



LABORATÓRIO NACIONAL
DE ENGENHARIA CIVIL

DEPARTAMENTO DE TRANSPORTES
Núcleo de Planeamento, Tráfego e Segurança

Proc. 0703/11/17772

MODELOS DE ESTIMATIVA DE FREQUÊNCIA DE ACIDENTES E VÍTIMAS PARA DADOS BIDIRECCIONAIS DE AUTO-ESTRADAS PORTUGUESAS – UMA ABORDAGEM BAYESIANA

Lisboa • Dezembro de 2009

I&D TRANSPORTES

RELATÓRIO 5/2010 – NPTS

SUMÁRIO

O estudo referente ao presente documento consiste na selecção, ajuste e validação de modelos de previsão de frequências de acidentes, mortes, feridos graves e mortos e feridos graves empregando métodos estatísticos Bayesianos. Os dados analisados foram registados em trechos (bidireccionais) de auto-estradas da rede portuguesa e medidos num período de cinco anos compreendido entre 1999 e 2003, inclusive. Nove variáveis foram medidas juntamente com os eventos considerados, nomeadamente o tráfego médio diário anual, comprimento do trecho, número de faixas de rodagem, largura de vias, existência de vias adicionais para veículos lentos, tipo e largura das bermas e do separador central. Os dados foram subsequentemente divididos em subgrupos constituídos por todos os trechos nos quais as variáveis com valores omissos foram excluídas e por conjuntos de trechos sem valores omissos. Numa segunda fase foram considerados unicamente trechos cujos valores de tráfego médio diário anual fossem maiores que 5000 veículos.

Os modelos estatísticos considerados incluíram o modelo de Poisson e modelos que assumem uma sobredispersão dos eventos baseados na distribuição de Poisson, isto é, os modelos Poisson-Gama e Poisson Log-Normal. Todos os modelos foram ajustados usando métodos Bayesianos. Foram assumidas distribuições *a priori* não informativas para os coeficientes dos parâmetros dos modelos. Consequentemente, as estimativas Bayesianas para esses coeficientes coincidem com os resultados obtidos através da abordagem *clássica*. No entanto, mesmo no caso não informativo, as simulações obtidas *a posteriori* são úteis para fazer inferências preditivas e para as verificações do ajuste dos modelos, que foram efectuadas.

O critério de informação da desviância (DIC) foi usado para selecção de variáveis e foram feitas verificações preditivas *a posteriori* para avaliar os ajustes dos vários modelos assim como análises exaustivas para garantir que os algoritmos de Monte Carlo via cadeias de Markov (MCMC) convergiam e, por consequência, se obtinham distribuições *a posteriori* estacionárias. Calcularam-se os números esperados para cada um dos diversos eventos, em cada subgrupo de dados, para valores médios de tráfego médio diário anual e de comprimento de trecho e obtiveram-se distribuições *a posteriori* para as estimativas dos coeficientes dos modelos obtidos, assim como intervalos de credibilidade para essas estimativas.

ABSTRACT

The study described in this document consists on the selection, fit and checking of accident, fatalities, killed and serious injured and serious injured prediction models obtained using Bayesian statistical techniques. The data analysed were measured on Portuguese motorways segments for a period of five years between 1999 and 2003, inclusive. Nine variables were measured together with the counting events chosen to be of interest, namely, average annual daily traffic (*AADT*), length of the segment, number of lanes, lane widths, existence of additional lanes and type and width of the shoulders and medians. The data were then divided into subgroups according to whether the variables with missing values were not taken into account for the analysis and considering only the motorway segments without missing values. In a second stage each subset was subdivided so that it was considered only the motorway sections with *AADT* values greater than 5000 vehicles were considered.

The statistical models obtained included the Poisson model and models which take the overdispersion into account but are based on the Poisson distribution, namely, the Poisson-Gamma and Poisson Log-Normal. All models were adjusted using Bayesian techniques. The *a priori* distribution for the model parameters were assumed to be non-informative. Therefore, the Bayesian estimates for those parameters were very similar with the ones obtained by the *classical* statistical methods. However, even in the non-informative case, the posterior simulations are useful for predictive inference and model checking.

The Deviance Information Criterion (DIC) was used for variable selection. *Posterior* predictive checking was performed for model validation as well as for checking the convergence of Markov chains via Monte Carlo methods (MCMC) to make sure that the obtained *posterior* distributions were stationary. The expected numbers for each of the events considered were calculated based on the average profiles of *AADT* and length of the motorway segments. *Posterior* distributions and credibility intervals were obtained for the coefficient estimates of each model obtained.

ÍNDICE

1 	INTRODUÇÃO.....	1
2 	DESCRIÇÃO DOS DADOS.....	4
2.1	Os dados das auto-estradas	4
3 	A ABORDAGEM BAYESIANA.....	6
3.1	Descrição Sucinta dos Métodos de Monte Carlo.....	7
3.2	Cadeias de Markov	8
3.3	Avaliação da Convergência da Cadeia de Markov.....	9
3.3.1	Trace plots, Erros de Monte Carlo e Diagnóstico de Autocorrelação	10
3.3.2	Diagnóstico de Gelman-Rubin.....	12
3.4	Validação e Selecção de Modelos	13
3.5	Critério DIC.....	14
3.6	Modelos Lineares Generalizados	15
3.6.1	Modelos de Regressão de Poisson.....	16
3.6.2	Modelo de Poisson-Gama.....	17
3.6.3	Modelo de Poisson Log-Normal	18
4 	MODELAÇÃO DOS DADOS EXCLUINDO AS VARIÁVEIS COM VALORES OMISSOS.....	19
4.1	Modelação do Número de Acidentes.....	19
4.1.1	Modelo de Poisson	19
4.1.1.1	Validação do Modelo.....	24
4.1.2	Modelo de Poisson-Gama	26
4.1.2.1	Validação do Modelo.....	30
4.1.3	Modelo de Poisson Log-Normal	32
4.1.3.1	Validação do Modelo.....	35
4.2	Modelação do Número de Mortos e Feridos Graves	37
4.2.1	Modelo de Poisson	37
4.2.1.1	Validação do Modelo.....	41
4.2.2	Modelo de Poisson-Gama.....	43
4.2.2.1	Validação do Modelo.....	46
4.2.3	Modelo de Poisson Log-Normal	48
4.2.3.1	Validação do Modelo.....	50
4.3	Discussão.....	53

5 | MODELAÇÃO DOS DADOS EXCLUINDO OS TRECHOS COM VALORES OMISSOS..... 54

5.1	Modelação do Número de Acidentes.....	54
5.1.1	Modelo de Poisson	54
5.1.1.1	Validação do Modelo.....	58
5.1.2	Modelo de Poisson-Gama	60
5.1.2.1	Validação do Modelo.....	64
5.1.3	Modelo de Poisson Log-Normal	66
5.1.3.1	Validação do Modelo.....	70
5.2	Modelação do Número de Mortos e Feridos Graves	72
5.2.1	Modelo de Poisson	73
5.2.1.1	Validação do Modelo.....	76
5.2.2	Modelo de Poisson-Gama	78
5.2.2.1	Validação do Modelo.....	81
5.2.3	Modelo de Poisson Log-Normal	83
5.2.3.1	Validação do Modelo.....	87
5.3	Discussão.....	90

6 | MODELAÇÃO DOS DADOS CUJOS TRECHOS TENHAM VALORES DE AADT SUPERIORES A 5000 E EXCLUINDO AS VARIÁVEIS COM VALORES OMISSOS..... 91

6.1	Modelação do Número de Acidentes.....	91
6.1.1	Modelo de Poisson	91
6.1.1.1	Validação do Modelo.....	96
6.1.2	Modelo de Poisson-Gama	98
6.1.2.1	Validação do Modelo.....	101
6.1.3	Modelo de Poisson Log-Normal	103
6.1.3.1	Validação do Modelo.....	106
6.2	Modelação do Número de Mortos.....	108
6.2.1	Modelo de Poisson	108
6.2.1.1	Validação do Modelo.....	111
6.2.2	Modelo de Poisson-Gama	113
6.2.2.1	Validação do Modelo.....	116
6.2.3	Modelo de Poisson Log-Normal	118
6.2.3.1	Validação do Modelo.....	120
6.3	Modelação do Número de Feridos Graves	122
6.3.1	Modelo de Poisson	122
6.3.1.1	Validação do Modelo.....	125
6.3.2	Modelo de Poisson-Gama	128
6.3.2.1	Validação do Modelo.....	130
6.3.3	Modelo de Poisson Log-Normal	132
6.3.3.1	Validação do Modelo.....	135
6.4	Modelação do Número de Mortos e Feridos Graves	137
6.4.1	Modelo de Poisson	137
6.4.1.1	Validação do Modelo.....	141
6.4.2	Modelo de Poisson-Gama	143
6.4.2.1	Validação do Modelo.....	145
6.4.3	Modelo de Poisson Log-Normal	147
6.4.3.1	Validação do Modelo.....	150
6.5	Discussão.....	152

7 | MODELAÇÃO DOS DADOS CONSTITUÍDOS POR TRECHOS SEM VALORES OMISSOS E COM VALORES DE AADT SUPERIORES 5000 153

7.1	Modelação do Número de Acidentes.....	153
7.1.1	Modelo de Poisson	153
7.1.1.1	Validação do Modelo.....	157
7.1.2	Modelo de Poisson-Gama	160
7.1.2.1	Validação do Modelo.....	163
7.1.3	Modelo de Poisson Log-Normal	165
7.1.3.1	Validação do Modelo.....	168
7.2	Modelação do Número de Mortos.....	171
7.2.1	Modelo de Poisson	171
7.2.1.1	Validação do Modelo.....	174
7.2.2	Modelo de Poisson-Gama	176
7.2.2.1	Validação do Modelo.....	178
7.2.3	Modelo de Poisson Log-Normal	180
7.2.3.1	Validação do Modelo.....	182
7.3	Modelação do Número de Feridos Graves	184
7.3.1	Modelo de Poisson	184
7.3.1.1	Validação do Modelo.....	188
7.3.2	Modelo de Poisson-Gama	190
7.3.2.1	Validação do Modelo.....	193
7.3.3	Modelo de Poisson Log-Normal	196
7.3.3.1	Validação do Modelo.....	199
7.4	Modelação do Número de Mortos e Feridos Graves	201
7.4.1	Modelo de Poisson	201
7.4.1.1	Validação do Modelo.....	204
7.4.2	Modelo de Poisson-Gama	206
7.4.2.1	Validação do Modelo.....	209
7.4.3	Modelo de Poisson Log-Normal	211
7.4.3.1	Validação do Modelo.....	214
7.5	Discussão.....	216

8 | CONCLUSÕES 217

9 | BIBLIOGRAFIA 221

ANEXOS 225

ANEXO I – Excerto do código de <i>WinBUGS</i> onde a distribuição dos eventos foi assumida como a distribuição de Poisson.....	227
---	------------

ANEXO II – Excerto do código de <i>WinBUGS</i> onde a distribuição dos eventos foi assumida como a distribuição Poisson-Gama.....	229
--	------------

ANEXO III – Excerto do código de <i>WinBUGS</i> onde a distribuição dos eventos foi assumida como a distribuição Poisson Log-Normal.....	231
---	------------

ÍNDICE DE QUADROS

Quadro 1 Número de trechos de auto-estrada, a soma dos seus comprimentos, volume de tráfego, somas do número de acidentes, de mortes e feridos graves ocorridos num período de cinco anos (1999 a 2003) para cada conjunto de dados analisados, assim como os capítulos do relatório a que se refere a análise de cada conjunto de dados.....	5
Quadro 2 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo as variáveis com valores omissos.	22
Quadro 3 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 2.	23
Quadro 4 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 2.....	24
Quadro 5 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo as variáveis com valores omissos.....	28
Quadro 6 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 5.	29
Quadro 7 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 5.....	29
Quadro 8 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo as variáveis com valores omissos.....	33
Quadro 9 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 8.	34
Quadro 10 Médias <i>a posteriori</i> e desvios padrões para o número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 8.....	34
Quadro 11 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo as variáveis com valores omissos.....	39
Quadro 12 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 11.	40
Quadro 13 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 11.	41
Quadro 14 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo as variáveis com valores omissos.....	45
Quadro 15 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 14.	45
Quadro 16 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 14.	46
Quadro 17 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo as variáveis com valores omissos.	49
Quadro 18 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 17.	49

Quadro 19 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 17.	50
Quadro 20 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo os trechos com valores omissos.	56
Quadro 21 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 20.	57
Quadro 22 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 20.....	57
Quadro 23 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo os trechos com valores omissos.....	62
Quadro 24 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 23.	63
Quadro 25 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 23.....	63
Quadro 26 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo os trechos com valores omissos.....	68
Quadro 27 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 26	69
Quadro 28 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 26.....	70
Quadro 29 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo os trechos com valores omissos.....	75
Quadro 30 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 29.	75
Quadro 31 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 29.	76
Quadro 32 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo os trechos com valores omissos.....	80
Quadro 33 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 32.	80
Quadro 34 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 32.	81
Quadro 35 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo os trechos com valores omissos.....	85
Quadro 36 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 35.	86
Quadro 37 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 35.	87

Quadro 38	Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	94
Quadro 39	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 38.	95
Quadro 40	Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 38.	96
Quadro 41	Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	100
Quadro 42	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 41.	100
Quadro 43	Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 41.	101
Quadro 44	Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	105
Quadro 45	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 44.	105
Quadro 46	Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 44.	106
Quadro 47	Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	110
Quadro 48	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 47.	110
Quadro 49	Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 47.	111
Quadro 50	Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	115
Quadro 51	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 50.	115
Quadro 52	Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 50.	115
Quadro 53	Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	119
Quadro 54	Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 53.	119
Quadro 55	Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 53.	120
Quadro 56	Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	124

Quadro 57 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 56.	124
Quadro 58 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 56.....	125
Quadro 59 Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	129
Quadro 60 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 59.	129
Quadro 61 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 59.....	130
Quadro 62 Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	134
Quadro 63 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 62.	134
Quadro 64 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 62.....	135
Quadro 65 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	139
Quadro 66 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 65.	139
Quadro 67 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 65.	140
Quadro 68 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	144
Quadro 69 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 68.	144
Quadro 70 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 68.	145
Quadro 71 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	148
Quadro 72 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 71.	149
Quadro 73 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 71.	149
Quadro 74 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	155
Quadro 75 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 74.	156

Quadro 76 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 74.....	157
Quadro 77 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.....	161
Quadro 78 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 77.	161
Quadro 79 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 77.....	162
Quadro 80 Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	166
Quadro 81 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 80.	167
Quadro 82 Médias <i>a posteriori</i> e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 80.....	168
Quadro 83 Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	172
Quadro 84 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 83.	173
Quadro 85 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 83.....	174
Quadro 86 Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.....	177
Quadro 87 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 86.	178
Quadro 88 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 86.....	178
Quadro 89 Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.....	182
Quadro 90 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 89.	182
Quadro 91 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 89.....	182
Quadro 92 Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.....	186
Quadro 93 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 92.	187
Quadro 94 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 92.....	188

Quadro 95 Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	192
Quadro 96 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 95.	192
Quadro 97 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 95.	193
Quadro 98 Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	197
Quadro 99 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 98.	198
Quadro 100 Médias <i>a posteriori</i> e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 98.	198
Quadro 101 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	202
Quadro 102 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 101.	203
Quadro 103 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 101.	204
Quadro 104 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	207
Quadro 105 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 104.	208
Quadro 106 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 104.	208
Quadro 107 Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.	212
Quadro 108 Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 107.	213
Quadro 109 Médias <i>a posteriori</i> e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 107.	213

