



Universidade do Porto

Faculdade de Engenharia

Departamento de Engenharia Civil

**A Segurança Rodoviária no Processo de Planeamento de Redes
de Transporte em Meio Urbano**

Sara Maria Pinho Ferreira
(Mestre em Vias de Comunicação)

Orientador: Prof. Américo Henrique Pires da Costa
(Professor Associado da Faculdade de Engenharia da Universidade do Porto)

Co-Orientador: Doutor João Paulo Lourenço Cardoso
(Investigador Principal do Laboratório Nacional de Engenharia Civil)

Dissertação para Doutoramento em Engenharia Civil
na Faculdade de Engenharia da Universidade do Porto

Março 2010

AO TIAGO E AO AFONSO

...

AGRADECIMENTOS

Ao Professor Américo Henrique Pires da Costa e ao Doutor João Cardoso que orientaram cientificamente o presente trabalho, exprimo o meu profundo reconhecimento pelas condições propiciadas, pelo empenho sempre demonstrado e pela forma amigável como sempre me apoiaram e incentivaram.

Não poderia ainda deixar de expressar os meus agradecimentos a todos aqueles que directa ou indirectamente contribuíram para que este trabalho se concretizasse. De um modo muito especial, quero expressar a mais sincera gratidão:

- ao Professor Arnaldo Sousa Melo, ex-Director da Secção de Vias de Comunicação, pelas condições propiciadas e pela amizade;
- ao Prof. António Couto pela disponibilidade e colaboração prestada ao longo deste trabalho;
- aos colegas da Secção de Vias de Comunicação, em particular, aos Profs. Carlos Rodrigues e José Pedro Tavares e ao Engº António Vasconcelos que em diferentes fases deste trabalho se disponibilizaram e que, por diversos modos, manifestaram o seu estímulo e encorajamento;
- à Câmara Municipal do Porto, na pessoa de Engº João Neves, por propiciar as necessárias condições ao levantamento dos dados de tráfego da cidade do Porto.

Finalmente, o meu agradecimento à família e amigos em especial ao Tiago e ao Afonso, pelo sacrifício e estímulo sempre manifestado.

RESUMO

No período entre 1975 e 2006, Portugal apresentou uma evolução na redução da sinistralidade rodoviária de cerca de 74%. Apesar dos resultados alcançados, Portugal ainda está abaixo da média europeia no que concerne ao número de mortos por milhão de habitantes. A ocorrência de um número elevado de vítimas, em particular de vítimas mortais, concentra-se significativamente no meio urbano (71% dos acidentes com vítimas e 47% das vítimas mortais).

Os acidentes rodoviários resultam da interacção entre o condutor, o veículo e a infra-estrutura. Com vista à redução da sinistralidade, foram desenvolvidas e aplicadas medidas aos três elementos, destacando-se neste trabalho a infra-estrutura. Se inicialmente essas medidas visavam sobretudo o tratamento de locais com um número elevado de acidentes e/ou vítimas, mais recentemente, e com o objectivo de prevenir a ocorrência de acidentes por todos indesejados, as medidas visam a antecipação do risco de acidente associado à infra-estrutura. Nesse sentido, surgem conceitos aplicados à fase de projecto e mais recentemente à fase de planeamento. Contudo, na fase de planeamento poucos elementos da rede viária estão ainda definidos e/ou detalhados sendo, por isso, mais difícil de estabelecer relações de causa/efeito a partir das quais se identificam factores de risco.

O objectivo do presente trabalho é contribuir para a análise/avaliação da sinistralidade em fase de planeamento da rede viária do meio urbano. Para tal, desenvolveram-se modelos de previsão de acidentes considerando dois tipos de abordagem – modelos de resposta quantitativa e qualitativa. A primeira tem sido intensamente aplicada no âmbito da sinistralidade e cuja resposta obtida pelos modelos corresponde a um valor estimado para a frequência de acidentes num determinado ano horizonte. Na segunda abordagem, menos aplicada no âmbito a sinistralidade, a resposta obtida corresponde à probabilidade de ocorrer determinadas categorias, às quais corresponde um grau/nível de sinistralidade.

Para o desenvolvimento e aplicação destes modelos, considerou-se o caso da cidade do Porto com dados dos acidentes ocorridos entre 2001 e 2005. A caracterização da rede (arcos e nós) e respectiva envolvente (ocupação do solo) foi obtida a partir do Plano Director Municipal, sendo os volumes de tráfego na rede estimados a partir de contagens reais no terreno e da aplicação de um modelo de afectação de tráfego.

ABSTRACT

From 1975 to 2006, Portugal was able to decrease the total number of road accidents in the country by 74%. Despite this, Portugal is still below the European average in terms of the number of road fatalities per million inhabitants. The majority of the number of road victims occurred in urban roads (71% of accidents with victims and 47% of fatalities).

Road accidents are a result of the interaction between the driver, the vehicle and the infrastructure. In order to reduce the number of road accidents, research and actions were developed and implemented in these three elements. This work provides a particular focus on the specific domain relating to infrastructures. Initial efforts were aimed to dealing with sites which had a higher number of road accidents and/or victims. More recently, the aim has been prevention through the analysis of potential hazardous elements during the road project and the transportation planning process phases. However, during the planning process phase there is a lack of available data related to the features of the road networks which makes it difficult to detect risk effects.

Thus, the objective of this work was to develop a series of models that permit the assessment of an urban road network during the planning process. These models were based on two different approaches namely, quantitative response models and qualitative response models. The former is usually applied to road safety and is used to estimate the number of accidents for the prognosis year. In the latter, which is not applied to road safety as often, the outcome is a probability of a discrete choice among a set of alternatives representing a certain degree of accident frequency.

In order to accomplish this goal, several models have been developed and applied to the concrete case of the city of Porto with accident data ranging from 2001 to 2005. Information relating to the road network, in particular - nodes and links and its environmental (land-use) were obtained from the Municipal Master Plan. The traffic flow was performed by actual counts and assignment traffic model.

ÍNDICE

1. INTRODUÇÃO	1
1.1 Preâmbulo	1
1.2 Objectivos e metodologia base	9
1.3 Estrutura da dissertação	10
1.4 Referências bibliográficas.....	11
 2. O PLANEAMENTO DO SISTEMA DE TRANSPORTES	 13
2.1 O processo de planeamento do sistema de transportes em meio urbano	14
2.2 O processo de tomada de decisão	23
2.3 Modelos de transporte	28
2.4 Estrutura organizacional	31
2.5 Conclusões	33
2.6 Referências bibliográficas.....	34
 3. A SEGURANÇA RODOVIÁRIA NO PROCESSO DE PLANEAMENTO DO MEIO URBANO	 37
3.1 Acções na infra-estrutura	39
3.2 A segurança rodoviária no planeamento dos transportes.....	46
3.3 Conclusões	56
3.4 Referências bibliográficas.....	57
 4. MODELOS DE PREVISÃO DE ACIDENTES	 61
4.1 Modelos de resposta quantitativa.....	62
4.2 Modelos de resposta qualitativa.....	78
4.3 Conclusões	88
4.4 Referências bibliográficas.....	93
 5. DESENVOLVIMENTO DE UM MODELO MATEMÁTICO DE APOIO À DECISÃO EM FASE DE PLANEAMENTO	 99
5.1 Descrição dos modelos.....	99

5.2	Dados considerados na aplicação.....	108
5.2.1	Dados da infra-estrutura	108
5.2.2	Dados de acidentes	113
5.2.3	Dados de tráfego	116
5.3	Modelos de resposta quantitativa.....	119
5.3.1	Modelos para arcos	123
5.3.2	Modelos para nós	153
5.4	Modelos de resposta qualitativa.....	170
5.4.1	Modelos para arcos	174
5.4.2	Modelos para nós	185
5.5	Conclusões	193
5.6	Referências bibliográficas.....	203
6.	CONCLUSÕES E PERSPECTIVAS FUTURAS	207
6.1	Síntese geral e principais conclusões	207
6.2	Desenvolvimento do estudo	214
6.3	Referências bibliográficas.....	216
	ANEXO 1 - GLOSSÁRIO DAS ABREVIATURAS MAIS UTILIZADAS	219
	ANEXO 2 - RESULTADOS DA APLICAÇÃO DOS MODELOS (LIMDEP 9.0).....	225
	ANEXO 2.I.....	227
	ANEXO 2.II	241
	ANEXO 2.III	245
	ANEXO 2.IV	249
	ANEXO 2.V	253
	ANEXO 2.VI.....	257
	ANEXO 2.VII	263
	ANEXO 2.VIII	267
	ANEXO 2.IX.....	271
	ANEXO 2.X	275
	ANEXO 2.XI.....	279
	ANEXO 2.XII	283

ANEXO 2.XIII	287
ANEXO 2.XIV	291
ANEXO 2.XV	295
ANEXO 2.XVI.....	299
ANEXO 2.XVII.....	303
ANEXO 2.XVIII	307
ANEXO 2.XIX.....	311
ANEXO 2.XX	315
ANEXO 2.XXI.....	319
ANEXO 2.XXII.....	325
ANEXO 2.XXIII	329
ANEXO 2.XXIV	335
ANEXO 2.XXV	341
ANEXO 2.XXVI.....	345

ÍNDICE DE FIGURAS

Figura 1.1 - Evolução do número de mortos, feridos e acidentes por população na UE.....	2
Figura 1.2 - Evolução do número de mortos por milhão de habitantes em acidentes rodoviários	3
Figura 1.3 - Organização e fluxos de informação da ENSR 2008-2015	5
Figura 2.1 - Etapas do processo de planeamento.....	17
Figura 2.2 - Processo de definição de metas	19
Figura 2.3 - Métodos de avaliação e medição dos impactes em diferentes países	22
Figura 2.4 - Estrutura do modelo de 4 passos	30
Figura 3.1. - Organigrama representativo do processo de análise a longo prazo incorporando a SR	51
Figura 4.1 - Gráfico representativo de um modelo ordenado para $J = 4$	83
Figura 5.1. - Exemplos de uma intersecção simples a); intersecção múltipla b); intersecção singular c)	109
Figura 5.2. - Classificação hierárquica da rede do Porto.....	112
Figura 5.3. - Classificação da ocupação do solo do concelho do Porto	112
Figura 5.4. - Organigrama da base de dados dos acidentes só com danos materiais	114
Figura 5.5. - Organigrama da base de dados dos acidentes com vítimas	115
Figura 5.6. - Histograma do número de acidentes nos arcos – total e com vítimas	125
Figura 5.7. - Gráfico comparativo da FIE(X) das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ com as funções Potência, Gama e Exponencial.....	126
Figura 5.8. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo BN.....	134
Figura 5.9. - Gráfico dos valores de “leverage” e de $2K/n$ referente à base de dados dos arcos.....	135
Figura 5.10. - Gráfico dos valores da função $Dfit_i$ das observações registadas nos arcos	136

Figura 5.11. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE sem as observações “outliers” e de influência	136
Figura 5.12. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo com heteroscedasticidade da equação (5.2).....	139
Figura 5.13. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo com função Translog da equação (5.4).....	148
Figura 5.14. - Gráfico comparativo da FIE(X) das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ com as funções Potência, Gama e Exponencial para os acidentes com vítimas	150
Figura 5.15. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo BN aplicado aos acidentes com vítimas.....	152
Figura 5.16. - Histograma do número de acidentes nos nós – total e com vítimas	155
Figura 5.17. - Gráfico comparativo da FIE(X) das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ com as funções Potência, Gama e Exponencial para o total dos acidentes	156
Figura 5.18. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ elaborados segundo o método CURE para o modelo BN.....	161
Figura 5.19. - Gráfico dos valores de “leverage” e de $2K/n$ referente à base de dados dos nós.....	162
Figura 5.20. - Gráfico dos valores da função Df_{iti} das observações registradas nos nós	162
Figura 5.21. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ elaborados segundo o método CURE para o modelo com heteroscedasticidade da equação (5.6).....	164
Figura 5.22. - Probabilidades do modelo ordenado estimado para os arcos e o efeito da variação do valor de x nas probabilidades de previsão.....	177
Figura 5.23. - Probabilidades do modelo <i>ordenado</i> estimado para os nós.....	188
Figura 5.24. - Gráficos dos resíduos acumulados dos modelos quantitativos e qualitativos aplicados aos arcos e nós	196
Figura 5.25. - Gráficos representativos da distribuição das probabilidades de ocorrer as três categorias	202

Figura 6.1. - Estrutura proposta para o processo de planeamento da SR em meio urbano 210

ÍNDICE DE QUADROS

Quadro 1.1 - Taxas de mortalidade por modo de transportes.....	6
Quadro 2.1 – Estrutura descritiva de avaliação	20
Quadro 4.1 - Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta quantitativa.....	89
Quadro 4.2 - Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta qualitativa.....	91
Quadro 5.1 - Classificação hierárquica dos arcos.....	110
Quadro 5.2 - Classificação hierárquica dos nós	111
Quadro 5.3 - Classificação da ocupação do solo	111
Quadro 5.4 - Distribuição dos acidentes ao longo dos 5 anos.....	115
Quadro 5.5 - Distribuição dos acidentes nos arcos e nos nós (5 anos).....	116
Quadro 5.6 - Proporções dos volumes de hora de ponta por zona	118
Quadro 5.7 - Medidas de avaliação da qualidade de ajuste dos modelos.....	121
Quadro 5.8 - Dados estatísticos das variáveis TMDA, comprimento e densidade de intersecções dos arcos.....	124
Quadro 5.9 - Dados estatísticos dos acidentes ocorridos nos arcos distribuídos por ano e gravidade e sua distribuição percentual pelas variáveis ocupação do solo e classificação hierárquica	124
Quadro 5.10 - Matriz de correlações entre as variáveis independentes da equação (5.2)	127
Quadro 5.11 - Resultados da estimação do modelo BN da equação (5.2)	128
Quadro 5.12 - Pseudo-elasticidade das “dummies” da equação (5.2).....	131
Quadro 5.13 - Resultados da estimação do modelo com heteroscedasticidade da equação (5.2).....	138
Quadro 5.14 - Resultados da estimação do modelo com PA da equação (5.2).....	141
Quadro 5.15 - Resultados da estimação do modelo ZIP da equação (5.2).....	143
Quadro 5.16 - Valores estimados e níveis de confiança para os diferentes modelos da equação (5.2) aplicados à base de dados dos arcos.....	144

Quadro 5.17 - Matriz de correlação entre as variáveis independentes cruzadas da equação (5.4).....	146
Quadro 5.18 - Resultados da estimação do modelo com função Translog da equação (5.4) ..	147
Quadro 5.19 - Resultados da estimação do modelo BN da equação (5.2) para os acidentes com vítimas.....	151
Quadro 5.20 - Dados estatísticos das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$	154
Quadro 5.21 - Dados estatísticos dos acidentes ocorridos nos nós distribuídos por ano e gravidade e sua distribuição percentual pelas variáveis ocupação do solo e classificação hierárquica	154
Quadro 5.22 - Matriz de correlações entre as variáveis independentes da equação (5.6)	157
Quadro 5.23 - Resultados da estimação do modelo BN da equação (5.6)	158
Quadro 5.24 - Resultados da estimação do modelo com heteroscedasticidade da equação (5.6).....	163
Quadro 5.25 - Resultados da estimação do modelo com PA da equação (5.6)	165
Quadro 5.26 - Resultados da estimação do modelo ZIP da equação (5.6)	167
Quadro 5.27 - Valores estimados e níveis de confiança para os diferentes modelos da equação (5.6) aplicados à base de dados dos nós	168
Quadro 5.28 - Matriz de correlação entre as variáveis independentes cruzadas da equação (5.7).....	170
Quadro 5.29 - Classe de categorias da variável de resposta com base na frequência de acidentes - arcos.....	174
Quadro 5.30 - Resultados da estimação do modelo ordenado probit – arcos.....	176
Quadro 5.31 - Matriz de previsões do modelo ordenado probit – arcos	179
Quadro 5.32 - Resultados da estimação do modelo multinomial logit – arcos	181
Quadro 5.33 - Matriz de previsões do modelo multinomial logit – arcos	183
Quadro 5.34 - Categorias da variável de resposta com base na ocorrência de vítimas – arcos	184
Quadro 5.35 - Categorias da variável de resposta com base na frequência - nós.....	185
Quadro 5.36 - Resultados da estimação do modelo ordenado probit – nós.....	187
Quadro 5.37 - Matriz de previsões do modelo ordenado probit – nós	189
Quadro 5.38 - Resultados da estimação do modelo multinomial logit – nós	190
Quadro 5.39 - Matriz de previsões do modelo multinomial logit – nós.....	191

Quadro 5.40 - Categorias da variável de resposta com base para o total dos acidentes dos modelos de nós	192
Quadro 5.41 - Probabilidades percentuais de ocorrer cada categoria em cenários hipotéticos de um arco.....	199
Quadro 5.42 - Probabilidades percentuais de ocorrer cada categoria em cenários hipotéticos de um nó	200

1. INTRODUÇÃO

1.1 Preâmbulo

Estima-se que desde a invenção do automóvel há um século atrás morreram em acidentes rodoviários cerca de 30 milhões de pessoas (GRSP, 2010). A dimensão mundial do problema da segurança rodoviária (SR) foi reconhecida pela Organização Mundial de Saúde (OMS), que em 1974 classificou, através de uma resolução, os acidentes rodoviários como um importante problema de saúde pública, do qual resultam custos indesejáveis ao nível dos recursos médicos. Nos países com baixo e médio rendimento *per capita* estima-se que entre 30% a 85% dos internamentos hospitalares devido a traumatismo são de vítimas de acidentes rodoviários (GRSP, 2010). O crescente investimento em infra-estruturas rodoviárias, em especial nos países em desenvolvimento tende a aumentar a exposição ao risco de morte e ferimentos por este tipo de acidentes. A menos que este investimento se faça a par de um plano de acção que assegure bons níveis de SR, esse aumento da exposição vai consequentemente gerar um aumento de mortos e feridos. Segundo a OMS, em 2020 os acidentes rodoviários serão, no mundo, a terceira causa de morte e incapacidade.

E, contudo, a maior parte dos acidentes rodoviários pode ser prevenida.

Os países com tradição no investimento rodoviário enfrentaram já esta problemática e iniciaram a sua acção para a mitigação dos acidentes. Esta acção consistiu num conjunto de medidas aplicadas às três componentes do sistema rodoviários – condutor, veículo e infra-estrutura. Sendo o âmbito deste trabalho a infra-estrutura, salienta-se que a este nível as primeiras medidas tiveram como objectivo corrigir os problemas detectados na infra-estrutura através da identificação e correcção dos locais onde se concentra um maior número de acidentes – “pontos negros”. Os programas de tratamento dos pontos negros reduziram, na Austrália, as vítimas mortais em cerca de 20 vidas por ano (Australian Transport Safety Bureau, 2001). A par, foi também sendo melhorada a qualidade das infra-estruturas nomeadamente ao nível da sinalização. A aplicação destas medidas revelou-se muito positiva mas na prática só podem ser aplicadas após a ocorrência de um número elevado de acidentes. Por esse facto, desenvolveu-se, complementarmente, uma visão mais

preventiva iniciando-se assim o desenvolvimento e aplicação de medidas de prevenção. Simultaneamente à aplicação destas medidas foi-se verificando uma crescente sensibilização das populações e dos governos para esta problemática reflectindo-se na criação de organismos que se responsabilizassem pela análise desta problemática e desenvolvimento das respectivas soluções. Nesse sentido, estes organismos passaram a dar relevância ao tema através, por exemplo, da definição de estratégias, objectivos e metas para a SR. Em muitos países, a SR faz parte da agenda política e têm vindo a realizar esforços para tornar as estradas mais seguras.

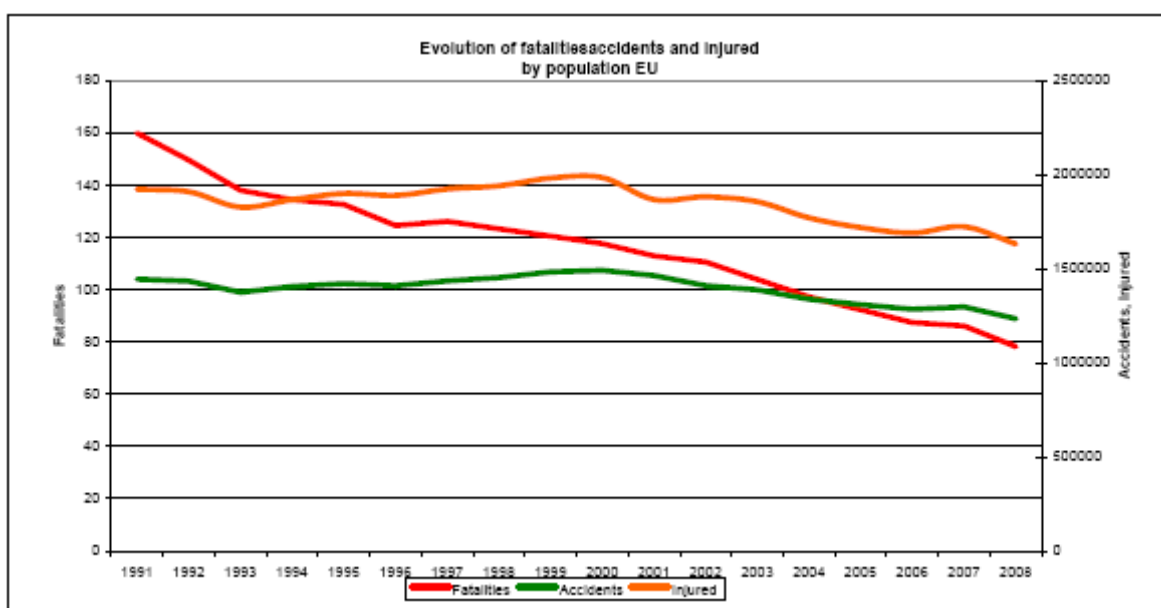


Figura 1.1 - Evolução do número de mortos, feridos e acidentes por população na UE (Fonte: CARE - base de dados dos acidentes rodoviários na UE ou publicações nacionais)

A União Europeia (UE) estabeleceu o objectivo, para os países integrados, de reduzirem para metade o número de mortos até 2010 relativamente a 2000. No ano 2000, cerca de 40 800 pessoas morreram devido a acidentes rodoviários na Europa. O número anual de mortos por milhão de habitantes, no conjunto dos países da Europa, corresponde a cerca de 110, variando entre um mínimo de 60 e um máximo de 200 (ETSC, 2003a).

A evolução destes números tem sido, de uma forma geral, muito favorável no conjunto dos países da UE principalmente no que respeita ao número de mortos, tal como é possível de observar pelo gráfico da Figura 1.1.

No caso de Portugal, os resultados alcançados desde 1975 permitiram passar o país do último lugar (Europa dos 15) para uma posição acima do meio da tabela em 2006 (Europa dos 27) aproximando-se assim da média europeia em termos de número de mortos por milhão de habitantes (ANSR, 2009), tal como se pode observar pelo gráfico da Figura 1.2.

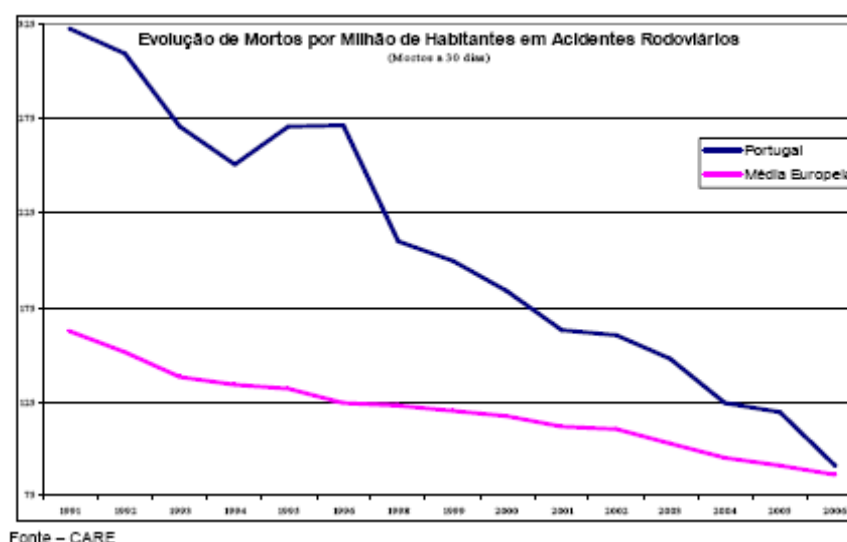


Figura 1.2 - Evolução do número de mortos por milhão de habitantes em acidentes rodoviários (Fonte: (ANSR, 2009))

Esta tendência para a diminuição dos acidentes neste período de tempo pode ser consequência do maior investimento de Portugal nesta área que, como país da UE, subscreveu o objectivo de reduzir para metade o número de mortos até 2010. Para tal, foi desenvolvido o Plano Nacional de Prevenção Rodoviária (PNPR) de 2003 (Ministério da Administração Interna, 2003) em que se estabeleceram vários objectivos considerados prioritários em três áreas principais – educação contínua do utente, ambiente rodoviário seguro e quadro legal e sua aplicação.

Tendo em conta que o objectivo estabelecido para 2010 foi já alcançado, foi aprovado¹ em 2009 um documento que define a Estratégia Nacional de Segurança Rodoviária (ENSR) para o período 2008-2015. Neste documento estabelece-se um novo objectivo: colocar Portugal entre os dez países da UE com mais baixa sinistralidade rodoviária, medida em mortos por milhão de habitantes. Para avaliar e monitorizar as acções definidas para

¹ Diário da República, 1ª Série – Nº 122 – 26 de Junho de 2009

alcançar este objectivo, estabeleceram-se dois períodos 2008-2011 e 2012-2015. Assim, convertendo o objectivo para estes dois períodos em número de mortos por milhão de habitantes, equivale a atingir em 2011 o valor de 78 mortos por milhão de habitante, ou seja, uma redução de 14,3% relativamente a 2006 e em 2015 o valor de 62 mortos por milhão de habitantes o que corresponde a uma redução de 31,9% em relação a 2006.

Para se atingirem estes objectivos, a ENSR estabeleceu um conjunto de objectivos estratégicos em diversas áreas de actuação, dos quais resultaram objectivos operacionais a colocar em prática através de um conjunto de acções chaves. Estas acções abrangem várias áreas de actuação, envolvendo, por isso, várias entidades sendo necessário coordenar esta equipa multidisciplinar. A Figura 1.3 resume a organização e fluxos de informação da ENSR a serem desenvolvidos na fase da sua implementação.

Nesta estratégia de actuação, particular atenção terá que ser dada ao meio urbano tendo em conta o ritmo inferior à média com que a diminuição da sinistralidade aí se tem verificado nos últimos anos em especial no período mais recente (ANSR, 2009). Em Portugal, no ano 2005, 43,1% das vítimas mortais ocorreram em meio urbano sendo que a média europeia (Europa dos 17) para o mesmo ano foi de 35,3% (ECRS, 2010).

O meio urbano é um ambiente rodoviário variado e complexo o que potencia a ocorrência de um maior número de conflitos. Características como: indefinição das hierarquias viárias; zonas históricas; partilha dos canais por diferentes meios de transporte com características muito diversas; grande ocupação do espaço, o que não permite criar novas vias ou segregar fluxos; diferentes movimentos e funções (mobilidade, acessibilidade, actividades económicas e sociais, etc.); existência de mobiliário urbano e outros objectos que diminuem a visibilidade e dispersam a atenção dos condutores tornam difícil a redução da sinistralidade no meio urbano, acrescentando o facto de existir pouca articulação entre a administração local, responsável pela actuação no meio urbano, e a administração central, responsável pela SR.

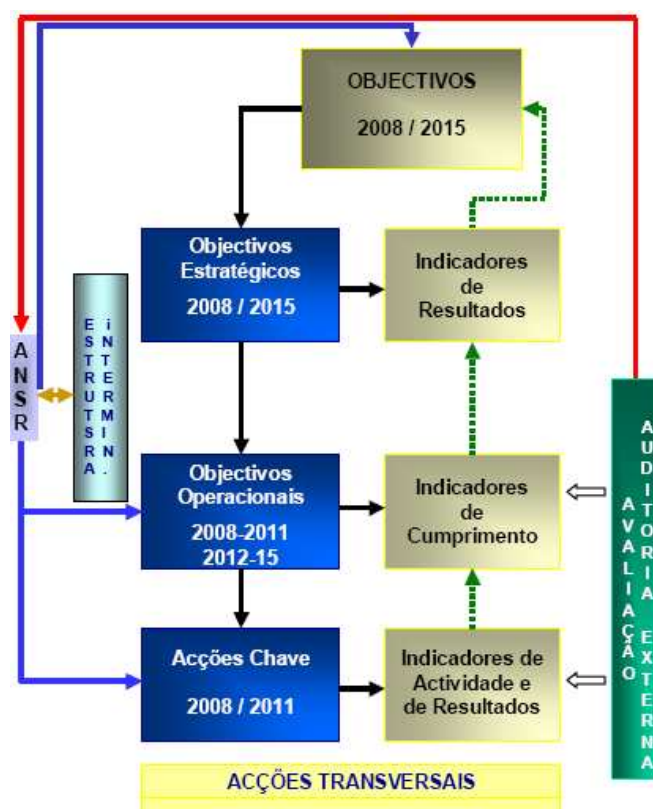


Figura 1.3 - Organização e fluxos de informação da ENSR 2008-2015 (Fonte: (ANSR, 2009))

A interacção muitas vezes conflituante entre os diferentes modos de transporte tem consequências nefastas nos utilizadores vulneráveis como sejam os peões e ciclistas. Pelos valores do Quadro 1.1 é possível comparar as taxas de mortalidade medidas pela distância percorrida e pela duração dos percursos efectuados. Em qualquer um dos dois tipos de taxa de mortalidade, o motociclo é o meio mais perigoso. Pelos valores das taxas apresentadas no mesmo quadro é bem evidente o maior risco dos três tipos de utilizadores vulneráveis relativamente aos do carro e autocarro, sendo que aqueles utilizadores viários são característicos do meio urbano.

Quadro 1.1 - Taxas de mortalidade por modo de transportes (Fonte:(ETSC, 2003b))

Modo de transporte	Mortos por 100 milhões de pessoa×km	Mortos por 100 milhões de pessoa×hora
Motociclo	13,8	440
A pé	6,4	25
Bicicleta	5,4	75
Carro	0,7	25
Autocarro	0,07	2

Nos países industriais, a intervenção na infra-estrutura rodoviária tem sido essencialmente na perspectiva de satisfazer as necessidades do tráfego crescente e da mobilidade. O meio urbano também acompanhou esta tendência. As vias foram construídas ou alargadas face ao crescimento do tráfego automóvel, muitas vezes com efeitos nefastos para os utentes vulneráveis e para os residentes dos meios urbanos. As velocidades máximas aumentaram, os passeios ficaram mais estreitos, o estacionamento invadiu os espaços inicialmente concebidos para os peões e ciclistas, surgiram novos e mais largos acessos urbanos, entre outros factores. Mais ainda, os já limitados espaços pedonais, que não foram apropriadamente mantidos, viram-se obstruídos por todo o tipo de mobiliário urbano e sinalização vertical.

Com a expansão das áreas urbanas a distância casa-trabalho aumentou, dificultando assim as deslocações a pé ou de bicicleta como meio de transporte diário. A política de transportes centrou-se, essencialmente, no veículo privado tendo sido proferido frases como a do Presidente da França, George Pompidou - “as cidades devem adaptar-se aos veículos privados” (Road Transport Research, 1998). Foram projectadas novas áreas residenciais tendo por base o princípio da segregação completa dos peões e veículos motorizados, primeiro na Suécia e posteriormente em alguns aglomerados britânicos. Este conceito de segregação completa começou por espalhar-se sem por em questão a prioridade geral dada ao veículo particular (Road Transport Research, 1998).

Contudo, nos finais dos anos 70, esta tendência começou a reverter-se quando se verificou que as ruas das vilas antigas e dos centros das cidades não podiam suportar um aumento indefinido de tráfego. Instituíram-se, então, em alguns países tais como França e Países Baixos, planos de gestão de tráfego, de modo a otimizar a utilização do espaço das ruas existentes. Tais planos também passavam pela redução do tráfego de automóveis privados

nos centros das cidades e, como consequência, o aumento da oferta de transportes públicos e das vias criadas para os ciclistas. As ruas pedonais começaram a surgir, essencialmente em áreas comerciais ou turísticas. No entanto, os movimentos pedonais foram apenas considerados como deslocações de curta duração correspondentes a trajectos entre os veículos e o destino final resultando, por exemplo, na construção de parques de estacionamento nas saídas das zonas pedestres.

Entretanto, a qualidade das áreas residenciais segregadas começou a ser questionada, dado que não facilitavam o planeamento dos transportes públicos e não permitiam qualquer tipo de actividades mistas. Face às novas exigências surgiu um novo conceito de integração de tráfego misto designado “*woonerf*” na Holanda, fundamentado na ideia de que nestes locais os condutores devem abrandar e dar prioridade aos utilizadores vulneráveis das vias, particularmente as crianças (Road Transport Research, 1998). Reverte-se assim a tendência da segregação e surge a ideia de que as ruas urbanas podem ter outra função para além de simplesmente suportar o tráfego motorizado.

Os técnicos profissionais tomaram consciência de que a publicidade e o reforço policial não eram a resposta para induzir um comportamento mais adequado dos condutores e que a infra-estrutura e o meio envolvente podiam desempenhar um papel muito importante neste contexto. Desta forma, começou a desenvolver-se o conceito de acalmia de tráfego.

Nos diferentes países da Europa, constatou-se que, ao longo do tempo, não só houve alterações na infra-estrutura, nos utilizadores e nos veículos como também se verificam diferenças substanciais entre os países. Essas diferenças têm origem no crescimento da mobilidade inerente a cada país e a forma eficiente como são resolvidos os consequentes efeitos no tráfego. A SR é considerada, em geral, como um problema nacional mas que se manifesta também a nível local. Contudo, em alguns países, temas como o registo dos acidentes e respectivo sistema de base de dados, assim como a discussão de medidas de actuação, nem sempre envolvem as entidades locais sendo, no entanto, estas entidades que, por um lado, têm acesso a informação importante para a análise das causas dos acidentes (por exemplo: medidas de exposição, caracterização da infra-estrutura) e, por outro lado, as responsáveis por actuar neste meio.

No meio urbano, a introdução de diferentes modos de transportes e respectivas vias, são definidas dentro de uma política de transportes a implementar visando dar resposta à distribuição da população no território que é caracterizada através da identificação do tipo de ocupação do solo. Esta ligação entre o sistema de transportes e a ocupação do solo é

reconhecida como importante para garantir as acessibilidades e a mobilidade do meio urbano, sendo por isso uma etapa a considerar para a introdução do conceito de SR no âmbito da prevenção dos acidentes.

A SR pode ser considerada na fase de planeamento quer como uma estratégia política traduzida em objectivos a cumprir para a redução da sinistralidade quer como critério a considerar na avaliação de diferentes alternativas planeadas para o sistema de transportes. Em qualquer uma destas situações o planeamento dos transportes tem como fundamento gerar informação útil para os decisores avaliarem as consequências dos diferentes cenários estabelecidos.

Para se poder gerar informação nesta fase, configurada em possíveis cenários, é necessário reproduzir a realidade através, por exemplo, de modelos matemáticos. Estes modelos permitem simular a realidade e assim avaliar as consequências de determinadas estratégias ou soluções. Os mais comumente utilizados nesta fase são os modelos de transporte que representam as diversas componentes da infra-estrutura de transporte (rede viária, linhas de transporte público, gestão e controlo de tráfego) e os fluxos de tráfego (utilizadores e veículos) de forma a reproduzir as condições operacionais do sistema viário. Através desta informação e de formulações matemáticas simula-se o comportamento dos utilizadores do sistema.

No âmbito da sinistralidade, vários modelos matemáticos têm sido desenvolvidos com o objectivo de prever a frequência da ocorrência de acidentes para determinadas características da rede. No entanto, estes modelos necessitam de uma grande quantidade de informação, tais como volume de tráfego e características geométricas e funcionais da rede, o que constituiu um obstáculo à sua aplicação em fase de planeamento, dado que nesta fase ainda não estão definidos tais elementos da rede. Nesse sentido, o desenvolvimento de modelos a aplicar em fase de planeamento constituiu um desafio que só recentemente alguns investigadores têm explorado, sendo neste campo que se desenvolve o presente trabalho.

1.2 Objectivos e metodologia base

O objectivo principal deste trabalho de investigação consiste em desenvolver uma ferramenta susceptível de ser usada na análise da sinistralidade na fase de planeamento do sistema de transportes do meio urbano. Os resultados da aplicação desta ferramenta poderão constituir uma fonte de informação complementando as já disponíveis para a tomada de decisão. Assim, com este trabalho pretende contribuir-se para que a sinistralidade seja um critério relevante a ser ponderado logo nas primeiras etapas das decisões acerca do sistema de transportes.

A ferramenta desenvolvida e que se descreve neste trabalho, consiste num conjunto de modelos matemáticos que, com base em informação disponível na fase de planeamento de um meio urbano, determinam um grau de gravidade, em termos de sinistralidade. Com esta indicação, é possível adicionar o critério da sinistralidade no processo de selecção da melhor alternativa de actuação na rede urbana.

Os modelos desenvolvidos aplicam-se à rede viária constituída por arcos e nós, utilizando-se para caracterizar estas entidades dados que se possam relacionar com a ocorrência de acidentes mas que também sejam possíveis de obter nesta fase. Sendo que a exposição, nomeadamente o tráfego (pois sem este não há acidentes), é uma variável fundamental para caracterizar a rede, propõe-se neste trabalho que esta informação seja obtida através de um modelo de transportes. As outras variáveis consideradas são definidas no planeamento de uma cidade de Portugal, através do Plano Director Municipal (PDM), e compreendem a identificação do tipo de ocupação do solo e do tipo de classificação viária.

Os modelos desenvolvidos neste trabalho correspondem a dois tipos de abordagem – modelos de resposta quantitativa e modelos de resposta qualitativa. Da primeira abordagem, a informação resultante da aplicação do modelo corresponde a um valor de acidentes previsto para um determinado ano. É a partir deste valor que se identifica a gravidade das diferentes alternativas em termos de sinistralidade. Esta abordagem tem sido habitualmente considerada no âmbito da sinistralidade, estando, por isso, numa fase de desenvolvimento avançada, existindo já um conjunto de variantes matemáticas aplicadas de forma a reduzir os erros e assim melhorar o ajuste à base de dados. Na segunda abordagem, o resultado compreende um código que identifica uma categoria de gravidade definida com base num intervalo de valores de frequência de acidentes. Esta abordagem, menos considerada no âmbito da sinistralidade mas amplamente estudada e aplicada no

planeamento de transportes, tem como objectivo analisar as consequências de decisões através de um resultado qualitativo.

Estas duas abordagens foram aplicadas neste trabalho, ao caso da cidade do Porto, com base nos dados de acidentes ocorridos entre 2001 e 2005. Para os dados do volume de tráfego consideraram-se os valores obtidos através de uma aplicação do modelo de transportes SATURN². Os restantes dados foram obtidos com base em informação disponível no PDM da cidade.

1.3 Estrutura da dissertação

O presente trabalho incide sobre o desenvolvimento e aplicação de modelos de previsão de acidentes (MPA). A especificidade dos modelos apresentados resulta da fase e do meio em cuja aplicação se inserem – fase de planeamento do meio urbano. Dadas as características específicas desta fase é apresentado neste trabalho uma abordagem aos diferentes modos de planeamento de transportes. Apresenta-se ainda uma análise evolutiva das diferentes ferramentas desenvolvidas e aplicadas no âmbito da SR de forma a enquadrar o presente trabalho. Nesse sentido, para além do presente capítulo, o trabalho foi estruturado em cinco capítulos.

No Capítulo 2., analisa-se o processo de planeamento do sistema de transportes. Apesar de existirem diferentes estruturas para o processo de planeamento, algumas etapas são comuns e recomendáveis. Em comum aos diversos processos, encontra-se o objectivo de, através deste processo, obter-se informação para apoio à tomada de decisão. O processo de decisão, tal como o planeamento, pode ser realizado de diferentes formas. Nesse sentido, descrevem-se alguns modelos conceptuais de tomada de decisão. Com este capítulo, pretende-se analisar possíveis formas de integrar a sinistralidade como um critério a ponderar nesta fase.

² modelação do tráfego da rede do Porto realizada no âmbito de um protocolo entre a Faculdade de Engenharia da Universidade do Porto e a Câmara Municipal do Porto

O terceiro capítulo centra-se no tema da SR na fase de planeamento. Inicia-se, no entanto, o capítulo com uma referência às diversas acções na infra-estrutura nos dois tipos de abordagem – acções *a priori* e *a posteriori*. Destas acções destaca-se a acção *a priori* recentemente proposta e designada de avaliação do impacte sobre a segurança rodoviária. No mesmo âmbito, é ainda analisado um conjunto de trabalhos que abordam a SR em fase de planeamento.

O quarto capítulo incide sobre os modelos matemáticos aplicados à modelação dos acidentes. É realizada uma análise aprofundada nos dois tipos de abordagens de modelação considerados neste trabalho. Por um lado, modelos de resposta quantitativa em que o resultado da aplicação do modelo corresponde a um valor indicativo do número de acidentes. Por outro lado, modelos de resposta qualitativa cujo resultado corresponde a uma categoria, de um conjunto de categorias definidas para classificar a gravidade através de um intervalo de número de frequências de acidentes. Para estas duas abordagens é apresentado o estado da arte.

A aplicação das duas abordagens anteriormente referidas é descrita no Capítulo 5. Os modelos são aplicados ao caso da cidade do Porto, cuja base de dados dos acidentes registados compreende quer os acidentes só com danos materiais quer os acidentes com vítimas, ocorridos entre 2001 e 2005. Apresentam-se os resultados das aplicações e face a estes resultados comparam-se as duas abordagens.

Finalmente, no Capítulo 6., expõem-se as principais conclusões do presente trabalho, traçando-se linhas orientadoras de trabalho futuro neste campo de aplicação.

1.4 Referências bibliográficas

ANSR. (2009). "Estratégia Nacional de Segurança Rodoviária 2008-2015." Diário da República 1ª série-nº122 - 26 de Junho de 2009, 4160-4188.

Australian Transport Safety Bureau (2001). *The National Road Safety Strategy 2002-2010*. Australian Transport Council.

ECRS,

http://ec.europa.eu/transport/road_safety/specialist/statistics/care_reports_graphics/index_en.htm, acedido em Dezembro de 2009.

ETSC (2003a). *Assessing risk and setting-targets in transport safety programmes*. Brussels: European Transport safety Council.

ETSC (2003b). *Transport Safety Performance in the EU. A statistical overview*. Brussels: European Transport Safety Council.

GRSP, www.grsroadsafety.org, Acedido em Dezembro 2009.

Ministério da Administração Interna (2003). *Plano Nacional de Prevenção Rodoviária*. Ministério da Administração Interna.

Road Transport Research (1998). *Safety of vulnerable road users*. Paris: Organisation for Economic Co-Operation and Development, DSTI/DOT/RTR/RS7(98)1/FINAL.

2. O PLANEAMENTO DO SISTEMA DE TRANSPORTES

O desempenho do sistema de transportes é um factor primordial na economia e no desenvolvimento social de uma área metropolitana. Um sistema de transportes não só proporciona oportunidades para a mobilidade de pessoas e bens como também influencia, a longo prazo, o padrão de crescimento e o nível de actividades económicas do território servido (Meyer, M. D. and Miller, E. J., 2001).

Nas últimas décadas o sistema de transportes em meio urbano tem merecido uma crescente atenção política motivada pela necessidade de melhorar a qualidade do ar, viabilizar os centros de actividades económicas, permitir a mobilidade de grupos sociais específicos (como por exemplo, os idosos) e promover o desenvolvimento sustentável da comunidade. Por estes factos, o planeamento do sistema de transportes é uma actividade importante que permite antecipar as consequências de determinada(s) acção(ões) avaliando os impactes daí decorrentes. O processo de planeamento deve, no entanto, ser visto como uma base para o processo de tomada de decisão tal como referido em (Meyer, M. D. and Miller, E. J., 2001):

“...to be effective, planning must be an integral and ongoing part of the decision-making process.”

Ou seja, o produto final do processo de planeamento deve servir como objecto de comunicação para os decisores dando informação útil acerca dos impactes na área, identificando diferentes alternativas de actuação e desenvolvendo estratégias de implementação.

Os critérios de escolha dos impactes a serem avaliados devem ser ajustados a factores identificados como consequências directas ou indirectas do sistema de transportes e ter em conta a gravidade social e económica dessas consequências.

Nesse sentido, a sinistralidade há já várias décadas que foi identificada como uma consequência nefasta para a sociedade. Por essa razão, não é aceitável que na actualidade se planeie uma acção no sistema de transportes sem avaliar os respectivos impactes na sinistralidade.

Esta realidade, provavelmente, é consequência da não existência de ferramentas de avaliação da sinistralidade em cenários futuros, assim como o desconhecimento acerca da forma de incluir essas ferramentas no processo habitual de decisão na fase de planeamento e tomada de decisão.

Para desenvolver uma ferramenta que contribua para a integração da segurança rodoviária (SR) no planeamento é necessário entender os processos de planeamento de transportes e de decisão acerca destes.

Apresenta-se a seguir, em 2.1 e 2.2, a descrição do processo de planeamento e de tomada de decisão, respectivamente, baseada em pesquisa bibliográfica e referindo apenas o que se julga necessário conhecer para a integração da segurança rodoviária. Embora se apresente estes dois processos separadamente, na prática eles estão interligados. Apresenta-se ainda, em 2.3, uma breve descrição dos modelos de transporte por ser uma ferramenta frequentemente utilizada e eventualmente integrável com a análise da sinistralidade. Por fim, em 2.4, faz-se referência às características básicas de uma estrutura organizacional responsável pelo processo de planeamento.

2.1 O processo de planeamento do sistema de transportes em meio urbano

O planeamento urbano dos transportes é o processo com o qual se (Meyer, M. D. and Miller, E. J., 2001):

- Estabelecem objectivos gerais para o meio urbano;
- Analisa o tipo de decisões a tomar para se cumprir esses objectivos;
- Avaliam as oportunidades e limitações do futuro relativamente às metas e medidas de desempenho desejadas para o sistema de transportes;
- Identificam as consequências a curto e longo prazo das várias alternativas definidas, tendo em conta as oportunidades e limitações identificadas;
- Relacionam as várias alternativas com as metas, objectivos ou medidas estabelecidas;
- Apresentam toda a informação aos decisores;

- Apoiam os decisores a estabelecer prioridades e a desenvolver programas de investimento.

Ou seja, de uma forma resumida, o processo de planeamento de transportes tem como objectivo gerar informação acerca das consequências de uma decisão (ou não decisão) e identificar acções alternativas e destas seleccionar a melhor acção. O processo de planeamento do sistema de transportes tem também um papel importante na aplicação de estratégias políticas, a partir das quais se definem quais os investimentos na infra-estrutura ou no sistema operacional de uma área.

O processo de planeamento de transportes tradicionalmente reflecte as preocupações políticas e questões da época em que se insere. Nesse sentido, por volta da década de cinquenta, os estudos na área dos transportes, geralmente realizados a uma escala macro (nível regional), focavam praticamente em exclusivo a expansão das vias rápidas respondendo assim ao rápido crescimento do automóvel particular. Nesse sentido, os técnicos da área desenvolveram várias ferramentas, conceitos e teorias para apoiar esta política de actuação no sistema de transportes. Gradualmente, a partir do fim da década de sessenta inícios da década de setenta, as perspectivas políticas foram-se modificando no sentido de equilibrar o sistema de transportes nas áreas metropolitanas havendo, no entanto, uma lacuna ao nível de ferramentas, conceitos e dados que fundamentassem o processo de planeamento de um sistema de transportes multimodal (o primeiro “*Transit Capacity Manual*” foi publicado em 1999). Ou seja, o planeamento evoluiu do estudo ao nível regional da expansão da infra-estrutura para o nível metropolitano do sistema de transportes multimodal (Meyer, M. D. and Miller, E. J., 2001). De facto, existe uma série de meios de transporte a integrar no sistema que promove a mobilidade e acessibilidade. Estes diferentes meios de transporte interagem de uma forma muitas vezes complexa sendo necessário criar um processo de coordenação dos vários modos de transportes.

A evolução do processo de planeamento de transportes está relacionada com a ocupação do solo, embora nem sempre de uma forma assumidamente integrada. De facto, na década de cinquenta foram realizados os primeiros esforços nos Estados Unidos da América (EUA) para estudar as inter-relações entre transportes e ocupação do solo, reconhecendo-se a necessidade da sua coordenação. Reconhecia-se assim um ciclo de relações em que, por um lado, a distribuição de usos do solo (áreas residenciais, comerciais, etc.) determinam as actividades realizadas (morar, ir às compras, etc.) que, por sua vez, determinam as

deslocações através do sistema de transportes. Por outro lado, a distribuição das infra-estruturas que suporta o sistema de transportes criam acessibilidades que, por sua vez, determinam as decisões de localização resultando em alterações na ocupação do solo. Veja-se o caso da influência da vulgarização do transporte individual e da expansão da rede rodoviária na descentralização da habitação e posteriormente do emprego, gerando deslocações de longas distâncias e movimentos mais complexos. Incrementou-se, assim a suburbanização e o declínio urbano nos centros das cidades. Contudo, nas últimas décadas tem-se desenvolvido uma consciência de que o crescimento do tráfego não é sustentável. Nesse sentido, procurou-se desenvolver novos sistemas de transportes públicos criando-se progressivamente algumas limitações ao uso indiscriminado do automóvel dentro das áreas urbanas.

Desta forma, o processo de planeamento tem um papel fundamental a desempenhar na implementação de políticas sustentáveis sendo reconhecida a importância da utilização de políticas integradas entre planeamento da ocupação do solo e transportes como forma de alcançar o desenvolvimento sustentável (Ramos, P. C. R., 2001).

Sendo este processo de planeamento bastante complexo e exigente tornou-se necessário desenvolver e aplicar ferramentas destacando-se os Sistemas de Informação Geográfica (SIG) e os modelos de simulação de tráfego. Os SIGs facilitam o planeamento eficiente da rede de transportes através da análise de diferentes tipos de dados. Os modelos de simulação de tráfego permitem ao planeador simular o movimento dos veículos na rede até à unidade de um veículo.

Contudo, o processo de planeamento não se baseia unicamente na utilização de técnicas. Durante décadas o processo de planeamento baseou-se numa abordagem racional composta por cinco etapas, tal como mostra a Figura 2.1 (Meyer, M. D. and Miller, E. J., 2001).

Definir objectivos é algo que se pratica há já algum tempo no planeamento de transportes em diversos países. Os objectivos podem ser o garante da aplicação do conceito de sustentabilidade urbana através da definição de um sistema de transportes e de uma ocupação do solo que assenta nas seguintes características fundamentais, (May, A. D., 2003):

- Eficiência económica;
- Protecção do ambiente natural;
- Ruas e zonas envolventes com vivência agradável;
- Segurança;

- Equidade e inserção social;
- Contribuição para o crescimento económico;
- Equidade entre as gerações.

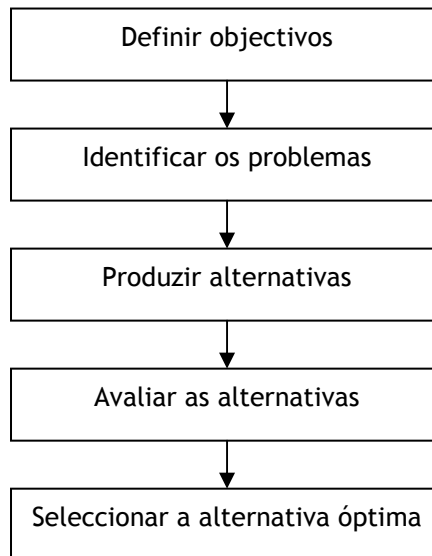


Figura 2.1 - Etapas do processo de planeamento (Fonte: (Meyer, M. D. and Miller, E. J., 2001))

Para garantir a aplicação deste conceito é necessário definir objectivos em várias áreas de actuação que envolvem diferentes actores e decisores. Para além do planeador e decisor existe outro tipo de intervenientes tais como empresas privadas, associações comunitárias e utilizadores viários que podem ter um papel relevante no sistema, devendo estes intervir sempre que possível na definição dos objectivos.

Contudo, os governantes são os intervenientes com maior poder de actuação, pois podem intervir na infra-estrutura bem como no sistema operacional de forma a garantir a mobilidade e a acessibilidade. São estes decisores políticos que, em geral, definem os objectivos e que também podem incluir os interesses dos diversos intervenientes através do processo de participação pública. Habitualmente, neste processo surgem conflitos devido a interesses distintos e muitas vezes incompatíveis. Para poder gerir esses conflitos e coordenar os diferentes interesses os decisores políticos terão de definir prioridades. Neste caso, o processo de planeamento pode ajudar a clarificar e a justificar essas prioridades.

Como os objectivos se traduzem geralmente por conceitos abstractos, torna-se difícil medir o desempenho das medidas implementadas. Neste sentido, May, A. D. (2003) apresenta indicadores de desempenho a partir dos quais é possível quantificar os objectivos e assim medir a eficiência do sistema de transportes. Os indicadores definidos são de três tipos:

- “*Outcome*” – indicador que avalia o resultado de uma determinada estratégia ou acção;
- “*Input*” – indicador que quantifica as medidas implementadas;
- “*Process*” – indicador que descreve a forma como o sistema está a responder.

Por exemplo, tendo definido como objectivo o aumento da segurança rodoviária, os acidentes serão os indicadores de “*outcome*”. Os indicadores de “*input*” podem ser o número de medidas de tratamento, tais como, colocação de passagens para peões, separador central, etc. ou o valor monetário dispendido nestas medidas. Os indicadores de “*process*”, neste caso, poderiam ser a velocidade dos veículos e o volume de peões e de veículos. Por este exemplo, depreende-se que os indicadores de “*input*” e de “*process*” não medem o impacto relativamente aos objectivos definidos mas podem ser um apoio para o estudo dos efeitos das medidas aplicadas para implementação de determinada estratégia.

Os objectivos podem ainda ser convertidos em valores fixos a atingir designando-se, pela literatura inglesa, de “*target*” e traduzido neste documento por “meta”. Para definir uma meta é importante ponderar o que esse valor representa, podendo também nesta fase recorrer-se ao processo de planeamento para avaliar as implicações de determinado valor. As vantagens de especificar uma meta são que esta torna claro se os objectivos foram ou não atingidos. Estas metas, que medem directamente os objectivos, são definidas para os indicadores de “*outcome*”. Contudo, também se podem definir metas para os indicadores de “*input*” e de “*process*” definindo, por exemplo, um orçamento a distribuir para um determinado número de medidas com vista a melhorar a segurança rodoviária, indicando um valor percentual de veículos que devem reduzir a velocidade ou um volume de tráfego a atingir.

Contudo, apesar das suas evidentes vantagens existem alguns riscos ao definir metas. Se estas são definidas apenas para alguns objectivos corre-se o risco de se estar a dar ênfase a esses objectivos em detrimento de outras áreas não incluídas, eventualmente, por não serem fáceis de quantificar ou por não serem politicamente relevantes. Por outro lado, se se atribuir metas a todos os objectivos também se pode estar a dar indicações erradas da

importância desses objectivos. Além disso, não é fácil defini-las de uma forma apropriada e coerente. Reconhece-se, no entanto, que são uma peça fundamental para orientar e estimular a actuação assim como a monitorização do desempenho.

Wong, S. C., *et al.* (2006) realizaram um estudo em que se analisou a relação entre a atribuição de metas para a segurança rodoviária e a redução do número de mortos. O estudo considerou dados de catorze países entre 1981 e 1999 que definiram metas para a segurança rodoviária. Os resultados desse estudo revelaram que a maioria dos países que estabeleceram metas diminuíram o número de vítimas mortais, tendo, por isso, os autores concluído que a definição de metas está fortemente associada a uma melhoria da SR.

Segundo May, A. D. (2003), as metas devem ser definidas seguindo o processo indicado na Figura 2.2.

Após a definição dos objectivos, e partindo destes, a etapa seguinte do processo de planeamento corresponde à identificação dos problemas físicos e operacionais da rede. Dessa análise resulta a proposta de uma ou mais alternativas de intervenção na rede. Para se seleccionar qual a melhor alternativa é necessário proceder a uma avaliação daquelas, recorrendo, em geral, aos seguintes métodos de avaliação (May, A. D., 2003) :

- Estruturas descritivas;
- Análise Custo-Benefício (ACB);
- Análise Multi-Critério (AMC).

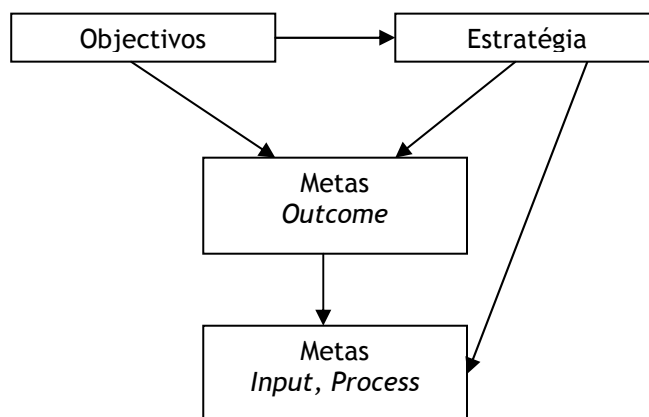


Figura 2.2 - Processo de definição de metas (Fonte: (May, A. D., 2003))

O método das estruturas descritivas corresponde a uma avaliação qualitativa cujos resultados não são convertidos em unidades monetárias nem se atribui pesos. Como

exemplo de um método de avaliação com base numa estrutura descritiva observe-se o Quadro 2.1. Neste quadro organiza-se a informação relativa a vários indicadores para diferentes intervenções.

Quadro 2.1 - Estrutura descritiva de avaliação (Fonte: (May, A. D., 2003))

Objectivos/Indicadores	Sem-intervenção	Intervenção 1	Intervenção 2	Etc.
Eficiência				
Ambiente				
Segurança				
Etc.				

No entanto, este método de avaliação nem sempre é compatível com a complexidade do meio urbano e a existência de várias variáveis muitas vezes inter-relacionadas, sendo, por isso, necessário recorrer à aplicação de técnicas mais sofisticadas e completas, tais como, a ACB e a AMC.

A ACB surgiu no final da década de sessenta a inícios da década de setenta e actualmente inclui as seguintes fases (Grant-Muller, S. M., et al., 2001):

- Identificar os impactes a serem incluídos;
- Especificar como é que os impactes incluídos devem ser formalmente definidos e medidos;
- Estimar os impactes;
- Atribuir um valor monetário à unidade do impacte estimado com base na valorização social e a aceitabilidade de custear determinados princípios ou valores;
- Definir o período temporal para o qual são determinados os custos, os benefícios e o valor actual líquido³ (VAL).

Esta avaliação caracteriza-se por comparar as diferentes alternativas com base na unidade monetária. Isto é, os impactes nos tempos de viagens, nos acidentes, no ambiente, etc. são

³ *net present value*, em inglês

convertidos em valores monetários. Os custos e os benefícios são determinados para cada ano futuro relativamente à alternativa de não intervir. Os benefícios líquidos são determinados em relação ao momento actual e, através do somatório destes valores para o período de avaliação, calcula-se o VAL do investimento. A melhor solução será aquela a que corresponde o maior valor do benefício líquido⁴.

As desvantagens identificadas nesta técnica são a dificuldade em atribuir um valor monetário a indicadores do tipo acidentes, ruído, etc., e o facto do valor final ser determinante na decisão, não se encorajando a discussão nem a valorização da participação pública (May, A. D., 2003).

A AMC surge para colmatar as desvantagens da ACB possibilitando aos decisores a atribuição de pesos aos vários critérios de avaliação. Este método de avaliação inclui as seguintes fases (Grant-Muller, S. M., et al., 2001):

- Identificar e definir os impactes a serem incluídos;
- Especificar o método de medição dos impactes e o processo de atribuição de uma pontuação;
- Determinar os pesos;
- Estabelecer um método de combinação dos pesos e pontuação dos impactes de forma a chegar a um valor para cada alternativa.

Através dos pesos reflecte-se a importância atribuída aos diferentes impactes, os quais podem ser definidos em função dos objectivos nacionais considerados para o sector dos transportes. A forma mais simples de utilizar esta técnica é determinar um único valor final para cada alternativa tal como é descrito em Sayers, T. M., *et al.* (2003). Com esta técnica também é possível determinar qual a alternativa mais estável às variações de pesos ou qual a alternativa menos provável de ser rejeitada por outro grupo de decisores que tenha um ponto de vista diferente, isto é, que atribuiria pesos diferentes, designando-se esta técnica de análise de sensibilidade (May, A. D., 2003).

Em Bristow, A. L. and Nellthorp, J. (2000) é apresentada uma tabela (ver Figura 2.3) que mostra a prática corrente, em alguns países da Europa (incluindo Portugal), na utilização de métodos de avaliação e determinação dos impactes. Os impactes indicados nesta tabela são de três tipos: impactes directos (onde se inclui a segurança); impactes ambientais; impactes sócio-económicos. Nos quatro países onde a AMC é utilizada esta aparece

⁴ *net benefit*, em inglês

integrada com uma ACB. Segundo esta tabela, todos os países analisados utilizam a ACB em pelo menos um sector dos transportes. Verifica-se também que, em vários países, como seja o caso de Portugal, a ACB é complementada com análises qualitativas ou quantitativas para incluir outros impactes que não foram convertidos em valor monetário por razões técnicas ou políticas. Este estudo refere ainda que estas metodologias de avaliação são utilizadas para diversos tipos de decisões como sejam: a atribuição de prioridades aos projectos de um programa; a escolha entre soluções alternativas; a avaliação económica se uma iniciativa privada representa ou não benefícios sociais; a escolha do momento ideal para aplicar determinado investimento.

	AUS	BEL	DEN	FIN	FRA	GER	GRE	IRL	ITA	NRL	POR	SPA	SWE	UK
	Road	Road	Road				Road						Road	Road
DIRECT IMPACTS														
Capital														
Construction Costs	MCA	MCA					MCA			MCA				
Disruption Costs		MCA					MCA							
Land and Property Costs		MCA					MCA							
Recurring														
Maintenance Costs	MCA	MCA					MCA			MCA				
Operating Costs							MCA			MCA				
Vehicle Operating Costs	MCA	MCA					MCA			MCA				
Revenues	MCA						MCA							
Passenger Cost Savings							MCA							
Time Savings	MCA	MCA					MCA			MCA				
Safety	MCA	MCA					MCA			MCA				
Service Level	MCA						MCA							
Information							MCA							
Enforcement							MCA							
Financing / Taxation							MCA							
ENVIRONMENTAL IMPACTS														
Noise	MCA	MCA					MCA			MCA				
Vibration		MCA					MCA							
Air Pollution - Local	MCA	MCA					MCA			MCA				
Air Pollution - Global	MCA	MCA					MCA			MCA				
Severance	MCA						MCA							
Visual Intrusion							MCA							
Loss of important Sites		MCA					MCA							
Resource Consumption							MCA							
Landscape	MCA						MCA							
Ground / Water Pollution	MCA						MCA							
SOCIO-ECONOMIC IMPACTS														
Land Use	MCA						MCA			MCA				
Economic Development	MCA	MCA					MCA			MCA				
Employment		MCA					MCA							
Economic & Social Cohesion							MCA							
International Traffic							MCA							
Interoperability							MCA							
Regional Policy		MCA					MCA			MCA				
Conformity to Sector Plans		MCA					MCA							
Peripherality/Distribution							MCA							

Key: ■ CBA (Monetised Impacts) ■ Measured impacts □ Qualitative Assessment ■ MCA - included in Multi-Criteria Analysis

Figura 2.3 - Métodos de avaliação e medição dos impactes em diferentes países (Fonte: (Bristow, A. L. and Nellthorp, J., 2000))

É importante referir que o processo de avaliação está cada vez mais exigente. O número de indicadores a considerar na avaliação aumentou e estes são mais complexos. Os impactes ambientais e socioeconómicos são cada vez mais valorizados. Se inicialmente o processo de avaliação se cingia ao nível do projecto, geralmente financiado pelo governo e centrado apenas num modo de transporte, actualmente o processo centra-se, sobretudo, ao nível do planeamento de uma área ou corredor e são consideradas as interações entre os vários modos de transporte (Bristow, A. L. and Nellthorp, J., 2000).

Como já foi referido, no processo de planeamento é necessário definir o período temporal a que este se deve referir. Segundo May, A. D. (2003), idealmente o processo deveria incluir

um planeamento a curto prazo (aproximadamente 5 anos) e a longo prazo (entre 15 a 20 anos) e ainda um mecanismo de relação entre estes dois planeamentos. O planeamento a longo prazo representa a direcção e estratégia política assumida pressupondo uma continuidade, embora dentro de um contexto de tomadas de decisões susceptíveis de reavaliações periódicas. O planeamento a curto prazo deve ter em conta as necessidades imediatas do funcionamento do sistema de transportes, centradas essencialmente nas mudanças operacionais e na gestão da procura. As decisões a curto prazo podem, no entanto, influenciar o desempenho futuro do sistema de transportes. É, por isso, necessário estabelecer um mecanismo que avalie as consequências a longo prazo no sistema de transportes devido às decisões de curto prazo.

Em Portugal foi, recentemente, publicado o Plano Estratégico de Transportes (PET) (MOPTC, 2009). Este plano inscreve-se no esforço nacional no sentido de se caminhar para o desenvolvimento sustentável. Nesse sentido, o PET define objectivos a atingir no ano horizonte (2020) e um conjunto de orientações, acções e projectos que viabilizarão o atingir os objectivos definidos.

2.2 O processo de tomada de decisão

O processo de planeamento de transportes começou a expandir-se em meados da década de cinquenta e inícios da década de sessenta com a aplicação de diversos estudos de transportes em meio urbano. Ainda na década de sessenta, o processo de planeamento foi formalizado em directivas governamentais baseando-se numa premissa de selecção puramente racional, isto é, a escolha do cenário futuro baseava-se apenas em otimizar indicadores de desempenho através da utilização de técnicas de análise e modelos informáticos (Meyer, M. D. and Miller, E. J., 2001).

Durante várias décadas este modelo serviu de base para o planeamento de sistemas de transportes nos EUA e no Reino Unido (RU). No entanto, destas experiências resultaram muitos casos negativos que foram sendo identificados, destacando-se, por exemplo, o planeamento a longo prazo baseado unicamente no crescimento exponencial do uso do veículo privado sem prestar atenção a outros meios de transporte e à interacção entre eles.

Desde então as abordagens de tomada de decisões têm-se vindo a diversificar, podendo-se estratificar em vários modelos conceptuais. Pela bibliografia é possível identificar vários modelos conceptuais de diferentes formas em função dos autores. No caso de May, A. D. (2003), os modelos conceptuais são simplificadamente agrupados em três:

- “*Vision-led*” – modelo que envolve apenas um decisor (normalmente o governante eleito) que define a estratégia política para a sua cidade, assim como as medidas necessárias à sua implementação. Este era o modelo mais utilizado no passado.
- “*Plan-led*” – modelo que compreende a identificação do problema e a definição de objectivos. Através de um procedimento normativo selecciona de entre várias alternativas a melhor. Este modelo surge diversas vezes no seguimento do modelo anterior.
- “*Consensus-led*” – modelo que se baseia num processo de discussão até se atingir um entendimento entre os vários decisores envolvidos.

Outros autores, tais como, Meyer, M. D. and Miller, E. J. (2001) e Ortúzar, J. and Willumsen, L. G. (2001), diferenciam o processo de tomada de decisão segundo cinco e seis modelos conceptuais, respectivamente, estando a diferença, nalguns casos, apenas na designação. Descrevem-se a seguir os modelos conceptuais segundo Ortúzar, J. and Willumsen, L. G. (2001).

Plano Director⁵

Existe uma longa tradição no desenvolvimento e aplicação deste modelo no planeamento dos transportes e da ocupação do solo. Neste modelo as decisões são baseadas em interpretações do plano director que estabelece as regras e o desempenho desejável. O plano director é preparado com cuidado com vista a direccionar o futuro do sistema de transportes, normalmente através da utilização de um dispendioso e único exercício de modelação de transportes do tipo implementado nas décadas de sessenta e setenta. Existe uma longa tradição no desenvolvimento e aplicação deste tipo de modelo no planeamento de transportes e de ocupação do solo.

O plano director é razoável de se aplicar quando a envolvente é estável, isto é, o tipo de problemas mantém-se constante. Tem a vantagem de informar a população do que é que vai ser realizado. A maior desvantagem corresponde à inadaptação do plano a rápidas

⁵ *master plans*, em inglês

alterações verificadas ao nível do ambiente económico, social e tecnológico, ou seja, a actualização da informação não é considerada no plano director sendo raramente recolhida ou utilizada para a tomada de decisão (não é realizada a monitorização do sistema). Alguns países desenvolvidos têm optado por este tipo de abordagem sendo, no entanto, inapropriada para as constantes e rápidas alterações da envolvente que ocorrem nos nossos dias.

Modelo Racional⁶

O modelo racional de decisão é o mais referido nos artigos sobre planeamento de transportes. Neste, a quantificação é essencial. O processo de decisão compreende a escolha entre várias soluções alternativas tendo por base a estimação do funcionamento operacional do sistema e a quantificação, em cada alternativa, dos benefícios e custos, e ainda outros critérios como a protecção do meio ambiente, a segurança, etc. Neste modelo a modelação de transportes é uma ferramenta central.

Em alguns casos é possível transformar a decisão numa estrutura matemática de programação pressupondo-se para isso que a função objectivo esteja bem definida, assim como as restrições a observar. Contudo, para a maior parte dos problemas reais, alguns elementos da função objectivo e restrições são difíceis de quantificar ou de converter em unidades de medição comuns, tais como, moeda ou tempo. Poderá ainda ser igualmente difícil incluir em alguns casos elementos probabilísticos. No entanto, neste processo recorre-se a uma grande quantidade de informação e conhecimento.

Na aplicação desta abordagem podem resultar alguns problemas, tais como a falta de sensibilidade para os interesses da população, o elevado custo e o distanciamento dos decisores relativamente à abordagem analítica do problema.

Este modelo tem falhado na obtenção de resultados em tempo e com precisão aceitáveis. A comprovar, refira-se a reacção adversa que se verificou nos anos 80's aos resultados da modelação de transportes.

⁶ *normative decision theory or substantive rationality*, em inglês

Teoria Comportamental⁷

Este modelo surgiu na tentativa de corrigir o modelo anterior, reconhecendo que na maior parte das vezes os decisores devem tentar obter a satisfação de um número elevado de utilizadores. Neste sentido, o processo termina quando é encontrada uma solução aceitável. O modelo baseia-se em estudos e conhecimentos que permitem chegar a uma solução que não é mais do que um aperfeiçoamento marginal da prática corrente. O modelo é, por isso, descrito como uma análise marginal da optimização do problema, iniciando o processo num cenário não-ótimo. Os aperfeiçoamentos, embora pequenos, são explorados e desenvolvidos. A modelação, neste caso, tem um papel mais restrito sendo, geralmente, bastante simples.

Decisão de Grupo⁸

Esta é uma abordagem comum em várias áreas onde predomina a organização de comissões. A tomada de decisão inicia-se por um processo de aprendizagem dentro de um grupo com autoridade de decisão e funções específicas. Os contributos individuais e do grupo relativamente aos seus conhecimentos e especialidades são canalizados para a decisão. Este modelo baseia-se em informação quantitativa e qualitativa mas não de uma forma sistemática. A capacidade persuasiva e o poder dos membros do grupo podem predominar em detrimento do valor intrínseco da decisão final. Por outro lado, um grupo de decisão deste tipo promove com facilidade a aceitação da decisão final, o que pode ser uma vantagem no contexto do planeamento. A modelação geralmente é orientada pelo grupo, podendo este mesmo advertir para a inclusão ou não de tópicos na tarefa de modelação, promovendo posteriormente a aceitabilidade dos resultados.

Decisão Adaptada⁹

Esta abordagem, em geral, é mais flexível que a anterior. Reconhece a interacção entre grupos não atribuindo um poder único na tomada de decisão. Cada grupo analisa o problema de uma forma distinta sendo, por isso, necessário negociar a decisão final. Este modelo é particularmente vantajoso em situações em que o problema contém várias variáveis com elevado grau de incerteza e interacções onde não é possível aplicar uma

⁷ *behavioural decision theory*, em inglês

⁸ *group decision making*, em inglês

⁹ *adaptive decision making*, em inglês

teoria normativa ou comportamental que determine relações de causa e efeito. Esta abordagem é comumente utilizada em decisões relativas a legislação e diplomacia.

Neste modelo, a modelação do sistema de transporte tem um papel instrumental reduzido. É realizado um estudo cujas recomendações são usadas como argumentos nas negociações, conferindo maior poder aos decisores.

Modelo Misto¹⁰

Este último modelo combina os vários modelos referidos anteriormente numa estratégia flexível. O modelo baseia-se na análise, persuasão, acordos e estratégias políticas em diferentes campos de acção e com diferentes objectivos. Os objectivos são, em geral, fixados e conhecidos mas podem ser alargados durante o processo de negociação de forma a incluir os interesses das partes em conflito.

É uma abordagem realística que aceita que, para problemas importantes, os objectivos e os campos de acção possam variar como parte do processo de decisão. A modelação de transportes normalmente desempenha um papel importante nesta abordagem.

Este modelo caracteriza-se pela flexibilidade e capacidade de adaptação, que se reflecte na inserção de novas variáveis e na célere análise de políticas e projectos inovadores.

Se analisarmos as decisões que têm sido tomadas nas últimas décadas ao nível nacional e internacional é difícil caracterizá-las por um único modelo conceptual. Em Portugal, o Estado mantém-se como o principal providenciador de infra-estruturas de transportes e gere através de empresas controladas uma parte importante da oferta de transporte público de passageiros e actua como regulador da actividade transportadora (MOPTC, 2009). O processo de decisão do sistema de transportes ao nível urbano está fortemente centrado na organização municipal, burocratizada e muito centralizada no órgão executivo (Camões, P. J. S., 1995). O modelo conceptual frequentemente associado ao sistema de transportes é o Plano Director Municipal em que a rede viária é coordenada com a ocupação do solo.

¹⁰ *mixed-mode decision-making strategies*, em inglês

2.3 Modelos de transporte

Como se pode concluir pelo atrás referido, em 2.2, em praticamente todos os modelos conceptuais de tomada de decisão, a modelação de transportes é incluída como uma ferramenta, diferindo apenas no grau de importância que lhe é atribuído.

Um modelo é uma representação matemática da realidade. No caso do modelo de transportes o objectivo é representar as diversas componentes da infra-estrutura de transporte (rede viária, linhas de transporte público, gestão e controlo de tráfego) e os fluxos de tráfego (utilizadores e veículos) de forma a reproduzir as condições operacionais do sistema viário. Os modelos são geralmente sequenciais, envolvendo uma série de etapas que correspondem às decisões dos utilizadores, sendo para tal necessário definir os vários pressupostos sobre como são tomadas as decisões (Tavares, J. P., 2003). O conjunto destes sub-modelos sequenciais, designa-se habitualmente de “Modelo de 4 Passos” (ver Figura 2.4).

Para a aplicação do modelo de transportes é necessário que a área em estudo seja desagregada em áreas mais pequenas de forma a corresponder a áreas de geração e atracção de viagens, já que é impraticável considerar a multiplicidade de viagens e origens/destinos existentes num meio urbano. Estas áreas, definidas considerando uma certa homogeneidade relativamente a actividades relacionadas com os transportes (ocupação do solo, infra-estrutura, etc.), são designadas de zonas e o respectivo centro de actividade, onde, supostamente, se iniciam e terminam as viagens, de centróide (Tavares, J. P., 2003). Assim, estas zonas são definidas tendo em conta vários critérios (Meyer, M. D. and Miller, E. J., 2001):

- Minimizar o número de viagens intrazonais;
- Atender às fronteiras físicas, políticas, jurisdicionais e históricas;
- Assegurar, para cada zona populacional, características socioeconómicas homogéneas;
- Considerar apenas zonas ligadas e evitar zonas que estejam completamente contidas noutras;
- Considerar, na medida do possível, um sistema zonal em que os valores do número de habitantes, área, viagens geradas e atraídas são aproximadamente iguais;

- Conjugar, na medida do possível, as fronteiras zonais com as zonas consideradas para os censos.

A escolha da dimensão das zonas é importante para a confiabilidade do modelo e ajustamento deste à realidade: se as zonas são de grande dimensão, a exactidão dos valores estimados é pequena para o fluxo de tráfego embora a matriz de viagens seja mais confiável; se as zonas forem de pequena dimensão, a confiança na matriz baixa para além de exigir recursos eventualmente incompatíveis (Tavares, J. P., 2003).

A aplicação do modelo é feita numa representação da infra-estrutura de transportes designada de rede. A rede é constituída por arcos que representam os segmentos de vias rodoviárias, por nós que identificam as intersecções e pelos centróides que se encontram ligados à rede pelos respectivos conectores (Tavares, J. P., 2003). Estes arcos e nós são caracterizados física e operacionalmente.

Com esta representação da rede é possível simular as viagens com início na zona de geração e término na zona de atracção. Esta simulação é executada através do denominado “Modelo de 4 Passos”¹¹. Este modelo está representado na Figura 2.4 e compreende as seguintes etapas (Tavares, J. P., 2003):

1. Geração de viagens – corresponde à determinação do número total de viagens em cada origem e em cada destino, para cada uma das zonas modeladas;
2. Distribuição de viagens – corresponde à estimação do número de viagens entre cada par de zonas;
3. Escolha modal – corresponde à repartição de viagens entre cada par de zonas pelos modos alternativos de transporte;
4. Afectação de tráfego – corresponde à atribuição das viagens entre cada par de zonas a um ou mais percursos, utilizando o modo de transporte respectivo da rede.

