [10] Dunn, P. K., Smyth, G. K. (1996) *Randomized Quantile Residuals*. Department of Mathematics. University of Queensland, Brisbane, Q 4072, Australia.
- [11] Dunn, P. K., Smyth, G. K. (1996) Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 236-244.
- [12] Faraway, J. J. (2006). *Extending the Linear Model with R: Generalized Linear Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC, 2^a edição.
- [13] Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (2008). *Longitudinal Data Analysis*. Chapman and Hall/CRC.
- [14] Harting, F. (2022) DHARMA- an R package for Residual Diagnostics of GLMMs. Theoretical Ecology. <https://theoreticalecology.wordpress.com/2016/08/28/dharma-an-r-package-for-residual-diagnostics-of-glmms/>
- [15] Harting, F. (2022). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models*. Theoretical Ecology. University of Regensburg.
- [16] Hedeker, D. (2005). *Generalized linear mixed models*. Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, New York.
- [17] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer New York.
- [18] Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics*, 15,3, 209-225.
- [19] Nogueira, A. S. (2015) *Avaliação de rodeiras de pavimentos rodoviários com recurso a varrimento laser*. Tese de Mestrado em Engenharia Civil: Estruturas e Construção. Universidade da Beira Interior, Covilhã.
- [20] R (2022) Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing,

- [21] Relhas, R. J. (2012) *Elaboração do Projeto de um Protótipo para a Avaliação do Coeficiente de Atrito entre Superfícies*. Tese de Mestrado em Engenharia Mecânica. Faculdade de Engenharia da Universidade do Porto, Porto.
- [22] Sinay, M. C. F.; Tamayo, A. S. (2005) *Segurança Viária: Uma Visão Sistémica*. In: Anais do Rio de Transportes III, Rio de Janeiro
- [23] Silveira, F. L. (2007) *Inclinações das Ruas e das Estradas*. Publicado na Revista Física na Escola 8^o volume, n^o 2. Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.
- [24] Stroup, W. W. (2016). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press.
- [25] Tvardíková, K. (2010). Bird abundances in primary and secondary growths in papua new guinea: a preliminary assessment. *Tropical Conservation Science*, 3, 373–388.
- [26] Zuur A., Ieno E., Walker N., Saveliev A., Smith G. (2009). *Mixed effects models and extensions in ecology with R*, Springer.