ÍNDICE DE FIGURAS

Figura 1 Exemplo de <i>trace plots</i> de duas cadeias de Markov, indicadas a azul e a vermelho, que aparentam convergir (gráfico superior) e de duas cadeias de Markov que não convergem (gráfico inferior).	11
Figura 2 Exemplos de gráficos de autocorrelação para parâmetros com (gráfico superior) e sem (gráfico inferior) autocorrelação.	12
Figura 3 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.....	20
Figura 4 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.....	20
Figura 5 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.....	21
Figura 6 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	23
Figura 7 Histograma de <i>Nac</i> (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidas da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos. 24	24
Figura 8 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão de 3000 conjuntos de dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	25
Figura 9 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	26
Figura 10 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	27
Figura 11 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	27
Figura 12 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	28
Figura 13 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson-Gama ajustado a <i>Nac</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	29
Figura 14 Histograma de <i>Nac</i> (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.	30
Figura 15 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	31

Figura 16 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	31
Figura 17 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	32
Figura 18 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	33
Figura 19 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>Nac</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	33
Figura 20 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>Nac</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	34
Figura 21 Histograma de <i>Nac</i> (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.	35
Figura 22 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	36
Figura 23 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	36
Figura 24 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	37
Figura 25 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	38
Figura 26 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	38
Figura 27 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	40
Figura 28 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.	42
Figura 29 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	42
Figura 30 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	43

Figura 31 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	44
Figura 32 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	44
Figura 33 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	44
Figura 34 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NKSI</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	45
Figura 35 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.	46
Figura 36 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	47
Figura 37 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	47
Figura 38 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	48
Figura 39 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	48
Figura 40 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo as variáveis com valores omissos.	49
Figura 41 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	50
Figura 42 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.	51
Figura 43 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	52
Figura 44 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	52
Figura 45 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>Nac</i> obtido do conjunto de dados excluindo os trechos com valores omissos.	55

Figura 46 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	55
Figura 47 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	56
Figura 48 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a Nac (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.	57
Figura 49 Histograma de Nac (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados excluindo os trechos com valores omissos.	58
Figura 50 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson (3000 simulações). As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	59
Figura 51 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	60
Figura 52 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	61
Figura 53 Gráficos das autocorrelações dos seis parâmetros β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos..	61
Figura 54 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	62
Figura 55 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson-Gama ajustado a Nac (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.	63
Figura 56 Histograma de Nac (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson-Gama ajustado ao conjunto de dados excluindo os trechos com valores omissos.....	64
Figura 57 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	65
Figura 58 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 18000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida dos dados replicados do modelo Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	65
Figura 59 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	67
Figura 60 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	67
Figura 61 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.....	68

Figura 62 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.	69
Figura 63 Histograma de Nac (no canto superior esquerdo) e dezanove réplicas (" $Nac.rep$ ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados excluindo os trechos com valores omissos.....	71
Figura 64 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	71
Figura 65 Razão entre a variância e a média de Nac (recta vertical) comparada com 18000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida dos dados replicados do modelo Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	72
Figura 66 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.	73
Figura 67 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.	74
Figura 68 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.	74
Figura 69 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a $NKSI$ (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.	76
Figura 70 Histograma de $NKSI$ (no canto superior esquerdo) e dezanove réplicas (" $NKSI.rep$ ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados excluindo os trechos com valores omissos.....	77
Figura 71 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	77
Figura 72 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	78
Figura 73 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.....	79
Figura 74 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.....	79
Figura 75 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a $NKSI$ obtido do conjunto de dados excluindo os trechos com valores omissos.....	80
Figura 76 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson-Gama ajustado a $NKSI$ (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.	81

Figura 77 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados excluindo os trechos com valores omissos.	82
Figura 78 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	82
Figura 79 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	83
Figura 80 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo os trechos com valores omissos.	84
Figura 81 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo os trechos com valores omissos.	84
Figura 82 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados excluindo os trechos com valores omissos.	85
Figura 83 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> (obtido do conjunto de dados excluindo os trechos com valores omissos) resultantes da simulação de três cadeias de Markov.	86
Figura 84 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados excluindo os trechos com valores omissos.	88
Figura 85 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	89
Figura 86 Razão entre a variância e a média de <i>NKSI</i> (recta vertical) comparada com 18000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida dos dados replicados do modelo Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	89
Figura 87 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>Nac</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	92
Figura 88 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	92
Figura 89 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	93
Figura 90 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>Nac</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	95

Figura 91 Histograma de N_{ac} (no canto superior esquerdo) e dezanove réplicas (“ $N_{ac.rep}$ ”) dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	97
Figura 92 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	98
Figura 93 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	98
Figura 94 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	99
Figura 95 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	99
Figura 96 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	100
Figura 97 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} (obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	101
Figura 98 Histograma de N_{ac} (no canto superior esquerdo) e dezanove réplicas (“ $N_{ac.rep}$ ”) dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	102
Figura 99 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	102
Figura 100 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	103
Figura 101 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	104
Figura 102 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	104
Figura 103 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.	105

Figura 104 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>Nac</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos)resultantes da simulação de três cadeias de Markov.....	106
Figura 105 Histograma de <i>Nac</i> (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	107
Figura 106 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	107
Figura 107 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	108
Figura 108 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	109
Figura 109 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	109
Figura 110 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	110
Figura 111 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos)resultantes da simulação de três cadeias de Markov.	111
Figura 112 Histograma de <i>NF</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NF.rep</i> ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	112
Figura 113 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	112
Figura 114 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	113
Figura 115 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	114
Figura 116 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	114
Figura 117 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	114

Figura 118 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NF</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos)resultantes da simulação de três cadeias de Markov.....	115
Figura 119 Histograma de <i>NF</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NF.rep</i> ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	116
Figura 120 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	117
Figura 121 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	117
Figura 122 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	118
Figura 123 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	118
Figura 124 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	119
Figura 125 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NF</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos)resultantes da simulação de três cadeias de Markov.....	119
Figura 126 Histograma de <i>NF</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NF.rep</i> ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	120
Figura 127 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	121
Figura 128 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	121
Figura 129 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	122
Figura 130 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	123
Figura 131 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	123

Figura 132 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.	125
Figura 133 Histograma de <i>NSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NSI.rep</i> ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	126
Figura 134 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	127
Figura 135 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	127
Figura 136 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	128
Figura 137 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	128
Figura 138 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	129
Figura 139 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	130
Figura 140 Histograma de <i>NSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NSI.rep</i> ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	131
Figura 141 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	131
Figura 142 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	132
Figura 143 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	132
Figura 144 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	133
Figura 145 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	133

Figura 146 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	134
Figura 147 Histograma de <i>NSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NSI.rep</i> ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	135
Figura 148 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	136
Figura 149 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	136
Figura 150 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	137
Figura 151 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	138
Figura 152 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	138
Figura 153 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	140
Figura 154 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	141
Figura 155 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	142
Figura 156 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	142
Figura 157 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	143
Figura 158 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	143

Figura 159 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	144
Figura 160 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	145
Figura 161 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	146
Figura 162 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	146
Figura 163 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	147
Figura 164 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	147
Figura 165 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	148
Figura 166 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.....	148
Figura 167 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.....	149
Figura 168 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de <i>AADT</i> maiores que 5000 e excluindo as variáveis com valores omissos.	150
Figura 169 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	151
Figura 170 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	151
Figura 171 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>Nac</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	154

Figura 172 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	154
Figura 173 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	155
Figura 174 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a Nac (obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	156
Figura 175 Histograma de Nac (no canto superior esquerdo) e dezanove réplicas (" $Nac.rep$ ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	158
Figura 176 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	159
Figura 177 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	159
Figura 178 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	160
Figura 179 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	160
Figura 180 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	161
Figura 181 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a Nac (obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	162
Figura 182 Histograma de Nac (no canto superior esquerdo) e dezanove réplicas (" $Nac.rep$ ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	163
Figura 183 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	164
Figura 184 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	164
Figura 185 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	165
Figura 186 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	166
Figura 187 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	166

Figura 188 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>Nac</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov	167
Figura 189 Histograma de <i>Nac</i> (no canto superior esquerdo) e dezanove réplicas (" <i>Nac.rep</i> ") dos números de acidentes obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	169
Figura 190 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	169
Figura 191 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	170
Figura 192 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	171
Figura 193 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	172
Figura 194 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	172
Figura 195 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NF</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	173
Figura 196 Histograma de <i>NF</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NF.rep</i> ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	174
Figura 197 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	175
Figura 198 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	176
Figura 199 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	176
Figura 200 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	177
Figura 201 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NF</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	177
Figura 202 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NF</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	178

Figura 203 Histograma de NF (no canto superior esquerdo) e dezanove réplicas (" $NF.rep$ ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	179
Figura 204 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	179
Figura 205 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	180
Figura 206 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	181
Figura 207 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	181
Figura 208 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	181
Figura 209 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a NF (obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	182
Figura 210 Histograma de NF (no canto superior esquerdo) e dezanove réplicas (" $NF.rep$ ") dos números de mortos obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	183
Figura 211 Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	183
Figura 212 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	184
Figura 213 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a NSI obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	185
Figura 214 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a NSI obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	185
Figura 215 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a NSI obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.	186
Figura 216 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a NSI (obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	187
Figura 217 Histograma de NSI (no canto superior esquerdo) e dezanove réplicas (" $NSI.rep$ ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.....	188

Figura 218	Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	189
Figura 219	Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	189
Figura 220	<i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	190
Figura 221	Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	191
Figura 222	Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	191
Figura 223	Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NSI</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	193
Figura 224	Histograma de <i>NSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NSI.rep</i> ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	194
Figura 225	Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.....	195
Figura 226	Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	195
Figura 227	<i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	196
Figura 228	Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	196
Figura 229	Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	197
Figura 230	Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NSI</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.	198
Figura 231	Histograma de <i>NSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NSI.rep</i> ") dos números de feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	199
Figura 232	Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.	200

Figura 233 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	200
Figura 234 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	201
Figura 235 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	202
Figura 236 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	202
Figura 237 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo de Poisson ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	203
Figura 238 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	204
Figura 239 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	205
Figura 240 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	205
Figura 241 <i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	206
Figura 242 Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	206
Figura 243 Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	207
Figura 244 Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson-Gama ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.....	208
Figura 245 Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	209
Figura 246 Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.....	210
Figura 247 Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	210

Figura 248	<i>Trace plots</i> (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	211
Figura 249	Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	211
Figura 250	Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a <i>NKSI</i> obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.....	212
Figura 251	Densidades <i>a posteriori</i> dos parâmetros β_j do modelo Poisson Log-Normal ajustado a <i>NKSI</i> (obtido do conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.	213
Figura 252	Histograma de <i>NKSI</i> (no canto superior esquerdo) e dezanove réplicas (" <i>NKSI.rep</i> ") dos números de mortos e feridos graves obtidos da distribuição preditiva <i>a posteriori</i> segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm <i>AADT</i> maior que 5000 e não possuem valores omissos.	214
Figura 253	Histogramas dos valores <i>T</i> : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O <i>p</i> é a probabilidade de <i>T</i> , obtido pelos dados replicados, ser superior ao <i>T</i> obtido pelos dados observados.	215
Figura 254	Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição <i>a posteriori</i> preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.	215

O presente documento descreve em detalhe a concepção, selecção e análise de modelos de estimativa de frequências de ocorrência de acidentes, mortos, feridos graves e mortos e feridos graves registados em auto-estradas portuguesas no período referente aos anos de 1999 a 2003. O estudo estatístico foi elaborado empregando exclusivamente uma abordagem Bayesiana e pode considerar-se como complementar do trabalho realizado em Azeredo Lopes e Cardoso (2009), no qual a abordagem efectuada foi a chamada *clássica*.

Os métodos Bayesianos são aplicados com a finalidade de estimar os parâmetros dos diversos modelos e para realizar diagnósticos dos modelos e comparações entre estes. No entanto, dado um modelo estatístico, e ao usar uma estrutura baseada na inferência *clássica* de verosimilhança, podem obter-se estimativas pontuais dos parâmetros dos modelos que sejam semelhantes entre si. Uma importante diferença entre a abordagem *clássica* e a Bayesiana reside no facto de, através desta última, se poderem obter medidas de incerteza para estimativas pontuais dos parâmetros (geralmente são considerados como tais os desvios padrões na abordagem *clássica*, e incertezas *a posteriori* ou intervalos de credibilidade na abordagem Bayesiana). É neste aspecto da análise estatística de dados de acidentes que a abordagem Bayesiana apresenta uma vantagem em relação aos métodos clássicos.

Métodos estatísticos Bayesianos têm sido aplicados em vários estudos referentes a acidentes de tráfego (Lan *et al.* (2009), Li *et al.* (2008), Carriquiry e Pawlovich (2005), Miaou e Lord (2003), Song *et al.* (2006), Shlüter *et al.* (1997)), no entanto não tinham ainda sido aplicados a dados registados na rede de estradas portuguesas.

Os modelos que foram empregues neste estudo consistem em modelos de regressão de Poisson e modelos hierárquicos Poisson-Gama e Poisson Log-Normal. Todos estes modelos constituem a estrutura probabilística mais usual para modelar frequências de acidentes rodoviários (Lord (2006), Lord *et al.* (2004), Tunaru (2002)). Os números dos vários eventos considerados foram assumidos como seguindo uma daquelas distribuições. Designam-se por eventos a ocorrência de um acontecimento do tipo analisado (acidente, morto, ferido grave ou a soma dos últimos dois).

As distribuições *a priori* (distribuições dos parâmetros desconhecidos para os quais o utilizador tem a possibilidade de inserir conhecimento prévio) foram assumidas como não informativas. Foi seguido um método de selecção de variáveis baseado no valor de DIC (critério de desviância de informação), após o qual se efectuaram verificações preditivas *a posteriori* da convergência do algoritmo de Monte Carlo via cadeias de Markov (MCMC), de maneira a garantir que se obteve uma distribuição *a posteriori* estacionária. Os modelos foram seguidamente validados, verificando-se a qualidade dos ajustes. Para o efeito efectuaram-se também verificações preditivas *a posteriori* e o cálculo de várias quantidades (testes de quantidades também chamados medidas de discrepância) que reflectissem características dos dados e que foram usadas para medir possíveis discrepâncias entre os modelos obtidos e os dados observados. Foram calculados os números esperados dos diversos eventos para

representações médias dos valores do tráfego médio diário anual (*AADT*) e do comprimento do trecho.

A organização do estudo seguiu a mesma sequência da análise descrita em Azeredo Lopes e Cardoso (2009), pois foram usados os mesmos conjuntos de dados, mantendo-se assim, no presente documento, a mesma nomenclatura das variáveis usadas nesse documento, para uma mais fácil comparação entre os resultados destes dois estudos.

Nas análises efectuadas foram usados os *softwares* estatísticos: *R: A Language and Environment for Statistical Computing*, mantido por R Development Core Team (2009) e o *software* desenvolvido para análises Bayesianas baseado numa linguagem de programação: *WinBUGS* (Spiegelhalter *et al.* (2003) e Lunn *et al.* (2000)). O *WinBUGS* é usado para gerar amostras aleatórias da distribuição *a posteriori* dos parâmetros de um modelo Bayesiano. O programa *R2WinBUGS: A Package for Running WinBUGS from R* (Sturtz *et al.* (2005)) foi utilizado para desencadear os programas de *WinBUGS* através do *R*.

Tal como em Azeredo Lopes e Cardoso (2009), o conjunto dos dados foi dividido de acordo com os valores do tráfego médio diário anual (*AADT*). Um dos conjuntos contém todos os trechos registados, o outro consiste nos trechos cujos valores de *AADT* registados foram superior a 5000 veículos. Cada um destes conjuntos foi novamente dividido em dois. No primeiro, os trechos de auto-estrada com valores omissos foram removidos. No segundo removeram-se as variáveis que continham valores omissos. Como era de esperar, os modelos finais obtidos através da abordagem Bayesiana, usando distribuições *a priori* não informativas foram semelhantes aos modelos obtidos empregando a abordagem *clássica* efectuada em Azeredo Lopes e Cardoso (2009).