¹¹ *four-stage-model*, em inglês

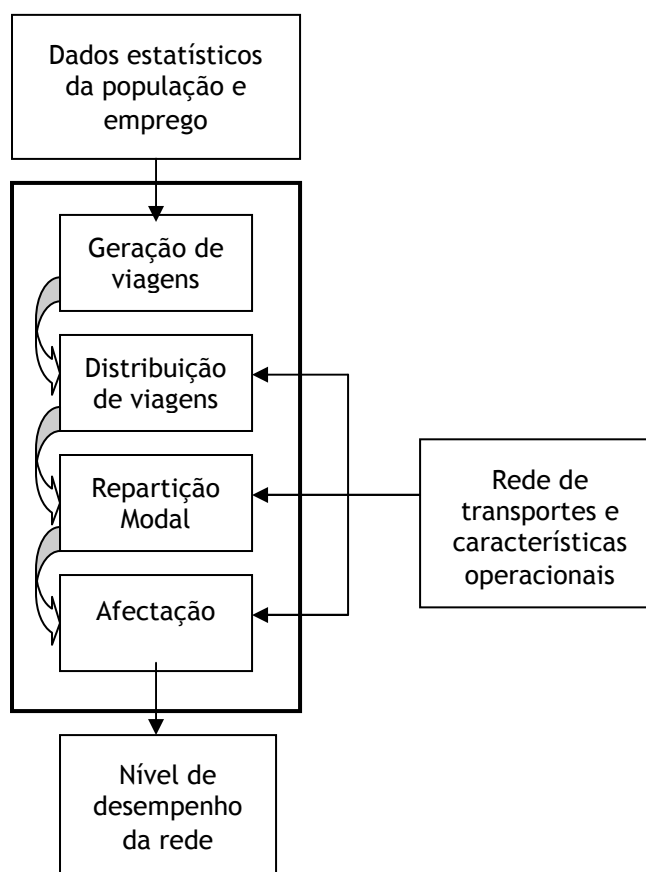


Figura 2.4 - Estrutura do modelo de 4 passos (Fonte: (Tavares, J. P., 2003))

Para o primeiro passo do modelo de transportes – geração e atracção de viagens, é necessária informação relativa ao uso do solo e dados socioeconómicos, de forma a estimar o número de pessoas que irão realizar deslocações de e para cada zona, ao longo de um determinado período de tempo (Tavares, J. P., 2003). Nesta primeira etapa são determinadas as zonas de Origem e Destino bem como o número de viagens entre estas zonas. No segundo passo do modelo – distribuição das viagens, procede-se à ligação entre estas zonas formando pares de viagens Origem-Destino (O-D). O terceiro passo consiste em repartir as viagens O-D pelos diferentes modos de transporte. A utilização de cada modo de transporte é determinada através da comparação da atractividade de cada modo de viajar. Finalmente, no último passo do modelo afecta-se o tráfego aos arcos através da modelação das decisões dos utilizadores do sistema de transporte na escolha do percurso. Nesta fase reproduz-se o nível de desempenho da rede com base em modelos de simulação. Esse nível de desempenho é caracterizado por (Tavares, J. P., 2003):

- volumes de tráfego, velocidades e atrasos;
- variáveis agregadas da rede (velocidade média de operação, atraso total, etc.);
- tempos de viagens entre zonas;
- arcos congestionados;
- percursos entre pares de zonas;

Com maior ou menor pormenorização a caracterização do nível de desempenho da rede é o produto final da modelação dos transportes.

O grau de importância atribuído aos resultados de um modelo de transportes no processo de planeamento e de tomada de decisão está relacionado com a disponibilidade dos dados necessários, do estado da arte da modelação e dos recursos existentes.

Antes da escolha da estrutura do modelo de transportes a aplicar é necessário identificar o modelo conceptual de tomada de decisão adoptado no país pelo governo ou pela unidade de decisão, tal como se descreve anteriormente em 2.2.

2.4 Estrutura organizacional

Para gerir coordenadamente o processo de planeamento e de tomada de decisão tem de existir uma base organizacional estável e abrangente de forma a integrar os diferentes interesses num sistema de transportes multimodal. Essa estrutura organizacional pode evitar a dispersão da responsabilidade que limita a capacidade de definir estratégias. Como são vários os intervenientes no sistema de transportes, e geralmente agem apenas em função dos seus interesses resultando daqui conflitos, é fundamental a estrutura organizacional ter capacidade de gerir esses conflitos. Para tal, deve basear-se em informação técnica que justifique as prioridades na acção com vista a determinado resultado. Esta estrutura terá ainda que satisfazer os objectivos políticos a alcançar, tais como políticas de ambiente e segurança.

De acordo com Meyer, M. D. and Miller, E. J. (2001), a estrutura organizacional de planeamento numa área metropolitana com um número de habitantes superior a 50 000, deve corresponder a cinco funções centrais:

1. Instituir e gerir um nível de acção multimodal e intergovernamental no processo de decisão numa área metropolitana;
2. Desenvolver, adoptar e actualizar um plano a longo prazo de avaliação do desempenho do sistema de transportes multimodal centrado em três tipos de desempenho: mobilidade e acessibilidade para pessoas e bens, manutenção e operacionalidade do sistema e qualidade de vida;
3. Desenvolver e acompanhar de uma forma contínua um programa analítico de avaliação das alternativas de transporte que suporte as decisões para a área metropolitana, e que deve ser adequado ao tamanho, complexidade da região e natureza dos transportes com vista a opções praticáveis e realísticas;
4. Desenvolver e acompanhar sistematicamente um programa multifacetado projectado para abranger todos os objectivos do plano do sistema de transportes elaborado;
5. Desenvolver e acompanhar um programa de envolvimento público proactivo que inclua o público em geral e a todos os subgrupos significativamente afectados.

Num sistema complexo e abrangente como é o sistema de transportes, muitas barreiras poderão surgir ao funcionamento de uma organização deste tipo. Essas barreiras devem ser identificadas e caracterizadas para ser possível ultrapassá-las. Segundo Meyer, M. D. *and* Miller, E. J. (2001), estas podem ser agrupadas em três tipos:

- Organizacionais
- Inter-jurisdicionais
- Meios

As primeiras são, por exemplo, a existência de várias organizações locais e estatais e legislação estatal que limite a flexibilidade de actuação da organização.

As barreiras inter-jurisdicionais correspondem à relutância de partilhar poder e autoridade assim como de formar parcerias. Por exemplo, a pouca integração entre as decisões sobre a ocupação do solo e as de transporte podem condicionar a actuação da organização.

Finalmente, podem ocorrer falhas importantes ao nível dos meios disponíveis, tais como, falta de meios humanos com formação e experiência no sector dos transportes, de informação disponível para análise da rede e de ferramentas adequadas à avaliação das alternativas por diferentes critérios.

É também fundamental o factor económico, isto é, o modelo de atribuição de fundos económicos e a sua distribuição. Este modelo económico está muito dependente da estrutura governamental e dos objectivos traçados para o sector do transporte.

Em Portugal, e ao nível do meio urbano, não existe uma instituição vocacionada para o processo de planeamento e de decisão do sistema de transportes. Existe, contudo, um quadro legislativo, desde 1990, para a implementação de Comissões Metropolitanas de Transportes (Autoridade) que controlem e coordenem o sistema de transportes nas áreas metropolitanas do Porto e Lisboa mas, apesar de terem sido implementadas, praticamente não exerceram funções até à data. Não existindo uma entidade única a gerir o sistema de transportes, as responsabilidades de actuação e a própria informação dispersam-se. As Câmaras Municipais são, talvez, as entidades para onde a maior parte da informação confluiu. São estas entidades que definem a rede urbana a par da ocupação do solo através da elaboração dos Planos Directores Municipais. É nesta fase que se decide o futuro da cidade. Este modelo corresponde ao modelo conceptual anteriormente referido e designado como Plano Director.

2.5 Conclusões

O processo de planeamento do sistema de transportes focava-se, inicialmente, na expansão da infra-estrutura rodoviária ao nível regional tendo, gradualmente evoluído para o nível metropolitano, adaptando-se a uma visão mais alargada do sistema de transportes no seu conjunto. Desde o início é reconhecida neste processo a interdependência existente entre o sistema de transportes e a ocupação do solo. E mais recentemente, com o objectivo de se alcançar um desenvolvimento sustentável, reconhece-se inclusive a necessidade de se implementar políticas integradas entre o sistema de transportes e a ocupação do solo.

O processo de planeamento do sistema de transportes é complexo e envolve um conjunto diverso de interesses. De forma a incluir os diferentes interesses, elabora-se uma estratégia alargada à qual se atribui objectivos para diversas áreas do sistema de transportes. Em Portugal, foi publicado recentemente o PET onde se perspectiva de uma forma integrada cada sector de actividade. Geralmente, aos objectivos considerados na estratégia, são designadas metas para as quais é necessário definir um conjunto de acções.

Para avaliar as consequências das diferentes alternativas de actuação no sistema de transportes é necessário identificar e determinar os impactes. As diferentes alternativas propostas nesta fase podem ser avaliadas através de um ou mais métodos de avaliação, geralmente uma análise de custo-benefício, uma análise multi-critério e uma estrutura descritiva.

Para avaliar os impactes recorre-se, geralmente, a ferramentas tais como a modelação de transportes. O modelo habitualmente considerado corresponde ao “Modelo de 4 Passos” que permite caracterizar o nível de desempenho da rede.

O propósito do processo de planeamento é gerar informação útil aos decisores, nomeadamente, acerca dos impactes das diferentes alternativas de actuação no sistema de transportes. A forma como esta informação é considerada depende do modelo conceptual utilizado no processo de decisão. Estes modelos são definidos e designados de forma distinta em função dos autores.

A gerir o processo de planeamento e de decisão, deve existir uma estrutura organizacional capaz de coordenar os vários interesses e de implementar a estratégia definida para o sistema de transportes.

2.6 Referências bibliográficas

Bristow, A. L. and Nellthorp, J. (2000). Transport project appraisal in the European Union. *Transport Policy*, 7, 51-60.

Camões, P. J. S. (1995) *Processos de tomada de decisão na organização municipal. Estudo exploratório de um caso*. Dissertação para obtenção do grau de mestre, Escola de Economia e Gestão, Universidade do Minho, Braga.

Grant-Muller, S. M., et al. (2001). Economic appraisal of European transport projects: the state-of-the-art revisited. *Transport Reviews*, 21, 237-261.

May, A. D. (2003). "A Decision Maker's Guidebook." PROSPECTS - Procedures for Recommending Optimal Sustainable Planning of European City Transport Systems,

Institute of Transport Studies, University of Leeds, ed., European Commission - 5th Framework-EESD, 52.

Meyer, M. D. and Miller, E. J. (2001). *Urban Transportation Planning*, McGraw Hill.

MOPTC (2009). *Plano Estratégico de Transportes 2008-2020*. Lisboa: Ministério das Obras Públicas, Transportes e Comunicações.

Ortúzar, J. and Willumsen, L. G. (2001). *Modelling Transport*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.

Ramos, P. C. R. (2001) *Estruturação espacial, transportes e desenvolvimento sustentável da área metropolitana do Porto*. Mestrado, Faculdade de Engenharia/ Faculdade de Arquitectura, Universidade do Porto, Porto.

Sayers, T. M., Jessop, A. T. and Hills, P. J. (2003). Multi-criteria evaluation of transport options-flexible, transparent and user-friendly? *Transport Policy*, **10**, 95-105.

Tavares, J. P. (2003) *Aplicabilidade e robustez de modelos de afectação de tráfego em redes urbanas*. Tese de Doutoramento (PhD), Faculdade de Engenharia, Universidade do Porto, Porto.

Wong, S. C., et al. (2006). Association between setting quantified road safety targets and road fatality reduction. *Accident Analysis and Prevention*, **38**, 997-1005.

3. A SEGURANÇA RODOVIÁRIA NO PROCESSO DE PLANEAMENTO DO MEIO URBANO

A sinistralidade rodoviária é um problema de escala mundial que atinge principalmente jovens com idades compreendidas entre os 18 e os 20 anos (ERF, 2006). Com a crescente motorização de países como a China, a Índia e alguns países africanos não se perspectiva que o actual cenário se reverta. A crescente motorização é normalmente acompanhada por um aumento da construção viária que, devido às taxas de crescimento económico elevadas destes países, poderá ocorrer de uma forma desenfreada, ou seja, sem um planeamento ou projecto que inclua, entre outros, a segurança rodoviária.

Os diversos esforços empreendidos nas últimas décadas nos países da União Europeia (UE) têm-se verificado, em geral, eficazes na diminuição do número de mortos e feridos. No relatório de 2006, realizado pela *European Union Road Federation* (ERF), é possível consultar uma série de tabelas e gráficos que de uma forma geral reflectem um decréscimo do número de mortos e feridos nos países da UE, sendo ainda previsível que alguns países atinjam o objectivo fixado para 2010, de diminuir para metade o número de mortos relativamente ao ano 2000. Em Portugal, foi aprovado em 2003 o Plano Nacional de Prevenção Rodoviária (PNPR) que estabeleceu este objectivo, bem como objectivos relativos a determinada parcela de população mais exposta. Como estes objectivos foram em geral alcançados, foi recentemente aprovada a Estratégia Nacional de Segurança Rodoviária (ENSR) 2008-2015, onde se definem novos objectivos a atingir em duas fases. Não obstante os resultados positivos é de salientar que essa redução partiu de um número elevado de mortos e feridos, continuando Portugal em 2006 a posicionar-se como um dos países com o maior número de mortos por 1 milhão de habitantes, estando acima da média da UE tal como é referido no documento “Estratégia Nacional de Segurança Rodoviária 2008-2015” (2009).

Em Portugal, o maior número de ocorrências continua a registar-se dentro das localidades (69,3% dos acidentes e 66% das vítimas) (DGV, 2006). Grande parte das actividades diárias da nossa sociedade actual é realizada no meio urbano. Em geral, as entidades locais têm como objectivo desenvolver um sistema de transportes que garanta a mobilidade e a

acessibilidade, mas muitas vezes sem considerar devidamente a segurança rodoviária e o meio ambiente. Para estas entidades poderem avaliar as consequências de uma determinada decisão é necessário recorrer a ferramentas que determinem os impactes das diferentes soluções definidas ainda em fase de planeamento. Desta forma é possível prevenir determinados efeitos negativos diminuindo consideravelmente os custos sociais e económicos. No caso da sinistralidade essa diminuição de custos advém não só da diminuição do número de acidentes, mas também da redução ou até eliminação da necessidade de implementar medidas de tratamento.

Tradicionalmente, e tal como referido no Capítulo 2, o planeamento em meio urbano é centrado na definição de níveis de ocupação do solo e da classificação hierárquica viária. Partindo desta definição, é estimada a procura nos transportes resultante da distribuição da população e do emprego. Neste processo não há uma análise das consequências ao nível da sinistralidade.

Com a rede urbana consolidada, a entidade local fica limitada à manutenção da infra-estrutura e ao melhoramento do funcionamento do sistema de transportes. Vários tipos de intervenção podem ocorrer: alteração funcional de uma via existente, introdução de um pólo gerador de tráfego, introdução de uma nova via, etc. Estas intervenções podem afectar significativamente o sistema de transportes e consequentemente a segurança rodoviária (SR). Em Portugal, e na maior parte dos países desenvolvidos, não há uma tradição, na fase de planeamento deste tipo de intervenções, em considerar a SR podendo ser, eventualmente, incluída de uma forma pontual em resposta a situações específicas e não de uma forma organizada e sistemática.

Como se pode concluir, apesar do meio urbano concentrar o maior número de acidentes e vítimas, continua a persistir uma lacuna ao nível da prevenção dos acidentes pela actuação na infra-estrutura. Esta pode dever-se, por um lado, à pouca consciência da influência da infra-estrutura na SR e, por outro lado, à falta de ferramentas de análise.

A seguir, em 3.1, descreve-se um conjunto de acções ao nível da infra-estrutura que têm vindo a ser desenvolvidas e aplicadas com vista à redução da sinistralidade e que se dividem em dois tipos de abordagens – acções *a priori* ou *a posteriori*.

Ainda neste capítulo, em 3.2, destaca-se das acções *a priori* a introdução da SR na fase de planeamento fazendo referência a um conjunto de trabalhos nesta área e, em 3.3, termina-se este capítulo com as conclusões.

3.1 Acções na infra-estrutura

Têm-se desenvolvido ao longo das últimas décadas diversas acções ao nível da infra-estrutura algumas das quais num número significativo de experiências de aplicação em países da Europa e, em geral, com resultados muito positivos (High Level Group Road Safety, 2003).

Estas acções podem ser classificadas num dos dois seguintes grupos (Cardoso, J. L. and Gomes, S. V., 2005):

- i. acções “*a posteriori*” – são fundamentadas no conhecimento geral existente sobre factores de risco e no conhecimento directo sobre indicadores de segurança rodoviária nos locais a intervir. Estas acções incluem medidas de tratamento das zonas de acumulação de acidentes (ZAA), de tratamento de um itinerário ou rua, de tratamento de grandes áreas, e a aplicação sistemática de um tipo de tratamento a uma determinada estrada ou itinerário ou a uma rede viária de determinada área geográfica;
- ii. acções “*a priori*”- são fundamentadas no conhecimento pré-existente de factores influentes na sinistralidade e consistem basicamente na realização de estudos de impacte sobre a segurança das intervenções planeadas e na execução de inspecções de segurança à rede rodoviária aberta ao tráfego.

Dos dois grupos de acção acima referidos destacam-se a seguir as seguintes acções:

- tratamento de zonas de acumulação de acidentes
- inspecção de segurança rodoviária (ISR)
- auditoria de segurança rodoviária (ASR)
- análise da segurança da rede rodoviária (ASRR)
- avaliação do impacte na segurança rodoviária (AISR)

Para ser possível a aplicação e estudo destas acções é fundamental a existência de uma base de dados de acidentes. Nesse sentido, a recolha e tratamento dos dados dos acidentes é da maior importância para a manutenção e desenvolvimento de infra-estruturas viárias seguras de forma eficiente. Sem uma base de dados confiável e actualizada a eficiência das acções acima referidas ficam comprometidas.

Para que o sistema de dados seja crível é necessário que a recolha dos dados seja organizada e que inclua a participação de várias instituições (Polícia de Segurança Pública, entidades administrativas da infra-estrutura e de saúde, etc.) para que seja possível, por exemplo, garantir a informação hospitalar na identificação das vítimas mortais. O registo deve ser uniforme com uma descrição do acidente completa e exacta destacando-se a importância da localização do acidente. Esta pode ser realizada de uma forma eficaz através da utilização das novas tecnologias tais como GPS, “*palm tops*” e Sistemas de Informação Geográfica (SIG). A base de dados deve estar actualizada sendo necessário, para tal, um financiamento estável garantindo assim a mesma qualidade dos dados de ano para ano.

O sistema de dados, tal como se descreve acima, só pode ser optimizado se estiver disponível quer para estudos de investigação quer para as entidades locais com responsabilidade na infra-estrutura.

Com uma base de dados deste tipo é mais fácil garantir a eficiência de qualquer uma das acções a seguir descritas.

Contudo, existe ainda um conjunto de obstáculos à formação de uma base de dados confiável. Destes obstáculos, destaca-se a não adopção de um sistema eficiente de registo do local do acidente, o que no caso particular do meio urbano é ainda mais gravoso, dado que a eficiência do processo de localização através, por exemplo, de um SIG está dependente do sistema toponímico da cidade. Salienta-se ainda o facto de não existir, até à data da recolha de dados para este trabalho, uma base de dados organizada em formato digital, gerando consequentemente perdas de tempo e erros resultantes do processo de inserção dos dados. Actualmente, em Portugal, a Autoridade Nacional de Segurança Rodoviária (ANSR) tem disponível uma base de dados em formato digital mas sem a localização precisa do acidente.

Tratamento de Zonas de Acumulação de Acidentes

Uma zona de acumulação de acidentes é uma zona geográfica onde, por influência de características da infra-estrutura rodoviária específicas à área, a frequência esperada de acidentes é superior ao expectável face à distribuição de acidentes nas áreas circundantes, nomeadamente em zonas aparentemente semelhantes (Cardoso, J. L., 1998).

Para a identificação das ZAA é necessário (Cardoso, J. L., 1998):

- definir uma unidade geográfica padrão de rede rodoviária;

- definir um período de contagem de ocorrências;
- estimar a frequência esperada de acidentes em situação de risco habitual;
- definir um limiar crítico de frequência esperada de acidentes.

A maior parte dos países europeus tem vindo a adoptar acções de tratamento sistemático de ZAA sendo que, em alguns deles como é o caso da Finlândia, Suécia e Reino Unido (RU), esta medida deixou de ter importância pois praticamente já não têm ZAA (High Level Group Road Safety, 2003). Em geral, os restantes países da Europa recorrem ao tratamento de ZAA, variando de país para país a sua definição que está dependente da base de dados de acidentes disponível. Destas experiências de tratamento sistemático de ZAA sabe-se que, quando incluídas em programas integrados de SR, têm uma relação custo-benefício muito favorável (Cardoso, J. L., 1998).

O tratamento de ZAA tem como objectivo reduzir os acidentes mas, como “a prevenção é melhor que a cura”, foram também desenvolvidas acções “*a priori*” que têm a vantagem de não ser necessário esperar pela ocorrência dos acidentes para serem aplicadas.

Inspecção de Segurança Rodoviária

A ISR é realizada por engenheiros de tráfego que examinam a rede viária existente (incluindo a área adjacente à faixa de rodagem) na perspectiva da segurança rodoviária e independentemente do número de acidentes ocorridos.

O objectivo é encontrar deficiências na infra-estrutura com implicações na segurança, de forma a definir medidas de tratamento que previnam os acidentes. Para tal, define-se um conjunto de indicadores de segurança relacionados com as características da infra-estrutura. A ISR destina-se a toda a rede viária nacional.

As inspecções rodoviárias são praticadas em Portugal, Itália, Grécia e Áustria numa perspectiva de manutenção da rede viária e são da responsabilidade das entidades regionais e locais. Mais recentemente, em Portugal, a ISR é incluída na ENSR como um objectivo operacional.

Curiosamente, no RU, com vinte anos de actividade na investigação da segurança rodoviária e de tratamento de ZAA, não aplicam a ISR. No entanto, existe uma proposta de incluir uma fase de pós-monitorização nas ASR que equivale à ISR. É o que acontece actualmente na Noruega. Na Alemanha existe um programa orientado de ISR aplicado

regularmente e dando relevância às intersecções, secções livres e áreas adjacentes à faixa de rodagem (High Level Group Road Safety, 2003).

Auditoria de Segurança Rodoviária

Segundo definição proposta Laboratório Nacional de Engenharia Civil (LNEC), a ASR é um conjunto de procedimentos destinados a incorporar, de modo explícito e formal, os conhecimentos e informações relativos à segurança rodoviária, no planeamento e projecto de estradas, com as finalidades de mitigar o risco de acidentes e de reduzir as respectivas consequências. Estes procedimentos podem ser aplicados aos projectos para construção de novos traçados rodoviários ou para intervenção em estradas existentes (Cardoso, J. L. and Roque, C. d. A., 2006).

A auditoria é realizada por um auditor ou uma equipa de auditores independente e qualificada que elabora um relatório onde indica os elementos potenciais de insegurança identificados no projecto e avalia o seu efeito na SR da utilização da obra.

A ASR actua desde as fases iniciais de concepção e projecto, quer na construção de novas vias quer à melhoria das vias existentes e suas zonas limítrofes. A ASR desenvolve-se em quatro fases, que acompanham o desenvolvimento do projecto e a sua execução em obra: estudo de viabilidade (programa base); estudo prévio; projecto de execução; projecto executado em obra, imediatamente após a conclusão dos trabalhos e antes da abertura ao tráfego (Bairrão, L. A., 1999).

O país com mais prática desta acção é o RU onde, desde 1990, a ASR é um requisito obrigatório para a rede nacional, quer para novas vias quer na reabilitação das existentes. A Dinamarca introduziu o primeiro manual e realizou projectos piloto em 1993, não sendo, no entanto, obrigatória a sua realização. Fora da Europa, e seguindo o exemplo do RU, a Austrália e Nova Zelândia têm adaptado os mesmos procedimentos com manuais de aplicação próprios (High Level Group Road Safety, 2003).

Em Portugal, no âmbito do projecto europeu SAFESTAR, foram realizados projectos piloto e elaborado um manual em 2002. O Plano Rodoviário Nacional de 2000 prevê a realização de ASR aos projectos de novas construções da Rede Rodoviária Nacional e o PNPR de 2003 refere a sua realização obrigatória no futuro em todos os casos de investimento rodoviário, dentro e fora da localidade e em empreendimentos (públicos e privados) com grande impacte na rede rodoviária. Mais recentemente, na ENSR, a realização de ASR é considerada como um objectivo operacional.

Análise da Segurança da Rede Rodoviária

A ASRR corresponde a uma metodologia de análise da rede na perspectiva da segurança rodoviária. Esta abordagem permite às entidades responsáveis pela administração da rede viária detectar quais as intervenções na infra-estrutura de determinados trechos de ligação da rede que poderão ter elevada eficiência em termos de custo-benefício. Este processo consiste em, após a selecção de um conjunto de intervenções possíveis, comparar o custo das mesmas com o potencial em reduzir um determinado custo de acidentes, analisando-se assim o rácio custo-benefício. Com os valores deste rácio, estabelece-se uma hierarquia com base no potencial em diminuir os custos de acidente, atribuindo-se assim prioridades nas intervenções à rede.

Esta análise não é praticada pela maior parte dos países europeus sendo que a metodologia referida baseia-se essencialmente nas abordagens consideradas pela Alemanha e pela França (High Level Group Road Safety, 2003). Contudo, existem diversas definições e abordagens de ASRR que variam de país para país e, tal como para as ZAA, não foi estabelecida uma definição normalizada internacional (Sorensen, M., 2007).

Avaliação do Impacte na Segurança Rodoviária

Apesar das acções ISR e ASR descritas anteriormente serem acções “*a priori*” nenhuma delas é específica de meio urbano e aplicável em fase de planeamento. Contudo, na última década têm surgindo alguns trabalhos que tentam dar resposta a esta lacuna, destacando-se um novo conceito introduzido como acção “*a priori*” – a avaliação do impacte na segurança rodoviária¹².

A AISR foi introduzida pelo instituto holandês SWOV – *Institute for Road Safety Research*, através do documento de Wegman, F. C. M., *et al.* (1994).

A AISR integra as ferramentas criadas ou adoptadas na Holanda no âmbito do desenvolvimento do conceito “sistema viário seguro sustentável” em que um dos princípios base é o de incorporar explicitamente a SR no planeamento e no processo de decisão (Wegman, F., *et al.*, 1994). Os princípios de um sistema viário seguro compreendem a (Wegman, F. and Aarts, L., 2006):

- Funcionalidade, proporcionando a utilização das vias correspondente à função atribuída;

¹² *Road Safety Impact Assessment*, em inglês

- Homogeneidade do sistema viário, prevenindo as grandes discrepâncias nas velocidades, nas massas dos veículos em circulação e nos sentidos;
- Previsibilidade da rede, evitando a existência de situações pouco usuais que contrariem as expectativas ao longo do percurso e resultando, por esse facto, num comportamento inapropriado dos utilizadores;
- Tolerância do meio envolvente e do comportamento do utente, proporcionando um comportamento coerente do utente com as condições do meio envolvente;
- Estado de vigilância (do utente) que permita ao utente avaliar as suas próprias capacidades e actuar de acordo.

A AISR inclui o procedimento legal e administrativo e a metodologia. O procedimento é desenvolvido de forma semelhante ao Estudo de Impacte Ambiental aplicado na Holanda, tendo sido já assegurado, através de duas disposições legais na lei ambiental, a incorporação da AISR. A metodologia está dividida em duas fases: 1ª - avaliação dos impactes na segurança rodoviária das diferentes soluções através da utilização de indicadores e tendo em conta diferentes tipos de via; 2ª - utilização de técnicas de auditoria para otimizar o projecto viário (similar à ASR referida anteriormente). Para descrever a aplicação de cada uma destas fases são elaborados dois relatórios cujos conteúdos devem seguir as indicações do documento de Wegman, F. C. M., *et al.* (1994).

A AISR é uma análise macro e mesoscópica da segurança rodoviária, sistemática e independente aplicada ao nível estratégico de decisão. A avaliação dos impactes é feita, segundo Wegman, F. C. M., *et al.* (1994), na redistribuição do tráfego aquando da alteração da rede e quando se altera as características físicas da rede.

A metodologia desenvolvida pelo SWOV resulta da constatação que, para diferentes tipos de vias, existe uma relação diferente entre o volume de tráfego e os acidentes rodoviários. Neste sentido, integram-se na análise os principais factores ao nível da infra-estrutura de ocorrência de acidentes: características geométricas da via e volumes de tráfego.

O método pode ser descrito sucintamente em três etapas (Wegman, F., *et al.*, 1994):

Etapla 1: determinar, ao nível nacional, os indicadores de segurança por categorias viárias (por exemplo, número de vítimas por quilómetro, número de vítimas por um milhão de quilómetros percorridos, número de vítimas por acidente com vítimas, etc.).

Etapla 2: calcular os indicadores de segurança na área em estudo e validar esses valores através dos indicadores nacionais obtidos na etapa 1.

Etapa 3: prever, para o ano horizonte, a rede viária e o respectivo volume de tráfego. Com estes dados, estima-se os indicadores de segurança. A partir destes valores, determinam-se os efeitos na segurança rodoviária das alterações na rede viária e no volume de tráfego.

Os resultados da AISR podem ser utilizados na fase de discussão pública e incorporados no processo de tomada de decisões.

A sinopse das respostas ao questionário realizado pela Comissão Europeia e publicado no documento High Level Group Road Safety (2003), refere que na Dinamarca este tipo de procedimento é obrigatório nos projectos de maior dimensão. O RU, Noruega, Alemanha, Suécia e Itália consideram a avaliação dos impactes como a primeira fase da ASR mas apenas para alguns tipos de projecto.

Em 2008, foi aprovada pelo Parlamento Europeu, uma directiva sobre a melhoria da segurança da infra-estrutura rodoviária. Esta directiva tem como objectivos garantir as condições de segurança em todas as fases das infra-estruturas rodoviárias (planeamento, concepção e funcionamento) e a rentabilidade e protecção do ambiente (MOPTC, 2009). Com a aplicação da nova directiva, os Estados Membros serão responsáveis pela realização de avaliações do impacte da SR, controlos da SR, inspecções de segurança, gestão da segurança da rede, bem como pela definição de normas ou requisitos técnicos.

Para analisar as diferentes abordagens desenvolvidas e aplicadas no âmbito da SR a nível europeu, foi desenvolvido o projecto RIPCORDER - ISEREST “*Road Infrastructure Safety Protection – Core Research and Development for Road Safety in Europe; Increasing Safety and Reliability of Secondary Roads for a Sustainable Surface Transport*”, co-financiado pela Comissão Europeia (www.ripcord-iserest.com). Este projecto teve como objectivo reunir e avaliar as várias abordagens desenvolvidas e aplicadas no âmbito da SR para desta forma tornar acessível esta informação aos países da Europa, facilitando o incremento de ferramentas que contribuam para o aumento da SR. No âmbito deste projecto foi elaborado o documento designado de “*Accident Prediction Models and Road Safety Impact Assessment: a state-of-the-art*” de Reurings, M., *et al.* (2005), que descreve o estado da arte dos modelos de previsão de acidentes (MPA) e da AISR.

O conceito de AISR, inicialmente desenvolvido por Wegman, F., *et al.* (1994), tem sido aprofundado em diversos trabalhos, referidos e analisados em Reurings, M., *et al.* (2005). Neste documento, e tendo em conta os trabalhos desenvolvidos, são considerados dois âmbitos de aplicação para a AISR: a avaliação de políticas de segurança na rede e a avaliação do impacte na segurança das alternativas planeadas. No primeiro caso, a AISR é

utilizada para avaliar as consequências na sinistralidade das acções a implementar no âmbito da segurança e assim avaliar a capacidade de se atingir determinadas metas. A AISR pode inclusive ser utilizada para definir metas realísticas no número de acidentes a diminuir. No segundo caso, a AISR serve para avaliar as consequências em termos de sinistralidade, das diferentes soluções funcionais e geométricas planeadas para a rede.

Partindo dos resultados dos vários trabalhos analisados são delineadas em Eenink, R., *et al.* (2007) algumas recomendações relativas à aplicação da AISR para as três entidades envolvidas na SR – autoridades rodoviárias, decisores políticos e investigadores. Destacam-se aqui as recomendações para os investigadores (Eenink, R., *et al.*, 2007):

- Aplicar os seguintes cinco passos:
 - o Definir a situação de base;
 - o Analisar cenário futuro sem a aplicação de medidas de segurança;
 - o Analisar cenário com medidas de segurança rodoviária;
 - o Efectuar análise custo-benefício
 - o Optimizar
- Ter consciência das limitações da AISR e comunicá-las aos utilizadores finais;
- Continuar a desenvolver a análise de dados através da utilização do SIG e integrar as consequências no ambiente e acessibilidades.

3.2 A segurança rodoviária no planeamento dos transportes

A AISR é uma acção do grupo de acções *a priori* referidas anteriormente, em 3.1, que se insere na fase de planeamento e que se encontra ainda em desenvolvimento. A introdução da SR na fase de planeamento é uma abordagem recente que tem sido analisada por diversos autores sugerindo diferentes formas de actuação. Vários estudos apontam para a importância de considerar a SR no planeamento do sistema de transportes através da definição de estratégias, objectivos e metas.

Sayed, T. *and* Leur, P. d. (2000) salientam que apesar do tratamento de ZAA ser muito importante e ter demonstrado bons resultados, é importante evoluir para uma abordagem proactiva da SR. Os autores apontam, no entanto, uma série de obstáculos que dificultam a análise da SR no planeamento dos transportes, tais como, a falta de uma metodologia e

ferramentas que avaliem a SR, a inexistência de um processo sistemático ou de uma estrutura que inclua a SR, ou de uma fase no tradicional processo de planeamento dos transportes que considere explicitamente a análise da SR. Neste trabalho, os autores analisam estes obstáculos e propõem formas de os ultrapassar. Segundo os mesmos, a actuação proactiva deve ser ao nível dos três elementos fundamentais que quantificam o risco de acidente: exposição, probabilidade e consequências. Para cada um dos três elementos, Sayed, T. and Leur, P. d. (2000) identificam factores que são decididos em fase de planeamento e que podem afectar esses três elementos:

- Exposição:
 - o Ocupação do solo;
 - o Estrutura e eficiência da rede viária;
 - o Escolha modal;
- Probabilidade:
 - o Capacidade de manobra dos veículos;
 - o Elementos geométricos do projecto;
 - o Funcionalidade viária;
 - o Conflitos de tráfego;
 - o Fricção da rede viária;
 - o Previsibilidade da via;
- Consequências:
 - o Protecção dos utilizadores vulneráveis;
 - o Redução das velocidades em áreas de risco;
 - o Envolvente viária.

Na mesma linha de trabalho, isto é, na actuação nos três elementos que quantificam a SR, o instituto holandês SWOV – *Institute for Road Safety Research*, publicou quatro relatórios de Hummel, T. (2001c; 2001a; 2001d; 2001b) que descrevem a forma de incluir a SR no planeamento da rede de transportes criando o conceito designado de “*Safer Transportation Network Planning*”. Cada relatório corresponde a uma das seguintes áreas de actuação:

- Planeamento da ocupação do solo;
- Gestão das acessibilidades;
- Gestão dos percursos;
- Planeamento das intersecções.

As duas primeiras áreas referidas estão muito interligadas e como tal, devem ser analisadas em conjunto, sendo que a gestão das acessibilidades tem como objectivo garantir os acessos apropriados, preservando, no entanto, a segurança e a capacidade da rede envolvente. Com a gestão dos percursos pretende influenciar-se a escolha dos percursos pelos utilizadores viários para que prefiram os mais seguros. Por fim, sendo as intersecções um elemento importante da rede viária, através da distribuição na rede e do *layout* das mesmas, define-se a velocidade das viagens, os custos de operação, a eficiência, a capacidade e a segurança da rede, etc. Para cada uma destas áreas são propostas acções que actuem nos três elementos referidos – exposição, probabilidade e consequências.

O planeamento da ocupação do solo é também considerado por Berkovitz, A. (2001) uma área importante de actuação para prevenção dos acidentes. Segundo esta autora, a relação entre a SR e a ocupação do solo é fundamental para a redução das vítimas, em especial dos utilizadores vulneráveis, tais como, peões e ciclistas. Berkovitz, A. (2001) recorda que os modelos urbanísticos implementados nas zonas residenciais logo após a 2ª Guerra Mundial foram influenciados pelo aumento exponencial do tráfego que gerou congestionamentos, pela evolução do automóvel nomeadamente na capacidade em atingir velocidades elevadas, a par do aumento da população nos meios urbanos provocando o crescimento da construção em altura. Com o tempo, tornou-se evidente que a crescente procura pela maior capacidade viária não é compatível com a imutável e consolidada urbanização. Assim, a solução é diversificar as viagens quer criando um maior número de percursos, quer diversificando os meios de transportes. Contudo, esta diversificação só é possível quando a ocupação do solo é compacta e mista. O sistema de transportes é pois uma consequência da ocupação do solo e para o alterar é necessário intervir na ocupação do solo. É por esse facto que a ocupação do solo e o sistema de transportes têm de ser planeados em conjunto sendo, por isso, recomendável que os agentes responsáveis pelo planeamento da rede de transportes trabalhem em conjunto com os técnicos e políticos responsáveis pelo planeamento do meio urbano (Berkovitz, A., 2001).

Mais recentemente foi publicada uma tese de doutoramento designada de “*Incorporating safety into transportation planning and decision making in mid-sized metropolitan areas*” de Gaines, D. L. (2007). Este trabalho teve como objectivo estudar sete casos de áreas metropolitanas de média dimensão (200 000 a 600 000 habitantes) dos Estados Unidos da América (EUA) de forma a entender os desafios e oportunidades do conceito “*Safety Conscious Planning*” neste contexto. Este conceito é uma abordagem proactiva para

reduzir e prevenir os acidentes rodoviários e as condições de insegurança no sistema de transportes através da integração de um conjunto de medidas de segurança no processo de planeamento de transportes ao nível federal, estadual, regional e local. Os resultados demonstraram que a maioria das áreas metropolitanas considera a SR no planeamento dos transportes a longo prazo (estratégias, metas e objectivos), embora com diferenças na forma de actuação das organizações sendo que algumas são mais activas do que outras. Por este facto, e com o objectivo de homogeneizar a acção das organizações, o autor recomenda uma série de medidas que visam envolver mais o governo no processo de planeamento destas mesmas organizações. As medidas propostas compreendem a criação de uma ordenação das qualificações resultante da avaliação do trabalho das mesmas nesta área, promovendo o desenvolvimento de mais ferramentas e formação neste âmbito e criando um fórum de colaboração entre todos os especialistas de segurança rodoviária.

A importância de quantificar os objectivos foi analisada por Wong, S. C., *et al.* (2006) através de um estudo já anteriormente referido no Capítulo 2. Os resultados deste estudo revelaram que a maioria dos países que estabeleceram metas diminuíram as vítimas mortais nesse período. Assim, os autores concluíram que a definição de metas está fortemente associada a uma melhoria da SR.

Nos EUA vários trabalhos têm sido desenvolvidos no sentido de criar ferramentas e contextualizá-las na fase de planeamento. No documento intitulado “*Considering Safety in the Transportation Planning Process*” de AECOM, *et al.* (2002) é referido a lacuna existente relativamente a uma metodologia que estime *a priori* o desempenho em termos de SR da rede de transportes, sugerindo três potenciais métodos:

- Avaliação por especialistas;
- Aplicação de modelos de análise da SR do projecto;
- Aplicação de modelos de previsão de acidentes a partir dos resultados da estimação do tráfego.

O primeiro método consiste na utilização dos conhecimentos de especialistas nas áreas do transporte e da segurança para a avaliação da situação.

O segundo corresponde à aplicação de modelos de previsão de acidentes com base em informação que caracteriza geométrica e funcionalmente a rede. Este método compreende, essencialmente a utilização de um software designado de IHSD - “*Interactive Highway Safety Design Model*”, desenvolvido para avaliar a segurança rodoviária de vias rurais na

fase de estudo prévio e/ou projecto de execução, ou seja, numa fase imediatamente posterior ao planeamento.

O terceiro método trata-se de avaliar a SR através da aplicação de modelos de previsão de acidentes em que as variáveis independentes são relativas ao tráfego cujos valores resultam da modelação do tráfego. Para este método é proposto um processo de análise técnica a longo prazo que incorpora potenciais ferramentas de análise da SR tal como indica o organigrama da Figura 3.1.

Segundo AECOM, *et al.* (2002) este processo poderá garantir a SR na rede com a introdução de ferramentas de análise da segurança nas fases de planeamento da ocupação do solo, de projecto de transporte e de definição da rede de transporte.

A primeira ferramenta proposta tem como função apoiar o planeamento da ocupação do solo de forma a contemplar critérios de segurança. A segunda ferramenta pretende apoiar o planeador a formular e a avaliar projectos que garantam a segurança no sistema de transportes. E, por último, duas ferramentas de análise da segurança da rede que se destinam a quantificar a sinistralidade e a analisar soluções alternativas à solução inicial, aplicadas em etapas distintas. Na primeira etapa, é determinado o número de acidentes através de um modelo de previsão para a solução considerada. Com os resultados obtidos por este modelo, avaliam-se os níveis de segurança atingidos sendo assim possível avaliar se, por exemplo, a solução analisada cumpre ou não os objectivos definidos. Se os resultados obtidos para a segurança não atingirem os valores desejados então entra-se na segunda etapa que corresponde a uma ferramenta de análise de medidas ou soluções alternativas de forma a melhorar o desempenho em termos de SR. Segundo os mesmos autores, esta ferramenta seria desenvolvida com base no conhecimento de medidas que melhoram a segurança da rede de transportes.

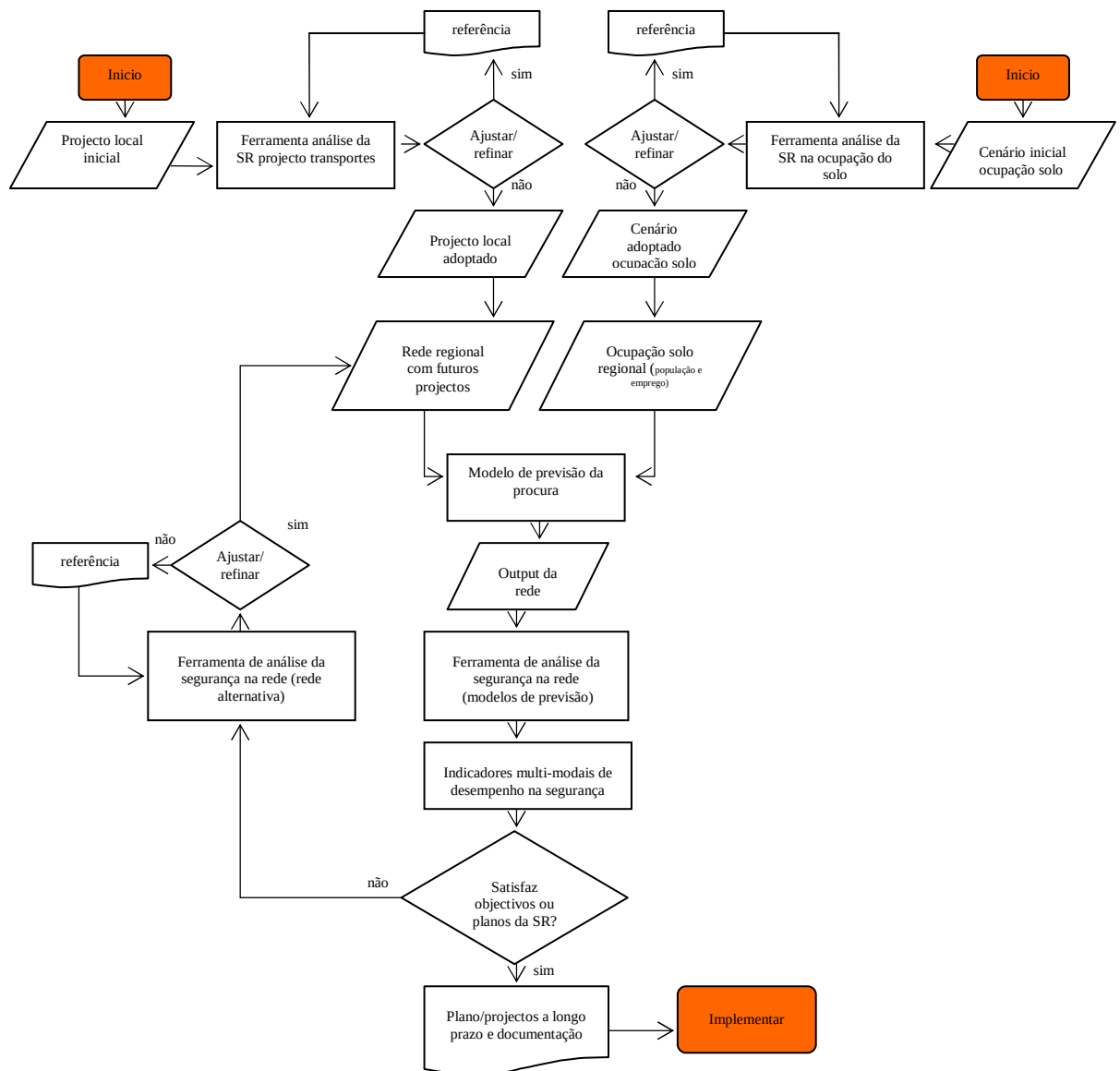


Figura 3.1. - Organograma representativo do processo de análise a longo prazo incorporando a SR (Fonte: (AECOM, et al., 2002))

Já em 2006 foi publicado o relatório com a designação - “*Incorporating Safety into Long-Range Transportation Planning*” de Washington, S., et al. (2006) que vai mais além do que o anterior, na medida em que descreve várias ferramentas já desenvolvidas e prontas a aplicar. Este relatório constitui um guia direccionado para os departamentos de transportes estatais, organizações metropolitanas de planeamento e agências que integram o planeamento de transportes regional e estadual. Como guia, refere as fases em que é

possível incluir a segurança no planeamento dos transportes a longo prazo: estratégia; metas e objectivos; definição de indicadores de desempenho; análise dos dados; ferramentas de análise; avaliação; plano e programa de desenvolvimento; monitorização. No anexo C deste documento descreve-se um conjunto de catorze ferramentas indicando para cada uma delas a fase de aplicação (reactiva ou proactiva), o nível de detalhe e os meios necessários para a respectiva aplicação. Destas ferramentas, destacam-se duas: SAFENET e PLANSAFE, ambas de duplo carácter - reactivo e pro-activo.

O SAFENET – “*Software for Accident Frequency Estimation for Networks*”, foi desenvolvido pelo Departamento de Transportes do RU. É um software com várias componentes em que cada uma delas é constituída por um modelo de previsão de acidentes correspondente a um determinado tipo de intersecção ou segmento de via. Os dados de tráfego para os modelos de previsão são obtidos pelo software CONTRAM.

O PLANSAFE – “*Planning Level Safety Prediction Model*”, foi desenvolvido no projecto do qual resultou o relatório de Washington, S., *et al.* (2006). Esta ferramenta pretende ser um apoio à tomada de decisão na fase de planeamento por permitir, por exemplo, conferir se determinadas metas e objectivos para a segurança poderão ser atingidos. Neste sentido, foram desenvolvidos vários modelos macro de previsão de acidentes cujos dados são agregados no mínimo numa zona do tipo da definida pelo habitual processo de zonamento para os modelos de transporte (ver 2.3), mas podem agrupar um conjunto de zonas equivalentes a uma região do tipo área metropolitana. Os dados agregados por zona correspondem a informação socio-económica, demográfica e do sistema de transportes (densidade de intersecções, comprimento da via por tipos de via, volume de tráfego, etc.). Este tipo de modelos de previsão de frequências de acidentes têm a finalidade de determinar o número de acidentes a ocorrer num determinado ano e podem ser utilizados em projectos de larga escala para informar acerca dos impactes na SR, para comparar diferentes soluções ou para estabelecer metas a atingir num determinado ano horizonte. Não devem, no entanto, ser utilizados como critério de selecção de alternativas de investimentos ou medidas de tratamento a implementar, pois estes modelos não têm a capacidade de incluir a complexidade das relações da rede com os utilizadores viários e, como tal, embora possam ser úteis para advertir não devem ser, no entanto, utilizados para restringir (Washington, S., *et al.*, 2006).

Das várias abordagens acima referidas para incluir a SR na fase de planeamento, ressalta em comum o facto de a maior parte considerar os modelos de previsão de acidentes (MPA) como uma ferramenta de análise quantitativa proposta.

Os MPA são constituídos por uma variável dependente – Y , que corresponde a um indicador de sinistralidade e por uma ou várias variáveis independentes que representam um conjunto de características dos locais em estudo, tais como volume de tráfego, comprimento, limite de velocidade, etc. - $X_1, X_2, \dots, X_n$, e que se relacionam com a variável dependente através de uma função:

$$Y = f(X_1, X_2, \dots, X_n) \quad (3.1)$$

No caso da SR, estes modelos têm vindo a ser desenvolvidos com dois objectivos, em termos de contexto de aplicação, distintos. Por um lado, a descrição do mecanismo que descreve as relações entre a infra-estrutura e os acidentes. Neste caso, o modelo é composto por um número elevado de variáveis independentes que representam determinadas características da infra-estrutura. Por outro lado, a previsão do número de acidentes para um determinado ano horizonte em que as variáveis independentes são normalmente factores de exposição que influenciam a frequência dos acidentes. São estes últimos modelos que se aplicam na fase de planeamento.

Dos MPA desenvolvidos distinguem-se três tipos quanto à área em estudo:

- Modelos micro: aplicados a entidades específicas, tais como uma intersecção ou secção de via, atendendo ao funcionamento específico dessa entidade;
- Modelos meso: aplicados à rede (arcos e nós) de uma determinada área, com dados relativos a factores de exposição;
- Modelos macro: com dados agregados numa zona e relativos a informação socioeconómica, demográfica e do sistema de transportes.

O primeiro tipo de modelos é aplicado quando se pretende analisar um local específico (intersecção ou secção de via) com o objectivo de avaliar o impacto na sinistralidade resultante de uma alteração do local em termos físicos e funcionais. Estes modelos são inseridos numa fase de decisão das características geométricas e funcionais de um determinado local. É o caso dos modelos que incorporam o software SAFENET.

O segundo tipo de modelos surgiu como ferramenta de aplicação específica na fase de planeamento para decisão entre alternativas. Como tal, são modelos que consideram como variáveis independentes dados disponíveis nesta fase. As variáveis independentes destes modelos são, em geral, indicadores que caracterizam a zona em estudo com informação que possa estar relacionada directa ou indirectamente com a ocorrência de acidentes (Levine, N., et al., 1995; Hadayeghi, A., et al., 2003; Sayed, T. and Lovegrove, G. R., 2005; Wedagama, D., et al., 2006; Wier, M., et al., 2009). As variáveis independentes mais utilizadas e com melhores resultados em termos de estimação matemática são: tipo e comprimento das vias, ocupação do solo, densidade da população e da população empregada, densidade de intersecções, volume de tráfego e limites de velocidade. Nos modelos macro os dados das variáveis são agregados por zona.

Ao nível meso, existem poucos trabalhos desenvolvidos destacando-se o de Lord, D. (2000) e o de Greibe, P. (2003). Nestes trabalhos a rede viária é dividida em duas entidades – arcos e nós desenvolvendo-se modelos distintos para cada uma destas entidades viárias. No trabalho de Lord, D. (2000) desenvolveu-se vários modelos para diferentes tipos de arcos e intersecções (por exemplo, número de vias no arco e tipo de sinalização da intersecção) e tendo como única variável independente o tráfego médio diário anual. No trabalho de Greibe, P. (2003) desenvolveu-se um modelo para cada entidade com um conjunto de variáveis independentes que caracterizam estas duas entidades (por exemplo, limite de velocidade, largura das vias, número de vias, número de ramos, tipo de sinalização, etc.).

Os MPA apresentados no documento de Reurings, M., *et al.* (2005) no âmbito do projecto RIPCORDER – ISEREST, estão divididos pelo tipo de envolvente, rural ou urbana, e em arcos ou intersecções. Para cada um destes modelos são descritos resumidamente vários trabalhos. Da análise destes trabalhos resulta um conjunto de recomendações para a aplicação dos MPA através do documento de Eenink, R., *et al.* (2007) tal como realizado para as AISR. Destacam-se mais uma vez as recomendadas para os investigadores:

- Definir um modelo simples para três a cinco tipos de vias e preferencialmente também para as intersecções entre estes tipos de vias, usando os seguintes métodos:
 - o Função básica:

$$E(\lambda) = \alpha Q^\beta$$
 em que $E(\lambda)$ é o valor esperado de acidentes e Q^β é a variável independente relativa ao tráfego
 - o Usar a modelação linear generalizada

o Assumir uma distribuição Binomial Negativa

- Se os dados assim o permitirem, expandir os modelos usando factores de modificação de acidentes ou adicionar variáveis explicativas.

No âmbito do mesmo projecto, foi ainda identificada um conjunto de critérios a considerar para a avaliação da qualidade dos modelos (Reurings, M., et al., 2005):

- A distribuição probabilística dos acidentes da base de dados original deve ser testada;
- A estrutura do termo dos resíduos deve ser testada;
- Devem ser desenvolvidos modelos distintos em função da gravidade dos acidentes;
- Devem ser desenvolvidos modelos distintos para cada tipo de entidade viária (intersecções, pontes, secções de via, etc.);
- Os dados relativos à exposição devem ser decompostos o máximo possível (volume de tráfego de veículos ligeiros, veículos pesados, peões etc.);
- A forma funcional a considerar para descrever a relação entre a variável dependente e as independentes deve ser seleccionada de uma forma claramente justificada, após teste de alternativas;
- As variáveis independentes devem ser introduzidas uma a uma no modelo;
- As correlações entre as variáveis devem ser testadas;
- A avaliação do ajuste do modelo deve incluir critérios que individualizem a componente sistemática explicada pelo modelo, a componente sistemática não explicada pelo modelo e a componente aleatória;
- A estrutura da componente sistemática não explicada pelo modelo deve ser analisada;
- Qualquer modelo deve identificar explicitamente as variáveis para as quais se pretende analisar uma interpretação causal;
- Nos modelos desenvolvidos para representar as relações causais devem ser considerados critérios operacionais explícitos (critérios que possam ser avaliados empiricamente);
- Deve ser analisada a hipótese de existirem variáveis omitidas no modelo que possam influenciar os resultados;
- Deve ser testada a eficiência do modelo na previsão da frequência dos acidentes;

- Os resultados estatísticos dos parâmetros estimados devem ser apresentados de uma forma que permita realizar uma síntese formal das conclusões obtidas com um conjunto de vários modelos (meta-análise).

No relatório desenvolvido por Chatterjee, A. and Wegmann, F. J. (2002), os autores consideram um desafio o desenvolvimento de modelos de previsão de acidentes para a fase de planeamento. Na fase de planeamento de, por exemplo, uma nova via, de um novo serviço de transporte público ou de uma alteração da ocupação do solo, não existem muitos dados disponíveis e como tal as variáveis independentes que possam integrar um modelo nesta fase são sempre em número reduzido. Este deve ser um factor a ter em conta quando se define as variáveis a introduzir no modelo, caso contrário, por falta de dados não será possível aplicá-lo na prática. Outro factor importante, segundo os autores, a ter em conta na selecção das variáveis para os modelos aplicados em fase de planeamento é que estas devem estar relacionadas de alguma forma com esta fase, isto é, devem estar relacionadas com características da rede que se definem nesta fase. Contudo, a maior parte das variáveis até agora estudadas que se relacionam de uma forma estatisticamente significativa com os acidentes são, em geral, relacionadas com características físicas da infra-estrutura que em fase de planeamento não estão ainda definidas. Assim, por estes factos, Chatterjee, A. and Wegmann, F. J. (2002) concluem que o desafio é desenvolver modelos de previsão com variáveis independentes orientadas para a fase de planeamento e que sejam estatisticamente significantes. Os autores recomendam, no entanto, que uma abordagem quantitativa do tipo dos modelos de previsão deve ser utilizada unicamente para comparar soluções alternativas e não para converter estes valores estimados em valores monetários para, por exemplo, uma análise de custo-benefício, pois subsiste ainda uma grande parcela de incerteza nos termos absolutos destes valores.

3.3 Conclusões

Nas últimas décadas, têm sido desenvolvidas e aplicadas várias acções ao nível da infra-estrutura com o objectivo de melhorar a SR. Estas podem ser divididas em dois tipos de acções – acções *a priori* e *a posteriori*. Tendo estas últimas, já vários anos de aplicação e,

em geral, com bons resultados, recentemente tem-se evoluído para o desenvolvimento e aplicação de acções *a priori*.

A SR pode ser incluída na fase de planeamento através da definição de estratégias, de objectivos e metas para o sistema de transportes ou ainda como critério de decisão entre cenários futuros. Em qualquer uma destas formas de introdução da SR no planeamento dos transportes, os modelos de previsão de acidentes podem desempenhar um papel importante como ferramenta de análise. Existem vários tipos de modelos em função da área de actuação. Os modelos macro (ao nível da área) e meso (ao nível da rede – arcos e nós) são os mais referidos para esta fase. Contudo, o desenvolvimento destes constitui um desafio na medida em que na fase de planeamento a informação é ainda muito generalista e, por isso, nem sempre relacionada directamente com as possíveis causas de ocorrência de acidentes.

3.4 Referências bibliográficas

AECOM, Bellomo-McGee Inc. and Ned Levine & Associates (2002). *Considering Safety in the Transportation Planning Process*. Federal Highway Administration.

ANSR. (2009). "Estratégia Nacional de Segurança Rodoviária 2008-2015." Diário da República 1ª série-nº122 - 26 de Junho de 2009, 4160-4188.

Bairrão, L. A. (1999) *Auditorias de Segurança Rodoviária*. Mestrado, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa.

Berkovitz, A. (2001). The marriage of safety and land-use planning: a fresh look at local roadways. *Public Roads*, **65**.

Cardoso, J. L. (1998). Definição e detecção de zonas de acumulação de acidentes. In: *Segurança Rodoviária - Avaliação e Redução da Sinistralidade*, Lisboa.

Cardoso, J. L. and Gomes, S. V. (2005). *Avaliação do impacte sobre a sinistralidade de medidas correctivas em trecho da EN6*. Lisboa: LNEC, 217/05-NPTS.

Cardoso, J. L. and Roque, C. d. A. (2006). Auditoria de Segurança Rodoviária ao projecto de estradas. Alguns ensinamentos da sua realização em Portugal. In: Lisboa.

Chatterjee, A. and Wegmann, F. J. (2002). *Assessing Safety Consequences of Alternative Highway Networks*. Knoxville, TN: Southeastern Transportation Center, The University of Tennessee.

DGV (2006). *Sinistralidade rodoviária. Elementos estatísticos de 2005*. Lisboa: Observatório de Segurança Rodoviária, DGV.

Eenink, R., et al. (2007). "Accident Prediction Models and Road Safety Impact Assessment: recommendations for using these tools." RiPCORD-iSEREST, Sixth Framework Programme, 20.

ERF (2006). *European Road Statistics 2006*. European Union Road Federation.

Gaines, D. L. (2007) *Incorporating safety into transportation planning and decision making in mid-sized metropolitan areas*. PhD, Georgia Institute of Technology, Georgia.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention*, **35**, 273-285.

Hadayeghi, A., Shalaby, A. and Persaud, B. (2003). Macro-level accident prediction models for evaluating the safety of urban transportation systems. *Transportation Research Record*, **1840**, 87-95.

High Level Group Road Safety (2003). *Road Infrastructure Safety Management*. European Commission DG Energy and Transport.

Hummel, T. (2001a). *Access management in Safer Transportation Network Planning*. Leidschendam: SWOV, D-2001-10.

Hummel, T. (2001b). *Intersection planning in Safer Transportation Network Planning*. Leidschendam: SWOV, D-2001-13.

Hummel, T. (2001c). *Land use planning in Safer Transportation Network Planning*. Leidschendam: SWOV, D-2001-12.

Hummel, T. (2001d). *Route management in Safer Transportation Network Planning*. Leidschendam: SWOV, D-2001-11.

Levine, N., Kim, K. E. and Nitz, L. H. (1995). Spatial Analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis and Prevention*, **27**, 675-685.

Lord, D. (2000) *The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models*. PhD, Department of Civil Engineering, University of Toronto, Toronto.

MOPTC (2009). *Plano Estratégico de Transportes 2008-2020*. Lisboa: Ministério das Obras Públicas, Transportes e Comunicações.

Reurings, M., et al. (2005). "Accident Prediction Models and Road Safety Impact Assessment: a state-of-the-art." RIPCORD-ISEREST: Road Infrastructure Safety Protection - Core Research and development for Road Safety in Europe; Increasing Safety and Reliability of Secondary Roads for a Sustainable Surface Transport, 135.

Sayed, T. and Leur, P. d. (2000). Developing a systematic framework for proactive road safety planning. In: *13th ICTCT workshop - Traffic Safety Measures*, Corfu - Greece.

Sayed, T. and Lovegrove, G. R. (2005). Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Canadian Journal of Civil Engineering*.

Sorensen, M. (2007). "Black spot management and safety analysis of road networks - best practice guidelines and implementation steps." RIPCORD-ISEREST: Road Infrastructure Safety Protection - Core Research and development for Road Safety in Europe; Increasing Safety and Reliability of Secondary Roads for a Sustainable Surface Transport, 99.

Washington, S., et al. (2006). *Incorporating safety into long-range transportation planning*. Washington D.C.: Transportation Research Board, NCHRP Report 546.

Wedagama, D., Bird, R. and Metcalfe, A. (2006). The influence of urban land-use on non-motorised transport casualties. *Accident Analysis and Prevention*, **38**, 1049-1057.

Wegman, F. and Aarts, L. (2006). *Advancing sustainable safety*, SWOV Institute for Road Safety Research, Leidschendam.

Wegman, F., et al. (1994). *Road Safety Impact Assessment: RIA*. Leidschendam: SWOV, R-94-20.

Wier, M., et al. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis and Prevention*, **41**, 137-145.

Wong, S. C., et al. (2006). Association between setting quantified road safety targets and road fatality reduction. *Accident Analysis and Prevention*, **38**, 997-1005.

4. MODELOS DE PREVISÃO DE ACIDENTES

Os modelos estatísticos podem desempenhar um papel fundamental no estudo da sinistralidade na medida em que é pouco viável realizar ensaios controlados para a descrição do mecanismo dos acidentes. Na maior parte dos casos, os modelos são desenvolvidos com o objectivo de medir efeitos parciais resultantes de um ensaio hipotético em que um determinado “tratamento” é aplicado, avaliando-se as alterações da frequência dos acidentes ou de outro indicador de sinistralidade (OCDE, 1997). No entanto, para estratégias ou acções políticas, normalmente pretendem-se realizar previsões para determinadas condições que constituem um cenário, e assim determinar as diferenças entre um cenário de referência e a solução alternativa que incluiu um conjunto de intervenções. Como ferramenta de apoio à decisão os modelos estatísticos podem ser bastante valiosos (OCDE, 1997).

O objectivo deste capítulo é descrever os vários modelos estatísticos desenvolvidos e aplicados recentemente. Ao longo dos últimos 20 anos, tem vindo a ser divulgado um conjunto considerável de estudos para análise da sinistralidade. Esses estudos têm, geralmente, em comum o facto de aplicarem modelos lineares generalizados (MLG). Estes modelos caracterizam-se por uma componente aleatória distribuída segundo uma função da família exponencial, por uma componente sistemática e por uma função de ligação das referidas componentes. A forma como os MLG são considerados e tratados tem muitas variantes que podem obviamente influenciar o resultado final da aplicação.

A abordagem considerada para a descrição dos modelos neste capítulo compreende a distinção em dois tipos de modelos em função da resposta a analisar. Assim, quando a resposta do modelo a analisar corresponde à frequência de acidentes registada num determinado local para um determinado período de tempo, os modelos designam-se de resposta quantitativa. Os modelos são de resposta qualitativa quando a variável de resposta corresponde a um código representativo de um grau de sinistralidade. Descreve-se a seguir, em 4.1. e em 4.2., as duas abordagens.

4.1 Modelos de resposta quantitativa

O estudo econométrico de um fenómeno consiste na formulação matemática de um modelo ajustado aos dados empíricos existentes com o objectivo de fundamentar uma teoria. Isto é, os modelos econométricos são modelos utilizados para estudar uma determinada teoria e não para explorar relações empíricas que possam eventualmente existir entre um conjunto de variáveis. Essa teoria deve ser fundamentada em algo para além de uma análise de dados (OCDE, 1997).

Os acidentes têm características específicas que se ajustam a uma análise com base na aplicação de modelos econométricos, em particular o facto dos acidentes serem eventos aleatórios mais do que qualquer outro fenómeno social pois a sua ocorrência é indesejada, gerando, no entanto, por esse facto, uma baixa probabilidade de ocorrerem. Este facto impossibilita a realização eficaz de experiências controladas para analisar o processo de ocorrência de acidentes. Assim, a base de dados dos acidentes resulta de um processo de registos não experimental, sendo que este processo não é isento de erros podendo os mesmos distorcer determinados efeitos. Estes erros só poderão ser evidenciados com a introdução no modelo de um conjunto relevante de variáveis independentes. O facto dos acidentes serem eventos aleatórios e imprevistos ao nível micro (é impossível determinar onde, quando e como vai ocorrer um acidente), não impede que o número de acidentes seja objecto de uma análise causal ou de intervenção política (OCDE, 1997). Através das características da infra-estrutura rodoviária, dos veículos e do comportamento dos condutores, a probabilidade de um acidente ocorrer pode ser influenciada, alterando a longo prazo a frequência dos acidentes.

Esta frequência de acidentes a longo prazo – número esperado de acidentes por unidade de tempo, resulta de um processo causal. Assim, considere-se o valor esperado da frequência de acidentes que ocorrem num determinado período de tempo t e num determinado local r – $\mu(r,t)$. Este número esperado de acidentes não é constante, isto é, varia no tempo e no espaço. A variação atribuída a vários factores causais é sistemática e pode, à partida, ser influenciada e controlada. Para efeitos de estratégia política da segurança rodoviária é esta componente sistemática que interessa estudar (OCDE, 1997).

O número esperado de acidentes é determinado considerando a relação:

$$\mu(r,t) = E[y(r,t)|X(r,t)] = f\{X(r,t)\} \quad (4.1)$$

em que $y(r,t)$ corresponde ao número de acidentes observado num determinado local e num determinado período de tempo e $X(r,t)$ corresponde a um conjunto de variáveis independentes que influenciam a ocorrência do acidente num determinado local e num determinado período de tempo – $x_1(r,t), x_2(r,t), \dots, x_n(r,t)$.

O cálculo do valor esperado de $\mu(r,t) = E[Y]$, função de determinadas variáveis independentes, é a base da aplicação dos modelos de previsão de acidentes (MPA).

Os MPA mais recentes são desenvolvidos com base nos modelos lineares generalizados.

Os MLG são uma generalização dos modelos clássicos lineares.

Simplificadamente, os modelos podem ser especificados em três partes (McCullagh, P. and Nelder, J. A., 1989):

1. componente aleatória - as componentes de Y (do qual resulta um conjunto de n observações y) são independentes com distribuição Normal em que $E[Y] = \mu$ e variância constante σ^2 ;
2. componente sistemática – variáveis independentes $x_1, x_2, \dots, x_p$ que produzem um factor de previsão linear¹³ η dado por $\eta = \sum_{j=1}^p x_j \beta_j$;
3. a ligação entre as componentes aleatória e sistemática – no caso dos modelos clássicos lineares estas componentes são iguais $\mu = \eta$, mas podemos generalizar considerando uma função de ligação $g(\cdot)$ ficando $\eta_i = g(\mu_i)$.

A generalização dos modelos lineares clássicos permite duas extensões. Por um lado, a distribuição da componente aleatória (ponto 1) pode assumir uma distribuição diferente da Normal, mais concretamente qualquer distribuição do tipo exponencial (Poisson, Binomial Negativa ou Gama). Esta extensão é uma grande vantagem dos MLG aplicados aos acidentes, na medida em que as distribuições exponenciais se aproximam das características dos valores das frequências dos acidentes – valores inteiros e positivos. Por outro lado, a função de ligação (ponto 3) pode ser qualquer função monótona e diferencial. No caso dos acidentes, em que se assume habitualmente uma distribuição exponencial do

¹³predictor linear, em inglês

tipo Poisson, a função de ligação é uma função logarítmica, garantindo assim que para qualquer combinação de parâmetros e variáveis independentes o modelo aditivo não altera a característica dos valores das frequências dos acidentes (números inteiros e positivos).

Tal como para os modelos lineares clássicos, assume-se para os MLG que a componente aleatória é única, a componente sistemática é linear, o termo de erro não deve ser correlacionado com qualquer variável independente e as observações são independentes, não devendo apresentar auto-correlação espacial ou temporal (Cardoso, J., 2007).

Assim, para se desenvolver um MPA é necessário:

1. definir as variáveis independentes x
2. determinar a função f
3. definir a distribuição probabilística que rege a componente aleatória
4. estimar os parâmetros da função $f(x)$

A seguir descrevem-se os procedimentos para estes quatro pontos que, em grande parte, se aplicam também aos modelos qualitativos descritos a seguir, em 4.2.

Variáveis independentes

As variáveis independentes podem ser de dois tipos: variáveis que descrevem a exposição ao risco de acidentes e variáveis que identificam factores de risco que influenciam a ocorrência de acidentes numa unidade de exposição (Reurings, M., et al., 2005).

As variáveis independentes mais utilizadas nos MPA para descrever factores de risco são, geralmente, representativas das características físicas e funcionais da rede, tais como a função hierárquica, as características geométricas e o tipo de controlo de tráfego (Fridstrom, L., et al., 1995; Lord, D., 2000; Greibe, P., 2003; Golob, T., et al., 2004; Polus, A., et al., 2005; Reurings, M., et al., 2005). A variável de exposição mais utilizada é um indicador do tráfego, normalmente o Tráfego Médio Diário Anual (TMDA). Esta variável tem sido referida em diversos trabalhos como a mais determinante para a ocorrência de acidentes. (Fridstrom, L., et al., 1995; OCDE, 1997; Lord, D., 2000; Lord, D., 2006).

O processo de selecção das variáveis passa muitas vezes por uma fase de experimentação, tendo idealmente por base uma teoria que a fundamente (Fridstrom, L., et al., 1995). Na prática, o que habitualmente acontece é que as escolhas das variáveis é limitada aos dados

disponíveis. No relatório de Reurings, M., *et al.* (2005) recomenda-se que as variáveis independentes sejam seleccionadas com base em estudos anteriores, que possam ser medidas e determinadas de uma forma fiável, e que não sejam correlacionadas com outra variável já incluída. No método utilizado por Fridstrom, L., *et al.* (1995) as variáveis são introduzidas uma a uma e apenas são incluídas no modelo se a relação com a variável dependente for estatisticamente significativa.

Reforçando estes princípios de escolha das variáveis, Maher, M. *and* Summersgill, I. (1995) referem que se deve ter em conta três factores:

- O nível de significância estatística – as variáveis são rejeitadas para níveis de confiança inferiores a 95% e para níveis de confiança superiores a 99% só serão rejeitadas se um estudo cuidadoso assim o justificar;
- A estabilidade do modelo – se as variáveis estão relacionadas entre si então, ao introduzir uma variável, os valores dos parâmetros das outras variáveis são afectados.
- O efeito da variável deve ser lógico – o valor do parâmetro deve ser ajustado aos conhecimentos empíricos e a estrutura do modelo deve fazer sentido.
- A dimensão do efeito e a facilidade de medição da variável – são preferíveis as variáveis que tenham um efeito preponderante e que sejam fáceis de medir.

Hauer, E. *and* Bamfo, J. (1997) propuseram um método, com base na análise dos resíduos, para avaliar a relação das variáveis independentes com a variável dependente – método CURE *CUmulative REsiduals*¹⁴. Assim, requer a necessidade de pré-estimar os parâmetros para determinar os valores previstos e com estes os resíduos de cada observação. De seguida, é elaborado um gráfico com os acumulados desses resíduos, ordenados por ordem crescente de acordo com determinada variável independente a analisar. O padrão esperado para os resíduos acumulados no gráfico corresponde a uma linha que deveria oscilar em torno da origem e terminar com um valor final acumulado próximo do valor nulo. Através do método CURE, é analisada a variação da linha para distinguir o que é um padrão que represente um bom ajuste do modelo e o que é um padrão de enviesamento sistemático do modelo. Para tal, considere-se que N é o número total de resíduos, n é um inteiro entre 1 e

¹⁴ Resíduos Acumulados, em português

N e $S(n)$ é a soma dos resíduos de 1 a n . Considere-se ainda a hipótese de $S(n)$ seguir a distribuição normal com média nula e variância calculada a partir do gráfico acumulado dos resíduos ao quadrado e que a soma dos resíduos de n a N segue a distribuição normal com média nula e variância calculada como o resto do gráfico de acumulados dos resíduos ao quadrado (de n a N). Então, é possível definir a função de densidade de probabilidade de que a soma dos resíduos atinja um determinado valor para n e seja nula para N . Essa função é normal, com valor médio nulo e variância calculada pela equação:

$$\sigma^{*2} = \sigma^2(n)[1 - \sigma^2(n)/\sigma^2(N)] \quad (4.2)$$

Com este método avalia-se a qualidade de ajuste do modelo e a influência das variáveis independentes que vão sendo sucessivamente introduzidas no modelo. Assim, por observação do gráfico, considera-se um modelo satisfatório quando os resíduos acumulados, ordenados de acordo com o valor da variável independente, oscilam em torno da origem, termine com um valor final acumulado próximo do valor nulo e não ultrapasse a banda limitada pelos valores $\pm 2\sigma^*$.

Função f

A forma funcional mais habitual e recentemente utilizada numa análise multivariada é o modelo de regressão linear:

$$\eta_i = \sum_j \beta_j x_{ji} \quad (4.3)$$

em que i corresponde ao número de observações, j ao número de variáveis independentes, η_i é a componente sistemática do modelo, β_j os parâmetros a estimar e x_{ji} as variáveis independentes. Note-se que as variáveis x_{ji} podem assumir qualquer forma, isto é, o modelo é linear para os parâmetros β_j mas não necessariamente nas variáveis x_{ji} .

Tendo em conta que o valor esperado de acidentes tem de ser necessariamente positivo, é conveniente garantir esta condição na função considerando por exemplo:

$$E[Y] = e^{\sum_j \beta_j x_j} \quad (4.4)$$

Sendo esta função linearizada através da função logarítmica.

Uma das formas mais recentemente utilizada é (Reurings, M., et al., 2005; Cardoso, J., 2007):

$$E[Y] = \alpha Q^\beta e^{\sum \gamma_j x_j} \quad (4.5)$$

em que o valor esperado da frequência de acidentes – $E[Y]$ é uma função multiplicativa do tráfego – Q , e de um conjunto de p factores de risco $x_j = x_1, x_2, x_3, \dots, x_p$, sendo os parâmetros a estimar α , β e γ_j .

A relação entre os acidentes e o tráfego (normalmente o TMDA) é dada pela função potência e a função exponencial é a forma utilizada para relacionar a frequência dos acidentes com as variáveis x_j . A função potência apresenta duas características fundamentais. Por um lado, reflecte a situação que quando não há tráfego não há acidentes, por outro lado, estabelece uma relação não linear entre os acidentes e o fluxo de tráfego (Lord, D., 2000). Só se o valor de β for igual a 1, então o valor esperado da frequência dos acidentes é proporcional ao volume de tráfego. Na maior parte dos modelos o valor de β é diferente da unidade, devido, eventualmente, a não se desagregar a variável dependente por tipos ou a alterações das características da circulação do tráfego (Cardoso, J., 2007).

Uma alternativa ao modelo base acima referido e representado pela equação (4.5), é considerar como única variável independente o tráfego e aumentar o número de modelos por categorias correspondentes às variáveis independentes não consideradas. De acordo com Hauer *and* Persaud (1996) citado em (Lord, D., 2000), é preferível separar os dados em diferentes modelos para cada variável independente do que criar um modelo com muitas variáveis, na medida em que simplifica o modelo e consequentemente a estimação dos parâmetros. A desvantagem deste processo é que quanto maior o número de divisões consideradas menor o número de dados para cada modelo sendo, por isso, difícil de garantir um número de dados suficientes para a modelação.

Ivan, J. N., *et al.* (1999) acrescentam ainda, com base nos resultados do trabalho por eles realizado, que a escolha da função a considerar nos modelos é difícil quando os acidentes não são desagregados por tipo.

Para além do método CURE atrás referido, Hauer, E. *and* Bamfo. J. (1997) apresentam no mesmo trabalho o método *Integrate-Differentiate* (ID). Este método tem como objectivo auxiliar na determinação da melhor forma funcional para o modelo tendo em conta os dados das variáveis.