Foi concluído, como também em Azeredo Lopes e Cardoso (2009), que é preferível usar o conjunto de dados nos quais os trechos com valores omissos foram descartados, do que o conjunto de dados descartando as variáveis com valores omissos. Isto porque o tipo de separador (que constitui uma variável) usado nas auto-estradas foi escolhido, após selecção de variáveis, para a maioria dos modelos em qualquer dos eventos considerados. A maior diferença resultante entre considerar todos os trechos de auto-estrada ou somente os trechos com valores de *AADT* superiores a 5000 foi que, no primeiro caso, a existência ou não de uma via adicional foi considerada de relevância, no segundo caso foi a largura da berma esquerda que foi relevante aquando da modelação dos vários tipos de eventos.

As vantagens de se terem usado técnicas Bayesianas consistiram em se ter podido escolher uma determinada distribuição para os coeficientes dos modelos considerados. A distribuição Normal com valor médio zero e desvio padrão elevado (para evidenciar a falta de conhecimentos prévios sobre os coeficientes) nos modelos de Poisson; e distribuições que partilham parâmetros desconhecidos, formando assim modelos hierárquicos, para os modelos Poisson-Gama e Poisson Log-Normal. Para todos os modelos foram obtidos intervalos de credibilidade de 95% e distribuições *a posteriori* para as estimativas dos respectivos coeficientes. Das análises efectuadas conclui-se que o modelo de regressão de Poisson (distribuição de Poisson) não é apropriado para modelar os diversos tipos de eventos considerados, pois não capta alguns dos atributos dos dados medidos pelos testes de

quantidades escolhidos, nomeadamente a dispersão existente nos dados. Os modelos obtidos após se terem assumido distribuições Poisson-Gama e Poisson Log-Normal foram considerados apropriados para modelar os eventos considerados.

O Capítulo 2 do presente documento consiste numa breve descrição dos dados analisados; no entanto, para uma descrição mais completa consultar Azeredo Lopes e Cardoso (2009). No Capítulo 3 faz-se uma breve introdução às definições e técnicas estatísticas empregues na análise Bayesiana; é feita uma descrição sucinta dos métodos de Monte Carlo via cadeias de Markov (MCMC). O Capítulo finaliza com a descrição detalhada dos modelos de Poisson, Poisson-Gama e Poisson Log-Normal empregues neste estudo e suas implementações em *WinBUGS*.

A análise detalhada e os resultados obtidos pelos modelos ajustados a trechos de auto-estradas, dos quais foram excluídas as variáveis com valores omissos, estão descritos no Capítulo 4. Os equivalentes para os dados excluindo trechos com valores omissos são apresentados no Capítulo 5. Os Capítulos 6 e 7 descrevem as análises efectuadas nos conjuntos de dados segundo os atributos dos dados analisados nos Capítulos 4 e 5, mas cujos valores de *AADT* são superiores a 5000 veículos. As conclusões gerais deste estudo encontram-se no Capítulo 8. Nos Anexos estão apresentados três excertos dos códigos de *WinBUGS* usados para ajustar os modelos de regressão de Poisson e os modelos hierárquicos de Poisson-Gama e Poisson Log-Normal.

2.1 Os dados das auto-estradas

Os dados analisados neste estudo foram obtidos em 227 trechos de auto-estradas portuguesas nos anos compreendidos entre 1999 e 2003, inclusive. Estes dados referem-se a várias características das auto-estradas, tais como tráfego e geometria do respectivo traçado, assim como o número de acidentes, feridos graves e mortes ocorridas em cada trecho de auto-estrada, considerando de forma agregada ambos os sentidos.

Visto estes dados terem sido já descritos e analisados, usando uma abordagem frequencista (clássica), em Azeredo Lopes e Cardoso (2009), indica-se este artigo como referência para mais detalhes sobre os dados incluindo vários gráficos, nomeadamente de densidades, e fotografias. No entanto segue-se uma breve descrição de todas as variáveis, incluindo os nomes originais (em inglês) que foram mantidos para facilitar a comparação entre o presente estudo e a análise anteriormente apresentada.

Para cada trecho de auto-estrada foram recolhidas as seguintes informações:

RoadName: a identificação do trecho da auto-estrada;

AADT: o tráfego médio diário anual (TMDA) em ambos os sentidos;

Length: o comprimento do trecho (medido entre nós rodoviários) em quilómetros;

Lane_Width: uma variável categórica que indica a largura da via em metros. Tem quatro categorias para larguras nos seguintes intervalos 'A'=[2.50, 3], 'B'=[3, 3.50], 'C'=[3.50, 3.75] e 'D'=[3.75, 4.30];

Number_of_Lanes: o número de vias no trecho de auto-estrada. É uma variável categórica com dois níveis: 'A' (quando existem duas vias), 'B' (quando existem três ou quatro vias);

Additional_Lane: uma variável categórica com 2 níveis indicando a existência (Y) ou não (N) de uma via adicional para veículos lentos.

Left_Shoulder_Width: variável categórica com 2 níveis indicando a largura da berma esquerda de cada faixa de rodagem. 'A' se a largura, em metros, está compreendida no intervalo]0, 0.50] e 'B' se está compreendida no intervalo]0.50, 2.50].

Right_Shoulder_Width: variável categórica representada por três classes indicando a largura, em metros, da berma direita de cada faixa de rodagem; as classes são [0, 1.5],]1.5, 2.5] e [2.5, -[.

Median_Type: variável categórica que indica o tipo de separador; tipo 'A' se for uma barreira de betão do tipo "New Jersey" ou "Double New Jersey" (DNJ;NJ), 'B' se "guarda metálica" (GM); e 'C' se "Vala", um separador com 11 metros de largura sem qualquer barreira.

Median_Width: variável categórica que indica a largura do separador; tipo 'A' se a largura do separador, em metros, estiver entre [0, 1.0], tipo 'B' se estiver em]1.0, 5.0] e 'C' se estiver no intervalo]5.0,-[.

Nac: o número de acidentes (em ambos os sentidos);

NF: o número de mortos (em ambos os sentidos);

NSI: o número de feridos graves (em ambos os sentidos);

NKSI: o número de mortes e feridos graves (em ambos os sentidos).

Nos capítulos seguintes, os modelos de predição de acidentes que irão ser seleccionados e ajustados usando uma abordagem Bayesiana, têm como variáveis de resposta o número de acidentes (*Nac*), o número de mortos (*NF*), número de feridos graves (*NSI*) ou o número de mortos e feridos graves (*NKSI*).

Os dados foram reagrupados tendo em conta vários critérios, nomeadamente, a exclusão de trechos e de variáveis com dados omissos (*missing values*), assim como o interesse em trechos de auto-estradas com valores de *AADT* (TMDA) maiores que 5000. Este último critério foi escolhido com o intuito de excluir trechos de auto-estrada que abriram ao tráfego em 2002, ou em 2003, e para os quais não foi possível obter informação sobre as datas de abertura.

Cada agrupamento de dados foi analisado separadamente. Dados cujas variáveis com valores omissos foram excluídas são analisados no Capítulo 4. O Capítulo 5 diz respeito a análises de dados sem trechos que possuíssem valores omissos e no Capítulo 6 consideram-se somente trechos com valores de *AADT* maiores que 5000, sem considerar as variáveis com valores omissos. Por fim, no Capítulo 7 os dados analisados incluem unicamente trechos sem valores omissos e com *AADT* maior que 5000.

O Quadro 1 contém sumários dos conjuntos de dados considerados em cada capítulo do presente relatório.

Quadro 1 | Número de trechos de auto-estrada, a soma dos seus comprimentos, volume de tráfego, somas do número de acidentes, de mortes e feridos graves ocorridos num período de cinco anos (1999 a 2003) para cada conjunto de dados analisados, assim como os capítulos do relatório a que se refere a análise de cada conjunto de dados.

	Número de Trechos	Soma dos comprimentos dos Trechos (km)	Volume de Tráfego (10^6 veículosxkm)	Número de Acidentes (Nac)	Número de Mortos (NF)	Número de Feridos Graves (NSI)
Capítulo 4	227	1671.67	59285	11400	642	1819
Capítulo 5	198	1574.88	52408	9686	566	1622
Capítulo 6	159	1111.88	56619	10522	573	1640
Capítulo 7	130	1015.09	49741	8808	497	1443

3 | A ABORDAGEM BAYESIANA

A estatística Bayesiana consiste numa maneira diferente de abordar a inferência estatística, comparativamente à que efectua a escola de pensamento *clássica*, também chamada *frequencista*. Esta diferença inclui uma abordagem diferente no cálculo de testes de hipóteses, *p*-values e intervalos de confiança, entre outros. Um dos aspectos mais importantes da estatística Bayesiana consiste no seguinte: antes de proceder à recolha dos dados existe geralmente alguma “opinião”, ou mesmo informação, acerca de determinada quantidade (geralmente algum parâmetro) que se queira estudar. Esta “informação” vai ser formulada em termos de uma distribuição chamada *a priori*. Após os dados terem sido recolhidos e observados, a “opinião” (ou informação) inicial vai ser actualizada (ou renovada) com o cálculo da distribuição *a posteriori*. Esta distribuição *a posteriori* é consequentemente resumida de modo a se proceder à inferência das quantidades (parâmetros) de interesse. Poderá também haver interesse em fazer inferências preditivas sobre uma nova amostra da população.

Introduzindo a notação usada no presente documento, considere-se $y = (y_1, \dots, y_n)$ um vector de observações com densidade amostral (verosimilhança) dada por $f(y | \theta)$, e seja $f(\theta)$ a distribuição *a priori* do parâmetro desconhecido θ . A inferência deste parâmetro é baseada na distribuição *a posteriori*, $f(\theta | y)$, que é calculada por intermédio do teorema de Bayes:

$$f(\theta | y) = \frac{f(y | \theta)f(\theta)}{f(y)} \propto f(y | \theta)f(\theta) \quad (3.1)$$

O símbolo \propto em, por exemplo $A \propto B$, significa que A é proporcional a B .

Na equação (3.1) $f(y)$ é dado por:

$$f(y) = \int f(y | \theta)f(\theta)d\theta \quad (3.2)$$

e representa a distribuição marginal de y , ou distribuição preditiva *a priori*. Depois de se observarem os dados y pode predizer-se um valor observável desconhecido y_i (por exemplo um valor futuro) resultante do mesmo processo. A distribuição de y_i chama-se distribuição preditiva *a posteriori*. É posterior porque é condicional no valor observado y e é preditiva porque é uma predição de um valor observável y_i .

$$\begin{aligned} f(y_i | y) &= \int f(y_i, \theta | y)d\theta . \\ &= \int f(y_i | y, \theta)f(\theta | y)d\theta \\ &= \int f(y_i | \theta)f(\theta | y)d\theta \end{aligned} \quad (3.3)$$

Na estatística Bayesiana ocorre frequentemente a impossibilidade de calcular analiticamente a distribuição *a posteriori*, $f(\theta | y)$, assim como outras quantidades de interesse, tais como $E(\theta | y)$ ou

$P(\theta \leq t | y)$, para um certo valor t . No entanto, pode sempre obter-se proporcionalmente a distribuição *a posteriori* ao multiplicar-se a distribuição *a priori* pela verosimilhança:

$$f(\theta | y) \propto f(\theta)f(y | \theta). \quad (3.4)$$

Neste caso, por não se poder obter analiticamente uma expressão para $\int f(\theta)f(y | \theta)d\theta$, também não haverá uma expressão para $f(\theta | y)$. Quando se usa uma distribuição *a priori* conjugada $f(\theta)$ pode-se “reconhecer” a forma da distribuição *a posteriori*. A propriedade que determinadas distribuições *a posteriori* possuem de seguirem a mesma forma paramétrica da distribuição *a priori* chama-se conjugação (Gelman *et al.* (2004)). A família conjugada é bastante conveniente matematicamente, no sentido em que a distribuição *a posteriori* segue uma forma paramétrica conhecida. No entanto, se houver informação que contradiga a família conjugada, será necessário usar uma distribuição *a priori* mais realista, embora mais inconveniente parametricamente. Para mais detalhes sobre as distribuições *a priori* a usar ver Gelman *et al.* (2004) e Paulino *et al.* (2003).

Uma análise Bayesiana requer o cálculo de valores esperados de funções de quantidades aleatórias como base para inferências, onde essas quantidades poderão ter distribuições *a posteriori* multivariadas ou possuir uma forma complexa, ou ambas (Spiegelhalter e Rice (2009)). Os métodos de integração numérica baseados em aproximações analíticas permitiram uma mudança radical nesses cálculos, facilitando-os, com a adopção de métodos indirectos, nomeadamente o método de Monte Carlo via cadeias de Markov (MCMC).

3.1 Descrição Sucinta dos Métodos de Monte Carlo

Os métodos de Monte Carlo englobam vários algoritmos que usam a simulação para solucionar diversos problemas estatísticos ou matemáticos. No presente caso ir-se-á usar um desses métodos, mais precisamente o método de Monte Carlo via Cadeias de Markov (MCMC), para estimar propriedades da distribuição de probabilidade que não se podem obter analiticamente.

Consideremos uma variável aleatória X com função de densidade de probabilidade $f_X(x)$. Suponhamos que se quer obter o valor médio de X , isto é:

$$\mu = \int_{-\infty}^{+\infty} xf_X(x)dx, \quad (3.5)$$

mas que não se pode resolver analiticamente o integral da equação (3.5). Se houver acesso a uma amostra aleatória X_1, \dots, X_N da distribuição de X , a estimativa mais óbvia de μ será:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i, \quad (3.6)$$

o valor médio amostral.

Dada uma amostra X_1, \dots, X_N pode estimar-se (com precisão arbitrária se N for suficientemente grande) qualquer medida sumária da pretendida distribuição de X . Por exemplo, se se quiser calcular $P(X \leq t)$, para um dado valor de t , mas não se pode calcular analiticamente o integral:

$$P(X \leq t) = \int_{-\infty}^t f_X(x) dx . \quad (3.7)$$

Usa-se então a estimativa:

$$\hat{P}(X \leq t) = \frac{1}{N} I(X_i \leq t) \quad (3.8)$$

Onde $I(\cdot)$ é a função indicadora.

No presente caso, produzindo uma amostra $\theta_1, \theta_2, \dots$ da distribuição *a posteriori* $f(\theta | y)$ e usando o método de Monte Carlo pode obter-se, com satisfatória precisão, desde que se use uma amostra suficientemente grande, qualquer quantidade de interesse de θ . Amostras da distribuição *a posteriori* podem ser obtidas (geradas) de várias maneiras sem contudo haver necessidade de ter um conhecimento exacto de $f(\theta | y)$. Os métodos directos incluem amostragem por rejeição, que gera propostas independentes para θ e aceita-as com uma taxa pela qual os valores retidos são proporcionais aos valores *a posteriori* desejados (Spiegelhalter e Rice (2009)).

O algoritmo que produz a amostra dos parâmetros θ para qualquer distribuição *a posteriori*, mesmo que seja conhecida somente pela constante de proporcionalidade, é baseado em cadeias de Markov e é conhecido por métodos de Monte Carlo via Cadeias de Markov (MCMC).

Consultar Ntzoufras (2009) para mais detalhes e exemplos de métodos de Monte Carlo em simulação e integração e correspondentes aplicações em inferência Bayesiana.

3.2 Cadeias de Markov

As realizações da distribuição *a posteriori* usadas em métodos de Monte Carlo não necessitam de ser independentes ou geradas directamente. Se a distribuição condicional de cada parâmetro é conhecida (condicionalmente a todos os outros parâmetros), uma maneira simples de gerar uma amostra possivelmente dependente é através da amostragem de Gibbs. Este algoritmo gera um parâmetro de cada vez; à medida que actualiza cada parâmetro, todo o espaço dos parâmetros é explorado. É por isso apropriado começar com múltiplos, e díspares, valores iniciais de maneira a verificar a convergência e, a longo prazo, assegurar que as “cadeias” de realizações produzidas irão reflectir os valores *a posteriori* pretendidos (Spiegelhalter e Rice (2009)).

A ideia principal de uma “actualização”, isto é, de produzir um conhecimento actualizado dos parâmetros $\theta = (\theta_1, \dots, \theta_p)$, consiste em combinar a informação *a priori* sobre a densidade dos parâmetros com a informação sobre os mesmos parâmetros fornecida pelos dados amostrais y . Por intermédio de métodos de Monte Carlo via cadeias de Markov (MCMC) podem obter-se repetidas amostras de θ . Especificamente, as densidades *a posteriori* $f(\theta | y)$ combinam pressupostos *a priori*

$f(\theta)$ em θ , com distribuições amostrais aplicáveis a diferentes tipos de dados observados y , a chamada verosimilhança.

Os MCMC são métodos que consistem na extração de valores θ de distribuições aproximadas, seguida da correcção dessas extracções para melhor aproximar a distribuição alvo *a posteriori*, $f(\theta|y)$. As amostras são extraídas sequencialmente, com distribuição dependente do último valor extraído, por consequência, as extracções formam uma cadeia de Markov. Tal como é definido em teoria de probabilidade, uma cadeia de Markov é uma sequência de variáveis aleatórias $\theta^1, \theta^2, \dots$ para as quais, e para cada t , a distribuição de θ^t , dados todos os θ anteriores, depende somente do valor mais recente, θ^{t-1} . O sucesso deste método deve-se, não à propriedade de Markov, mas sim ao facto de as distribuições aproximadas serem melhoradas em cada passo da simulação, no sentido em que convergem para a distribuição alvo (Gelman *et al.*, 2004).

Paulino *et al.* (2003) afirmam que a ideia básica por detrás dos métodos de Monte Carlo via cadeias de Markov (MCMC) é a de transformar o problema estático em consideração num problema de natureza dinâmica, construindo para o efeito um processo estocástico temporal, artificial, que seja fácil de simular, e que convirja para a distribuição original. Este processo temporal é, em geral, uma cadeia de Markov homogénea cuja distribuição de equilíbrio é a distribuição que se pretende simular. Para implementar este método há necessidade de saber construir cadeias de Markov com distribuições de equilíbrio específicas. Vários algoritmos foram desenvolvidos para esse efeito.

Exemplos de algoritmos de cadeias de Markov muito úteis em vários problemas multidimensionais são o algoritmo de Metropolis-Hasting e algoritmo de Gibbs. Para mais detalhes sobre a construção e implementação destes algoritmos consultar Gelman *et al.* (2004) e Paulino *et al.* (2003). A aplicação dos algoritmos sob o ponto de vista do *software* WinBUGS encontra-se explicada em Congdon (2003) e Ntzoufras (2009).

3.3 Avaliação da Convergência da Cadeia de Markov

O algoritmo da cadeia de Markov converge quando atinge a sua distribuição de equilíbrio, isto é o seu objectivo. Se a convergência ocorre, a amostra obtida (a amostra simulada) provém da distribuição que se pretende. Verificar e certificar que a convergência do algoritmo ocorre é essencial para produzir resultados confiáveis da distribuição *a posteriori* pretendida.

Segundo Gelman *et al.* (2004) a simulação iterativa acrescenta dois problemas à inferência por simulação. O primeiro consiste no facto de que se as iterações não forem em número suficientemente grande, isto é, se não se prolongarem, as simulações consequentes podem não ser representativas da distribuição alvo pretendida. Por outro lado, mesmo quando as simulações atingiram aproximadamente a convergência, as primeiras iterações não deixam de estar influenciadas pela aproximação inicial ao invés de estarem influenciadas pela distribuição alvo. Para evitar esta situação existe a prática de descartar as primeiras iterações, prática essa chamada de *burn-in*. Usualmente também se procede a um “desbaste” de iterações ao reter todas as *k*-ésimas iterações de cada

simulação e descartar as restantes; é chamada de *thin* ou *thinning rate* igual a k . Quando se efectua o *thin* reduz-se o espaço de memória utilizada assim como a correlação entre as iterações.

Para os diversos modelos ajustados neste estudo foram usados diferentes números de iterações, de *burn-in* e de *thinning rates*, de maneira a que cada algoritmo atingisse a convergência.

O segundo problema dos valores da simulação iterativa consiste nas suas inter-correlações; a inferência baseada em simulações de valores correlacionados é geralmente menos precisa do que se for baseada em valores independentes entre si. Para evitar estes problemas Gelman *et al.* (2004) aconselham um planeamento de simulações que permita acompanhar a convergência, em particular simular múltiplas sequências, cujos valores iniciais estejam dispersos pelo espaço dos parâmetros. Os autores sugerem também verificar a convergência de todas as quantidades de interesse, comparando a variação "*within*" e "*between*" de sequências de simulações até estas variações serem iguais. Somente quando a distribuição de cada sequência simulada se aproximar da distribuição de todas as sequências conjuntamente, isto é, quando as distribuições se interligarem, se poderá dizer que todas elas se estão a aproximar da distribuição alvo. O algoritmo também pode ter de ser alterado quando a eficiência da simulação for muito baixa, isto é, quando a simulação requerer muito tempo para obter uma convergência (aproximada) das inferências *a posteriori* para as quantidades de interesse. O tipo de alterações a fazer nessas circunstâncias é descrito no Capítulo 11 de Gelman *et al.* (2004).

Verificar a convergência do algoritmo MCMC consiste em verificar a convergência dos parâmetros, ou de um conjunto de parâmetros, se o modelo os possuir em número elevado, estimados pelo algoritmo. Após o algoritmo convergir, as amostras das distribuições condicionais irão ser usadas para sumariar a distribuição *a posteriori* dos parâmetros de interesse. Segundo o manual *online* do *software* WinBUGS (ver Spiegelhalter *et al.* (2003)), avaliar a convergência requer bastante cuidado, sendo muito difícil afirmar que a cadeia (simulação) converge; é apenas possível diagnosticar quando claramente não converge.

3.3.1 *Trace plots*, Erros de Monte Carlo e Diagnóstico de Autocorrelação

A verificação da convergência do algoritmo MCMC quando se utiliza o *software* WinBUGS consiste na análise de várias quantidades obtidas pelo *output* deste programa, que serão seguidamente detalhadas. Segundo Ntzoufras (2009) existem várias possibilidades de verificar a convergência a mais simples consiste em verificar os erros de Monte Carlo (erros de MC). Estes erros medem a variação do valor médio dos parâmetros devido à simulação. Há indicação de que o valor médio *a posteriori* foi estimado com elevada precisão quando os erros de MC são mais baixos do que os desvios padrões da distribuição *a posteriori* dos parâmetros.

Uma vez atingida a convergência, as amostras devem apresentar-se com uma dispersão (*scatter*) aleatória à volta de um valor médio estável. Os gráficos do traço (*trace plots*) (ver Figura 1 com dois exemplos deste tipo de gráficos) apresentam a relação entre as iterações e os valores obtidos de

cada parâmetro. Pode-se assumir a convergência do algoritmo se todos os valores se encontrarem dentro de uma zona sem apresentar periodicidades e tendências. No gráfico superior da Figura 1 podem observar-se *trace plots* de duas cadeias (indicadas a azul e a vermelho) para as quais parece haver razoável convergência, no gráfico inferior as duas cadeias claramente não convergem.

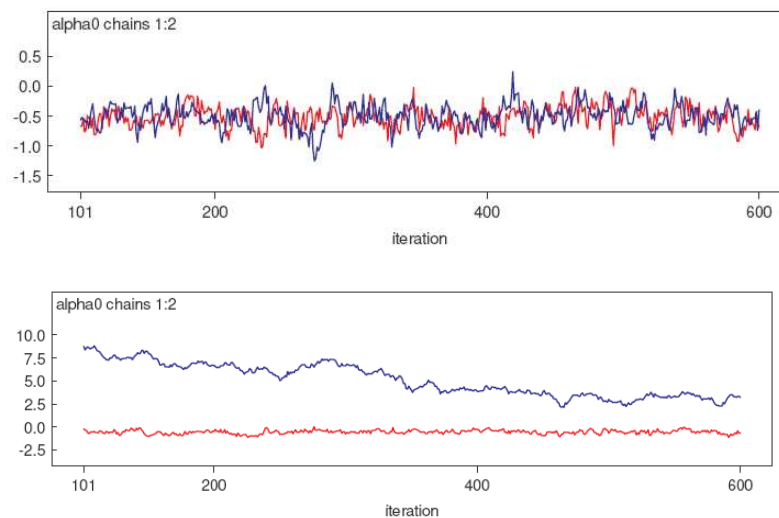


Figura 1| Exemplo de *trace plots* de duas cadeias de Markov, indicadas a azul e a vermelho, que aparentam convergir (gráfico superior) e de duas cadeias de Markov que não convergem (gráfico inferior).

A autocorrelação refere-se ao padrão (*pattern*) da correlação serial da cadeia de Markov, onde as várias sequências dos parâmetros são correlacionadas. A causa da autocorrelação deve-se ao facto de os parâmetros no modelo serem correlacionados. A autocorrelação revela a rapidez da convergência; valores altos ou baixos indicam uma convergência lenta ou rápida, respectivamente. A autocorrelação é mais bem analisada graficamente. Na

Figura 2 estão representados dois gráficos com os valores da autocorrelação de dois parâmetros de um dado modelo. O gráfico superior indica a presença de correlação (note-se a presença de duas cadeias simuladas), e o gráfico inferior indica ausência de correlação.

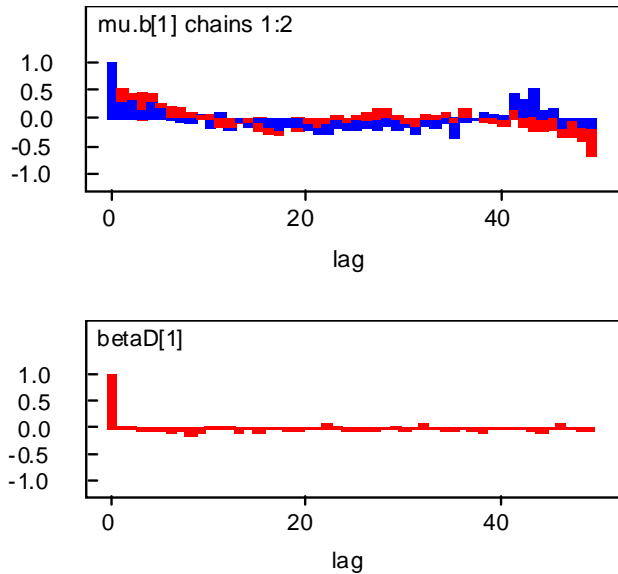


Figura 2| Exemplos de gráficos de autocorrelação para parâmetros com (gráfico superior) e sem (gráfico inferior) autocorrelação.

3.3.2 Diagnóstico de Gelman-Rubin

Gelman *et al.* (2004) argumentam que para melhor identificar uma não convergência se deverão simular múltiplas sequências de cadeias de Markov com valores iniciais sobredispersos. O objectivo é verificar se o comportamento das cadeias é semelhante. Quando várias cadeias são simuladas em paralelo com diferentes valores iniciais pode calcular-se o diagnóstico de convergência de Gelman-Rubin. Este diagnóstico consiste em obter, e posteriormente comparar, as variabilidades *entre* e *inter*-amostrais, isto é, a variabilidade *between-sample* e *within-sample*.

O *software* WinBUGS calcula a estatística de Gelman-Rubin retirando n amostras de m parâmetros e calculando as seguintes estatísticas:

Variância inter-amostrai, W :

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_j^i - \bar{\theta}_j)^2 \quad (3.9)$$

Variância entre amostras, B :

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta})^2 \quad (3.10)$$

Variância estimada, $\hat{V}(\theta)$:

$$\hat{V}(\theta) = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B \quad (3.11)$$

Sendo a estatística de Gelman-Rubin dada por R :

$$R = \frac{\hat{V}(\theta)}{W} \quad (3.12)$$

Antes de se atingir a convergência, W subestima a variância *a posteriori* em θ porque ainda não explorou completamente a distribuição alvo. $V(\theta)$, por outro lado, sobrestima a variância em θ porque os valores iniciais estão sobredispersos relativamente à distribuição alvo. Uma vez atingida a convergência, W e $\hat{V}(\theta)$ devem ser equivalentes porque as variações intra e entre as cadeias devem coincidir, conseqüentemente R deve ser aproximadamente igual a um. Para mais detalhes sobre estas estatísticas consultar Ntzoufras (2009) e Gelman *et al.* (2004).

Segundo Ntzoufras (2009) todos os testes de diagnóstico de convergência são “alarmes” que alertam quando alguma anomalia não esperada é detectada no *output* do algoritmo de MCMC. Como cada teste de diagnóstico é construído para detectar problemas diferentes, o autor sugere que, na maioria dos casos, todos os diagnósticos devam ser empregues de forma a garantir, com maior precisão, a convergência do algoritmo.

No entanto, os testes baseados nas estatísticas de Gelman-Rubin são considerados como sendo os mais formais e têm por isso resultados mais fidedignos do que os outros diagnósticos de convergência usados.

3.4 Validação e Selecção de Modelos

Segundo Gelman *et al.* (2004) uma análise Bayesiana adequada deverá incluir uma validação do ajuste do modelo aos dados, assim como uma medida da plausibilidade do modelo, tendo em conta os objectivos que se pretendem em cada estudo.

Para verificar se o modelo é coerente com os dados faz-se um controlo da distribuição preditiva *a posteriori*. Um modelo bem ajustado irá gerar dados semelhantes aos dados observados. Ou seja, os dados observados devem ser coerentes com os dados provenientes da distribuição preditiva *a posteriori* (equação (3.3)). A técnica para verificar a adequação do modelo aos dados usada por Gelman *et al.* (2004) consiste em recolher valores de simulações da distribuição preditiva *a posteriori* de dados replicados e comparar essas amostras com os dados observados. Quaisquer diferenças sistemáticas entre as simulações e os dados indicam potenciais deficiências do modelo. A discrepância entre o modelo e os dados pode ser medida por “testes de quantidades” (*test quantities*), também chamados “medidas de discrepância” (*discrepancy measures*), $T(y|\theta)$, que são um sumário escalar de parâmetros e dados. Estas medidas são usadas como padrão (*standard*) quando se comparam os dados com as simulações preditivas. O não ajustamento (*lack-of-fit*) dos dados com respeito à distribuição preditiva *a posteriori* (equação (3.3)) pode ser medido pela probabilidade da área da cauda da distribuição, ou seja o *p*-value do teste de quantidade.

Na abordagem Bayesiana os testes de quantidade podem ser funções de parâmetros desconhecidos θ , como também de dados, isto porque o teste de quantidade é avaliado por intermédio de várias recolhas de valores da distribuição *a posteriori* dos parâmetros desconhecidos θ . O *p*-value Bayesiano, p_B , é definido por Gelman *et al.* (2004) como a probabilidade do dados replicados (*y.rep*) serem mais extremos que os dados observados, tal como foram medidos pelo teste de quantidade:

$$p_B = P(T(y.rep, \theta) \geq T(y, \theta | y)) \quad (3.13)$$

Na prática a distribuição *a posteriori* preditiva é calculada por simulação. Se obtivermos L simulações da densidade *a posteriori* de θ e retiramos um valor de $y.rep$ da distribuição preditiva por cada valor simulado de θ ; fica-se assim com L valores da distribuição *a posteriori* conjunta, $f(y.rep, \theta | y)$. A validação pela distribuição preditiva *a posteriori* consiste na comparação entre os testes de quantidades obtidos pelos dados observados, $T(y, \theta^t)$ e os testes de quantidades obtidos pela distribuição preditiva *a posteriori*, $T(y.rep^t, \theta^t)$, para $t=1, \dots, L$. O p -value estimado é a proporção das L simulações para as quais o teste de quantidade obtido pela distribuição preditiva *a posteriori* é maior ou igual ao valor observado, isto é:

$$p_B = T(y.rep^t, \theta^t) \geq T(y, \theta^t), \quad t = 1, \dots, L \quad (3.14)$$

Valores de p_B aproximadamente iguais a 0.5 indicam um modelo coerente com os dados observados; por outro lado valores extremos (próximo de 0 ou de 1) sugerem incoerências entre as predições do modelo e os dados observados. Congdon (2005) sugere valores entre 0.10 e 0.90 como valores limites aceitáveis para p_B . Para mais detalhes consultar Gelman *et al.* (2004) e Congdon (2003, 2005).