Simplificadamente, o método consiste em separar cada variável independente numa série de “classes” (correspondendo cada “classe” a um local observado), sendo posteriormente

estas “classes” ordenadas por ordem crescente dos seus valores. Associando-se a cada “classe” o acumulado dos acidentes de todas as “classes” precedentes estabelece-se a “função integral empírica” (FIE). Assim, para cada local a fronteira esquerda das “classes” é localizada a meio entre o local corrente e o local anterior. A fronteira direita é localizada a meio entre o local corrente e o local seguinte. A altura da “classe” é o número de acidentes ocorridos no local. Assim, o valor de FIE na fronteira direita da “classe” corrente é o acumulado dos acidentes de todas as “classes” desde o valor mais pequeno até à fronteira. O objectivo do método é comparar o gráfico FIE obtido da forma referida com outros gráficos já estabelecidos de funções conhecidas (função potência, gama, polinomial, etc.) O gráfico com a forma mais similar ao gráfico FIE indica a relação mais apropriada entre a variável dependente e as variáveis independentes. Este método deve, contudo, ser utilizado apenas como uma investigação preliminar sendo necessário utilizar dados adicionais para confirmar as formas funcionais inicialmente seleccionadas.

Distribuição probabilística da componente aleatória

Para estimar os parâmetros de um modelo é necessário especificar a distribuição probabilística da componente aleatória. As características da distribuição associada a esta componente afectam fortemente a eficiência das técnicas de estimação e a estimativa de intervalos de confiança (OCDE, 1997).

A ocorrência de um acidente pode ser comparada a um ensaio de Bernoulli que consiste num processo binomial, isto é, o resultado de cada ensaio só pode ser de dois tipos: “sucesso” ou “insucesso”. A ocorrência de um acidente corresponde a “sucesso”, entendido como um ensaio que produz um resultado indesejável (Cardoso, J., 2007). Os ensaios de Bernoulli são independentes entre si, ou seja, o resultado de um ensaio não afecta os resultados dos outros ensaios. Supondo que a probabilidade de “sucesso” é constante para todos os ensaios, o modelo probabilístico mais apropriado para representar uma série de ensaios de Bernoulli é a distribuição binomial dada por (Lord, D., et al., 2005b):

$$P(Y = n) = \binom{N}{n} p^n (1 - p)^{N-n} \quad (4.6)$$

em que $n=0,1,2,\dots,N$. Nesta equação, n é definido como o número de acidentes (“sucesso”). A média e a variância da distribuição binomial é $E(Y)=Np$ e $VAR(Y)=Np(1-p)$, respectivamente.

Numa situação típica de ocorrência de acidentes em que a probabilidade de ocorrência é baixa e o número de ensaios é elevado (por exemplo: número de veículos que entram numa intersecção, quilómetros percorridos, etc.) a distribuição binomial pode ser aproximada pela distribuição de Poisson (Lord, D., et al., 2005b). A distribuição de Poisson é uma distribuição probabilística utilizada para descrever a probabilidade do número de ocorrências num intervalo contínuo (tempo ou espaço). A unidade de medida é contínua, mas a variável aleatória (número de ocorrências) é discreta (número inteiro e positivo). Os “insucessos” não são contáveis, isto é, não podemos determinar a probabilidade dos acidentes não ocorrerem.

Tendo como base estes princípios teóricos, a distribuição de Poisson foi frequentemente considerada no desenvolvimento de MPA. A distribuição de Poisson especifica que a observação y_i tem uma probabilidade de ocorrer segundo a distribuição probabilística de Poisson. A distribuição de Poisson é caracterizada pelo parâmetro λ_i que por sua vez é função de um conjunto de variáveis independentes x_i (Greene, W. H., 2008). A forma funcional que caracteriza a distribuição de Poisson é:

$$P[Y = y_i] = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \quad (4.7)$$

Esta distribuição é caracterizada por uma média $\mu_i = E[Y] = \lambda_i$ e variância $Var[y_i] = \lambda_i$, ou seja, a média é igual à variância.

A função de ligação estabelecida que relaciona o parâmetro de Poisson λ_i às variáveis independentes é a função logarítmica:

$$\ln \lambda_i = \sum_j \beta_j x_{ij} \quad (4.8)$$

Tal como se referiu anteriormente para os ensaios de Bernoulli, a distribuição de Poisson também assenta em duas hipóteses: a probabilidade de ocorrência de um evento associado a um local é constante num determinado período de tempo e os eventos ocorridos em intervalos diferentes são probabilisticamente independentes (Cardoso, J., 2007).

Contudo, nem sempre é possível admitir a constância da probabilidade das ocorrências. Por um lado, tem-se verificado uma tendência para a diminuição dos acidentes ao longo do tempo reflectindo uma diminuição do risco de acidente e, por outro lado, diversos factores levam a pôr em causa a homogeneidade do risco individual de cada entidade de um grupo homogéneo de entidades, dando origem a valores da variância diferentes do valor esperado (por sua vez igual ao valor médio) (Cardoso, J., 2007).

Para ultrapassar este problema, tem vindo a ser cada vez mais frequente o recurso à distribuição Binomial Negativa (BN)¹⁵. Nesta distribuição considera-se que os acidentes resultam de um processo cujas probabilidades de sucesso variam em cada evento. Desta forma, a distribuição BN é caracterizada por uma distribuição de Poisson para a componente aleatória e uma distribuição Gama para o termo do erro u_i , com média 1,0 e variância α (Greene, W. H., 2008):

$$\ln \mu_i = \sum x_i \beta + \varepsilon_i = \ln \lambda_i + \ln u_i \quad (4.9)$$

A forma funcional que caracteriza a distribuição BN é¹⁶:

$$\begin{aligned} \Pr[Y = y_i] &= \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i + 1)} r_i^\theta (1 - r_i)^{y_i} \\ r_i &= \lambda_i / (\theta + \lambda_i) \\ \theta &= 1 / \alpha \\ \text{Var}[y_i] &= E[y_i] \{1 + \alpha E[y_i]\} \end{aligned} \quad (4.10)$$

em que α é o parâmetro de sobre-dispersão, $\text{Var}[y_i]$ é a variância da distribuição e $E[y_i]$ é o valor esperado da frequência de acidentes (igual a μ_i) e $\Gamma(.)$ é a função Gama. Quando o parâmetro de sobre-dispersão α se aproxima de zero ($\theta \rightarrow \infty$), a regressão reverte para a distribuição de Poisson (Lord, D. and Park, P. Y. J., 2008).

Como referido anteriormente, a principal causa para se verificar uma dispersão (variância diferente da média) tem sido atribuída à natureza aleatória do processo de ocorrência de acidentes na medida em que os acidentes resultam de um processo de ensaios de Bernoulli com desigual probabilidade, ou seja, mesmo nos locais com iguais características (físicas e funcionais), os valores da média diferem entre si (Lord, D., 2006; Lord, D. and Park, P. Y. J., 2008). Contudo, outros factores podem ser apontados como possíveis causas, tais como a forma como se agrupam os dados, a não consideração da correlação temporal, o facto de

¹⁵ também designada de distribuição Poisson-Gama

¹⁶ existem duas famílias de funções para o modelo BN. Para se obter a outra função, substitui-se θ por $\theta_i = \theta \lambda_i$.

não se quantificarem todas as variáveis independentes, por não terem sido observadas ou por não serem observáveis, resultando num problema de omissão na componente sistemática da regressão (Lord, D., 2006).

No caso dos acidentes esta dispersão é, geralmente, do tipo sobre-dispersão em que a variância é superior à média.

O valor do parâmetro de sobre-dispersão α pode ser utilizado para avaliar a capacidade do modelo de explicar a componente sistemática do processo de ocorrência dos acidentes. Como uma das possíveis causas da sobre-dispersão é a omissão de variáveis na componente sistemática, então quanto menor o valor de α melhor o modelo. Existem, no entanto, vários indicadores da qualidade de ajuste do modelo que utilizam o valor do parâmetro da sobre-dispersão para calcular a percentagem da componente sistemática explicada pelo modelo (Fridstrom, L., et al., 1995; Reurings, M., et al., 2005).

Partindo do princípio que as opções acerca da natureza da sobre-dispersão podem afectar as estimativas, através do método da máxima verosimilhança, dos parâmetros do modelo, Hauer, E. (2001) analisou a sobre-dispersão na modelação dos acidentes em arcos. O autor verificou que ao considerar o parâmetro de sobre-dispersão constante para todos os arcos, a estimação dos parâmetros do modelo é condicionada pelo comprimento dos arcos. Assim, a influência nos resultados de estimação de um determinado arco é tanto maior quanto menor for o seu comprimento. Esta condicionante resulta do facto, segundo o autor, de um segmento de via (arco) poder ser interpretado como um conjunto de subsecções que diferem nas características e que não estão representadas no modelo. Como solução para ultrapassar este facto, Hauer, E. admitiu que o parâmetro de sobre-dispersão pode variar de arco para arco em função do comprimento destes – $\alpha_i = \alpha L_i$.

Este assunto foi também recentemente analisado por Lord, D., and Park, P. (2008) em modelos para intersecções em meio rural, tendo como única variável independente o tráfego do ramo principal e secundário das intersecções. Lord, D., and Park, P. (2008) compararam um modelo de distribuição BN tradicional (sem variação do parâmetro de sobre-dispersão) e um modelo de distribuição BN Generalizada (com variação do parâmetro de sobre-dispersão). Verificou que este último tem melhores resultados nos indicadores de qualidade de ajuste do modelo. Verificou ainda que a escolha da função para o modelo tem grande influência no parâmetro de sobre-dispersão (valor e significância estatística). O mesmo processo foi aplicado por Mitra, S. and Washington, S.

(2007) em modelos para intersecções de meio rural tendo chegado a conclusões semelhantes.

Esta heterogeneidade do parâmetro de sobre-dispersão (variação do parâmetro de local para local) pode ser representada por (Greene, W. H., 2008):

$$\alpha_i = \alpha \exp(\delta z_i) \quad (4.11)$$

em que α_i é o parâmetro de sobre-dispersão a variar em função das variáveis z_i , constantes ao longo do tempo.

A heterogeneidade entre locais pode-se reflectir em vários parâmetros do modelo para além da sobre-dispersão. A heterogeneidade dos dados ao longo do tempo e entre locais (dados em painel¹⁷) tem sido analisada e aplicada com base em diferentes abordagens que a seguir se resumem.

A primeira abordagem, designada de efeito fixo¹⁸, reflecte a heterogeneidade dos dados através do parâmetro β_0 , afecto à constante do modelo e tem a seguinte formulação geral (Greene, W. H., 2008):

$$\begin{aligned} \ln \mu_{it} &= \beta x_{it} + \beta_{0i} + \varepsilon_{it} \\ \beta_{0i} &= z_i \beta_0 \end{aligned} \quad (4.12)$$

em que β_{0i} está relacionado com as variáveis z_i . Esta formulação reflecte as diferenças entre grupos de locais (obtidos pela correlação com as variáveis independentes) que são capturadas no termo da constante - β_{0i} . Cada β_{0i} é tratado como um parâmetro a ser estimado.

Se, contudo, os locais têm efeitos específicos não-correlacionados com as variáveis independentes, então será mais apropriado modelar o termo da constante como distribuído aleatoriamente. Esta abordagem designa-se de efeito aleatório¹⁹ (Greene, W. H., 2008):

$$\begin{aligned} \ln \mu_{it} &= \beta x_{it} + \beta_0 + v_i + \varepsilon_{it} \\ \beta_0 &= E[z_i \beta_0] \\ v_i &= \{z_i \beta_0 - E[z_i \beta_0]\} \end{aligned} \quad (4.13)$$

¹⁷ *panel data*, em inglês

¹⁸ *fixed effect*, em inglês

¹⁹ *random effects*, em inglês

em que o termo da constante β_0 é a média da heterogeneidade não observada e v_i é a heterogeneidade aleatória específica de cada observação i que se considera constante no tempo.

Esta heterogeneidade do termo da constante pode ainda ser estendida a todos os parâmetros das variáveis independentes do modelo designando-se, neste caso, de parâmetros aleatórios²⁰. Esta abordagem tem como pressuposto que, de local para local, a média dos valores das variáveis independentes pode variar. Neste caso, a regressão tem a seguinte formulação (Greene, W. H., 2008):

$$\ln \mu_{it} = x_{it} (\beta + h_i) + (\beta_0 + v_i) + \varepsilon_{it} \quad (4.14)$$

em que h_i é um vector aleatório que traduz a variação dos parâmetros entre locais.

Das três abordagens referidas, esta última tem vindo a ser recentemente aplicada na modelação dos acidentes (Gkritza, K. and Mannering, F. L., 2008; Milton, J., et al., 2008). Anastasopoulos, P. C. and Mannering, F. L. (2009) aplicaram a distribuição Binomial Negativa com parâmetros aleatórios a um conjunto de arcos de vias de meio rural. Verificaram que algumas das variáveis independentes do modelo apresentaram uma variação significativa entre os arcos, o que leva a concluir que a não consideração dos parâmetros aleatórios induz uma incorrecta análise da influência de determinada variável na frequência dos acidentes.

Outra característica frequentemente verificada nos dados de acidentes é a não ocorrência de acidentes em determinados locais durante um determinado período de tempo. Ou seja, se os locais a estudar forem seleccionados aleatoriamente, uma preponderância de zero acidentes poderá ser observada dado que os acidentes são acontecimentos raros (Qin, X., et al., 2004). Esta sobre representação de zeros nos dados pode, erradamente, ser considerada como uma dispersão.

Com o objectivo de diferenciar a origem dos *zero* acidentes registados nos locais, vários autores analisaram e aplicaram uma abordagem designada de zero inflacionado²¹ (ZI). Surge assim a teoria que o processo de geração de acidentes pode ser o resultado de dois estados diferentes. O estado “normal” corresponde à ocorrência de acidentes segundo um determinado número esperado constante por unidade de tempo. O outro estado corresponde à frequência esperada de acidentes nula – estado “zero acidentes” (Cardoso, J., 2007). Para

²⁰ *random parameters*, em inglês

²¹ *zero-inflated*, em inglês

o primeiro estado a distribuição probabilística pode ser uma das habitualmente consideradas – Poisson ou Binomial Negativa, tomando a designação de Zero-Inflacionado Poisson (ZIP) e Zero-Inflacionado Binomial Negativa (ZIBN), respectivamente.

Este modelo de probabilidades binário pode ser representado por uma função probabilística em que q_i é a probabilidade do local i se situar no estado de “zero acidentes” e $(1-q_i)$ é a probabilidade de ocorrer o estado “normal”, segundo uma distribuição de Poisson ou BN (Kumara, S. P. and Chin, H. C., 2003). Matematicamente este duplo estado pode ser representado da seguinte forma (Kumara, S. P. and Chin, H. C., 2003):

$$P(n_i) = \begin{cases} q_i + (1-q_i)R_i(0) & n_i = 0 \\ (1-q_i)R_i(n_i) & n_i > 0 \end{cases} \quad (4.15)$$

em que n_i é o número anual de acidentes registados no local, $R_i(n_i)$ é a probabilidade de ocorrência de n_i acidentes (distribuição de Poisson ou BN).

Os modelos ZI podem ser formulados de duas formas: assumir as mesmas variáveis independentes para os dois estados ou considerar diferentes variáveis. Assim, por exemplo considerando o caso da distribuição BN, o modelo ZIBN (τ) é formulado assumindo que as variáveis independentes que afectam o valor esperado μ_i do estado normal estão relacionadas com a probabilidade q_i (estado “zero acidentes”) por um parâmetro τ tal que $\text{logit}(q_i) = \tau\beta x_i$, em que β são os parâmetros a serem estimados e x_i as variáveis independentes consideradas.

Existe contudo, a dificuldade de avaliar qual a causa do excesso de zeros, ou seja, se tem origem numa natural heterogeneidade e como tal captada pela sobre-dispersão (regressão BN), ou se resulta de um processo de dois estados (regressão ZIBN). Nesse sentido, Vuong propôs um teste para avaliar se existe ou não um regime de separação de estados:

$$V = N^{1/2} \bar{m} / s_m \quad (4.16)$$

em que $\bar{m}$, s_m , e N são a média, desvio padrão e o número de observações, respectivamente de m_i , sendo que $m_i = \log[f_1(n_i) / f_2(n_i)]$ em que $f_1(.)$ e $f_2(.)$ são as funções de densidade das distribuições ZIBN(τ) e BN respectivamente. O teste Vuong considera a distribuição Normal e, por isso, o seu valor pode ser comparado com o valor crítico da distribuição Normal. Assim, para valores do teste Vuong superiores a 1,96 a escolha do modelo é

favorável à distribuição ZI. Valores inferiores ou iguais a -1,96 favorecem a opção da distribuição BN. Para os valores entre 1,96 e -1,96, o teste é inconclusivo.

A escolha do modelo mais ajustável depende assim do teste Vuong e do teste *t-student* do parâmetro de sobre-dispersão.

Kumara, S. P. and Chin, H. C. (2003) aplicaram o modelo ZIBN a intersecções para identificar factores geométricos e do tráfego que influenciem a ocorrência de acidentes. O objectivo dos autores ao aplicar o modelo ZIBN é distinguir as intersecções virtualmente seguras, isto é, diferenciar as intersecções que têm um registo de zero ocorrências resultado de um período excepcional ou do não registo de acidentes, das intersecções com registo de zero acidentes devido às suas características geométrica e funcionais “seguras”. Das intersecções seleccionadas para o estudo, 80,5% não têm registo de nenhum acidente no período em estudo. Da aplicação do modelo ZIBN resultou um valor de 4,98 pelo teste Vuong. Este valor é indicativo de que o modelo ZINB é mais ajustado por detectar os zeros em excesso, não correspondendo estes a uma sobre-dispersão. O parâmetro τ revelou ser estatisticamente significativo, indicando que as variáveis que afectam o valor esperado λ_i (calculado pela distribuição BN) estão implicitamente relacionadas com as variáveis do estado de zero acidentes.

Carson, J. and Mannering, F. (2001) também aplicaram o modelo ZIBN para analisar a relação entre a frequência de acidentes devido ao gelo no pavimento e a colocação de sinalização de aviso de gelo. Os autores desenvolveram modelos para três tipos de categorias de vias – inter-estaduais, artérias principais e artérias secundárias²². Apenas no modelo para as vias inter-estaduais é que o ZIBN revelou ser mais apropriado relativamente à regressão BN.

Apesar destes modelos ZI serem recentemente aplicados em diversos estudos, alguns autores colocam algumas reservas na sua aplicação aos modelos de acidentes. Para aprofundar esta questão, e assumindo a dificuldade na selecção do modelo a aplicar, Lord, D., et al. (2005b) compararam as várias abordagens possíveis de aplicar – Poisson, BN, ZIP e ZIBN. Considerando os princípios teóricos que fundamentam o processo de geração de acidentes (o acidente é o resultado “sucesso” de um ensaio de Bernoulli), os autores simularam um conjunto de dados. Desta análise, concluíram que o “excesso” de zeros observados não é indicativo da existência de um processo de geração de acidentes de dois

²² *interstate, principal arterial, minor arterial*, em inglês

estados, mas devido à presença de uma ou mais das seguintes condições: locais ou unidades de tempo demasiado pequenas; não participação de acidentes efectivamente ocorridos, originando registos incompletos; locais heterogéneos, com pequena exposição e elevado risco de acidente; ou omissão de importantes variáveis independentes. Assim, no desenvolvimento de um modelo estas hipóteses deverão ser analisadas e eventualmente corrigidas.

Estimação dos parâmetros da função $f(x)$

Para regressões não lineares, como é o caso da regressão de Poisson e da BN, a técnica de estimação dos parâmetros usualmente aplicada é o método da máxima verosimilhança. A ideia-base do método da máxima verosimilhança consiste em seleccionar, entre todos os valores possíveis dos parâmetros populacionais, aqueles que tornem mais verosímil a ocorrência de uma amostra idêntica àquela que efectivamente se obteve (Guimarães, R. C. and Cabral, J. S., 1997).

Assim, a probabilidade de ocorrência dos valores observados de uma amostra aleatória simples idêntica àquela que efectivamente se obteve a partir da população é dada por (distribuições discretas) (Guimarães, R. C. and Cabral, J. S., 1997):

$$L = \text{Pr obabilidade}(Y_1 = y_1, \dots, Y_N = y_N | \beta_1, \beta_2, \dots, \beta_R) \quad (4.17)$$

As estimativas de máxima verosimilhança dos parâmetros $\beta_1, \beta_2, \dots, \beta_R$ são os valores destes parâmetros que maximizam a função de verosimilhança. Estas estimativas podem ser obtidas resolvendo o sistema de equações definido por (Guimarães, R. C. and Cabral, J. S., 1997):

$$\frac{\partial L}{\partial \beta_i} = 0 \quad (i=1, \dots, R) \quad (4.18)$$

No caso, por exemplo, da regressão de Poisson, a função logarítmica de verosimilhança é (Greene, W. H., 2008):

$$\ln L = \sum_{i=1}^n [-\lambda_i + y_i x_i \beta - \ln y_i!] \quad (4.19)$$

Nestas distribuições o processo de estimativa dos parâmetros implica a resolução de sistemas de equações não lineares, iterativamente até que os critérios de convergência sejam satisfeitos (Cardoso, J., 2007).

No caso de se verificar sobre-dispersão (distribuição BN), as estimativas serão quase-verosimilhanes (Cardoso, J., 2007). O procedimento é idêntico tomando para a função de probabilidade a equação (4.10).

Para proporcionar um maior conhecimento das implicações dos resultados da estimação dos parâmetros, calcula-se a elasticidade para determinar os impactes das variáveis independentes na variável de resposta. A elasticidade corresponde a uma estimação do impacte de uma variável no valor percentual esperado da frequência dos acidentes e é interpretada como o efeito no valor esperado da frequência dos acidentes μ_i resultante de uma alteração de 1% na variável.

Assim, a elasticidade de uma variável contínua é definida por (Washington, S., et al., 2003):

$$E_{x_{ij}}^{\mu_i} = \frac{\partial \mu_i}{\mu_i} \times \frac{x_{ij}}{\partial x_{ij}} = \beta_j x_{ij} \quad (4.20)$$

em que E representa a elasticidade, x_{ij} é o valor da variável independente j para a observação i , β_j é o parâmetro a estimar da variável j e μ_i é o valor esperado da frequência de acidentes para a observação i .

Quando as variáveis independentes x_j são “*dummys*”, isto é, não são contínuas e tomam valores de 0 ou 1, os parâmetros destas variáveis medem aproximadamente o efeito relativo no valor esperado $E[Y]$ e designa-se de pseudo-elasticidade (Fridstrom, L., et al., 1995; Washington, S., et al., 2003):

$$E_{x_{ij}}^{\mu_i} = \frac{e^{\beta_j} - 1}{e^{\beta_j}} \quad (4.21)$$

4.2 Modelos de resposta qualitativa

Os modelos referidos anteriormente são modelos aplicados a dados discretos (números inteiros maiores que zero) que representam, no caso da sinistralidade, um número de acidentes, desagregados ou não, por tipo. Contudo, existe outro tipo de modelos probabilísticos aplicados também a dados discretos mas de resposta qualitativa²³, isto é, a variável de resposta tem um valor que corresponde a um código que identifica uma categoria à qual é associado um determinado significado. As categorias podem ser apresentadas de diversas formas tais como:

- decisão entre duas escolhas qualitativas - “sim” - 0 e “não” - 1;
- ordenação das respostas qualitativas - 0 – “baixo”, 1 – “médio”, 2 – “alto”, em que a diferença da resposta entre 0 e 1 pode não ser necessariamente a mesma entre 1 e 2;
- categorização sem seguir nenhuma ordem de valor - 0 – “engenheiro”, 1- “advogado”, 2- “professor”.

Este tipo de modelos probabilísticos tem a seguinte estrutura geral (Greene, W. H., 2008):

$$P(\text{evento } j \text{ ocorrer}) = P(Y=j) = F[\text{factores relevantes}, \text{parâmetros}] \quad (4.22)$$

em que um “evento” corresponde a uma categoria entre um conjunto de possíveis categorias.

Este tipo de modelos foi amplamente aplicado na análise dos transportes, no estudo, por exemplo, da repartição modal. Só recentemente, foram também aplicados ao estudo da sinistralidade, fundamentalmente na modelação da gravidade das vítimas dos acidentes, considerando um código para os diferentes níveis de gravidade.

O modelo de resposta qualitativa inicialmente aplicado ao estudo da sinistralidade foi o modelo binário²⁴. Este modelo aplica-se quando a variável de resposta tem dois resultados possíveis. No caso da sinistralidade, a variável de resposta habitualmente considerada para o modelo binário é uma das duas hipóteses: “acidente com vítima” ou “acidente sem

²³ *qualitative response models*, em inglês

²⁴ *binary model*, em inglês

vítima”. Utiliza-se o código 0 para identificar um “insucesso” e o código 1 para “sucesso”, em que sucesso significa a ocorrência de um determinado evento a estudar.

O modelo binário é representado matematicamente por (Greene, W. H., 2008):

$$\begin{aligned} P(Y=1|x) &= F(x,\beta) \\ P(Y=0|x) &= 1 - F(x,\beta) \end{aligned} \quad (4.23)$$

em que x corresponde às variáveis independentes, β os respectivos parâmetros e $F(.)$ é uma função específica para assegurar que $0 \leq P(Y) \leq 1$.

Para o cálculo da probabilidade da componente aleatória qualquer distribuição probabilística contínua é suficiente. As mais usualmente aplicadas são, no entanto, as distribuições simétricas normal e logística (Greene, W. H., 2008).

Quando a probabilidade é calculada pela distribuição normal acumulada o modelo designa-se de probit e tem a seguinte representação matemática (Greene, W. H., 2008):

$$P(Y=1|x) = \int_{-\infty}^{x\beta} \phi(t) dt = \Phi(x\beta) \quad (4.24)$$

em que a função $\Phi(.)$ é a notação considerada para a função da distribuição normal acumulada.

O modelo cuja componente aleatória tem uma distribuição logística é designado de modelo logit. Este modelo tem sido recentemente muito aplicado devido, essencialmente, à fácil interpretação dos seus parâmetros. Este modelo é representado matematicamente por (Greene, W. H., 2008):

$$P(Y=1|x) = \frac{e^{x\beta}}{1 + e^{x\beta}} = \Lambda(x\beta) \quad (4.25)$$

A notação $\Lambda(.)$ identifica a função de distribuição logística acumulada.

As duas distribuições referidas são ambas contínuas e simétricas diferindo apenas nas “caudas”, isto é, a distribuição logística decresce quando a probabilidade de “sucesso” tende para zero e cresce quando a probabilidade de sucesso tende para um mais rapidamente que a distribuição normal. Os resultados obtidos por estes dois modelos não diferem muito excepto quando a amostra de dados contenha um conjunto muito reduzido de respostas $Y=1$ ou $Y=0$, e uma ampla variação dos valores relativos a uma variável independente com um efeito significativo na resposta do modelo. Por questões de

conveniência matemática, poderão existir razões práticas para preferir um dos referidos modelos, sendo, no entanto, difícil justificar essa escolha com base em considerações teóricas (Greene, W. H., 2008).

O modelo binário é construído a partir de uma regressão latente²⁵ em que um conjunto de variáveis independentes x explica a decisão entre as duas alternativas e os respectivos parâmetros β reflectem o impacte na probabilidade das variações dos valores de x (Greene, W. H., 2008).

Nos modelos de resposta qualitativa, a regressão latente reflecte uma variável não observada y^* , obtida por:

$$y^* = x\beta + \varepsilon \quad (4.26)$$

em que $x\beta$ é designado de função índex. O termo de erro aleatório não observado ε , no caso dos modelos probit, é normalmente distribuído na amostra, com média zero e variância um. No caso dos modelos logit, o ε assume a distribuição logística padrão com média zero e variância $\pi^2/3$.

A variável observada é y , tal que:

$$y = 1 \text{ se } y^* > 0,$$

$$y = 0 \text{ se } y^* \leq 0. \quad (4.27)$$

Nesta expressão assumem-se duas simplificações. Por um lado, considera-se que a variância de ε é conhecida, mas de facto não é possível determiná-la a partir dos dados. Esta normalização não altera os valores observados pois se considerarmos $y^* = x\beta + \sigma\varepsilon$ e portanto $(y^*/\sigma) = x(\beta/\sigma) + \varepsilon$, resulta no mesmo modelo com os mesmos dados. Por outro lado, considera-se o limite de zero para y^* . Esta última simplificação também não produz alterações significativas no modelo se se considerar um termo da constante (o que é aconselhável neste tipo de modelos). Isto é, se considerássemos a em vez de zero e um termo de constante teríamos a expressão:

$\beta_0 + x\beta + \varepsilon > a$ que é igual a $(\beta_0 - a) + x\beta + \varepsilon > 0$, ou seja como o termo da constante é desconhecido o termo $(\beta_0 - a)$ mantém-se desconhecido (Greene, W. H., 2008).

A estimação dos modelos binários é geralmente baseada no método da máxima verosimilhança (Greene, W. H., 2008). Cada observação é tratada como um ensaio único

²⁵ *latent regression*, em inglês

de uma distribuição de Bernoulli. Ao modelo com probabilidade de “sucesso” - $F(x\beta)$, e para um conjunto de n observações independentes, corresponde a função logarítmica de verosimilhança (Greene, W. H., 2008):

$$\ln L = \sum_{i=1}^n \{y_i \ln F(x\beta) + (1 - y_i) \ln [1 - F(x\beta)]\} \quad (4.28)$$

Na análise dos modelos probabilísticos não lineares, como é o caso do modelo binário, é importante ter em conta que os parâmetros β estimados não representam o efeito marginal das variáveis independentes consideradas no modelo. Por esse facto, quando se pretende analisar o efeito da variação dos valores das variáveis na probabilidade calculam-se os efeitos marginais. Assim, generalizando, os efeitos marginais são calculados da seguinte forma:

$$\frac{\partial F(x\beta)}{\partial x} = \left\{ \frac{dF(x\beta)}{d(x\beta)} \right\} \beta = f(x\beta) \beta \quad (4.29)$$

em que $f(.)$ é a função densidade correspondente à função de distribuição (ou de probabilidade acumulada) $F(.)$. Assim, no caso do modelo probit, o efeito marginal é dado pela equação:

$$\frac{\partial F(x\beta)}{\partial x} = \phi(x\beta) \beta \quad (4.30)$$

em que $\phi(t)$ é a densidade normal padrão.

Para o modelo logit o efeito marginal corresponde a:

$$\frac{\partial F(x\beta)}{\partial x} = \Lambda(x\beta) [1 - \Lambda(x\beta)] \beta \quad (4.31)$$

Os efeitos marginais variam em função dos diferentes valores de x . Assim, para interpretar as estimações do modelo, é habitual calcular os efeitos marginais para o valor médio das variáveis independentes x .

O modelo binário tem, no entanto, a limitação de considerar apenas duas categorias de resposta. Para modelos de resposta qualitativa com múltiplas categorias e com uma determinada natureza ordenada, utiliza-se o modelo ordenado²⁶ (MO). Este modelo é de decisão multivariada em que a variável dependente assume valores que correspondem a

²⁶ *ordered model*, em inglês

uma ordenação das categorias, não necessariamente de forma linear, mas definindo uma ordem para os possíveis resultados.

Tal como os modelos binários, os modelos ordenados têm uma longa história no contexto dos transportes, aplicados, por exemplo, ao estudo do mecanismo comportamental associado à decisão de aquisição de veículo próprio (Bhat, C. R. and Pulugurta, V., 1997). No entanto, na análise da sinistralidade só muito recentemente têm vindo a ser aplicados (Eluru, N. and Bhat, C. R., 2007). No caso da sinistralidade, estes modelos são utilizados para a análise da gravidade dos acidentes quando estes são registados em múltiplas categorias tais como: “danos materiais”, “ferido leve”, “ferido grave” e “morto”. Ao considerar as várias categorias existentes para a classificação da gravidade dos acidentes e definindo uma ordenação dessas categorias, o MO introduz mais informação no estudo da sinistralidade. Nesse sentido, a aplicação desta metodologia tem crescido nos últimos tempos reconhecendo a natureza seriada da gravidade dos acidentes registados.

O modelo ordenado, tal como o modelo binário, pode ser utilizado com uma estrutura logit ou probit para a componente aleatória da regressão latente, conforme o tipo de distribuição considerada, isto é, distribuição logística ou normal, respectivamente. Tal como no modelo binário, o MO é estruturado à volta de uma regressão latente do tipo:

$$y^* = x\beta + \varepsilon \quad (4.32)$$

Como no caso dos modelos binários, y^* é não observado. O que se observa é representado por y :

$$y = 0 \text{ se } y^* \leq 0$$

$$y = 1 \text{ se } 0 < y^* \leq \mu_1$$

$$y = 2 \text{ se } \mu_1 < y^* \leq \mu_2$$

...

$$y = J \text{ se } \mu_{J-1} \leq y^* \quad (4.33)$$

em que μ são parâmetros desconhecidos a estimar juntamente com os parâmetros β . O gráfico da Figura 4.1 mostra as implicações da estrutura do MO para um caso em que $J=3$.

No caso dos modelos probit, o termo de erro ε é normalmente distribuído na amostra. Assim, as probabilidades para o modelo probit ordenado são (Greene, W. H., 2008):

$$\begin{aligned}
 P(y=0|x) &= \Phi(-x\beta), \\
 P(y=1|x) &= \Phi(\mu_1 - x\beta) - \Phi(-x\beta), \\
 P(y=2|x) &= \Phi(\mu_2 - x\beta) - \Phi(\mu_1 - x\beta), \\
 &\dots \\
 P(y=J|x) &= 1 - \Phi(\mu_{J-1} - x\beta)
 \end{aligned} \tag{4.34}$$

Para que todas as probabilidades sejam positivas é necessário que $0 < \mu_1 < \mu_2 < \dots < \mu_{J-1}$.

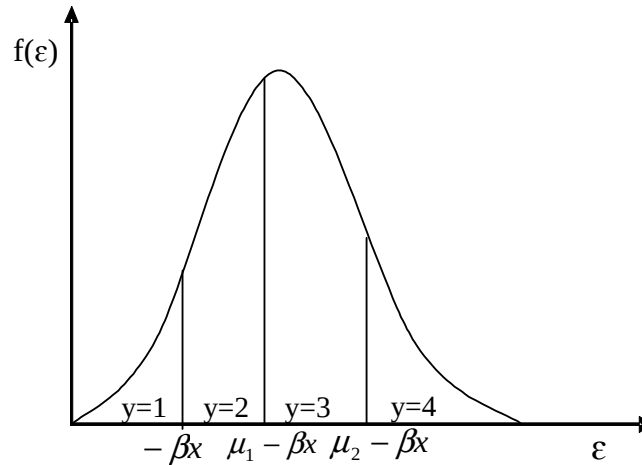


Figura 4.1 - Gráfico representativo de um modelo ordenado para $J = 4$

No caso do modelo logit ordenado a formulação é a mesma alterando-se apenas o tipo de distribuição para a distribuição logística - $\Lambda(\cdot)$.

Para a estimação dos parâmetros desconhecidos do MO utiliza-se o método da máxima verossimilhança (Yamamoto, T., et al., 2008):

$$\max \ln L = \sum_{i=1}^N \ln P(y_i = j_i | x_i) = \sum_{j=1}^J \sum_{i \in (j_i=j)} \ln \{ \Phi(\mu_j - \beta x_i) - \Phi(\mu_{j-1} - \beta x_i) \} \tag{4.35}$$

É de realçar que também no MO os efeitos marginais resultantes da variação dos valores das variáveis independentes x não são necessariamente iguais aos parâmetros β . Para o caso apresentado na Figura 4.1 em que $J = 3$, os efeitos marginais resultantes das variações dos valores das variáveis x são (Greene, W. H., 2008):

$$\begin{aligned}
\frac{\partial P(y=0|x)}{\partial x} &= -\phi(x\beta)\beta, \\
\frac{\partial P(y=1|x)}{\partial x} &= [\phi(-x\beta) - \phi(\mu_1 - x\beta)]\beta, \\
\frac{\partial P(y=2|x)}{\partial x} &= [\phi(\mu_1 - \beta x) - \phi(\mu_2 - \beta x)]\beta, \\
\frac{\partial P(y=3|x)}{\partial x} &= \phi(\mu_2 - \beta x)\beta
\end{aligned} \tag{4.36}$$

Como se pode constatar da equação (4.36), a variação dos efeitos marginais da primeira e última categoria está directamente relacionada com os sinais dos parâmetros β . Nas categorias intermédias o efeito do sinal dos parâmetros é mais ambíguo.

A aplicação do MO à sinistralidade tem, no entanto, uma condicionante. A existência de uma regressão latente (equação (4.32)) para todas as categorias de resposta pressupõe considerar as mesmas variáveis independentes e o mesmo tipo de efeito na variável dependente, assumindo como tal uma correlação entre as diferentes categorias. Por esse facto, alguns investigadores optaram por considerar modelos não ordenados²⁷ que possibilitam considerar para as várias categorias diferentes relações entre a variável dependente e as variáveis independentes, pressupondo uma independência entre as categorias. Estes modelos, no entanto, e tal como o nome indica, não atribuem uma ordenação às categorias. É o caso do modelo multinomial²⁸ (MM). Este modelo de resposta não ordenada é fundamentado num modelo utilitário aleatório²⁹ do tipo (Carson, J. and Mannering, F., 2001):

$$S_{jn} = \beta_j X_n + \varepsilon_{jn} \tag{4.37}$$

em que β_j são os parâmetros a estimar, X_n são as variáveis independentes que representam as características observáveis que determinam a categoria e ε_{jn} é o termo de erro aleatório não observado. A equação (4.37) é designada de função utilitária, termo utilizado na economia para designar a capacidade para estimular o consumo. Neste caso, a função utilitária pode ser entendida como a função que reúne as condições, através das variáveis independentes, para influenciar a ocorrência de determinada categoria de sinistralidade.

²⁷ *unordered model*, em inglês

²⁸ *multinomial model*, em inglês

²⁹ *random utility model*, em inglês

A probabilidade de um evento n corresponder à categoria j é expressa pela probabilidade de S_{jn} ser maior que todas as outras categorias S_{Jn} , isto é:

$$P_n(j) = P(S_{jn} > S_{Jn}) \forall J \neq j \quad (4.38)$$

Assim, se substituirmos na equação (4.38) a equação (4.37), a primeira pode ser reescrita da seguinte forma:

$$P_n(j) = P(\beta_j X_n + \varepsilon_{jn} > \beta_J X_n + \varepsilon_{Jn}) \forall J \neq j \quad (4.39)$$

Também no MM podemos distinguir dois tipos de estrutura – probit e logit. O modelo probit é menos utilizado por implicar o cálculo de integrais múltiplos (Greene, W. H., 2008). No modelo logit, a distribuição habitualmente considerada é uma distribuição Gumbel (também designada de distribuição de valores extremos). É também uma distribuição exponencial, tal como a distribuição logística, mas assimétrica. Assim, no caso do modelo multinomial logit (MML), e considerando que J termos de erro ε são independentes e com a mesma distribuição, a probabilidade de um indivíduo (ou observação) n escolher a categoria j é calculada por:

$$P_n(j) = \frac{\exp(\beta_j X_n)}{\sum_J \exp(\beta_J X_n)} \quad (4.40)$$

Esta equação determina um conjunto de $J+1$ probabilidades. Os valores dos parâmetros β podem ser estimados através do método padrão de máxima verosimilhança. Assim, a função logarítmica de verosimilhança a ser maximizada é:

$$\ln L = \sum_{n=1}^N \sum_{j=1}^J y_{nj} \ln P_n(j) \quad (4.41)$$

em que $P_n(j)$ é dado pela equação (4.40) e y_{nj} é definido por:

$y_{nj} = 1$ se a categoria j é escolhida pelo indivíduo n ,

$y_{nj} = 0$ caso contrário.

Como em qualquer modelo não linear é necessário algum cuidado na interpretação dos parâmetros estimados. Isto é especialmente verdade no caso dos modelos multinomiais em que, ao contrário do MO, não se verifica uma correspondência directa entre o sinal do parâmetro estimado e o efeito marginal. Assim, para analisar a relação das variáveis independentes com a variável dependente calculam-se os efeitos marginais. O efeito

marginal permite determinar a variação estimada da probabilidade de um indivíduo n escolher a categoria j , quando a variável independente varia. Para cada variável independente o efeito marginal é calculado por:

$$\frac{\partial P_n(j)}{\partial X_n} = P_n(j) \left[\beta_j - \sum_{k=0}^J P_n(k) \beta_k \right] = P_n(j) [\beta_j - \bar{\beta}] \quad (4.42)$$

Como atrás se referiu o MML pressupõe a independência entre as diferentes categorias. No entanto, para algumas aplicações é importante considerar uma correlação entre as categorias. Com esse objectivo, foi desenvolvida uma extensão do MML designado de modelo hierárquico³⁰ logit. A estrutura deste modelo consiste em agrupar as várias categorias J em subgrupos G permitindo que a variância seja diferente consoante os grupos mantendo, no entanto, a suposição de independência dentro dos grupos. O processo de decisão consiste na observação n escolher uma categoria de um conjunto de G categorias e a seguir fazer uma escolha específica j dentro da categoria escolhida g , $g=1, \dots, G$. Assim, para dois níveis de decisão a forma matemática para o modelo hierárquico logit é (Abdel-Aty, M. and Abdelwahab, H., 2004):

$$\begin{aligned} P_n(j, g) &= P_n(g) P_n(j|g), j = 1, \dots, J-1 \\ P_n(j|g) &= \frac{\exp(\beta_j X_n | g)}{\sum_{j \in g} \exp(\beta_j X_n | g)}, \\ P_n(g) &= \frac{\exp(\gamma_g M_g + \tau_g I_g)}{\sum_{g=1}^G \exp(\gamma_g M_g + \tau_g I_g)}, \\ P_n(J) &= 1 - \sum_{j=1}^{J-1} P_n(j) \end{aligned} \quad (4.43)$$

em que $P_n(j, g)$ é a probabilidade de uma observação n pertencer à categoria j do grupo g , $P_n(j|g)$ é a probabilidade condicional de uma observação n pertencer à categoria j sabendo que esta pertence ao grupo g , $P_n(g)$ é a probabilidade da observação n pertencer ao grupo g , X_n são as características medidas que determinam a posição da observação, M_g são as características do grupo, $I_g = \ln \sum_{j \in g} \exp(\beta_j X_n | g)$ e é designado de valor inclusivo do grupo g ; β_j e γ_g são os parâmetros a serem estimados e τ_g é o parâmetro do valor inclusivo do grupo g . Ao restringir todos os valores inclusivos a 1, o modelo hierárquico logit é reduzido ao MML. O modelo hierárquico pode ser facilmente estendido para mais do que

³⁰ *nested*, em inglês

dois níveis. A complexidade do modelo aumenta geometricamente com o número de níveis. Contudo, a aplicação do modelo tem-se mostrado bastante flexível e como tal é muito aplicado na modelação de transportes (Abdel-Aty, M. and Abdelwahab, H., 2004)

A estimação dos parâmetros do modelo hierárquico logit pode ser realizada de duas formas distintas. O primeiro método designa-se de máxima verosimilhança com informação limitada³¹ e é realizado em várias fases: estima-se β tratando a decisão dentro da subdivisão, ou seja, como se se tratasse de um MML simples; calculam-se os valores inclusivos para todas as subdivisões do modelo e por fim estimam-se os parâmetros considerando a decisão entre as subdivisões como um MML simples.

O segundo método designa-se de máxima verosimilhança com informação completa³² e consiste em estimar o modelo numa única fase. Em geral, o modelo resultante é mais eficiente (Abdel-Aty, M. and Abdelwahab, H., 2004).

Para analisar as variáveis incluídas no modelo, tal como nos modelos anteriores, são calculados os efeitos marginais. A fórmula de cálculo do efeito marginal, para uma estrutura de dois níveis, de um atributo r da decisão j de um grupo g na probabilidade da categoria j^* do grupo g^* é:

$$\frac{\partial \ln P(j^*, g^*)}{\partial r_{(j,g)}} = \beta_{r(j,g)} \{1_{(g=g^*)} [1_{(j=j^*)} - P(j^*|g)] + \tau_g [1_{g=g^*} - P(g^*)] P(j|g)\} \quad (4.44)$$

em que $1_{(g=g^*)} = 1$, se $g=g^*$, e 0 , caso contrário, $1_{(j=j^*)} = 1$, se $j=j^*$, e 0 , caso contrário, $P(j)$ é a probabilidade da categoria j , $P(g)$ é a probabilidade do grupo g , e $\beta_{r(j,g)}$ é o parâmetro estimado correspondente ao atributo r .

Tal como nos modelos de resposta quantitativa referidos em 4.1, também nos modelos de resposta qualitativa aqui referidos se aplica a análise dos dados em painel. É o caso dos modelos com efeito fixo e efeito aleatório que consideram a heterogeneidade do termo de erro (Greene, W. H., 2008):

$$y_{it}^* = x\beta + v_{it} + u_i; i=1,...,n; t=1,...,T_i \quad (4.45)$$

em que no caso, por exemplo, dos modelos binários teríamos:

$$y_{it} = 1 \text{ se } y_{it}^* > 0, \text{ senão } y_{it} = 0$$

³¹ *limited information maximum likelihood*, em inglês

³² *full information maximum likelihood*, em inglês

e em que u_i representa a heterogeneidade individual específica não observada. Mais uma vez, a distinção entre os modelos de efeito fixo e aleatório é feita pela relação entre u_i e x_{it} . Assim, se u_i é não relacionado com x_{it} de tal forma que a distribuição condicional $f(u_i | x_{it})$ é independente de x_{it} então o modelo é de efeito aleatório. Se a distribuição não é condicional de tal forma que u_i pode ser correlacionado com x_{it} então o modelo é de efeito fixo.

Esta heterogeneidade pode ser estendida a todos os parâmetros do modelo de forma a permitir que esses parâmetros sejam distribuídos continuamente entre observações. Esta abordagem designa-se de parâmetros aleatórios ou modelos mistos³³.

A inclusão da heterogeneidade do termo do erro ou dos parâmetros nos modelos é particularmente importante nos modelos ordenados, devido à rigidez do modelo em fixar as mesmas características (isto é, variáveis independentes) para todas as categorias.

4.3 Conclusões

O estudo da sinistralidade pode ser realizado mediante duas abordagens diferentes em função do tipo de resposta pretendida. Considerando a natureza numérica da frequência dos acidentes, há já algumas décadas que se tem desenvolvido e aplicado modelos de resposta quantitativa. No âmbito deste trabalho, foi realizada uma selecção de bibliografia considerada representativa da aplicação dos modelos de resposta quantitativa e que se pode consultar no Quadro 4.1.

Os modelos de resposta quantitativa aplicam-se diferenciando o local das ocorrências, geralmente, pelo tipo geometria (nós ou arcos) e pelo tipo de hierarquia viária (auto-estrada, vias distribuidoras, etc.), e destinam-se principalmente à análise de factores de risco de acidentes (características geométricas, tráfego, etc.) através da introdução de um conjunto de variáveis independentes. Pelo Quadro 4.1 pode-se constatar que a variável independente comum a todos os modelos é a variável de exposição que quantifica o volume de tráfego, normalmente através do TMDA. Pela sua importância nestes modelos, Lord, D., *et al.* (2005a) estudaram as diferentes formas de a considerar nos modelos

³³ *mixed models*, em inglês

chegando à conclusão que, no entanto, considerar o tráfego como única variável não é a melhor forma de caracterizar o processo de ocorrências. A variável dependente mais considerada corresponde ao número de acidentes, eventualmente desagregado por tipologia e pela gravidade do acidente através do registo das vítimas.

Quadro 4.1 - Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta quantitativa

Bibliografia	Objectivo	Entidades	Variável dependente	Fontes dos dados	Variáveis independentes	Estrutura do erro
(Fridstrom et al., 1995)	•analisar a influência da aleatoriedade nos acidentes	•arcos	•nº acid c/ vítimas	dados de 4 países - Dinamarca (1977-1987), Finlândia (1975-1987), Noruega (1973-1986), Suécia (1976-1987)	•exposição	•Poisson
					•registo acidentes	
			•nº acid mortais		•legislação	•BN
			•nº vítimas		•condições atmosféricas	
(Cardoso, J., 1996)	•analisar factores de risco nos acidentes	•arcos (meio rural)	•nº acid c/ vítimas (excepto atropelamentos e colisões c/ animais)	DGV (1988-1993)	•tendencia temporal	•BN
					•identificação do país e região e da época do ano	
					•velocidade	
(Mountain et al., 1996)	•desenvolver um modelo para arcos das vias principais com intersecções com vias secundárias	•arcos-estradas com 1 ou 2 vias (meio urbano e rural)	•nº acid c/ vítimas	várias províncias do Reino Unido (períodos de 5 a 15 anos)	•características geométricas	•BN
					•comportamento condutor	
					•tráfego	
(Poch and Mannering, 1996)	•analisar factores de risco nos acidentes	•nós (meio urbano)	•nº acid total	Washington, USA (1987-1993)	•nº intersecções no arco	•Poisson
			•nº acid por colisão traseira		•comprimento do arco	
			•nº acidentes por colisão lateral e por mudança de direcção			
(Turner and Nicholson, 1997)	•analisar qual o modelo que melhor reflecte o funcionamento do nó	•nós (meio urbano)	•nº acid total	New Zealand (período de 5 anos)	•tráfego	•BN
			•nº acid total por tipo de movimento no nó		•características geométricas e funcionais da via	
(Mountain et al., 1998)	•analisar o efeito da tendencia temporal nos acidentes em diferentes tipos de modelos	•nós	•nº acid c/ vítimas	várias províncias do Reino Unido (períodos de 5 a 15 anos entre 1980-1994)	•tráfego	•BN
			•nº acid por tipo (condições pavimento e condições de iluminação)		•tipo de controlo	
					•nº ramos	
(Ivan et al., 1999)	•analisar factores de risco nos acidentes	•arcos- estradas com 2 vias (meio rural)	•nº acid com 1 veic	Connecticut Department of Transportation, USA (Out 1990-Set 1995)	•limite de velocidade	•Poisson
			•nº acid com mais do que 1 veic			
(Ivan et al., 2000)	•analisar factores de risco nos acidentes	•arcos- estradas com 2 vias (meio rural)	•taxa acid c/ 1 veic	Connecticut Department of Transportation, USA (Out 1990-Set 1995)	•tráfego	•Poisson
			•taxa acid c/ mais do que 1 veic		•densidade tráfego	
					•ocupação do solo	
					•condições de iluminação e altura do dia	

Quadro 4.1 – Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta quantitativa (continuação)

Bibliografia	Objectivo	Entidades	Variável dependente	Fontes dos dados	Variáveis independentes	Estrutura do erro
(Sawalha and Sayed, 2001)	•analisar factores de risco nos acidentes	•arcos - vias distribuidoras (meio urbano)	•nº acid total	Greater Vancouver Regional District, British Columbia, Canada (1994-1996)	•tráfego •características geométricas	•Poisson •BN
(Carson and Mannering, 2001)	•avaliar o efeito da sinalização de pavimento c/ gelo	•arcos	•nº acid devido ao gelo	Washington State Department of Transportation, USA (1993-1995)	•tráfego •características geométricas •condições atmosféricas •características temporais e espaciais	•ZIP •ZIBN
(Lee and Mannering, 2002)	•analisar factores de risco na frequência dos acidentes por despiste com saída da via (auto-estrada)	•arcos - auto-estrada	•nº acid por despiste com saída da via	Washington State Department of Transportation, USA (1994-1996)	•características temporais •características geométricas da via e da envolvente	•ZIP •ZIBN
(Greibe, 2003)	•desenvolver modelos simples de previsão de acidentes em meio urbano	•nós (meio urbano) •arcos (meio urbano)	•nº acid total	official accident statistics database, Denmark (nós:1987-1991; arcos:1990-1994)	•tráfego •características geométricas da via	•Poisson
(Kumara and Chin, 2003)	•analisar factores de risco nos acidentes •distinguir entre os locais verdadeiramente ou virtualmente seguros através de uma análise dos	•nós 3 ramos	•nº acid total	National Road Accident Database, Singapura (1992-2000)	•tráfego •características geométricas e funcionais da via •tendência temporal	•BN •ZIBN
(Lord and Persaud, 2004)	•criar modelos de previsão de acidentes para uma rede digital urbana em fase de planeamento	•nós (meio urbano) •arcos (meio urbano)	•nº acid total	City of Toronto Transportation Department, Canada (1990-1995)	•tráfego	•BN
(Qin et al., 2004)	•analisar a relação entre a exposição e os acidentes	•arcos - estradas de 2 vias	•nº acid com 1 veic •nº acid com mais do que 1 veic na mesma direcção •nº acid com mais do que 1 veic em direcções •nº acid com mais do que 1 veic que se intersectam	Highway Safety Information System, Michigan, USA (1994-1997)	•tráfego •características geométricas da via	•ZIP
(Lord et al., 2005)	•avaliar como o tráfego (em diferentes variáveis) influência a ocorrência de acidentes •analisar diferentes formas funcionais para os modelos	•arcos - auto-estrada (meio urbano e meio rural)	•nº acid total •nº acid mortais •nº acid c/ 1 veic •nº acid c/ mais do que 1 veic	Quebec, Canada (1994-1998)	•tráfego •comprimento do arco	•BNG (com variação da dispersão)
(Caliendo et al., 2007)	•desenvolver um modelo de previsão para estradas de vias múltiplas	•arcos - auto-estradas com múltiplas vias	•nº acid total •nº acid c/ vítimas	Motorway Management Agency, Italy (1999-2003)	•tráfego •características geométricas da via •distância de visibilidade •atrito do pavimento •condições atmosféricas	•Poisson •BN •Multinomial Negativa c/ parâmetro dispersão proporcional do arco
(Mitra and Washington, 2007)	•analisar o efeito da variação do parâmetro de dispersão	•nós (meio rural)	•nº acid total	Georgia, USA (1996-1997)	•tráfego •características geométricas da via •condições de iluminação	•BNG (com variação da dispersão) •modelação Bayesiana
(Lord and Park, 2008)	•avaliar o efeito do parâmetro de dispersão na estimação EBayes •comparar os modelos BN e BN-sobredispersão fixa e variável	•nós 3 ramos (meio rural)	•nº acid total	Highway Safety Information System, California, USA (1997-2001)	•tráfego	•BN •BNG (com variação da dispersão)
(Anastasopoulos and Mannering, 2009)	•analisar modelos com parâmetros aleatórios	•arcos (meio rural)	•nº acid total	Indiana State Patrol, USA (1995-1999)	•tráfego •características geométricas da via •características do pavimento	•BN •BN c/ parâmetros aleatórios

Ao longo dos últimos anos em que estes modelos foram estudados, evoluiu-se da aplicação da distribuição de Poisson para uma distribuição BN na definição da estrutura do erro. Nos anos mais recentes, algumas extensões a esta distribuição foram ainda analisadas, tal como se pode observar pelo Quadro 4.1 na coluna referente à estrutura do erro, como é exemplo a variação do parâmetro de dispersão e a aleatoriedade dos parâmetros.

A segunda abordagem, geralmente considerada nos modelos de transporte, é mais recente na aplicação ao estudo da sinistralidade. Nesta abordagem, os modelos aplicam-se a uma resposta não observada e em que esta traduz uma categoria de sinistralidade - modelos de resposta qualitativa. O Quadro 4.2 apresenta alguns estudos com a aplicação destes modelos probabilísticos.

Quadro 4.2 - Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta qualitativa

Bibliografia	Objectivo	Entidades	Variável dependente	Fonte dos dados	Variáveis independentes	Modelo
(Carson and Mannering, 2001)	• analisar a eficiência da sinalização de presença de gelo no pavimento na redução de acidentes	• arcos	• vítimas: categorias - só com danos materiais; ferido; morto	Washington State Department of Transportation, USA (1993-1995)	• características do tráfego • características da via • identificação da região	• modelo multinomial logit
(Kockelman and Kweon, 2002)	• analisar factores de risco na gravidade no total dos acidentes, nos acidentes que envolvem 1 veículo e 2 veículos		• vítimas condutores: categorias: só com danos materiais; ferido ligeiro; ferido grave; morto	National Automotive Sampling System GES, USA (1998)	• características do condutor • características do veículo • características do acidente • condições atmosféricas	• modelo ordenado probit
(Lee and Mannering, 2002)	• analisar factores de risco na frequência e gravidade dos acidentes por despiste com saída da via (auto-estrada)	• arcos - auto-estrada	• vítimas: categorias: só com danos materiais; possibilidade de ferimentos; ferimentos; incapacitante; morto	Washington State Department of Transportation, USA (1994-1996)	• condições atmosféricas • características temporais • características do condutor • características da via • características da envolvente à via	• modelo nested logit
(Abdel-Aty, 2003)	• analisar factores de risco • comparar o modelo ordenado com os modelos multinomial e nested	• arcos • intersecções sinalizadas • zona de portagens	vítimas condutores: categorias: só com danos materiais; possibilidade de ferimentos; com ferimentos; ferido grave/morto	Florida, USA (1996-1997 para os arcos e intersecções sinalizadas; 1999-2000 para zona de portagens)	• características do condutor • utilização do cinto segurança • características do veículo • características do acidente • factores ambientais	• modelo ordenado probit • modelo multinomial logit • modelo nested logit
(Abdel-Aty and Abdelwahab, 2004)	• analisar a influência nos acidentes por colisão traseira da invisibilidade provocada pela geometria de determinados veículos (veic pesados ligeiros, SUV)		• acidentes por colisão traseira entre 2 veículos: categorias - veic ligeiro colide com veic ligeiro; veic ligeiro colide com veic pesado; veic pesado colide com veic ligeiro; veic pesado colide com veic pesado	General Estimates System, USA (2000)	• características e acção do condutor • tipo de veículo • condições atmosféricas • visibilidade do condutor • características da via	• modelo nested logit
(Sze and Wong, 2007)	• analisar factores de risco	• meio urbano	• vítimas peões: categorias - morto/ferido grave; ferido ligeiro	Traffic Accident Database System, Hong Kong, China (1991-2004)	• características do peão • características da lesão do peão • trajectória do peão • comportamento do peão • características temporais • características do tráfego e do local	• modelo binário logit

Quadro 4.2 – Referências bibliográficas de estudos da sinistralidade com base em modelos de resposta qualitativa (continuação)

Bibliografia	Objectivo	Entidades	Variável dependente	Fonte dos dados	Variáveis independentes	Modelo
(Qi et al., 2007)	•criar um modelo de previsão de acidentes para gestão de tráfego em tempo real	•meio urbano	•nº acidentes em auto-estradas: categorias - não acidente; 1 acidente; >1 acidente (por intervalo de tempo)	Hampton Roads Smart Traffic Center Freeway Incident Response Team, USA (Jul 1999-Dez 2001 para calibrar modelo e 2002-2003 para testar o modelo)	•condições atmosféricas •características do fluxo de tráfego •características geométricas	•modelo ordenado probit com efeito aleatório
(Eluru and Bhat, 2007)	•analisar a endogeneidade da variável uso de cinto de segurança e a gravidade das vítimas •analisar os efeitos de factores não observados na gravidade das vítimas		•vítimas: categorias - só com danos materiais; possibilidade de ferimentos; não incapacitante; incapacitante;morto	General Estimates System, USA (2003)	•características do condutor •características do veículo •características da via •factores ambientais •características do acidente	•modelo binário-ordenado com parâmetros aleatórios
(Eluru et al., 2008)	•analisar factores de risco •comparar o modelo generalizado ordenado c/ efeito aleatório com o modelo ordenado padrão		•vítimas peões: categorias - morto; com incapacidade; sem incapacidade; sem ferimentos •vítimas ciclistas: categorias - morto; com incapacidade; sem incapacidade; sem ferimentos	General Estimates System, USA (2004)	•características do peão e do ciclista •características do condutor •características do veículo •características da via •factores ambientais •características do acidente	•modelo generalizado ordenado com efeito aleatório •modelo ordenado
(Yamamoto et al., 2008)	•analisar a influência de registos incompletos na gravidade das vítimas de acidentes de 1 veículo com um obstáculo fixo		•vítimas condutores: categorias - só com danos materiais; possibilidade de ferimentos; com ferimentos; com incapacidade; morto	Washington State Department of Transportation, USA (Jan 1993-Jul 1996)	•condições da via •tipo de obstáculo fixo •sistemas de segurança do veículo •características do veículo •características do condutor	Binário probit sequencial modelo ordenado probit
(Wang and Abdel-Aty, 2008)	•analisar a gravidade dos acidentes com viragem à esquerda •comparar o modelo ordenado padrão com o modelo generalizado ordenado	•nós	•vítimas: categorias - só com danos materiais; possibilidade de ferimentos; não incapacitante; incapacitante;morto	Florida Department of Transportation Safety Office, USA (2000-2005)	•características do condutor •características do veículo •características da via •factores ambientais •características do acidente	•modelo generalizado ordenado logit •modelo ordenado logit
(Kim et al., 2008)	•analisar a hipótese da variância das características dos peões não observadas aumentar com a idade dos peões		•vítimas peões: categorias - possibilidade de ferimento ou não ferido; não incapacitante; incapacitante; morto	Highway Safety Research Center at the University of North Caroline, USA (1997-2000)	•características do peão •características do condutor •características temporais •características do controlo de tráfego •características da ocupação do solo •factores ambientais •características do veículo	•modelo multinomial com heteroscedasticidade •modelo multinomial
(Milton et al., 2008)	•analisar a distribuição da gravidade das vítimas em acidentes na auto-estrada	•arcos	•vítimas: categorias - só com danos materiais; possibilidade de ferimentos; com ferimentos	Washington State Department of Transportation, USA (1990-1994)	•características do tráfego •características da via •condições atmosféricas	•modelo multinomial com parâmetros aleatórios
(Gkritza and Mannering, 2008)	•analisar o uso do cinto de segurança nos acidentes com 1 ou 2 ocupantes		•veic c/ 1 ocupante: categorias - usa cinto segurança; não usa cinto segurança •veic c/ 2 ocupantes: categorias - 2 ocupantes usam cinto; nenhum usa; apenas o condutor usa; apenas o passageiro usa	Indiana Criminal Justice Institute, USA (Dez 2006)	•características do veículo •características da via •características dos ocupantes do veículo •características temporais	•modelo multinomial com parâmetros aleatórios

Deste quadro verifica-se que os modelos ordenados e os multinomiais são os mais aplicados. Segundo Wang, X. and Abdel-Aty, M. (2008), como os modelos ordenados têm uma estrutura simples são, por isso, mais fáceis de aplicar e analisar. O MO tem ainda a vantagem de permitir ordenar a gravidade dos acidentes por níveis que são correlacionados entre si. Contudo, a rigidez do modelo ao considerar os mesmos factores (variáveis independentes) para todos os níveis de gravidade, levou a que alguns autores optassem por aplicar modelos multinomiais. Abdel-Aty, M. (2003) aplicou, para a mesma base de dados, os três modelos habitualmente utilizados no estudo da gravidade dos acidentes - modelos ordenado, multinomial e hierárquico, com o objectivo de comparar os resultados. O autor concluiu que o modelo que produziu piores resultados foi o MML. Os resultados do modelo hierárquico foram ligeiramente melhores do que os do MO mas, tendo em conta que este último é de mais fácil aplicação e interpretação, o autor seleccionou este modelo para o estudo.

De salientar que a comparação entre os resultados destes modelos não pode ser feita pelos parâmetros estimados, pois estes têm estruturas diferentes. Essa comparação deve ser ao nível agregado através do efeito de elasticidade das variáveis independentes (Bhat, C. R. and Pulugurta, V., 1997).

Deste capítulo, pode concluir-se que existe um conjunto de técnicas estatísticas que permitem desenvolver modelos com elevada qualidade de ajuste, sendo, contudo, fundamental não esquecer que um modelo “correcto” deve, não só descrever o processo real que gera os acidentes mas também, e crucial, corresponder ao objectivo final pretendido (OCDE, 1997). Acrescente-se o facto do desenvolvimento do modelo “correcto” estar ainda condicionado à qualidade e quantidade de informação disponível.

4.4 Referências bibliográficas

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, **34**, 597-603.

Abdel-Aty, M. and Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. *Accident Analysis and Prevention*, **36**, 447-456.

Anastasopoulos, P. C. and Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters. *Accident Analysis and Prevention*, **41**, 153-159.

Bhat, C. R. and Pulugurta, V. (1997). A comparison of two alternative behavioral choice mechanism for household auto ownership decisions. *Transportation Research Part B*, **32**, 61-75.

Caliendo, C., Guida, M. and Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis and Prevention*, **39**, 657-670.

Cardoso, J., (1996) *Estudo das relações entre as características da estrada, a velocidade e os acidentes rodoviários. Aplicação a estradas de duas vias e dois sentidos*. Tese de Doutoramento (PhD), IST e LNEC, Lisboa.

Cardoso, J. (2007) *Métodos racionais de apoio à intervenção da engenharia em segurança rodoviária*. Programa de investigação para obtenção do título "Habilitado para o exercício de funções de Coordenação de Investigação Científica", LNEC, Lisboa.

Carson, J. and Mannering, F. (2001). The effect of ice warning signs on ice-accident frequencies and severities. *Accident Analysis and Prevention*, **33**, 99-109.

Eluru, N. and Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention*, **39**, 1037-1049.

Eluru, N., Bhat, C.R. and Hensher, D.A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, **40**, 1033-1054.

- Fridstrom, L., et al. (1995). Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, **27**, 1-20.
- Gkritza, K. and Mannering, F. L. (2008). Mixed logit analysis of safety-belt use in single and multi-occupant vehicles. *Accident Analysis and Prevention*, **40**, 443-451.
- Golob, T., Recker, W. and Alvarez, V. (2004). Freeway safety as a function of traffic flow. *Accident Analysis and Prevention*, **36**, 933-946.
- Greene, W. H. (2008). *Econometric Analysis*, Pearson International Edition, New Jersey.
- Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention*, **35**, 273-285.
- Guimarães, R. C. and Cabral, J. S. (1997). *Estatística*, McGraw-Hill, Lisboa etc.
- Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention*, **33**, 799-808.
- Hauer, E. and Bamfo, J. (1997). Two tools for finding what function links the dependent variable to the explanatory variables. In: *ICTCT*, Lund.
- Ivan, J. N., Pasupathy, R. K. and Ossenbruggen, P. J. (1999). Differences in causality factors for single and multi-vehicle crashes on two-lane roads. *Accident Analysis and Prevention*, **31**, 695-704.
- Ivan, J.N., Wang, C. and Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention*, **32**, 787-795.
- Kim, J.K., Ulfarsson, G.F. and Shankar, V.N. (2008). Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. *Accident Analysis and Prevention*, **40**, 1695-1702.

Kockelman, K.M. and Kweon, Y. J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis and Prevention*, **34**, 313-321.

Kumara, S. P. and Chin, H. C. (2003). Modeling accident occurrence at signalized Tee intersections with special emphasis on excess zeros. *Accident Analysis and Prevention*, **4**, 53-57.

Lee, J. and Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis and Prevention*, **34**, 149-161.

Lord, D. (2000) *The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models*. PhD, Department of Civil Engineering, University of Toronto, Toronto.

Lord, D. and Persaud, B.N. (2004). Estimating the safety performance of urban road transportation networks. *Accident Analysis and Prevention*, **36**, 609-620.

Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, **38**, 751-766.

Lord, D., Guikema, S. D. and Geedipally, S. R. (2008). Application of the Conway-Maxwell-Poisson generalized linear model for analysing motor vehicle crashes. *Accident Analysis and Prevention*, **40**, 1123-1134.

Lord, D., Manar, A. and Vizioli, A. (2005a). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis and Prevention*, **37**, 185-199.

Lord, D. and Park, P. Y. J. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis and Prevention*, **40**, 1441-1457.

- Lord, D., Washington, S. P. and Ivan, J. (2005b). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, **37**, 35-46.
- Maher, M. and Summersgill, I. (1995). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, **28**, 281-296.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman & Hall, London, etc.
- Milton, J., Shnakar, V. and Mannering, F. (2008). Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis and Prevention*, **40**, 260-266.
- Mitra, S. and Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis and Prevention*, **39**, 459-468.
- Mountain, L., Fawaz, B. and Jarrett, D. (1996). Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, **28** (6), 695-707.
- Mountain, L., Maher, M. and Fawaz, B. (1998). The influence of trend on estimates of accidents at junctions. *Accident Analysis and Prevention*, **30**(5), 641-649.
- OCDE (1997). *Road safety principles and models: review of descriptive, predictive, risk and accident consequence models*. Paris: Organisation for Economic Co-operation and Development, OCDE/GD(97)153.
- Poch, M. and Mannering, F. (1996). Negative Binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, **122**(2), 105-113.
- Polus, A., Pollatschek, M. and Farah, H. (2005). Impact of infrastructure characteristics on road crashes on two-lane highways. *Traffic Injury Prevention*, **6**, 240-247.

Qi, Y., Smith, B. L. and Guo, J. (2007). Freeway accident likelihood prediction using a panel data analysis approach. *Journal of Transportation Engineering*, **133**, 149-156.

Qin, X., Ivan, J. and Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, **36**, 183-191.

Reurings, M., et al. (2005). "Accident Prediction Models and Road Safety Impact Assessment: a state-of-the-art." RIPCORDER-ISEREST: Road Infrastructure Safety Protection - Core Research and development for Road Safety in Europe; Increasing Safety and Reliability of Secondary Roads for a Sustainable Surface Transport, 135.

Sawalha, Z. and Sayed, T. (2001). Evaluating safety of urban arterial roadways. *Journal of Transportation Engineering*, **127**(2), 151-158.

Sze, N.N. and Wong, S.C. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis and Prevention*, **39**, 1267-1278.

Turner, S. and Nicholson, A. (1997). Intersection accident estimation: the role of intersection location and non-collision flows. *Accident Analysis and Prevention*, **30**(4), 505-517.

Wang, X. and Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis and Prevention*, **40**, 1674-1682.

Washington, S., Karlaftis, M. and Mannering, F. (2003). *Statistical and econometric methods for transportation data analysis*.

Yamamoto, T., Hashiji, J. and Shankar, V. N. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident Analysis and Prevention*, **40**, 1320-1329.

5. DESENVOLVIMENTO DE UM MODELO MATEMÁTICO DE APOIO À DECISÃO EM FASE DE PLANEAMENTO

Como se pode depreender do Capítulo 3., em particular de 3.2, a fase de planeamento é preponderante para uma actuação proactiva na infra-estrutura, tendo em vista a diminuição da sinistralidade em meio urbano. Esta actuação deve estar de acordo com o modo de planear o meio urbano. Tendo em conta a realidade do planeamento urbanístico e viário praticado em Portugal, pretende-se neste trabalho desenvolver um modelo econométrico que incorpore variáveis que resultem de informação habitualmente tratada nesta fase.

Embora tenham sido já realizados alguns trabalhos no contexto do planeamento tal como referido em 3.2, essas ferramentas são, no entanto, difíceis de aplicar nos modelos correntes de planeamento dos transportes. Para determinar qual o melhor modelo a considerar como ferramenta de decisão para a fase de planeamento, consideraram-se várias hipóteses de modelação agrupadas em dois tipos de abordagens em função do tipo de resposta do modelo – resposta quantitativa e qualitativa.

Assim, na primeira parte deste capítulo, será feita uma breve descrição dos modelos a aplicar. Na segunda parte, descrevem-se os dados considerados na modelação. Na terceira parte, apresentam-se os resultados da aplicação e respectiva análise, divididos pelos dois tipos de abordagens acima referidas.

5.1 Descrição dos modelos

Os modelos desenvolvidos neste trabalho têm como objectivo constituir uma ferramenta de apoio à decisão em fase de planeamento. Os modelos são do tipo mesoscópico, isto é, aplicam-se à rede viária desagregada em dois tipos de entidades – arcos e nós, sendo que os nós correspondem às intersecções e os arcos são os eixos viários que ligam cada par de nós. Este nível de aplicação tem a vantagem, em relação aos modelos macro que agregam os dados ao nível da área, de se poder identificar mais concretamente os locais

problemáticos sendo, por isso, mais fácil de ponderar possíveis medidas mitigadoras ou correctivas de determinado cenário. Estes dois tipos de abordagem têm sido pouco aplicados e surgem normalmente associados a uma fase proactiva. Existem ainda os modelos micro, que se utilizam, em geral, para analisar a configuração geométrica de locais específicos e o respectivo funcionamento. Esta abordagem é frequentemente utilizada para analisar medidas de tratamento a implementar em determinado local. Esta abordagem micro, amplamente aplicada e desenvolvida nas últimas décadas, é geralmente usada numa fase reactiva para tratamento de determinados locais da rede ou, eventualmente, na fase de projecto para analisar configurações alternativas (fase proactiva).

Os modelos meso são normalmente diferenciados em função da entidade rodoviária – arcos ou nós, pois as características associadas às mesmas são distintas, homogeneizando-se assim características intergrupais, factor importante no caso dos modelos transversais (modelos aplicados a dados de vários locais). A distinção entre estas duas entidades é física e funcional e consequentemente os acidentes que ocorrem nestas entidades são também de natureza distinta. Por estes factos, generalizou-se a separação da modelação da rede nestes dois tipos de entidades (Brude, U. and Larsson, J., 1993; Poch, M. and Mannering, F., 1996; Turner, S. and Nicholson, A., 1997; Lord, D., 2000; Chatterjee, A. and Wegmann, F. J., 2002; Kumara, S. P. and Chin, H. C., 2003; Jonsson, T., 2005).

Para relacionar os acidentes com a exposição e factores de risco associados à fase de planeamento, analisou-se um conjunto de possíveis variáveis independentes representativas destes factores a considerar nos modelos. Dessa análise, e em comum às duas entidades viárias, resultou a selecção de uma variável de exposição e duas variáveis que traduzem factores de risco – volume de tráfego, ocupação do solo e classificação hierárquica das vias.

O volume de tráfego é uma variável de exposição fundamental na medida em que é a circulação dos veículos que gera os acidentes, isto é, sem tráfego não há acidentes. Esta variável tem sido referida em diversos trabalhos de modelação como a mais determinante para a ocorrência de acidentes sendo muitas vezes considerada como variável única do modelo (Fridstrom, L., et al., 1995; OCDE, 1997; Lord, D., 2000; Lord, D., 2006). Nos modelos desenvolvidos por Greibe, P. (2003) o tráfego (volume e suas interacções) foi identificado como a variável independente mais preponderante, representando cerca de

90% e 30% da componente sistemática dos modelos para arcos e nós, respectivamente. Esta variável de exposição pode ser considerada de diversas formas sendo a mais utilizada o tráfego médio diário anual.

A forma como esta variável é incluída nos modelos de sinistralidade difere em função do tipo de entidade. No caso dos nós, a variável relativa ao volume de tráfego pode assumir várias formas. A mais simples relaciona o fluxo de tráfego com os acidentes considerando o total do fluxo de entrada no nó. No entanto, e embora esta relação tenha o mérito da simplicidade, ela não traduz o conflito do tráfego da intersecção. Para avaliar qual o modelo que melhor representa a relação entre os acidentes e o fluxo de tráfego na intersecção, Turner, S. and Nicholson, A. (1997) consideraram três tipos de modelos: tipo 1 - modelos que relacionam o total de acidentes com a soma do volume de tráfego de entrada na intersecção; tipo 2 – modelos que relacionam o total de acidentes com o produto do volume de tráfego dos ramos de entrada que se intersectam (ramos principal e secundário); tipo 3 – modelos que relacionam os acidentes de determinada natureza em função das manobras de conflito e o volume de tráfego desses movimentos conflituantes. Para cada um destes três tipos de modelos foram consideradas diferentes hipóteses para as funções que relacionam os acidentes com o tráfego. Todos estes modelos foram desenvolvidos e analisados para a generalidade dos nós de uma rede urbana e não para aplicação individual em intersecções. Dos resultados obtidos, os autores concluíram que os modelos do tipo 1 são os menos adequados. Este resultado vem de encontro ao expectável já que neste tipo de modelo não são tidos em conta os mecanismos de conflito relacionados com a ocorrência dos acidentes. Os modelos do tipo 3 são mais adequados, pois são baseados em funções que descrevem melhor os conflitos do tráfego e distinguem a natureza dos acidentes. No entanto, para a aplicação dos modelos do tipo 3 é necessário um grau de informação nem sempre possível de obter, o que justifica a adopção, em geral, dos modelos do tipo 2 (Turner, S. and Nicholson, A., 1997). A separação do tráfego de entrada dos ramos principais e secundários, tal como considerada nos modelos do tipo 2, tem sido aplicada em diversos trabalhos (Mountain, L., et al., 1998; Lord, D., 2000; Greibe, P., 2003).

No caso dos arcos é consensual considerar-se o total do volume de tráfego presente numa secção média do arco, dado pela soma dos volumes de tráfego que circulam nos dois sentidos.

Para se obterem os volumes de tráfego, e em particular quando a abordagem é proactiva para toda a rede viária de um meio urbano, é de destacar a importância de dispor de modelos matemáticos de tráfego que permitam prever os volumes de tráfego associados a cada arco da rede. Actualmente existem vários programas de cálculo de modelação do tráfego, geralmente utilizados como uma ferramenta de gestão e planeamento do tráfego e que podem ser um complemento ao estudo da sinistralidade, tal como anteriormente referido anteriormente, em 2.3.

Para além do tráfego, existem factores de risco relacionados com o funcionamento e geometria dos locais que podem ser considerados em fase de planeamento através de variáveis que identifiquem a ocupação do solo e a classificação hierárquica da via. Estas são definidas durante o processo de elaboração do Plano Director Municipal (PDM).