No presente estudo o ajuste dos modelos aos dados observados foi feito com base nas distribuições *a posteriori* preditivas dos seguintes testes de quantidades: a maior observação, $max_j y_j$, o valor médio, $mean(y_j)$, o desvio padrão, $sd(y_j)$ e um critério sugerido por Congdon (2005) para aplicação em dados de contagem com sobredispersão, uma vez que nos dados relativos ao número de acidentes existe sobredispersão (ver Azeredo Lopes e Cardoso (2007, 2009)); critério esse que consiste na razão da variância sobre a média, $var(y_j)/mean(y_j)$.

A distribuição preditiva *a posteriori* para cada teste é aproximada pelo histograma dos valores obtidos após L simulações dos parâmetros e dos dados preditivos (replicados). Cada distribuição (histograma) é comparada com o valor observado do teste estatístico. O modelo ajustado deverá reproduzir as suas características nas réplicas que dele forem extraídas.

Para verificar se a sobredispersão está adequadamente modelada, Congdon (2005) sugere a comparação do valor médio das desviâncias amostrais, ou seja, o valor médio de D^t , onde:

$$D^t = 2\{\sum (y_i + 0.5) \log\left(\frac{y_i + 0.5}{\mu_i^{(t)} + 0.5}\right) - (y_i - \mu_i^{(t)})\} \quad (3.15)$$

(em que y_i são os valores observados e μ_i são os números esperados obtidos pelo modelo), com os graus de liberdade disponíveis, $n-p$, onde p é dado pelo número de covariáveis do modelo.

3.5 Critério DIC

O critério da desviância de informação (DIC) foi introduzido por Spiegelhalter *et al.* (2002) como uma medida de comparação e adequação de modelos. É dado pela seguinte expressão:

$$DIC = \overline{2D(\theta_m, m)} - D(\bar{\theta}_m, m) = D(\bar{\theta}_m, m) + 2p_m \quad (3.16)$$

onde $D(\theta_m, m)$ é a medida de desviância usual envolvendo o logaritmo da verosimilhança:

$$D(\theta_m, m) = -2 \log f(y | \theta_m, m) \quad (3.17)$$

e $\overline{D(\theta_m, m)}$ é o seu valor médio *a posteriori*. p_m é interpretado como sendo o número de parâmetros “efectivos” para o modelo m e é dado por:

$$p_m = \overline{D(\theta_m, m)} - D(\bar{\theta}_m, m) \quad (3.18)$$

Onde $\bar{\theta}_m$ é o valor médio *a posteriori* dos parâmetros do modelo m .

O cálculo de DIC possibilita a selecção de variáveis, pois valores mais baixos de DIC indicam um modelo mais bem ajustado. Modelos não encaixados (*non-nested*) podem também ser comparados pelo critério de DIC. Para mais detalhes sobre o DIC consultar Ntzoufras (2009). Segundo este autor, um bom começo para seleccionar as variáveis explicativas (ou covariáveis) é ajustar o modelo completo (*full model*) e depois adoptar como modelo inicial o modelo que inclui as variáveis para as quais a distribuição *a posteriori* dos respectivos coeficientes está afastada de zero, ou seja, cujos intervalos de credibilidade de 95% não incluam o valor zero. Seguidamente escolhe-se o modelo que produza o menor valor de DIC.

3.6 Modelos Lineares Generalizados

A classe dos modelos lineares generalizados unifica as abordagens necessárias para analisar dados para os quais a suposição de uma relação linear entre X e y , ou a suposição de uma variação normal, não é apropriada. Um modelo linear generalizado é especificado em três níveis:

1. O preditor linear, $\eta = X\beta$
2. A função de ligação $g(\cdot)$ que relaciona o preditor linear com o valor médio da variável resultante, $\mu = g^{-1}(\eta) = g^{-1}(X\beta)$
3. A componente aleatória que especifica a distribuição da variável y com valor médio $E(y | X) = \mu$. Em geral a distribuição de y dado X pode também depender de um parâmetro de dispersão ϕ .

Consequentemente, a média da distribuição de y , dado X , é determinada por $X\beta$: $E(y | X) = g^{-1}(X\beta)$.

Os modelos lineares generalizados podem ter distribuições *a priori* informativas ou não informativas, assim como podem ser modelos hierárquicos para os quais a distribuição *a priori* de β depende de hiperparâmetros desconhecidos. Os modelos hierárquicos, também denominados *multilevel*, são usados quando existe informação em vários níveis das unidades observadas.

3.6.1 Modelos de Regressão de Poisson

Os dados na forma de contagens aparecem em diversas aplicações e são frequentemente modelados usando um modelo de Poisson, modelo este que desempenha um papel fundamental na análise deste tipo de dados (Amaral Turkman e Silva, 2000). No modelo linear generalizado de Poisson, usualmente chamado modelo de regressão de Poisson, ou modelo de Poisson log-linear, assume-se que os dados y_i seguem uma distribuição de Poisson com valores médios λ_i ,

$$y_i \sim \text{Poisson}(\lambda_i), i=1, \dots, N \quad (3.19)$$

e por conseguinte variâncias iguais a λ_i . A função de ligação é geralmente escolhida como sendo o logaritmo, de maneira a que:

$$\log \lambda_i = \beta X_i \quad (3.20)$$

A distribuição dos dados $y=(y_1, \dots, y_N)$ é dada em Gelman *et al.* (2004) por:

$$p(y | \beta) = \prod_{i=1}^N \frac{1}{y_i!} e^{-\exp(\eta_i)} (\exp(\eta_i))^{y_i} \quad (3.21)$$

Onde $\eta_i = (X\beta)_i$ é o preditor linear para o elemento i .

Na prática, os dados de Poisson podem estar mais dispersos do que a respectiva densidade assume (Congdon, 2005). Se não se corrigir a variação extra (sobredispersão), a precisão dos parâmetros β_i irá estar inflacionada e os correspondentes intervalos de credibilidade vão ser bastante estreitos (Cameron e Trivedi, 1998).

Dados de contagem resultantes de um presumível modelo de Poisson apresentam amiúde uma variância amostral maior do que é esperada neste tipo de modelos (Congdon, 2006). Esta sobredispersão pode ocorrer devido à omissão de covariáveis ou de alguma forma de agrupamento (*clustering*) das unidades originais (por exemplo, os dados pertencem a determinadas unidades, mas exibem efeitos de agrupamento (*clustering effects*) porque as unidades estão agrupadas a um nível superior). Congdon (2006) afirma ser preferível usar um modelo que tenha em atenção a sobredispersão, especialmente se o interesse é focado na significância dos parâmetros de regressão.

Uma alternativa que tem em conta a sobredispersão na densidade de Poisson (quando se verifica que a variância observada excede a forma assumida pela função densidade de probabilidade) consiste na modificação desta densidade envolvendo uma mistura de distribuições no valor médio da distribuição de Poisson. De acordo com Congdon (2005) esta modificação pode também ser feita em conjunção com a adopção de funções de ligação alternativas ou de misturas de distribuição e de funções de ligação. As generalizações das misturas de distribuição e de funções de ligação proporcionam maior robustez nas inferências, assim como uma densidade compatível com os dados (Gelman *et al.*, 2004). Outra motivação para usar modelos de mistura é o facto de estes modelos reunirem informações sobre as várias unidades (neste caso trechos rodoviários) quando as contagens em cada unidade variam consideravelmente; ao modelar as taxas para unidades individuais em termos de uma hiperdensidade conjunta, ir-se-ão obter estimativas mais precisas para cada unidade que tendem “suavemente” (*smooth*) para a média conjunta (Congdon, 2005).

No presente documento os valores de *Nac*, *NF*, *NKSI* e *NSI* recolhidos em cada trecho de auto-estrada *i* num período de cinco anos são denotados por y_i , considerados independentes e modelados por uma distribuição de Poisson com valor médio λ_i , ou seja:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \beta_0 + \sum_{j=1}^p \beta_j X_{ij} \\ \beta_0, \beta_j &\sim \text{Normal}(0, 10^3) \end{aligned} \quad (3.22)$$

λ_i denota o número esperado de cada tipo de evento considerado num período de cinco anos e X_j , $j=1, \dots, p$, são as p covariáveis usadas. Assumindo para os parâmetros β_0 e β_j uma distribuição Normal com valor médio igual a zero e um grande valor para a variância, resultando assim numa distribuição *a priori* não informativa.

3.6.2 Modelo de Poisson-Gama

Numa análise de regressão pode representar-se a heterogeneidade não observada em dados de contagem usando um efeito aleatório multiplicativo φ tal que:

$$y_i \sim \text{Poisson}(\theta_i) \quad (3.23)$$

onde

$$\theta_i = \lambda_i \varphi_i = \exp(\beta X_i) \varphi_i \quad (3.24)$$

Assumindo uma distribuição Gama (conjugada para a distribuição de Poisson) para os efeitos aleatórios φ_i :

$$\varphi_i \sim \text{Ga}(\delta, \gamma) \quad (3.25)$$

A restrição $\delta = \gamma$ é usada frequentemente. A distribuição conjugada Gama conduz a um modelo Poisson-Gama ou a um modelo Binomial Negativo (Congdon, 2006). Este último decorre de se integrar a seguinte expressão em função de φ_i :

$$P(y_i | X_i) = \int P(y_i | X_i, \varphi_i) P(\varphi_i) d\varphi_i \quad (3.26)$$

Obtém-se, assim, uma distribuição Binomial Negativa (onde $1/\delta$ é o parâmetro de dispersão) para distribuição marginal dos dados, y_i (para mais detalhes consultar Congdon (2005, 2006)).

No caso presente optou-se por implementar os modelos Poisson-Gama no seguinte modelo hierárquico:

$$\begin{aligned}
y_i &\sim \text{Poisson}(\theta_i) \\
\theta_i &= \lambda_i \varphi_i = \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \varphi_i \\
\beta_0, \beta_j &\sim \text{Normal}(0, 10^3) \\
\varphi_i &\sim \text{Gama}(\delta, \delta) \\
\delta &\sim \text{Gama}(10^{-1}, 10^{-1})
\end{aligned} \tag{3.27}$$

Onde θ_i denota o número esperado de cada tipo de evento considerado.

3.6.3 Modelo de Poisson Log-Normal

A alternativa principal ao modelo Poisson-Gama consiste em adicionar um erro aleatório, ε_i , a $\log(\theta_i)$ de maneira a que:

$$y_i \sim \text{Poisson}(\theta_i) \tag{3.28}$$

$$\log(\theta_i) = X_i \beta + \varepsilon_i \tag{3.29}$$

Onde ε_i pode ter uma densidade paramétrica tal como a distribuição Normal:

$$\varepsilon_i \sim \text{Normal}(0, \sigma^2) \tag{3.30}$$

Congdon (2005) refere que mantendo a ligação *log* e assumindo $\varepsilon_i \sim N(0, \alpha)$, a variância da distribuição *a posteriori* é aproximada a $V(y_i | X_i) = \theta_i + \alpha \theta_i^2$, onde α é o parâmetro de dispersão.

Ao modelo acima dá-se usualmente o nome de Poisson Log-Normal.

O modelo hierárquico Poisson Log-Normal foi implementado com os seguintes níveis:

$$\begin{aligned}
y_i &\sim \text{Poisson}(\theta_i) \\
\log(\theta_i) &= \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i \\
\beta_0, \beta_j &\sim \text{Normal}(0, 10^3) \\
\varepsilon_i &\sim \text{Normal}(0, \sigma^2) \\
\sigma &\sim \text{Gama}(10^{-3}, 10^{-3})
\end{aligned} \tag{3.31}$$

Onde $1/\sigma$ dá o valor do parâmetro de dispersão.

4 | MODELAÇÃO DOS DADOS EXCLUINDO AS VARIÁVEIS COM VALORES OMISSOS

A análise estatística descrita no presente capítulo consiste na modelação, selecção e ajuste de modelos a dados obtidos nas auto-estradas portuguesas, aos quais foram removidas as variáveis em que um ou mais valores eram omissos. Consequentemente, as variáveis elegíveis para serem inseridas nos modelos foram $\log(AADT)$, $\log(Length)$, $Additional_Lane$, $Lane_Width$ e $Number_of_Lanes$. Os dados analisados compreenderam 227 trechos de auto-estrada. Tal como no Capítulo 4 de Azeredo Lopes e Cardoso (2009), foi decidido modelar os dados assumindo como variáveis resposta o número de acidentes (Nac) e o número de mortos e feridos graves (NKS).

4.1 Modelação do Número de Acidentes

A descrição detalhada das características de todas as variáveis consideradas nesta secção encontra-se na secção 4.1 de Azeredo Lopes e Cardoso (2009).

4.1.1 Modelo de Poisson

Uma vez ajustado o modelo completo (*full model*) ao conjunto de equações (3.22), verificou-se que os intervalos de credibilidade de 95% para os parâmetros β_i não continham o valor zero. Seguidamente procedeu-se ao ajuste de modelos obtidos removendo cada uma das variáveis. O menor valor de DIC foi obtido no modelo completo, ou seja, não se obteve um melhor ajuste após a remoção de variáveis. A equação na segunda linha do conjunto de equações em (3.22) é representada por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Lane_Width_i + \beta_5 Number_of_Lanes_i \quad (4.1)$$

O algoritmo de MCMC composto por três cadeias converge após 30000 iterações das quais 20000 são descartadas (*burn-in*) com um *thinning rate* igual a 10, resultando numa amostra final de dimensão 3000.

A observação dos *trace plots* na Figura 3 para os parâmetros β_j mostra que, para cada uma das três cadeias de Markov (denotadas a azul, vermelho e verde), os valores das estimativas dos parâmetros tendem a aproximarem-se, uma vez que estes valores se situam sempre entre a “mesma zona” sem se verificarem periodicidades ou tendências, não havendo, por isso, razões para suspeitar da não convergência do algoritmo.

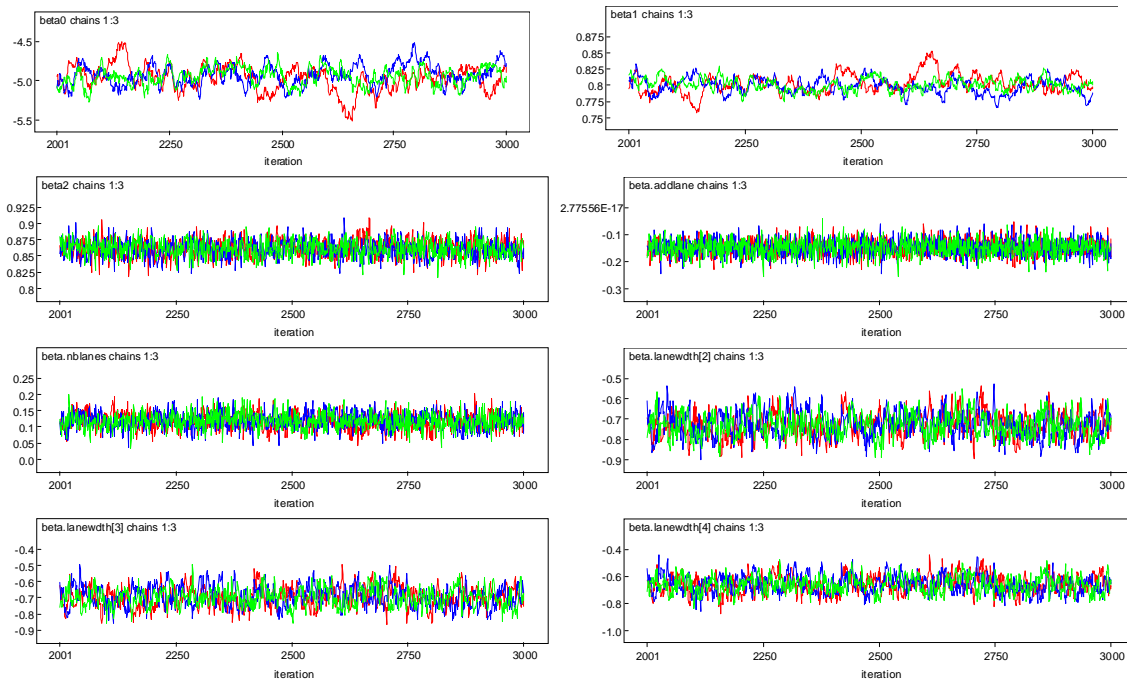


Figura 3] Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *Nac* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações das estimativas dos parâmetros β_j reproduzidos na Figura 4 indicam uma convergência lenta de β_0 e β_1 , mas mais rápida para os restantes parâmetros.