Relativamente à ocupação do solo, vários autores têm abordado a importância deste factor na segurança rodoviária. É o caso de Berkovitz, A. (2001) que analisa a relação estreita entre a ocupação do solo e a segurança rodoviária, em particular no que se refere aos peões e ciclistas. Neste artigo, a autora salienta os benefícios de uma ocupação do solo mista que promova a utilização de meios de transporte não poluentes e a diminuição das distâncias e tempos de viagem. A mesma autora refere, no entanto, a falta de ferramentas de medição da segurança rodoviária no planeamento da ocupação do solo para apoio aos decisores. Ivan, J. N., *et al.* (2000), desenvolveram um modelo, usando a regressão de Poisson, para analisar a influência da ocupação do solo na taxa de acidentes em estradas de duas vias usando como variável de exposição o volume de tráfego correspondente a cada hora. Neste artigo, os autores concluem que os resultados obtidos indicam a necessidade de aprofundar a influência da ocupação do solo nos acidentes referindo que o estudo aponta para que esta seja uma variável explicativa da acumulação de acidentes em determinados segmentos de estrada. Também Wedagama, D., *et al.* (2006) se debruçaram sobre a relação entre a ocupação do solo e os acidentes com vítimas não motorizadas (peões e ciclistas). Estes autores lembram que as teorias de gestão da segurança rodoviária (SR) no meio urbano apontam para as políticas de ocupação do solo como uma das estratégias de prevenção e redução dos acidentes. Tendo como base a ideia de que diferentes tipos de ocupação do solo geram diferentes volumes de viagens e diferentes comportamentos do condutor que por sua vez são potenciais causas de acidentes, os mesmos autores desenvolveram um modelo macro (acidentes agregados por áreas) para analisar a influência da ocupação do

solo nos acidentes com vítimas peões e ciclistas, dividindo a análise em dois períodos de tempo – horário laboral e horário fora do período laboral. Os resultados demonstraram que, por exemplo, durante o horário laboral, as vítimas por atropelamento estão associadas a áreas com um número elevado de edifícios de ocupação colectiva, tais como, hospitais, escolas, instituições religiosas, etc. e de comércio e no horário fora do período laboral com as actividades de lazer nocturnas (bares, discotecas, clubes, etc.). Mais recentemente, Dissanayake, D., *et al.* (2009) concluíram existir uma relação entre a ocupação do solo e os acidentes com crianças vítimas por atropelamento.

Embora a maior parte dos trabalhos acima referidos se centrem em exclusivo nos acidentes com peões e ciclistas, considerou-se neste trabalho a extensão desta relação a todo tipo de acidentes pois à ocupação do solo está associada a caracterização da mobilidade (afluência de peões, transportes públicos, etc.) e das acessibilidades (número e tipo de acessos aos edifícios marginais à via) dos locais de uma forma geral.

Para relacionar os acidentes com factores de risco associados às características físicas e funcionais dos locais, considerou-se a classificação hierárquica da via. Como é sabido, esta classificação foi definida para agrupar as vias em termos funcionais às quais deverão corresponder determinadas características geométricas. À data não se conhecem trabalhos que incluam a classificação hierárquica como variável independente na modelação dos acidentes. Existem, contudo, modelos ajustados a determinado tipo de vias, considerando-se assim de uma forma indirecta a classificação hierárquica das vias.

Para os modelos aplicados aos arcos, para além das variáveis independentes referidas, também é habitual considerar como variáveis de exposição o comprimento do arco e o número de intersecções do arco com vias de acesso local. No caso do comprimento dos arcos, verifica-se que nem sempre a relação com os acidentes é linear, tal como demonstram os resultados de Mountain, L., *et al.* (1996). No mesmo estudo, os autores concluíram ainda que a presença de intersecções com vias de acesso local³⁴ é um factor influente na frequência de acidentes. Nesse sentido, os autores consideram que se deve incluir nos modelos dos arcos uma variável relativa ao número de intersecções por quilómetros de via ou modelar os acidentes ocorridos nestas intersecções separadamente. Qualquer uma das duas hipóteses é preferível a não considerar este factor de exposição. No entanto, a segunda hipótese só é aplicável se houver dados do tráfego para as vias de acesso local, o que nem sempre acontece (Mountain, L., *et al.*, 1996).

³⁴ *minor junctions*, em inglês

Nos modelos com dados temporais (dados obtidos para determinados períodos de tempo) há ainda a considerar a tendência de variação dos acidentes ao longo do tempo - tendência temporal³⁵. Este factor é de especial importância quando se pretende realizar previsões de acidentes para um período de tempo relativamente longo. A tendência temporal resulta do efeito de vários factores, tais como, a aplicação de estratégias, políticas ou programas educacionais de segurança rodoviária a nível local ou nacional, factores económicos, condições temporais, práticas no registo dos acidentes, etc. Normalmente, tem-se registado em diversos países uma tendência para a diminuição dos acidentes ao longo do tempo. É o caso verificado no estudo realizado por Fridstrom, L., *et al.* (1995) que consideraram a tendência temporal como uma variável independente e analisaram o seu efeito em dados de diversos países do Norte da Europa. Nos modelos considerados, os parâmetros estimados para esta variável apresentaram valores negativos, com significância estatística (nível de confiança superior a 95% pelo teste *t-student*), revelando assim um decréscimo nos acidentes ao longo do tempo. Resultados idênticos foram obtidos por outros autores (Mountain, L., *et al.*, 1998; Greibe, P., 2003; Kumara, S. P. and Chin, H. C., 2003). No caso dos acidentes com vítimas essa tendência nem sempre se verifica tal como demonstra o trabalho realizado por Mountain, L., *et al.* (1996).

Contudo, os dados temporais podem sofrer do fenómeno de correlação temporal. Este fenómeno pode ocorrer devido a erros de medição ou da forma de cálculo dos dados considerados para as variáveis (por exemplo: é comum no cálculo do tráfego médio diário anual (TMDA) considerar para os anos em que faltam dados uma interpolação, induzindo assim à partida uma correlação nesses dados). Para analisar este fenómeno, Maher, M. and Summersgill, I. (1995) consideraram um modelo desagregado no tempo (cada ano corresponde a uma unidade) e compararam com o modelo agregado. Os autores concluíram que não existem diferenças significativas nas estimativas dos parâmetros das variáveis, verificando-se, no entanto, que os valores do parâmetro de sobredispersão no caso da distribuição Binomial Negativa (BN) são menores nos modelos ajustados aos dados desagregados. Estas diferenças nos valores do parâmetro de sobredispersão podem modificar a significância estatística de algumas variáveis e com isso influenciar a decisão na escolha dessas variáveis. Quando os modelos são desagregados por anos, não é garantida a independência do termo de erro. Isto representa que os erros, para o mesmo

³⁵ *time-trend*, em inglês

local em diferentes anos, podem estar altamente correlacionados. A estrutura do erro passa a ser uma mistura dos erros entre locais³⁶ e do próprio local³⁷. Por estes factos, Maher, M. and Summersgill, I. (1995) concluíram que é preferível considerar modelos agregados desde que o período de tempo não seja excessivamente largo, para dessa forma a interpretação da estrutura do termo de erro ser mais plausível. No entanto, os autores reconhecem que há situações em que é conveniente considerar os dados desagregados por anos, nomeadamente quando se pretende considerar a possibilidade dos níveis de risco de sinistralidade se alterarem ao longo do tempo. Nesse sentido, assumindo a importância de considerar a tendência temporal, os autores ajustaram um modelo desagregado aos dados por ano, mas cujos parâmetros foram estimados tratando a correlação temporal através de um método iterativo que constitui uma abordagem do método marginal apresentado por McCullagh, P. and Nelder, J. A. (1989). Os modelos do método marginal decorrem directamente dos modelos lineares generalizados (MLG) em que se considera que a função de ligação não é completamente conhecida, sendo necessário estimar um ou mais parâmetros através de um método iterativo.

Também Poch, M. and Mannering, F. (1996) ajustaram um modelo com os dados desagregados para comparar com o modelo agregado. Tal como Maher, M. and Summersgill, I. (1995), também concluíram que a possível correlação existente entre os anos em estudo não afecta significativamente os parâmetros estimados.

Para analisar a tendência temporal e o fenómeno da correlação temporal, Mountain, L. *et al.* (1998) definiram um modelo que considera o efeito da tendência temporal através da introdução de uma variável independente relacionada com os acidentes através de uma função geométrica simples. Embora esta função seja uma simplificação da realidade na medida em que a variação de ano para ano pode não ser a mesma, tem a vantagem da simplicidade e de ser mais adaptável aos dados existentes. Os autores verificaram que para os dados nacionais do Reino Unido (RU) correspondentes ao período de 1975-1995, ocorreu um decréscimo anual de 2%, ou seja, no total do período o decréscimo acumulado foi de 33%, concluindo assim que não é de desprezar o efeito desta variável no número dos acidentes. Para o tratamento da correlação temporal, tendo em conta que o modelo considera os dados desagregados por ano, Mountain, L. *et al.* (1998) estimaram os

³⁶ *between-site*, em inglês

³⁷ *within-site*, em inglês

parâmetros através de um método iterativo proposto por Atkinson que, tal como o aplicado por Maher, M. and Summersgill, I. (1995), constitui uma abordagem do método marginal.

A aplicação do método marginal tradicional, tal como nos dois casos acima referidos, embora apropriada para os modelos de previsão de acidentes (MPA), tem a desvantagem de ser necessário conhecer o tipo e a extensão da correlação temporal (Lord, D. and Persaud, B., 2000). Os modelos de transição são uma alternativa ao método marginal embora pouco atractiva já que implica uma extensa manipulação matemática e nem sempre se tem verificado como apropriada para a modelação dos acidentes.

Para analisar o tratamento da correlação temporal com base no método “*Generalized Estimating Equations*” (GEE) e também a importância da tendência temporal nos modelos, Lord, D. and Persaud, B. (2000) compararam cinco modelos: modelo 1 – com tendência temporal, aplicação do GEE e dados desagregados por ano; modelo 2 – com tendência temporal, sem aplicação do GEE e dados desagregados por ano; modelo 3 – sem tendência temporal, com aplicação do GEE e dados desagregados por ano; modelo 4 – sem tendência temporal, sem aplicação do GEE e dados de acidentes desagregados por ano; modelo 5 – sem tendência temporal, sem aplicação do GEE, mas com os dados agregados num único período temporal. Os parâmetros foram estimados considerando a distribuição BN. Os resultados da aplicação dos modelos mostraram que os modelos 1 e 2, 3 e 4 têm valores dos parâmetros iguais entre si. Ou seja, o tratamento da correlação temporal não altera os valores dos parâmetros estimados. No entanto, o tratamento da correlação temporal nos modelos 1 e 3, contribuiu para aumentar quase para o dobro os valores dos desvios padrão. Ou seja, em teoria, algumas variáveis independentes podem ficar estatisticamente insignificantes (nível de confiança inferior a 95% pelo teste *t-student*) por se tratar a correlação temporal. Esta conclusão vai de encontro às conclusões dos trabalhos referidos anteriormente. Pelos resultados dos valores dos parâmetros de sobredispersão dos modelos 3 e 5, pode concluir-se ainda que, mesmo que não seja do interesse para a aplicação considerar a tendência temporal no modelo, continua a ser benéfico considerar o GEE. Relativamente à tendência temporal considerada no modelo 1, verificou-se que os valores do termo da constante do modelo que reflecte a variação anual dos acidentes não são estatisticamente diferentes nos seis anos considerados, ou seja, os dados não demonstram uma tendência clara nos níveis de risco de acidentes. Nesse caso, o modelo 3, com um único valor para a constante, seria a opção preferível, justificada ainda pelo facto dos

parâmetros de sobredispersão não diferirem no valor. Contudo, o autor refere que, por exemplo, para estudos antes e depois não é de desprezar a diferença existente entre os valores do termo da constante de ano para ano que reflectem a variação dos níveis de risco. Para este trabalho, e dado que se pretende desenvolver um modelo de previsão em que o efeito do tempo assume especial importância, optou-se por incluir a tendência temporal como uma variável do modelo. Será ainda analisada a possibilidade de existir correlação temporal nos dados considerados.

Assim, separando a modelação dos arcos e nós, as possíveis variáveis independentes consideradas nos modelos são, respectivamente:

- $TMDA_{Tot}$; $Comp$; $DenInt$; ANO ; $OcSolo_k$; $ClHier_g$
- $TMDA_{Princ}$; $TMDA_{Sec}$; ANO ; $OcSolo_k$; $Classe_h$

em que $TMDA$ são variáveis contínuas de exposição (total no arco, ou nos ramos principal e secundário de um nó), ANO é a variável da tendência temporal, $OcSolo$ é uma variável “dummy” que identifica a ocupação do solo e $ClHier$ e $Classe$ são variáveis “dummys” que identificam a classificação hierárquica das vias dos arcos e nós, respectivamente. Esta última distinção é necessária na medida em que um nó pode conter ramos de vias com classificações hierárquicas diferentes. No caso dos arcos são ainda consideradas duas variáveis contínuas de exposição – $Comp$ e $DenInt$, comprimento do arco e densidade de intersecções com vias de acesso local, respectivamente.

Para a modelação destas funções consideraram-se, tal como referido anteriormente, duas abordagens distintas em função do tipo de resposta dos modelos - quantitativa e qualitativa. Para a primeira abordagem considera-se como resposta do modelo a frequência dos acidentes. Na segunda abordagem a variável de resposta corresponde a uma categoria de entre uma classe de categorias que está relacionada com o número de acidentes observados. Para a modelação dos acidentes segundo estas duas abordagens, utilizou-se o programa de cálculo LIMDEP versão 9.0, tendo como base de dados de acidentes e da rede a do concelho do Porto e que a seguir se descreve.

5.2 Dados considerados na aplicação

Os dados considerados correspondem à informação dos acidentes, do tráfego e da rede viária do concelho do Porto no período de 2001 a 2005. Para obter estes dados foi necessário contactar diferentes instituições públicas e realizar todo o processo de organização e tratamento dos dados que a seguir se descreve.

5.2.1 Dados da infra-estrutura

A rede considerada neste trabalho está estruturada em arcos e nós correspondentes às vias distribuidoras principais e locais. Os arcos resultam da intersecção entre as vias distribuidoras principais e locais e entre estas e os ramos das vias colectoras (via de cintura interna, avenida AEP, via norte e A3). Os nós foram inicialmente agrupados em função da sua geometria e funcionamento tendo sido consideradas as seguintes classes: intersecções simples; intersecções múltiplas; intersecções singulares; rotundas (ver Figura 5.1).

As intersecções simples são aquelas cujos eixos das vias dos ramos se intersectam num único ponto a partir do qual se mede o raio de influência da intersecção nos acidentes. As intersecções múltiplas apresentam dois ou mais pontos de intersecção entre os eixos das vias dos ramos a partir dos quais se mede a distância de influência da intersecção nos acidentes. O funcionamento destes dois tipos de intersecções decorre dentro dos padrões normais. As intersecções singulares surgiram pela dificuldade em integrar algumas intersecções numa das anteriores devido ao seu grau de complexidade em termos funcionais e/ou geométricos. Estas intersecções ocupam, em geral, uma grande área e são, em alguns casos, geometricamente idênticas às rotundas mas com um funcionamento distinto destas. Normalmente, são constituídas por cinco ou mais ramos cujos ângulos entre si tornam difícil a escolha dos ramos principais e secundários. Algumas apresentam ainda um funcionamento muito difícil de caracterizar. Devido a estas especificidades e ao número reduzido destas (8 intersecções), excluíram-se estas intersecções do estudo assim como as rotundas (16 rotundas). O objectivo foi o de não introduzir uma elevada heterogeneidade entre os nós considerados para o estudo que poderá afectar os resultados do modelo. As intersecções simples e múltiplas foram agrupadas por não apresentarem

grandes diferenças em termos funcionais assegurando assim um número de dados suficiente para a aplicação. Esta diferenciação em intersecções simples e múltiplas foi essencialmente necessária para determinar os acidentes que pertencem à intersecção, já que o processo de medição do raio de influência é diferente.

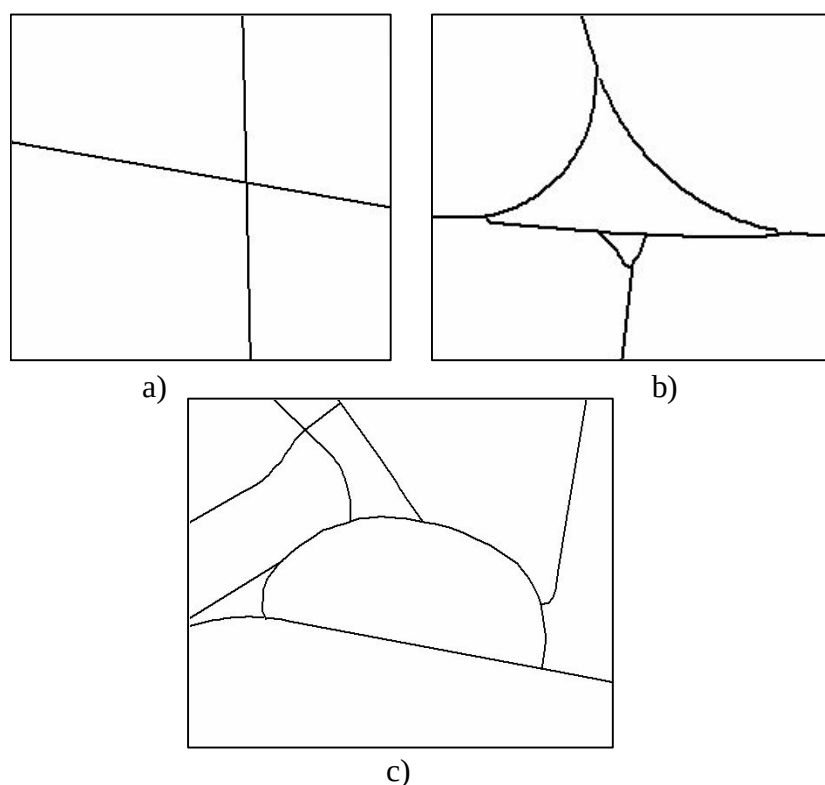


Figura 5.1. - Exemplos de uma intersecção simples a); intersecção múltipla b); intersecção singular c)

Assim, a rede resultante, após as considerações acima mencionadas, compreende 280 nós e 457 arcos. No entanto, devido à falta de dados de tráfego em algumas destas entidades, este número reduziu-se para 211 nós e 396 arcos.

Para afectar os acidentes às duas entidades rodoviárias, considerou-se como raio de influência da intersecção o valor de 20 metros. Assim, os acidentes localizados até 20 metros do centro da intersecção são quantificados na base de dados de acidentes ocorridos nos nós. Os acidentes localizados fora dos 20 metros, são quantificados na base de dados de acidentes ocorridos nos arcos.

Atendendo a que, tal como referido anteriormente, a ocupação do solo e a classificação hierárquica viária vão integrar o modelo proposto, estabeleceu-se uma classificação dos arcos e dos nós tendo em conta estas variáveis.

Para a classificação hierárquica da rede considerou-se a definida pela Câmara Municipal do Porto (CMP) e que consta no documento “Planta de Hierarquia Rodoviária” realizado no âmbito do Plano Director Municipal³⁸. As designações para as vias consideradas pela CMP foram alteradas de forma a coincidir com as habitualmente utilizadas na bibliografia nacional e internacional de referência.

Para este trabalho consideraram-se unicamente as vias distribuidoras principais e locais, excluindo-se as vias colectoras pela sua especificidade e as vias de acesso por falta de dados.

Desta forma, foram identificados 199 arcos de vias distribuidoras principais e 197 arcos de vias distribuidoras locais.

Quadro 5.1 - Classificação hierárquica dos arcos

Designação das vias considerada pela CMP	Designação das vias considerada no estudo
Rede Nacional (IP1, IC1, IC23 e IC29)	Via colectora (IC23, IC1, IP1)
Eixo urbano estruturante e de articulação intermunicipal	Distribuidora principal
Eixo urbano complementar ou estruturante local	Distribuidora local
-	Acesso local

Relativamente aos nós, a classificação não é tão evidente na medida em que o nó é constituído no mínimo por três ramos que podem corresponder a vias com classificações hierárquicas diferentes. Assim, foi criada uma nova classificação hierárquica para os nós que considera três classes em função da conjugação entre as vias classificadas segundo o Quadro 5.1. Estas três novas classes foram designadas de A, B e C, tal como indica o Quadro 5.2 onde se mostra igualmente o número de nós presentes em cada uma delas.

³⁸ Ratificado pela Resolução do Conselho de Ministros n.º 19/2006 publicada no Diário da República - I Série- B, n.º 25 de 3 de Fevereiro de 2006. Consultado em www.cm-porto.pt em 2008

Quadro 5.2 - Classificação hierárquica dos nós

Classes hierárquicas	Classificação dos ramos			Nº Nós
A	Via colectora $\cap$ Dist. Principal		Via colectora $\cap$ Dist. Local	12
B	Dist. Principal $\cap$ Dist. Principal	Dist. Principal $\cap$ Dist. Local	Dist. Principal $\cap$ Acesso Local	128
C	Dist. Local $\cap$ Dist. Local		Dist. Local $\cap$ Acesso. Local	71

Para a identificação do tipo de ocupação do solo considerou-se a classificação definida pela CMP e que consta no documento intitulado “Planta de qualificação do solo” do mesmo PDM. As várias formas de ocupação territorial definidas nesta planta foram agrupadas em cinco categorias, tal como apresentado no Quadro 5.3, de forma a serem suficientemente representativas através de um conjunto de características específicas comuns e que estas se possam reflectir na sinistralidade. Estas cinco categorias foram diferenciadas após testar outras hipóteses nomeadamente a junção das categorias 3 e 4 e das categorias 1 e 5.

Quadro 5.3 - Classificação da ocupação do solo

Categorias	Classificação da CMP		Nº Nós	Nº Arcos
Ocupação Solo1	Área de frente urbana consolidada	Área de frente urbana contínua em consolidação	118	217
Ocupação Solo2	Área de habitação do tipo unifamiliar	Área de edificação isolada com prevalência da habitação colectiva	46	88
Ocupação Solo3	Área empresarial		6	12
Ocupação Solo4	Área de equipamento existente		11	23
Ocupação Solo5	Área histórica		30	56

Os arcos e nós foram classificados nestas cinco categorias com a distribuição indicada no lado direito do Quadro 5.3.

Apresenta-se também nas Figuras 5.2 e 5.3 a distribuição das vias classificadas e das cinco classes de ocupação do solo no concelho do Porto, respectivamente.

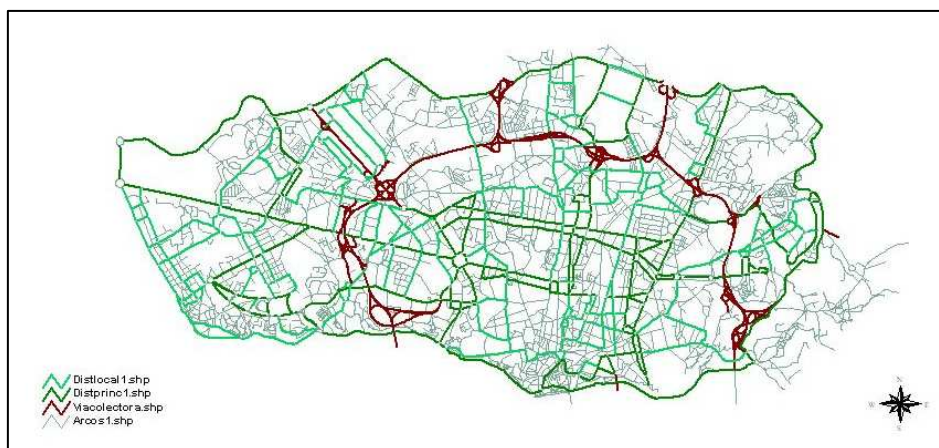


Figura 5.2. - Classificação hierárquica da rede do Porto

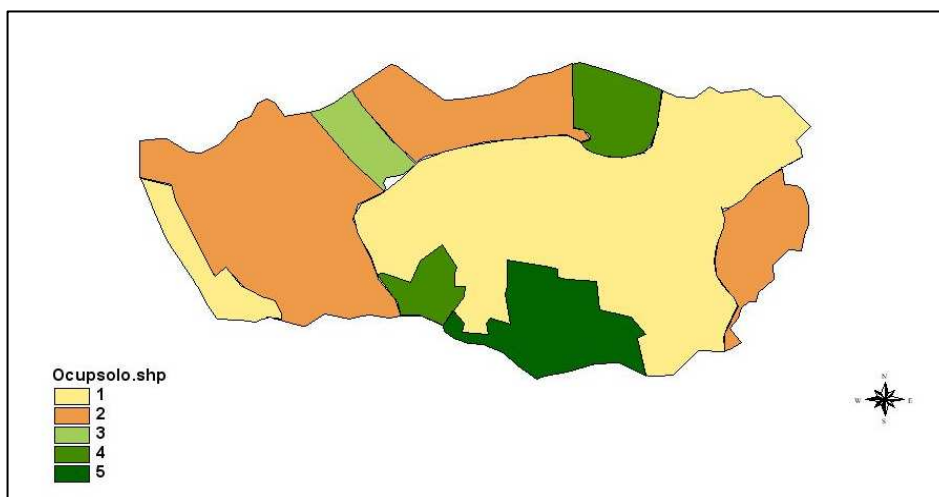


Figura 5.3. - Classificação da ocupação do solo do concelho do Porto

Para além desta informação qualitativa dos nós e dos arcos foram ainda obtidos dados para os nós sobre o tipo de funcionamento (com/sem sinalização luminosa), o número de ramos e o número de movimentos permitidos. Para os arcos considerou-se ainda o seu comprimento e o número de intersecções com as vias de acesso local. Esta informação foi obtida quer pela análise da rede digitalizada quer por observação *in situ* dos locais.

5.2.2 Dados de acidentes

Os dados dos acidentes rodoviários utilizados neste trabalho resultam da informação dos Boletins Estatísticos de Acidentes de Viação (BEAV). Os BEAV foram facultados pelo Governo Civil do Porto que até ao ano 2005 procedia ao seu arquivo. A informação considerada para o trabalho corresponde aos anos de 2001, 2002, 2003, 2004 e 2005. Os BEAV foram fornecidos em formato papel tendo sido necessário elaborar uma base de dados usando para tal um programa informático que organiza os dados, neste caso o Microsoft® Access.

A informação inserida para cada acidente corresponde à data e hora, localização (nome da rua e número de polícia ou ponto de referência), natureza do acidente, número e tipo de veículos intervenientes e, no caso de acidentes com vítimas, o número de vítimas e a sua gravidade (mortos, feridos graves e feridos leves).

Em Janeiro de 2004 o BEAV sofreu algumas alterações, destacando-se o número de campos que identificam a natureza do acidente. Até 2004, a natureza do acidente era identificada por um de doze campos tendo, a partir desta data, passado a existir dezassete campos. Relativamente ao processo de registo da ocorrência, o preenchimento do BEAV só é completo no caso dos acidentes com vítimas, registando-se apenas, no caso de acidente só com danos materiais, os campos de “identificação do acidente”. Estes campos são os equivalentes aos seleccionados para o trabalho perdendo-se, no entanto, informação precisa do tipo específico da natureza do acidente, pois passa a ser identificado apenas em três categorias (despiste, colisão e atropelamento). Por estes factos, foi necessário compatibilizar a informação dos diferentes BEAVs.

Em todo o processo que vai desde o registo dos acidentes à elaboração da base de dados podem surgir uma série de erros que, para o trabalho em questão, são potencialmente prejudiciais para os resultados da modelação dos acidentes. No registo da ocorrência, os erros podem ser resultantes de um incorrecto preenchimento ou da omissão de dados. Estes são de maior gravidade para o trabalho quando se trata da localização do acidente, já que sem esta informação não é possível relacionar os acidentes com as características do local da ocorrência. Como se poderá depreender pelos organigramas apresentados nas Figuras 5.4 e 5.5, este tipo de erro é muito frequente (52% dos acidentes não têm registo do número de polícia). Também o facto de ser um grande volume de dados (28 559 acidentes) a inserir para formato digital favorece a ocorrência de enganos.

O número de acidentes a analisar ficou ainda reduzido aos acidentes que se localizam nas vias da rede em estudo (vias distribuidoras principais e distribuidoras locais) como referido pelos organigramas das Figuras 5.4 e 5.5.

Dos 28 559 acidentes inseridos, 23 662 são acidentes só com danos materiais e 4 897 são acidentes com vítimas. Nos organigramas das Figuras 5.4 e 5.5, são apresentados os vários níveis de perdas de informação para estes dois tipos de acidentes. Do número inicial de acidentes inseridos ao número de acidentes considerados no trabalho houve uma perda de cerca de 66% e de 63% nos acidentes com danos materiais e nos acidentes com vítimas, respectivamente.

De salientar ainda que este resultado final beneficiou de um trabalho individualizado para alguns acidentes que apesar de terem a informação necessária não foram localizados automaticamente pelo Sistema de Informação Geográfica (SIG). Nestes casos, foi necessário estudar os dados quer dos acidentes quer do local para conseguir atribuir uma possível localização do acidente.

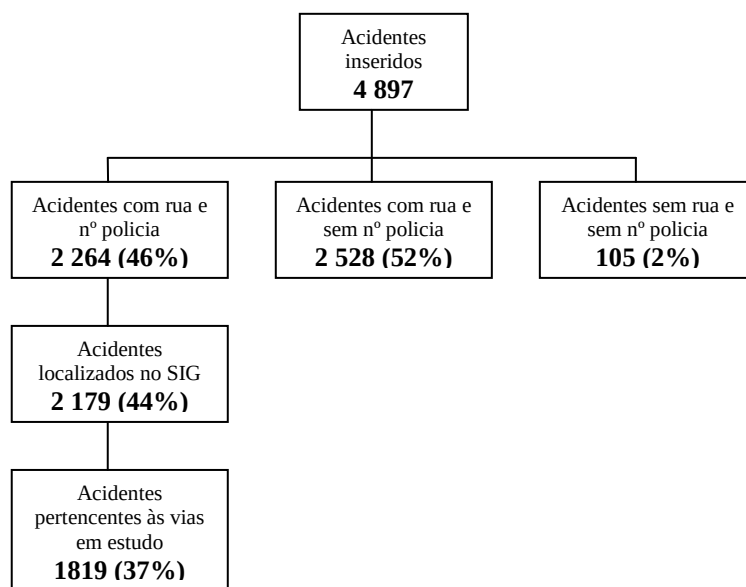


Figura 5.4. - Organograma da base de dados dos acidentes com vítimas

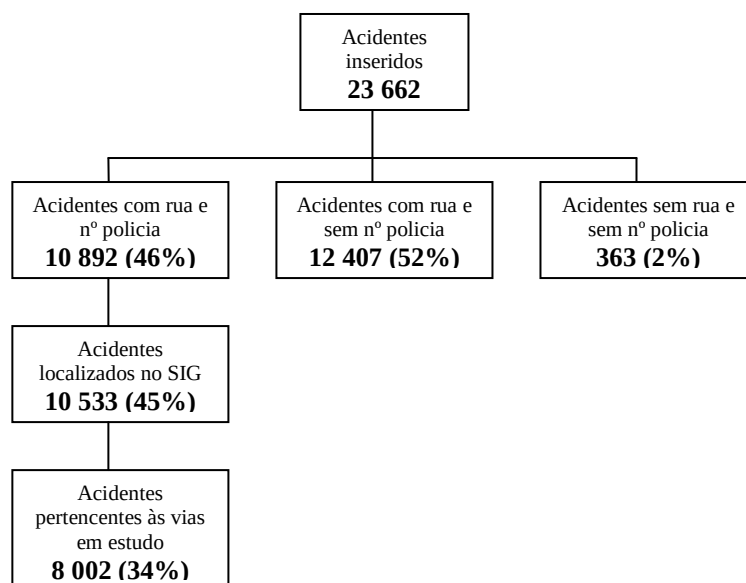


Figura 5.5. - Organograma da base de dados dos acidentes só com danos materiais

Ao longo dos cinco anos em estudo, os acidentes sofreram uma variação da frequência que aponta para uma tendência para a sua diminuição, tal como se pode observar pelo Quadro 5.4. A exceção é o ano de 2004, onde se verificou um aumento em relação ao ano anterior.

Quadro 5.4 - Distribuição dos acidentes ao longo dos 5 anos

Nº Acidentes	2001	2002	2003	2004	2005
Com Vítimas	445	327	343	374	329
Danos Materiais	1700	1706	1447	1731	1418

Estes acidentes estão distribuídos na rede pelos arcos e nós considerados para o estudo tal como indica o Quadro 5.5.

Quadro 5.5 - Distribuição dos acidentes nos arcos e nos nós (5 anos)

Acidentes	Arcos	Nós
c/ Vítimas	1183	447
Só c/ Danos Materiais	4467	1582
Total	5650	2029

Como se pode observar pelo Quadro 5.5 os acidentes concentram-se maioritariamente nos arcos (74%), onde 21% são acidentes com vítimas.

5.2.3 Dados de tráfego

Para o meio urbano não existe uma entidade responsável pela recolha e arquivo da informação relativa aos tráfegos rodoviário e pedonal. Não sendo possível recorrer a uma única fonte de informação completa, foi necessário estimar o TMDA a partir de duas fontes. A principal fonte de informação resultou da modelação de tráfego a partir do programa de cálculo SATURN – “*Simulation and Assignment of Traffic to Urban Roads Networks*”. A segunda fonte de informação corresponde a valores de tráfego medidos por contadoras existentes em algumas zonas da rede viária.

O SATURN é um modelo de análise de redes desenvolvido pelo *Institute for Transport Studies*, da Universidade de Leeds. Este modelo, através da definição da rede e da matriz Origem-Destino, afecta as viagens aos percursos e simula macroscopicamente o comportamento do tráfego na rede calculando os volumes totais nos arcos e os indicadores de funcionamento da rede (tempos, atrasos, distâncias percorridas, etc) (Tavares, J. P., 2003).

Os volumes de tráfego considerados neste trabalho resultaram da modelação do tráfego da rede do Porto com o SATURN realizada no âmbito de um protocolo entre a Faculdade de Engenharia da Universidade do Porto e a CMP. Os resultados obtidos na aplicação do SATURN no âmbito deste projecto correspondem aos volumes da hora de ponta (VHP)

para o ano de 2005. Estes resultados foram afectos à rede considerada para este estudo. Devido a algumas incompatibilidades nas classificações hierárquicas da rede utilizada neste trabalho e a do SATURN, resultantes de não se ter considerado a mesma fonte de classificação, verificou-se que, por um lado, grande parte dos arcos abrangiam dois ou mais arcos da rede do SATURN e, por outro lado, alguns arcos não estavam incluídos na rede do SATURN não sendo possível, por isso, obter dados. No primeiro caso, considerou-se a média dos VHP dos respectivos arcos do SATURN.

Os valores dos VHP foram afectos de forma distinta aos arcos e nós. Nos arcos adicionou-se os VHP dos dois sentidos sempre que era o caso. Nos nós considerou-se a soma dos volumes de entrada dos ramos opostos, tendo-se assim obtido dois valores distintos que correspondem aos ramos principal e secundário.

Para complementar a informação do SATURN recorreu-se ainda aos valores do tráfego medidos por equipamento de recolha automática de dados de tráfego (espiras electromagnéticas vulgarmente conhecidas por contadoras) instaladas em algumas zonas do Porto e cujos dados são arquivados na CMP. As contadoras registam permanentemente o número de veículos que passam. Existem 71 contadoras instaladas em 7 zonas que correspondem às zonas mais centrais e movimentadas da rede viária. Em 2004 e em 2005 foram instaladas mais 4 e 2 contadoras respectivamente, mas cujos dados não foram considerados no trabalho por estarem bastante incompletos.

Os valores do tráfego medidos pelas contadoras foram fornecidos em dois tipos de formato – digital e papel. Os dois primeiros anos (2001 e 2002) foram cedidos em formato digital com os valores de hora a hora. Destes valores calculou-se o volume total diário (VTD). Para os restantes anos (2003, 2004 e 2005) apenas estavam disponíveis os volumes totais diários arquivados em papel tendo-se, por isso, inseridos estes dados num ficheiro Microsoft® Excel.

Devido a eventuais falhas ou avarias das contadoras não foi possível, em alguns anos, obter os dados de todos os meses.

Com os dados das contadoras de 2001 e 2002 (os únicos para os quais foi possível obter valores de hora a hora) determinou-se a proporção do volume da hora de ponta relativamente ao volume total diário. Como se pode ver no Quadro 5.6 as médias aritméticas das percentagens para as 7 zonas são iguais nos dois anos – 5,9%.

Quadro 5.6 - Proporções dos volumes de hora de ponta por zona

Zona	VHP/VTD %	
	2001	2002
1	5,3	5,3
2	6,0	5,5
3	6,0	5,8
4	5,9	6,7
5	6,2	6,3
6	6,2	6,0
7	5,4	5,6
Média	5,9	5,9

Identificou-se para cada entidade (arcos e nós) a zona correspondente e, ao valor do VHP obtido no SATURN, atribui-se a referida proporção. Nas entidades que não se inserem numa das zonas (a maior parte) aplicou-se o valor médio, ou seja 5,9%, e, no caso das entidades abrangerem mais do que uma zona, considerou-se a média das respectivas zonas. Desta forma obteve-se para cada entidade o TMDA para o ano de 2005.

O TMDA para os cinco anos em estudo foi também calculado com base nos dados das contadoras. A partir do TMDA de cada contadora procedeu-se a uma normalização dos valores tendo como referência o ano de 2005. Para tal, calculou-se para cada contadora a relação entre o valor anual e a média dos cinco anos (por exemplo: $TMDA_{2001}/TMDA_{média5}$). Em seguida, e considerando o valor médio das contadoras para cada zona, dividiu-se os valores obtidos para cada ano pelo valor do ano de 2005. Assim, obteve-se a variação do tráfego relativamente ao ano de 2005 em cada zona.

O valor do TMDA de 2005 de cada entidade, determinado a partir do SATURN, multiplicado pelos valores de variação do tráfego de cada ano, originou o TMDA dos anos de 2001, 2002, 2003 e 2004.

5.3 Modelos de resposta quantitativa

Nos modelos de resposta quantitativa a variável de resposta corresponde à frequência dos acidentes. Esta frequência de acidentes pode ser desagregada, por exemplo, em função das consequências do acidente ou da natureza do acidente. No primeiro caso, geralmente consideram-se os acidentes com vítimas (mortos, feridos graves e feridos leves) e o total dos acidentes (acidentes com vítimas e acidentes só com danos materiais). A utilização dos dados dos acidentes com vítimas tem a vantagem de estes serem à partida mais próximos dos valores reais na medida em que o registo destes é usual havendo, em geral, um maior cuidado no preenchimento da informação acerca da ocorrência. No entanto, a utilização de dados referentes apenas aos acidentes com vítimas tem a desvantagem de corresponder a um menor número, relativamente à frequência do total de acidentes. Acresce ainda o facto de se associar à ocorrência de vítimas uma sobredispersão maior resultante de diversos factores que condicionam a existência, o número e a gravidade das vítimas tais como as características do veículo e as condições físicas das vítimas. Nesse sentido, dificilmente se poderá considerar estas ocorrências como independentes (Cardoso, J., 2007).

A desagregação dos acidentes pela sua natureza permite analisar de uma forma mais clara as relações entre a variável de resposta e as variáveis independentes porém, para ser aplicável, é necessária informação desagregada (por exemplo, movimentos do tráfego). Acresce o facto de esta desagregação reduzir o número de observações por modelo. Este tipo de desagregação é mais utilizado em modelos cujo objectivo é determinar as causas dos acidentes ou destinados a utilização na fase de projecto.

A agregação dos acidentes permite obter resultados estatísticos satisfatórios pois o número de observações é maior, com a vantagem de eliminar os erros resultantes da diferenciação dos acidentes pelas suas consequências ou pela sua natureza.

Assim, ponderando as vantagens e desvantagens e tendo em conta os objectivos do trabalho, optou-se por considerar duas variáveis de resposta – o total dos acidentes e os acidentes com vítimas (mortos, feridos graves e feridos leves).

Normalmente no caso de estudos da sinistralidade, o processo de modelação inicia-se considerando para a variável dependente as distribuições de Poisson e BN. Aliás, quando se considera a distribuição BN, e tendo em conta que o cálculo do parâmetro de dispersão é um processo iterativo que se inicia com os resíduos obtidos pelo modelo com distribuição

de Poisson, o programa de cálculo utilizado também apresenta os resultados da estimação deste último.

A selecção das variáveis independentes a manter na estrutura do modelo é realizada com base nos níveis de confiança dos parâmetros estimados pelo teste *t-student*. As variáveis cujos parâmetros são estimados com um nível de confiança igual ou superior a 95% não são rejeitadas. Para valores inferiores a 95% é feita uma análise para ponderar se a variável deverá ser ou não incluída. Se a presença da variável sem significância estatística (nível de confiança inferior a 95%) não afectar os níveis de confiança das restantes variáveis, optou-se por manter a variável, já que o facto de não ter significância estatística nestes modelos não lhe retira a possibilidade de ser uma variável influente noutra base de dados.

Para avaliar os resultados de diferentes modelos aplicados e assim seleccionar o melhor, é necessário dispor de elementos que avaliem a qualidade de ajuste desses modelos. De notar a dificuldade em definir o significado de melhor modelo já que o modelo mais ajustado à base de dados pode não ser o modelo com melhor desempenho em termos de previsão. Tal como referido por McCullagh, P. and Nelder, J. A. (1989), é possível conseguir um modelo com um ajuste perfeito à base de dados se considerarmos um número de variáveis igual ao número de observações. Contudo, não é isso que se pretende num modelo, pelo contrário o objectivo é desenvolver modelos simples através de um conjunto parcimonioso de variáveis. Os mesmos autores referem ainda que uma importante propriedade do modelo é o âmbito em que se insere, isto é, o conjunto de condições existentes a partir das quais se atinge bons resultados de previsões da variável de resposta.

Para este trabalho seleccionou-se um conjunto de medidas de avaliação da qualidade do ajuste tendo em conta que os modelos desenvolvidos são MLG, cuja estimação dos parâmetros é realizada pelo método da máxima verosimilhança. Essas medidas estão apresentadas no Quadro 5.7 e estão divididas por duas colunas que diferenciam as medidas determinadas e apresentadas pelo programa de cálculo e as que foram calculadas por aplicação da fórmula de cálculo.

Quadro 5.7 - Medidas de avaliação da qualidade de ajuste dos modelos

LIMDEP 9.0	Aplicação da fórmula de cálculo
AIC = - 2ML + 2K ML é o logaritmo do valor máximo da função de verosimilhança; K – nº de parâmetros do modelo a analisar (Ivan, J. N., et al., 1999)	Freeman-Tuckey: $R^2_{FT} = \frac{\sum (f_i - \bar{f})^2 - \sum \hat{e}_i^2}{\sum (f_i - \bar{f})^2 - n}$ f – variável transformada; $\hat{e}_i$ – desvios da variável transformada; n – nº de observações (Cardoso, J., 2007)
	Índice Elvik: $R^2_{PE} = \frac{(\hat{\alpha}^0 - \hat{\alpha}^m) \sum_i (\hat{y}_i^*)^2}{\hat{\alpha}^0 \sum_i (\hat{y}_i^*)^2}$ $\hat{\alpha}^0$ - parâmetro dispersão para o modelo sem variáveis independentes (só com constante); $\hat{\alpha}^m$ - parâmetro dispersão para o modelo a analisar; $\hat{y}_i^*$ - valores ajustados pelo modelo de referência (Fridstrom, L., et al., 1995)

As medidas de avaliação do ajuste tradicionais, tais como o R^2 ou o $R^2_{ponderado}$, desenvolvidas para modelos com uma distribuição de Poisson da variável de resposta, revelaram-se desajustadas e inadequadas. Inadequadas na medida em que para baixos valores esperados da variável de resposta a distribuição de Poisson é assimétrica. Desajustadas porque engloba na variável não explicada pelo modelo quer a componente sistemática resultante das causas não directamente consideradas quer a componente aleatória (Cardoso, J., 2007).

A forma mais intuitiva de avaliar a qualidade do ajuste é considerar a função logaritmica de verosimilhança maximizada. Contudo, esta medida favorece os modelos com um maior número de variáveis. Assim, com o objectivo de seleccionar um modelo com um bom ajuste mas com um número reduzido de parâmetros, Akaike em 1974 desenvolveu a medida AIC – “Akaike Information Criterion”, que se baseia num processo de

minimização que não envolve testes estatísticos (ver Quadro 5.7). Nesse sentido, e tendo em conta que a função logaritmica de verosimilhança cresce com o aumento do número de parâmetros do modelo, deve-se seleccionar o modelo com o menor valor de AIC (Ivan, J. N., et al., 1999; Lopes, S. M. d. A. and Cardoso, J. L., 2007).

O parâmetro de dispersão das regressões BN obtido pelo método iterativo de estimação por máxima verosimilhança pode ser utilizado como medida da quantidade de variação não explicada pelo modelo. Quando num modelo o parâmetro de dispersão apresenta um valor estatisticamente significativo (nível de confiança superior a 95% pelo teste *t-student*) maior que zero (sobredispersão), tal significa que a quantidade de variação não explicada é maior do que a quantidade correspondente à variação aleatória de um modelo de Poisson perfeitamente especificado, indicando que nem todo o erro é puramente aleatório. O modelo não explica toda a variação sistemática mas representa essa parte não explicada como variação aleatória. Assim, o parâmetro de dispersão pode ser utilizado para avaliar se as variáveis independentes consideradas no modelo explicam toda a variação explicável (componente sistemática) (Fridstrom, L., et al., 1995). Nesse sentido, quanto maior o valor do parâmetro de dispersão menor é a componente sistemática explicada pelo modelo.

Com base no mesmo princípio, ou seja, determinar a quantidade de variação sistemática explicada pelo modelo, foram ainda desenvolvidas várias medidas de avaliação do ajuste. Fridstrom, L., *et al.* (1995), após aplicação e análise de várias medidas, recomenda o uso do índice de Elvik R^2_{PE} e da medida designada de Freeman-Tukey R^2_{FT} , cujas fórmulas se apresentam no Quadro 5.7.

Tal como sugerido por Hauer, E. and Bamfo, J. (1997) e já anteriormente referido em 4.1, aplicou-se ainda neste trabalho o método CURE. O método consiste em elaborar um gráfico com o acumulado dos resíduos (diferença entre o valor observado e estimado) em ordem crescente de cada uma das variáveis independentes separadamente. Assim, a curva dos resíduos acumulados do gráfico indicará um bom ajuste do modelo se esta oscilar e terminar no valor zero e não ultrapassar as linhas fronteiras correspondentes a duas vezes o desvio padrão ($\pm 2 \cdot \sigma$).

Da análise de um vasto conjunto de trabalhos realizados no âmbito da modelação dos acidentes, verificou-se que as medidas utilizadas na avaliação da qualidade do ajuste do modelo não são comuns aos vários trabalhos. Segundo Lord, D. (2000), não existe um critério manifestamente superior a outro e, como tal, a escolha dos critérios a aplicar é uma

decisão individual. Contudo, ao considerar mais do que uma medida de avaliação minimiza-se o risco de se estabelecer conclusões com base em relações subjectivas (Fridstrom, L., et al., 1995).

Os modelos de resposta quantitativa desenvolvidos e aplicados neste trabalho foram divididos em dois tipos – modelos para arcos e modelos para nós. Descrevem-se a seguir, separadamente, os modelos aplicados e os respectivos resultados.

5.3.1 Modelos para arcos

Os modelos a aplicar aos arcos baseiam-se na relação entre a variável de resposta e as variáveis independentes:

$$E[Y] = f(TMDA_{Tot}; Comp; DenInt; ANO; OcSolo_k; ClHier_g) \quad (5.1)$$

em que $E[Y]$ é o valor esperado da frequência de acidentes. As variáveis independentes são de dois tipos: variáveis contínuas e “*dummys*”. Estas últimas são a $OcSolo_k$ e $ClHier_g$. A variável $OcSolo_k$ é composta por cinco “*dummys*”: $OcSolo_k = 1$ para ocupação do solo do tipo k ; 0 para ocupação do solo diferente de k , com $k = 1$ a 5. A variável $ClHier_g$ é composta por duas “*dummys*”: $ClHier_g = 1$ para a classificação viária do tipo g ; 0 para classificação hierárquica viária diferente de g , com $g = 1$ a 2. As variáveis contínuas são o tráfego médio diário anual ($TMDA$), o comprimento do arco ($Comp$), a densidade de intersecções com vias de acesso local ($DenInt$) e a variável da tendência temporal (ANO).

Considerou-se para aplicação do modelo dois tipos de ocorrências – acidentes com vítimas e total dos acidentes. Estes acidentes estão distribuídos por 396 arcos em que a cada arco correspondem cinco observações referentes aos cinco anos em estudo o que equivale no total a 1980 observações. No Quadro 5.8 são apresentados dados estatísticos das variáveis $Comp$, $DenInt$ e $TMDA_{Tot}$. Neste quadro, é possível constatar o intervalo alargado entre os valores mínimo e máximo do $TMDA_{Tot}$ o que corresponde à elevada heterogeneidade funcional da rede. No Quadro 5.9 são apresentados os número de acidentes discriminados em acidentes com vítimas e total dos acidentes ao longo dos cinco anos, sendo ainda apresentada a distribuição percentual dos acidentes segundo o tipo de ocupação do solo e a classificação hierárquica dos arcos.

Quadro 5.8 - Dados estatísticos das variáveis TMDA, comprimento e densidade de intersecções dos arcos

Ano	TMDA (min-máx)	Comp* (min-máx)	Dens Inter* (min-máx)
2001	149-56 380	3,7-3 342,8	0,000-0,032
2002	142-53 817		
2003	147-55 739		
2004	144-54 458		
2005	169-64 068		

*comprimento dos arcos em metros; densidade de intersecções corresponde ao nº de intersecções a dividir pelo comprimento em metros

Da distribuição dos acidentes pelo tipo de ocupação do solo apresentada no Quadro 5.9, verifica-se que a classe de ocupação do solo 1 é a que se destaca com o maior número total de acidentes (51%) e a classe 3 com o menor número total dos acidentes (4%).

Quadro 5.9 - Dados estatísticos dos acidentes ocorridos nos arcos distribuídos por ano e gravidade e sua distribuição percentual pelas variáveis ocupação do solo e classificação hierárquica

Ano	Gravidade do Acidente	Nº Acidentes (min-máx-total)	OcSolo % acidentes					Classif. Hier. % acidentes	
			1	2	3	4	5	Dist Princ	Dist Loc
2001	C/ Vítimas	0-10-301	48	24	2	4	21	57	43
	Total	0-27-1277	52	19	4	6	19	56	44
2002	C/ Vítimas	0-10-220	48	18	2	8	24	61	39
	Total	0-23-1194	54	17	4	7	18	59	41
2003	C/ Vítimas	0-7-218	45	31	1	7	16	56	44
	Total	0-19-966	47	21	5	7	20	56	44
2004	C/ Vítimas	0-10-237	48	29	1	7	15	56	44
	Total	0-27-1193	50	22	4	5	18	56	44
2005	C/ Vítimas	0-9-207	44	17	4	8	27	53	47
	Total	0-18-1020	51	19	4	6	20	55	45

Nos acidentes com vítimas verifica-se a mesma proporção de acidentes segundo a ocupação do solo – 47% de acidentes no tipo 1 e 2% no tipo 3. Para a classificação hierárquica considerada, a distribuição dos acidentes é mais equitativa: 56% e 44% para o total de acidentes, 57% e 43% para os acidentes com vítimas, nas vias distribuidoras principais e locais, respectivamente.

O histograma dos acidentes (total e com vítimas) apresentado na Figura 5.6 demonstra que o maior número de observações concentra-se no número de acidentes nulo, sendo esta concentração mais preponderante nos acidentes com vítimas. A maior predominância de valores zero acidentes nas observações é um factor, em geral, comum às diversas bases de dados utilizadas em trabalhos internacionais e que motivou alguns estudos, nomeadamente a análise da aplicação de modelos zero inflacionado (ZI) cuja distribuição probabilística dos acidentes é baseada na existência de dois estados diferentes no processo de geração de acidentes (ver 4.1).

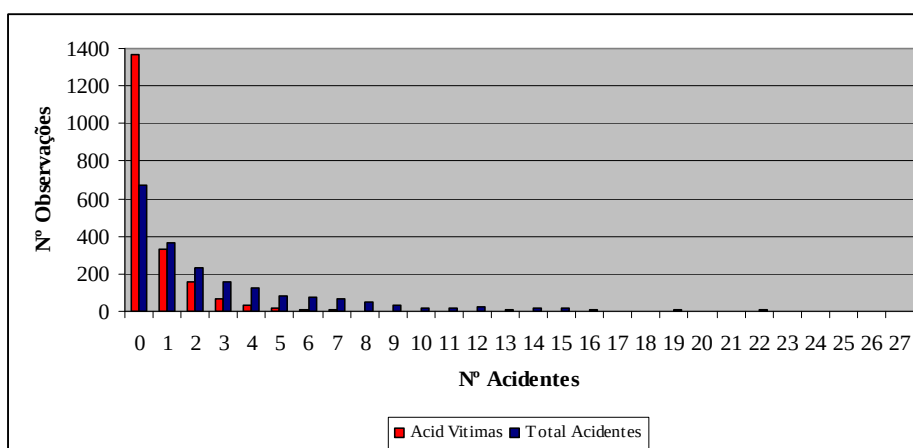


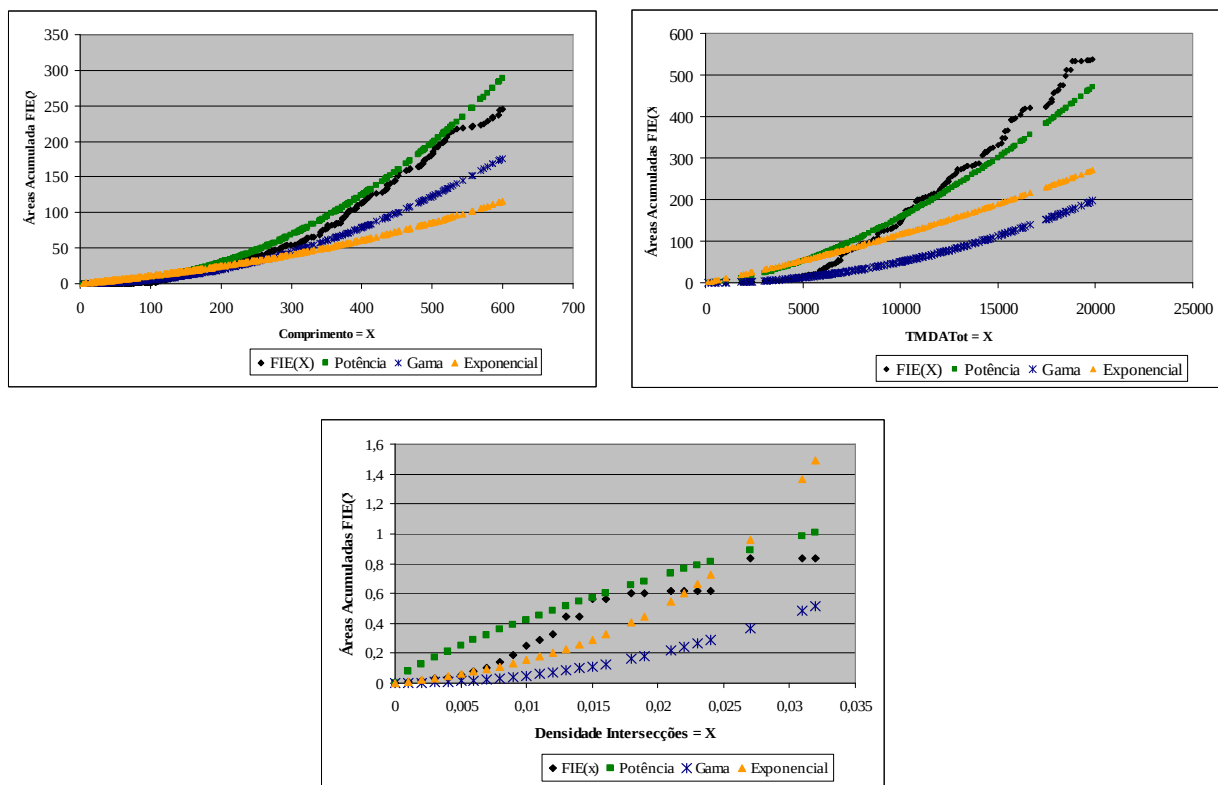
Figura 5.6. - Histograma do número de acidentes nos arcos – total e com vítimas

Total dos Acidentes

Para estabelecer a estrutura do modelo a aplicar ao total dos acidentes é necessário seleccionar a forma funcional que relaciona as variáveis independentes com a variável de resposta. Para orientar essa selecção utilizou-se o método “*Integrate-Differentiate*” (ID). Como já anteriormente referido em 4.1, este método aplica-se às diferentes variáveis separadamente. Para tal, determinou-se a “função integral empírica” (FIE) para as variáveis contínuas $TMDA_{Tot}$, $Comp$ e $DenInt$, resultando nos gráficos da Figuras 5.7.

Nesses gráficos são também apresentadas as curvas das funções a comparar - Potência, Gama e Exponencial.

Comparando a $FIE(X)$ com as curvas das funções Potência, Gama e Exponencial pode concluir-se que no caso do $TMDA_{Tot}$ e do $Comp$ a função que mais se aproxima é a função Potência. No entanto, como o método serve apenas de orientação analisaram-se os resultados para as três hipóteses (ver Anexo 2.I). Contudo, esses resultados não evidenciam qual a melhor opção, pois todos os parâmetros estimados para o tráfego e para o comprimento dos arcos em qualquer umas das três funções têm um nível de confiança superior a 95%. Assim, optou-se por considerar a função Potência por ser a função habitualmente utilizada para as variáveis tráfego e comprimento na maior parte dos modelos referidos na bibliografia (Mountain, L., et al., 1996; Lord, D., 2000; Lopes, S. M. d. A. and Cardoso, J. L., 2007; Lord, D. and Park, P. Y. J., 2008).



$$\text{Potência } f_1(x) = \alpha x^\beta \quad \text{Gama } f_2(x) = \alpha x^{\beta-1} e^{-\beta x} \quad \text{Exponencial } f_3(x) = e^{-\beta x}$$

Figura 5.7. - Gráfico comparativo da $FIE(X)$ das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ com as funções Potência, Gama e Exponencial

Pelo gráfico da Figura 5.7, pode concluir-se que a curva da função Gama é a que mais se afasta da FIE(X) da variável *DenInt*, não sendo evidente qual das funções Potência ou Exponencial se aproxima mais. De facto, aplicando o modelo para as três hipóteses, os resultados excluem a hipótese da função Gama dado que o valor estimado do primeiro parâmetro corresponde a um nível de confiança claramente inferior a 95% (ver Anexo 2.I). Pelo contrário, os valores estimados para os parâmetros das funções Potência e Exponencial correspondem a um nível de confiança superior a 95%. Assim, baseado no trabalho de Mountain, L., *et al.* (1996) optou-se pela função Exponencial.

A variável ANO é incluída no modelo para captar o efeito da tendência temporal, tendo-se considerado a forma funcional geométrica simples utilizada no trabalho de Mountain, L., *et al.* (1998).

Desta primeira análise concluiu-se que a estrutura do modelo a aplicar ao total dos acidentes é:

$$E[Y] = \beta_0 TMDA_{Tot}^{\beta_1} Comp^{\beta_2} \gamma^{ANO} e^{\beta_3 DenInt} e^{\beta_{4k} OcSolo_k + \beta_{5g} ClHier_g} \quad (5.2)$$

em que as variáveis são as mesmas anteriormente apresentadas e descritas na equação (5.1).

Realizou-se ainda uma análise das correlações entre as variáveis independentes através da matriz de correlações cujos valores são apresentados no Quadro 5.10.

Quadro 5.10 - Matriz de correlações entre as variáveis independentes da equação (5.2)

	<i>TMDA_{Tot}</i>	<i>Comp</i>	<i>DenInt</i>	<i>ANO</i>	<i>OcSolo2</i>	<i>OcSolo3</i>	<i>OcSolo4</i>	<i>OcSolo5</i>	<i>DistLoc</i>
<i>TMDA_{Tot}</i>	1,00	-	-	-	-	-	-	-	-
<i>Comp</i>	0,11	1,00	-	-	-	-	-	-	-
<i>DenInt</i>	0,05	0,23	1,00	-	-	-	-	-	-
<i>ANO</i>	0,07	0,00	0,00	1,00	-	-	-	-	-
<i>OcSolo2</i>	-0,02	0,05	0,04	0,00	1,00	-	-	-	-
<i>OcSolo3</i>	-0,04	0,03	-0,02	0,00	-0,09	1,00	-	-	-
<i>OcSolo4</i>	0,07	0,11	0,01	0,00	-0,13	-0,04	1,00	-	-
<i>OcSolo5</i>	0,10	-0,03	0,09	0,00	-0,22	-0,07	-0,10	1,00	-
<i>DistLoc</i>	-0,32	0,04	0,10	0,00	-0,06	0,15	-0,10	0,06	1,00

Dos valores apresentados no Quadro 5.10 constata-se que os valores da matriz são bastantes inferiores a 1,00 demonstrando assim não existir correlações significativas entre variáveis.

Modelo Binomial Negativo – modelo base

A primeira modelação do total de acidentes corresponde a um MLG aplicado à estrutura representada pela equação (5.2) com distribuição BN para a variável de resposta (o programa de cálculo apresenta também os resultados para a distribuição de Poisson). Os resultados desta primeira aplicação são apresentados no Quadro 5.11 (no Anexo 2.II apresentam-se todos os resultados da aplicação).

Quadro 5.11 - Resultados da estimação do modelo BN da equação (5.2)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-7,318	0,327	0,0000
β_1	0,344	0,030	0,0000
β_2	0,910	0,029	0,0000
$\ln(\gamma)$	-0,056	0,016	0,0003
β_3	37,635	4,515	0,0000
$\beta_{4,2}$	-0,336	0,055	0,0000
$\beta_{4,3}$	0,403	0,142	0,0044
$\beta_{4,4}$	-0,315	0,096	0,0011
$\beta_{4,5}$	0,207	0,072	0,0042
$\beta_{5,2}$	-0,169	0,049	0,0006
α	0,437	0,029	0,0000
AIC	3,69395		
R^2_{FT}	66,9%		
R^2_{PE}	72,3%		

Da análise do Quadro 5.11 verifica-se que todos os parâmetros estimados das variáveis independentes têm um nível de confiança superior a 95% (pelo teste *t-student*), incluindo o parâmetro de dispersão (α), sugerindo assim que a distribuição BN é apropriada, tal como demonstra o teste de sobredispersão³⁹. O valor da medida AIC apresentado no mesmo quadro será utilizado no final deste subcapítulo para comparação dos diferentes modelos

³⁹ Valor superior a $\chi^2_{0,05GL=1}(3,84)$ rejeitando-se assim a hipótese da média igual à variância

aplicados ao total dos acidentes, não sendo, por isso, útil para a análise dos resultados nesta fase. As medidas R^2_{FT} e R^2_{PE} calculadas indicam a percentagem da componente sistemática explicada pelo modelo. Estes valores são próximos e situam-se na ordem dos 70%, resultado bastante satisfatório para um modelo de previsão que inclui variáveis independentes de carácter muito generalista na caracterização dos locais.

Para avaliar a importância de cada uma das variáveis no modelo analisaram-se os valores de AIC e do parâmetro de dispersão. Com a introdução individual de cada uma das variáveis verificou-se que o valor da medida AIC e do valor do parâmetro de dispersão diminuíram em qualquer caso, traduzindo assim uma melhoria do ajuste e, como tal, a importância destas variáveis no modelo.

Importa agora analisar se os valores estimados para os parâmetros apontam para uma relação das variáveis com os acidentes de acordo com o intuitivamente expectável e o referenciado na bibliografia. Considerando o valor da elasticidade obtido através da equação (4.20) e que no caso das funções logarítmicas corresponde ao valor estimado do parâmetro, verifica-se que, para o tráfego - $TMDA_{Tot}$, o valor da elasticidade indica que a um aumento de 10% desta variável corresponde um aumento de 3,5% do número de acidentes. Este valor traduz uma relação não-linear entre o tráfego e os acidentes e em concordância com os resultados apresentados em diversos trabalhos (Mountain, L., et al., 1996; Lord, D., 2000; Greibe, P., 2003), sendo contudo o valor da elasticidade obtido mais baixo do que o referido por esses autores. Esta diferença pode resultar da presença das outras variáveis independentes incluídas no modelo que, eventualmente, captam efeitos que nalguns autores são representados pela variável $TMDA$.

O valor da elasticidade da variável $Comp$ é próximo de 1 revelando uma relação quase linear entre os acidentes e o comprimento do arco (um aumento de 10% no comprimento do arco resulta num aumento de 9,1% dos acidentes). Este resultado é coerente com o apresentado por Mountain, L., et al. (1996).

O valor positivo do parâmetro da variável $DenInt$ indica que aumentando a densidade das intersecções no arco o número de acidentes também aumenta, tal como seria expectável. Como esta variável está relacionada com os acidentes por uma função exponencial, a elasticidade é calculada segundo (4.20) considerando o valor médio da variável $DenInt$. Assim, para o valor médio de 0,0046 a elasticidade é de 0,173, ou seja, para um aumento de 10% da variável $DenInt$ o acréscimo nos acidentes é de 1,7%. Este valor é baixo comparando, por exemplo, ao obtido por (Mountain, L., et al., 1996) (valor na ordem dos

3%) e ao valor da elasticidade do *Comp*. Este resultado pode advir do facto de à maior parte dos arcos corresponder um número de intersecções nulo (40% dos arcos).

O parâmetro estimado para a tendência temporal (γ) representada pela variável *ANO* é negativo indicando um decréscimo dos acidentes no período considerado. O valor estimado revela uma diminuição de ano para ano na ordem dos 5% ($\gamma = 0,95$), o que no total dos cinco anos corresponde a uma diminuição de 23%⁴⁰. Esta tendência para a diminuição dos acidentes era já patente no Quadro 5.4 e é coerente com outros trabalhos (Karlaftis, M. G. and Tarko, A. P., 1997; Mountain, L., et al., 1998; Kumara, S. P. and Chin, H. C., 2003). A forma funcional utilizada neste modelo para a tendência temporal é uma forma simplista em que se considera o efeito do tempo igual de ano para ano. Para analisar se esta simplificação não se afasta demasiado da realidade, estimou-se o mesmo modelo mas considerando para a tendência temporal quatro “*dummys*”, uma para cada ano e considerando o ano 2001 como base. Pelos resultados apresentados no Anexo 2.III, verifica-se que, de ano para ano, há sempre um decréscimo relativamente ao primeiro ano (2001), sendo esse decréscimo mais acentuado no ano de 2003. De referir que os valores estimados para os anos de 2002 e 2004 estão bastante abaixo do nível de confiança de 95%.

O facto da variação do decréscimo não ser igual de ano para ano tem especial importância quando se pretende realizar um estudo antes-depois ou quando há interesse por parte de determinadas entidades em identificar anos de elevado risco de sinistralidade para investigar possíveis causas (Lord, D. and Persaud, B., 2000). Contudo, o âmbito deste trabalho não se enquadra em nenhuma das situações referidas.

Numa base de dados temporal, isto é, com dados obtidos para um determinado período de tempo, existe a possibilidade de haver correlação temporal. Para aferir a existência de correlação temporal no modelo, procedeu-se a um teste de razão de verosimilhança tal como aplicado por Poch, M. and Mannering, F. (1996) para testar a hipótese dos parâmetros estimados pelo modelo com os dados dos cinco anos serem diferentes dos parâmetros estimados pelos cinco modelos referentes a cada ano⁴¹. Os resultados do teste apresentados no Anexo 2.IV apontam para a existência de correlação temporal na medida em que se obteve um nível de confiança de 85,7% para a hipótese de os parâmetros

⁴⁰ percentagem calculada a partir do valor $(1-0,95^5)$

⁴¹ Para $\chi = -2[3652,48 - (-765,11 - 748,03 - 662,63 - 737,51 - 710,76)] = 56,88$ e K graus de liberdade $(9 \times 5 - 9) = 36$ o valor α da probabilidade é 0,015

estimados no modelo para cada ano serem diferentes dos parâmetros estimados no modelo com os cinco anos. Esta correlação poderá resultar da variável $TMDA_{Tot}$ dado que, para além da variável ANO , esta é a única que varia com o tempo. De facto, tal como descrito anteriormente em 5.2.3., a base de informação do tráfego partiu do ano de 2005. Os valores para os outros anos foram calculados considerando, para a evolução do tráfego nos cinco anos, as mesmas das obtidas para as contadoras existentes na cidade.

Para avaliar se esta correlação temporal afecta os valores estimados dos parâmetros aplicou-se o mesmo modelo aos dados agregados nos cinco anos, isto é, a cada local corresponde uma observação igual à soma dos acidentes dos cinco anos e cujo tráfego corresponde ao valor médio dos cinco anos, eliminando-se desta forma a correlação temporal. Pela análise dos resultados apresentados em Anexo 2.V é possível concluir que os valores dos parâmetros do modelo agregado são muito próximos do modelo desagregado. As diferenças residem nos níveis de confiança das “*dummys*”. Este resultado confirma os resultados apresentados por Mountain, L., *et al.* (1998) e Lord, D. and Persaud, B. (2000) que concluíram que a correlação temporal afecta essencialmente os níveis de confiança da estimação. Nesse sentido, considerou-se o modelo simplificado sem tratamento da correlação temporal.

Para analisar os valores estimados dos parâmetros das “*dummys*” é necessário calcular a pseudo-elasticidade através da equação (4.21). Os valores calculados são apresentados no Quadro 5.12.

Quadro 5.12 - Pseudo-elasticidade das “*dummys*” da equação (5.2)

“ <i>dummys</i> ”	pseudo-elasticidade
$OcSolo_2$	-0,400
$OcSolo_3$	0,332
$OcSolo_4$	-0,370
$OcSolo_5$	0,187
$ClHier_2$	-0,184

Analisando esses valores e considerando o tipo de ocupação do solo 1 como sendo a base de comparação (é o tipo de ocupação de solo predominante como se pode concluir pelo Quadro 5.9), verifica-se que mantendo-se todas as outras características representadas pelas outras variáveis, o tipo de ocupação do solo 2 e 4 representam um decréscimo nos acidentes de 40% e 37%, respectivamente em relação ao tipo de solo 1. Dado que a ocupação do solo do tipo 2 corresponde a áreas de habitação do tipo unifamiliar e de edificação isolada, ou seja, áreas essencialmente habitacionais que geram um menor número de conflitos entre o tráfego motorizado e entre o mesmo e o tráfego pedonal, pode considerar-se que o resultado está de acordo com as características referidas. A ocupação do solo do tipo 4 identifica áreas de equipamento que no caso do Porto estão concentradas em duas zonas (ver Figura 5.3): uma das zona abrange um pólo universitário e um hospital e a outra também um pólo universitário e um pavilhão desportivo. A diminuição do risco de acidentes nestas zonas pode dever-se a factores, tais como, pouca densidade de construção, infra-estruturas viárias adequadas e a sazonalidade do tráfego (períodos de tempo com elevado congestionamento e número de conflitos em oposição a outros períodos de tempo com reduzido congestionamento e conflitos de tráfego). A ocupação do solo tipo 3 - área empresarial, concentra-se numa única zona do Porto. Esta zona caracteriza-se por dois tipos de funções: durante o dia verifica-se essencialmente um elevado tráfego misto (pesado e ligeiro) para acesso às empresas e com um número elevado de estacionamento para cargas e descargas muitas vezes em situação ilegal; durante a noite é uma zona de diversão nocturna (bares, discotecas, restaurantes, etc.) e como tal com potenciais comportamentos de maior risco por parte dos condutores. Nesse sentido, o valor do parâmetro estimado, indicativo de aumento de acidentes nos arcos situados nesta zona, está também de acordo com o expectável. Finalmente, a ocupação do solo do tipo 5 com a designação de área histórica, concentra-se no centro da cidade onde se situam comércio e serviços. Esta é uma zona caracterizada por vias geometricamente pouco adaptadas às funções atribuídas, havendo também uma elevada densidade de construção. Caracteriza-se ainda por ser uma zona com elevada afluência de peões em conjunto com a existência de várias paragens de transporte público. Por todos estes factores, seria de esperar um aumento dos acidentes nos arcos desta zona, tal como revelado pelo valor estimado.

Relativamente à classificação hierárquica das vias, e tendo como base de comparação as vias distribuidoras principais, verifica-se um decréscimo nos acidentes nos arcos das vias distribuidoras locais em relação às distribuidoras principais. Este decréscimo não contraria as características apontadas por esta classe hierárquica na medida em que estas vias são, em geral, geometricamente mais urbanas, isto é, com características que se traduzem num comportamento do condutor de menor risco, nomeadamente na prática de velocidades mais reduzidas.

Para aferir da qualidade de ajuste do modelo aplicou-se ainda o método CURE. Este método consiste na elaboração de gráficos de resíduos acumulados dos acidentes, ordenados de acordo com uma variável independente. Se o acumulado dos resíduos não apresentar o padrão esperado – oscilar em torno do eixo do zero, terminar próximo do valor nulo e estar contido entre as bandas definidas por $\pm 2\sigma$, pode indicar uma eventual deficiência conceptual do modelo. Assim, elaboraram-se os gráficos para as variáveis *Comp*, *TMDA_{Tot}* e *DenInt* apresentados na Figura 5.8.

O gráfico dos resíduos acumulados em ordem ao *TMDA_{Tot}* da Figura 5.8 mostra que a linha dos resíduos acumulados excede a banda $\pm 2\sigma$ para os intervalos de valores do tráfego entre 10 000 e 25 000 e acima de 43 000, e não termina no valor nulo. Destes factos pode concluir-se que o modelo aplicado não se ajusta a todos os valores do tráfego em especial para valores elevados. O facto do intervalo de valores do tráfego ser muito extenso, denotado pela diferença entre o valor mínimo e máximo do Quadro 5.8 e que reflecte também a heterogeneidade deste meio urbano, parece aconselhar o desenvolvimento de dois modelos distintos em função dos valores de tráfego em análise. Contudo, esta solução pode beneficiar o ajuste desta variável mas prejudicar o de outras variáveis.

O gráfico da Figura 5.8, relativo à variável *Comp*, apresenta uma curva segundo o padrão desejado com excepção do valor final dos resíduos acumulados não terminar no valor nulo. Verifica-se que apenas nos valores elevados do comprimento dos arcos, mais concretamente os valores de 2 743, 2 820 e 3 343 metros, os resíduos acumulados se afastam da banda delimitadora $\pm 2\sigma$. Estes valores são uma excepção podendo, por esse facto, corresponder a observações “outliers”.

Também no caso da variável *DenInt*, para valores mais elevados desta variável os resíduos acumulados ultrapassam a banda da curva $\pm 2\sigma$.

Pelos gráficos da Figura 5.8 verifica-se que para valores elevados das três variáveis analisadas os resíduos acumulados excedem as curvas $\pm 2\sigma$ e afastam-se do valor nulo.

Para analisar se existe uma eventual deficiência conceptual do modelo, nomeadamente na forma funcional considerada para as variáveis independentes (com excepção das “*dummys*”), elaboraram-se os mesmos gráficos para comparar as diferentes hipóteses (função Potência, Gama e Exponencial). Os gráficos CURE resultantes são apresentados no Anexo 2.VI. Como se pode observar pelos respectivos gráficos, os problemas detectados no gráfico da Figura 5.8 não são resolvidos e em alguns caso o problema é até agravado.

Analisando os valores dos resíduos para as observações cujo valor de $TMDA_{Tot}$ é superior a 43 000, é possível detectar observações cujos valores previstos de acidentes são substancialmente sobrestimados contribuindo assim para os valores negativos elevados dos resíduos. Nesse sentido, realizou-se uma análise aos dados de forma a identificar a eventual presença de “*outliers*” e a sua influência na base de dados.

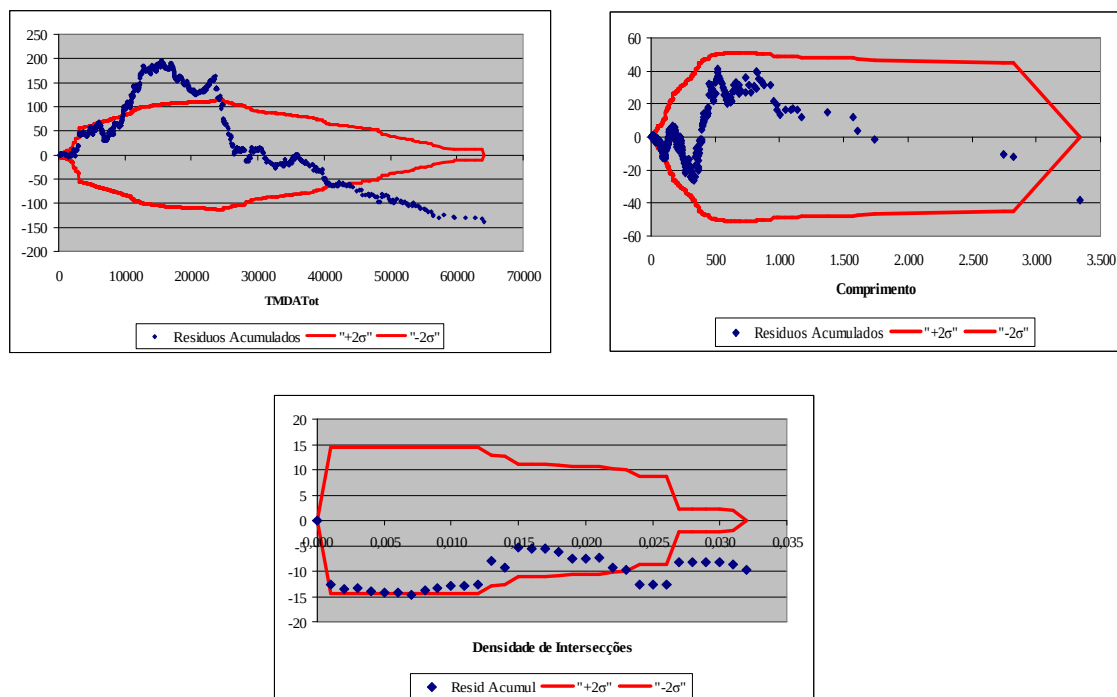


Figura 5.8. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, *Comp* e *DenInt* elaborados segundo o método CURE para o modelo BN

Os “*outliers*” são definidos como valores observados significativamente diferentes dos restantes valores da base de dados. Estes ocorrem quer porque são valores genuinamente diferentes dos outros valores da base de dados quer devido, por exemplo, a erros no registo

e/ou na recolha da base de dados. Para identificar a presença de “outliers” na base de dados considerou-se um procedimento baseado na estatística “leverage”. No caso de modelos com várias variáveis independentes, o processo de identificação dos “outliers” consiste em determinar a distância de cada ponto (observação) ao centroide formado pelos outros pontos (observações). A aplicação deste processo tem como base a construção de uma matriz designada de “hat matrix” que cruza o vector dos valores observados com o vector dos valores previstos. Os valores da diagonal desta matriz, designados de valores “leverage” da base de dados, são comparados com $2K/n$ (K parâmetros e n observações). Os valores superiores a $2K/n$ correspondem aos “outliers”. Os valores “leverage” para cada observação estão representados na Figura 5.9, assim como o valor corresponde a $2K/n$.

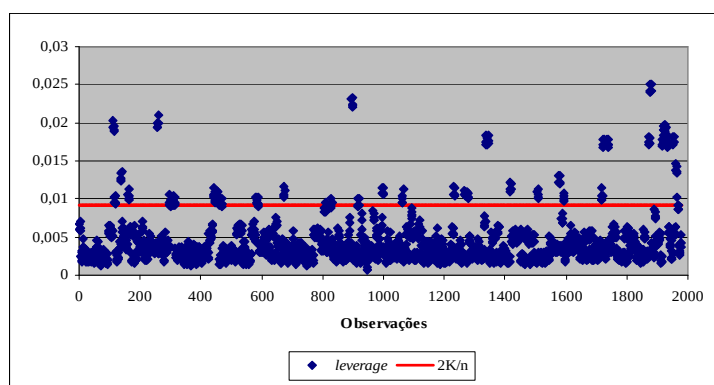


Figura 5.9. - Gráfico dos valores de “leverage” e de $2K/n$ referente à base de dados dos arcos

Os pontos do gráfico acima da barra vermelha correspondem aos “outliers”. Nestas condições, foram identificados 208 “outliers” na base de dados. Embora os pontos identificados como “outliers” tenham um elevado valor de “leverage”, não implica necessariamente que influenciem o valor da estimação dos parâmetros da regressão. Para analisar a influência das observações na estimação é necessário complementar a análise dos “outliers” com uma análise de influência. Para tal, considerou-se a função Df_{iti} que mede o efeito da observação no valor ajustado associado (Greene, W., 2007). O gráfico da Figura 5.10 mostra os valores resultantes desta função. Do cruzamento das duas análises resultou um total de 13 pontos identificados como “outliers” e cujos valores da função Df_{iti} são maiores que 0,2 ou menores que -0,2.

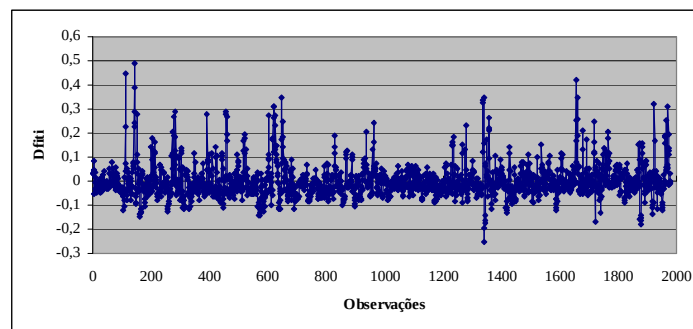


Figura 5.10. - Gráfico dos valores da função $Dfiti$ das observações registradas nos arcos

Estes pontos foram retirados da base de dados e realizou-se um reajuste do modelo cujos resultados são apresentados no Anexo 2.VII. Pelos resultados verifica-se que não houve uma significativa alteração dos valores dos parâmetros estimados em relação ao modelo inicial.

Os gráficos elaborados segundo o método CURE para estes resultados são apresentados na Figura 5.11.

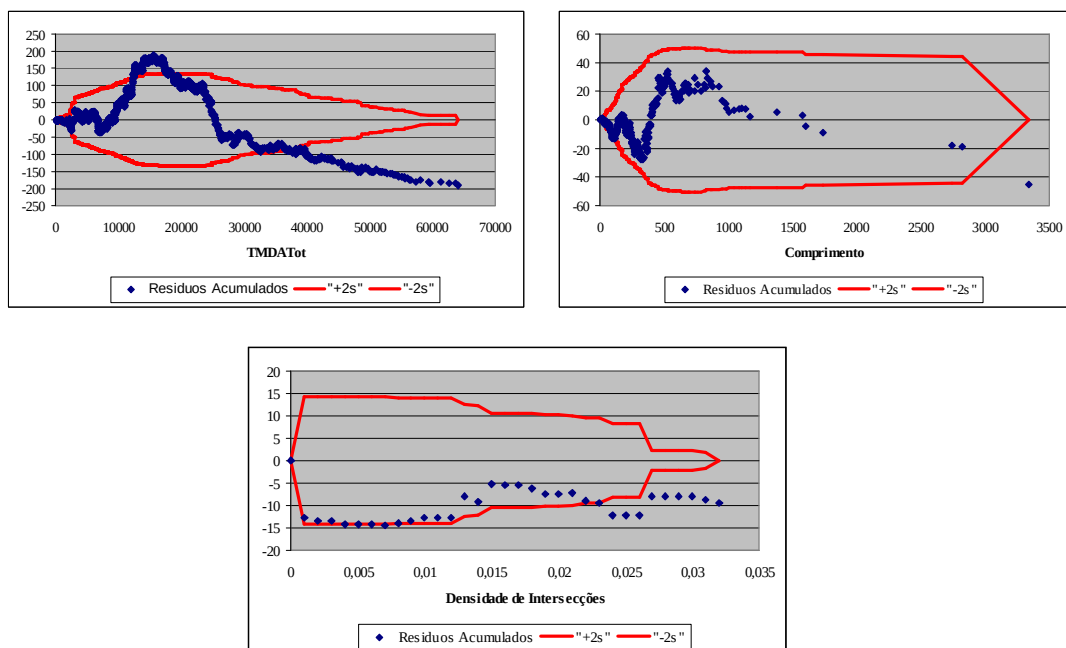


Figura 5.11. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE sem as observações “outliers” e de influência

Como se pode verificar pela Figura 5.11 os gráficos dos resíduos acumulados não apresentam alterações significativas em comparação com os gráficos da Figura 5.8.

Mantendo a mesma estrutura da equação (5.2) mas de forma a melhorar o ajuste do modelo consideraram-se modelos alternativos cujas aplicações a seguir se descreve.

Modelo com heteroscedasticidade

Para a aplicação dos MLG assume-se que o termo do erro é homoscedástico⁴², isto é, variância constante entre observações – $E(\varepsilon^2) = \sigma^2$. Contudo, esta suposição pode ser violada, particularmente na análise de dados espaciais designando-se esse fenómeno de heteroscedasticidade⁴³ - $E(\varepsilon^2) = \sigma_i^2$ ($i = 1, 2, \dots, n$). A heteroscedasticidade pode ocorrer quando factores não directamente observados pelo modelo variam entre observações de uma forma tendenciosa que se pode reflectir na variância que acompanha a variação de algumas variáveis do modelo. Assim, analisando as variáveis incluídas no modelo e o seu efeito nos acidentes, pode considerar-se a hipótese de, por exemplo, a variável $TMDA_{Tot}$ estar correlacionada com outros factores não observados pelo modelo, tais como número e tipo de conflitos de tráfego, ou ainda, a variável $Comp$ na medida que ao aumento do comprimento dos arcos se pode associar um conjunto de características geométricas ou de comportamentos do condutor, tais como maior variação da velocidade de circulação. Estes factos podem estar na origem dos elevados valores dos resíduos acumulados verificados pelos gráficos CURE da Figura 5.8.

Um dos métodos possíveis para eliminar a heteroscedasticidade presente no modelo base (distribuição BN) consiste em dividir a regressão por um factor c_i em que $\sigma_i^2 = c_i \sigma^2$. O quadrado deste factor é determinado pelo método dos mínimos quadrados ponderados⁴⁴, particularizando que a variância é proporcional a uma das variáveis do modelo. Assim, para a aplicação deste método analisaram-se várias hipóteses para a variável de ponderação sendo que as únicas que se verificaram estarem correlacionadas com o erro foram as variáveis $TMDA_{Tot}$ e $Comp$. Seleccionou-se, das duas hipóteses, o modelo ponderado pela variável $TMDA_{Tot}$ por apresentar nos gráficos CURE um padrão dos resíduos acumulados mais satisfatórios.

Pelos resultados apresentados no Quadro 5.13 (ver no Anexo 2.VIII os resultados completos) verifica-se que os valores estimados dos parâmetros são muito próximos dos

⁴² homoscedastic, em inglês

⁴³ heteroskedasticity, em inglês

⁴⁴ weighted least squares, em inglês

valores estimados pelo modelo base, incluindo o valor do parâmetro de dispersão. Os valores calculados das medidas R^2_{FT} e R^2_{PE} correspondem a uma percentagem da componente sistemática explicada pelo modelo na ordem dos 68%, ou seja, ligeiramente inferior aos valores das mesmas medidas calculadas para o modelo base.

Para comparar estes dois modelos com o modelo base aplicou-se ainda o teste de rácio verosimilhança⁴⁵ cujo resultado indica que o modelo base é preferível. A medida AIC reforça também este resultado.

Quadro 5.13 - Resultados da estimação do modelo com heteroscedasticidade da equação (5.2)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-6,456	0,410	0,0000
β_1	0,287	0,038	0,0000
β_2	0,843	0,026	0,0000
$\ln(\gamma)$	-0,047	0,015	0,0013
β_3	36,350	4,185	0,0000
$\beta_{4,2}$	-0,246	0,051	0,0000
$\beta_{4,3}$	0,444	0,174	0,0107
$\beta_{4,4}$	-0,207	0,094	0,0275
$\beta_{4,5}$	0,229	0,067	0,0006
$\beta_{5,2}$	-0,110	0,051	0,0323
α	0,429	0,026	0,0000
AIC	4,14063		
R^2_{FT}	67,9%		
R^2_{PE}	69,1%		

Contrariamente ao que indicam o teste do rácio de verosimilhança e a medida AIC, os gráficos dos resíduos acumulados (CURE) para esta aplicação apresentam resultados mais próximos dos padrões desejados. Estes gráficos são apresentados na Figura 5.12.

Pelo gráfico da Figura 5.12 verifica-se que para valores de tráfego acima dos 53 000 os resíduos acumulados continuam a ultrapassar a banda $\pm 2\sigma$, embora com valores mais próximos de zero relativamente ao gráfico do modelo base. Os gráficos relativos às variáveis *Comp* e *DenInt* apresentam um padrão idêntico aos anteriores mas com uma

⁴⁵ $\chi^2 = 535,90$ e $\chi^2 = 884,42$ para a variável de ponderação *Comp* e $TMDA_{Tot}$, respectivamente superiores a $\chi^2_{GL=1}(0,01) = 6,63$

aproximação dos resíduos acumulados ao valor nulo mantendo-se, no entanto, algum afastamento nos valores mais altos desta variável.

Pode concluir-se assim que, pelos gráficos CURE, o ajuste dos valores previstos aos observados é melhor, denotando-se um eventual efeito da variável $TMDA_{Tot}$ nos resíduos da regressão, embora a correcção da heteroscedasticidade não se tenha repercutido nas medidas de avaliação do ajuste consideradas.

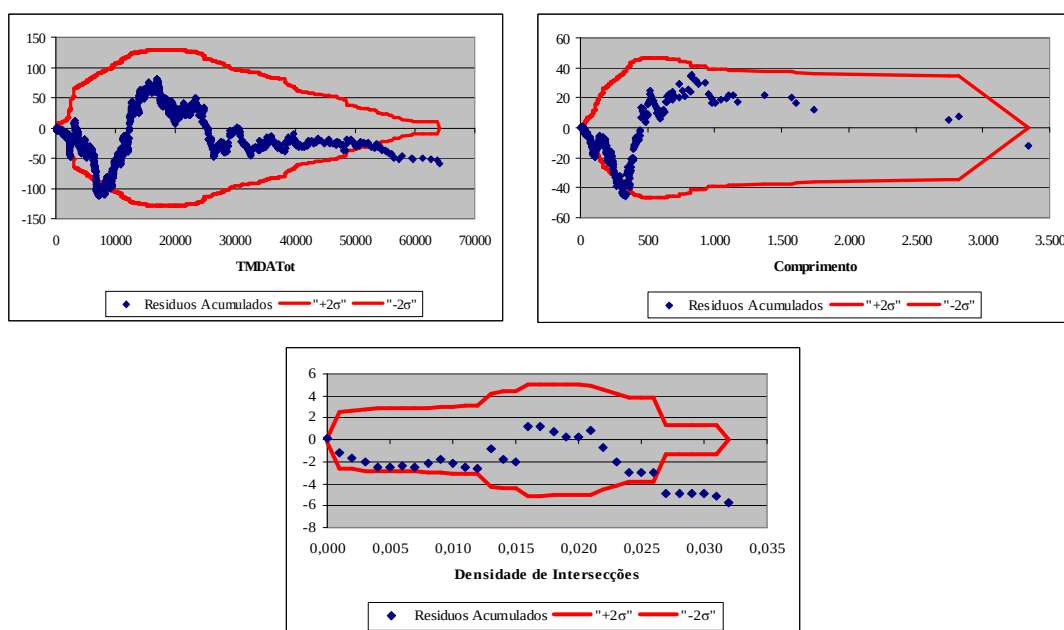


Figura 5.12. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo com heteroscedasticidade da equação (5.2)

Modelo com Parâmetros Aleatórios

Os dados em painel⁴⁶ (dados espaciais e temporais) são, geralmente, considerados no estudo da sinistralidade. Este tipo de dados permite analisar a diferença de comportamento entre entidades, reflectindo assim a heterogeneidade inerente à ocorrência de acidentes. Para analisar a heterogeneidade dos dados em painel, utilizam-se modelos específicos, tais como o modelo com parâmetros aleatórios (PA), que permitem flexibilizar a modelação dos acidentes ajustando-se assim melhor aos dados.

⁴⁶ *panel data*, em inglês

Nesse sentido, considerou-se a aplicação de um modelo com parâmetros aleatórios ao total dos acidentes. A heterogeneidade dos dados é considerada no modelo através de uma distribuição probabilística que representa a variação do valor estimado dos parâmetros em torno do valor médio tal como mostra a equação:

$$\beta_i = \beta_i^0 + \gamma_i \nu_i \quad (5.3)$$

em que β_i^0 é o valor médio a ser estimado, γ_i é o factor de escala e ν_i é a variação aleatória definida por uma distribuição probabilística (normalmente a distribuição normal, uniforme ou triangular), para cada variável i .

O método de estimação utilizado foi o da máxima verosimilhança baseada na simulação⁴⁷ (os parâmetros estimados são aqueles que maximizam a função logarítmica de verosimilhança simulada) através da “*Halton draws*”. Várias hipóteses foram testadas variando o tipo de distribuição considerada para a variável de resposta (Poisson ou BN), as variáveis a considerar com parâmetros fixos ou aleatórios e o tipo de distribuição probabilística para esta variação aleatória. A hipótese seleccionada corresponde ao modelo com distribuição de Poisson para a variável de resposta e uma distribuição normal para todos os parâmetros com variação aleatória. Considerando a significância estatística do factor de escala estimado para o PA, constatou-se que todas as variáveis, excepto as “*dummys*” *OcSolo3* e a *OcSolo4*, têm uma variação aleatória entre observações.

Não se optou pelo modelo com distribuição BN para a variável de resposta, por este apresentar um valor do parâmetro de dispersão muito elevado (40,672 com um valor de $P[Z > z] = 0,0067$).

Os resultados da hipótese seleccionada são apresentados no Quadro 5.14 (ver resultados completos no Anexo 2.IX). Para as variáveis com PA os resultados compreendem o valor médio e o factor de escala do parâmetro $i - \beta_i^0$ e γ_i , respectivamente da equação (5.3). Este factor de escala corresponde ao desvio padrão no caso da distribuição normal.

Pelo Quadro 5.14 constata-se que os valores médios dos parâmetros e os factores de escala dos PA foram estimados com um nível de confiança superior a 95%. Os valores médios estimados dos PA não alteraram a relação positiva ou negativa com os acidentes obtida pelo modelo base. Analisando, por exemplo, o resultado da variável “*dummy*” *OcSolo2*

⁴⁷ *simulation-based maximum likelihood*, em inglês

com parâmetro regido pela distribuição normal cujo valor médio é -0,546 e o desvio padrão é 0,413 (91% da distribuição é inferior a zero), pode concluir-se que na maior parte dos arcos resulta um decréscimo das ocorrências de acidentes e apenas para um pequeno número de arcos (9%) se verifica um aumento dos acidentes para esta “dummy”. Recordar-se que no modelo base o parâmetro estimado para esta variável é igual a -0,336.

Quadro 5.14 - Resultados da estimação do modelo com PA da equação (5.2)

Total dos Acidentes					
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]	Factor Escala	P[Z>z]
$\ln(\beta_0)$	-6,474	0,192	0,0000	0,113	0,0000
β_1	0,190	0,018	0,0000	0,066	0,0000
β_2	0,988	0,018	0,0000	0,043	0,0000
$\ln(\gamma)$	-0,063	0,009	0,0000	0,067	0,0000
β_3	45,253	3,273	0,0000	25,658	0,0000
$\beta_{4,2}$	-0,546	0,039	0,0000	0,413	0,0000
$\beta_{4,3}$	0,388	0,069	0,0000	-	-
$\beta_{4,4}$	-0,384	0,058	0,0000	-	-
$\beta_{4,5}$	0,139	0,038	0,0002	0,120	0,0003
$\beta_{5,2}$	-0,278	0,030	0,0000	0,155	0,0000
AIC	3,39317				

Pelo cálculo do teste do rácio de verossimilhança⁴⁸, o modelo com PA considerado é preferível ao modelo base. O valor de AIC comprova mais uma vez o resultado do teste. Dada a dificuldade de determinar os valores previstos para as observações com base em parâmetros aleatórios, sendo praticável apenas para os valores médios dos mesmos, não é razoável elaborar os gráficos CURE e calcular a medida R^2_{FT} .

Pelos resultados desta aplicação pode concluir-se que na maior parte das variáveis independentes incluídas no modelo se verifica uma variação aleatória entre observações que o modelo PA traduz com um melhor ajuste do que o modelo base.

Modelo Zero-Inflacionado

Como referido anteriormente, no histograma da Figura 5.6 é possível observar uma concentração das observações no número de acidentes nulo. O registo de zero acidentes num determinado período de tempo pode indicar que o local é seguro, ou, pelo contrário,

⁴⁸ $\chi^2 = 13,15$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

que esse registo resulta de um período de tempo excepcional ou de não se ter efectuado nenhum registo nesse mesmo período (Kumara, S. P. and Chin, H. C., 2003). Com o objectivo de traduzir estes possíveis acontecimentos em simultâneo com a ocorrência de determinado número de acidentes, aplicou-se à base de dados um modelo binário designado de modelo zero-inflacionado.

Este modelo descreve a ocorrência dos acidentes como resultado de um de dois estados: por um lado, um estado cujo resultado é sempre zero acidentes – estado “zero acidentes”, por outro lado, um segundo estado que corresponde ao processo aleatório de ocorrência de acidentes habitualmente descrito pela distribuição probabilística de Poisson ou BN e que pode gerar zero ou outro número de acidentes – estado normal. Matematicamente este duplo estado é representado pela equação (4.15). Esta mesma equação pode ser traduzida de uma forma simplificada: considere-se z um indicador binário em que $z = 0$ para o estado “zero acidentes” e $z = 1$ para o estado normal e y^* corresponde ao resultado do processo de Poisson ou BN do estado normal; então o número de acidentes y observado é igual a $z \times y^*$. Assim, para o modelo estruturado segundo a equação (5.2) e considerando o total dos acidentes, aplicou-se a regressão Zero-Inflacionado Poisson (ZIP) e Zero-Inflacionado Binomial Negativa (ZIBN) com distribuição de Poisson e BN para a variável de resposta, respectivamente. Para estas aplicações, formularam-se várias hipóteses para caracterizar os dois estados, nomeadamente a selecção das variáveis independentes que afectam os estados e do tipo de distribuição probabilística do binário que descreve a probabilidade de ocorrer determinado estado (geralmente a distribuição probit ou logit).

Para avaliar as várias hipóteses, utilizou-se o teste Vuong formulado pela equação (4.16). Segundo este teste, para valores de V inferiores a -1,96 (valor crítico da distribuição Normal) os modelos são rejeitados.

Das várias hipóteses testadas, a hipótese com maior valor do teste Vuong (6,76) corresponde a um modelo com distribuição de Poisson para variável de resposta (ZIP), com distribuição probabilística probit para a probabilidade de ocorrer determinado estado e com as mesmas variáveis independentes a condicionar os dois estados de geração de acidentes. As variáveis independentes que afectam o valor esperado do estado normal estão relacionadas com a probabilidade q_i do estado “zero acidentes” por um parâmetro τ tal que $\text{logit}(q_i) = \tau \beta x_i$, em que β são os parâmetros a serem estimados e x_i as variáveis independentes consideradas.

Dos resultados apresentados no Quadro 5.15 (ver resultados completos no Anexo 2.X) verifica-se que o valor estimado do parâmetro τ apresenta um nível de confiança superior a 95%, assim como todos os outros parâmetros estimados. Se compararmos os valores dos parâmetros estimados pelo modelo ZIP com os valores estimados pelo modelo base, verifica-se que o sinal positivo ou negativo da elasticidade mantém-se havendo apenas ligeiras alterações nos valores absolutos.

O modelo ZIP aplicado, tal como se pode constatar do Anexo 2.X, determinou um número de observações previstas de zero acidentes igual a 576,3, sendo que o número de registos observados na base de dados deste mesmo valor nulo de ocorrências é de 669. Já pela aplicação do modelo Poisson o número de observações previstas para essas ocorrências é de 334. Como se pode depreender, o modelo ZIP prevê um número de observações de zero ocorrências mais próximo do número observado do que o modelo Poisson.

Quadro 5.15 - Resultados da estimação do modelo ZIP da equação (5.2)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-4,924	0,137	0,0000
β_1	0,252	0,013	0,0000
β_2	0,677	0,010	0,0000
$\ln(\gamma)$	-0,053	0,006	0,0000
β_3	37,003	2,027	0,0000
$\beta_{4,2}$	-0,238	0,024	0,0000
$\beta_{4,3}$	0,373	0,057	0,0000
$\beta_{4,4}$	-0,265	0,040	0,0000
$\beta_{4,5}$	0,127	0,023	0,0000
$\beta_{5,2}$	-0,152	0,020	0,0000
τ	-1,434	0,074	0,0000
R^2_{FT}	54,2%		

A percentagem da componente sistemática explicada pelo modelo ZIP é de 54%, segundo a medida R^2_{FT} , valor este inferior aos valores obtidos pelos modelos anteriores. O teste do rácio de verosimilhança⁴⁹ aponta como preferível o modelo base em detrimento do ZIP.

⁴⁹ $\chi^2 = 449,58$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

Análise comparativa dos modelos

As várias alternativas ao modelo base apresentadas anteriormente visam tentar melhorar o ajuste aos dados, analisando-se dessa forma qual o modelo que traduz melhor a ocorrência do total dos acidentes nos arcos.

Para melhor comparar os vários modelos aplicados resumem-se a seguir, no Quadro 5.16, os resultados obtidos.

Os valores apresentados no Quadro 5.16 para o modelo com PA correspondem ao valor médio no caso das variáveis com parâmetros aleatórios. A distribuição probabilística da variável de resposta corresponde à distribuição BN para os dois primeiros modelos e à distribuição de Poisson nos dois últimos modelos.

Quadro 5.16 - Valores estimados e níveis de confiança para os diferentes modelos da equação (5.2) aplicados à base de dados dos arcos

Parâmetro	Total dos Acidentes							
	Modelo base		Heterosc ($TMDA_{Tot}$)		Modelo PA		ZIP	
	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]
$\ln(\beta_0)$	-7,318	0,0000	-6,456	0,0000	-6,474*	0,0000	-4,924	0,0000
β_1	0,344	0,0000	0,287	0,0000	0,190*	0,0000	0,252	0,0000
β_2	0,910	0,0000	0,843	0,0000	0,988*	0,0000	0,677	0,0000
$\ln(\gamma)$	-0,056	0,0003	-0,047	0,0013	-0,063*	0,0000	-0,053	0,0000
β_3	37,635	0,0000	36,350	0,0000	45,253*	0,0000	37,003	0,0000
$\beta_{4,2}$	-0,336	0,0000	-0,246	0,0000	-0,546*	0,0000	-0,238	0,0000
$\beta_{4,3}$	0,403	0,0044	0,444	0,0107	0,388	0,0000	0,373	0,0000
$\beta_{4,4}$	-0,315	0,0011	-0,207	0,0275	-0,384	0,0000	-0,265	0,0000
$\beta_{4,5}$	0,207	0,0042	0,229	0,0006	0,139*	0,0002	0,127	0,0000
$\beta_{5,2}$	-0,169	0,0006	-0,110	0,0323	-0,278*	0,0000	-0,152	0,0000
α	0,437	0,0000	0,429	0,0000	-	-	-	-
AIC	3,69395		4,14063		3,39317		-	
R^2_{FT}	66,9%		67,9%		-		54,2%	
R^2_{PE}	72,3%		69,1%		-		-	
função logaritmica de verosimilhança ⁵⁰	-3646,01		-4088,22		-3341,24		-3870,90	

*valores médios

Pelo Quadro 5.16 constata-se que os valores estimados dos parâmetros mantêm os mesmos sinais e a mesma ordem de grandeza em qualquer um dos modelos aplicados e todos

⁵⁰ *log likelihood function*, em inglês

determinados com níveis de confiança superiores a 95%. Se no caso das variáveis contínuas este resultado está de acordo outros trabalhos que têm demonstrado a importância das mesmas na modelação dos acidentes, já no caso das variáveis “dummy” este facto vem salientar a importância da presença destas no modelo para caracterizar a ocorrência de acidentes.

O facto de, no modelo ZIP e no modelo com PA, a distribuição de Poisson ser a que corresponde aos melhores resultados, pode indiciar que estes modelos, cada qual com princípios distintos, tratam a dispersão dos dados que o modelo BN capta como sobredispersão. Contudo, tal como referido anteriormente em 4.1, o modelo ZIP baseia-se em princípios conceptuais que suscitam dúvidas quanto à sua aplicabilidade à modelação dos acidentes (Lord, D., et al., 2005).

Pelos valores da medida AIC pode concluir-se que o modelo com PA é o mais ajustado aos dados. Esta medida não é determinável para o modelo ZIP, não se podendo assim comparar com este. Os vários testes de rácio de verosimilhança realizados aos diferentes modelos em comparação com o modelo base indicam que apenas o modelo com PA é preferível ao modelo base.

Embora o modelo PA seja o mais ajustados aos dados, é importante salientar que um modelo bem ajustado pode não ser o melhor modelo de previsão tal como demonstrou Partyka, S. C. (1991) referido em Reurings, M., *et al*, (2005). A aplicação do modelo PA na previsão da frequência de acidentes num determinado ano, obriga a basear o cálculo no valor médio dos parâmetros aleatórios, o que aumenta o intervalo de confiança do resultado.

Modelo com função Translog

Tendo-se verificado pelo gráfico CURE do modelo base um desajuste dos valores previstos para determinados intervalos de valores das variáveis contínuas analisadas, considerou-se ainda a hipótese das elasticidades das variáveis não serem constantes. Nesse sentido, considerou-se ainda a modelação dos acidentes com base numa nova equação que traduz a flexibilidade das relações entre variáveis independentes e a variável de resposta.

O modelo base aplicado à equação (5.2) é um modelo rígido na medida em que pressupõe que a elasticidade das variáveis sejam independentes da magnitude da própria ou das outras variáveis. Dado que a curva dos resíduos acumulados apresentada nos gráficos da Figura 5.8 indicia a existência de um comportamento distinto dos resíduos para valores

mais elevados das mesmas variáveis, considerou-se a análise de um modelo com função Translog. A função Translog atende ao efeito do cruzamento das variáveis contínuas entre si e assim flexibiliza a estrutura do modelo pois não restringe a variável a um único valor de elasticidade. No caso das variáveis consideradas no estudo, e mantendo as formas funcionais das mesmas tal como representadas pela equação (5.2), a equação do modelo que formula o efeito do cruzamento entre essas variáveis é:

$$\ln E[Y] = \ln(\beta_0) + \beta_1 \ln(TMDA_{Tot}) + \beta_2 \ln(Comp) + \beta_3 DenInt + \beta_4 [\ln(TMDA_{Tot})]^2 + \beta_5 [\ln(TMDA_{Tot}) DenInt] + \beta_6 [\ln(TMDA_{Tot}) \ln(Comp)] + \beta_7 [\ln(Comp)]^2 + \beta_8 [\ln(Comp) DenInt] + \beta_9 (DenInt)^2 + (\ln \gamma) ANO + \beta_{10,k} OcSolo_k + \beta_{11,g} ClHier_g \quad (5.4)$$

em que k identifica a ocupação do solo e varia entre 1 e 5, e g identifica a classe hierárquica do arco com o valor 1 ou 2.

Para estas variáveis realizou-se uma análise das correlações através da matriz de correlações cujos resultados são apresentados no Quadro 5.17.

Quadro 5.17 - Matriz de correlação entre as variáveis independentes cruzadas da equação (5.4)

	$TMDA_{Tot}$	$Comp$	$DenInt$	$TMDA_{Tot}^* TMDA_{Tot}$	$TMDA_{Tot}^* DenIn$	$TMDA_{Tot}^* Comp$	$Comp^* Comp$	$Comp^* DenIn$	$DenIn^* DenIn$
$TMDA_{Tot}$	1,00	-	-	-	-	-	-	-	-
$Comp$	0,11	1,00	-	-	-	-	-	-	-
$DenIn$	0,05	0,23	1,00	-	-	-	-	-	-
$TMDA_{Tot}^* TMDA_{Tot}$	0,99	0,11	0,03	1,00	-	-	-	-	-
$TMDA_{Tot}^* DenIn$	0,11	0,24	0,99	0,09	1,00	-	-	-	-
$TMDA_{Tot}^* Comp$	0,53	0,90	0,21	0,53	0,24	1,00	-	-	-
$Comp^* Comp$	0,14	0,99	0,21	0,14	0,22	0,90	1,00	-	-
$Comp^* DenIn$	0,06	0,36	0,97	0,05	0,97	0,33	0,35	1,00	-
$DenIn^* DenIn$	0,02	-0,02	0,87	0,01	0,86	-0,01	-0,03	0,76	1,00

Analisando a matriz de correlação das variáveis apresentada no Quadro 5.17, verifica-se uma forte possibilidade de correlação entre algumas das variáveis cruzadas. Por esse facto, e de forma a evitar a existência de fenómenos de colineariedade, optou-se por aplicar

novamente o modelo retirando as variáveis cruzadas com um valor de correlação superior a 0,90, ou seja, todas as variáveis cruzadas excepto a variável que cruza o $TMDA_{Tot}$ com o $Comp$. Admitiu-se para a aplicação a mesma distribuição probabilística do modelo base, ou seja, a distribuição BN. Os resultados são apresentados no Quadro 5.18 (no Anexo 2.XI apresentam-se todos os resultados da aplicação)

Quadro 5.18 - Resultados da estimação do modelo com função Translog⁵¹ da equação (5.4)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-14,295	1,443	0,0000
β_1	1,068	0,151	0,0000
β_2	2,146	0,262	0,0000
$\ln(\gamma)$	-0,055	0,016	0,0004
β_3	36,200	4,552	0,0000
β_6	-0,128	0,027	0,0000
$\beta_{10,2}$	-0,314	0,055	0,0000
$\beta_{10,3}$	0,397	0,142	0,0053
$\beta_{10,4}$	-0,316	0,095	0,0009
$\beta_{10,5}$	0,205	0,071	0,0040
$\beta_{11,2}$	-0,182	0,049	0,0002
α	0,423	0,029	0,0000
AIC	3,68803		
R^2_{FT}	68,7%		
R^2_{PE}	73,2%		

Pelo quadro verifica-se que todos os parâmetros estimados têm significância estatística. Os valores das medidas R^2_{FT} e R^2_{PE} apontam para uma elevada percentagem da componente sistemática explicada pelo modelo, com valores ligeiramente superiores aos do modelo base.

Pelos valores dos parâmetros estimados das variáveis $TMDA_{Tot}$ e $TMDA_{Tot} * Comp$ pode concluir-se que o aumento do volume de tráfego, não atendendo ao efeito do comprimento, aumenta o risco de acidente de uma forma praticamente linear, no entanto, a magnitude deste efeito tende a diminuir para valores mais elevados do comprimento do arco, tal como demonstra o valor negativo do parâmetro da variável $\ln(TMDA_{Tot}) \times \ln(Comp)$. O mesmo se

⁵¹ dado que o modelo não considera todas as variáveis cruzadas, a designação Translog não é a mais correcta, sendo apenas considerada por simplificação

pode concluir para o efeito da variável *Comp*, ou seja, com o aumento do comprimento do arco, aumenta o risco de acidente, no entanto, a magnitude deste efeito tende a diminuir com o aumento do volume de tráfego. Relativamente aos valores dos parâmetros estimados das outras variáveis, verifica-se que estes não se alteraram significativamente em relação aos valores estimados pelo modelo base. A exceção é o valor do parâmetro estimado para o termo da constante β_0 , devendo-se provavelmente ao facto do modelo incluir mais uma variável, captando parte do efeito parcialmente absorvido no termo da constante dos modelos anteriormente analisados.

O valor de AIC é muito próximo do valor do modelo base, sendo que o teste do rácio de verosimilhança⁵² indica que o modelo Translog é melhor do que o modelo base. Os gráficos CURE apresentados na Figura 5.13 também apontam para um melhor ajuste deste modelo aos dados, embora, mantendo-se o afastamento, relativamente ao valor zero, dos resíduos acumulados para valores elevados das variáveis.

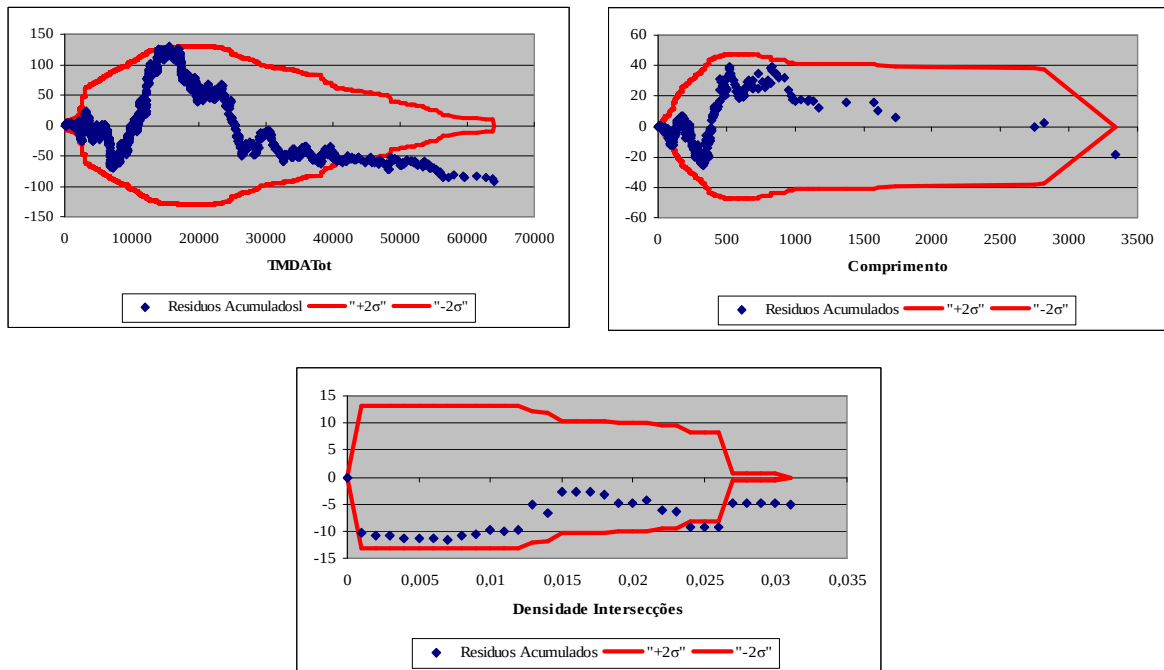


Figura 5.13. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, *Comp* e *DenInt* elaborados segundo o método CURE para o modelo com função Translog da equação (5.4)

⁵² $\chi^2 = 13,7$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

Concluiu-se assim que este modelo, cuja diferença em relação ao modelo base reside na inclusão da variável que cruza o volume de tráfego com o comprimento do arco, ajusta-se melhor aos dados, eventualmente por traduzir melhor a variação do efeito do tráfego e do comprimento no risco de acidente.

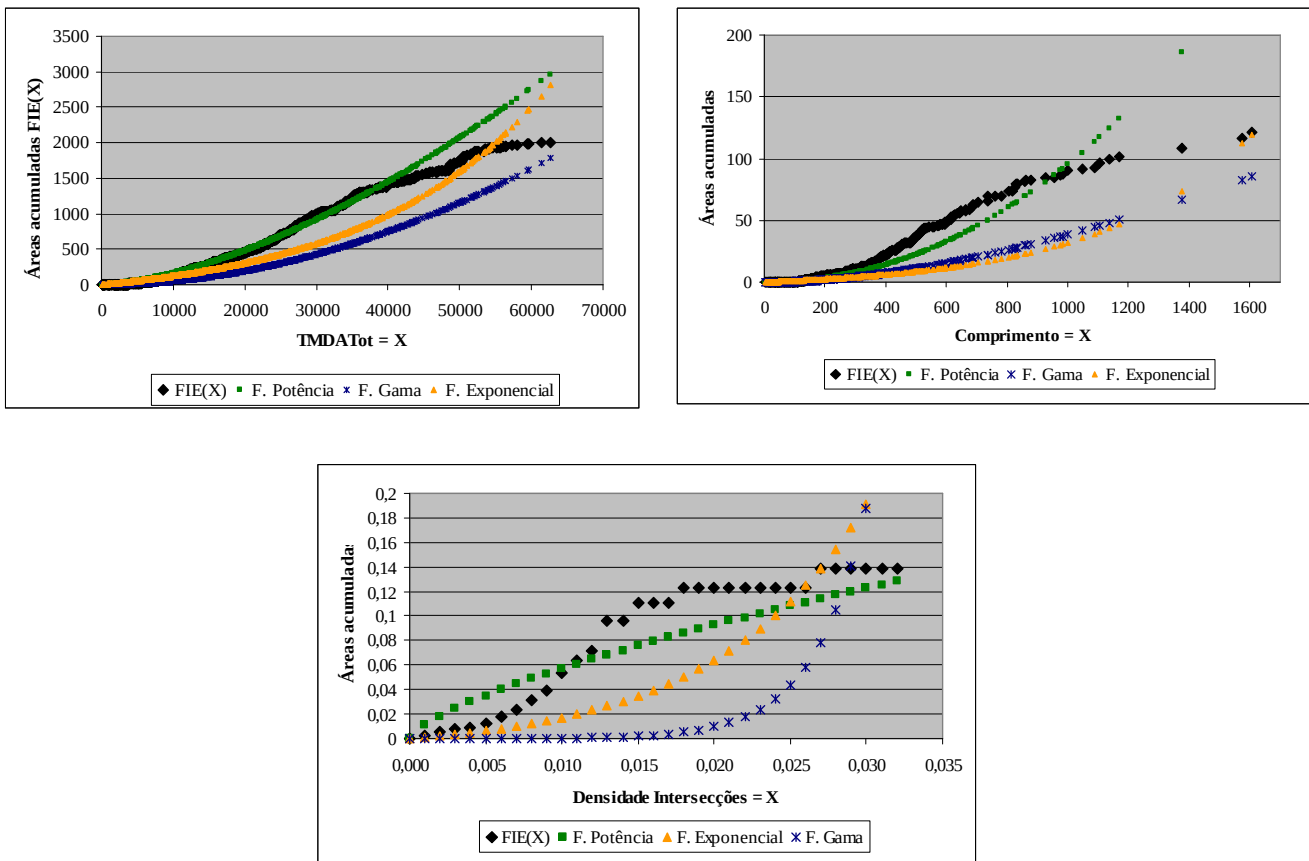
Assim, no caso da base de dados dos arcos para o total dos acidentes, e segundo as medidas de avaliação do ajuste, os modelos que melhor se aproximam aos dados registados são o modelo PA e o modelo Translog.

Contudo, tendo em conta que o objectivo deste trabalho é desenvolver uma ferramenta simples que determine o grau de sinistralidade representado por uma frequência de acidentes para um conjunto de soluções alternativas planeadas, o modelo que melhor se enquadra neste contexto é o modelo Translog. O modelo PA embora apresente o melhor valor de AIC, não é um modelo simples de aplicar e de interpretar, sendo necessária uma análise profunda para a sua aplicação.

Acidentes com Vítimas

O mesmo procedimento foi aplicado aos acidentes com vítimas. Para analisar a forma funcional a considerar para as variáveis independentes elaboraram-se os gráficos FIE(X) apresentados na Figura 5.14.

Pelos gráficos da Figura 5.14 pode concluir-se que a função que mais se aproxima da curva FIE(X) das variáveis $TMDA_{Tot}$ e $Comp$ é a função Potência. No caso da variável $DenInt$ a escolha da função não é tão evidente. Embora a curva da função Potência se aproxime mais da FIE(X), a forma da curva assemelha-se mais à da função Exponencial pelo menos até ao valor de 0,020 da variável. Optou-se, por isso, pela função Exponencial. Esta selecção da forma funcional das variáveis contínuas é coincidente com as consideradas para o total dos acidentes, pelo que neste caso o modelo tem a mesma estrutura da equação (5.2).



$$\text{Potência } f_1(x) = \alpha x^\beta \quad \text{Gama } f_2(x) = \alpha x^{\beta-1} e^{-\beta x} \quad \text{Exponencial } f_3(x) = e^{-\beta x}$$

Figura 5.14. - Gráfico comparativo da FIE(X) das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ com as funções Potência, Gama e Exponencial para os acidentes com vítimas

A modelação dos acidentes com vítimas considerando a distribuição BN para a variável de resposta gerou os resultados apresentados no Quadro 5.19 (no Anexo 2.XII apresentam-se os resultados completos).

Pelo Quadro 5.19 verifica-se que o valor do parâmetro de dispersão é baixo para um nível de confiança superior a 95%, confirmando assim a escolha da distribuição BN. Os valores dos parâmetros das variáveis contínuas foram estimados com um nível de confiança superior a 95%. Esses valores têm os mesmos sinais e uma ordem de grandeza próxima dos valores estimados para o total dos acidentes. Contudo, no caso das variáveis “dummies”, o nível de confiança abaixo dos 95% para os valores estimados dos parâmetros, não permite tirar conclusões acerca das mesmas. A exceção verifica-se no valor estimado do parâmetro da variável que identifica a ocupação do solo 5 ($\beta_{4,5}$). Daqui se pode concluir

que possivelmente os dados não são em número suficiente para se analisar a distribuição dos acidentes com vítimas na rede.

Quadro 5.19 - Resultados da estimação do modelo BN da equação (5.2) para os acidentes com vítimas

Acidentes com Vítimas			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-11,051	0,606	0,0000
β_1	0,510	0,055	0,0000
β_2	0,989	0,052	0,0000
$\ln(\gamma)$	-0,091	0,026	0,0004
β_3	42,816	8,892	0,0000
$\beta_{4,2}$	-0,129	0,095	0,1758
$\beta_{4,3}$	-0,291	0,222	0,1906
$\beta_{4,4}$	-0,188	0,184	0,3076
$\beta_{4,5}$	0,281	0,113	0,0125
$\beta_{5,2}$	0,015	0,080	0,8554
α	0,512	0,084	0,0000
AIC	1,71707		
R^2_{FT}	-		
R^2_{PE}	78,2%		

Para analisar a hipótese de incluir as “*dummies*” mesmo sem significância estatística, calculou-se o teste do rácio de verosimilhança⁵³ para comparar os modelos com “*dummies*” e sem “*dummies*”. O resultado do teste não rejeita a hipótese considerada de incluir as “*dummies*” indicando, eventualmente, que o modelo com as “*dummies*” é estatisticamente melhor.

Relativamente às medidas calculadas e apresentadas no Quadro 5.19, não foi possível obter um valor de R^2_{FT} aceitável dado que este excedeu o valor 1,00 (100%). Este resultado pode ser interpretado como um sobre-ajuste do modelo à base de dados, significando que o modelo explicou mais do que havia a explicar com a componente sistemática, resultado de uma eventual incorrecta interpretação do erro da regressão (Fridstrom, L., et al., 1995). Contudo, para a medida R^2_{PE} obteve-se o valor de 78,2%, indicando uma elevada percentagem da componente sistemática explicada pelo modelo.

⁵³ $\chi^2 = 3,87$ inferior a $\chi^2_{GL=4}(0,05) = 9,49$

A fim de analisar os resultados com base nos resíduos acumulados e em função das variáveis contínuas, elaboraram-se gráficos segundo o método CURE apresentados na Figura 5.15.

Pelo gráfico da Figura 5.15 relativo à variável $TMDA_{Tot}$ verifica-se que os resíduos acumulados se encontram, em geral, dentro dos limites das curvas $\pm 2\sigma$ e que o valor final é próximo de zero. O mesmo se pode concluir acerca dos resíduos acumulados da variável $Comp$, excepto o valor final que se distancia do valor nulo. O gráfico da variável $DenInt$ apresenta uma curva afastada do padrão desejável para valores da variável acima dos 0,021.

Embora o modelo aparente um bom ajuste à base de dados, os níveis de confiança abaixo dos 95% na estimação dos parâmetros das “dummies” não permitem tirar conclusões. Tendo-se verificado que ao retirar estas “dummies” os valores estimados e respectivos níveis de confiança das outras variáveis não se alteraram significativamente, optou-se por manter as mesmas.

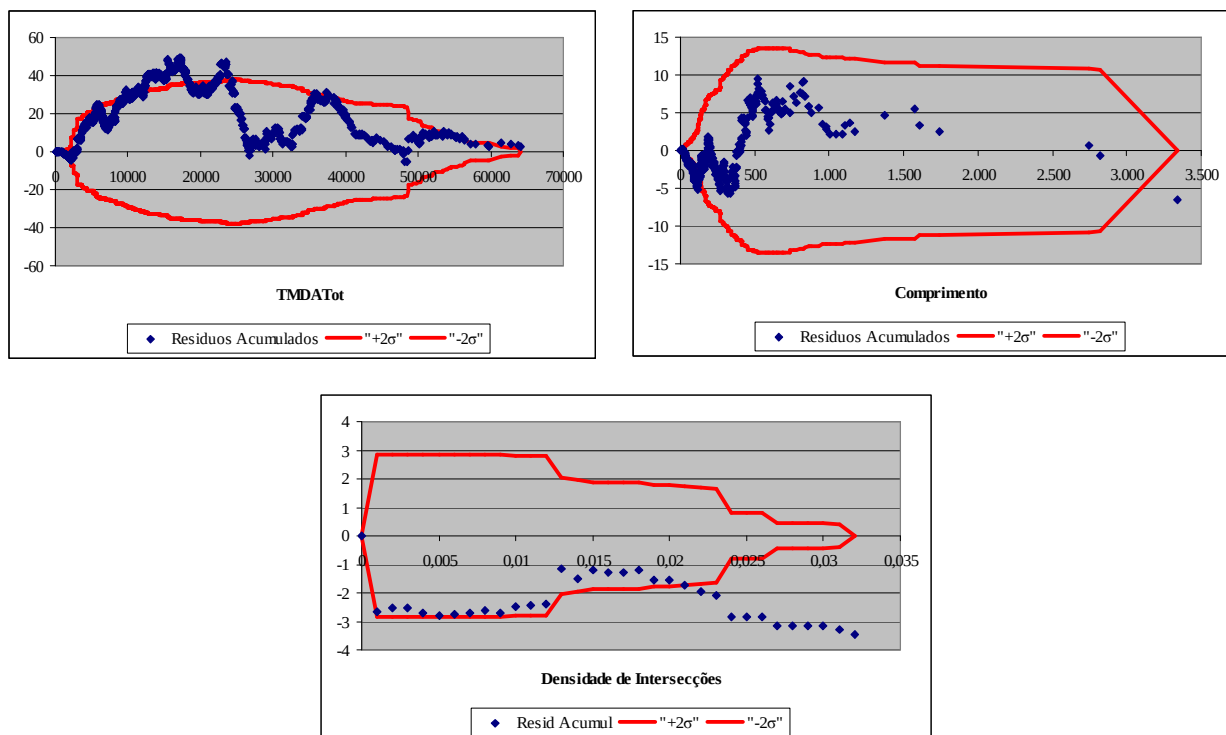


Figura 5.15. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Tot}$, $Comp$ e $DenInt$ elaborados segundo o método CURE para o modelo BN aplicado aos acidentes com vítimas

Na medida em que os números de acidentes com vítimas podem distribuir-se na rede com um grau de aleatoriedade maior do que o total dos acidentes e que factores, tais como, as características dos veículos e das vítimas podem influenciar a ocorrência dos acidentes, seria expectável uma maior dificuldade em obter relações fiáveis entre as variáveis independentes e a frequência de acidentes com vítimas. Em fase de planeamento é, assim, difícil de incluir, através da introdução de um maior número de variáveis independentes, factores relevantes para a ocorrência dos acidentes com vítimas. No caso dos acidentes com vítimas será necessário considerar um maior período de tempo para a base de dados e variáveis que quantifique, por exemplo, os volumes de peões dado que, em geral, uma percentagem elevada de acidentes com vítimas em meio urbano corresponde a atropelamentos. Assim, optou-se neste trabalho, por apenas desenvolver um modelo para o total dos acidentes.

5.3.2 Modelos para nós

Os modelos a desenvolver para nós baseiam-se na relação entre a variável de resposta e as variáveis independentes:

$$E[Y] = f(TMDA_{Princ}; TMDA_{Sec}; ANO; OcSolo_k; Classe_h) \quad (5.5)$$

Estas variáveis independentes são de dois tipos: variáveis contínuas e “*dummys*”. Estas últimas correspondem à $OcSolo_k$ e $Classe_g$. A “*dummy*” $OcSolo_k$ é composta por cinco categorias ($k=5$), tais como as consideradas para o modelo dos arcos. A “*dummy*” $Classe_g$ é composta por três categorias ($g=3$) definidas pela classificação hierárquica dos ramos do nó segundo o apresentado no Quadro 5.2 ($Classe_h = 1$ para classe do tipo h ; 0 para classe diferente de h , com $h=1$ a 3).

Os dados considerados correspondem ao número anual de acidentes totais e corporais distribuídos por 211 nós, o que equivale, para os cinco anos em estudo, a um total de 1055 observações.

No Quadro 5.20 são apresentados os valores das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ ao longo dos cinco anos. Do quadro é possível constatar que, tal como para os arcos, os valores mínimo e máximo do tráfego quer nos ramos principais quer nos ramos secundários são bastante diferentes e no caso do $TMDA_{Sec}$ o valor mínimo é zero. Este último verifica-se

em nós cujos ramos secundários são de sentido único da saída da intersecção, o que acontece em 34 nós.

Quadro 5.20 - Dados estatísticos das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$

Ano	$TMDA_{Princ}$ (min-máx)	$TMDA_{Sec}$ (min-máx)
2001	298-62 942	0-32 882
2002	285-60 081	0-32 882
2003	295-62 227	0-31 072
2004	288-60 797	0-26 547
2005	339-71 525	0-30 167

O Quadro 5.21 apresenta dados estatísticos diferenciados em acidentes com vítimas e total dos acidentes ao longo dos cinco anos e ainda a distribuição percentual dos acidentes segundo o tipo de ocupação do solo e a classe hierárquica dos nós.

Quadro 5.21 - Dados estatísticos dos acidentes ocorridos nos nós distribuídos por ano e gravidade e sua distribuição percentual pelas variáveis ocupação do solo e classificação hierárquica

Ano	Gravidade do Acidente	Nº Acidentes (min-máx-total)	OcSolo % acidentes					Classe Hier. % acidentes		
			1	2	3	4	5	A	B	C
2001	C/ Vítimas	0-4-108	48	24	0	5	23	3	80	17
	Total	0-11-481	52	21	3	6	18	5	75	20
2002	C/ Vítimas	0-4-73	45	26	0	6	23	1	81	18
	Total	0-13-432	51	23	4	4	18	4	73	23
2003	C/ Vítimas	0-5-89	51	26	3	1	19	3	73	24
	Total	0-10-366	52	21	4	4	19	5	72	23
2004	C/ Vítimas	0-4-95	53	24	2	5	16	3	72	25
	Total	0-10-424	52	20	4	4	20	6	70	24
2005	C/ Vítimas	0-4-82	49	28	0	2	21	5	61	34
	Total	0-9-326	50	20	4	4	22	5	64	31

Da distribuição percentual dos acidentes pelo tipo de ocupação do solo apresentada no Quadro 5.21, verifica-se que a classe de ocupação do solo 1 é a que se destaca com o maior número total de acidentes (52%) e a classe 3 com o menor número total dos acidentes (4%). Nos acidentes com vítimas verifica-se a mesma proporção de acidentes segundo a ocupação do solo – 49% de acidentes no tipo 1 e 1% no tipo 3. Nas classes de hierarquia viária consideradas verifica-se uma maior concentração de acidentes na classe B quer no total dos acidentes quer nos acidentes com vítimas (71% e 74%, respectivamente). A classe A apresenta uma concentração significativamente menor de acidentes (3% e 5% nos acidentes com vítimas e no total dos acidentes, respectivamente). Como se pode depreender dos valores percentuais referidos, a distribuição dos acidentes com vítimas pelas diferentes classes hierárquicas e de ocupação do solo é muito semelhante à do total dos acidentes.

Foi ainda elaborado o histograma apresentado na Figura 5.16, com o número de observações relativas à ocorrência de um número N de acidentes.

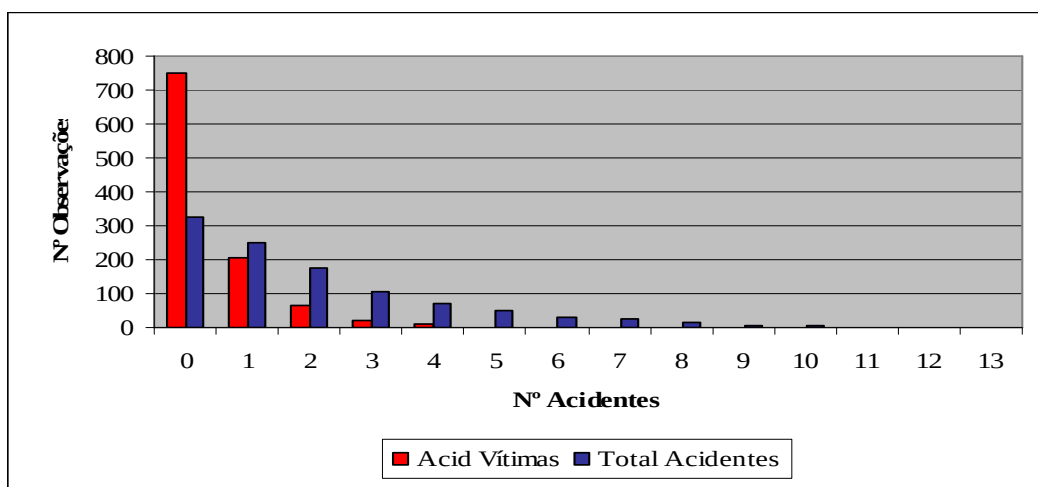


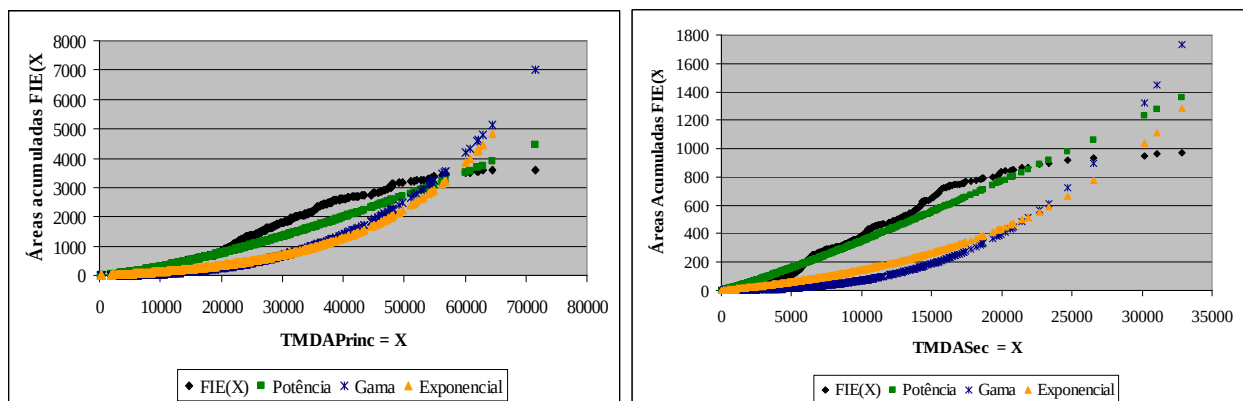
Figura 5.16. - Histograma do número de acidentes nos nós – total e com vítimas

Tal como no caso dos arcos, o maior número de observações concentra-se nos locais com zero acidentes, sendo mais preponderante nos acidentes com vítimas. No entanto, o número de observações em que se registou um acidente, no caso do total dos acidentes, não é muito inferior ao de zero acidentes, comparativamente aos acidentes com vítimas.

Total dos Acidentes

Para definir a estrutura do modelo a aplicar ao total dos acidentes utilizou-se o método ID para seleccionar a forma funcional que relaciona as variáveis contínuas com a variável de resposta. Para tal, determinou-se a FIE(X) para as variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ resultando nos gráficos da Figura 5.17, onde são também apresentadas as curvas das funções Potência, Gama e Exponencial.

Comparando as FIE(X) com as curvas das funções Potência, Gama e Exponencial a função que mais se aproxima da FIE(X) das duas variáveis é a função Potência. Testando as várias alternativas através da modelação da base de dados, verifica-se que as variáveis apresentam níveis de confiança bastante inferiores a 95% no caso da função Gama, não sendo de excluir por este critério a função Exponencial. Por ser também a função mais utilizada para o tráfego, optou-se por considerar a função Potência.



$$\text{Potência } f_1(x) = \alpha x^\beta \quad \text{Gama } f_2(x) = \alpha x^{\beta-1} e^{-\beta x} \quad \text{Exponencial } f_3(x) = e^{-\beta x}$$

Figura 5.17. - Gráfico comparativo da FIE(X) das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ com as funções Potência, Gama e Exponencial para o total dos acidentes

Importa lembrar a existência de nós (34) com o $TMDA_{Sec}$ cujo tráfego de entrada nos ramos é igual a zero. Este factor, por um lado, não permite a aplicação da função logaritmica a um valor zero e, por outro lado, torna o conjunto dos nós em análise heterogéneo. O facto de existir menos um fluxo de entrada na intersecção pode alterar o risco de ocorrência de acidentes. Assim, realizou-se uma análise específica a esta característica dos 34 nós através

da introdução de uma “dummy” que identifica os nós com $TMDA_{Sec}$ igual a zero. Pelos resultados desta aplicação apresentados em Anexo 2.XIII, verifica-se que o parâmetro estimado para esta “dummy” não tem significância estatística (nível de confiança superior a 95% pelo teste *t-student*). Ao comparar os valores estimados para as outras variáveis com o modelo sem a “dummy”, verifica-se que estes são muito idênticos. Assim, pode concluir-se que a presença de ramos com $TMDA_{Sec}$ zero na intersecção não altera significativamente o risco de acidente, considerando-se assim o valor zero nestes casos após transformação logaritmica.

Introduziu-se ainda no modelo a variável *ANO* para captar o efeito da tendência temporal. Com esta primeira análise definiu-se a estrutura a considerar para a modelação da base de dados:

$$E[Y] = \beta_0 TMDA_{Princ}^{\beta_1} TMDA_{Sec}^{\beta_2} \gamma^{ANO} e^{\beta_{3k} OcSolo_k + \beta_{4h} Classe_h} \quad (5.6)$$

Para estas variáveis analisaram-se as correlações entre as variáveis através da matriz de correlações, cujos valores são apresentados no Quadro 5.22.

Quadro 5.22 - Matriz de correlações entre as variáveis independentes da equação (5.6)

	$TMDA_{Princ}$	$TMDA_{Sec}$	<i>ANO</i>	<i>OcSolo2</i>	<i>OcSolo3</i>	<i>OcSolo4</i>	<i>OcSolo5</i>	<i>ClasseA</i>	<i>ClasseC</i>
$TMDA_{Princ}$	1,00	-	-	-	-	-	-	-	-
$TMDA_{Sec}$	0,06	1,00	-	-	-	-	-	-	-
<i>ANO</i>	0,11	0,02	1,00	-	-	-	-	-	-
<i>OcSolo2</i>	-0,01	-0,04	0,00	1,00	-	-	-	-	-
<i>OcSolo3</i>	-0,10	0,01	0,00	-0,09	1,00	-	-	-	-
<i>OcSolo4</i>	0,09	-0,09	0,00	-0,12	-0,04	1,00	-	-	-
<i>OcSolo5</i>	0,17	0,12	0,00	-0,21	-0,07	-0,10	1,00	-	-
<i>ClasseA</i>	0,04	0,01	0,00	0,12	0,20	0,03	-0,04	1,00	-
<i>ClasseC</i>	-0,31	0,01	0,00	-0,08	0,12	-0,17	0,08	-0,17	1,00

Dos valores apresentados no Quadro 5.22, verifica-se que estes valores são inferiores a 0,31, demonstrando assim não existir correlações significativas entre variáveis.

Modelo BN – modelo base

A modelação da base de dados, considerando um MLG aplicado à estrutura representada pela equação (5.6) com distribuição BN para a variável de resposta, resultou nos valores estimados apresentados no Quadro 5.23 (no Anexo 2.XIV apresentam-se todos os resultados da aplicação).

Da análise do Quadro 5.23, verifica-se que ao valor estimado do parâmetro de dispersão corresponde um nível de confiança superior a 95% que em conjunto com o teste de sobredispersão⁵⁴ sugere que a distribuição BN é apropriada.

Quadro 5.23 - Resultados da estimação do modelo BN da equação (5.6)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-3,140	0,500	0,0000
β_1	0,359	0,052	0,0000
β_2	0,095	0,011	0,0000
$\ln(\gamma)$	-0,102	0,023	0,0000
$\beta_{3,2}$	-0,038	0,084	0,6522
$\beta_{3,3}$	0,694	0,219	0,0015
$\beta_{3,4}$	-0,098	0,168	0,5606
$\beta_{3,5}$	0,236	0,102	0,0203
$\beta_{4,1}$	-0,388	0,136	0,0044
$\beta_{4,3}$	-0,420	0,082	0,0000
α	0,469	0,052	0,0000
<i>AIC</i>	3,56877		
R^2_{FT}	33,9%		
R^2_{PE}	37,5%		

As medidas R^2_{FT} e R^2_{PE} calculadas apresentam valores na ordem dos 35%. Este valor indica que a percentagem da componente sistemática explicada pelo modelo é reduzida. Tal como já verificado no trabalho de Greibe, P. (2003), os modelos para nós apresentam, em geral, uma menor percentagem explicada da componente sistemática. O mesmo autor concluiu que a modelação dos acidentes ocorridos nos arcos é menos complexa do que a dos acidentes ocorridos nos nós, justificando este facto com o padrão dos acidentes mais uniforme e uma exposição referente ao tráfego mais simples no caso dos arcos ou ainda

⁵⁴ valor superior a $\chi^2_{0,05GL=1}(3,84)$ rejeitando-se assim a hipótese da média igual à variância

devido à não utilização de variáveis independentes adequadas no caso dos modelos dos nós.

Das variáveis consideradas para caracterizar os nós em fase de planeamento nem todas apresentam um nível de confiança superior a 95%. É o caso das “*dummys*” que identificam a ocupação do solo 2 e 4, sugerindo que não existe uma variação significativa entre estes tipos de ocupação do solo e a ocupação do solo 1 (base de comparação).

Dos valores estimados com significância estatística pode concluir-se que os nós localizados nas zonas de ocupação do solo 3 e 5 produzem um aumento significativo nos acidentes relativamente à zona de ocupação do solo 1. A pseudo-elasticidade destas “*dummys*” é de 0,500 e 0,210 para a ocupação do solo 3 e 5, respectivamente indicando que mantendo-se todas as outras características representadas pelas outras variáveis, o tipo de ocupação do solo 3 e 5 representam um acréscimo nos acidentes de 50% e 21%, respectivamente.

Esta relação coincide com a obtida no modelo dos arcos. Porém, é de certa forma compreensível que o que caracteriza a ocupação do solo, desde o tipo e volume de habitação ao tipo e uso de equipamento existente, não seja tão preponderante nos nós. A influência destas características não se sobrepõe a outras, tais como, a geometria e funcionamento. No caso da ocupação do solo 3 (área empresarial) a influência da localização do nó manifesta-se provavelmente devido à presença de tráfego de veículos pesados e ao elevado número de conflitos na zona. Na zona identificada como ocupação do solo 5 é, eventualmente, a presença de um elevado tráfego pedonal que se manifesta no atravessamento nas intersecções que determina a maior sinistralidade.

Os parâmetros estimados das “*dummys*” que identificam a classe hierárquica dos nós apresentam um nível de confiança superior a 95%. Os valores estimados para as classes A e C apontam para a diminuição dos acidentes nos nós com estas classificações em relação aos nós da classe B (base de comparação por ter o maior número de nós). A pseudo-elasticidade destas classes é muita semelhante indicando um decréscimo nos acidentes de 47% e 52% nos nós das classes A e C, respectivamente. Como se pode observar pelo Quadro 5.2, estas classes identificam nós em que pelo menos um dos ramos tem uma classificação hierárquica muito diferente – via colectora e via de acesso local para a classe A e C, respectivamente. No caso da classe A, em que pelo menos um dos ramos é de uma via de acesso de uma via colectora (normalmente a Via de Cintura Interna do Porto), é possível que esta diminuição de acidentes resulte da existência de um conjunto de

características dos nós que diminuem os pontos de conflitos, nomeadamente a sinalização luminosa. Na classe C, em que pelo menos um dos ramos é de uma via de acesso local, o menor risco de acidentes pode estar relacionado com o menor número de conflitos e com uma geometria mais urbana dos nós com esta classificação.

O parâmetro estimado para a tendência temporal dos acidentes ao longo dos cinco anos em estudo indica um decréscimo de 40%⁵⁵ no total do período considerado. Para analisar se este decréscimo se distribui de uma forma idêntica em cada ano, estimou-se o mesmo modelo, mas considerando para a tendência temporal quatro “*dummys*” tendo como ano base 2001. Pelos resultados apresentados no Anexo 2.XV, verifica-se que de ano para ano há sempre um decréscimo relativamente ao ano 2001 e ao ano anterior, excepto no ano 2004 cujo decréscimo é inferior ao do ano 2003. De referir, no entanto, que os valores estimados para os anos de 2002 e 2004 têm um nível de confiança inferior a 95%.

Para analisar a existência de correlação temporal considerou-se o teste de razão de verosimilhança tal como aplicado aos arcos. Neste caso, a possibilidade de existir correlação é maior em relação aos arcos, na medida em que se obteve um nível de confiança de 95%⁵⁶ em afirmar que os parâmetros estimados no modelo para cada ano são diferentes dos parâmetros estimados no modelo com os cinco anos. Para avaliar se esta possível correlação temporal afecta os valores estimados dos parâmetros aplicou-se o mesmo modelo mas para os dados agregados nos cinco anos, isto é, considerou-se para cada local a soma dos acidentes dos cinco anos e a média do tráfego eliminando-se desta forma a correlação temporal. Pela análise dos resultados apresentados no Anexo 2.XVI é possível concluir que os valores dos parâmetros do modelo agregado são muito próximos do modelo desagregado verificando-se algumas diferenças em alguns níveis de confiança. Tal como para o modelo dos arcos considerou-se o modelo sem tratamento da correlação temporal.

Realizou-se uma análise para verificar se a elasticidade das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ corresponde a uma relação destas com os acidentes de acordo com o intuitivamente expectável. O valor da elasticidade é determinado através da equação (4.20) o que para funções logarítmicas, como é o caso das duas variáveis do tráfego, corresponde ao valor estimado do parâmetro. Verifica-se que para o tráfego principal e secundário o valor da

⁵⁵ Percentagem calculada a partir do valor de $(1-0,9^5)$

⁵⁶ Para $\chi^2 = -2[-1881,81 - (-403,28 - 389,79 - 347,52 - 375,61 - 337,80)] = 55,62$ e K graus de liberdade $(10 \cdot 5 - 10) = 40$ o valor α da probabilidade é 0,05

elasticidade indica que a um aumento de 10% destas variáveis corresponde um aumento de 3,6% e 1%, respectivamente no risco de acidentes. No modelo aplicado por Turner, S. and Nicholson, A. (1997), tendo como únicas variáveis o tráfego principal e secundário da intersecção, os valores estimados dos parâmetros para estas variáveis situam-se num intervalo de 0,2 a 0,9, ou seja, valores mais altos do que os apresentados no Quadro 5.23 para as mesmas variáveis.

Para analisar o ajuste do modelo aos dados de tráfego elaboraram-se, segundo o método CURE, os gráficos dos resíduos acumulados apresentados na Figura 5.18, ordenados segundo as duas variáveis contínuas – $TMDA_{Princ}$ e $TMDA_{Sec}$.

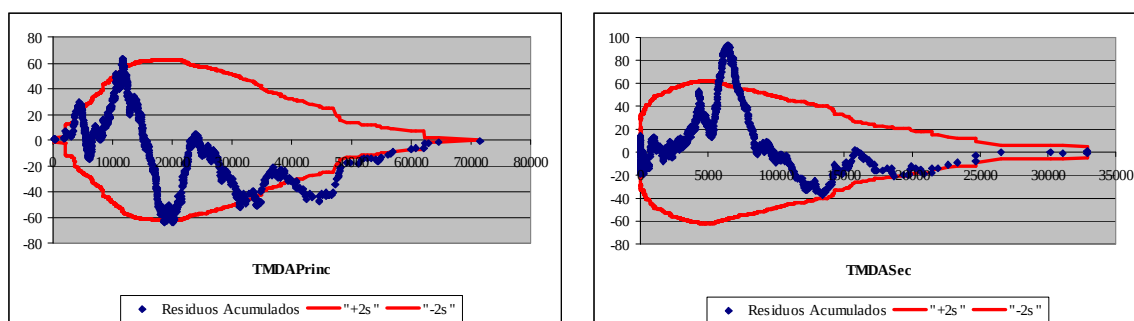


Figura 5.18. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ elaborados segundo o método CURE para o modelo BN

A curva dos resíduos acumulados ordenados segundo o $TMDA_{Princ}$ termina no valor nulo e situa-se dentro das bandas $\pm 2\sigma$, excepto no intervalo de valores do tráfego entre 40 000 e 50 000, aproximadamente. Também para o caso da variável $TMDA_{Sec}$ a curva dos resíduos acumulados apresenta um padrão indicativo de um bom ajuste do modelo aos dados com excepção do intervalo de valores do tráfego entre 6 000 e 7 000, aproximadamente.

Realizou-se uma análise à base de dados de forma a identificar a eventual presença de “outliers” e a sua influência na base de dados utilizando o método aplicado anteriormente aos arcos. Os valores “leverage” para cada observação estão representados na Figura 5.19, assim como o valor corresponde a $2K/n$.

Os pontos do gráfico acima da barra vermelha correspondem aos “outliers”. Nestas condições, foram identificados 130 “outliers” na base de dados.

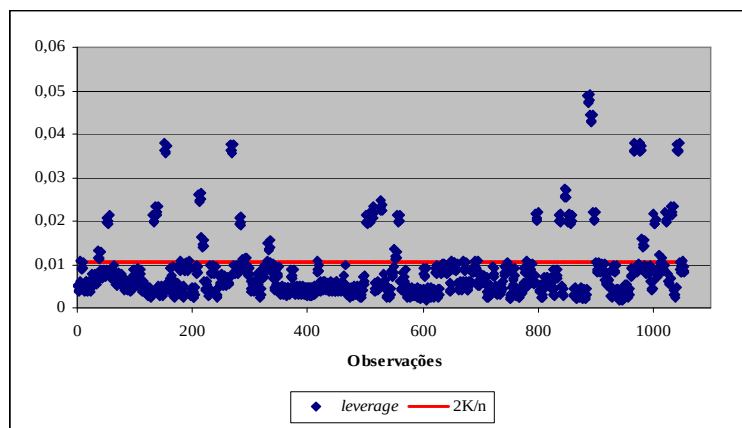


Figura 5.19. - Gráfico dos valores de “leverage” e de $2K/n$ referente à base de dados dos nós

Cruzou-se esta informação com os valores da função $Dfiti$. Dos “outliers” identificados, 23 apresentam valores da função $Dfiti$ maiores que 0,2 ou menores que -0,2. O gráfico da Figura 5.20 mostra os valores resultantes da função $Dfiti$.

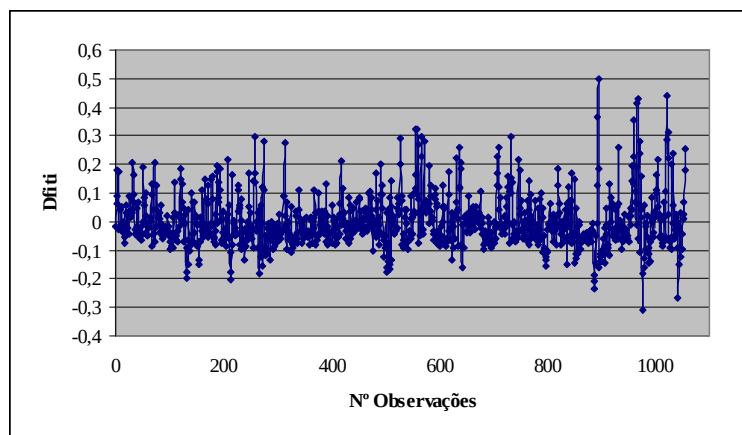


Figura 5.20. - Gráfico dos valores da função $Dfiti$ das observações registadas nos nós

Os resultados do modelo BN aplicado ao total dos acidentes sem estes 23 pontos demonstram que as diferenças nos valores estimados são muito reduzidas (ver Anexo 2.XVII), considerando-se, por isso, a seguir a mesma base de dados.

Assim, tal como se procedeu para a base de dados dos arcos apresentam-se a seguir alternativas ao modelo base, mantendo as mesmas formas funcionais da equação (5.6).

Modelo com heteroscedasticidade

Considerando a possibilidade da variância não ser constante entre observações, aplicou-se um modelo com heteroscedasticidade. Assim, após a análise de várias hipóteses quanto à variável que induz a heteroscedasticidade presente no modelo, seleccionou-se o modelo ponderado pela variável $TMDA_{Princ}$, cujos resultados são apresentados no Quadro 5.24 (resultados completos no Anexo 2.XVIII).

Quadro 5.24 - Resultados da estimação do modelo com heteroscedasticidade da equação (5.6)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-3,278	0,510	0,0000
β_1	0,373	0,053	0,0000
β_2	0,095	0,011	0,0000
$\ln(\gamma)$	-0,102	0,023	0,0000
$\beta_{3,2}$	-0,026	0,083	0,7549
$\beta_{3,3}$	0,701	0,220	0,0015
$\beta_{3,4}$	-0,092	0,165	0,5777
$\beta_{3,5}$	0,218	0,099	0,0286
$\beta_{4,1}$	-0,376	0,134	0,0050
$\beta_{4,3}$	-0,421	0,081	0,0000
α	0,457	0,051	0,0000
AIC	3,59769		
R^2_{FT}	34,0%		
R^2_{PE}	39,1%		

Pelos resultados constata-se que os valores estimados dos parâmetros são muito próximos dos valores estimados pelo modelo base, incluindo o valor do parâmetro de dispersão. Neste modelo as “dummies” ocupação do solo 2 e 4 não apresentam níveis de confiança superiores a 95%, tal como no modelo base.

Os valores das medidas R^2_{FT} e R^2_{PE} apresentados no quadro são praticamente iguais aos do modelo base. Para comparar este modelo com o modelo base calculou-se o teste do rácio

de verosimilhança⁵⁷. Segundo este teste e o valor de AIC, o modelo base é estatisticamente melhor.

Os gráficos dos resíduos acumulados (CURE) apresentados na Figura 5.21 mostram que não houve alterações significativas relativamente aos gráficos da Figura 5.18 do modelo base.

Verifica-se, assim, que este modelo não apresenta um melhor ajuste aos dados relativamente ao modelo base, não denotando, por isso, a existência de heteroscedasticidade relacionada com as variáveis do modelo.