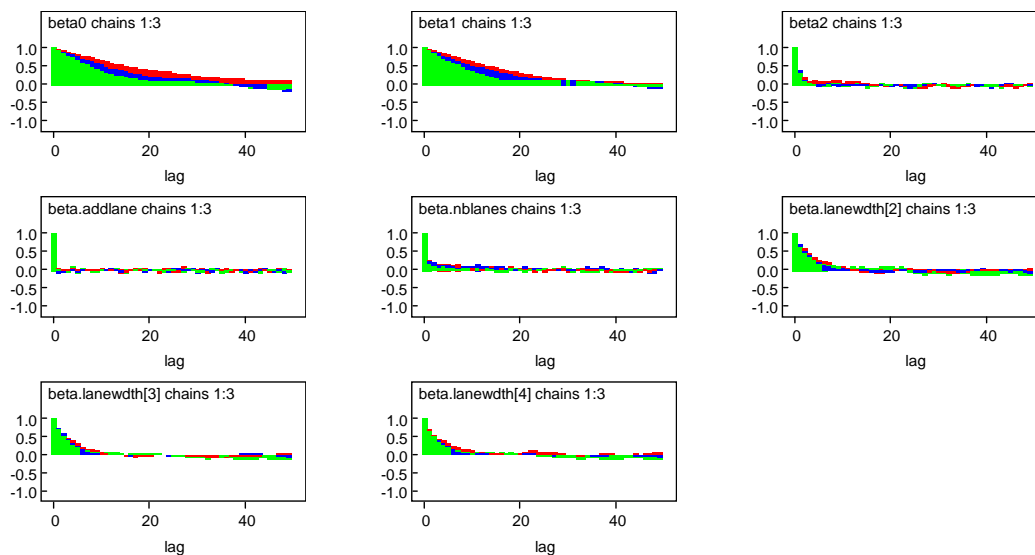


Figura 4] Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *Nac* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 5. Pode observar-se que R (a vermelho nos gráficos e dado pela equação (3.12)), se aproxima, em todos os parâmetros, para o valor 1 e que, ambas as variâncias W (linha azul) e \hat{V}

(linha verde), estabilizam à medida que o número de iterações aumenta. Factos que indicam uma convergência da simulação iterativa.

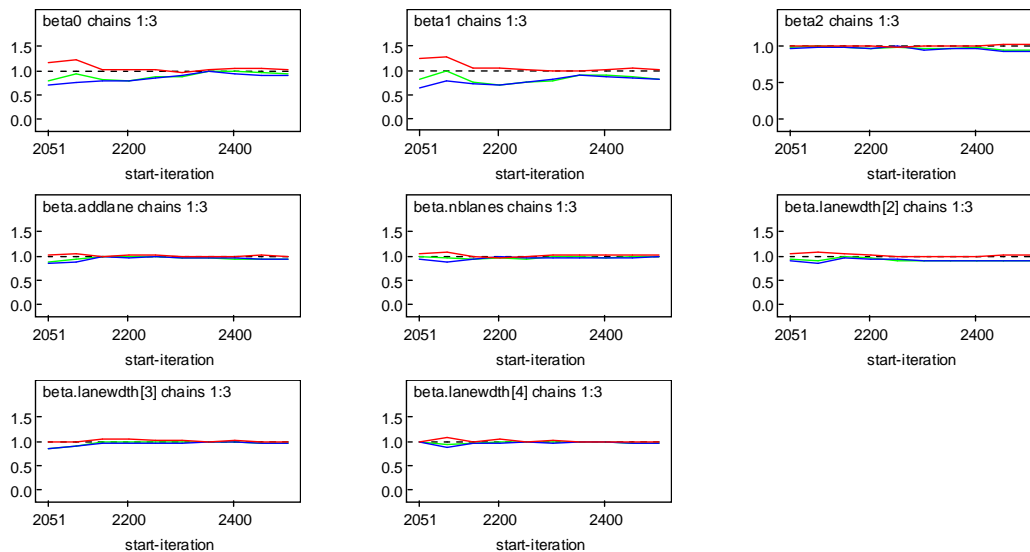


Figura 5| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados excluindo as variáveis com valores omissos.

As equações dos números esperados de acidentes, para um período de cinco anos, estão representadas no Quadro 2, de acordo com os níveis das covariáveis *Additional_Lane*, *Number_of_Lanes* e *Lane_Width*. Pode constatar-se que o número esperado de acidentes é menor em trechos com via adicional (*Additional_Lane*='Yes') e com duas vias (*Number_of_Lanes*='A'). Segundo o modelo de Poisson, os trechos por ordem crescente de números esperados de acidentes têm largura de via (*Lane_Width*) dos tipos 'B', 'C', 'D' e 'A' (ver secção 2.1).

O Quadro 3 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% para as estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pela equação (4.1).

Os intervalos de credibilidade de 95% não contêm o valor zero, o que indica que todas as variáveis do modelo têm um efeito relevante para a predição do número de acidentes (ver Congdon, 2005). As densidades *a posteriori* dos parâmetros do modelo exibidas nos gráficos da Figura 6 também mostram que os valores médios de todas as densidades se afastaram consideravelmente do valor zero (a média assumida pela distribuição *a priori* para todos os parâmetros).

No entanto, devido ao facto de os erros de Monte Carlo não serem muito baixos em comparação com os desvios padrões (isto é, menores que 0.1% dos desvios padrões posteriores, como sugere Ntzoufras, 2009) para alguns dos parâmetros, deve ter-se em conta que as densidades *a posteriori* (ver Figura 6) desses parâmetros não foram estimadas com muita precisão.

Pela leitura das médias no Quadro 3 pode afirmar-se que o número de acidentes num trecho com uma via adicional é menor em cerca de 14% do que num trecho nas mesmas condições, mas sem via adicional.

Quadro 2| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo as variáveis com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='A'</i> <i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 7.083 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'B'	$\hat{\lambda}_i = 3.418 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'C'	$\hat{\lambda}_i = 3.537 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'D'	$\hat{\lambda}_i = 3.666 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='B'</i> <i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 7.974 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'B'	$\hat{\lambda}_i = 3.848 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'C'	$\hat{\lambda}_i = 3.982 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'D'	$\hat{\lambda}_i = 4.127 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='A'</i> <i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 6.107 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'B'	$\hat{\lambda}_i = 2.947 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'C'	$\hat{\lambda}_i = 3.050 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'D'	$\hat{\lambda}_i = 3.161 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='B'</i> <i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 6.875 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'B'	$\hat{\lambda}_i = 3.318 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'C'	$\hat{\lambda}_i = 3.433 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$
'D'	$\hat{\lambda}_i = 3.558 \times 10^{-3} \times AADT_i^{0.801} \times Length_i^{0.861}$

Um trecho que tenha 3 ou 4 vias (isto é, *Number_of_Lanes='B'*) tem um número esperado de acidentes cerca de 13% superior aos trechos com duas vias (*Number_of_Lanes='A'*). Trechos com largura de via no intervalo]3.00, 3.50] (isto é, *Lane_Width='B'*) têm uma redução no número esperado

de acidentes em cerca de 52% em relação a trechos, em iguais condições, mas com largura de via entre]0, 3.00] ($Lane_Width='A'$). Os trechos com largura de via nos intervalos]3.50, 3.75] e]3.75, $+\infty$ [têm uma redução de 50% e 48%, respectivamente, em relação a trechos com largura de via no intervalo]0, 3.00].

Quadro 3| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 2.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-4.950	0.138	1.094×10^{-2}	-5.220	-4.698
β_1 ($\log(AADT)$)	0.801	0.012	9.898×10^{-4}	0.777	0.826
β_2 ($\log(Length)$)	0.861	0.013	3.824×10^{-4}	0.836	0.886
β_3 (Additional_Lane='Yes')	-0.148	0.029	5.789×10^{-4}	-0.204	-0.093
β_4 (Lane_Width='B')	-0.729	0.059	2.647×10^{-3}	-0.842	-0.607
β_5 (Lane_Width='C')	-0.694	0.061	2.828×10^{-3}	-0.806	-0.571
β_6 (Lane_Width='D')	-0.659	0.063	2.914×10^{-3}	-0.778	-0.533
β_7 (Number_of_Lanes='B')	0.118	0.025	8.838×10^{-4}	0.069	0.168

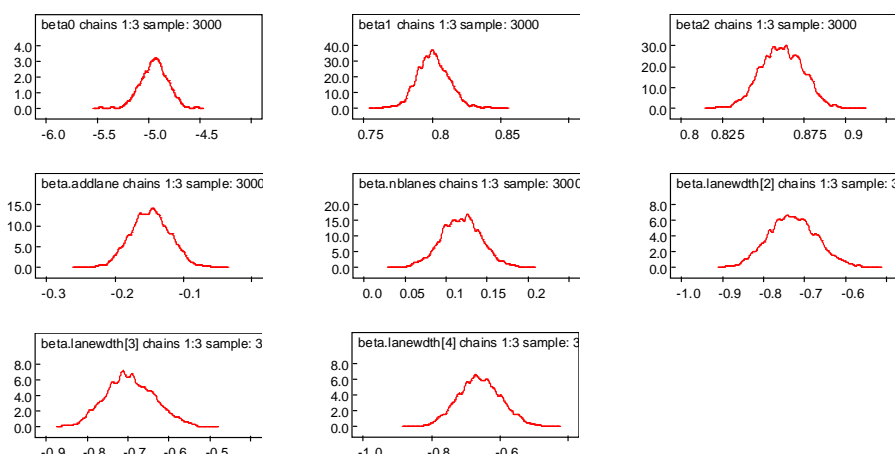


Figura 6| Densidades *a posteriori* dos parâmetros β_i do modelo de Poisson ajustado a N_{ac} (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de acidentes para os diversos níveis das covariáveis do modelo foi calculado para uma representação média de $\log(AADT)$ e $\log(Length)$, isto é, para valores médios de $\log(AADT)$ e $\log(Length)$. Os valores das médias *a posteriori* e desvios padrões para estas representações (correspondentes a um trecho típico de auto-estrada) estão exibidos no Quadro 4.

Para um trecho com duas vias e com via adicional, cuja largura de via toma valores compreendidos entre 3.00 a 3.50 metros, esperam-se 23.35 acidentes em cinco anos. Um trecho de auto-estrada com 3 ou 4 vias, mas sem via adicional e com largura de via entre os 0 e 3.00 metros, é esperado ter, num período de cinco anos, 63.23 acidentes. Note-se, no entanto, os elevados valores dos desvios padrões (relativos a algumas médias) que indicam uma grande variação para estas médias.

Quadro 4| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 2.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	56.17 (3.318)	'A'	63.23 (3.661)
'B'	27.07 (0.751)	'B'	30.47 (0.879)
'C'	28.01 (0.563)	'C'	31.54 (0.918)
'D'	29.03 (0.702)	'D'	32.68 (0.947)
<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='_Yes'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	48.45 (3.103)	'A'	54.54 (3.538)
'B'	23.35 (0.845)	'B'	26.28 (1.056)
'C'	24.15 (0.733)	'C'	27.20 (1.090)
'D'	25.04 (0.889)	'D'	28.19 (1.182)

4.1.1.1 Validação do Modelo

Para Gelman *et al.* (2004) se o modelo se encontra bem ajustado, então os conjuntos de dados obtidos através desse modelo (dados replicados) devem ser semelhantes aos dados observados.

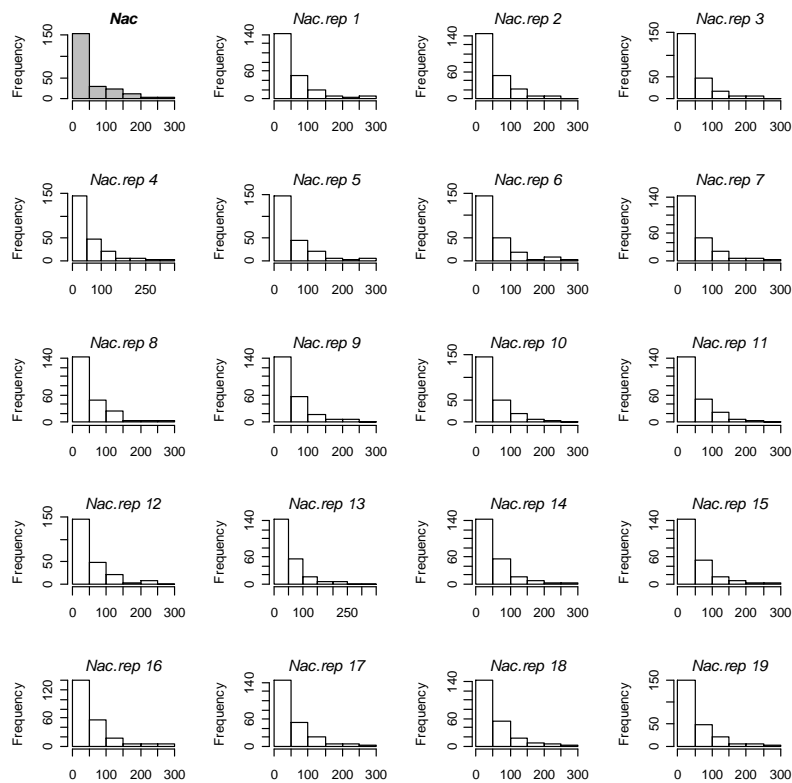


Figura 7| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas ("*Nac.rep*") dos números de acidentes obtidas da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

A Figura 7 é composta por vinte histogramas, estando no canto superior esquerdo, a cinzento, o histograma dos números de acidentes (Nac) observados em cada um dos 227 trechos estudados. Os restantes dezanove histogramas foram obtidos de dados replicados da distribuição *a posteriori* preditiva, cada um com dimensão 227 (denotados por $Nac.rep$). A comparação dos dezanove histogramas com o histograma dos dados observados mostra que vários histogramas das réplicas diferem substancialmente do histograma de Nac , nomeadamente, os histogramas $Nac.rep$ 2, 12 e 16. Há assim suspeitas de que o modelo de Poisson não seja representativo do número de acidentes.

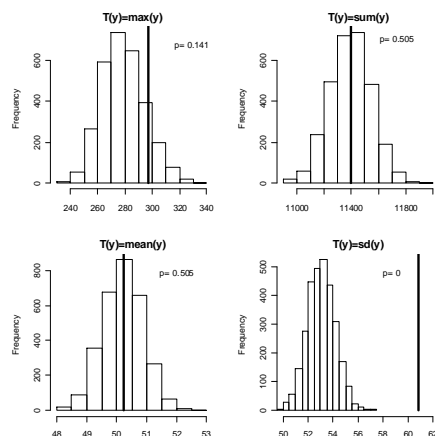


Figura 8| Histogramas dos valores T : máximo, soma, média e desvio padrão de 3000 conjuntos de dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Para verificar a ocorrência de possíveis discrepâncias entre os dados observados e os dados replicados pelo modelo escolheram-se as medidas referentes ao valor máximo, soma, média e desvio padrão dos números de acidentes. A Figura 8 mostra os histogramas das quatro medidas obtidas de cada um de 3000 conjuntos de dados replicados (cada conjunto com dimensão 227, igual à dimensão dos dados observados). As rectas verticais representam a medida obtida no conjunto dos dados observados, isto é, referentes a Nac .