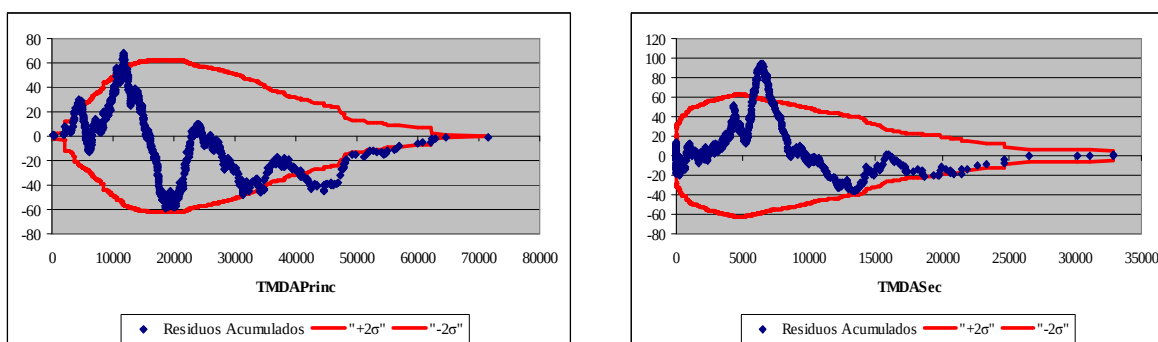


Figura 5.21. - Gráficos dos resíduos acumulados das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ elaborados segundo o método CURE para o modelo com heteroscedasticidade da equação (5.6)

Modelo com Parâmetros Aleatórios

Para analisar a influência da heterogeneidade dos dados entre locais e ao longo do tempo nas diferentes variáveis consideradas, aplicou-se um modelo com PA cuja variação do respectivo valor estimado é representada pela equação (5.3).

Após considerar várias hipóteses de configuração para este modelo, seleccionou-se o modelo com distribuição de Poisson para a variável de resposta e distribuição probabilística normal para a variação aleatória dos parâmetros e cujos resultados são apresentados no Quadro 5.25 (ver no Anexo 2.XIX os resultados completos). A selecção dos parâmetros com variação aleatória teve como base a significância estatística dos valores médios e dos factores de escala estimados para cada variável. O modelo

⁵⁷ $\chi^2 = 30,504$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

seleccionado apresenta cinco parâmetros aleatórios, três parâmetros fixos e dois cujos parâmetros estimados não são estatisticamente significantes.

Como se pode verificar pelo quadro os valores estimados para os parâmetros das “dummys” ocupação do solo 2 e 4 não apresentam níveis de confiança superiores a 95%, tal como nos modelos anteriores. Todos os outros parâmetros e factores de escala foram estimados com um nível de confiança superior a 95%.

A hipótese da variável de resposta ter uma distribuição BN foi preterida por apresentar um valor do parâmetro de dispersão muito elevado e sem significância estatística (36,026 com um valor de $P[Z > z] = 0,1496$).

Quadro 5.25 - Resultados da estimação do modelo com PA da equação (5.6)

Total dos Acidentes					
Parâmetro	Valor estimado	Desvio padrão	$P[Z > z]$	Factor Escala	$P[Z > z]$
$\ln(\beta_0)$	-1,295	0,365	0,0004	0,094	0,0000
β_1	0,148	0,038	0,0001	0,067	0,0000
β_2	0,090	0,009	0,0000	-	-
$\ln(\gamma)$	-0,106	0,015	0,0000	0,071	0,0000
$\beta_{3,2}$	-0,060	0,060	0,3201	-	-
$\beta_{3,3}$	0,558	0,145	0,0001	-	-
$\beta_{3,4}$	0,069	0,111	0,5338	-	-
$\beta_{3,5}$	0,348	0,064	0,0000	-	-
$\beta_{4,1}$	-0,896	0,162	0,0000	1,093	0,0000
$\beta_{4,3}$	-0,512	0,060	0,0000	0,164	0,0005
<i>AIC</i>	3,35521				

Os valores dos parâmetros estimados com significância estatística por este modelo não alteram a relação positiva ou negativa das variáveis com os acidentes identificada pelo modelo base. Inclusive os valores estimados pelos dois modelos para as mesmas variáveis são próximos, destacando-se, no entanto, uma maior diferença no valor da “dummy” da *ClasseA*, à qual no modelo com PA corresponde um menor risco de acidentes, isto é, um valor negativo mais afastado de zero. De notar, contudo, que o parâmetro desta “dummy” neste modelo é variável segundo uma distribuição Normal, com uma média de -0,896 e desvio padrão de 1,093 (79% da distribuição é inferior a 0).

O teste do rácio de verosimilhança⁵⁸ e o valor da medida AIC sugerem que o modelo com PA é estatisticamente melhor em relação ao modelo base. Assim, pode concluir-se que a maior parte das variáveis independentes da base de dados analisada se distribuem de uma forma heterogénea entre observações.

Modelo Zero-Inflacionado

No histograma da Figura 5.16 verifica-se uma concentração de observações no valor nulo de acidentes. Para analisar a hipótese do número de acidentes da base de dados ser gerado por um modelo binário, considerou-se a aplicação de modelos ZI. Assim, para o modelo com a estrutura (5.6) e para o total dos acidentes, considerou-se a regressão ZIP e ZIBN. Formularam-se várias hipóteses relativamente à selecção das variáveis independentes que afectam os dois estados e à distribuição probabilística que descreve a probabilidade de cada estado.

Considerou-se o teste Vuong para avaliar os resultados das diferentes aplicações rejeitando-se os modelos cujo valor V do teste é inferior a -1,96.

Das várias hipóteses testadas, o modelo seleccionado corresponde à distribuição de Poisson para a variável de resposta e distribuição probit para a probabilidade de ocorrer determinado estado. Considera ainda as mesmas variáveis independentes a condicionar os dois estados de geração de acidentes. O valor do teste Vuong para este modelo é de 4,7965. O modelo ZIP seleccionado determinou um número de observações previstas de zero acidentes igual a 308,8, sendo que o número de registos observados deste mesmo valor nulo de ocorrências é de 324, tal como se pode constar dos resultados apresentados no Anexo 2.XX. Pela aplicação do modelo Poisson o número de observações previstas para essas ocorrências é de 191,3. Como se pode depreender, o modelo ZIP prevê um número de observações de zero ocorrências mais próximo do número de valores observados do que o modelo Poisson, inflacionando, assim, o número de zeros em relação a este.

⁵⁸ $\chi^2 = 233,31$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

Quadro 5.26 - Resultados da estimação do modelo ZIP da equação (5.6)

Total dos acidentes			
Parâmetro	Valor estimado	Desvio padrão	P[Z>z]
$\ln(\beta_0)$	-1,894	0,277	0,0000
β_1	0,264	0,029	0,0000
β_2	0,067	0,007	0,0000
$\ln(\gamma)$	-0,075	0,012	0,0000
$\beta_{3,2}$	-0,001	0,045	0,9942
$\beta_{3,3}$	0,506	0,107	0,0000
$\beta_{3,4}$	-0,092	0,099	0,3515
$\beta_{3,5}$	0,156	0,050	0,0018
$\beta_{4,1}$	-0,257	0,076	0,0007
$\beta_{4,3}$	-0,310	0,046	0,0000
τ	-1,031	0,078	0,0000
R^2_{FT}	18,7%		

Dos resultados apresentados no Quadro 5.26 (ver no Anexo 2.XX os resultados completos) verifica-se que o valor estimado do parâmetro τ apresenta um nível de confiança superior a 95%, assim como todos os parâmetros estimados com excepção, uma vez mais, para as “dummies” *OcSolo2* e *OcSolo4*.

Se compararmos os valores dos parâmetros estimados pelo modelo ZIP com os valores do modelo base verifica-se que não há alteração dos sinais dos parâmetros, mantendo-se, por isso, o mesmo tipo de relação das variáveis com os acidentes.

O teste do rácio de verosimilhança⁵⁹ indica que o modelo ZIP, ao contrário do verificado nos modelos da base de dados dos arcos, é estatisticamente melhor do que o modelo base. No entanto, o valor da medida R^2_{FT} é inferior ao valor da mesma medida calculado para o modelo base, indicando assim uma menor percentagem da componente sistemática explicada pelo modelo ZIP relativamente ao modelo base.

Análise comparativa dos modelos

Para comparar os vários modelos aplicados à base de dados dos nós, apresentam-se no Quadro 5.27 os resultados obtidos.

⁵⁹ $\chi^2 = 96,228$ superior a $\chi^2_{GL=1}(0,01) = 6,63$

Quadro 5.27 - Valores estimados e níveis de confiança para os diferentes modelos da equação (5.6) aplicados à base de dados dos nós

Parâmetro	Total dos Acidentes							
	Modelo base		Heteroc ($TMDA_{Princ}$)		Modelo PA		ZIP	
	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]
$\ln\beta_0$	-3,140	0,0000	-3,278	0,0000	-1,295*	0,0004	-1,894	0,0000
β_1	0,359	0,0000	0,373	0,0000	0,148*	0,0001	0,264	0,0000
β_2	0,095	0,0000	0,095	0,0000	0,090	0,0000	0,067	0,0000
$\ln(\gamma)$	-0,102	0,0000	-0,102	0,0000	-0,106*	0,0000	-0,075	0,0000
$\beta_{3,2}$	-0,038	0,6522	-0,026	0,7549	-0,060	0,3201	-0,001	0,9942
$\beta_{3,3}$	0,694	0,0015	0,701	0,0015	0,558	0,0001	0,506	0,0000
$\beta_{3,4}$	-0,098	0,5606	-0,092	0,5777	0,069	0,5338	-0,092	0,3515
$\beta_{3,5}$	0,236	0,0203	0,218	0,0286	0,348	0,0000	0,156	0,0018
$\beta_{4,1}$	-0,388	0,0044	-0,376	0,0050	-0,896*	0,0000	-0,257	0,0007
$\beta_{4,3}$	-0,420	0,0000	-0,421	0,0000	-0,512*	0,0000	-0,310	0,0000
α	0,469	0,0000	0,457	0,0000	-	-	-	-
AIC	3,56877		3,59769		3,35521		-	
R^2_{FT}	33,9%		34,0%		-		18,7%	
R^2_{PE}	37,5%		39,1%		-		-	
função logaritmica de verosimilhança	-1 871,528		-1 886,780		-1 754,873		-1 919,647	

*valores médios

Os valores apresentados no quadro para os parâmetros estimados com variação aleatória pelo modelo PA correspondem ao valor médio.

Pelo Quadro 5.27 verifica-se que os valores estimados para os parâmetros das variáveis mantêm os sinais e a mesma ordem de grandeza, e que no caso das “dummies” *OcSolo2* e *OcSolo4*, estes não apresentam um nível de confiança superior a 95% em qualquer um dos modelos. Para estas variáveis não é possível concluir acerca da variação do risco de acidentes nestes nós relativamente aos situados em zonas do tipo de ocupação do solo 1. Os valores estimados pelos quatros modelos para as outras “dummies” são muito próximos e com níveis de confiança superiores a 95%, apenas com uma alteração mais significativa na “dummy” *ClasseA* do modelo com PA, sendo que este valor corresponde ao valor médio da distribuição normal. Esta estabilidade dos valores entre os modelos mostra a coerência das variáveis consideradas nos modelos para caracterizar o risco de acidentes ocorridos nos nós.

Tal como verificado na base de dados dos arcos, a distribuição de Poisson é a distribuição considerada para a variável de resposta no caso dos modelos PA e ZIP, podendo este facto

indiciar que estes modelos tratam a dispersão dos dados que o modelo BN capta como sobredispersão.

Segundo os valores da medida AIC, o modelo estatisticamente mais ajustado é o modelo PA, não sendo esta medida aplicada ao modelo ZIP. O teste de razão de verosimilhança calculado para comparar o modelo ZIP com o modelo base indica que o modelo ZIP é estatisticamente melhor embora a percentagem da componente sistemática explicada por este modelo é menor do que a percentagem calculada para o modelo base.

Tal como na modelação dos acidentes ocorridos nos arcos, também a aplicação destes modelos variantes ao modelo base dos nós teve como objectivo melhorar o ajuste à base de dados, mantendo a estrutura inicial representada pela equação (5.6). O modelo com PA e o modelo ZIP apresentam melhores resultados segundo as medidas de qualidade de ajuste. A aplicação do modelo PA tem recentemente vindo a ser aplicada e, em geral, com bons resultados. O facto de se verificarem níveis de confiança superiores a 95% na estimação dos factores de escala demonstra que a variação do efeito dessas variáveis não deve ser desprezada. Contudo, a aplicação e a interpretação dos resultados deste modelo é mais complexa. O mesmo se pode afirmar para o modelo ZIP, sendo, também, a sua aplicação mais controversa tal como referido anteriormente em 4.1.

O facto de estes dois modelos de aplicação mais complexa melhorarem o ajuste do modelo, não é garantia que o resultado da previsão seja o mais correcto. Se acrescentarmos o facto de que, a previsão para um determinado ano com base no modelo PA se baseia no valor médio dos parâmetros aleatórios induzindo um maior intervalo de confiança no resultado final. No entanto, para melhor aferir esta conclusão será necessário aplicar os modelos referidos a um conjunto de soluções alternativas para um determinado ano prognóstico e comparar com os valores registados para esse mesmo ano, e assim seleccionar o modelo que melhor se adapta aos objectivos pretendidos.

Embora nos gráficos CURE do modelo base não se tenha verificado, em intervalos de valores das variáveis analisadas, um afastamento dos resíduos acumulados para além dos limites aceitáveis, aplicou-se, à semelhança do realizado para a base de dados dos arcos, o modelo com função Translog.

Modelo com função Translog

No caso do modelo dos nós, a função Translog implica apenas o cruzamento das variáveis $TMDA_{Princ}$ e $TMDA_{Sec}$ tal como exposto na equação abaixo:

$$\ln E[Y] = \ln(\beta_0) + \beta_1 \ln(TMDA_{Princ}) + \beta_2 \ln(TMDA_{Sec}) + \beta_3 [\ln(TMDA_{Princ})]^2 + \beta_4 [\ln(TMDA_{Princ}) \ln(TMDA_{Sec})] + \beta_5 [\ln(TMDA_{Sec})]^2 + (\ln \gamma) ANO + \beta_{6,k} OcSolo_k + \beta_{7,h} Classe_h \quad (5.7)$$

em que k identifica a ocupação do solo com valores de 1 a 5 e h a classe hierárquica do nó com valores de 1 a 3.

Contudo, a matriz de correlações, apresentada no Quadro 5.28 revela a existência de correlações entre essas variáveis com valores acima de 0,90. Por esse facto, não se aplicou o modelo Translog.

Quadro 5.28 - Matriz de correlação entre as variáveis independentes cruzadas da equação (5.7)

	$TMDA_{Princ}$	$TMDA_{Sec}$	$TMDA_{Princ}$ $*TMDA_{Princ}$	$TMDA_{Sec}$ $*TMDA_{Sec}$	$TMDA_{Princ}$ $*TMDA_{Sec}$
$TMDA_{Princ}$	1,00	-	-	-	-
$TMDA_{Sec}$	0,59	1,00	-	-	-
$TMDA_{Princ} * TMDA_{Princ}$	0,99	0,04	1,00	-	-
$TMDA_{Sec} * TMDA_{Sec}$	0,13	0,97	0,12	1,00	-
$TMDA_{Princ} * TMDA_{Sec}$	0,18	0,99	0,17	0,97	1,00

Pode-se assim concluir que não é possível assumir uma relação entre as variáveis cruzadas que influenciem o risco de ocorrer acidentes nos nós.

5.4 Modelos de resposta qualitativa

Nos modelos de resposta qualitativa a variável de resposta é discreta e o valor por ela assumido corresponde a um código que identifica uma categoria à qual é associado um

determinado significado. No caso do estudo da sinistralidade, a maior parte dos modelos já desenvolvidos considera para este código diferentes graus de gravidade das vítimas (por exemplo: morto, ferido grave e ferido leve). Contudo, tendo em conta que o número de acidentes com vítimas é menor do que o número total dos acidentes e que à gravidade das vítimas estão associados factores para além da infra-estrutura difíceis de quantificar em fase de planeamento, optou-se neste trabalho por considerar apenas a modelação do total dos acidentes, estabelecendo, no entanto, dois grupos de categorias num dos quais se diferencia a ocorrência de acidentes com vítimas. Assim, no primeiro grupo, considerou-se para os códigos da variável de resposta categorias de graus de sinistralidade relacionadas com intervalos de frequência de acidentes num ano. Pretendeu-se ainda que estes graus de sinistralidade correspondessem a três categorias de avaliação dos cenários – bom, médio e mau, para desta forma se obter categorias que se distingam entre si não só pela gravidade em termos de sinistralidade a que correspondam, mas também para que se possa associar efeitos diferentes das variáveis independentes na variável de resposta. Nesse sentido, os intervalos de frequência foram definidos considerando a distribuição do número de acidentes pelas observações e após uma análise dos resultados a várias hipóteses. Assim, estabeleceu-se que na primeira categoria o intervalo de frequências varia entre zero e dois acidentes. Esta categoria corresponde ao melhor cenário pois considera a não ocorrência de acidentes ou a eventual existência de um ou dois acidentes devido, eventualmente, a motivos fortuitos e como tal difíceis de prever e de mitigar. Na segunda categoria, o intervalo inicia-se no valor seguinte ao da última categoria, ou seja, três acidentes sendo que o limite superior é dado pelo número de acidentes ao qual corresponde a existência de 10% de observações com valores acima deste número (percentil 90). A terceira categoria corresponde aos valores acima do último valor da categoria anterior. Dado que a distribuição dos acidentes é diferente nos arcos e nos nós estes intervalos serão também distintos.

O segundo grupo de categorias, baseado igualmente no total dos acidentes mas com o objectivo de evidenciar a importância da ocorrência de vítimas, considerou-se três categorias, neste caso iguais para arcos e nós: zero acidentes, acidentes sem vítimas e acidentes com vítimas.

Dos vários modelos de resposta qualitativa existentes e referidos em 5.2, optou-se por aplicar o modelo ordenado e o modelo multinomial. Excluiu-se o modelo binário por

considerar apenas duas categorias e o modelo hierárquico por ser estruturalmente complexo para o objectivo pretendido.

No modelo MO as categorias da variável de resposta são ordinais definindo assim uma ordenação dos graus de sinistralidade. Esta característica da variável de resposta dos MO pressupõe uma dependência entre as categorias. A probabilidade de ocorrer determinada categoria é calculada por comparação com outra categoria (geralmente a primeira do ranking). Por este facto, o modelo tem uma estrutura simples em que, através de uma regressão latente, se estabelecem as relações entre a variável de resposta e as variáveis independentes e em que os valores dos parâmetros estimados são os mesmos em qualquer categoria. Contudo, esta característica dos modelos ordenados pode ser uma simplificação da realidade. Como alternativa ao MO surge o MM, que ultrapassa esta simplificação ao considerar para cada categoria uma função “utilitária” que estabelece as relações entre a variável de resposta e as variáveis independentes. A estrutura do MM pressupõe uma independência entre as categorias, ou seja, a ocorrência de uma categoria não é influenciada por outra. Consequentemente, as categorias não são susceptíveis de ordenação.

Nos dois modelos acima referidos, o cálculo da probabilidade de ocorrer determinada categoria considera uma distribuição probabilística contínua e simétrica para a componente aleatória, normalmente as distribuições normal ou logística. No primeiro caso o modelo designa-se de probit e no segundo caso de logit, sendo a estimação realizada pelo método da máxima verosimilhança em ambos os casos.

Os valores estimados para os parâmetros das variáveis consideradas são analisados quanto à sua significância estatística com base nos níveis de confiança através do teste *t-student* (considera-se o valor estimado com significância estatística para níveis de confiança superiores a 95%). Contudo, é importante ter em conta que os parâmetros estimados não representam directamente a magnitude do efeito na probabilidade das variáveis independentes consideradas no modelo. Consideraram-se, assim, os efeitos marginais calculados com base no valor médio das variáveis independentes, a partir dos quais se analisou o efeito das variações dos valores das variáveis independentes na probabilidade de ocorrer cada categoria. Através destes valores realizou-se ainda uma análise comparativa entre os dois modelos aplicados.

Para avaliar e comparar os resultados das aplicações com base nestes dois tipos de modelos de resposta qualitativa considerou-se a medida designada de McFadden Pseudo R^2_{adj} ajustado. Esta medida enquadra-se nas medidas designadas de Pseudo- R^2 que foram desenvolvidas para modelos cujo método de estimação é o da máxima verosimilhança. A designação “pseudo” resulta da semelhança com a medida R^2 em termos de escala, ou seja, com valores a variar entre 0 e 1, e em que quanto maior o valor maior a qualidade do ajuste. A medida McFadden Pseudo R^2_{adj} também designada de índice de rácio de verosimilhança ajustado, baseia-se no rácio entre o valor da função logarítmica de verosimilhança do modelo a analisar (modelo com as variáveis independentes consideradas) ajustado ao número de parâmetros do modelo e o valor da função logarítmica de verosimilhança do modelo reduzido (modelo só com o termo da constante), tal como expresso pela equação abaixo:

$$R^2_{adj} = 1 - \frac{\ell(\beta) - k}{\ell(0)} \quad (5.8)$$

em que $\ell(\beta)$ e $\ell(0)$ correspondem, respectivamente, à função logarítmica de verosimilhança do modelo completo e função logarítmica de verosimilhança do modelo reduzido (modelo só com o termo da constante), e k é o número de parâmetros do modelo. O ajuste ao número de parâmetros tem como objectivo penalizar o modelo que inclua um número muito elevado de parâmetros. Como o MM implica a estimação de um conjunto elevado de parâmetros é importante considerar uma medida que tenha em conta este facto. Para além desta medida considerou-se ainda a medida AIC já anteriormente definida no Quadro 5.7.

Para testar o desempenho estatístico dos dois modelos MO e MM, teve-se em conta ainda o teste do rácio de verosimilhança ajustado⁶⁰ de Ben-Akiva *and* Lerman's (1985) aplicado por Bhat, C. R. *and* Pulugurta, V. (1997). Este teste determina se os valores da medida McFadden Pseudo R^2_{adj} são significativamente diferentes pelo que a hipótese nula a considerar é de os seus valores serem iguais. Neste caso se τ corresponder à diferença entre os dois valores então a probabilidade de τ ter ocorrido com probabilidade desprezável é menor do que $\Phi\left\{-\left[-2\tau\ell(0) + (K_2 - K_1)\right]^{0.5}\right\}$ (em que $\Phi\{.\}$ é o acumulado da função

⁶⁰ *adjusted likelihood ratio test*, em inglês

distribuição normal normalizada, isto é, com média 0 e desvio padrão 1; em que K_1 e K_2 são o número de parâmetros estimados pelos dois modelos). Um valor muito pequeno da probabilidade desta ocorrência indica que a diferença τ é estatisticamente significativa e que o modelo com o valor mais elevado de McFadden Pseudo R^2_{adj} é preferível.

Para comparar estes dois modelos relativamente à qualidade das previsões considerou-se ainda a análise da matriz de previsões apresentada nos resultados da aplicação. Nesta matriz comparam-se os valores observados com os valores previstos pelo modelo para cada categoria.

Descrevem-se a seguir os modelos de resposta qualitativa desenvolvidos e aplicados neste trabalho, divididos em dois tipos de aplicação – modelos para arcos e modelos para nós.

5.4.1 Modelos para arcos

Os modelos de resposta qualitativa a aplicar aos arcos consideram as formas funcionais para as variáveis independentes anteriormente identificadas e representadas na equação (5.2). Para a variável de resposta definiram-se três categorias às quais correspondem intervalos de frequência anuais de acidentes. Como se referiu anteriormente a primeira categoria considera a não ocorrência de acidentes ou a ocorrência de um número reduzido (1 ou 2) devido, eventualmente, a condições pontuais. Para definir o intervalo da segunda categoria definiu-se o limite superior como sendo o número de acidentes correspondente ao percentil 90 das observações registadas. No caso da base de dados dos arcos esse valor corresponde a 8 acidentes. Apresentam-se no Quadro 5.29 os intervalos de cada categoria e o respectivo número de observações associado a cada uma delas.

Quadro 5.29 - Classe de categorias da variável de resposta com base na frequência de acidentes - arcos

Categoria	Nº Acidentes	Nº Observações
0	0;1;2	1260
1	3;4;5;6;7;8	557
2	≥ 9	163

Como seria de esperar o número de observações é mais pequeno nas categorias de maior índice.

Apresentam-se a seguir os resultados das aplicações para os dois modelos – MO e MM.

Modelo Ordenado

O modelo ordenado é estruturado a partir da regressão latente em que as variáveis independentes e respectivas formas funcionais são as mesmas da equação (5.2):

$$y^* = x\beta + \varepsilon \quad (5.9)$$

em que y^* é não observado, $x\beta$ é designado de função índice e ε é o termo de erro aleatório. Esta componente não observada y^* traduz assim o impacto das variáveis no local. O observado é representado por y e está relacionado com y^* , ou seja, com os factores de impacto, da seguinte forma:

$$\begin{aligned} y &= 0 \text{ se } y^* \leq 0 \\ y &= 1 \text{ se } 0 < y^* \leq \mu \\ y &= 2 \text{ se } y^* > \mu \end{aligned} \quad (5.10)$$

em que μ é um parâmetro desconhecido a estimar juntamente com os parâmetros β .

Assim, a cada categoria corresponde uma probabilidade calculada considerando uma distribuição probabilística contínua e simétrica, geralmente a distribuição probit ou logit.

No caso da distribuição probit o cálculo das probabilidades é representado por:

$$\begin{aligned} P(y=0|x) &= \Phi(-x\beta), \\ P(y=1|x) &= \Phi(\mu - x\beta) - \Phi(-x\beta), \\ P(y=2|x) &= 1 - \Phi(\mu - x\beta) \end{aligned} \quad (5.11)$$

em que $\Phi(.)$ representa a função de distribuição normal acumulada.

No caso da distribuição logit representa-se pelo símbolo - $\Lambda(.)$ a função de distribuição logística acumulada. A diferença entre as duas distribuições reside essencialmente nas

“caudas”, isto é, a curva da distribuição logística decresce e cresce mais rapidamente do que a curva da distribuição normal. A escolha entre uma das duas distribuições não tem uma fundamentação teórica podendo apenas existir razões práticas que justifique a opção por uma delas. Neste trabalho consideraram-se as duas hipóteses e com base nas medidas de qualidade de ajuste referidas anteriormente seleccionou-se uma delas. Como seria expectável, os resultados obtidos por estes dois modelos não diferem muito. Os valores das medidas AIC e McFadden Pseudo R^2_{adj} para os dois modelos são praticamente iguais (ver Anexo 2.XXI), mas apontando para o modelo com distribuição probit (MOP) como o modelo com melhor qualidade do ajuste.

Apresentam-se no Quadro 5.30 os valores estimados dos parâmetros e respectivos níveis de confiança segundo o teste *t-student*, os efeitos marginais das variáveis para cada categoria e os valores das medidas da qualidade de ajuste consideradas.

Como se pode constatar do Quadro 5.30 todos os parâmetros foram estimados com um nível de confiança superior a 95%.

Quadro 5.30 - Resultados da estimação do modelo ordenado probit – arcos

Parâmetro	Valor estimado	Desvio padrão	P[Z>z]	Efeitos marginais		
				Categoria		
				0	1	2
$\ln(\beta_0)$	-9,984	0,505	0,0000	-	-	-
β_1	0,393	0,045	0,0000	-0,1292	0,1163	0,0129
β_2	1,107	0,047	0,0000	-0,3641	0,3278	0,0364
$\ln(\gamma)$	-0,056	0,022	0,0119	0,0184	-0,0166	-0,0018
β_3	37,562	6,799	0,0000	-12,3608	11,1263	1,2346
$\beta_{4,2}$	-0,287	0,082	0,0005	0,0893	-0,0813	-0,0080
$\beta_{4,3}$	0,614	0,175	0,0005	-0,2285	0,1900	0,0384
$\beta_{4,4}$	-0,494	0,132	0,0002	0,1381	-0,1279	-0,0102
$\beta_{4,5}$	0,299	0,093	0,0012	-0,1042	0,0917	0,0126
$\beta_{5,2}$	-0,186	0,069	0,0076	0,0610	-0,0548	-0,0061
μ	1,614	0,062	0,0000	-	-	-
Mc Fadden R^2_{adj}				0,3059		
AIC				1,18091		
função logaritmica de verosimilhança				-1158,103		

Os valores dos parâmetros estimados pelo MO indicam o efeito das variáveis independentes na propensão latente de ocorrer determinada categoria de sinistralidade num local. Os sinais dos valores estimados dos parâmetros apresentados no quadro são coerentes com os resultados dos modelos quantitativos. Contudo, os parâmetros dos modelos probabilísticos não lineares, como é o caso do MO, não traduzem a magnitude e o sentido do efeito das variáveis independentes nas respectivas categorias de sinistralidade. Por esse facto, analisaram-se os efeitos marginais das variáveis em cada categoria e que estão apresentados no Quadro 5.30. Os efeitos marginais são determinados considerando a equação (4.36) para os valores médios das variáveis. Desta equação verifica-se que os efeitos marginais correspondem à multiplicação do coeficiente da variável a analisar pela função de densidade da distribuição normal (no caso do modelo probit) nos pontos que limitam as categorias.

A Figura 5.22 ilustra o cálculo das probabilidades do MO. A probabilidade associada a cada categoria é calculada pela equação (5.11) e corresponde à área do gráfico da Figura 5.22 delimitada pelos valores da regressão latente e dos parâmetros μ .

Nesta figura é também perceptível o efeito nas probabilidades associado à alteração do valor da variável x (curva a tracejado). Assim, aumentando o valor de uma das variáveis x , mantendo-se β e μ constantes, é equivalente a deslocar a distribuição para a direita, tal como mostra a curva a tracejado da figura.

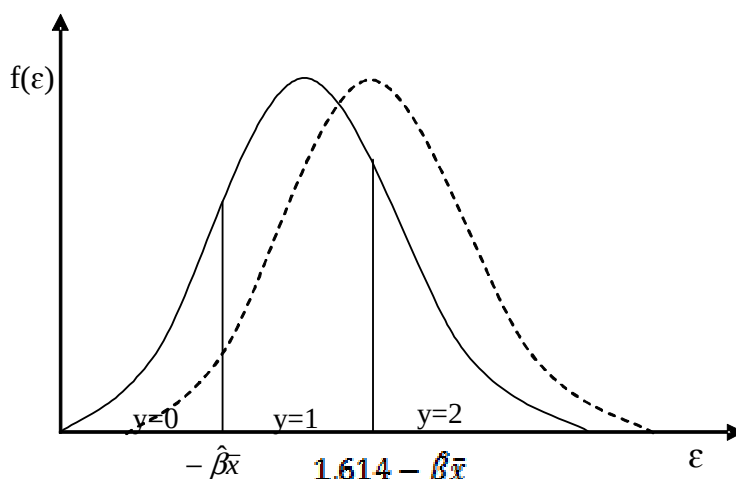


Figura 5.22. - Probabilidades do modelo ordenado estimado para os arcos e o efeito da variação do valor de x nas probabilidades de previsão (a tracejado)

Da equação (4.36) é perceptível, tal como seria necessariamente de esperar, o facto de na primeira categoria o sinal do valor do efeito marginal ser contrário ao sinal do parâmetro β da variável em análise, assim como na última categoria o valor do efeito marginal ter sempre o mesmo sinal que o referido parâmetro. No caso das categorias intermédias o sinal resultante para o efeito marginal é flutuante.

De notar que a soma dos efeitos marginais das três categorias em cada variável é igual a zero. Tal decorre do facto da soma das probabilidades ser necessariamente igual a 1.

No caso das “*dummys*” o efeito marginal deve ser analisado comparando os dois valores possíveis – 0 ou 1. Assim, no caso das “*dummys*” os efeitos marginais correspondem à diferença entre as duas probabilidades associadas aos dois valores da “*dummy*” – 0 e 1⁶¹, isto é, sem a variável e com a variável, respectivamente.

Analisando os valores dos efeitos marginais verifica-se que para as variáveis contínuas relativas ao tráfego, ao comprimento e à densidade de intersecções o efeito destas é positivo nas duas últimas categorias sendo maior na categoria 1. Este efeito vem de encontro ao expectável, pois o aumento dos valores destas variáveis produz um aumento da frequência de acidentes, tal como os modelos quantitativos demonstraram. O efeito temporal, medido pela variável *ANO*, é negativo nas duas últimas categorias traduzindo a tendência de diminuição dos acidentes ao longo do tempo nos locais com maior número de acidentes.

Relativamente à ocupação do solo, os tipos de ocupação 3 (área empresarial) e 5 (zona histórica) têm um efeito positivo nas categorias 1 e 2 ocorrência de acidentes. Tal como anteriormente demonstrado pelos modelos quantitativos, os locais situados nestes dois tipos de ocupação de solo têm maiores frequências de ocorrência de acidentes. Nos locais situados em zonas de ocupação do solo do tipo 2 e 4 o efeito é contrário, isto é, verifica-se uma maior probabilidade nestes locais de ocorrer a categoria 0. Nos arcos das vias classificadas como distribuidoras locais verifica-se um efeito marginal negativo nas últimas categorias traduzindo assim a menor frequência de acidentes nos arcos com esta classificação. Das várias “*dummys*” consideradas a que apresenta maiores valores absolutos nas diferentes categorias dos efeitos marginais é a variável que identifica os locais situados na ocupação do solo do tipo 3.

⁶¹ $Pr[y|x=1] - Pr[y|x=0]$

Apresenta-se a seguir no Quadro 5.31 a matriz de previsões em que a coluna do total observado corresponde à distribuição das observações registadas nas três categorias e a última linha corresponde às observações previstas pelo modelo e que corresponde ao maior valor da probabilidade prevista (Greene, W., 2007).

Quadro 5.31 - Matriz de previsões do modelo ordenado probit – arcos

	Total Observado	0	1	2
0	1260	1101	155	4
1	557	249	283	25
2	163	12	99	52
Total Previsto	1980	1362	537	81

Os valores a negrito do Quadro 5.31 representam as observações previstas para a respectiva categoria coincidentes com as observações registadas na mesma. Assim, por exemplo na categoria 0 verifica-se que 1101 das observações previstas para esta categoria são coincidentes com as observações registadas, ou seja, 87% das observações registadas nesta categoria foram previstas pelo modelo. No caso das categorias 1 e 2 esse valor percentual é de 51% e 32%, respectivamente. No total a percentagem de observações previstas correctamente é de 73%.

O modelo ordenado, por considerar a mesma regressão para todas as categorias, é um modelo rígido e como tal pode estar a simplificar a realidade e, por esse facto, o ajuste do modelo aos dados ser menor.

Modelo Multinomial

O modelo multinomial (modelo de resposta não ordenada) pode ser fundamentado por um modelo utilitário aleatório. O termo “utilitário” no âmbito da economia designa a capacidade para estimular o consumo e pode ser reinterpretado para o estudo da sinistralidade substituindo consumo por ocorrência de acidentes. Assim, para um local i

onde existe a possibilidade de ocorrerem $J+1$ categorias, considere-se que a “utilidade” da categoria j é dada pela equação:

$$S_{ij} = \beta X_{ij} + \varepsilon_{ij} \quad (5.12)$$

Assim, a cada categoria corresponde uma função utilitária.

Se no local i ocorrer a categoria j , então assume-se que S_{ij} é o máximo entre as J funções utilitárias. Neste caso, o modelo estatístico é conduzido pela probabilidade de um local i corresponder à categoria j e é expressa pela probabilidade de S_{ij} ser maior que todas as outras categorias S_{iJ} , isto é:

$$P(j) = P(\beta_j X_i + \varepsilon_{ij} > \beta_J X_i + \varepsilon_{iJ}) \forall J \neq j \quad (5.13)$$

Pelo facto de neste modelo existir uma função utilitária para cada categoria, é possível analisar o efeito de cada variável independente nas diferentes categorias de uma forma independente.

Para a aplicação do MM, considerou-se a distribuição probabilística logit por ser a habitualmente mais aplicada em diversas áreas, nomeadamente na área dos transportes. A distribuição probit, por implicar cálculos de integrais múltiplos complexos, tem-se verificado limitadora (Greene, W. H., 2008).

A distribuição logit aplicada aos MM é, normalmente, uma distribuição Gumbel (também designada de distribuição de valores extremos) que, tal como a distribuição logística, é também uma distribuição exponencial mas, assimétrica. Em geral, considera-se por defeito esta distribuição no modelo multinomial, designando-se de modelo multinomial logit.

Apresentam-se no Quadro 5.32 os valores estimados dos parâmetros e respectivos níveis de confiança segundo o teste *t-student*, assim como os valores das medidas de ajuste (no Anexo 2.XXII apresentam-se os resultados completos).

Quadro 5.32 - Resultados da estimação do modelo multinomial logit – arcos

Parâmetro	Prob[Y=1]		Prob[Y=2]		Efeitos marginais		
	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Categoria		
					0	1	2
$\ln\beta_0$	-16,329	0,0000	-36,038	0,0000	2,5828	-1,3709	-1,2119
β_1	0,651	0,0000	1,149	0,0000	-0,0987	0,0646	0,0341
β_2	1,763	0,0000	3,839	0,0000	-0,2780	0,1497	0,1283
$\ln(\gamma)$	-0,062	0,1547	-0,208	0,0079	0,0108	-0,0027	-0,0081
β_3	27,211	0,0386	252,915	0,0000	-7,1182	-4,4294	11,5475
$\beta_{4,2}$	-0,442	0,0054	-0,762	0,0128	0,0668	-0,0445	-0,0223
$\beta_{4,3}$	0,890	0,0138	2,588	0,0000	-0,1499	0,0530	0,0969
$\beta_{4,4}$	-1,045	0,0001	-0,424	0,3319	0,1379	-0,1533	0,0155
$\beta_{4,5}$	0,469	0,0104	1,123	0,0003	-0,0755	0,0363	0,0392
$\beta_{5,2}$	-0,015	0,9110	-1,184	0,0000	0,0192	0,0388	-0,0580
<i>Mc Fadden R^2_{adj}</i>	0,3162						
<i>AIC</i>	1,16251						
<i>função logaritmica de verosimilhança</i>	-1130,881						

Os resultados relativos à estimação dos parâmetros não consideram a categoria 0 por a mesma ser tomada como referência. No MM os valores estimados dos parâmetros representam o efeito das variáveis independentes na “utilidade” de cada categoria em alternativa à categoria 0.

No Quadro 5.32 verifica-se que na categoria 1 o efeito temporal medido pela variável ANO e a classificação hierárquica das vias não têm significância estatística. Na categoria 2 apenas a “dummy” que identifica a ocupação do solo 4 (área de equipamento) não tem significância estatística. Nesse sentido, não é possível aferir sobre a influência destas variáveis nas referidas categorias.

Ao compararmos os valores dos parâmetros estimados para as duas categorias, verifica-se que têm os mesmos sinais mas com ordem de grandezas diferentes em que esta última é maior na categoria 2, ou seja, o efeito das variáveis é maior na ocorrência de um número elevado de acidentes representado pela categoria 2. Os sinais dos parâmetros estimados são

mais uma vez coerentes com os sinais dos mesmos parâmetros obtidos nos modelos de resposta quantitativa.

Por estes factos, pode concluir-se que existem diferenças substanciais garantidas pelas variáveis consideradas no modelo que distingam a ocorrência das três categorias.

Os valores estimados dos parâmetros são, contudo, difíceis de interpretar quanto à magnitude e direcção dos efeitos nas três categorias sendo, por isso, habitual analisarem-se as relações das variáveis independentes com a variável de resposta através dos efeitos marginais determinados segundo a equação (4.42). Tal como no MOP, a soma dos efeitos marginais das diferentes categorias para cada variável é igual a zero.

No Quadro 5.32 é possível constatar que os efeitos marginais das variáveis na categoria 0 têm sinais contrários aos valores dos parâmetros estimados das mesmas, tal como verificado no MOP. Os efeitos marginais das variáveis *ANO*, *TMDA_{Tot}* e *Comp* têm uma variação entre as categorias idêntica aos efeitos marginais resultantes da aplicação do MOP havendo contudo diferenças nos valores. A densidade de intersecções tem um efeito negativo na categoria 1 ao contrário do verificado no MOP havendo ainda grandes diferenças na grandeza dos valores. Relativamente às “*dummys*” constata-se que apenas a ocupação do solo 2 tem a mesma variação dos valores dos efeitos do MOP, ou seja, da categoria 0 para a categoria 1 há uma mudança de sinal e desta última para a categoria 2 o sinal mantém-se e o valor absoluto diminui. As restantes “*dummys*” apresentam diferenças no aumento ou diminuição do efeito na mudança de categoria relativamente ao MOP e no caso da *OcSolo4* e *DisLocal* verifica-se sinais diferentes na categoria 2 e 1, respectivamente. Ao contrário dos efeitos marginais resultantes da aplicação do MOP, neste caso não se destaca nenhuma “*dummy*” cujo valor absoluto dos efeitos marginais seja superior aos das restantes. A “*dummy*” *OcSolo3* apresenta maiores valores para a categoria 0 e 2 e a “*dummy*” *OcSolo4* apresenta o maior valor na categoria 1.

A análise comparativa entre o MOP e o MML sugere existirem diferenças nos efeitos marginais que justifiquem uma avaliação estatística para determinar qual a estrutura mais ajustada aos dados.

A matriz de previsões permite avaliar a eficácia dos modelos na previsão das observações nas respectivas categorias. Apresenta-se a seguir, no Quadro 5.33, a matriz de previsões do MML.

Quadro 5.33 - Matriz de previsões do modelo multinomial logit – arcos

	Total Observados	0	1	2
0	1260	1119	140	1
1	557	265	256	36
2	163	15	79	69
Total Previsto	1980	1399	475	106

De notar que neste quadro, tal como no Quadro 5.31, os valores a negrito correspondem ao número de observações previstas coincidente com as observações registadas. Comparando esta matriz com a matriz do MOP apresentada no Quadro 5.31, pode concluir-se que as diferenças não são muito elevadas, sendo que os valores percentuais do total de previsões correctas são iguais – 73%. A categoria 0 continua a ser aquela com maior percentagem (89%).

Para comparar o ajuste dos modelos aos dados analisaram-se os valores das medidas AIC e McFadden Pseudo R^2_{adj} , apresentados no Quadro 5.30 e no Quadro 5.32. Embora os valores sejam muito próximos verifica-se que o valor de AIC é menor e o valor de McFadden Pseudo R^2_{adj} é maior no MML indicando-se assim um melhor ajuste com este modelo.

Realizou-se ainda o teste do rácio de verosimilhança ajustado. O resultado do teste - $\Phi\{-6,6669\} = 1,3 \times 10^{-11}$, indica o limite máximo da probabilidade da diferença de valores de McFadden Pseudo R^2_{adj} , entre o MOP e o MML - τ , ser desprezável. Sendo o valor desse limite baixo, conclui-se que a diferença τ é significativa e que o modelo com o valor mais elevado de McFadden Pseudo R^2_{adj} é preferível ou seja, o MML.

Considerou-se ainda neste trabalho a aplicação dos modelos de resposta qualitativa a uma variável de resposta constituída por três categorias em que se distingue a ocorrência de acidentes com vítimas: zero acidentes, acidentes sem vítimas e acidentes com vítimas.

Esta classificação, descrita no Quadro 5.34, permite atribui maior relevância aos acidentes com vítimas.

Quadro 5.34 - Categorias da variável de resposta com base na ocorrência de vítimas -
arcos

Categoria	Acidentes	Nº Observações
0	0 acidentes	669
1	Sem vítimas	695
2	Com vítimas	616

Verifica-se pelo quadro acima que a distribuição do número de observações pelas diferentes categorias é muito semelhante.

Para esta classe de categorias aplicaram-se os MO e MM cujos resultados são apresentados no Anexo 2.XXIII. Destes resultados é possível constatar que no MO (probit ou logit) a variável *ANO* não tem significância e apenas as “*dummys*” *OcSolo2* e *OcSolo5* têm significância estatística. No caso do MML, na categoria 1 apenas uma “*dummy*” tem significância estatística (*OcSolo2*) e na categoria 2 as “*dummys*” *OcSolo2* e *OcSolo5* são as únicas com significância estatística.

Destes resultados pode concluir-se que o modelo desenvolvido com um conjunto de variáveis que caracterizam a rede em fase de planeamento não permite de uma forma substancial caracterizar locais identificados pelas três categorias – zero acidentes, acidentes sem vítimas e acidentes com vítimas. Esta constatação pode evidenciar o facto destas ocorrências estarem fortemente relacionadas com combinações de factores humanos (capacidades físicas e psíquicas), do veículo (existência de equipamentos de segurança, sistemas de travagem, etc.) e da infra-estrutura que não estão relacionadas com as variáveis independentes consideradas.

Nos modelos de resposta quantitativa desenvolveram-se modelos variantes ao modelo BN para melhorar o ajuste aos dados. Nos modelos qualitativos a variável de resposta é uma categoria que, no caso em estudo, corresponde a um intervalo de frequência de acidentes, pelo que, para o objectivo em estudo, perde relevância uma eventual melhoria do ajuste, nomeadamente ao nível da eficiência dos parâmetros estimados, na medida em que não se traduz directamente num resultado substancialmente diferente.

Também o facto dos valores estimados dos parâmetros pelos diversos modelos variantes ao modelo BN não mostrarem uma variação significativa entre eles, apesar de em alguns

casos o ajuste estatístico ter sido melhorado significativamente, reforça o facto de não se justificar a aplicação de modelos variantes ao MML.

5.4.2 Modelos para nós

Os modelos de resposta qualitativa a aplicar aos nós consideram as mesmas formas funcionais para as variáveis independentes anteriormente identificadas e representadas na equação (5.6). Para a variável de resposta consideraram-se três categorias às quais corresponde um número ou intervalo de acidentes identificados para cada categoria. No Quadro 5.35 definem-se as três categorias e refere-se o número de observações associadas a cada uma delas. Como se referiu anteriormente, a primeira categoria considera a não ocorrência de acidentes ou a ocorrência de um número reduzido (1 ou 2). A segunda categoria tem como limite superior o valor correspondente ao percentil 90 das observações registadas. No caso da base de dados dos nós esse valor corresponde a cinco acidentes.

Quadro 5.35 - Categorias da variável de resposta com base na frequência - nós

Categoria	Nº Acidentes	Nº Observações
0	0;1;2	748
1	3;4;5	225
2	≥ 6	82

Como se pode constatar pelo Quadro 5.35 a maior concentração de observações corresponde à categoria 1 (71%).

Considerando as três categorias aplicaram-se os modelos ordenado e multinomial que se descrevem a seguir.

Modelo Ordenado

O modelo ordenado é estruturado a partir da regressão latente com as mesmas variáveis e formas funcionais da equação (5.6):

$$y^* = x\beta + \varepsilon \quad (5.14)$$

em que y^* é não observado, $x\beta$ é designado de função índice e ε é o termo de erro aleatório. As categorias y relacionam-se com y^* da seguinte forma:

$$\begin{aligned} y &= 0 \text{ se } y^* \leq 0 \\ y &= 1 \text{ se } 0 < y^* \leq \mu \\ y &= 2 \text{ se } y^* > \mu \end{aligned} \quad (5.15)$$

em que μ é um parâmetro desconhecido a estimar juntamente com os parâmetros β . Assim, a cada categoria corresponde uma probabilidade calculada considerando a distribuição probit ou logit. O cálculo das probabilidades da distribuição probit é representado por:

$$\begin{aligned} P(y=0|x) &= \Phi(-x\beta), \\ P(y=1|x) &= \Phi(\mu - x\beta) - \Phi(-x\beta), \\ P(y=2|x) &= 1 - \Phi(\mu - x\beta) \end{aligned} \quad (5.16)$$

em que $\Phi(\cdot)$ representa a função de distribuição normal acumulada, e que no caso da distribuição logit se representa a função de distribuição logística acumulada pelo símbolo $\Lambda(\cdot)$. Neste trabalho consideraram-se as duas hipóteses. Como os valores das medidas AIC e McFadden Pseudo R^2_{adj} e da matriz de previsões para os dois modelos são praticamente iguais (ver Anexo 2.XXIV), optou-se por considerar a mesma distribuição anteriormente seleccionada para o modelo dos arcos - modelo ordenado probit (MOP).

Apresentam-se no Quadro 5.36 os valores estimados dos parâmetros e respectivos níveis de confiança segundo o teste *t-student*, os efeitos marginais das variáveis em cada categoria e os valores das medidas da qualidade de ajuste consideradas.

Quadro 5.36 - Resultados da estimação do modelo ordenado probit – nós

Parâmetro	Valor estimado	Desvio padrão	P[Z>z]	Efeitos marginais		
				Categoria		
				0	1	2
$\ln(\beta_0)$	-7,804	1,239	0,0000	-	-	-
β_1	0,630	0,126	0,0000	-0,1132	0,0766	0,0366
β_2	0,206	0,030	0,0000	-0,0370	0,0250	0,0120
$\ln(\gamma)$	-0,178	0,051	0,0004	0,0372	-0,0251	-0,0120
$\beta_{3,2}$	0,039	0,190	0,8390	-0,0048	0,0032	0,0016
$\beta_{3,3}$	1,504	0,408	0,0002	-0,3180	0,1588	0,1592
$\beta_{3,4}$	-0,103	0,318	0,7461	0,0317	-0,0219	-0,0098
$\beta_{3,5}$	0,526	0,202	0,0091	-0,1054	0,0671	0,0383
$\beta_{4,1}$	-0,526	0,324	0,1047	0,0856	-0,0610	-0,0246
$\beta_{4,3}$	-0,943	0,183	0,0000	0,1692	-0,1176	-0,0516
μ	1,755	0,112	0,0000	-	-	-
Mc Fadden R^2_{adj}				0,0900		
AIC				1,40878		
função logaritmica de verosimilhança				-732,1338		

Como se pode constatar do Quadro 5.36 a estimação dos parâmetros corresponde a um nível de confiança superior a 95%, excepto para os parâmetros $\beta_{3,2}$, $\beta_{3,4}$ e $\beta_{4,1}$. Os dois primeiros parâmetros correspondem às variáveis de ocupação do solo do tipo 2 e 4, respectivamente e, tal como já anteriormente referido nos modelos de resposta quantitativa, não é possível diferenciar o efeito das mesmas na sinistralidade da ocupação do solo 1. O parâmetro $\beta_{4,1}$ corresponde à variável de classificação hierárquica dos ramos do nó (neste caso à classe A). Ao contrário do que se verificou nos modelos quantitativos, no MOP esta variável não tem significância estatística que valide o valor estimado do parâmetro que traduz a relação desta variável com as categorias definidas. Mais à frente, com a aplicação do MM, poder-se-á concluir se a falta de significância estatística resulta da aplicação do MOP ou de qualquer modelo qualitativo que relaciona esta variável com as três categorias definidas.

Estes valores estimados dos parâmetros pelo MO indicam o efeito das variáveis independentes na propensão latente de ocorrer determinada categoria de sinistralidade. A probabilidade associada a cada categoria é calculada pela equação (5.16) e pode ser representada pelo gráfico da Figura 5.23 e em que esta corresponde à área delimitada pela curva da distribuição e pela recta correspondente aos valores da regressão latente e do parâmetro μ .

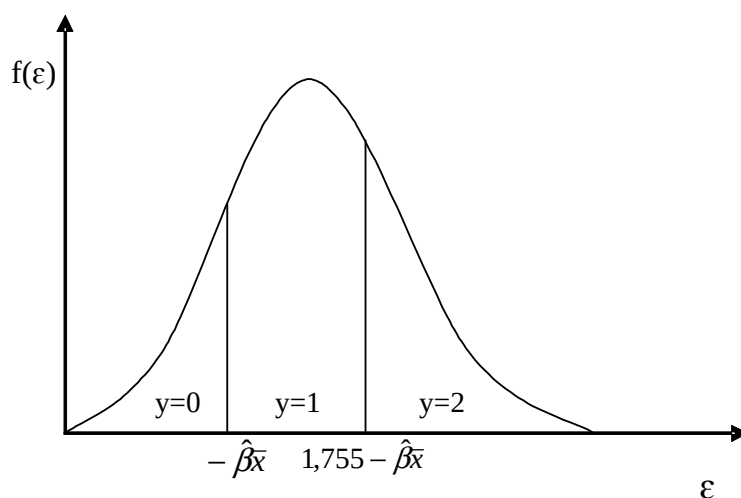


Figura 5.23. - Probabilidades do modelo ordenado estimado para os nós

Consideraram-se mais uma vez os efeitos marginais para analisar a magnitude e o sentido do efeito das variáveis independentes nas respectivas categorias.

Pela análise do Quadro 5.36 verifica-se que os tráfegos dos ramos principal e secundário têm um efeito idêntico nas três categorias embora esse efeito seja mais preponderante para o tráfego do ramo principal. Para os dois tipos de volumes de tráfego, verifica-se que na categoria 0 o efeito é negativo (sinal contrário ao da parâmetro estimado) e na categoria 1 o efeito é maior do que nas restantes categorias. Estas relações entre as variáveis e as categorias são uma característica do MOP pois nele se considera uma única regressão latente e com uma determinada distribuição probabilística que neste caso é a normal. Por este facto, a variação dos efeitos marginais entre as três categorias para as outras variáveis obedece a estas características. Das três “*dummys*” com significância estatística, a que produz um maior efeito em valor absoluto em qualquer uma das categorias é a “*dummy*” *OcSolo3*. A *ClasseC* relativamente à *ClasseB* tem um efeito preponderante nas várias categorias sendo positivo na categoria 0.

O efeito temporal medido pela variável *ANO* é positivo na categoria 0 e negativo na categoria 1 e 2, concluindo-se assim, que o efeito do tempo produz uma diminuição do número de acidentes, tal como constatado pelo modelo BN.

Apresenta-se a seguir, no Quadro 5.37, a matriz de previsões resultante da aplicação do MOP.

Quadro 5.37 - Matriz de previsões do modelo ordenado probit – nós

	Total Observados	0	1	2
0	748	739	9	0
1	225	217	8	0
2	82	74	7	1
Total Previsto	1055	1030	24	1

Os valores a negrito do Quadro 5.37 representam as observações previstas para a respectiva categoria coincidentes com as observações registadas na mesma. Pelo quadro constata-se que o modelo não previu eficazmente as observações para as categorias 1 e 2, tendo previsto correctamente 99% das observações registadas no caso da categoria 0. No total foram previstas correctamente 71% das observações registadas.

Apresenta-se a seguir a aplicação do MM ajustado a esta base de dados.

Modelo Multinomial

O modelo multinomial é estruturado por uma função utilitária para cada categoria j :

$$S_{ij} = \beta X_{ij} + \varepsilon_{ij} \quad (5.17)$$

Se no local i ocorrer a categoria j , então assume-se que S_{ij} é o máximo entre as J funções utilitárias:

$$P(j) = P(S_{ij} > S_{iJ}) \forall J \neq j \quad (5.18)$$

Pelo facto de neste modelo existir uma função utilitária para cada categoria, é possível analisar o efeito de cada variável independente nas diferentes categorias.

Aplicou-se à base de dados dos acidentes ocorridos nos nós o MML. No Quadro 5.38 são apresentados os resultados dos valores estimados dos parâmetros e respectivos níveis de confiança segundo o teste *t-student*, assim como os valores das medidas de ajuste (apresentam-se no Anexo 2.XXV os resultados completos).

Os valores dos parâmetros estimados das categorias 1 e 2 representam o efeito das variáveis independentes na “utilidade” de cada categoria em alternativa à categoria 0.

Do Quadro 5.38 verifica-se que nas duas categorias as “*dummys*” *OcSolo2* e *OcSolo4* não têm significância estatística, tal como anteriormente verificado no modelo de resposta quantitativa. Na categoria 1 a estimação do parâmetro da variável *ANO* também não apresenta significância estatística e na categoria 2 as “*dummys*” *OcSolo5* e *ClasseA* igualmente não apresentam significância estatística. Comparando os valores dos parâmetros estimados com significância estatística verifica-se que os valores absolutos são sempre maiores na categoria 2 indicando assim um efeito maior destas variáveis nesta categoria.

Quadro 5.38 - Resultados da estimação do modelo multinomial logit – nós

Parâmetro	Prob[Y=1]		Prob[Y=2]		Efeitos marginais		
	Valor estimado	P[Z>z]	Valor estimado	P[Z>z]	Categoria		
					0	1	2
$\ln(\beta_0)$	-6,878	0,0000	-12,801	0,0000	1,4783	-0,7950	-0,6833
β_1	0,479	0,0007	1,031	0,0000	-0,1090	0,0523	0,0116
β_2	0,196	0,0000	0,243	0,0000	-0,0368	0,0252	0,0116
$\ln(\gamma)$	-0,087	0,1258	-0,368	0,0000	0,0277	-0,0057	-0,0220
$\beta_{3,2}$	-0,066	0,7639	0,090	0,7772	0,0049	-0,0122	0,0073
$\beta_{3,3}$	1,224	0,0093	2,171	0,0014	-0,2585	0,1439	0,1146
$\beta_{3,4}$	0,455	0,1713	-30,155	1,0000	1,2463	0,7239	-1,9703
$\beta_{3,5}$	0,566	0,0129	0,468	0,1804	-0,0963	0,0782	0,0181
$\beta_{4,1}$	-0,836	0,0362	-0,180	0,7131	0,1200	-0,1264	0,0064
$\beta_{4,3}$	-0,716	0,0003	-1,677	0,0000	0,1688	-0,0753	-0,0935
<i>Mc Fadden R²_{adj}</i>	0,0926						
<i>AIC</i>	1,40084						
<i>função logaritmica de verosimilhança</i>	-718,9424						

Destes resultados pode concluir-se que o efeito de tempo só se manifesta nos locais com um número elevado de acidentes e que os nós localizadas na zona histórica e central do Porto (ocupação do solo do tipo 5) e os nós cujos ramos são classificados como via colectora que se intersecta com ramos de via distribuidora principal ou local (*ClasseA*), não influenciam a ocorrência de um número elevado de acidentes.

Para melhor interpretar os efeitos das variáveis na variável de resposta considerou-se a análise dos efeitos marginais apresentados no mesmo quadro. Assim, constata-se que o efeito do tempo vai aumentando à medida que se passa da categoria 0 para a categoria 2. O efeito do volume de tráfego dos ramos principais aumenta com as categorias e no caso do volume de tráfego dos ramos secundários o efeito é maior na categoria 1. Nas “*dummys*” os efeitos marginais são muito idênticos aos verificados no MOP, quer na variação entre as categorias, quer na ordem de grandeza dos valores. Aliás, em todas as variáveis com significância estatística verifica-se que os valores dos dois modelos têm ordem de grandeza muito semelhante, embora nestas se verifique variações ligeiramente diferentes.

Apresenta-se a seguir, no Quadro 5.39, a matriz de previsões do MML.

Quadro 5.39 - Matriz de previsões do modelo multinomial logit – nó

	Total Observados	0	1	2
0	748	737	11	0
1	225	207	13	5
2	82	70	9	3
Total Previsto	1055	1014	33	8

De notar que os valores a negrito correspondem ao número de observações previstas coincidente com as observações registadas. Pelo quadro, constata-se que tal como no MOP à categoria 0 corresponde a maior percentagem de previsões correctas (99%). Nas categorias 1 e 2 esse valor reduz-se significativamente, sendo a percentagem total de previsões correctas de 71%. Este último valor é igual ao verificado pelo Quadro 5.37 do MOP.

Para comparar o ajuste dos modelos aos dados analisou-se os valores das medidas AIC e McFadden Pseudo R^2_{adj} apresentados no Quadro 5.36 e no Quadro 5.38. Tal como se verificou nos modelos aplicados à base de dados dos arcos, os valores destas medidas são muito próximos verificando-se, no entanto, que o valor de AIC é menor e o valor de McFadden Pseudo R^2_{adj} é maior no MML.

Aplicou-se o teste do rácio de verosimilhança ajustado para determinar se os valores da medida McFadden Pseudo R^2_{adj} são significativamente diferentes. O resultado da teste - $\Phi\{-3,7729\} = 8,1 \times 10^{-5}$, indica o limite máximo da probabilidade da diferença nos valores de McFadden Pseudo R^2_{adj} entre o MOP e o MML - τ , ser desprezável. Sendo o valor desse limite baixo, concluiu-se que a diferença τ é estatisticamente significativa e que o modelo com o valor mais elevado de McFadden Pseudo R^2_{adj} é preferível ou seja, o MML. Considerou-se ainda outra classe de categorias de forma a ter em conta a ocorrência de vítimas. A classe de categorias considerada é a mesma anteriormente aplicada aos arcos, sendo que o número de observações distribuído pelas categorias está apresentado no Quadro 5.40.

Quadro 5.40 - Categorias da variável de resposta com base na ocorrência de vítimas - nós

Categoria	Acidentes	Nº Observações
0	0 acidentes	324
1	Sem vítimas	428
2	Com vítimas	303

Verifica-se pelo quadro acima que a distribuição do número de observações pelas categorias 0 e 2 é muito idêntica sendo o número de observações maior na categoria 1.

Para esta classe de categorias aplicaram-se os modelos ordenado e multinomial, cujos resultados são apresentados no Anexo 2.XXVI. Destes resultados é possível constatar que no MOP as “dummies” *OcSolo2* e *OcSolo4* não têm significância estatística e a “dummy” *OcSolo3* só é significativa para um nível de confiança superior a 10%.

No caso do MML, na categoria 1, das “dummies” consideradas, apenas a *OcSolo3* e *ClasseA* têm significância estatística e na categoria 2 as “dummies” *OcSolo5*, *ClasseA* e *ClasseC* têm significância estatística.

Destes resultados pode concluir-se que, tal como verificado na base de dados dos arcos, para esta classe de categorias não é possível, de uma forma relevante, relacionar as três categorias com as variáveis que caracterizam a rede em fase de planeamento. Outros factores, tais como, o factor humano e as características dos veículos, deverão ter uma influência na ocorrência dos acidentes com vítimas que não deverão ser desprezados.

5.5 Conclusões

Os modelos de resposta quantitativa têm vindo a ser amplamente utilizados para a análise das causas dos acidentes nomeadamente no que à infra-estrutura diz respeito. Tendo como base o teorema de Bernoulli generalizou-se a aplicação da distribuição de Poisson para descrever a ocorrência de acidentes num determinado local. Contudo, esta distribuição tem a limitação de impor para a sua correcta aplicação, uma base de dados em que o valor médio dos acidentes registados seja igual à variância. Por este facto, a aplicação de modelos evoluiu para a eventual consideração da distribuição BN, cuja característica principal reside no facto de permitir que a média seja diferente da variância.

Nas bases de dados consideradas para a modelação dos arcos e nós, verificou-se que esta distribuição é a mais ajustada. Outras variantes do modelo BN (modelo base) foram consideradas com o objectivo de tentar melhorar o ajuste aos dados. Em alguns desses modelos, e considerando os valores das medidas de avaliação do ajuste usadas, verificou-se um melhor ajuste estatístico. Contudo, é de salientar que os valores estimados para os parâmetros das variáveis independentes consideradas não alteraram o sentido do efeito na variável de resposta (isto é, o sinal do valor estimado) e os valores absolutos também não se alteraram substancialmente. Nesse sentido, o modelo base com distribuição BN, por ser de simples aplicação, é uma ferramenta aplicável na fase de planeamento para apoio à decisão.

Os modelos de resposta qualitativa têm sido fundamentalmente aplicados no estudo dos transportes e em alguns estudos de sinistralidade considerando a gravidade das vítimas. Nestes modelos é necessário definir uma classe de categorias como variável de resposta. Essas categorias devem traduzir diferentes efeitos das variáveis independentes e graus de sinistralidade que identifiquem situações de gravidade distintas. Nesse sentido, várias

hipóteses, baseadas em intervalos de frequência de acidentes, foram analisadas tendo-se optado por considerar uma classe de três categorias em que a primeira categoria corresponde a um grau de sinistralidade baixo cujo intervalo de frequências traduza a eventual existência de acidentes resultante de causas fortuitas e não devido a características dos locais. Para delimitar as duas categorias seguintes, de forma a que estas correspondam a graus distintos e ordenados por ordem crescente de gravidade, considerou-se o número de acidentes correspondente a 10% das observações com maior número de acidentes.

Para esta classe de categorias aplicaram-se dois modelos qualitativos – modelo ordenado e modelo multinomial. O modelo ordenado, como o próprio nome indica, estabelece uma ordenação das categorias e é estruturado por uma regressão latente que relaciona as variáveis independentes com a variável de resposta y^* . Esta característica do MO impõe que os valores dos parâmetros estimados sejam os mesmos, qualquer que seja a categoria. No MM a variável de resposta não obedece a uma ordenação sendo, por isso, estruturado por uma função “utilitária” para cada categoria que representa as relações entre as variáveis independentes e a variável de resposta.

Das aplicações efectuadas, quer para a base de dados dos arcos quer para a base de dados dos nós, as medidas de avaliação do ajuste indicam ser preferível o MM. De facto, contrariando a aparente vantagem de ter à partida a existência de uma ordenação subjacente às três categorias consideradas, verificou-se que no MM os valores estimados e as significâncias estatísticas são diferentes nas categorias e os efeitos marginais variam de forma distinta do MO. Estas diferenças foram mais substanciais nos modelos aplicados aos arcos. Assim, estes factos contrariam a hipótese de considerar uma única regressão para as categorias tal como o MO estabelece.

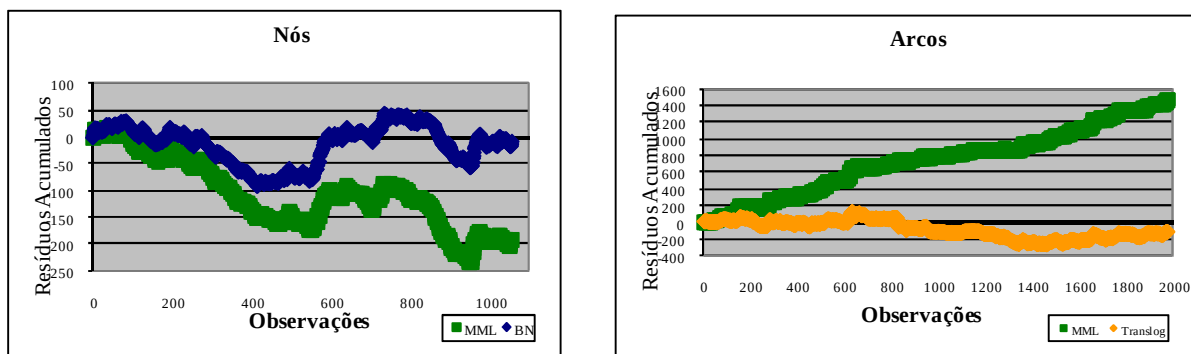
Considerou-se ainda a análise de uma classe de três categorias que diferenciam a não ocorrência de acidentes, a ocorrência de acidentes sem vítimas e a ocorrência de acidentes com vítimas. Os resultados desta aplicação indicam que das “*dummys*” incluídas no modelo para descrever a rede, poucas têm significância estatística. Isto é mais evidente no caso do modelo dos arcos. Este resultado evidencia que, no caso das referidas categorias, a classificação viária da rede e da ocupação do solo, não é muito eficiente a identificar os locais onde poderão ocorrer as classes definidas. De facto, os acidentes com vítimas estão, em geral, associados a factores humanos e dos veículos que não foram incluídos no modelo.

A selecção do modelo a aplicar como ferramenta para avaliar a sinistralidade da rede urbana em fase de planeamento deve ter em conta que o objectivo dessa avaliação é comparar soluções de intervenção alternativas elaboradas, no panorama nacional, na fase de definição do PDM. O modelo a aplicar nesta fase deve ser simples e o resultado dessa aplicação deve indicar um nível de sinistralidade inerente a cada solução de intervenção para assim seleccionar a melhor em termos de segurança rodoviária, ou ponderar essas soluções com base no critério da sinistralidade ou eventualmente mitigar o cenário seleccionado nos locais identificados pelo modelo. Assim, analisando os dois tipos de abordagem – modelo de resposta quantitativa e de resposta qualitativa, sabe-se que no primeiro obtém-se um valor previsto de frequência de acidentes para um determinado ano e do segundo, o resultado final compreende uma probabilidade associada à ocorrência das três categorias às quais corresponde um intervalo de frequência de acidentes. Este último, comparativamente ao primeiro, por não indicar um único valor previsto, assume de certa forma um intervalo de incerteza que pode traduzir a dificuldade de encontrar um modelo perfeito com uma estrutura funcional perfeita dando ainda uma margem de incerteza a todo o processo de obtenção dos dados dos acidentes desde o registo até à localização destes. Acrescentando o facto de o modelo a aplicar na fase de planeamento incluir variáveis independentes que caracterizam a rede de uma forma generalista, entende-se assim que o modelo de resposta qualitativa se compatibiliza melhor com a função de uma ferramenta aplicada nesta fase. Não obstante, o facto do resultado dos modelos de resposta quantitativa poder ser integrado num intervalo de frequência de acidentes definindo igualmente níveis de gravidade de sinistralidade. Acresce ainda que, das duas hipóteses estudadas como modelos de resposta qualitativa, o MM possibilita que se estabeleça relações entre as variáveis independentes e a variável de resposta por níveis de gravidade (assumindo independência entre categorias), o que enriquece a informação acerca da sinistralidade da rede viária, facto importante quando se pretende melhorar o cenário analisado/seleccionado ou propor medidas mitigadoras para determinados locais.

Para melhor aferir os resultados dos dois tipos de modelos, realizou-se uma análise comparativa dos mesmos, a fim de se conhecer qual o que poderá conduzir a melhores estimativas. Para tal, consideraram-se os valores previstos para cada observação pelos modelos quantitativo e qualitativo seleccionados que correspondem, no caso dos arcos, ao modelo com função Translog e MML, respectivamente e no caso dos nós ao modelo BN e MML, respectivamente. Para transformar a variável de resposta dos modelos MML num

valor comparável ao da variável de resposta do modelo quantitativo respectivo, foi necessário determinar, nesse caso, a esperança matemática do número de acidentes a partir dos valores das probabilidades associadas a cada categoria e um valor representativo do intervalo de frequência considerado para cada categoria. Para este último, dado que os intervalos de frequência definidos para cada categoria são diferentes e com uma dispersão muito irregular, optou-se por considerar para cada categoria o valor da mediana. Para os valores previstos pelos modelos quantitativo e qualitativo calculou-se os resíduos acumulados, tal como mostra a Figura 5.24.

Figura 5.24. - Gráficos dos resíduos acumulados dos modelos quantitativos e qualitativos aplicados aos arcos e nós



Como se pode verificar pelos gráficos da Figura 5.24, os resíduos acumulados obtidos pelo modelo quantitativo, quer dos arcos quer dos nós, estão mais próximos do valor nulo.

Contudo, relembra-se que a transformação dos resultados qualitativos dados pelos modelos MML em valores quantitativos que possam ser comparáveis é muito simplista (utilização da mediana), pelo que atendendo aos fundamentos anteriormente referidos para os modelos qualitativos, continuam a ser estes os modelos propostos como ferramenta de previsão da sinistralidade no contexto do planeamento da rede viária urbana e ainda, para análise da influência das diferentes variáveis incluídas no modelo.

Assim, considerando o MML aplicado aos arcos e nós, analisa-se a seguir a influência das variáveis independentes na variável de resposta.

Neste modelo, verifica-se também que a presença das variáveis relativas ao tráfego, quer no caso dos modelos de arcos quer nos modelos de nós, é preponderante pois a significância destas variáveis é sempre elevada em todas as categorias do MM. No caso

dos arcos, verifica-se ainda a importância da variável relativa ao comprimento dos arcos e à densidade de intersecções, embora esta última apresente um efeito significativamente distinto entre as categorias consideradas verificado pelos valores estimados dos parâmetros desta variável. O efeito temporal medido pela variável *ANO* só é significativo na categoria 2 quer no modelo dos nós quer no modelo dos arcos (número de acidentes maior ou igual a nove e número de acidentes maior ou igual a seis, no caso dos arcos e dos nós, respectivamente), ou seja, os acidentes tendem a diminuir no tempo de uma forma gradual nos locais onde ocorre um número elevado de acidentes. Esta maior dispersão no tempo nos locais onde se regista um menor número de acidentes pode advir do facto de que estes possam resultar de maioritariamente de factores humanos. Contudo, é de salientar que este efeito da variável *ANO* reflecte-se nos resultados do modelo aplicado a qualquer uma das soluções de intervenção a analisar na fase de planeamento, não interferindo, por isso, na comparação dessas diferentes soluções. Contudo, se se retirar esta variável, as probabilidades associadas a cada categoria podem ser de tal forma alteradas que representem irrealistamente a sinistralidade de cada solução de intervenção.

As “*dummys*” consideradas para caracterizar a envolvente, a geometria e funcionamento dos locais, são comuns aos modelos dos arcos e nós. Verificou-se que, no caso dos nós, as “*dummys*” que identificam a ocupação do solo tipo 2 e 4 não têm significância estatística. Este facto verificou-se em todos os modelos de resposta qualitativa ou quantitativa, não sendo por isso possível analisar os efeitos destas “*dummys*” na variável de resposta de forma diferenciada dos efeitos da variável de ocupação do solo 1. Contudo, estas envolventes, no caso dos arcos, têm o efeito de reduzir o risco de acidentes principalmente na categoria intermédia. Já a ocupação do solo do tipo 3 e 5 que identificam uma zona empresarial e uma zona histórica no centro da cidade, respectivamente, têm o efeito de aumentar o risco de acidentes quer nos nós quer nos arcos. Destes dois tipos de envolvente, a zona empresarial é a de maior risco quer nos arcos quer nos nós sendo os valores estimados dos parâmetros muito idênticos nestes dois tipos de locais.

As “*dummys*” consideradas para identificar o tipo de classificação hierárquica das vias são diferentes nos modelos das duas entidades, pois no caso dos nós os ramos podem ter classificações diferentes. Nos arcos, as vias classificadas de distribuidora local são de menor risco de acidente sendo este efeito maior nos arcos onde ocorre um número elevado de acidentes (categoria 2). Nos nós em que os ramos que se intersectam são do tipo via colectora com via distribuidora (*ClasseA*), o risco de acidentes diminui na categoria 1

(número de acidentes entre três e cinco inclusive), e na categoria 2 este efeito não tem significância estatística. Quando as vias de todos os ramos do nó são distribuidoras local ou quando as vias distribuidoras local se intersectam com vias de acesso local (*ClasseC*), o risco de acidentes também diminui, neste caso em qualquer uma das categorias. Os nós de maior risco são aqueles cujos ramos são de vias distribuidoras principal ou quando as vias distribuidoras local se intersectam com vias de acesso local (*ClasseB*). Os nós classificados de classe B são os nós predominantes na rede em estudo. Esta classe intermédia pode induzir um maior risco de acidente porque, por um lado, nestes nós se verifica, em geral, um maior número de conflitos relativamente aos da classe C mas, por outro lado, não têm a importância na rede dos nós da classe A que, por esse facto, serão nós, provavelmente, melhor sinalizados ou que induzem no condutor comportamentos de menor risco.

Embora a análise acima apresentada sobre a influência das variáveis no risco de acidentes seja baseada no MM, pode generalizar-se aos restantes modelos, mesmo aos modelos quantitativos já que, com algumas variações, todos eles apontam no mesmo sentido do efeito de risco de ocorrer acidentes.

Para melhor analisar os efeitos na variável de resposta das diferentes variáveis independentes consideraram-se cenários hipotéticos para dois locais – um arco e um nó, para um período temporal de dez anos com duas alternativas para o volume de tráfego: aumentar cerca de 30%, valor este que corresponde à taxa de crescimento observada nos cinco anos da base de dados, ou uma redução de 20%, num cenário hipotético de aplicação de medidas de redução do volume de tráfego num troço da rede urbana. Analisou-se ainda a possibilidade de nesses locais se considerar um conjunto de transformações que alterem as características geométricas e funcionais de tal forma, que uma via inicialmente classificada como distribuidora local passe a ser considerada como via distribuidora principal. Por fim, analisou-se o efeito da introdução de um pólo comercial numa zona habitacional pouco densa, com implicações na classificação do tipo de ocupação do solo, em que os locais analisados, inicialmente situados numa zona de ocupação do solo do tipo 2, passam a ser classificados como zona de ocupação do solo do tipo 4.

Assim, definiram-se dois locais de referência caracterizando-os com base nas variáveis consideradas e, para esses valores das variáveis, determinaram-se as probabilidades de ocorrer cada uma das três categorias de sinistralidade. A partir destes valores de base das probabilidades, analisaram-se as alterações produzidas pelas referidas modificações das

variáveis, tal como se apresenta no Quadro 5.41 e 5.42. A cada linha do quadro corresponde o efeito nas probabilidades de uma das variáveis e no final, a última linha, corresponde ao acumulado dos efeitos das variáveis que sofreram variações. No caso da diminuição do volume de tráfego em cerca de 20%, analisou-se a influência desta variável sem considerar as alterações nas outras variáveis, por as mesmas não serem coerentes com a diminuição do tráfego.

Quadro 5.41 - Probabilidades percentuais de ocorrer cada categoria em cenários hipotéticos de um arco

		Prob(y=0)%	Prob(y=1)%	Prob(y=2)%
Arco referência: $TMDA_{Tot} - 40\ 000$; $Comp - 615$; $DenInt - 0,01$; $OcSolo - 2$; $ClHier - DisLocal$		11	55	34
$TMDA_{Tot}$	+30%	9	53	38
ANO	10	22	67	11
$OcSolo$	2 para 4	12	34	54
$ClasHier$	DisLocal para DisPrinc	6	31	63
Efeitos acumulados		17	32	51
$TMDA_{Tot}$	-20%	13	56	31

Do Quadro 5.41 pode verificar-se que, atendendo às características do arco na situação de referência, corresponde uma probabilidade maior de ocorrer a categoria 1 (número de acidentes entre três e oito inclusive), cerca de 55%, sendo que a categoria 0 e 2 têm probabilidades de ocorrer de 11% e 34%, respectivamente, isto é, para este tipo de arco a probabilidade de ocorrência da categoria 1 ultrapassa os 50%, reservando-se para as outras duas categorias uma probabilidade igual a 45%. Com o aumento de 30% do tráfego, verifica-se um aumento da probabilidade de ocorrer a última categoria, diminuindo as restantes, embora com valores muito próximos aos anteriormente verificados. No caso da variável ANO , que traduz a tendência da frequência dos acidentes ao longo do tempo, verifica-se um efeito mais preponderante no aumento da probabilidade de ocorrerem as categoria 0 e 1, representando assim a tendência para a diminuição dos acidentes ao longo do tempo. A simulação da alteração da ocupação do solo do tipo 2 para o tipo 4, conduz a um aumento significativo da probabilidade de ocorrer um número elevado de acidentes (categoria 2). Embora, tal como se referiu anteriormente, estes dois tipos de ocupação do

solo diminuem o risco de acidentes relativamente à ocupação do solo¹, esta redução é maior na ocupação do solo tipo 2, no caso da probabilidade da categoria 2. As intervenções ao nível da infra-estrutura do arco que alterem as características geométricas e funcionais deste, mudando assim a sua classificação hierárquica de distribuidora local para principal, traduzem-se numa maior probabilidade de ocorrer a categoria 2, aumentando substancialmente assim o risco de acidente. Finalmente, o efeito total das alterações das variáveis referidas repercute-se num aumento relativamente à situação inicial, das categorias 0 e 2, passando a última (categoria 2) a ser a de maior probabilidade de ocorrência de acidentes. Assim, das variações simuladas para as referidas variáveis, verifica-se que o aumento da probabilidade de ocorrer a categoria 2 é maior quando se altera o tipo de classificação hierárquica, mais especificamente, quando a via distribuidora local passa a distribuidora principal. Pelo contrário, o aumento da probabilidade de ocorrer a categoria 0, é maior com o aumento do número de anos. A partir do Quadro 5.41 é possível ainda analisar o efeito da diminuição do volume de tráfego em cerca de 20%. Tal como seria de esperar, esta variação do tráfego reflecte-se num aumento das categorias 0 e 1, embora a diferença dos valores relativamente à situação inicial seja residual. A mesma análise foi ainda realizada para um nó hipotético e cujos resultados estão no Quadro 5.42.

Quadro 5.42 - Probabilidades percentuais de ocorrer cada categoria em cenários hipotéticos de um nó

		Prob(y=0)%	Prob(y=1)%	Prob(y=2)%
Nó referência: $TMDA_{Princ} - 56\ 600$; $TMDA_{Sec} - 6\ 200$; $OcSolo - 1$; $ClHier - ClasseC$		58	28	14
$TMDA_{Princ}$ e $TMDA_{Sec}$	+30%	53	30	17
ANO	10	81	18	1
$OcSolo$	1 para 5	43	39	18
$ClasHier$	ClasseC para ClasseB	21	37	42
Efeitos acumulados		49	46	5
$TMDA_{Princ}$ e $TMDA_{Sec}$	-20%	63	26	11

Do Quadro 5.42 pode verificar-se que ao nó de referência considerado, corresponde uma probabilidade maior de ocorrer a categoria 0, cerca de 58%, sendo que as categorias 1 e 2

têm probabilidades de ocorrência de 28% e 14%, respectivamente, isto é, para este tipo de nó a probabilidade de ocorrer até dois acidentes é maior do que a soma das probabilidades de ocorrer um número de acidentes superior a dois. Com o aumento de 30% no tráfego, verifica-se um aumento da probabilidade de ocorrência das categorias 1 e 2, diminuindo a categoria 0, embora com valores muito próximos aos anteriormente verificados. No caso da variável *ANO*, verifica-se um efeito preponderante no aumento da probabilidade de ocorrer a categoria 0 (de 58% para 81%), demonstrando assim, tal como no arco referido anteriormente, a tendência para diminuir os acidentes com o tempo. A alteração na ocupação do solo do tipo 1 para o tipo 5, conduz a um aumento significativo da probabilidade de ocorrência da categoria 1 e 2. Considerando-se ainda a existência de intervenções ao nível da infra-estrutura de um ou mais ramos do nó, de tal forma que as características geométricas e funcionais sejam alteradas com a consequente reclassificação como nó da *ClasseB*. Esta mudança originou um aumento das probabilidades de ocorrerem as categorias 1 e 2, sendo nesta última significativamente maior. Assim, tal como se concluiu para os arcos, as variações simuladas isoladamente para cada uma das referidas variáveis, conduziram a um aumento da probabilidade de ocorrência das categorias 0 e 2, em função do aumento do número de anos e da alteração do tipo de classificação hierárquica, respectivamente.

Quanto ao efeito da diminuição do volume de tráfego em cerca de 20%, verifica-se pelo Quadro 5.42 que, tal como seria de esperar, esta variação do tráfego se reflecte num aumento da categoria 0 (de 58% para 63%) e na diminuição das restantes categorias.

Analisando os resultados do MML para todas as observações destacam-se, das seis combinações de distribuição das probabilidades de ocorrer em cada categoria, resultantes da aplicação do modelo, duas combinações que estão representadas nos gráficos da Figura 5.25. As distribuições das probabilidades de ocorrer cada uma das três categorias representadas pelos gráficos da Figura 5.25 são as mais comuns, em particular a distribuição do gráfico da esquerda. Este último, abrange um conjunto de locais (arcos e nós) que embora com características, representadas pelas variáveis consideradas no modelo, muito distintas, têm o mesmo efeito no risco de ocorrer determinada categoria. Pelo contrário, no caso dos locais cuja categoria com maior probabilidade de ocorrer corresponde à categoria 1 (gráfico da direita), é possível encontrar características em comum. Assim, no caso dos arcos verifica-se que os mesmos se caracterizam por comprimentos inferiores a um quilómetro, por uma baixa densidade de intersecções e

elevados volumes de tráfego (o volume de tráfego máximo da base de dados insere-se nestes locais). Em termos de ocupação do solo, os arcos com maior probabilidade de ocorrer a categoria 1 estão localizados em áreas de ocupação do solo do tipo 1 e 2. Relativamente à classificação viária já não se distingue claramente o tipo de arcos verificando-se que o número de arcos classificados como vias distribuidoras locais é muito próximo do número de arcos classificados como vias distribuidoras principais.

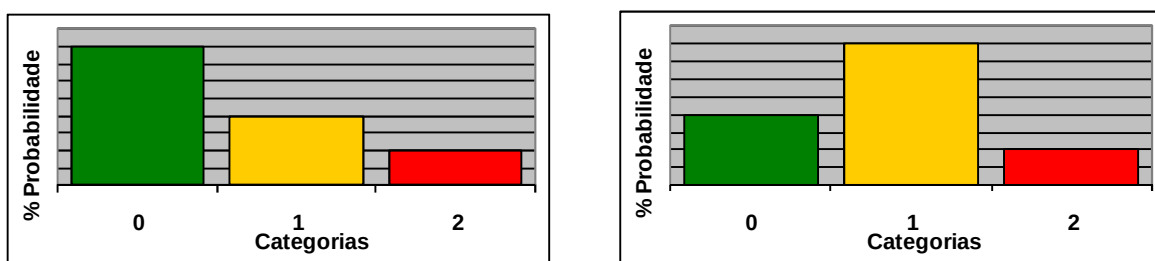


Figura 5.25. - Gráficos representativos da distribuição das probabilidades de ocorrer as três categorias

No caso dos nós com maior probabilidade de ocorrer a categoria 1, sendo estes em menor número a dispersão das características é menor, destacando-se os nós onde o volume de tráfego é elevado (ramos principal e secundário), situados em áreas de ocupação do solo do tipo 2 e classificados essencialmente na *ClasseB*.

Dos resultados da análise dos cenários hipotéticos e das combinações das probabilidades de ocorrer cada categoria predominantes na rede viária em estudo, pode concluir-se que é possível definir acções quer a um nível mais localizado quer a um conjunto de locais às quais corresponde um risco de ocorrência de acidentes. Embora os MML desenvolvidos neste trabalho se refiram a uma análise global da rede permitem, no entanto, flexibilizar a análise e consequente actuação a um nível mais microscópico (arcos e nós).

5.6 Referências bibliográficas

Berkovitz, A. (2001). The marriage of safety and land-use planning: a fresh look at local roadways. *Public Roads*, **65**.

Bhat, C. R. and Pulugurta, V. (1997). A comparison of two alternative behavioral choice mechanism for household auto ownership decisions. *Transportation Research Part B*, **32**, 61-75.

Brude, U. and Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? *Accident Analysis and Prevention*, **25**, 499-509.

Cardoso, J. (2007) *Métodos racionais de apoio à intervenção da engenharia em segurança rodoviária*. Programa de investigação para obtenção do título "Habilitado para o exercício de funções de Coordenação de Investigação Científica", LNEC, Lisboa.

Chatterjee, A. and Wegmann, F. J. (2002). *Assessing Safety Consequences of Alternative Highway Networks*. Knoxville, TN: Southeastern Transportation Center, The University of Tennessee.

Dissanayake, D., Aryaija, J. and Wedagama, D. M. P. (2009). Modelling the effects of land use and temporal factors on child pedestrian casualties. *Accident Analysis and Prevention*, **41**, 1016-1024.

Fridstrom, L., et al. (1995). Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, **27**, 1-20.

Greene, W. (2007). "LIMDEP version 9.0." *Econometric Modeling Guide - Manual*, Econometric Software, Inc.

Greene, W. H. (2008). *Econometric Analysis*, Pearson International Edition, New Jersey.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention*, **35**, 273-285.

Hauer, E. and Bamfo, J. (1997). Two tools for finding what function links the dependent variable to the explanatory variables. In: *ICTCT*, Lund.

Ivan, J. N., Pasupathy, R. K. and Ossenbruggen, P. J. (1999). Differences in causality factors for single and multi-vehicle crashes on two-lane roads. *Accident Analysis and Prevention*, **31**, 695-704.

Ivan, J. N., Wang, C. and Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention*, **32**, 787-795.

Jonsson, T. (2005) *Predictive models for accidents on urban links. A focus on vulnerable road users*. PhD, Department of Technology and Society Traffic Engineering, Lund Institute of Technology, Lund.

Karlaftis, M. G. and Tarko, A. P. (1997). Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention*, 425-433.

Kumara, S. P. and Chin, H. C. (2003). Modeling accident occurrence at signalized Tee intersections with special emphasis on excess zeros. *Accident Analysis and Prevention*, **4**, 53-57.

Lopes, S. M. d. A. and Cardoso, J. L. (2007). *Accident prediction models for Portuguese single carriageway roads*. Lisboa: LNEC.

Lord, D. (2000) *The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models*. PhD, Department of Civil Engineering, University of Toronto, Toronto.

Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, **38**, 751-766.

Lord, D. and Park, P. Y. J. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis and Prevention*, **40**, 1441-1457.

Lord, D. and Persaud, B. (2000). Accident Prediction Models with and without trend: application of the generalized estimating equations (GEE) procedure. In: *79th Annual Meeting*, Washington.

Lord, D., Washington, S. P. and Ivan, J. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, **37**, 35-46.

Maher, M. and Summersgill, I. (1995). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, **28**, 281-296.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman & Hall, London, etc.

Mountain, L., Fawaz, B. and Jarrett, D. (1996). Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, **28**, 695-707.

Mountain, L., Maher, M. and Fawaz, B. (1998). The influence of trend on estimates of accidents at junctions. *Accident Analysis and Prevention*, **30**, 641-649.

OCDE (1997). *Road safety principles and models: review of descriptive, predictive, risk and accident consequence models*. Paris: Organisation for Economic Co-operation and Development, OCDE/GD(97)153.

Partyka, S. C. (1991). Simple models of fatality trends revisited seven years later. *Accident Analysis and Prevention*, **23**, 423-430.

Poch, M. and Mannering, F. (1996). Negative Binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, **122**, 105-113.

Qi, Y., Smith, B. L. and Guo, J. (2007). Freeway accident likelihood prediction using a panel data analysis approach. *Journal of Transportation Engineering*, **133**, 149-156.

Reurings, M., et al. (2005). "Accident Prediction Models and Road Safety Impact Assessment: a state-of-the-art." RIPCORDER-ISEREST: Road Infrastructure Safety Protection - Core Research and development for Road Safety in Europe; Increasing Safety and Reliability of Secondary Roads for a Sustainable Surface Transport, 135.

Tavares, J. P. (2003) *Aplicabilidade e robustez de modelos de afectação de tráfego em redes urbanas*. Tese de Doutoramento (PhD), Faculdade de Engenharia, Universidade do Porto, Porto.

Turner, S. and Nicholson, A. (1997). Intersection accident estimation: the role of intersection location and non-collision flows. *Accident Analysis and Prevention*, **30**, 505-517.

Wedagama, D., Bird, R. and Metcalfe, A. (2006). The influence of urban land-use on non-motorised transport casualties. *Accident Analysis and Prevention*, **38**, 1049-1057.

6. CONCLUSÕES E PERSPECTIVAS FUTURAS

6.1 Síntese geral e principais conclusões

Ao longo das últimas décadas têm sido desenvolvidas diversas acções ao nível da infra-estrutura com vista a melhorar a segurança rodoviária (SR), algumas das quais num número significativo de experiências de aplicação em países da Europa e, em geral, com resultados muito positivos. Essas acções dividem-se em dois tipos segundo a fase de aplicação – acções *a priori* e *a posteriori*. As acções *a posteriori* são aplicadas quando se verifica um valor elevado para determinado(s) indicador(es) de sinistralidade e, como tal, correspondem a uma fase posterior à ocorrência dos acidentes. Para estas acções, existe um conjunto de ferramentas que têm vindo a ser desenvolvidas e aperfeiçoadas nomeadamente com a experiência de aplicação em diversos países. Contudo, as acções *a posteriori* têm a desvantagem de apenas serem aplicadas após a ocorrência de um determinado número de acidentes indesejáveis. Nesse sentido, tem-se evoluído para o desenvolvimento e aplicação de acções *a priori* que têm como objectivo a prevenção dos acidentes e das vítimas e baseiam-se num conhecimento pré-existente das causas da sinistralidade. Destas acções destacam-se as mais recentes – a auditoria de segurança rodoviária e a avaliação do impacte da segurança rodoviária. As Auditorias de Segurança Rodoviária (ASR) analisam a SR da infra-estrutura viária em fase de projecto, considerando essencialmente, as suas características geométricas. Para montante da fase de projecto foi desenvolvido o conceito de Avaliação do Impacte da Segurança Rodoviária (AISR). Esta acção *a priori* tem como objectivo incorporar explicitamente a SR no planeamento e no processo de decisão. Esta acção consiste na estimação de um número de acidentes para um ano horizonte com base na aplicação de modelos de previsão de acidentes. A AISR pode ter dois tipos de aplicação: avaliar políticas de SR na rede ou avaliar o impacte na SR de diferentes soluções planeadas para um determinado cenário. No primeiro caso, a AISR avalia as consequências na sinistralidade das acções a implementar no âmbito da SR e assim analisa-se a capacidade de se atingirem determinadas metas e objectivos. Pode ainda ser utilizada

para definir metas realísticas quanto ao número de acidentes a diminuir. No segundo caso, a AISR é aplicada para apoio à decisão da solução a implementar.

Saliente-se ainda o facto de nenhuma das acções acima referidas serem específicas do meio urbano. Pese embora o facto de ser neste meio que se concentra um número elevado de acidentes e de vítimas (no caso de Portugal, em 2008, cerca de 71% dos acidentes com vítimas e 47% das vítimas mortais ocorreram em meio urbano segundo o último relatório anual (ANSR, 2008)), não existe disponível, para a zona urbana, um número significativo de ferramentas devido às suas características específicas que dificultam o desenvolvimento e aplicação de modelos.

A importância de considerar, na fase de planeamento, um conjunto de critérios para garantir a sustentabilidade do sistema tem sido assumida em diversas áreas dos transportes nomeadamente nas questões ambientais asseguradas, por exemplo, pela Análise de Impacte Ambiental. Contudo, ao considerar na fase de planeamento determinados critérios pode correr-se o risco de preterir outros igualmente importantes e muitas vezes com consequências inter-relacionadas. Nesse sentido, é importante garantir que os critérios de escolha dos impactes a serem avaliados sejam ajustados a factores identificados como consequências directas ou indirectas do sistema de transportes e sejam tidas em conta as gravidades social e económica dessas consequências. Assim sendo, é fundamental considerar a sinistralidade como um critério preponderante a avaliar na fase de planeamento, sendo para tal necessário a aplicação de ferramentas ajustadas a esta fase.

Embora o processo de planeamento seja realizado por técnicos especializados, a tomada de decisão, apoiada na informação obtido pelo processo de planeamento, é essencialmente da responsabilidade dos governantes. São também estes que, em geral, definem os objectivos, tendo em consideração os interesses dos diversos intervenientes que são muitas vezes divergentes. Por esse facto, compete ainda aos governantes coordenar os diferentes interesses e gerir eventuais conflitos que possam advir da tomada de decisão. Nesse sentido, é importante estabelecer prioridades e critérios justificados e clarificados com base no processo de planeamento.

Existem vários modelos conceptuais de tomada de decisão definidos por diversos autores. No caso de Portugal e para o meio urbano, o modelo que ainda prevalece é o do Plano Director Municipal. Neste modelo as decisões são baseadas em interpretações do plano

director, que estabelece as regras e o desempenho desejável. Este modelo é um modelo tradicional no planeamento de transportes e da ocupação do solo. A desvantagem apontada para este modelo é a incapacidade de se adaptar às rápidas alterações verificadas ao nível do ambiente económico, social e tecnológico.

A modelação dos transportes é uma ferramenta frequentemente utilizada nos diversos modelos conceptuais de tomada de decisão incluindo o plano director. O objectivo de um modelo de transportes é representar as diversas componentes da infra-estrutura de transportes e os fluxos de tráfego de forma a reproduzir as condições do sistema viário. Para a aplicação deste modelo é necessário definir e caracterizar áreas de geração e atracção de viagens através de informação relativa ao uso do solo e a indicadores socioeconómicos. O produto final do modelo dos transportes corresponde à reprodução do nível de desempenho da rede para um conjunto seleccionado de critérios de avaliação, caracterizado pelos volumes de tráfego, velocidades, tempos de viagens, percursos, etc.

Assim, é neste contexto que se propõe a aplicação de uma ferramenta de análise da SR da rede viária do meio urbano em fase de planeamento. Propõe-se ainda que essa ferramenta seja integrada no processo de planeamento e de tomada de decisão com uma estrutura simples, adaptada à realidade do país, tal como mostra o organigrama da Figura 6.1.

O processo de planeamento representado pelo organigrama inicia-se com a definição do Plano Director Municipal (PDM) onde se identificam o tipo de ocupação do solo e a hierarquia viária da rede urbana. Esta informação é utilizada simultaneamente no modelo de transportes, a partir do qual se caracteriza o tráfego, e na ferramenta de análise da SR que determina o nível de sinistralidade do cenário. O *output* do modelo de transportes é também integrado na análise da SR. O resultado obtido pela ferramenta de análise da SR na rede quantifica o impacte na SR que, em conjunto com outros impactes avaliados (poluição, desempenho do tráfego, etc.), indica qual das soluções analisadas é a melhor, podendo qualquer delas ser “otimizada” repetindo-se o processo até à solução final a implementar.

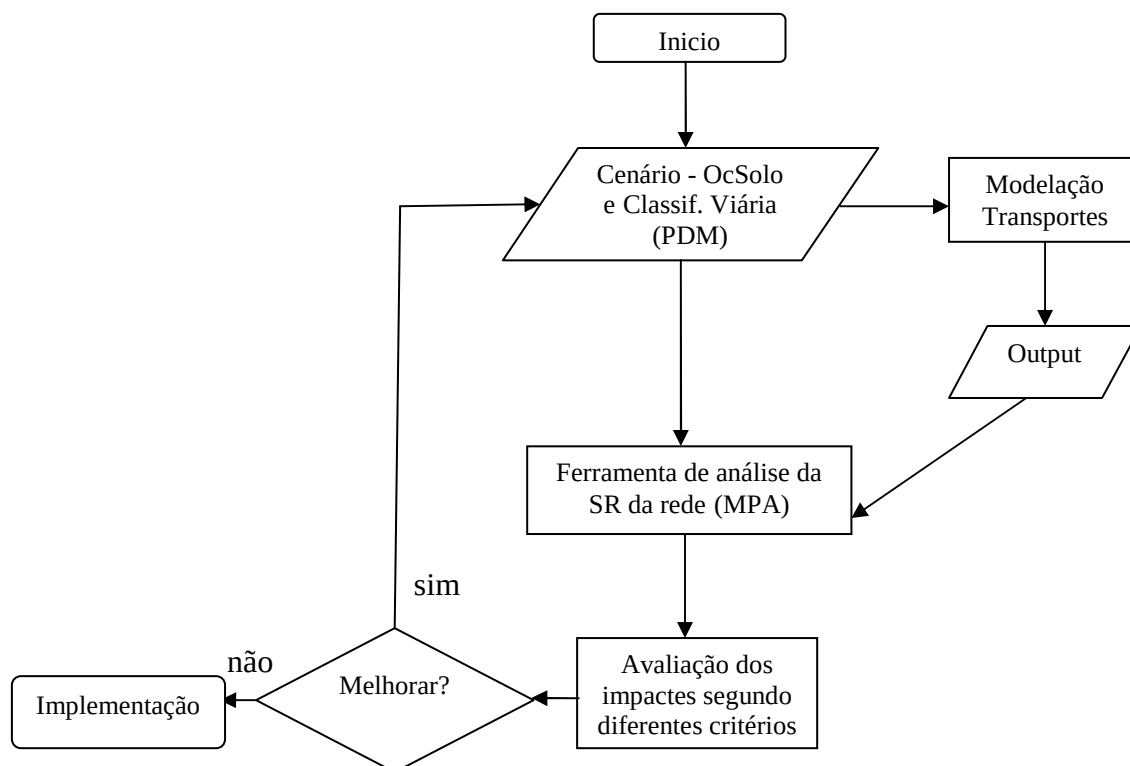


Figura 6.1. - Estrutura proposta para o processo de planeamento da SR em meio urbano

A ferramenta de análise da SR da rede proposta no processo de planeamento descrito no organigrama da Figura 6.1, compreende um modelo de previsão de acidentes. Embora exista já um conjunto de modelos matemáticos desenvolvidos e aplicados na modelação dos acidentes, a maior parte deles baseia-se na caracterização geométrica e funcional da rede, informação que ainda não está disponível na fase de planeamento. Nesta fase os elementos são ainda pouco precisos podendo, por isso, a informação obtida pelo modelo de transportes ser um elemento adicional importante no desenvolvimento dos modelos de previsão de acidentes (MPA).

Os MPA recentemente aplicados são, em geral, modelos lineares generalizados constituídos por uma componente aleatória normalmente com distribuição probabilística Poisson ou Binomial Negativa (BN). A resposta deste modelo corresponde a um valor de frequência de acidentes eventualmente decomposto por tipo ou gravidade do acidente. A aplicação destes modelos aos acidentes tem sido bastante aprofundada havendo já um conjunto de recomendações para a aplicação destes modelos. Estes modelos são, no

entanto, na maior parte das vezes desenvolvidos para explicar as ocorrências num determinado período de tempo em função de variáveis que caracterizam a infra-estrutura geométrica e de tráfego.

Assim, neste trabalho, centrado no objectivo de desenvolver uma ferramenta de apoio ao processo de planeamento do sistema de transportes do meio urbano, considerou-se a análise e aplicação dos modelos lineares generalizados (MLG), designados, neste trabalho, pelo tipo de variável de resposta, como modelos de resposta quantitativa. Para além destes modelos, consideraram-se ainda os modelos de resposta qualitativa, menos aplicados no âmbito da sinistralidade e cujo resultado corresponde a uma categoria indicativa de um determinado grau de sinistralidade. Existem vários tipos de modelos de resposta qualitativa cuja diferença reside, entre outras, na relação entre as categorias consideradas.

Considerou-se para a aplicação destes modelos o nível mesoscópico que compreende a rede principal do meio urbano dividida em arcos e nós. Este nível possibilita, por um lado, a análise global de um cenário do meio urbano e, por outro lado, localizar os pontos da rede de maior risco sendo assim mais fácil propor alterações ao cenário ou a implementação de medidas mitigadoras. Dadas as diferenças estruturais e funcionais dos arcos e dos nós consideraram-se modelos distintos para cada uma destas entidades viárias. As variáveis independentes incluídas nos modelos para caracterizar o cenário foram seleccionadas tendo em conta os factores decididos no planeamento do meio urbano na fase de elaboração do PDM – ocupação do solo e classificação hierárquica da rede. Estas variáveis, de uma forma indirecta, caracterizam a rede em termos funcionais e geométricos bem como a envolvente. Considerou-se ainda, à semelhança dos modelos habitualmente aplicados à modelação dos acidentes, o volume de tráfego mais precisamente o tráfego médio diário anual pois, em diversos estudos realizados, esta variável tem-se revelado como a variável de exposição mais relevante para explicar a ocorrência de acidentes. Considerou-se ainda a tendência temporal para captar factores difíceis de quantificar e que influenciam globalmente a frequência dos acidentes (educação, tempo meteorológico, programas de prevenção, etc.). No caso dos modelos dos arcos, incluiu-se ainda o comprimento e a densidade de intersecções com vias de acesso.

Para a aplicação dos modelos desenvolvidos neste trabalho, considerou-se o caso do concelho do Porto com os dados de sinistralidade relativos ao total dos acidentes (acidentes

só com danos materiais e acidentes com vítimas) ocorridos entre 2001 e 2005 (cinco anos). A caracterização da rede foi obtida com base na informação do PDM de 2006. Os valores do tráfego foram obtidos a partir dos resultados de um modelo de afectação do tráfego designado de SATURN.

Ao longo do processo de elaboração da base de dados, várias dificuldades foram encontradas. Essas dificuldades centraram-se essencialmente na obtenção dos dados e na qualidade destes e surgem desde o momento do registo até à localização dos acidentes num Sistema de Informação Geográfica (SIG). Também os valores do tráfego resultante da aplicação do modelo de afectação não corresponderam ao desejado, quer por não serem compatíveis com o mesmo período de tempo da base de dados dos acidentes quer por a rede do modelo de transportes apresentar diferenças relativamente à utilizada no modelo de acidentes, resultantes de uma classificação viária diferente. Contudo, estas incompatibilidades entre o modelo de transportes e o modelo de acidentes resultaram do facto destes modelos terem sido aplicados em momentos e com objectivos diferentes. Assim sendo, e tal como se sugere na estrutura do processo de planeamento apresentada na Figura 6.1, se os dois modelos forem aplicados de uma forma coordenada estas dificuldades serão ultrapassadas.

Para a primeira aplicação dos modelos de resposta quantitativa verificou-se ser a distribuição probabilística Binomial Negativa a mais ajustada para representar a componente aleatória.

Várias variantes ao modelo BN foram ainda consideradas com o objectivo de melhorar o ajuste aos dados. Estas variantes ao modelo BN compreenderam a análise da heteroscedasticidade, a função Translog, a aleatoriedade dos parâmetros das variáveis independentes (modelos de parâmetros aleatórios) e a existência de dois estados de geração de acidentes que inflaciona o número de zeros (modelos zero-inflacionado). De facto, em alguns desses modelos, e analisando os valores das medidas de avaliação do ajuste, verificou-se um melhor ajuste estatístico. Contudo, é de salientar que os valores estimados para os parâmetros das variáveis independentes consideradas não alteraram de forma relevante o sentido do efeito na variável de resposta (isto é, o sinal do valor estimado) e nos valores absolutos também não se verificaram alterações significativas.

Para a aplicação dos modelos de resposta qualitativa foi necessário definir as categorias da resposta do modelo. Foram analisadas várias hipóteses baseadas em intervalos de frequência de acidentes tendo-se optado por considerar uma classe de três categorias em que a primeira categoria corresponde a um grau de sinistralidade baixo (zero, um ou dois acidentes) de tal forma que traduza uma eventual ocorrência de acidentes resultante de condições pontuais e não devido a características dos locais. Para delimitar as duas categorias seguintes, de forma a que estas correspondam a graus distintos e ordenados por ordem crescente de gravidade, considerou-se o número de acidentes correspondente a 10% dos locais com maior número de acidentes (nove e cinco acidentes para os modelos dos arcos e nós respectivamente).

Para esta classe de categorias aplicaram-se dois modelos qualitativos – modelo ordenado e modelo multinomial. O primeiro estabelece uma ordenação das categorias e é estruturado por uma regressão latente que relaciona as variáveis com a variável de resposta y^* . Existe, portanto, uma relação entre as categorias que consequentemente não são independentes entre si, sendo os valores dos parâmetros estimados os mesmos, qualquer que seja a categoria, determinados pela única regressão estabelecida. No modelo multinomial (MM), a variável de resposta não obedece a uma ordenação sendo, por isso, estruturado por uma função “utilitária” para cada categoria que representa as relações entre as variáveis independentes e a variável de resposta, ou seja, as categorias são independentes entre si.

Das aplicações efectuadas, quer para a base de dados dos arcos quer para a base de dados dos nós, as medidas de avaliação do ajuste indicam ser preferível o MM. De facto, contrariando à partida a ordenação subjacente às três categorias consideradas, verificou-se que no MM os valores estimados e as significâncias estatísticas são diferentes nas três categorias e os efeitos marginais variam de forma distinta do modelo ordenado (MO). Estas diferenças foram mais significativas nos modelos aplicados aos arcos. Estes factos contrariam a hipótese de considerar uma única regressão para as categorias tal como o MO estabelece.

Analisando os dois tipos de abordagem aplicados – modelo de resposta quantitativa e de resposta qualitativa, verifica-se que as mesmas são distintas quer no tipo de resposta obtida quer na forma como se definem as relações entre a variável de resposta e as variáveis independentes. Nos modelos de resposta quantitativa, o resultado obtido corresponde a um valor estimado de frequência de acidentes para um determinado ano horizonte enquanto que, nos modelos de resposta qualitativa, o resultado final compreende uma probabilidade

associada à ocorrência de cada categoria (sendo que a cada uma destas corresponde um intervalo de frequência de acidentes). Este último resultado, por tomar a forma de intervalo, é mais abrangente quer por assumir de certa forma uma margem de incerteza que traduz a dificuldade em desenvolver um modelo perfeito, quer por se compatibilizar melhor com a forma generalista que caracteriza a fase de planeamento. A classificação de um cenário em fase de planeamento por um grau de gravidade é uma forma de considerar a sinistralidade como critério a analisar no planeamento do sistema de transportes e para informar os decisores das consequências de determinada solução de intervenção. Acrescente-se ainda o facto de que, das duas abordagens analisadas, os modelos de resposta qualitativa são mais flexíveis por considerarem os efeitos na variável de resposta distintos em função da categoria, em particular, o MM que permite que se estabeleçam relações entre as variáveis independentes e as diferentes categorias de uma forma independente. Nesse sentido, estes modelos acrescentam informação relativa ao comportamento das diferentes entidades viárias no âmbito da sinistralidade.

Assim, considerando as características do modelo de resposta qualitativa acima referidas e tendo em conta o objectivo e condições de aplicação de uma ferramenta de análise da segurança da rede em fase de planeamento, entende-se que o modelo de resposta qualitativa se enquadra melhor.

6.2 Desenvolvimento do estudo

Para fundamentar a aplicabilidade dos modelos desenvolvidos neste trabalho será necessário validá-los quanto à capacidade de previsão, isto é, com base nas relações entre a rede e os acidentes determinadas para o período de tempo entre 2001 e 2005 verificar se o modelo consegue prever as ocorrências para um período de tempo distinto com uma aproximação aceitável. Para isso será necessário reunir uma nova base de dados e assim validar os modelos apresentados neste trabalho.

É importante também validar a aplicação dos modelos a outros meios urbanos nomeadamente na significância estatísticas das variáveis consideradas.

Estes dois pontos são ainda mais importantes no caso dos modelos de resposta qualitativa uma vez que são de aplicação recente na área da sinistralidade, sendo necessário aprofundar a respectiva aplicabilidade a outras bases de dados.

Os vários modelos desenvolvidos e aplicados têm em comum as variáveis independentes consideradas. Estas variáveis foram seleccionadas tendo em conta o contexto de aplicação e o objectivo final, mas também a disponibilidade dos dados na fase de planeamento. Embora os resultados tenham, em geral, validado a escolha destas variáveis, julga-se ser importante incluir variáveis que caracterizem os vários modos de transporte existente no meio urbano, em especial os movimentos pedonais. Futuros trabalhos de investigação deverão assim centrar-se no desenvolvimento de modelos que caracterizem a diversidade do sistema de transportes em meio urbano. No caso particular dos modelos desenvolvidos para os nós, a qualidade do ajuste verificou-se ser inferior à obtida para os modelos dos arcos. De facto, os nós são caracterizadas por vários elementos que no seu conjunto podem interferir significativamente na ocorrência dos acidentes. De forma a caracterizar melhor os nós, melhorando assim o ajuste à base de dados, deverão ser analisadas outras possíveis variáveis a incluir no modelo, nomeadamente variáveis que caracterizem o nó de forma a distinguir diferentes tipologias.

Apesar da estrutura proposta para o processo de planeamento (ver Figura 6.1) ser uma estrutura simples, adaptada ao modelo conceptual de tomada de decisão habitualmente utilizado em Portugal para o meio urbano, será importante aprofundar e caracterizar melhor a forma de integrar uma ferramenta de análise da sinistralidade. Esta estrutura não será imutável dado que no futuro vários critérios irão adquirir substancial importância no processo de planeamento, nomeadamente os factores ambientais (poluição sonora, poluição atmosférica, etc.). Nesse sentido, é importante adaptar a estrutura do processo, de forma a não excluir a segurança rodoviária como um dos critérios a ponderar.

É indiscutível a importância que a SR assumiu nas últimas décadas. Os esforços desenvolvidos na área da investigação têm sido bastante frutíferos, comprovados pela diminuição do número de acidentes mas sobretudo do número de vítimas, pese embora o aumento do uso do transporte rodoviário individual. A actuação para a SR iniciou-se com medidas para eliminar as ocorrências estando actualmente, nos países desenvolvidos, a evoluir para a prevenção dos acidentes através, por exemplo, das ASR. Na sequência desta evolução este é o momento ideal para introduzir o conceito da SR na fase de planeamento.

6.3 Referências bibliográficas

ANSR (2008). *Sinistralidade rodoviária. Elementos estatísticos de 2008*. Lisboa: Observatório de Segurança Rodoviária, ANSR.

ANEXOS

ANEXO 1

GLOSSÁRIO DAS ABREVIATURAS MAIS UTILIZADAS

Abreviatura	Designação	
	Português	Inglês
ACB	Análise custo-benefício	-
AIA	Análise do Impacte Ambiental	-
AIC	Crítério de Informação de Akaike	<i>Akaike's Information Criterion</i>
AISR	Avaliação do Impacte da Segurança Rodoviária	-
AMC	Análise multi-critério	-
ANSR	Autoridade Nacional de Segurança Rodoviária	-
ASR	Auditoria de Segurança Rodoviária	-
BEAV	Boletim Estatístico de Acidentes de Viação	-
BN	Binomial Negativa	-
CARE	Base de dados de acidentes da UE	<i>Community database on accidents on the roads in Europe</i>
CMP	Câmara Municipal do Porto	-
CURE	-	<i>Cumulative Residuals</i>
ENSR	Estratégia Nacional de Segurança Rodoviária	-
ERF	-	<i>European Union Road Federation</i>
EUA	Estados Unidos da América	-
FIE	Função integral empírica	-
GEE	-	<i>Generalized Estimating Equations</i>
ASRR	Análise da segurança da rede rodoviária	-

Abreviatura	Designação	
	Português	Inglês
ID	-	<i>Integrate-Differentiate</i>
IHSD	-	<i>Interactive Highway Safety Design Model</i>
ISR	Inspecção de segurança rodoviária	-
LNEC	Laboratório Nacional de Engenharia Civil	-
MLG	Modelos lineares generalizados	-
MM	Modelo multinomial	<i>Multinomial model</i>
MML	Modelo multinomial logit	-
MO	Modelo ordenado	<i>Ordered model</i>
MOP	Modelo ordenado probit	-
MPA	Modelo de previsão de acidentes	
O-D	Origem-Destino	-
OMS	Organização Mundial de Saúde	-
PA	Parâmetros aleatórios	-
PDM	Plano Director Municipal	-
PET	Plano Estratégico de Transportes	-
PLANSAFE	-	<i>Planning Level Safety Prediction Model</i>
PNPR	Plano Nacional de Prevenção Rodoviária	-
RU	Reino Unido	-
SAFENET	-	<i>Software for Accident Frequency Estimation for Networks</i>

Abreviatura	Designação	
	Português	Inglês
SATURN	-	<i>Simulation and Assignment of Traffic to Urban Roads Networks</i>
SIG	Sistema de Informação Geográfica	-
SR	Segurança Rodoviária	<i>Road Safety</i>
SWOV	-	<i>Institute for Road Safety Research</i>
TMDA	Tráfego Médio Diário Anual	-
UE	União Europeia	-
VAL	Valor actual líquido	-
VHP	Volume da hora de ponta	-
VTD	Volume Total Diário	-
ZAA	Zona de acumulação de acidentes	-
ZI	Zero-inflaccionado	-
ZIBN	Zero-inflaccionado Binomial Negativo	-
ZIP	Zero-inflaccionado Poisson	-

ANEXO 2

RESULTADOS DA APLICAÇÃO DOS MODELOS (LIMDEP 9.0)

ANEXO 2.I

**MODELO BN (ARCOS) – FORMAS FUNCIONAIS DAS VARIÁVEIS
INDEPENDENTES**

Análise das três possíveis funções (Gama, Potência, Exponencial) para a variável Comp

```
-->
NEGBIN; Lhs=ACIDTOT; Rhs=ONE, LTMDTOT, LLENGTH, LENGTH, DENSINT, ANO, OCPSOL2, OCP
...
, OCPSOL4, OCPSOL5, DISLOCAL$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:12:11PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -3999.608
| Number of parameters        11
| Info. Criterion: AIC =      4.05112
|   Finite Sample: AIC =      4.05119
| Info. Criterion: BIC =      4.08218
| Info. Criterion: HQIC =     4.06253
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3687829
| Chi squared                 4673.470
| Degrees of freedom          10
| Prob[ChiSq > value] =       .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4414.33253  RsqP= .5809
| G - squared = 4070.77698  RsqD= .5345
| Overdispersion tests: g=mu(i) : 10.141
| Overdispersion tests: g=mu(i)^2: 9.433
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.73329231	.24115099	-32.068	.0000	
LTMDTOT	.30535211	.01937964	15.756	.0000	9.28075747
LLENGTH	1.06924688	.03076740	34.753	.0000	5.27644440
LENGTH	-.00040719	.456863D-04	-8.913	.0000	313.182126
DENSINT	44.8272769	3.18337611	14.082	.0000	.00457041
ANO	-.05959849	.00940075	-6.340	.0000	3.00000000
OCPSOL2	-.26938103	.03623659	-7.434	.0000	.22222222
OCPSOL3	.44359978	.06763430	6.559	.0000	.03030303
OCPSOL4	-.26663475	.05732519	-4.651	.0000	.05808081
OCPSOL5	.23140624	.03723293	6.215	.0000	.14141414
DISLOCAL	-.19438002	.03026333	-6.423	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:12:11PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        21
| Log likelihood function     -3636.318
| Number of parameters        12
| Info. Criterion: AIC =      3.68517
|   Finite Sample: AIC =      3.68525
| Info. Criterion: BIC =      3.71905
| Info. Criterion: HQIC =     3.69762
| Restricted log likelihood    -3999.608
+-----+
```

```

| McFadden Pseudo R-squared      .0908316
| Chi squared                    726.5813
| Degrees of freedom              1
| Prob[ChiSqd > value] =         .0000000
| NegBin form 2; Psi(i) = theta

```

```

+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
+-----+-----+-----+-----+-----+
| Constant | -8.21364177 | .37039447 | -22.175 | .0000 | 9.28075747
| LTMDTOT | .35616950 | .02933574 | 12.141 | .0000 | 5.27644440
| LLENGTH | 1.08044601 | .04799869 | 22.510 | .0000 | 313.182126
| LENGTH | -.00042618 | .00011140 | -3.826 | .0001 | .00457041
| DENSINT | 34.2945939 | 4.53445033 | 7.563 | .0000 | 3.00000000
| ANO | -.05693449 | .01562267 | -3.644 | .0003 | .22222222
| OCPSOL2 | -.31104010 | .05472067 | -5.684 | .0000 | .03030303
| OCPSOL3 | .42152396 | .13843389 | 3.045 | .0023 | .05808081
| OCPSOL4 | -.33226682 | .09699610 | -3.426 | .0006 | .14141414
| OCPSOL5 | .23778493 | .07432334 | 3.199 | .0014 | .49747475
| DISLOCAL | -.18334677 | .04938654 | -3.712 | .0002 |
+-----+-----+-----+-----+-----+
| Dispersion parameter for count data model
| Alpha | .42197805 | .02843907 | 14.838 | .0000

```

```

-->
NEGBIN; Lhs=ACIDTOT; Rhs=ONE, LTMDTOT, LLENGTH, DENSINT, ANO, OCPSOL2, OCPSOL3
, OCPSOL4, OCPSOL5, DISLOCAL$

```

```

+-----+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:17:15PM.
| Dependent variable      ACIDTOT
| Weighting variable      None
| Number of observations   1980
| Iterations completed    7
| Log likelihood function  -4043.441
| Number of parameters    10
| Info. Criterion: AIC =   4.09438
|   Finite Sample: AIC =   4.09444
| Info. Criterion: BIC =   4.12262
| Info. Criterion: HQIC =  4.10476
| Restricted log likelihood -6336.343
| McFadden Pseudo R-squared .3618652
| Chi squared             4585.805
| Degrees of freedom       9
| Prob[ChiSqd > value] =   .0000000
+-----+-----+

```

```

+-----+-----+
| Poisson Regression
| Chi- squared = 4354.45312 RsqP= .5866
| G - squared = 4158.44284 RsqD= .5244
| Overdispersion tests: g=mu(i) : 10.319
| Overdispersion tests: g=mu(i)^2: 10.270
+-----+-----+

```

```

+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
+-----+-----+-----+-----+-----+
| Constant | -6.50884674 | .19775825 | -32.913 | .0000 | 9.28075747
| LTMDTOT | .29704715 | .01941162 | 15.303 | .0000 | 5.27644440
| LLENGTH | .83506373 | .01568882 | 53.227 | .0000 | 313.182126
| DENSINT | 50.8228470 | 2.98282184 | 17.039 | .0000 | .00457041
| ANO | -.05930893 | .00940289 | -6.308 | .0000 | 3.00000000
| OCPSOL2 | -.31292194 | .03611472 | -8.665 | .0000 | .22222222
| OCPSOL3 | .44724187 | .06752801 | 6.623 | .0000 | .03030303
| OCPSOL4 | -.25631823 | .05733330 | -4.471 | .0000 | .05808081
| OCPSOL5 | .13916193 | .03630462 | 3.833 | .0001 | .14141414
| DISLOCAL | -.16262804 | .03025423 | -5.375 | .0000 | .49747475

```

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:17:16PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       1980
| Iterations completed        19
| Log likelihood function      -3646.012
| Number of parameters         11
| Info. Criterion: AIC =      3.69395
|   Finite Sample: AIC =      3.69402
| Info. Criterion: BIC =      3.72501
| Info. Criterion:HQIC =      3.70536
| Restricted log likelihood    -4043.441
| McFadden Pseudo R-squared   .0982899
| Chi squared                  794.8591
| Degrees of freedom           1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.31780925	.32664391	-22.403	.0000	
LTMDTOT	.34359618	.02958332	11.615	.0000	9.28075747
LENGTH	.91033691	.02872089	31.696	.0000	5.27644440
DENSINT	37.6348042	4.51520898	8.335	.0000	.00457041
ANO	-.05639366	.01567479	-3.598	.0003	3.00000000
OCPSOL2	-.33633805	.05469448	-6.149	.0000	.22222222
OCPSOL3	.40339596	.14160945	2.849	.0044	.03030303
OCPSOL4	-.31476027	.09637287	-3.266	.0011	.05808081
OCPSOL5	.20663985	.07213841	2.864	.0042	.14141414
DISLOCAL	-.16936993	.04927230	-3.437	.0006	.49747475
-----+Dispersion parameter for count data model					
Alpha	.43664021	.02949054	14.806	.0000	

--> **NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LENGTH,DENSINT,ANO,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5,DISLOCAL\$**

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:18:15PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       1980
| Iterations completed        7
| Log likelihood function      -4821.958
| Number of parameters         10
| Info. Criterion: AIC =      4.88077
|   Finite Sample: AIC =      4.88082
| Info. Criterion: BIC =      4.90900
| Info. Criterion:HQIC =      4.89114
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .2389998
| Chi squared                  3028.770
| Degrees of freedom           9
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 5888.66080  RsqP= .4409
+-----+

```

```
| G - squared = 5715.47749 RsqD= .3464 |
| Overdispersion tests: g=mu(i) : 12.686 |
| Overdispersion tests: g=mu(i)^2: 14.848 |
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-2.75940065	.18917940	-14.586	.0000	
LTMDTOT	.35821090	.01934221	18.520	.0000	9.28075747
LENGTH	.00092470	.193110D-04	47.884	.0000	313.182126
DENSINT	56.2157717	2.34900504	23.932	.0000	.00457041
ANO	-.06117519	.00938777	-6.516	.0000	3.00000000
OCPSOL2	-.36380225	.03674703	-9.900	.0000	.22222222
OCPSOL3	.46319091	.06737430	6.875	.0000	.03030303
OCPSOL4	-.13991399	.05732428	-2.441	.0147	.05808081
OCPSOL5	-.06035489	.03813383	-1.583	.1135	.14141414
DISLOCAL	-.03933434	.03032901	-1.297	.1947	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:18:16PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        20
| Log likelihood function     -3869.602
| Number of parameters        11
| Info. Criterion: AIC =      3.91980
|   Finite Sample: AIC =      3.91987
| Info. Criterion: BIC =      3.95086
| Info. Criterion:HQIC =      3.93121
| Restricted log likelihood    -4821.958
| McFadden Pseudo R-squared   .1975040
| Chi squared                 1904.712
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.12908138	.33591523	-9.315	.0000	
LTMDTOT	.35074774	.03481889	10.073	.0000	9.28075747
LENGTH	.00194371	.691206D-04	28.121	.0000	313.182126
DENSINT	57.4344641	5.16277388	11.125	.0000	.00457041
ANO	-.05635900	.01851805	-3.043	.0023	3.00000000
OCPSOL2	-.33140698	.06339538	-5.228	.0000	.22222222
OCPSOL3	.39739691	.18154141	2.189	.0286	.03030303
OCPSOL4	-.16084372	.11322517	-1.421	.1554	.05808081
OCPSOL5	.11260767	.08263123	1.363	.1730	.14141414
DISLOCAL	-.11399969	.05713550	-1.995	.0460	.49747475
-----Dispersion parameter for count data model					
Alpha	.73682530	.04134328	17.822	.0000	

**Análise das três possíveis funções (Gama, Potência, Exponencial) para a variável
DenInt**

```
-->
NEGBIN; Lhs=ACIDTOT; Rhs=ONE, LTMDTOT, LLENGTH, DENSINT, LDENINT, ANO, OCPSOL2, OC
...
, OCPSOL4, OCPSOL5, DISLOCAL$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:26:17PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4040.700
| Number of parameters        11
| Info. Criterion: AIC =      4.09263
|   Finite Sample: AIC =      4.09269
| Info. Criterion: BIC =      4.12369
| Info. Criterion: HQIC =     4.10404
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3622978
| Chi squared                 4591.286
| Degrees of freedom          10
| Prob[ChiSq > value] =       .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4337.57264  RsqP= .5882
| G - squared = 4152.96101  RsqD= .5251
| Overdispersion tests: g=mu(i) : 10.367
| Overdispersion tests: g=mu(i)^2: 10.192
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.55384416	.19817823	-33.070	.0000	
LTMDTOT	.29395756	.01943035	15.129	.0000	9.28075747
LLENGTH	.85897992	.01864983	46.058	.0000	5.27644440
DENSINT	53.9542310	3.24877625	16.608	.0000	.00457041
LDENINT	.02005678	.00853523	2.350	.0188	-3.08310935
ANO	-.05918586	.00940367	-6.294	.0000	3.00000000
OCPSOL2	-.32144519	.03635546	-8.842	.0000	.22222222
OCPSOL3	.47051831	.06825862	6.893	.0000	.03030303
OCPSOL4	-.24139176	.05767953	-4.185	.0000	.05808081
OCPSOL5	.13943557	.03630629	3.841	.0001	.14141414
DISLOCAL	-.16802470	.03035008	-5.536	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:26:18PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        22
| Log likelihood function     -3645.850
| Number of parameters        12
| Info. Criterion: AIC =      3.69480
|   Finite Sample: AIC =      3.69488
| Info. Criterion: BIC =      3.72868
| Info. Criterion: HQIC =     3.70725
+-----+
```

Restricted log likelihood		-4040.700			
McFadden Pseudo R-squared		.0977184			
Chi squared		789.7012			
Degrees of freedom		1			
Prob[ChiSqd > value] =		.0000000			
NegBin form 2; Psi(i) = theta					

+-----+-----+-----+-----+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

Constant	-7.34681847	.33660270	-21.826	.0000	
LTMdTOT	.34244838	.02962946	11.558	.0000	9.28075747
LLENGTH	.92117423	.03599760	25.590	.0000	5.27644440
DENSINT	39.1617871	5.16992805	7.575	.0000	.00457041
LDENINT	.00797660	.01375209	.580	.5619	-3.08310935
ANO	-.05633743	.01573705	-3.580	.0003	3.00000000
OCPSOL2	-.33681383	.05496032	-6.128	.0000	.22222222
OCPSOL3	.41028567	.14201130	2.889	.0039	.03030303
OCPSOL4	-.30812605	.09682335	-3.182	.0015	.05808081
OCPSOL5	.20640602	.07238035	2.852	.0043	.14141414
DISLOCAL	-.17012813	.04943748	-3.441	.0006	.49747475

+-----+Dispersion parameter for count data model					
Alpha	.43626589	.02945333	14.812	.0000	

-->

```
NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,ANO,OCPSOL2,OCPSOL3
,OCPSOL4,OCPSOL5,DISLOCAL$
```

```

Poisson Regression
Maximum Likelihood Estimates
Model estimated: Mar 10, 2010 at 05:27:25PM.
Dependent variable          ACIDTOT
Weighting variable          None
Number of observations       1980
Iterations completed         7
Log likelihood function      -4043.441
Number of parameters         10
Info. Criterion: AIC =      4.09438
Finite Sample: AIC =      4.09444
Info. Criterion: BIC =      4.12262
Info. Criterion:HQIC =      4.10476
Restricted log likelihood    -6336.343
McFadden Pseudo R-squared   .3618652
Chi squared                  4585.805
Degrees of freedom           9
Prob[ChiSqd > value] =      .0000000

```

```

+-----+
| Poisson Regression |
| Chi- squared = 4354.45312   RsqP=   .5866 |
| G - squared = 4158.44284   RsqD=   .5244 |
| Overdispersion tests: g=mu(i) : 10.319 |
| Overdispersion tests: g=mu(i)^2: 10.270 |
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.50884674	.19775825	-32.913	.0000	
LTMDTOT	.29704715	.01941162	15.303	.0000	9.28075747
LLENGTH	.83506373	.01568882	53.227	.0000	5.27644440
DENSINT	50.8228470	2.98282184	17.039	.0000	.00457041
ANO	-.05930893	.00940289	-6.308	.0000	3.00000000
OCPSOL2	-.31292194	.03611472	-8.665	.0000	2.22222222
OCPSOL3	.44724187	.06752801	6.623	.0000	.03030303

OCPSOL4	-.25631823	.05733330	-4.471	.0000	.05808081
OCPSOL5	.13916193	.03630462	3.833	.0001	.14141414
DISLOCAL	-.16262804	.03025423	-5.375	.0000	.49747475

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:27:25PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        19
| Log likelihood function     -3646.012
| Number of parameters        11
| Info. Criterion: AIC =      3.69395
|   Finite Sample: AIC =      3.69402
| Info. Criterion: BIC =      3.72501
| Info. Criterion:HQIC =      3.70536
| Restricted log likelihood   -4043.441
| McFadden Pseudo R-squared  .0982899
| Chi squared                 794.8591
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+
+-----+ +-----+ +-----+ +-----+ +-----+
|Variable| Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X|
+-----+ +-----+ +-----+ +-----+ +-----+
| Constant | -7.31780925 | .32664391 | -22.403 | .0000 | 9.28075747
| LTMDTOT  | .34359618   | .02958332 | 11.615  | .0000 | 5.27644440
| LLENGTH  | .91033691   | .02872089 | 31.696  | .0000 | 5.27644440
| DENSINT  | 37.6348042  | 4.51520898 | 8.335   | .0000 | .00457041
| ANO      | -.05639366  | .01567479 | -3.598  | .0003 | 3.00000000
| OCPSOL2  | -.33633805  | .05469448 | -6.149  | .0000 | .22222222
| OCPSOL3  | .40339596   | .14160945 | 2.849   | .0044 | .03030303
| OCPSOL4  | -.31476027  | .09637287 | -3.266  | .0011 | .05808081
| OCPSOL5  | .20663985   | .07213841 | 2.864   | .0042 | .14141414
| DISLOCAL | -.16936993  | .04927230 | -3.437  | .0006 | .49747475
+-----+ +-----+
| Dispersion parameter for count data model
| Alpha      | .43664021 | .02949054 | 14.806 | .0000

```

-->
NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LLENGTH,LDENiNT,ANO,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5,DISLOCAL\$

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:28:14PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4167.415
| Number of parameters        10
| Info. Criterion: AIC =      4.21961
|   Finite Sample: AIC =      4.21967
| Info. Criterion: BIC =      4.24785
| Info. Criterion:HQIC =      4.22998
| Restricted log likelihood   -6336.343
| McFadden Pseudo R-squared  .3422996
| Chi squared                 4337.856
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
+-----+

```

+-----+ Poisson Regression Chi-squared = 4588.41650 RsqP= .5644 G-squared = 4406.39133 RsqD= .4961 Overdispersion tests: g=mu(i) : 10.641 Overdispersion tests: g=mu(i)^2: 11.652 +-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.05713858	.19386420	-31.244	.0000	
LTMDTOT	.29798704	.01935417	15.397	.0000	9.28075747
LLENGTH	.77550625	.01793602	43.237	.0000	5.27644440
LDENINT	-.03559000	.00752161	-4.732	.0000	-3.08310935
ANO	-.05899458	.00940073	-6.276	.0000	3.00000000
OCPSOL2	-.33746071	.03648298	-9.250	.0000	.22222222
OCPSOL3	.28696005	.06734143	4.261	.0000	.03030303
OCPSOL4	-.32558644	.05750025	-5.662	.0000	.05808081
OCPSOL5	.16388277	.03635382	4.508	.0000	.14141414
DISLOCAL	-.07454756	.02993575	-2.490	.0128	.49747475

Normal exit from iterations. Exit status=0.

+-----+ Negative Binomial Regression Maximum Likelihood Estimates Model estimated: Mar 10, 2010 at 05:28:14PM. Dependent variable ACIDTOT Weighting variable None Number of observations 1980 Iterations completed 17 Log likelihood function -3670.623 Number of parameters 11 Info. Criterion: AIC = 3.71881 Finite Sample: AIC = 3.71888 Info. Criterion: BIC = 3.74987 Info. Criterion: HQIC = 3.73022 Restricted log likelihood -4167.415 McFadden Pseudo R-squared .1192088 Chi squared 993.5849 Degrees of freedom 1 Prob[ChiSqd > value] = .0000000 NegBin form 2; Psi(i) = theta +-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.17366557	.33572034	-21.368	.0000	
LTMDTOT	.34834909	.02958357	11.775	.0000	9.28075747
LLENGTH	.88458773	.03594882	24.607	.0000	5.27644440
LDENINT	-.04133293	.01277486	-3.235	.0012	-3.08310935
ANO	-.05825595	.01604736	-3.630	.0003	3.00000000
OCPSOL2	-.34171624	.05701184	-5.994	.0000	.22222222
OCPSOL3	.29534730	.15300763	1.930	.0536	.03030303
OCPSOL4	-.37373138	.10007504	-3.735	.0002	.05808081
OCPSOL5	.25929929	.07160950	3.621	.0003	.14141414
DISLOCAL	-.13663404	.05057213	-2.702	.0069	.49747475
-----Dispersion parameter for count data model					
Alpha	.48027648	.03090034	15.543	.0000	

**Análise das três possíveis funções (Gama, Potência, Exponencial) para a variável
 $TMDA_{Tot}$**

```
-->
NEGBIN; Lhs=ACIDTOT; Rhs=ONE, LTMDTOT, TMDATOT, LLENGTH, DENSINT, ANO, OCPSOL2, OC
...
, OCPSOL4, OCPSOL5, DISLOCAL$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:30:03PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       1980
| Iterations completed        7
| Log likelihood function      -4039.299
| Number of parameters         11
| Info. Criterion: AIC =      4.09121
|   Finite Sample: AIC =      4.09128
| Info. Criterion: BIC =      4.12227
| Info. Criterion: HQIC =     4.10262
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3625189
| Chi squared                  4594.088
| Degrees of freedom           10
| Prob[ChiSq > value] =       .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4373.65791  RsqP= .5848
| G - squared = 4150.15949  RsqD= .5254
| Overdispersion tests: g=mu(i) : 10.145
| Overdispersion tests: g=mu(i)^2: 10.111
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.38519065	.37174339	-19.866	.0000	
LTMDTOT	.40261351	.04243929	9.487	.0000	9.28075747
TMDATOT	-.723052D-05	.254993D-05	-2.836	.0046	15240.3849
LLENGTH	.83584959	.01567142	53.336	.0000	5.27644440
DENSINT	50.0388727	2.99805943	16.690	.0000	.00457041
ANO	-.05923864	.00940730	-6.297	.0000	3.00000000
OCPSOL2	-.29491747	.03663258	-8.051	.0000	.22222222
OCPSOL3	.45869480	.06766194	6.779	.0000	.03030303
OCPSOL4	-.24492299	.05745316	-4.263	.0000	.05808081
OCPSOL5	.13564896	.03633274	3.734	.0002	.14141414
DISLOCAL	-.17203204	.03038976	-5.661	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:30:04PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       1980
| Iterations completed        20
| Log likelihood function      -3643.601
| Number of parameters         12
| Info. Criterion: AIC =      3.69253
|   Finite Sample: AIC =      3.69261
| Info. Criterion: BIC =      3.72641
| Info. Criterion: HQIC =      3.70497
+-----+
```

Restricted log likelihood		-4039.299			
McFadden Pseudo R-squared		.0979622			
Chi squared		791.3973			
Degrees of freedom		1			
Prob[ChiSqd > value] =		.0000000			
NegBin form 2; Psi(i) = theta					

+-----+-----+-----+-----+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

Constant	-8.26220508	.47493621	-17.396	.0000	
LTMdTOT	.45756153	.05282908	8.661	.0000	9.28075747
TMDATOT	-.883074D-05	.374931D-05	-2.355	.0185	15240.3849
LLENGTH	.91390640	.02859793	31.957	.0000	5.27644440
DENSINT	36.7421640	4.49331127	8.177	.0000	.00457041
ANO	-.05667771	.01568129	-3.614	.0003	3.00000000
OCPSOL2	-.31527782	.05592776	-5.637	.0000	.22222222
OCPSOL3	.41213095	.14228344	2.897	.0038	.03030303
OCPSOL4	-.29487977	.09651418	-3.055	.0022	.05808081
OCPSOL5	.21339203	.07202620	2.963	.0030	.14141414
DISLOCAL	-.18283721	.04943630	-3.698	.0002	.49747475

+-----+Dispersion parameter for count data model					
Alpha	.43337794	.02939160	14.745	.0000	

```
-->
NEGBIN; Lhs=ACIDTOT; RhS=ONE, TMDATOT, LLENGTH, DENSINT, ANO, OCPSOL2, OCPSOL3
, OCPSOL4, OCPSOL5, DISLOCAL$
```

```

Poisson Regression
Maximum Likelihood Estimates
Model estimated: Mar 10, 2010 at 05:30:39PM.
Dependent variable          ACIDTOT
Weighting variable          None
Number of observations       1980
Iterations completed         7
Log likelihood function      -4091.654
Number of parameters         10
Info. Criterion: AIC =      4.14308
Finite Sample: AIC =      4.14314
Info. Criterion: BIC =      4.17132
Info. Criterion:HQIC =      4.15346
Restricted log likelihood    -6336.343
McFadden Pseudo R-squared   .3542563
Chi squared                  4489.380
Degrees of freedom           9
Prob[ChiSq > value] =      .0000000

```

```

+-----+
| Poisson Regression |
| Chi- squared = 4448.09251   RsqP=   .5777 |
| G - squared = 4254.86779   RsqD=   .5134 |
| Overdispersion tests: g=mu(i) : 10.533 |
| Overdispersion tests: g=mu(i)^2: 10.304 |
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-4.04136387	.10315715	-39.177	.0000	
TMDATOT	.144582D-04	.113432D-05	12.746	.0000	15240.3849
LLENGTH	.85114649	.01574765	54.049	.0000	5.27644440
DENSINT	52.9091670	2.98335633	17.735	.0000	.00457041
ANO	-.05549347	.00940531	-5.900	.0000	3.00000000
OCPSOL2	-.33791772	.03638084	-9.288	.0000	.22222222
OCPSOL3	.42579712	.06759720	6.299	.0000	.03030303
OCPSOL4	-.26896517	.05746979	-4.680	.0000	.05808081

```

OCPSOL5 |      .16678003      .03619989      4.607      .0000      .14141414
DISLOCAL|     -.19689875     .03036907     -6.484      .0000      .49747475

```

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:30:40PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        19
| Log likelihood function     -3675.015
| Number of parameters        11
| Info. Criterion: AIC =      3.72325
|   Finite Sample: AIC =      3.72331
| Info. Criterion: BIC =      3.75431
| Info. Criterion:HQIC =      3.73466
| Restricted log likelihood   -4091.654
| McFadden Pseudo R-squared  .1018266
| Chi squared                 833.2782
| Degrees of freedom          1
| Prob[ChiSqd > value] =     .00000000
| NegBin form 2; Psi(i) = theta
+-----+
+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+
| Constant | -4.43362893 | .17535001      | -25.284   | .0000    | 15240.3849
| TMDATOT  | .182135D-04 | .216239D-05    | 8.423     | .0000    | 5.27644440
| LLENGTH  | .91889250    | .02946945      | 31.181    | .0000    | .00457041
| DENSINT  | 39.5612651   | 4.60377683     | 8.593     | .0000    | 3.00000000
| ANO      | -.04990320   | .01568860      | -3.181    | .0015    | .22222222
| OCPSOL2  | -.36254303   | .05592996      | -6.482    | .0000    | .03030303
| OCPSOL3  | .39474423    | .14029311      | 2.814     | .0049    | .05808081
| OCPSOL4  | -.33066538   | .09656468      | -3.424    | .0006    | .14141414
| OCPSOL5  | .23003998    | .07409294      | 3.105     | .0019    | .49747475
| DISLOCAL | -.20822333   | .05056652      | -4.118    | .0000    |
+-----+-----+-----+-----+-----+
|-----+-----+-----+-----+-----+
| Dispersion parameter for count data model
| Alpha    | .46108618    | .03069300     | 15.023    | .0000
+-----+-----+-----+-----+-----+

```

```

-->
NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,ANO,OCPSOL2,OCPSOL3
,OCPSOL4,OCPSOL5,DISLOCAL$

```

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 05:31:13PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4043.441
| Number of parameters        10
| Info. Criterion: AIC =      4.09438
|   Finite Sample: AIC =      4.09444
| Info. Criterion: BIC =      4.12262
| Info. Criterion:HQIC =      4.10476
| Restricted log likelihood   -6336.343
| McFadden Pseudo R-squared  .3618652
| Chi squared                 4585.805
| Degrees of freedom          9
| Prob[ChiSqd > value] =     .00000000
+-----+

```

+-----+					
Poisson Regression					
Chi- squared = 4354.45312 RsqP= .5866					
G - squared = 4158.44284 RsqD= .5244					
Overdispersion tests: g=mu(i) : 10.319					
Overdispersion tests: g=mu(i)^2: 10.270					
+-----+					
+-----+					
Variable Coefficient Standard Error b/St.Er. P[Z >z] Mean of X					
+-----+					
Constant	-6.50884674	.19775825	-32.913	.0000	
LTMDTOT	.29704715	.01941162	15.303	.0000	9.28075747
LLENGTH	.83506373	.01568882	53.227	.0000	5.27644440
DENSINT	50.8228470	2.98282184	17.039	.0000	.00457041
ANO	-.05930893	.00940289	-6.308	.0000	3.00000000
OCPSOL2	-.31292194	.03611472	-8.665	.0000	.22222222
OCPSOL3	.44724187	.06752801	6.623	.0000	.03030303
OCPSOL4	-.25631823	.05733330	-4.471	.0000	.05808081
OCPSOL5	.13916193	.03630462	3.833	.0001	.14141414
DISLOCAL	-.16262804	.03025423	-5.375	.0000	.49747475

Normal exit from iterations. Exit status=0.

+-----+					
Negative Binomial Regression					
Maximum Likelihood Estimates					
Model estimated: Mar 10, 2010 at 05:31:13PM.					
Dependent variable ACIDTOT					
Weighting variable None					
Number of observations 1980					
Iterations completed 19					
Log likelihood function -3646.012					
Number of parameters 11					
Info. Criterion: AIC = 3.69395					
Finite Sample: AIC = 3.69402					
Info. Criterion: BIC = 3.72501					
Info. Criterion:HQIC = 3.70536					
Restricted log likelihood -4043.441					
McFadden Pseudo R-squared .0982899					
Chi squared 794.8591					
Degrees of freedom 1					
Prob[ChiSqd > value] = .00000000					
NegBin form 2; Psi(i) = theta					
+-----+					
+-----+					
Variable Coefficient Standard Error b/St.Er. P[Z >z] Mean of X					
+-----+					
Constant	-7.31780925	.32664391	-22.403	.0000	
LTMDTOT	.34359618	.02958332	11.615	.0000	9.28075747
LLENGTH	.91033691	.02872089	31.696	.0000	5.27644440
DENSINT	37.6348042	4.51520898	8.335	.0000	.00457041
ANO	-.05639366	.01567479	-3.598	.0003	3.00000000
OCPSOL2	-.33633805	.05469448	-6.149	.0000	.22222222
OCPSOL3	.40339596	.14160945	2.849	.0044	.03030303
OCPSOL4	-.31476027	.09637287	-3.266	.0011	.05808081
OCPSOL5	.20663985	.07213841	2.864	.0042	.14141414
DISLOCAL	-.16936993	.04927230	-3.437	.0006	.49747475
+-----+Dispersion parameter for count data model					
Alpha	.43664021	.02949054	14.806	.0000	

ANEXO 2.II

MODELO BN (ARCOS) – EQUAÇÃO (5.2)

-->

NEGBIN; Lhs=ACIDTOT; RhS=ONE, LTMDTOT, LLENGTH, DENSINT, ANO, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, DISLOCALS

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 08:15:12PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4043.441
| Number of parameters        10
| Info. Criterion: AIC =      4.09438
|   Finite Sample: AIC =      4.09444
| Info. Criterion: BIC =      4.12262
| Info. Criterion: HQIC =     4.10476
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3618652
| Chi squared                 4585.805
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi-squared = 4354.45312 RsqP= .5866
| G-squared = 4158.44284 RsqD= .5244
| Overdispersion tests: g=mu(i) : 10.319
| Overdispersion tests: g=mu(i)^2: 10.270
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.50884674	.19775825	-32.913	.0000	
LTMDTOT	.29704715	.01941162	15.303	.0000	9.28075747
LLENGTH	.83506373	.01568882	53.227	.0000	5.27644440
DENSINT	50.8228470	2.98282184	17.039	.0000	.00457041
ANO	-.05930893	.00940289	-6.308	.0000	3.00000000
OCPSOL2	-.31292194	.03611472	-8.665	.0000	.22222222
OCPSOL3	.44724187	.06752801	6.623	.0000	.03030303
OCPSOL4	-.25631823	.05733330	-4.471	.0000	.05808081
OCPSOL5	.13916193	.03630462	3.833	.0001	.14141414
DISLOCAL	-.16262804	.03025423	-5.375	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 08:15:13PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed       19
| Log likelihood function     -3646.012
| Number of parameters        11
| Info. Criterion: AIC =      3.69395
|   Finite Sample: AIC =      3.69402
| Info. Criterion: BIC =      3.72501
| Info. Criterion: HQIC =     3.70536
| Restricted log likelihood    -4043.441
| McFadden Pseudo R-squared   .0982899
| Chi squared                 794.8591
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+
```

+-----+ Variable	+-----+ Coefficient	+-----+ Standard Error	+-----+ b/St.Er.	+-----+ P[Z >z]	+-----+ Mean of X
Constant	-7.31780925	.32664391	-22.403	.0000	
LTMDTOT	.34359618	.02958332	11.615	.0000	9.28075747
LLENGTH	.91033691	.02872089	31.696	.0000	5.27644440
DENSINT	37.6348042	4.51520898	8.335	.0000	.00457041
ANO	-.05639366	.01567479	-3.598	.0003	3.00000000
OCPSOL2	-.33633805	.05469448	-6.149	.0000	.22222222
OCPSOL3	.40339596	.14160945	2.849	.0044	.03030303
OCPSOL4	-.31476027	.09637287	-3.266	.0011	.05808081
OCPSOL5	.20663985	.07213841	2.864	.0042	.14141414
DISLOCAL	-.16936993	.04927230	-3.437	.0006	.49747475
-----+Dispersion parameter for count data model					
Alpha	.43664021	.02949054	14.806	.0000	

ANEXO 2.III

MODELO BN (ARCOS) – TENDÊNCIA TEMPORAL


```
-->
NEGBIN; Lhs=ACIDTOT; RhS=ONE, LTMDTOT, LLENGTH, DENSINT, DUMMYA2, DUMMYA3, DUMMYA
...
, OCPSOL4, OCPSOL5, DISLOCAL$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:23:58PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4023.189
| Number of parameters        13
| Info. Criterion: AIC =      4.07696
|   Finite Sample: AIC =      4.07705
| Info. Criterion: BIC =      4.11367
| Info. Criterion: HQIC =     4.09044
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3650613
| Chi squared                 4626.308
| Degrees of freedom          12
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4278.95266 RsqP= .5937
| G - squared = 4117.93930 RsqD= .5291
| Overdispersion tests: g=mu(i) : 10.423
| Overdispersion tests: g=mu(i)^2: 10.464
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.57177419	.19947658	-32.945	.0000	
LTMDTOT	.29925451	.01946558	15.374	.0000	9.28075747
LLENGTH	.83478156	.01568705	53.215	.0000	5.27644440
DENSINT	50.8177172	2.98261529	17.038	.0000	.00457041
DUMMYA2	-.04796164	.04027362	-1.191	.2337	.20000000
DUMMYA3	-.28487559	.04264239	-6.681	.0000	.20000000
DUMMYA4	-.07763338	.04027041	-1.928	.0539	.20000000
DUMMYA5	-.28174305	.04213420	-6.687	.0000	.20000000
OCPSOL2	-.31304498	.03611628	-8.668	.0000	.22222222
OCPSOL3	.44739250	.06752648	6.625	.0000	.03030303
OCPSOL4	-.25642965	.05733307	-4.473	.0000	.05808081
OCPSOL5	.13906093	.03630431	3.830	.0001	.14141414
DISLOCAL	-.16120124	.03026956	-5.326	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:23:58PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        22
| Log likelihood function     -3637.560
| Number of parameters        14
| Info. Criterion: AIC =      3.68844
|   Finite Sample: AIC =      3.68855
| Info. Criterion: BIC =      3.72798
| Info. Criterion: HQIC =     3.70297
| Restricted log likelihood    -4023.189
| McFadden Pseudo R-squared   .0958517
| Chi squared                 771.2594
+-----+
```

Degrees of freedom 1					
Prob[ChiSqd > value] = .0000000					
NegBin form 2; Psi(i) = theta					
+-----+-----+-----+-----+-----+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
+-----+-----+-----+-----+-----+-----+					
Constant	-7.36940112	.32696838	-22.539	.0000	
LTMDTOT	.34545199	.02964557	11.653	.0000	9.28075747
LLENGTH	.90941639	.02870349	31.683	.0000	5.27644440
DENSINT	37.9585420	4.49259813	8.449	.0000	.00457041
DUMMYA2	-.03398695	.06503571	-.523	.6013	.20000000
DUMMYA3	-.30845423	.07192338	-4.289	.0000	.20000000
DUMMYA4	-.08918844	.06812697	-1.309	.1905	.20000000
DUMMYA5	-.25519907	.07039387	-3.625	.0003	.20000000
OCPSOL2	-.33547264	.05505277	-6.094	.0000	.22222222
OCPSOL3	.40959353	.13973788	2.931	.0034	.03030303
OCPSOL4	-.30856652	.09498303	-3.249	.0012	.05808081
OCPSOL5	.20806341	.07163022	2.905	.0037	.14141414
DISLOCAL	-.16770657	.04932468	-3.400	.0007	.49747475
-----+Dispersion parameter for count data model					
Alpha	.42749317	.02928731	14.597	.0000	

ANEXO 2.IV

ANÁLISE DA CORRELAÇÃO TEMPORAL (ARCOS)

Teste de razão de verossimilhança (Poch, M. and Mannering, F., 1996)

Modelo BN - Arcos

$$\chi^2 = -2*[-3646,012-(-765,1122-748,0281-662,6334-737,5068-710,7568)] = 43,9494$$
$$K_g(\text{graus de liberdade}) = (5*9-10) = 35$$

$$\text{Dist. } \chi^2 = 0,1427$$

$$(1-0,1427)*100 = 85,7\%$$

Poch, M. and Mannering, F. (1996). Negative Binomial Analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, **122**, 105-113.

ANEXO 2.V

MODELO BN AGREGADO NOS 5 ANOS (ARCOS)

--> NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5,DISLOCALS

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:29:58PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       396
| Iterations completed        6
| Log likelihood function      -1956.157
| Number of parameters         9
| Info. Criterion: AIC =      9.92504
|   Finite Sample: AIC =      9.92621
| Info. Criterion: BIC =     10.01552
| Info. Criterion:HQIC =      9.96088
| Restricted log likelihood     -4258.726
| McFadden Pseudo R-squared    .5406708
| Chi squared                  4605.138
| Degrees of freedom           8
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 2500.55290 RsqP= .7087
| G - squared = 2510.92879 RsqD= .6471
| Overdispersion tests: g=mu(i) : 8.231
| Overdispersion tests: g=mu(i)^2: 8.447
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-5.45678500	.20770414	-26.272	.0000	
LTMDTOT	.33984466	.02054742	16.540	.0000	9.29908834
LLENGTH	.82733875	.01567962	52.765	.0000	5.27644440
DENSINT	50.5931636	2.97832109	16.987	.0000	.00457041
OCPSOL2	-.31069675	.03612918	-8.600	.0000	.22222222
OCPSOL3	.45518508	.06749956	6.744	.0000	.03030303
OCPSOL4	-.25334076	.05732426	-4.419	.0000	.05808081
OCPSOL5	.14155730	.03626472	3.903	.0001	.14141414
DISLOCAL	-.13601299	.03052558	-4.456	.0000	.49747475

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:29:58PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       396
| Iterations completed        20
| Log likelihood function      -1258.915
| Number of parameters        10
| Info. Criterion: AIC =      6.40866
|   Finite Sample: AIC =      6.41011
| Info. Criterion: BIC =      6.50920
| Info. Criterion:HQIC =      6.44849
| Restricted log likelihood     -1956.157
| McFadden Pseudo R-squared    .3564345
| Chi squared                  1394.484
| Degrees of freedom           1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

ANEXO 2

```

+-----+-----+-----+-----+-----+
Constant | -6.46621354      .58396825    -11.073    .0000
LTMDTOT  | .39495794        .05672973     6.962     .0000    9.29908834
LLENGTH  | .93515129        .04758135    19.654     .0000    5.27644440
DENSINT  | 29.4439129      8.10145931     3.634     .0003    .00457041
OCPSOL2  | -.36169662       .10144017    -3.566     .0004    .22222222
OCPSOL3  | .39459675        .26345168     1.498     .1342    .03030303
OCPSOL4  | -.34948859       .21224678    -1.647     .0996    .05808081
OCPSOL5  | .20785846        .15421839     1.348     .1777    .14141414
DISLOCAL | -.16348452       .10103885    -1.618     .1057    .49747475
-----+-----+-----+-----+-----+
+Dispersion parameter for count data model
Alpha    | .51491327        .04282280    12.024     .0000