Pela observação da Figura 8 depende-se que os dados replicados pelo modelo de Poisson capturam bem a variação correspondente à soma e ao valor médio dos dados (ver histogramas no canto superior direito e no canto inferior esquerdo da Figura 8) onde a probabilidade p_B (ver equação 3.14) de que essas medidas obtidas nos dados replicados sejam superiores às dos dados observados, são iguais a 0.505 em ambos os casos. No entanto, os dados replicados pelo modelo de Poisson não parecem reflectir os mesmos valores máximos e desvios padrões obtidos nos dados observados (ver histogramas no canto superior esquerdo e no canto inferior direito da Figura 8). Neste caso as probabilidades p_B não se encontram entre os limites de 0.10 e 0.90 sugeridos por Congdon (2005), embora o p_B relativo a $T=max(y)$ se encontre próximo.

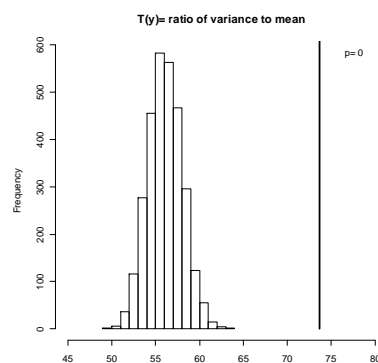


Figura 9| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Um critério sugerido por Congdon (2005) para verificar se a sobredispersão em dados de contagem é devidamente tida em conta por um determinado modelo consiste no cálculo da razão entre a variância e o valor médio dos dados observados que é depois comparada com as mesmas razões obtidas dos dados replicados. Existe sobredispersão quando a variância amostral é maior do que o valor médio. A Figura 9 mostra o histograma das razões obtidas entre as variâncias e os valores médios de 3000 simulações de dados obtidos pelo modelo de regressão de Poisson. A recta vertical denota a mesma razão obtida nos dados referentes aos números de acidentes observados. O valor de p_B dado pela equação (3.14) é igual a zero. Como se pode depreender o modelo de Poisson assume razões entre variância e valor médio das diversas simulações muito inferiores ao valor obtido pelos dados (dados onde existe sobredispersão, ver Azeredo Lopes e Cardoso, 2007, 2009). Obviamente, o modelo de Poisson expresso pelas equações no Quadro 2 não está a ter em consideração a sobredispersão que os dados possuem, o que indica que o modelo de Poisson não é apropriado para modelar os números de acidentes no presente estudo.

A média das desviâncias amostrais dada pela equação (3.15) resulta em $\bar{D}=3661.0$, um valor bastante elevado quando comparado com $n-p$ (222), e mais uma vez revelador de que a sobredispersão não está a ser bem modelada.

Este modelo produz um valor de DIC (ver equação 3.16) igual a 4853.85.

4.1.2 Modelo de Poisson-Gama

Seguindo a estratégia de Ntzoufras (2009), foi ajustado o modelo completo aos números de acidentes assumindo uma distribuição de Poisson-Gama. Verificou-se que os intervalos de credibilidade de 95% para as estimativas dos parâmetros das variáveis *Number_of_Lanes* e *Lane_Width* continham zero, o que indica que estas variáveis não são relevantes. O modelo com as covariáveis $\log(AADT)$, $\log(Length)$ e *Additional_Lane* produz o menor valor de DIC, mesmo tendo em conta a remoção de

Additional_Lane assim como a inserção das outras variáveis rejeitadas inicialmente. O logaritmo do número esperado de acidentes é dado pela seguinte equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i \quad (4.2)$$

Para que o algoritmo de MCMC convirja são necessárias 50000 iterações, das quais 40000 são descartadas (*burn-in*) com um *thinning rate* igual a 10 iterações. Os resultados das quantidades *a posteriori* são obtidos de uma amostra com dimensão 3000. Três cadeias de Markov foram simuladas simultaneamente.

Foram utilizados gráficos de várias quantidades com o intuito de verificar se o algoritmo atingiu a distribuição alvo (também chamada distribuição de equilíbrio).

A Figura 10, inclui quatro *trace plots* dos parâmetros do modelo considerado em (4.2). Pode observar-se que os gráficos das iterações correspondentes a β_0 e β_1 (na primeira linha da Figura 10) não parecem coincidir para o mesmo valor, o que evidencia que o algoritmo parece não convergir apropriadamente. No entanto, para β_2 e β_3 não se observam tendências particulares nos *trace plots*.

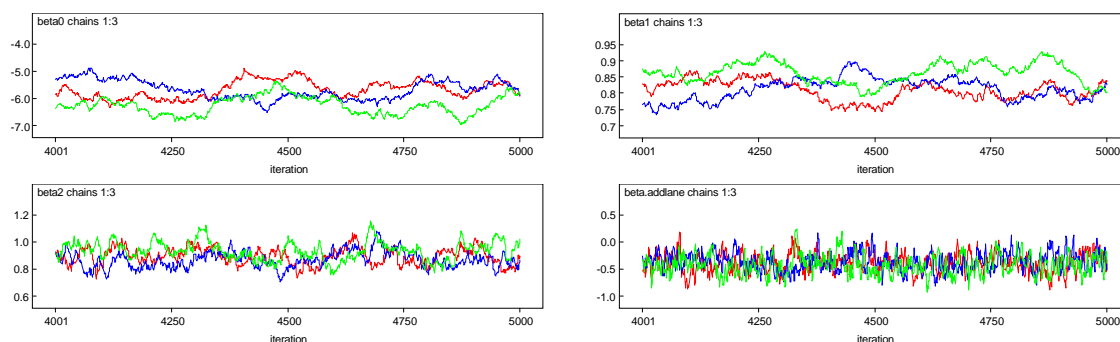


Figura 10| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *Nac* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações para os parâmetros β_j reproduzidos na Figura 11 indicam uma convergência lenta de β_0 e β_1 e uma convergência mais rápida de β_3 .

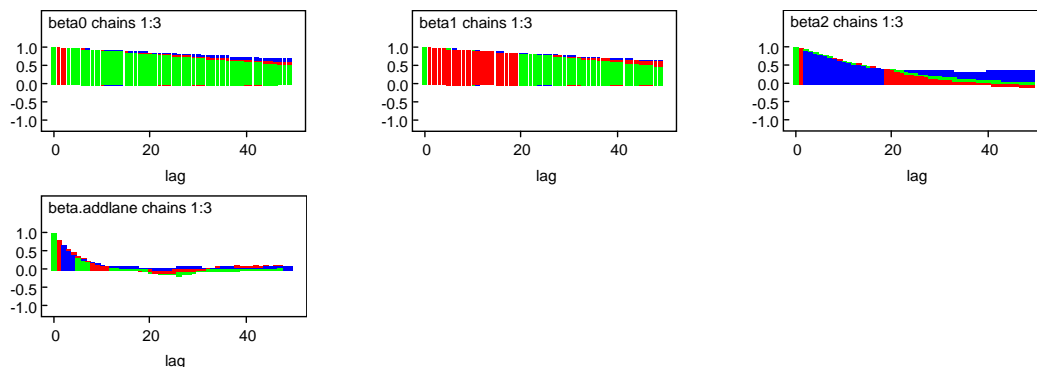


Figura 11| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *Nac* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin, W , \hat{V} e R (correspondentes, respectivamente às equações (3.9), (3.11) e (3.12)) estão representados na Figura 12. Para β_0 , β_1 e β_2 o valor da estatística R correspondente converge para 1.

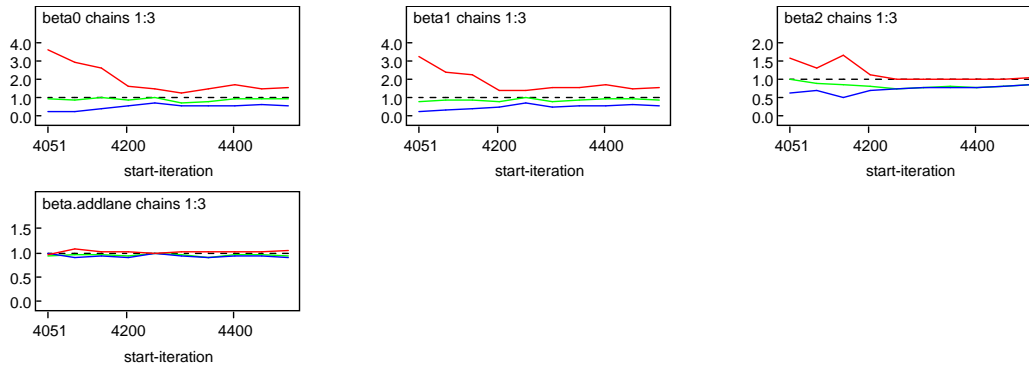


Figura 12| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados excluindo as variáveis com valores omissos.

Uma vez que os factores de redução de escala de Gelman-Rubin (ver Figura 12) mostram que a convergência foi atingida, não houve necessidade de repetir a simulação do algoritmo com um maior número de iterações.

As equações do modelo estimado *a posteriori* para os números esperados de acidentes, num período de cinco anos, de acordo com os dois níveis da covariável *Additional_Lane*, estão apresentadas no Quadro 5. Pode observar-se que o número esperado de acidentes é menor nos trechos com via adicional (*Additional_Lane*='Yes').

Quadro 5| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo as variáveis com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 2.809 \times 10^{-3} \times AADT_i^{0.828} \times Length_i^{0.897}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 1.896 \times 10^{-3} \times AADT_i^{0.828} \times Length_i^{0.897}$

O Quadro 6 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% para as estimativas da densidade *a posteriori* dos parâmetros β_j do modelo apresentado no Quadro 5. Os intervalos de credibilidade de 95% não contêm zero o que indica que todas as covariáveis são úteis no modelo. As médias *a posteriori* das estimativas dos parâmetros β_j afastaram-se bastante de zero (média *a priori* assumida), como também pode ser constatado nos gráficos das densidades *a posteriori* na Figura 13. Os erros de Monte Carlo, embora não suficientemente menores que 0.1% dos desvios padrões são no entanto valores baixos.

Quadro 6| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 5.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-5.875	0.430	4.326×10^{-2}	-6.735	-5.116
β_1 (log(AADT))	0.828	0.040	4.048×10^{-3}	0.755	0.909
β_2 (log(Length))	0.897	0.070	5.924×10^{-3}	0.772	1.047
β_3 (Additional_Lane='Yes')	-0.393	0.180	1.009×10^{-2}	-0.734	-3.310×10^{-2}

Pela interpretação das estimativas do parâmetro β_j correspondente a *Additional_Lane* pode afirmar-se que a presença de uma via adicional num trecho de auto-estrada reduz o número esperado de acidentes em cerca de 33% em comparação com um trecho em iguais condições, mas sem via adicional.

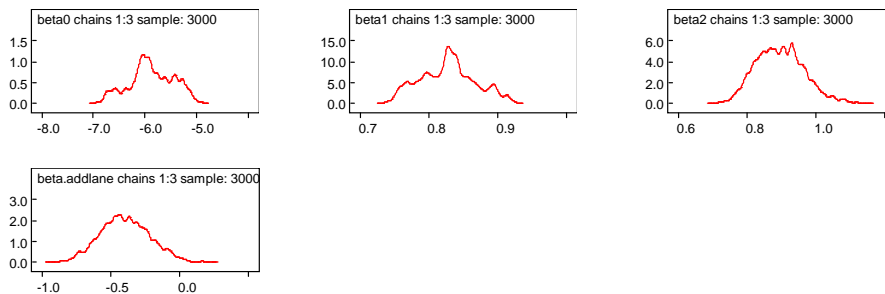


Figura 13| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson-Gama ajustado a *Nac* (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de acidentes para os dois níveis de *Additional_Lane*, e para uma representação média das restantes variáveis, foi calculado e é dado pelos valores indicados no Quadro 7.

Quadro 7| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 5.

	média (desvio padrão)
<i>Additional_Lane</i> ='No'	30.55 (1.801)
<i>Additional_Lane</i> ='Yes'	20.88 (3.571)

Para um trecho típico de auto-estrada (com um valor médio de *AADT* e de *Length*) sem via adicional, são esperados, para um período de cinco anos, cerca de 31 acidentes. Para um trecho equivalente, mas com via adicional, são esperados cerca de 21 acidentes, no mesmo período de tempo.

4.1.2.1 Validação do Modelo

A Figura 14 mostra dezanove histogramas obtidos de dados replicados pelo modelo Poisson-Gama ajustado ao número de acidentes, assim como o histograma dos dados observados (a cinzento no canto superior esquerdo). Ao contrário do que se observou na Figura 7 (histogramas obtidos de dados replicados pelo modelo de Poisson), os dezanove histogramas da Figura 14 assemelham-se bastante ao histograma dos dados observados o que significa que o modelo de Poisson-Gama está mais bem ajustado a estes dados do que o modelo de Poisson.

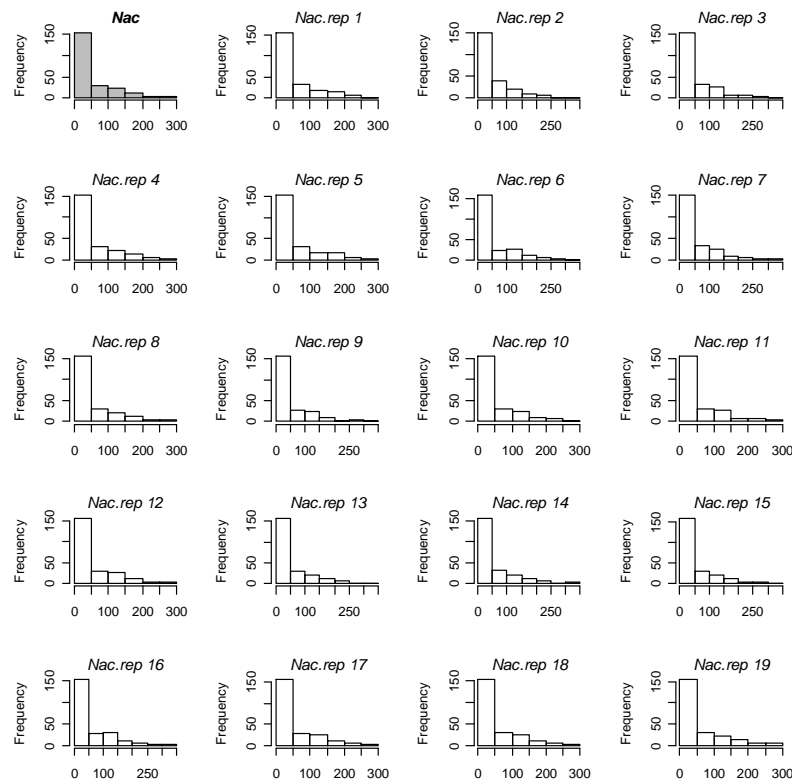


Figura 14| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

Os histogramas na Figura 15 consistem nos valores das quatro medidas *T* (ver secção 3.4) em cada um dos 3000 hipotéticos conjuntos de dados replicados pelo modelo de Poisson-Gama, comparados com os valores equivalentes de *T* para os dados observados (as rectas verticais nos histogramas), assim como as probabilidades p_B de que as medidas *T* dos dados replicados sejam maiores do que a mesma medida *T* obtida dos dados observados.

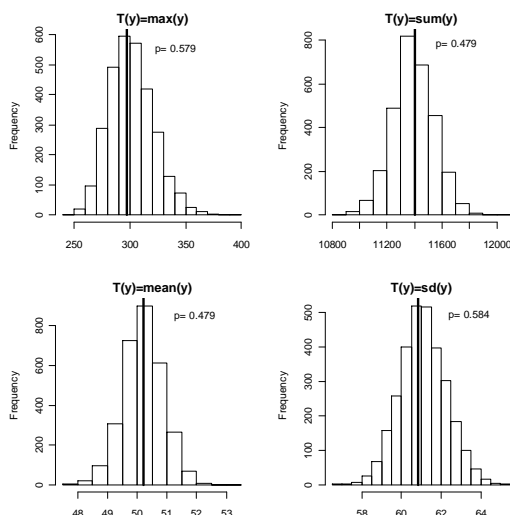


Figura 15| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo Poisson-Gama capta claramente as variações consideradas pelas medidas T calculadas, visto os histogramas se situarem em torno das rectas verticais e os p_B terem valores aproximadamente iguais a 0.5.