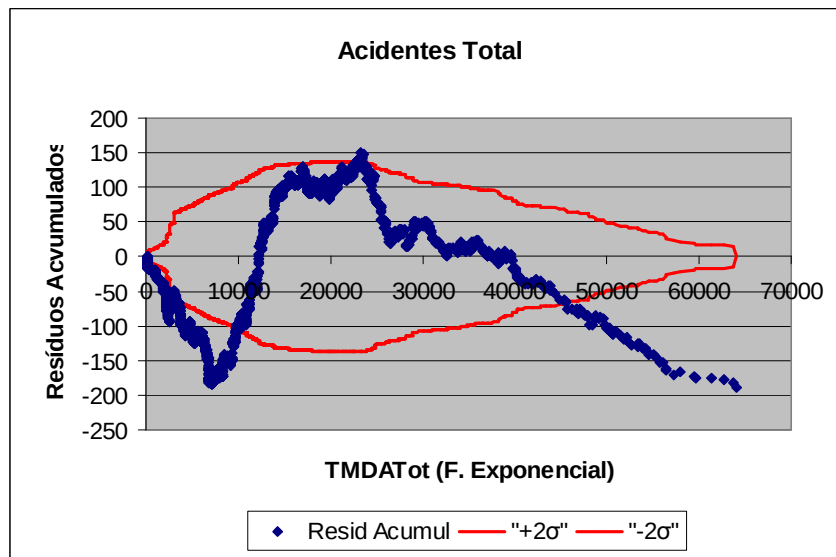
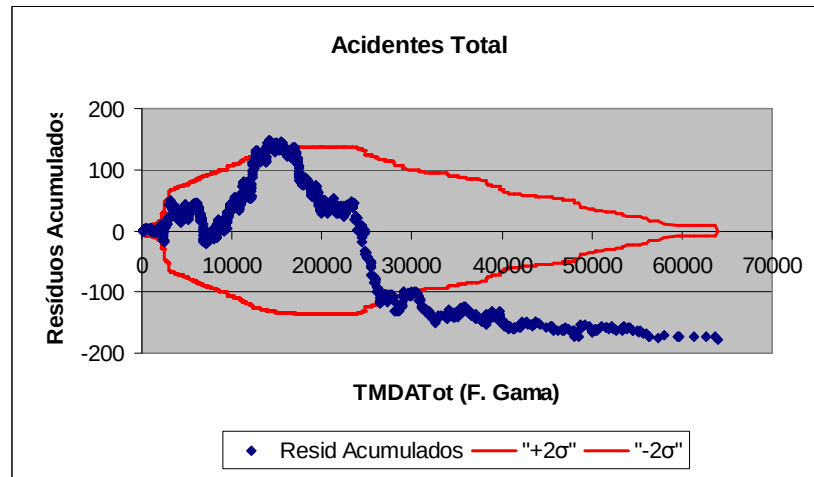
```

ANEXO 2.VI
GRÁFICOS CURE (ARCOS)

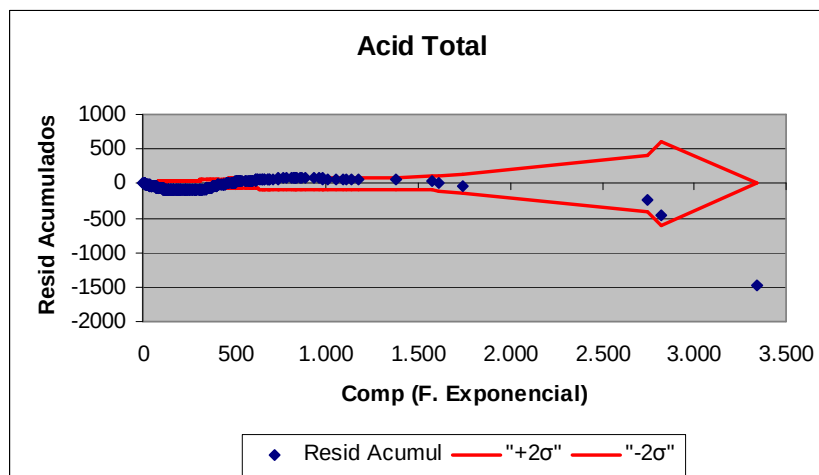
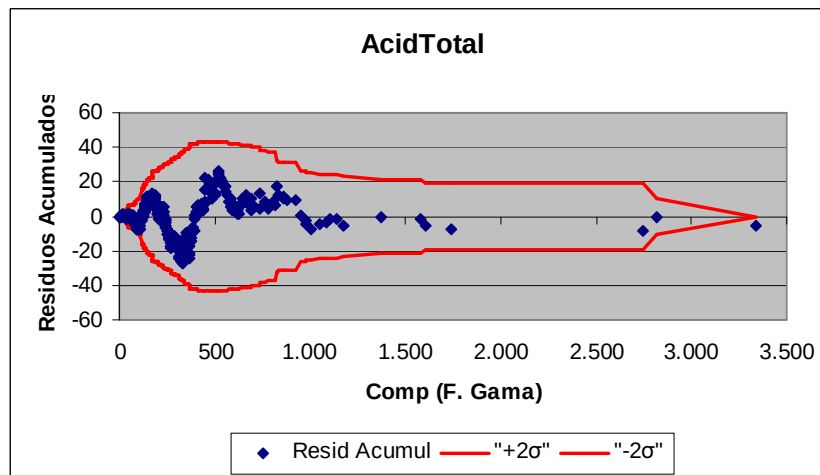
Gráficos de resíduos acumulados segundo o método CURE (Hauer, E. and Bamfo, J., 1997)

Modelo BN - Arcos

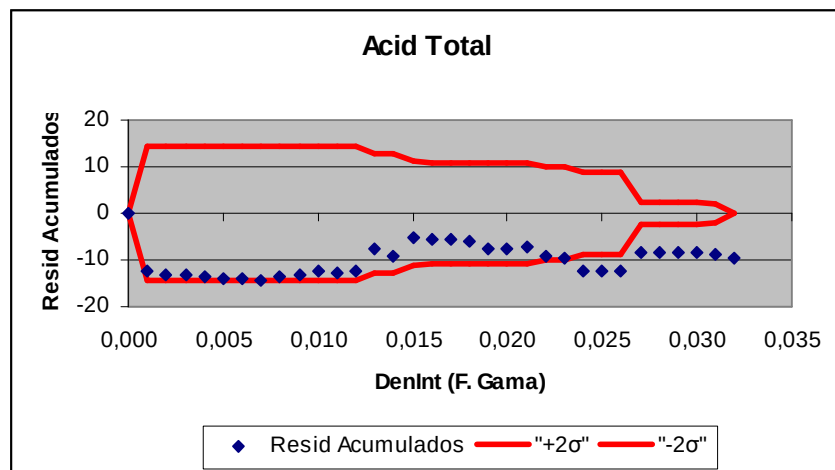
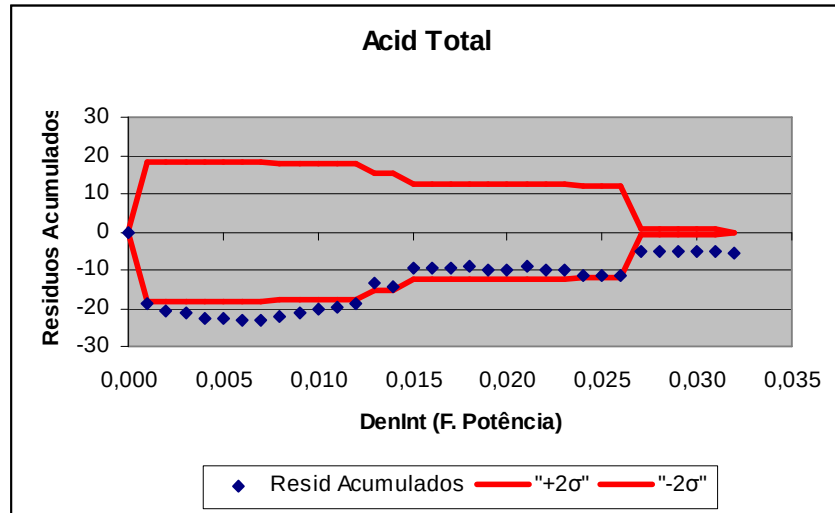
- Variável independente - $TMDA_{Tot}$



- Variável independente - *Comp*



- Variável independente - *DenInt*



Hauer, E. and Bamfo, J. (1997). Two tools for finding what function links the dependent variable to the explanatory variables. In: *ICTCT*, Lund.

ANEXO 2.VII

MODELO BN SEM AS OBSERVAÇÕES DE INFLUÊNCIA (ARCOS)

--> NEGBIN;Lhs=ACIDTOT;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5,DISLOCALS

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:36:09PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1967
| Iterations completed        7
| Log likelihood function     -3997.470
| Number of parameters        9
| Info. Criterion: AIC =      4.07369
|   Finite Sample: AIC =      4.07373
| Info. Criterion: BIC =      4.09924
| Info. Criterion:HQIC =      4.08308
| Restricted log likelihood    -6190.325
| McFadden Pseudo R-squared   .3542390
| Chi squared                 4385.709
| Degrees of freedom          8
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 4342.31360  RsqP= .5783
| G - squared = 4119.23501  RsqD= .5157
| Overdispersion tests: g=mu(i) : 9.736
| Overdispersion tests: g=mu(i)^2: 9.716
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.47452958	.19892898	-32.547	.0000	
LTMDTOT	.27658720	.01957662	14.128	.0000	9.27745471
LLENGTH	.83415088	.01582140	52.723	.0000	5.26973608
DENSINT	49.7571873	3.02773945	16.434	.0000	.00456242
OCPSOL2	-.30793724	.03611093	-8.528	.0000	.22369090
OCPSOL3	.31093475	.08195317	3.794	.0001	.02694459
OCPSOL4	-.37716310	.06240592	-6.044	.0000	.05592272
OCPSOL5	.14330941	.03639222	3.938	.0001	.14184037
DISLOCAL	-.17274276	.03058074	-5.649	.0000	.49669548

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:36:10PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1967
| Iterations completed        18
| Log likelihood function     -3602.984
| Number of parameters        10
| Info. Criterion: AIC =      3.67360
|   Finite Sample: AIC =      3.67366
| Info. Criterion: BIC =      3.70199
| Info. Criterion:HQIC =      3.68403
| Restricted log likelihood    -3997.470
| McFadden Pseudo R-squared   .0986840
| Chi squared                 788.9728
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

ANEXO 2

+-----+-----+-----+-----+-----+					
Constant	-7.30279691	.32856429	-22.226	.0000	
LTMDTOT	.32558022	.02940586	11.072	.0000	9.27745471
LLENGTH	.91021804	.02916044	31.214	.0000	5.26973608
DENSINT	36.6871675	4.62709631	7.929	.0000	.00456242
OCPSOL2	-.33667509	.05471622	-6.153	.0000	.22369090
OCPSOL3	.34908475	.14962779	2.333	.0196	.02694459
OCPSOL4	-.42069323	.10629781	-3.958	.0001	.05592272
OCPSOL5	.20294433	.07272212	2.791	.0053	.14184037
DISLOCAL	-.18147709	.04923516	-3.686	.0002	.49669548
-----+Dispersion parameter for count data model					
Alpha	.44163586	.02939503	15.024	.0000	

ANEXO 2.VIII

MODELO COM HETEROSCEDASTICIDADE (ARCOS)

-->

NEGBIN; Lhs=ACIDTOT; RhS=ONE, LTMDTOT, LLENGTH, DENSINT, ANO, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, DISLOCAL; Wts=TMDATOT\$

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:42:43PM.
| Dependent variable          ACIDTOT
| Weighting variable          TMDATOT
| Number of observations      1980
| Iterations completed        6
| Log likelihood function     -4563.764
| Number of parameters        10
| Info. Criterion: AIC =      4.61996
|   Finite Sample: AIC =      4.62002
| Info. Criterion: BIC =      4.64820
| Info. Criterion: HQIC =     4.63034
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .2797481
| Chi squared                 3545.160
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4771.79355  RsqP= .6595
| G - squared = 4649.44362  RsqD= .6419
| Overdispersion tests: g=mu(i) : 10.542
| Overdispersion tests: g=mu(i)^2: 10.574
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-5.98098515	.23169704	-25.814	.0000	
LTMDTOT	.26574691	.02213346	12.007	.0000	9.28075747
LLENGTH	.78708870	.01334001	59.002	.0000	5.27644440
DENSINT	48.9538976	2.73722931	17.884	.0000	.00457041
ANO	-.05471378	.00820049	-6.672	.0000	3.00000000
OCPSOL2	-.19454603	.03106247	-6.263	.0000	.22222222
OCPSOL3	.50015718	.06474344	7.725	.0000	.03030303
OCPSOL4	-.14599767	.04684437	-3.117	.0018	.05808081
OCPSOL5	.18553560	.03262954	5.686	.0000	.14141414
DISLOCAL	-.08985285	.02936890	-3.059	.0022	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 10:42:43PM.
| Dependent variable          ACIDTOT
| Weighting variable          TMDATOT
| Number of observations      1980
| Iterations completed       19
| Log likelihood function     -4088.223
| Number of parameters        11
| Info. Criterion: AIC =      4.14063
|   Finite Sample: AIC =      4.14070
| Info. Criterion: BIC =      4.17169
| Info. Criterion: HQIC =     4.15204
| Restricted log likelihood    -4563.764
| McFadden Pseudo R-squared   .1041993
| Chi squared                 951.0818
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+
```

ANEXO 2

+-----+ Variable	+-----+ Coefficient	+-----+ Standard Error	+-----+ b/St.Er.	+-----+ P[Z >z]	+-----+ Mean of X
Constant	-6.45616336	.41022990	-15.738	.0000	
LTMDTOT	.28696277	.03780787	7.590	.0000	9.28075747
LLENGTH	.84270669	.02609509	32.294	.0000	5.27644440
DENSINT	36.3500738	4.18503680	8.686	.0000	.00457041
ANO	-.04719352	.01464856	-3.222	.0013	3.00000000
OCPSOL2	-.24582475	.05053986	-4.864	.0000	.22222222
OCPSOL3	.44386562	.17397312	2.551	.0107	.03030303
OCPSOL4	-.20699491	.09388457	-2.205	.0275	.05808081
OCPSOL5	.22942128	.06718819	3.415	.0006	.14141414
DISLOCAL	-.10953359	.05117315	-2.140	.0323	.49747475
-----+Dispersion parameter for count data model					
Alpha		.42855981	.02601442	16.474	.0000

ANEXO 2.IX

MODELO COM PARÂMETROS ALEATÓRIOS (ARCOS)


```
-->
poisson;Lhs=ACIDtot;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,ano,OCPSOL2,OCPSOL3
,OCPSOL4,ocpsol5,DISLOCAL;pds=5;RPM;Fcn=ONE(n),LTMDTOT(n),LLENGTH(n),DENS
...
ocpsol5(n),DISLOCAL(n);pts=200;halton;list$
```

```
+-----+
| Poisson Regression Start Values for ACIDTO |
| Maximum Likelihood Estimates               |
| Model estimated: Mar 10, 2010 at 10:45:16PM. |
| Dependent variable                         ACIDTOT |
| Weighting variable                         None     |
| Number of observations                     1980     |
| Iterations completed                      10        |
| Log likelihood function                   -10654.14  |
| Number of parameters                      10        |
| Info. Criterion: AIC =                    10.77186  |
|   Finite Sample: AIC =                    10.77192  |
| Info. Criterion: BIC =                    10.80010  |
| Info. Criterion:HQIC =                    10.78223  |
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
OCPSOL3	.44724185	.06752801	6.623	.0000	.03030303
OCPSOL4	-.25631824	.05733330	-4.471	.0000	.05808081
Constant	-6.50884654	.19775824	-32.913	.0000	
LTMDTOT	.29704714	.01941162	15.303	.0000	9.28075747
LLENGTH	.83506371	.01568882	53.227	.0000	5.27644440
DENSINT	50.8228442	2.98282182	17.039	.0000	.00457041
ANO	-.05930893	.00940289	-6.308	.0000	3.00000000
OCPSOL2	-.31292193	.03611472	-8.665	.0000	.22222222
OCPSOL5	.13916194	.03630462	3.833	.0001	.14141414
DISLOCAL	-.16262804	.03025423	-5.375	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Random Coefficients Poisson Model         |
| Maximum Likelihood Estimates               |
| Model estimated: Mar 10, 2010 at 10:54:39PM. |
| Dependent variable                         ACIDTOT |
| Weighting variable                         None     |
| Number of observations                     1980     |
| Iterations completed                      61        |
| Log likelihood function                   -3341.237  |
| Number of parameters                      18        |
| Info. Criterion: AIC =                    3.39317  |
|   Finite Sample: AIC =                    3.39334  |
| Info. Criterion: BIC =                    3.44399  |
| Info. Criterion:HQIC =                    3.41184  |
| Restricted log likelihood                  -10654.14  |
| McFadden Pseudo R-squared                 .6863908    |
| Chi squared                              14625.81   |
| Degrees of freedom                       8          |
| Prob[ChiSqd > value] =                    .00000000  |
| Sample is 5 pds and 396 individuals.       |
| POISSON regression model                  |
| Simulation based on 200 Halton draws       |
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----+Nonrandom parameters					
OCPSOL3	.38846627	.06907600	5.624	.0000	.03030303
OCPSOL4	-.38448248	.05821504	-6.605	.0000	.05808081
-----+Means for random parameters					

ANEXO 2

Constant	-6.47384999	.19189158	-33.737	.0000	
LTMDTOT	.18989110	.01844590	10.294	.0000	9.28075747
LLENGTH	.98804873	.01770403	55.809	.0000	5.27644440
DENSINT	45.2529010	3.27362610	13.823	.0000	.00457041
ANO	-.06306815	.00928238	-6.794	.0000	3.00000000
OCPSOL2	-.54566782	.03919139	-13.923	.0000	.22222222
OCPSOL5	.13920040	.03787647	3.675	.0002	.14141414
DISLOCAL	-.27835403	.03007596	-9.255	.0000	.49747475
-----+Scale parameters for dists. of random parameters					
Constant	.11307850	.01369627	8.256	.0000	
LTMDTOT	.06604199	.00172324	38.324	.0000	
LLENGTH	.04307092	.00231670	18.591	.0000	
DENSINT	25.6584107	1.83976057	13.947	.0000	
ANO	.06692815	.00451517	14.823	.0000	
OCPSOL2	.41335513	.03255860	12.696	.0000	
OCPSOL5	.11968695	.03285764	3.643	.0003	
DISLOCAL	.15521999	.02111162	7.352	.0000	

Implied standard deviations of random parameters

Matrix S.D_Beta has 8 rows and 1 columns.

	1
1	.11308
2	.06604
3	.04307
4	25.65841
5	.06693
6	.41336
7	.11969
8	.15522

ANEXO 2.X
MODELO ZIP (ARCOS)

-->

poisson;Lhs=ACIDtot;Rhs=ONE,LTMDTOT,LLENGTH,DENSINT,ano,OCPSOL2,OCPSOL3,
OCPSOL4,ocpsol5,DISLOCAL;ZIP\$

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 10, 2010 at 11:02:41PM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations       1980
| Iterations completed        7
| Log likelihood function      -4043.441
| Number of parameters         10
| Info. Criterion: AIC =      4.09438
|   Finite Sample: AIC =      4.09444
| Info. Criterion: BIC =      4.12262
| Info. Criterion:HQIC =      4.10476
| Restricted log likelihood     -6336.343
| McFadden Pseudo R-squared    .3618652
| Chi squared                  4585.805
| Degrees of freedom           9
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4354.45312  RsqP= .5866
| G - squared = 4158.44284  RsqD= .5244
| Overdispersion tests: g=mu(i) : 10.319
| Overdispersion tests: g=mu(i)^2: 10.270
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-6.50884674	.19775825	-32.913	.0000	
LTMDTOT	.29704715	.01941162	15.303	.0000	9.28075747
LLENGTH	.83506373	.01568882	53.227	.0000	5.27644440
DENSINT	50.8228470	2.98282184	17.039	.0000	.00457041
ANO	-.05930893	.00940289	-6.308	.0000	3.00000000
OCPSOL2	-.31292194	.03611472	-8.665	.0000	.22222222
OCPSOL3	.44724187	.06752801	6.623	.0000	.03030303
OCPSOL4	-.25631823	.05733330	-4.471	.0000	.05808081
OCPSOL5	.13916193	.03630462	3.833	.0001	.14141414
DISLOCAL	-.16262804	.03025423	-5.375	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Zero Altered Poisson      Regression Model
| Logistic distribution used for splitting model.
| ZAP term in probability is F[tau x ln LAMBDA]
| Comparison of estimated models
|   Pr[0|means]      Number of zeros      Log-likelihood
| Poisson            .16870  Act.= 669 Prd.= 334.0  -4043.44114
| Z.I.Poisson        .29104  Act.= 669 Prd.= 576.3  -3870.89571
| Note, the ZIP log-likelihood is not directly comparable.
| ZIP model with nonzero Q does not encompass the others.
| Vuong statistic for testing ZIP vs. unaltered model is      6.7571
| Distributed as standard normal. A value greater than
| +1.96 favors the zero altered Z.I.Poisson model.
| A value less than -1.96 rejects the ZIP model.
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-4.92442667	.13665342	-36.036	.0000	
LTMDTOT	.25162560	.01295581	19.422	.0000	9.28075747

-----+Poisson/NB/Gamma regression model

ANEXO 2

LLENGTH		.67728390	.01027677	65.904	.0000	5.27644440
DENSINT		37.0031359	2.02664563	18.258	.0000	.00457041
ANO		-.05341180	.00623843	-8.562	.0000	3.00000000
OCPSOL2		-.23778061	.02391677	-9.942	.0000	.22222222
OCPSOL3		.37325350	.05665010	6.589	.0000	.03030303
OCPSOL4		-.26519017	.03981289	-6.661	.0000	.05808081
OCPSOL5		.12658335	.02349031	5.389	.0000	.14141414
DISLOCAL		-.15202210	.01971853	-7.710	.0000	.49747475
-----+Zero inflation model						
Tau		-1.43430277	.07397655	-19.389	.0000	

ANEXO 2.XI

MODELO COM FUNÇÃO TRANSLOG (ARCOS)


```
-->
NEGBIN;Lhs=ACidTot;Rhs=ONE,LTMDTOT,DENSINT,llength,ltmd_co,ANO,OCPSOL2,OC
...
,OCPSOL4,OCPSOL5,DISLOCAL$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:06:11AM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        7
| Log likelihood function     -4015.979
| Number of parameters        11
| Info. Criterion: AIC =      4.06766
|   Finite Sample: AIC =      4.06772
| Info. Criterion: BIC =      4.09872
| Info. Criterion:HQIC =      4.07907
| Restricted log likelihood    -6336.343
| McFadden Pseudo R-squared   .3661992
| Chi squared                 4640.728
| Degrees of freedom          10
| Prob[ChiSq > value] =       .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 4608.41276 RsqP= .5625
| G - squared = 4103.51899 RsqD= .5307
| Overdispersion tests: g=mu(i) : 8.506
| Overdispersion tests: g=mu(i)^2: 7.990
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-16.1087523	1.35292617	-11.907	.0000	
LTMDTOT	1.27982758	.13798752	9.275	.0000	9.28075747
DENSINT	47.5552427	3.05773750	15.552	.0000	.00457041
LLENGTH	2.46490996	.22574468	10.919	.0000	5.27644440
LTMD_CO	-.16621993	.02292708	-7.250	.0000	49.0872742
ANO	-.05753794	.00941402	-6.112	.0000	3.00000000
OCPSOL2	-.28037383	.03626126	-7.732	.0000	.22222222
OCPSOL3	.42801309	.06784404	6.309	.0000	.03030303
OCPSOL4	-.26737540	.05732646	-4.664	.0000	.05808081
OCPSOL5	.14644688	.03627223	4.037	.0001	.14141414
DISLOCAL	-.18675460	.03050853	-6.121	.0000	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:06:12AM.
| Dependent variable          ACIDTOT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        21
| Log likelihood function     -3639.146
| Number of parameters        12
| Info. Criterion: AIC =      3.68803
|   Finite Sample: AIC =      3.68811
| Info. Criterion: BIC =      3.72191
| Info. Criterion:HQIC =      3.70047
| Restricted log likelihood    -4015.979
| McFadden Pseudo R-squared   .0938335
| Chi squared                 753.6669
| Degrees of freedom          1
| Prob[ChiSq > value] =       .0000000
+-----+
```

NegBin form 2; Psi(i) = theta					
+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
+-----+					
Constant	-14.2951994	1.44317298	-9.905	.0000	
LTMDTOT	1.06839691	.15092722	7.079	.0000	9.28075747
DENSINT	36.1996362	4.55190229	7.953	.0000	.00457041
LLENGTH	2.14550137	.26232815	8.179	.0000	5.27644440
LTMD_CO	-.12811978	.02741542	-4.673	.0000	49.0872742
ANO	-.05493523	.01561924	-3.517	.0004	3.00000000
OCPSOL2	-.31374607	.05468554	-5.737	.0000	.22222222
OCPSOL3	.39670633	.14232199	2.787	.0053	.03030303
OCPSOL4	-.31599290	.09481247	-3.333	.0009	.05808081
OCPSOL5	.20496345	.07127076	2.876	.0040	.14141414
DISLOCAL	-.18166001	.04932655	-3.683	.0002	.49747475
-----+Dispersion parameter for count data model					
Alpha	.42330632	.02932162	14.437	.0000	

ANEXO 2.XII

MODELO BN – ACIDENTES COM VÍTIMAS (ARCOS)

-->

NEGBIN; Lhs=ACIDvIT; RhS=ONE, LTMDTOT, LLENGTH, DENSINT, ANO, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, DISLOCAL\$

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:09:15AM.
| Dependent variable          ACIDVIT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        8
| Log likelihood function     -1740.677
| Number of parameters        10
| Info. Criterion: AIC =      1.76836
|   Finite Sample: AIC =      1.76842
| Info. Criterion: BIC =      1.79660
| Info. Criterion:HQIC =      1.77873
| Restricted log likelihood    -2378.100
| McFadden Pseudo R-squared   .2680388
| Chi squared                 1274.846
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 2453.37409  RsqP= .4850
| G - squared = 1991.54808  RsqD= .3903
| Overdispersion tests: g=mu(i) : 6.937
| Overdispersion tests: g=mu(i)^2: 7.795
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-10.6788963	.46946209	-22.747	.0000	
LTMDTOT	.50365712	.04565863	11.031	.0000	9.28075747
LLENGTH	.92486194	.03446162	26.837	.0000	5.27644440
DENSINT	51.4193721	6.77482054	7.590	.0000	.00457041
ANO	-.09240580	.02049973	-4.508	.0000	3.00000000
OCPSOL2	-.06657016	.07520691	-.885	.3761	.22222222
OCPSOL3	-.18389140	.20232783	-.909	.3634	.03030303
OCPSOL4	-.12810740	.12206488	-1.050	.2939	.05808081
OCPSOL5	.22915975	.07844626	2.921	.0035	.14141414
DISLOCAL	.03530119	.06741383	.524	.6005	.49747475

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:09:16AM.
| Dependent variable          ACIDVIT
| Weighting variable          None
| Number of observations      1980
| Iterations completed        18
| Log likelihood function     -1688.896
| Number of parameters        11
| Info. Criterion: AIC =      1.71707
|   Finite Sample: AIC =      1.71713
| Info. Criterion: BIC =      1.74813
| Info. Criterion:HQIC =      1.72848
| Restricted log likelihood    -1740.677
| McFadden Pseudo R-squared   .0297478
| Chi squared                 103.5627
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+
```

+-----+ Variable	+-----+ Coefficient	+-----+ Standard Error	+-----+ b/St.Er.	+-----+ P[Z >z]	+-----+ Mean of X
Constant	-11.0509799	.60631380	-18.227	.0000	
LTMDTOT	.51047269	.05521785	9.245	.0000	9.28075747
LLENGTH	.98903296	.05231271	18.906	.0000	5.27644440
DENSINT	42.8159038	8.89190968	4.815	.0000	.00457041
ANO	-.09146447	.02582930	-3.541	.0004	3.00000000
OCPSOL2	-.12885103	.09518374	-1.354	.1758	.22222222
OCPSOL3	-.29100944	.22233294	-1.309	.1906	.03030303
OCPSOL4	-.18823439	.18449273	-1.020	.3076	.05808081
OCPSOL5	.28134435	.11265741	2.497	.0125	.14141414
DISLOCAL	.01464109	.08034117	.182	.8554	.49747475
-----+Dispersion parameter for count data model					
Alpha	.51166837	.08415566	6.080	.0000	

ANEXO 2.XIII

MODELO COM “DUMMY” $TMDA_{sec}$ (NÓS)

--> NEGBIN;Lhs=ACIDTOTA;Rhs=ONE, LTMDP, LTMS, dummytmd\$

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:23:42AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1055
| Iterations completed        7
| Log likelihood function     -2049.907
| Number of parameters        4
| Info. Criterion: AIC =      3.89366
|   Finite Sample: AIC =      3.89370
| Info. Criterion: BIC =      3.91247
| Info. Criterion:HQIC =      3.90079
| Restricted log likelihood    -2207.506
| McFadden Pseudo R-squared   .0713925
| Chi squared                 315.1989
| Degrees of freedom          3
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 2193.84172  RsqP= .1062
| G - squared = 2116.28864  RsqD= .1296
| Overdispersion tests: g=mu(i) : 8.134
| Overdispersion tests: g=mu(i)^2: 7.169
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-4.28391030	.36593473	-11.707	.0000	
LTMDP	.43680130	.03751868	11.642	.0000	9.60164716
LTMS	.09458112	.02265359	4.175	.0000	6.95332183
DUMMYTMD	-.01309503	.21039970	-.062	.9504	.16113744

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:23:43AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1055
| Iterations completed        9
| Log likelihood function     -1903.777
| Number of parameters        5
| Info. Criterion: AIC =      3.61853
|   Finite Sample: AIC =      3.61859
| Info. Criterion: BIC =      3.64205
| Info. Criterion:HQIC =      3.62745
| Restricted log likelihood    -2049.907
| McFadden Pseudo R-squared   .0712860
| Chi squared                 292.2593
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-4.06721053	.49379295	-8.237	.0000	
LTMDP	.40918962	.04961774	8.247	.0000	9.60164716
LTMS	.10068253	.03345934	3.009	.0026	6.95332183
DUMMYTMD	.03872329	.29120147	.133	.8942	.16113744

-----+Dispersion parameter for count data model

ANEXO 2

Alpha		.54741551	.05486584	9.977	.0000

ANEXO 2.XIV

MODELO BN (NÓS) – EQUAÇÃO (5.6)

--> NEGBIN;Lhs=ACIDTOTA;Rhs=ONE, LTMDP, LTMDs, ANO, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, CLASSEA, CLASSECS

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:31:06AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations       1055
| Iterations completed        7
| Log likelihood function      -1988.374
| Number of parameters        10
| Info. Criterion: AIC =      3.78839
|   Finite Sample: AIC =      3.78859
| Info. Criterion: BIC =      3.83541
| Info. Criterion:HQIC =      3.80621
| Restricted log likelihood    -2207.506
| McFadden Pseudo R-squared   .0992668
| Chi squared                 438.2640
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 1997.50985  RsqP= .1862
| G - squared = 1993.22349  RsqD= .1802
| Overdispersion tests: g=mu(i) : 8.418
| Overdispersion tests: g=mu(i)^2: 7.691
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.28675732	.38947659	-8.439	.0000	
LTMDP	.37370358	.03950482	9.460	.0000	9.60164716
LTMDs	.09422341	.00874381	10.776	.0000	6.95332183
ANO	-.09938093	.01571526	-6.324	.0000	3.00000000
OCPSOL2	-.00854442	.05874694	-.145	.8844	.21800948
OCPSOL3	.71389891	.12894428	5.536	.0000	.02843602
OCPSOL4	-.13355481	.10815416	-1.235	.2169	.05213270
OCPSOL5	.19429862	.06215201	3.126	.0018	.14218009
CLASSEA	-.34686192	.10647878	-3.258	.0011	.05687204
CLASSECS	-.42477118	.05742611	-7.397	.0000	.33649289

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:31:06AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations       1055
| Iterations completed        16
| Log likelihood function      -1871.528
| Number of parameters        11
| Info. Criterion: AIC =      3.56877
|   Finite Sample: AIC =      3.56901
| Info. Criterion: BIC =      3.62050
| Info. Criterion:HQIC =      3.58838
| Restricted log likelihood    -1988.374
| McFadden Pseudo R-squared   .0587647
| Chi squared                 233.6923
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

ANEXO 2

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.13979242	.49995372	-6.280	.0000	
LTMDP	.35865177	.05204163	6.892	.0000	9.60164716
LTMDS	.09517090	.01145181	8.311	.0000	6.95332183
ANO	-.10245261	.02344619	-4.370	.0000	3.00000000
OCPSOL2	-.03780880	.08388591	-.451	.6522	.21800948
OCPSOL3	.69371100	.21911165	3.166	.0015	.02843602
OCPSOL4	-.09787068	.16818381	-.582	.5606	.05213270
OCPSOL5	.23632048	.10179582	2.322	.0203	.14218009
CLASSEA	-.38774069	.13605919	-2.850	.0044	.05687204
CLASSEC	-.41982314	.08153715	-5.149	.0000	.33649289
-----+Dispersion parameter for count data model					
Alpha	.46917765	.05221467	8.986	.0000	

ANEXO 2.XV

MODELO BN (NÓS) – TENDÊNCIA TEMPORAL


```
-->
NEGBIN; Lhs=ACIDTOTA; Rhs=ONE, LTMDP, LTMDs, dummya2, dummya3, dummya4, dummya5, 0
...
, OCPSOL4, OCPSOL5, CLASSEA, CLASSECS
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:33:05AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1055
| Iterations completed        7
| Log likelihood function     -1981.250
| Number of parameters        13
| Info. Criterion: AIC =      3.78057
|   Finite Sample: AIC =      3.78090
| Info. Criterion: BIC =      3.84170
| Info. Criterion: HQIC =     3.80374
| Restricted log likelihood   -2207.506
| McFadden Pseudo R-squared  .1024939
| Chi squared                 452.5120
| Degrees of freedom          12
| Prob[ChiSqd > value] =     .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 1979.38838 RsqP= .1936
| G - squared = 1978.97552 RsqD= .1861
| Overdispersion tests: g=mu(i) : 8.353
| Overdispersion tests: g=mu(i)^2: 7.635
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.45636220	.39476695	-8.755	.0000	
LTMDP	.37999491	.03970467	9.571	.0000	9.60164716
LTMDs	.09448469	.00874965	10.799	.0000	6.95332183
DUMMYA2	-.07598995	.06633557	-1.146	.2520	.20000000
DUMMYA3	-.27965394	.06936674	-4.032	.0001	.20000000
DUMMYA4	-.14404740	.06663733	-2.162	.0306	.20000000
DUMMYA5	-.47656805	.07209351	-6.610	.0000	.20000000
OCPSOL2	-.00936574	.05875949	-.159	.8734	.21800948
OCPSOL3	.71572549	.12894392	5.551	.0000	.02843602
OCPSOL4	-.13400485	.10816075	-1.239	.2154	.05213270
OCPSOL5	.19191337	.06217821	3.087	.0020	.14218009
CLASSEA	-.34613935	.10646579	-3.251	.0011	.05687204
CLASSECS	-.42174472	.05746002	-7.340	.0000	.33649289

Normal exit from iterations. Exit status=0.

```
+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:33:05AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1055
| Iterations completed        19
| Log likelihood function     -1867.761
| Number of parameters        14
| Info. Criterion: AIC =      3.56732
|   Finite Sample: AIC =      3.56770
| Info. Criterion: BIC =      3.63316
| Info. Criterion: HQIC =     3.59228
| Restricted log likelihood   -1981.250
| McFadden Pseudo R-squared  .0572816
| Chi squared                 226.9783
+-----+
```

Degrees of freedom 1					
Prob[ChiSqd > value] = .0000000					
NegBin form 2; Psi(i) = theta					
+-----+-----+-----+-----+-----+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
+-----+-----+-----+-----+-----+-----+					
Constant	-3.31703871	.50653391	-6.549	.0000	
LTMDP	.36625413	.05211689	7.028	.0000	9.60164716
LTMDS	.09525491	.01141181	8.347	.0000	6.95332183
DUMMYA2	-.08495323	.09176654	-.926	.3546	.20000000
DUMMYA3	-.31011771	.10077237	-3.077	.0021	.20000000
DUMMYA4	-.16316624	.09914933	-1.646	.0998	.20000000
DUMMYA5	-.48459122	.10444829	-4.640	.0000	.20000000
OCPSOL2	-.03880266	.08406487	-.462	.6444	.21800948
OCPSOL3	.69493975	.21876665	3.177	.0015	.02843602
OCPSOL4	-.09840359	.16751381	-.587	.5569	.05213270
OCPSOL5	.23326251	.10211946	2.284	.0224	.14218009
CLASSEA	-.38948033	.13712383	-2.840	.0045	.05687204
CLASSEC	-.41478561	.08174095	-5.074	.0000	.33649289
-----+Dispersion parameter for count data model					
Alpha	.46014858	.05171023	8.899	.0000	

ANEXO 2.XVI

MODELO BN AGREGADO NOS 5 ANOS (NÓS)

--> NEGBIN;Lhs=ACIDTOTA;Rhs=ONE, LTMDP, LTMDs, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, CLASSEA, CLASSECS

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:42:35AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      211
| Iterations completed        6
| Log likelihood function     -885.9705
| Number of parameters        9
| Info. Criterion: AIC =      8.48313
|   Finite Sample: AIC =      8.48738
| Info. Criterion: BIC =      8.62610
| Info. Criterion:HQIC =      8.54092
| Restricted log likelihood    -1112.416
| McFadden Pseudo R-squared   .2035620
| Chi squared                 452.8913
| Degrees of freedom          8
| Prob[ChiSqd > value] =      .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 1021.41413  RsqP= .3201
| G - squared = 1012.68631  RsqD= .3090
| Overdispersion tests: g=mu(i) : 7.047
| Overdispersion tests: g=mu(i)^2: 6.462
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.01278763	.42899849	-7.023	.0000	
LTMDP	.47566526	.04299662	11.063	.0000	9.62185410
LTMDs	.09807661	.00881854	11.122	.0000	6.97109819
OCPSOL2	-.01864896	.05900229	-.316	.7519	.21800948
OCPSOL3	.74780905	.12891856	5.801	.0000	.02843602
OCPSOL4	-.13410317	.10824446	-1.239	.2154	.05213270
OCPSOL5	.16862879	.06229400	2.707	.0068	.14218009
CLASSEA	-.33440017	.10627106	-3.147	.0017	.05687204
CLASSECS	-.37375837	.05798734	-6.446	.0000	.33649289

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:42:35AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      211
| Iterations completed        16
| Log likelihood function     -658.7709
| Number of parameters       10
| Info. Criterion: AIC =      6.33906
|   Finite Sample: AIC =      6.34427
| Info. Criterion: BIC =      6.49792
| Info. Criterion:HQIC =      6.40327
| Restricted log likelihood    -885.9705
| McFadden Pseudo R-squared   .2564415
| Chi squared                 454.3993
| Degrees of freedom          1
| Prob[ChiSqd > value] =      .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

ANEXO 2

+-----+-----+-----+-----+-----+					
Constant	-2.70967085	.99029219	-2.736	.0062	
LTMDP	.44331595	.10062185	4.406	.0000	9.62185410
LTMDS	.09897894	.01893414	5.228	.0000	6.97109819
OCPSOL2	-.04876334	.14373112	-.339	.7344	.21800948
OCPSOL3	.70383837	.37213270	1.891	.0586	.02843602
OCPSOL4	-.03149791	.29175674	-.108	.9140	.05213270
OCPSOL5	.23776214	.18408426	1.292	.1965	.14218009
CLASSEA	-.43995574	.21693504	-2.028	.0426	.05687204
CLASSEC	-.36762274	.15078884	-2.438	.0148	.33649289
-----+Dispersion parameter for count data model					
Alpha	.46769067	.05959285	7.848	.0000	

ANEXO 2.XVII

MODELO BN SEM AS OBSERVAÇÕES DE INFLUÊNCIA (NÓS)

--> NEGBIN;Lhs=ACIDTOTA;Rhs=ONE, LTMDP, LTMS, ANO, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5, CLASSEA, CLASSECS

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:45:10AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1032
| Iterations completed        7
| Log likelihood function     -1927.960
| Number of parameters        10
| Info. Criterion: AIC =      3.75574
|   Finite Sample: AIC =      3.75595
| Info. Criterion: BIC =      3.80360
| Info. Criterion:HQIC =      3.77390
| Restricted log likelihood   -2122.827
| McFadden Pseudo R-squared  .0917958
| Chi squared                 389.7334
| Degrees of freedom          9
| Prob[ChiSqd > value] =     .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi- squared = 1933.27609 RsqP= .1841
| G - squared = 1933.24447 RsqD= .1678
| Overdispersion tests: g=mu(i) : 8.031
| Overdispersion tests: g=mu(i)^2: 7.537
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-2.74282227	.38181270	-7.184	.0000	
LTMDP	.32752848	.03749279	8.736	.0000	9.62785121
LTMS	.07354597	.00758482	9.696	.0000	7.65329303
ANO	-.10010679	.01614213	-6.202	.0000	3.00872093
OCPSOL2	-.01470994	.05952330	-.247	.8048	.21899225
OCPSOL3	.68998500	.16734125	4.123	.0000	.02034884
OCPSOL4	-.17354184	.12200169	-1.422	.1549	.04748062
OCPSOL5	.29891491	.06161313	4.851	.0000	.14437984
CLASSEA	-.81201001	.15159707	-5.356	.0000	.04651163
CLASSECS	-.46684762	.05702695	-8.186	.0000	.33914729

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:45:10AM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1032
| Iterations completed        16
| Log likelihood function     -1816.465
| Number of parameters        11
| Info. Criterion: AIC =      3.54160
|   Finite Sample: AIC =      3.54185
| Info. Criterion: BIC =      3.59425
| Info. Criterion:HQIC =      3.56158
| Restricted log likelihood   -1927.960
| McFadden Pseudo R-squared  .0578310
| Chi squared                 222.9917
| Degrees of freedom          1
| Prob[ChiSqd > value] =     .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-2.58599744	.49726290	-5.200	.0000	
LTMDP	.31040452	.05076899	6.114	.0000	9.62785121
LTMDS	.07431932	.01052765	7.059	.0000	7.65329303
ANO	-.09985685	.02345123	-4.258	.0000	3.00872093
OCPSOL2	-.05160247	.08365807	-.617	.5373	.21899225
OCPSOL3	.69546745	.36938295	1.883	.0597	.02034884
OCPSOL4	-.17830519	.20929806	-.852	.3943	.04748062
OCPSOL5	.34617948	.09966158	3.474	.0005	.14437984
CLASSEA	-.81575872	.19967597	-4.085	.0000	.04651163
CLASSEC	-.45872161	.07828265	-5.860	.0000	.33914729
-----+Dispersion parameter for count data model					
Alpha	.46902620	.05317297	8.821	.0000	

ANEXO 2.XVIII

MODELO COM HETEROSCEDASTICIDADE (NÓS)

-->

NEGBIN; Lhs=ACIDTOTA; Rhs=ONE, ANO, LTMDP, LTMDS, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5
L5
, CLASSEA, CLASSEB; Wts=LTMDP\$

```

+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:47:56AM.
| Dependent variable          ACIDTOTA
| Weighting variable          LTMDP
| Number of observations      1055
| Iterations completed        7
| Log likelihood function     -2002.047
| Number of parameters        10
| Info. Criterion: AIC =      3.81431
|   Finite Sample: AIC =      3.81451
| Info. Criterion: BIC =      3.86133
| Info. Criterion: HQIC =     3.83213
| Restricted log likelihood    -2207.506
| McFadden Pseudo R-squared   .0930731
| Chi squared                 410.9189
| Degrees of freedom          9
| Prob[ChiSq > value] =       .0000000
+-----+

```

```

+-----+
| Poisson Regression
| Chi-squared = 1992.18307  RsqP= .2022
| G-squared = 1993.76900  RsqD= .2087
| Overdispersion tests: g=mu(i) : 8.369
| Overdispersion tests: g=mu(i)^2: 7.586
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.42512699	.39796884	-8.607	.0000	
ANO	-.09918083	.01551634	-6.392	.0000	3.00000000
LTMDP	.38783185	.04017023	9.655	.0000	9.60164716
LTMDS	.09419268	.00858713	10.969	.0000	6.95332183
OCPSOL2	-.00030665	.05800678	-.005	.9958	.21800948
OCPSOL3	.72166140	.13111673	5.504	.0000	.02843602
OCPSOL4	-.12710406	.10594763	-1.200	.2303	.05213270
OCPSOL5	.17601361	.06148878	2.863	.0042	.14218009
CLASSEA	-.33302153	.10472713	-3.180	.0015	.05687204
CLASSEB	-.42671255	.05772222	-7.393	.0000	.33649289

Normal exit from iterations. Exit status=0.

```

+-----+
| Negative Binomial Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 10:47:57AM.
| Dependent variable          ACIDTOTA
| Weighting variable          LTMDP
| Number of observations      1055
| Iterations completed       16
| Log likelihood function     -1886.780
| Number of parameters        11
| Info. Criterion: AIC =      3.59769
|   Finite Sample: AIC =      3.59793
| Info. Criterion: BIC =      3.64942
| Info. Criterion: HQIC =     3.61730
| Restricted log likelihood    -2002.047
| McFadden Pseudo R-squared   .0575745
| Chi squared                 230.5336
| Degrees of freedom          1
| Prob[ChiSq > value] =       .0000000
| NegBin form 2; Psi(i) = theta
+-----+

```

+-----+					
+-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
+-----+					
Constant	-3.27821981	.51004641	-6.427	.0000	
ANO	-.10211287	.02312947	-4.415	.0000	3.00000000
LTMDP	.37276256	.05280368	7.059	.0000	9.60164716
LTMDS	.09508949	.01129705	8.417	.0000	6.95332183
OCPSOL2	-.02587824	.08290929	-.312	.7549	.21800948
OCPSOL3	.70125531	.22024112	3.184	.0015	.02843602
OCPSOL4	-.09178575	.16484268	-.557	.5777	.05213270
OCPSOL5	.21756568	.09937287	2.189	.0286	.14218009
CLASSEA	-.37612880	.13398165	-2.807	.0050	.05687204
CLASSEC	-.42075181	.08127810	-5.177	.0000	.33649289
-----+Dispersion parameter for count data model					
Alpha	.45692493	.05103047	8.954	.0000	

ANEXO 2.XIX

MODELO COM PARÂMETROS ALEATÓRIOS (NÓS)


```
--> poisson;Lhs=ACIDtota;Rhs=ONE,LTMDP,LTMD5,ano,OCPSOL2,OCPSOL3
      ,OCPSOL4,ocpsol5,classeA,classeC,pds=5;RPM;Fcn=ONE(n),LTMDP(n),ano
(n),cla...
```

```
+-----+
Poisson Regression Start Values for ACIDTO
Maximum Likelihood Estimates
Model estimated: Nov 26, 2009 at 06:00:06PM.
Dependent variable      ACIDTOTA
Weighting variable      None
Number of observations   1055
Iterations completed    10
Log likelihood function  -3493.851
Number of parameters    10
Info. Criterion: AIC =   6.64237
Finite Sample: AIC =   6.64257
Info. Criterion: BIC =   6.68940
Info. Criterion: HQIC =  6.66020
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
LTMD5	.09422339	.00874380	10.776	.0000	6.95332183
OCPSOL2	-.00854437	.05874694	-.145	.8844	.21800948
OCPSOL3	.71389886	.12894428	5.536	.0000	.02843602
OCPSOL4	-.13355481	.10815416	-1.235	.2169	.05213270
OCPSOL5	.19429864	.06215201	3.126	.0018	.14218009
Constant	-3.28675661	.38947657	-8.439	.0000	
LTMDP	.37370353	.03950482	9.460	.0000	9.60164716
ANO	-.09938093	.01571526	-6.324	.0000	3.00000000
CLASSEA	-.34686188	.10647877	-3.258	.0011	.05687204
CLASSEB	-.42477117	.05742611	-7.397	.0000	.33649289

Normal exit from iterations. Exit status=0.

```
+-----+
Random Coefficients Poisson Model
Maximum Likelihood Estimates
Model estimated: Nov 26, 2009 at 06:01:44PM.
Dependent variable      ACIDTOTA
Weighting variable      None
Number of observations   1055
Iterations completed    27
Log likelihood function  -1754.873
Number of parameters    15
Info. Criterion: AIC =   3.35521
Finite Sample: AIC =   3.35565
Info. Criterion: BIC =   3.42575
Info. Criterion: HQIC =  3.38195
Restricted log likelihood -3493.851
McFadden Pseudo R-squared .4977252
Chi squared             3477.955
Degrees of freedom      5
Prob[ChiSq > value] =   .0000000
Sample is 5 pds and 211 individuals.
POISSON regression model
Simulation based on 200 Halton draws
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----Nonrandom parameters					
LTMD5	.09031506	.00908401	9.942	.0000	6.95332183
OCPSOL2	-.05952096	.05986354	-.994	.3201	.21800948

OCPSOL3	.55840480	.14509812	3.848	.0001	.02843602
OCPSOL4	.06894386	.11080600	.622	.5338	.05213270
OCPSOL5	.34751876	.06435525	5.400	.0000	.14218009
-----Means for random parameters					
Constant	-1.29476170	.36479102	-3.549	.0004	
LTMDP	.14823375	.03772863	3.929	.0001	9.60164716
ANO	-.10597051	.01532634	-6.914	.0000	3.00000000
CLASSEA	-.89593779	.16245442	-5.515	.0000	.05687204
CLASSEC	-.51197339	.06000208	-8.533	.0000	.33649289
-----Scale parameters for dists. of random parameters					
Constant	.09385813	.02282730	4.112	.0000	
LTMDP	.06663872	.00269842	24.695	.0000	
ANO	.07056599	.00734509	9.607	.0000	
CLASSEA	1.09318444	.14347961	7.619	.0000	
CLASSEC	.16398938	.04734162	3.464	.0005	

Implied standard deviations of random parameters

Matrix S.D_Beta has 5 rows and 1 columns.

1	
+-----	
1	.09386
2	.06664
3	.07057
4	1.09318
5	.16399

ANEXO 2.XX
MODELO ZIP (NÓS)


```
--> poisson;Lhs=ACIDtota;Rhs=ONE,LTMDP,LTMDS,ano,OCPSOL2,OCPSOL3
,OCPSOL4,ocpsol5,classeA,classeC;ZIP=normal$
```

```
+-----+
| Poisson Regression
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 03:03:22PM.
| Dependent variable          ACIDTOTA
| Weighting variable          None
| Number of observations      1055
| Iterations completed        7
| Log likelihood function     -1988.374
| Number of parameters        10
| Info. Criterion: AIC =      3.78839
|   Finite Sample: AIC =      3.78859
| Info. Criterion: BIC =      3.83541
| Info. Criterion:HQIC =      3.80621
| Restricted log likelihood    -2207.506
| McFadden Pseudo R-squared   .0992668
| Chi squared                  438.2640
| Degrees of freedom           9
| Prob[ChiSqd > value] =      .0000000
+-----+
```

```
+-----+
| Poisson Regression
| Chi- squared = 1997.50985  RsqP= .1862
| G - squared = 1993.22349  RsqD= .1802
| Overdispersion tests: g=mu(i) : 8.418
| Overdispersion tests: g=mu(i)^2: 7.691
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-3.28675732	.38947659	-8.439	.0000	
LTMDP	.37370358	.03950482	9.460	.0000	9.60164716
LTMDS	.09422341	.00874381	10.776	.0000	6.95332183
ANO	-.09938093	.01571526	-6.324	.0000	3.00000000
OCPSOL2	-.00854442	.05874694	-.145	.8844	.21800948
OCPSOL3	.71389891	.12894428	5.536	.0000	.02843602
OCPSOL4	-.13355481	.10815416	-1.235	.2169	.05213270
OCPSOL5	.19429862	.06215201	3.126	.0018	.14218009
CLASSEA	-.34686192	.10647878	-3.258	.0011	.05687204
CLASSEB	-.42477118	.05742611	-7.397	.0000	.33649289

Normal exit from iterations. Exit status=0.

```
+-----+
| Zero Altered Poisson      Regression Model
| Normal distribution used for splitting model.
| ZAP term in probability is F[tau x ln LAMBDA]
| Comparison of estimated models
| Pr[0|means]      Number of zeros      Log-likelihood
| Poisson          .18137      Act.= 324 Prd.= 191.3      -1988.37402
| Z.I.Poisson      .29269      Act.= 324 Prd.= 308.8      -1919.64718
| Note, the ZIP log-likelihood is not directly comparable.
| ZIP model with nonzero Q does not encompass the others.
| Vuong statistic for testing ZIP vs. unaltered model is      4.7965
| Distributed as standard normal. A value greater than
| +1.96 favors the zero altered Z.I.Poisson model.
| A value less than -1.96 rejects the ZIP model.
+-----+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-1.89369840	.27720916	-6.831	.0000	
LTMDP	.26428527	.02881326	9.172	.0000	9.60164716
LTMDS	.06652911	.00673713	9.875	.0000	6.95332183

ANEXO 2

ANO		-.07455566	.01245855	-5.984	.0000	3.00000000
OCPSOL2		-.00032664	.04462353	-.007	.9942	.21800948
OCPSOL3		.50597609	.10706189	4.726	.0000	.02843602
OCPSOL4		-.09214280	.09890018	-.932	.3515	.05213270
OCPSOL5		.15578307	.04995053	3.119	.0018	.14218009
CLASSEA		-.25715561	.07586096	-3.390	.0007	.05687204
CLASSEC		-.31014774	.04648744	-6.672	.0000	.33649289
-----+Zero inflation model						
Tau		-1.03098765	.07799419	-13.219	.0000	

ANEXO 2.XXI

MODELO MOP E MOL (ARCOS)


```
-->
ORDERED;Lhs=catC0_2;Rhs=ONE,ANO,LTMDTOT,DENSINT,LLENGTH,OCPSOL2,OCPSOL3
,OCPSOL4,OCPSOL5,DISLOCALS
```

Normal exit from iterations. Exit status=0.

```
+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 03:06:44PM.
| Dependent variable          CATC0_2
| Weighting variable          None
| Number of observations      1980
| Iterations completed        21
| Log likelihood function     -1158.103
| Number of parameters        11
| Info. Criterion: AIC =      1.18091
|   Finite Sample: AIC =      1.18098
| Info. Criterion: BIC =      1.21197
| Info. Criterion:HQIC =      1.19232
| Restricted log likelihood    -1682.965
| McFadden Pseudo R-squared   .3118671
| Chi squared                 1049.722
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
| Underlying probabilities based on Normal
+-----+
```

```
+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq Y Count Freq Y Count Freq
| 0 1260 .636 1 557 .281 2 163 .082
+-----+
```

```
+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+-----+-----+-----+-----+-----+-----+
+-----+Index function for probability
Constant| -9.98359988 | .50470145 | -19.781 | .0000 |
ANO     | -.05602098  | .02227833 | -2.515  | .0119 | 3.000000000
LTMDTOT | .39270880   | .04462056 | 8.801   | .0000 | 9.28075747
DENSINT | 37.5621541  | 6.79899585 | 5.525   | .0000 | .00457041
LLENGTH | 1.10653791  | .04717981 | 23.454  | .0000 | 5.27644440
OCPSOL2 | -.28653365  | .08188388 | -3.499  | .0005 | .22222222
OCPSOL3 | .61367935   | .17534631 | 3.500   | .0005 | .03030303
OCPSOL4 | -.49355151  | .13246010 | -3.726  | .0002 | .05808081
OCPSOL5 | .29893261   | .09259514 | 3.228   | .0012 | .14141414
DISLOCAL| -.18550690  | .06944410 | -2.671  | .0076 | .49747475
+-----+Threshold parameters for index
Mu(1)   | 1.61390324  | .06159125 | 26.203  | .0000 |
```

```
+-----+
+---+
| Cross tabulation of predictions. Row is actual, column is predicted.
| Model = Probit . Prediction is number of the most probable cell.
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
| Actual|Row Sum| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 0 | 1260 | 1101 | 155 | 4 |
| 1 | 557 | 249 | 283 | 25 |
| 2 | 163 | 12 | 99 | 52 |
+-----+-----+-----+-----+-----+-----+-----+
+---+
|Col Sum| 1980 | 1362 | 537 | 81 | 0 | 0 | 0 | 0 | 0 | 0 |
+-----+-----+-----+-----+-----+-----+-----+
0 |
```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
                                     ---+

-->
ORDERED;Lhs=catC0_2;Rhs=ONE,ANO,LTMDTOT,DENSINT,LLENGTH,OCPSOL2,OCPSOL3
,OCPSOL4,OCPSOL5,DISLOCAL;LOGITS$
Normal exit from iterations. Exit status=0.
+-----+
| Ordered Probability Model |
| Maximum Likelihood Estimates |
| Model estimated: Mar 11, 2010 at 03:21:33PM. |
| Dependent variable          CATC0_2 |
| Weighting variable          None |
| Number of observations      1980 |
| Iterations completed        22 |
| Log likelihood function     -1161.609 |
| Number of parameters        11 |
| Info. Criterion: AIC =      1.18445 |
|   Finite Sample: AIC =      1.18452 |
| Info. Criterion: BIC =      1.21551 |
| Info. Criterion:HQIC =      1.19586 |
| Restricted log likelihood    -1682.965 |
| McFadden Pseudo R-squared   .3097842 |
| Chi squared                 1042.712 |
| Degrees of freedom          9 |
| Prob[ChiSqd > value] =      .00000000 |
| Underlying probabilities based on Logistic |
+-----+

+-----+
| Ordered Probability Model |
| Cell frequencies for outcomes |
| Y Count Freq Y Count Freq Y Count Freq |
| 0 1260 .636 1 557 .281 2 163 .082 |
+-----+

+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+-----+-----+-----+-----+-----+-----+
-----+Index function for probability
Constant| -17.7158484 | .93800095 | -18.887 | .0000 |
ANO     | -.09013643 | .03898587 | -2.312 | .0208 | 3.000000000
LTMDTOT | .70138640 | .07994173 | 8.774 | .0000 | 9.28075747
DENSINT | 63.8908109 | 12.0824692 | 5.288 | .0000 | .00457041
LLENGTH | 1.95398705 | .08869088 | 22.031 | .0000 | 5.27644440
OCPSOL2 | -.51344734 | .14566719 | -3.525 | .0004 | .22222222
OCPSOL3 | 1.08306913 | .29529274 | 3.668 | .0002 | .03030303
OCPSOL4 | -.89557560 | .23046369 | -3.886 | .0001 | .05808081
OCPSOL5 | .54225561 | .15871950 | 3.416 | .0006 | .14141414
DISLOCAL| -.29143429 | .12122255 | -2.404 | .0162 | .49747475
-----+Threshold parameters for index
Mu(1)   | 2.83621640 | .11521208 | 24.617 | .0000 |

+-----+
---+
| Cross tabulation of predictions. Row is actual, column is predicted. |
| Model = Logistic . Prediction is number of the most probable cell. |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+
| Actual|Row Sum| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
9 |
+-----+-----+-----+-----+-----+-----+-----+-----+
---+
| 0 | 1260 | 1095 | 160 | 5 |
| 1 | 557 | 245 | 287 | 25 |

```


	2	163	12	99	52								
+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+
---+													
Col Sum	1980	1352	546	82	0	0	0	0	0	0	0	0	0
0													
+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+
---+													

ANEXO 2.XXII
MODELO MML (ARCOS)

Notes: Entropy computed as $\sum(i)\sum(j)P_{fit}(i,j)*\log P_{fit}(i,j)$.
 Normalized entropy is computed against M0.
 Entropy ratio statistic is computed against M0.
 $BIC = 2*craterion - \log(N)*degrees\ of\ freedom$.
 If the model has only constants or if it has no constants,
 the statistics reported here are not useable.

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

		Predicted				
		0	1	2		
Actual	0	1119	140	1		1260
	1	265	256	36		557
	2	15	79	69		163
	Total	1399	475	106		1980

ANEXO 2.XXIII

**MODELO MML E MOP – CLASSES DE ACIDENTES COM VÍTIMAS
(ARCOS)**

Notes: Entropy computed as $\sum(i)\sum(j)P_{fit}(i,j)*\log P_{fit}(i,j)$.
 Normalized entropy is computed against M0.
 Entropy ratio statistic is computed against M0.
 $BIC = 2* criterion - \log(N)*degrees\ of\ freedom$.
 If the model has only constants or if it has no constants,
 the statistics reported here are not useable.

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Predicted				
Actual	0	1	2	Total
0	429	188	52	669
1	167	332	196	695
2	45	193	378	616
Total	641	713	626	1980

-->
ORDERED;Lhs=categor3;Rhs=ONE,ANO,LTMDTOT,DENSINT,LLENGTH,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5,DISLOCAL\$
 Normal exit from iterations. Exit status=0.

```

+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 03:11:09PM.
| Dependent variable          CATEGOR3
| Weighting variable          None
| Number of observations      1980
| Iterations completed        19
| Log likelihood function     -1726.703
| Number of parameters        11
| Info. Criterion: AIC =      1.75526
|   Finite Sample: AIC =      1.75532
| Info. Criterion: BIC =      1.78632
| Info. Criterion:HQIC =      1.76667
| Restricted log likelihood    -2172.779
| McFadden Pseudo R-squared   .2053020
| Chi squared                 892.1516
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .0000000
| Underlying probabilities based on Normal
+-----+

```

```

+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq Y Count Freq Y Count Freq
| 0 669 .337 1 695 .351 2 616 .311
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----+Index function for probability					
Constant	-6.41689690	.35148186	-18.257	.0000	
ANO	-.03067615	.01908807	-1.607	.1080	3.000000000
LTMDTOT	.33197687	.03288072	10.096	.0000	9.28075747
DENSINT	19.0190603	5.48077789	3.470	.0005	.00457041
LLENGTH	.74843627	.03271036	22.881	.0000	5.27644440
OCPSOL2	-.34649741	.06928785	-5.001	.0000	.22222222
OCPSOL3	-.06805486	.15662434	-.435	.6639	.03030303
OCPSOL4	-.08843591	.11652199	-.759	.4479	.05808081
OCPSOL5	.27252339	.08261641	3.299	.0010	.14141414
DISLOCAL	-.04882792	.05898639	-.828	.4078	.49747475

```

-----+Threshold parameters for index
Mu(1)   |      1.22666497      .04020277      30.512      .0000

+-----+
---+
|      Cross tabulation of predictions. Row is actual, column is predicted.
|      Model = Probit      .      Prediction is number of the most probable cell.
+-----+
---+
| Actual|Row Sum|  0  |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |
9 |-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      0|      669|    406|    222|     41|
|      1|      695|    142|    365|    188|
|      2|      616|     30|    224|    362|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+
| Col Sum|    1980|    578|    811|    591|     0|     0|     0|     0|     0|
0 |-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
---+

```


ANEXO 2.XXIV
MODELO MOP E MOL (NÓS)


```
-->
ORDERED;Lhs=catC0_2;Rhs=ONE,ANO,LTMDP,LTMDS,OCPSOL2,OCPSOL3,OCPSOL4,OCPSO
L5
```

,CLASSEA,CLASSECS

Normal exit from iterations. Exit status=0.

```
+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 03:15:45PM.
| Dependent variable          CATC0_2
| Weighting variable          None
| Number of observations       1055
| Iterations completed        16
| Log likelihood function     -731.0597
| Number of parameters        11
| Info. Criterion: AIC =      1.40675
|   Finite Sample: AIC =      1.40699
| Info. Criterion: BIC =      1.45848
| Info. Criterion:HQIC =      1.42636
| Restricted log likelihood   -814.3763
| McFadden Pseudo R-squared  .1023073
| Chi squared                 166.6334
| Degrees of freedom          9
| Prob[ChiSqd > value] =      .00000000
| Underlying probabilities based on Normal
+-----+
```

```
+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq Y Count Freq Y Count Freq
| 0 748 .709 1 225 .213 2 82 .077
+-----+
```

```
+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+-----+-----+-----+-----+-----+-----+
+-----+Index function for probability
Constant| -4.31374795 | .68304940 | -6.315 | .0000 |
ANO | -.11484264 | .02933277 | -3.915 | .0001 | 3.000000000
LTMDP | .34964978 | .06966376 | 5.019 | .0000 | 9.60164716
LTMDS | .11419112 | .01580257 | 7.226 | .0000 | 6.95332183
OCPSOL2 | .01480862 | .10822474 | .137 | .8912 | .21800948
OCPSOL3 | .84499427 | .24149320 | 3.499 | .0005 | .02843602
OCPSOL4 | -.10094665 | .18740522 | -.539 | .5901 | .05213270
OCPSOL5 | .30617981 | .11745832 | 2.607 | .0091 | .14218009
CLASSEA | -.29029246 | .18510526 | -1.568 | .1168 | .05687204
CLASSECS | -.56046695 | .10310870 | -5.436 | .0000 | .33649289
+-----+Threshold parameters for index
Mu(1) | .98880327 | .05952188 | 16.612 | .0000
```

```
+-----+
+---+
| Cross tabulation of predictions. Row is actual, column is predicted.
| Model = Probit . Prediction is number of the most probable cell.
+-----+
+---+
| Actual|Row Sum| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
+-----+
| 9 |
+-----+
+---+
| 0 | 748 | 739 | 9 | 0 |
| 1 | 225 | 217 | 8 | 0 |
| 2 | 82 | 74 | 7 | 1 |
+-----+
+---+
```

```

|Col Sum| 1055| 1030| 24| 1| 0| 0| 0| 0| 0| 0|
0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+
-->
ORDERED; Lhs=catC0_2; Rhs=ONE, ANO, LTMDP, LTMD5, OCPSOL2, OCPSOL3, OCPSOL4, OCPSO
L5
, CLASSEA, CLASSEC; logit$
Normal exit from iterations. Exit status=0.
+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 03:18:09PM.
| Dependent variable CATC0_2
| Weighting variable None
| Number of observations 1055
| Iterations completed 17
| Log likelihood function -732.1338
| Number of parameters 11
| Info. Criterion: AIC = 1.40878
| Finite Sample: AIC = 1.40902
| Info. Criterion: BIC = 1.46051
| Info. Criterion: HQIC = 1.42839
| Restricted log likelihood -814.3763
| McFadden Pseudo R-squared .1009884
| Chi squared 164.4851
| Degrees of freedom 9
| Prob[ChiSqd > value] = .0000000
| Underlying probabilities based on Logistic
+-----+

+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq Y Count Freq Y Count Freq
| 0 748 .709 1 225 .213 2 82 .077
+-----+

+-----+-----+-----+-----+-----+-----+
|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+-----+-----+-----+-----+-----+-----+
-----+Index function for probability
Constant| -7.80399610 1.23889676 -6.299 .0000
ANO | -.17833865 .05051710 -3.530 .0004 3.000000000
LTMDP | .62960829 .12563598 5.011 .0000 9.60164716
LTMD5 | .20565304 .02965056 6.936 .0000 6.95332183
OCPSOL2 | .03851118 .18956087 .203 .8390 .21800948
OCPSOL3 | 1.50401206 .40810010 3.685 .0002 .02843602
OCPSOL4 | -.10309902 .31847019 -.324 .7461 .05213270
OCPSOL5 | .52597035 .20175900 2.607 .0091 .14218009
CLASSEA | -.52598180 .32414829 -1.623 .1047 .05687204
CLASSEC | -.94317500 .18250964 -5.168 .0000 .33649289
-----+Threshold parameters for index
Mu(1) | 1.75520212 .11192608 15.682 .0000

+-----+
---+
| Cross tabulation of predictions. Row is actual, column is predicted.
| Model = Logistic . Prediction is number of the most probable cell.
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+
| Actual|Row Sum| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
9 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+
| 0 | 748 | 736 | 12 | 0 |
| 1 | 225 | 210 | 15 | 0 |

```


RESULTADOS DA APLICAÇÃO DOS MODELOS

[illegible]

ANEXO 2.XXV
MODELO MML (NÓS)

-->

```
MLOGIT;Lhs=catC0_2;Rhs=ONE,ANO,LTMDP,LTMDS,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5
```

, CLASSEA, CLASSECS\$

```
Normal exit from iterations. Exit status=0.
```

```

+-----+
| Multinomial Logit Model  

| Maximum Likelihood Estimates  

| Model estimated: Mar 11, 2010 at 05:35:26PM.  

| Dependent variable CATC0_2  

| Weighting variable None  

| Number of observations 1055  

| Iterations completed 31  

| Log likelihood function -718.9424  

| Number of parameters 20  

| Info. Criterion: AIC = 1.40084  

| Finite Sample: AIC = 1.40161  

| Info. Criterion: BIC = 1.49489  

| Info. Criterion:HQIC = 1.43649  

| Restricted log likelihood -814.3763  

| McFadden Pseudo R-squared .1171866  

| Chi squared 190.8679  

| Degrees of freedom 18  

| Prob[ChiSqd > value] = .0000000

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X

-----+Characteristics in numerator of Prob[Y = 1]					
Constant	-6.87777676	1.38505575	-4.966	.0000	
ANO	-.08741155	.05709282	-1.531	.1258	3.00000000
LTMDP	.47850999	.14077201	3.399	.0007	9.60164716
LTMDS	.19576147	.03302739	5.927	.0000	6.95332183
OCPSOL2	-.06552902	.21816011	-.300	.7639	.21800948
OCPSOL3	1.22440403	.47071549	2.601	.0093	.02843602
OCPSOL4	.45540914	.33291055	1.368	.1713	.05213270
OCPSOL5	.56642976	.22768250	2.488	.0129	.14218009
CLASSEA	-.83553317	.39896426	-2.094	.0362	.05687204
CLASSEC	-.71565839	.19811472	-3.612	.0003	.33649289

	Characteristics	in numerator of	Prob[Y = 2]		
Constant	-12.8008402	2.45053084	-5.224	.0000	
ANO	-.36784470	.09019655	-4.078	.0000	3.00000000
LTMDP	1.03130286	.24671446	4.180	.0000	9.60164716
LTMDS	.24332200	.05888428	4.132	.0000	6.95332183
OCPSOL2	.09001787	.31811465	.283	.7772	.21800948
OCPSOL3	2.17052395	.67992783	3.192	.0014	.02843602
OCPSOL4	-30.1547130	.150588D+07	.000	1.0000	.05213270
OCPSOL5	.46776426	.34923362	1.339	.1804	.14218009
CLASSEA	-.18012165	.48987240	-.368	.7131	.05687204
CLASSEC	-1.67693878	.40892801	-4.101	.0000	.33649289

Information Statistics for Discrete Choice Model.								
	M=Model			MC=Constants Only			M0=No Model	
Criterion F (log L)	-718.94237			-814.37633			-1159.03596	
LR Statistic vs. MC	190.86792			.00000			.00000	
Degrees of Freedom	18.00000			.00000			.00000	
Prob. Value for LR	.00000			.00000			.00000	
Entropy for probs.	718.94237			814.37633			1159.03596	
Normalized Entropy	.62029			.70263			1.00000	
Entropy Ratio Stat.	880.18718			689.31926			.00000	
Bayes Info Criterion	1.48169			1.66261			2.31600	
BIC(no model) - BIC	.83430			.65338			.00000	
Pseudo R-squared	.11719			.00000			.00000	
Pct. Correct Pred.	71.37441			.00000			33.33333	
Means:	y=0	y=1	y=2	y=3	y=4	y=5	y=6	y>=7
Outcome	.7090	.2133	.0777	.0000	.0000	.0000	.0000	.0000

```

| Pred.Pr      .7090  .2133  .0777  .0000  .0000  .0000  .0000  .0000 |
| Notes: Entropy computed as Sum(i)Sum(j)Pfit(i,j)*logPfit(i,j). |
|         Normalized entropy is computed against M0. |
|         Entropy ratio statistic is computed against M0. |
|         BIC = 2*criterion - log(N)*degrees of freedom. |
|         If the model has only constants or if it has no constants, |
|         the statistics reported here are not useable. |
+-----+

```

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Actual	Predicted			Total
	0	1	2	
0	737	11	0	748
1	207	13	5	225
2	70	9	3	82
Total				1055

ANEXO 2.XXVI

**MODELO MML E MOP – CLASSES DE ACIDENTES COM VÍTIMAS
(NÓS)**

-->

```
MLOGIT;Lhs=catVit;Rhs=ONE,ANO,LTMDP,LTMS,OCPSOL2,OCPSOL3,OCPSOL4,OCPSOL5
,CLASSEA,CLASEC$
```

```
Normal exit from iterations. Exit status=0.
```

Multinomial Logit Model					
Maximum Likelihood Estimates					
Model estimated: Mar 11, 2010 at 05:37:11PM.					
Dependent variable		CATVIT			
Weighting variable		None			
Number of observations		1055			
Iterations completed		5			
Log likelihood function		-1069.710			
Number of parameters		20			
Info. Criterion: AIC =		2.06580			
Finite Sample: AIC =		2.06657			
Info. Criterion: BIC =		2.15985			
Info. Criterion:HQIC =		2.10145			
Restricted log likelihood		-1146.641			
McFadden Pseudo R-squared		.0670926			
Chi squared		153.8623			
Degrees of freedom		18			
Prob[ChiSqd > value] =		.0000000			

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----+Characteristics in numerator of Prob[Y = 1]					
Constant	-4.73460126	1.18045427	-4.011	.0001	
ANO	-.10865596	.05523760	-1.967	.0492	3.00000000
LTMDP	.48856945	.12164190	4.016	.0001	9.60164716
LTMDS	.11904145	.02289035	5.201	.0000	6.95332183
OCPSOL2	-.09100048	.19549743	-.465	.6416	.21800948
OCPSOL3	1.55373456	.52373122	2.967	.0030	.02843602
OCPSOL4	.14516534	.34279819	.423	.6720	.05213270
OCPSOL5	.50423438	.26682365	1.890	.0588	.14218009
CLASSEA	-.99040319	.33040461	-2.998	.0027	.05687204
CLASSEC	-.31129449	.17313828	-1.798	.0722	.33649289
-----+Characteristics in numerator of Prob[Y = 2]					
Constant	-8.28670138	1.39325870	-5.948	.0000	
ANO	-.14584860	.06130778	-2.379	.0174	3.00000000
LTMDP	.78036169	.14225633	5.486	.0000	9.60164716
LTMDS	.18204004	.02840466	6.409	.0000	6.95332183
OCPSOL2	.21687486	.21430522	1.012	.3115	.21800948
OCPSOL3	.99430920	.66107275	1.504	.1326	.02843602
OCPSOL4	-.22608040	.41393982	-.546	.5850	.05213270
OCPSOL5	.91279134	.27720200	3.293	.0010	.14218009
CLASSEA	-1.38061155	.39508108	-3.495	.0005	.05687204
CLASSEC	-.55611697	.19826835	-2.805	.0050	.33649289

Information Statistics for Discrete Choice Model.								
M=Model				MC=Constants Only		M0=No Model		
Criterion F (log L)	-1069.70951		-1146.64066		-1159.03596			
LR Statistic vs. MC	153.86229		.00000		.00000			
Degrees of Freedom	18.00000		.00000		.00000			
Prob. Value for LR	.00000		.00000		.00000			
Entropy for probs.	1069.70951		1146.64066		1159.03596			
Normalized Entropy	.92293		.98931		1.00000			
Entropy Ratio Stat.	178.65290		24.79061		.00000			
Bayes Info Criterion	2.14666		2.29250		2.31600			
BIC(no model) - BIC	.16934		.02350		.00000			
Pseudo R-squared	.06709		.00000		.00000			
Pct. Correct Pred.	48.62559		.00000		33.33333			
Means:	y=0	y=1	y=2	y=3	y=4	y=5	y=6	y>=7
Outcome	.3071	.4057	.2872	.0000	.0000	.0000	.0000	.0000
Pred.Pr	.3071	.4057	.2872	.0000	.0000	.0000	.0000	.0000

Notes: Entropy computed as $\sum(i)\sum(j)P_{fit}(i,j)*\log P_{fit}(i,j)$.
 Normalized entropy is computed against M0.
 Entropy ratio statistic is computed against M0.
 $BIC = 2* criterion - \log(N)*degrees\ of\ freedom$.
 If the model has only constants or if it has no constants,
 the statistics reported here are not useable.

Frequencies of actual & predicted outcomes
 Predicted outcome has maximum probability.

Predicted				
Actual	0	1	2	Total
0	151	155	18	324
1	88	280	60	428
2	45	176	82	303
Total	284	611	160	1055

-->
ORDERED; Lhs=catVit; Rhs=ONE, ANO, LTMDP, LTMS, OCPSOL2, OCPSOL3, OCPSOL4, OCPSOL5
, CLASSEA, CLASSECS

Normal exit from iterations. Exit status=0.

```

+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Mar 11, 2010 at 05:37:11PM.
| Dependent variable          CATVIT
| Weighting variable          None
| Number of observations       1055
| Iterations completed        15
| Log likelihood function      -1076.599
| Number of parameters         11
| Info. Criterion: AIC =       2.06180
|   Finite Sample: AIC =       2.06204
| Info. Criterion: BIC =       2.11353
| Info. Criterion: HQIC =      2.08141
| Restricted log likelihood     -1146.641
| McFadden Pseudo R-squared    .0610844
| Chi squared                  140.0838
| Degrees of freedom           9
| Prob[ChiSq > value] =        .0000000
| Underlying probabilities based on Normal
+-----+

```

```

+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq Y Count Freq Y Count Freq
| 0   324 .307  1   428 .405  2   303 .287
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----+Index function for probability					
Constant	-2.88080930	.54222516	-5.313	.0000	
ANO	-.05892236	.02491365	-2.365	.0180	3.000000000
LTMDP	.32303033	.05573676	5.796	.0000	9.60164716
LTMS	.07733817	.01119144	6.910	.0000	6.95332183
OCPSOL2	.07761481	.08977995	.865	.3873	.21800948
OCPSOL3	.40674892	.21560091	1.887	.0592	.02843602
OCPSOL4	-.07826394	.16038688	-.488	.6256	.05213270
OCPSOL5	.34969860	.10695188	3.270	.0011	.14218009
CLASSEA	-.58750755	.16007453	-3.670	.0002	.05687204

```

CLASSEC |    -.23616187    .08093356    -2.918    .0035    .33649289
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Mu(1)   |    1.15754495    .04754687    24.345    .0000
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
| Cross tabulation of predictions. Row is actual, column is predicted.
|
| Model = Probit      . Prediction is number of the most probable cell.
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
| Actual|Row Sum|  0  |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |
9 |-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
|      0|    324|   124|   181|    19|
|      1|    428|    74|   279|    75|
|      2|    303|    39|   173|    91|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
| Col Sum|   1055|   237|   633|   185|    0|    0|    0|    0|    0|    0|
0 |-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+

```