A medida de discrepância usada para avaliar a “captação” da sobredispersão dos dados consiste na razão entre a variância e a média dos dados. O histograma da Figura 16 foi obtido pelos dados resultantes dessa razão retirados de cada uma de 3000 simulações de conjuntos retirados da distribuição preditiva *a posteriori* dos valores dos números de acidentes registados, comparada com a razão obtida pelos próprios dados observados (recta vertical).

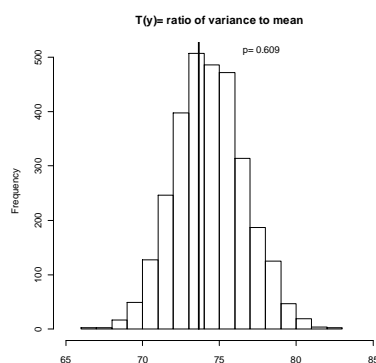


Figura 16| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Uma vez que o histograma engloba a recta vertical e $p_B=0.609$ há razões para crer que o modelo de Poisson-Gama capta a sobredispersão existente nos dados. O que não acontecia com o modelo de

Poisson (ver Figura 9). A média das desviâncias amostrais (ver equação (3.15)) resultou num $\bar{D}=223.6$ o que comparado com $n-p$ (224) indica um bom ajuste do modelo Poisson-Gama no que refere à modelação da sobredispersão. A estimativa do parâmetro de dispersão é igual a 0.6. Este modelo produz um valor DIC igual a 1525.31, consideravelmente menor que o valor DIC obtido pelo modelo de Poisson (4853.85).

4.1.3 Modelo de Poisson Log-Normal

O modelo de Poisson Log-Normal dado pelas equações (3.31) foi ajustado aos dados referentes aos números de acidentes assumindo $\log(AADT)$, $\log(\text{Length})$, Additional_Lane , Number_of_Lanes e Lane_Width como covariáveis. Deste conjunto de variáveis, Number_of_Lanes e Lane_Width foram removidas devido ao facto de os intervalos de credibilidade de 95% para as estimativas dos correspondentes parâmetros β_i conterem o valor zero, o que significa que estas variáveis não influenciam o número de acidentes. O processo recomendado por Ntzoufras (2009) para escolher a remoção e o adicionamento de covariáveis foi empregue e o modelo obtido com o menor valor de DIC contém as variáveis $\log(AADT)$, $\log(\text{Length})$ e Additional_Lane e é dado pela seguinte equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (4.3)$$

Todos os resultados obtidos resultaram de 40000 iterações do algoritmo de MCMC das quais 30000 foram inicialmente rejeitadas (*burn-in*) com *thinning rate* de 10 iterações. A amostra final consistiu em 3000 valores retidos.

A Figura 17 mostra os *trace plots* dos quatro parâmetros β_j . Pela observação da figura pode-se concluir que as estimativas de β_0 e β_1 não convergem tão facilmente como as de β_2 e β_3 .

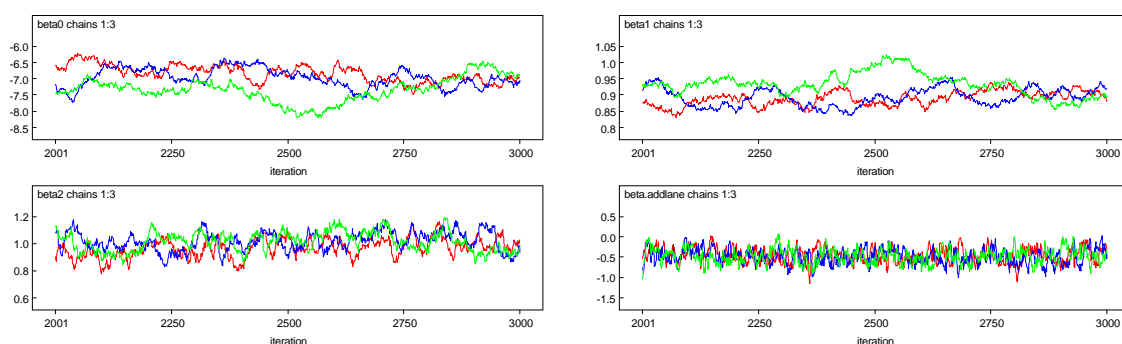


Figura 17| *Trace plots* (iteraões vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a *Nac* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Na Figura 18 estão exibidos os gráficos das autocorrelações dos parâmetros β_j . Os parâmetros β_0 , β_1 e β_2 convergem mais lentamente do que β_3 .

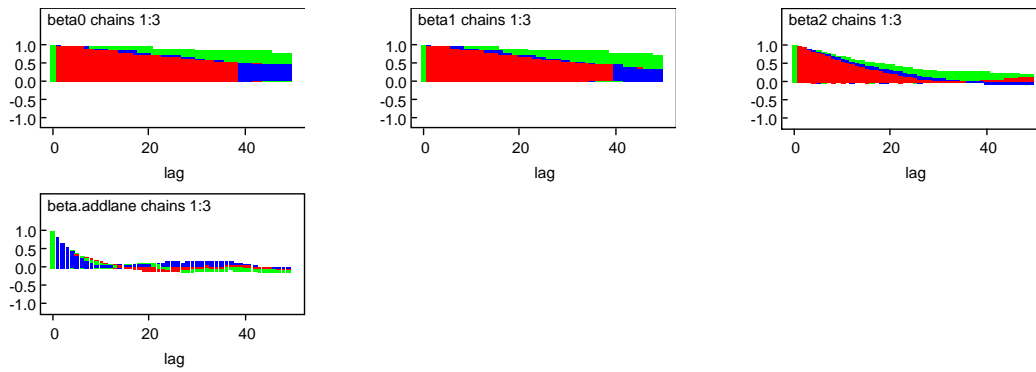


Figura 18| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo as variáveis com valores omissos.

A Figura 19 contém os gráficos das estatísticas de Gelman-Rubin, para os parâmetros β_j . Em todos eles R (ver equação (3.12)) converge para 1, o que indica que o algoritmo de MCMC converge, embora o número de iterações usado não tenha sido suficiente para esclarecer se os valores de R estabilizam no valor 1.

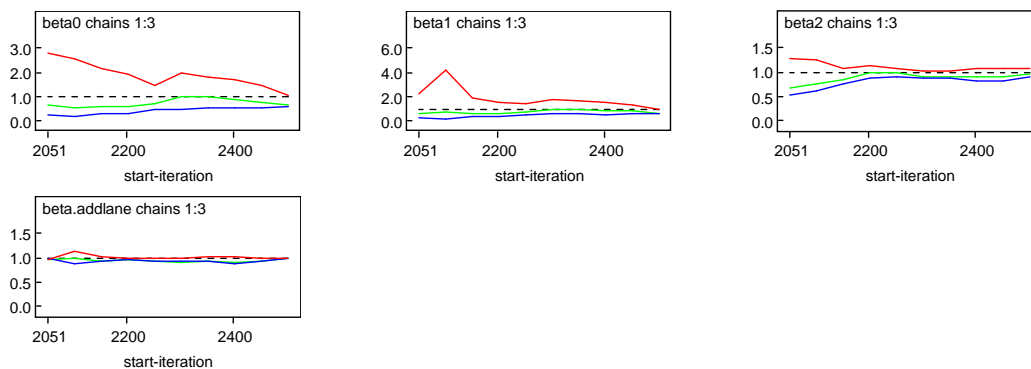


Figura 19| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo as variáveis com valores omissos.

As equações que dão o número esperado de acidentes, de acordo com os níveis da variável *Additional_Lane*, encontram-se no Quadro 8. Mais uma vez se pode constatar que o número esperado de acidentes é menor em trechos com presença de via adicional.

Quadro 8| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo as variáveis com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 8.650 \times 10^{-4} \times AADT_i^{0.905} \times Length_i^{0.998}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 5.367 \times 10^{-4} \times AADT_i^{0.905} \times Length_i^{0.998}$

As estimativas pontuais, assim como intervalos de credibilidade de 95% para os parâmetros da equação (4.3) são exibidos no Quadro 9.

Quadro 9| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 8.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.052	0.376	3.768×10^{-2}	-7.971	-6.456
β_1 (log(AADT))	0.905	0.035	3.514×10^{-3}	0.849	0.995
β_2 (log(Length))	0.998	0.074	6.199×10^{-3}	0.857	1.140
β_3 (Additional_Lane='Yes')	-0.478	0.183	9.322×10^{-3}	-0.818	-0.112

Os valores médios de todos os parâmetros “afastaram-se” de zero (o valor médio da distribuição *a priori* assumida), como também se pode observar pelas densidades *a posteriori* dos mesmos parâmetros na Figura 20.

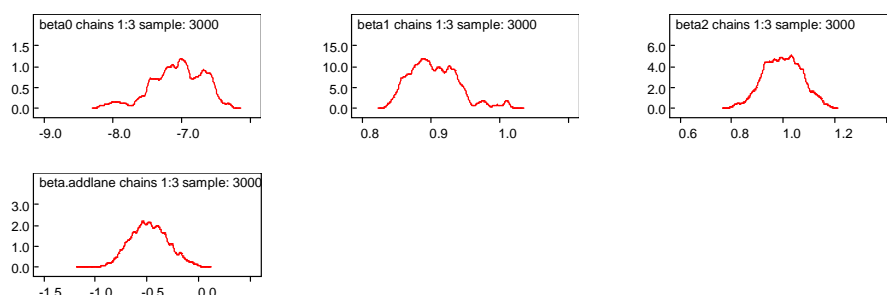


Figura 20| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *Nac* (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

A variável *Additional_Lane* tem um efeito negativo no número de acidentes; um trecho com via adicional reduz em cerca de 38% o número esperado de acidentes em relação a um trecho, em iguais condições, mas sem via adicional.

O número esperado de acidentes para valores médios de *log(AADT)* e *log(Length)* num período de cinco anos é de 23 e 14.45 em trechos sem e com via adicional, respectivamente, como se pode ver no Quadro 10.

Quadro 10| Médias *a posteriori* e desvios padrões para o número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 8.

	média (desvio padrão)
<i>Additional_Lane</i> ='No'	23.00 (1.382)
<i>Additional_Lane</i> ='Yes'	14.45 (2.513)

4.1.3.1 Validação do Modelo

As dezanove réplicas da distribuição *a posteriori* preditiva obtida pelo modelo Poisson Log-Normal estão dispostas em histogramas na Figura 21. Todas elas se assemelham muito ao histograma obtido pelos dados observados (no canto superior esquerdo da Figura 21), o que indica que as inferências resultantes do modelo Poisson Log-Normal são confiáveis.

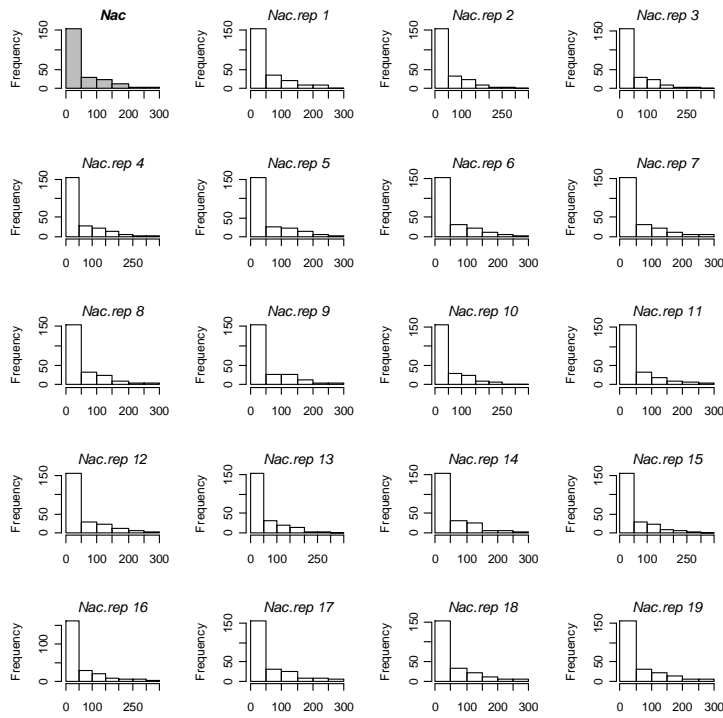


Figura 21| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

Para verificar o ajuste do modelo aos dados observados as distribuições *a posteriori* preditivas das medidas T são examinadas da seguinte maneira: as distribuições *a posteriori* preditivas das medidas T são “aproximadas” pelos histogramas de valores retirados a 3000 simulações dos parâmetros. São depois comparadas as distribuições dos valores de T , obtidas dos valores replicados, com os valores observados de T . Os resultados encontram-se exibidos na Figura 22 e na Figura 23, a última para T igual à razão entre a variância e a média.

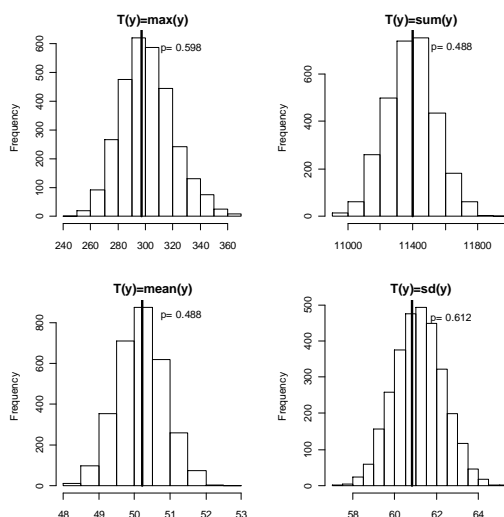


Figura 22| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Observação destas figuras sugere que o modelo Poisson Log-Normal gera valores preditivos semelhantes aos dados observados no estudo, isto é, as observações registadas em cinco anos são representativas das observações preditivas geradas pelo modelo.

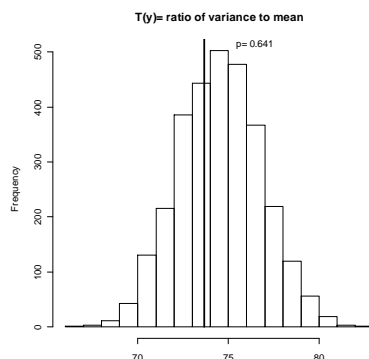


Figura 23| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais resulta num valor $\bar{D}=242.1$ (valor superior ao obtido pelo modelo Poisson-Gama) indicando um razoável ajuste do modelo ($n-p=224$). A estimativa do parâmetro de dispersão é 0.646. O valor DIC obtido é igual a 1556.52, valor ligeiramente mais elevado do que o resultante do modelo Poisson-Gama.

4.2 Modelação do Número de Mortos e Feridos Graves

A descrição detalhada das características de *NKSI*, aqui considerada como variável resposta, encontra-se na secção 4.2 de Azeredo Lopes e Cardoso (2009).

4.2.1 Modelo de Poisson

O modelo de Poisson ajustado aos números de mortos e feridos graves de acordo com as equações (3.22) que produz o menor valor de DIC contém as variáveis $\log(AADT)$, $\log(Length)$, $Additional_Lane$ e $Number_of_Lanes$. O número esperado de mortos e feridos graves, $\hat{\lambda}_i$ é dado por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Lane_Width_i + \beta_5 Number_of_Lanes_i \quad (4.4)$$

Os resultados apresentados foram obtidos de uma amostra de dimensão 3000 resultante de 30000 iterações das quais 20000 foram descartadas com um *thinning rate* igual a 10.

Os gráficos dos trace plots na Figura 24 não evidenciam suspeitas de não convergência.

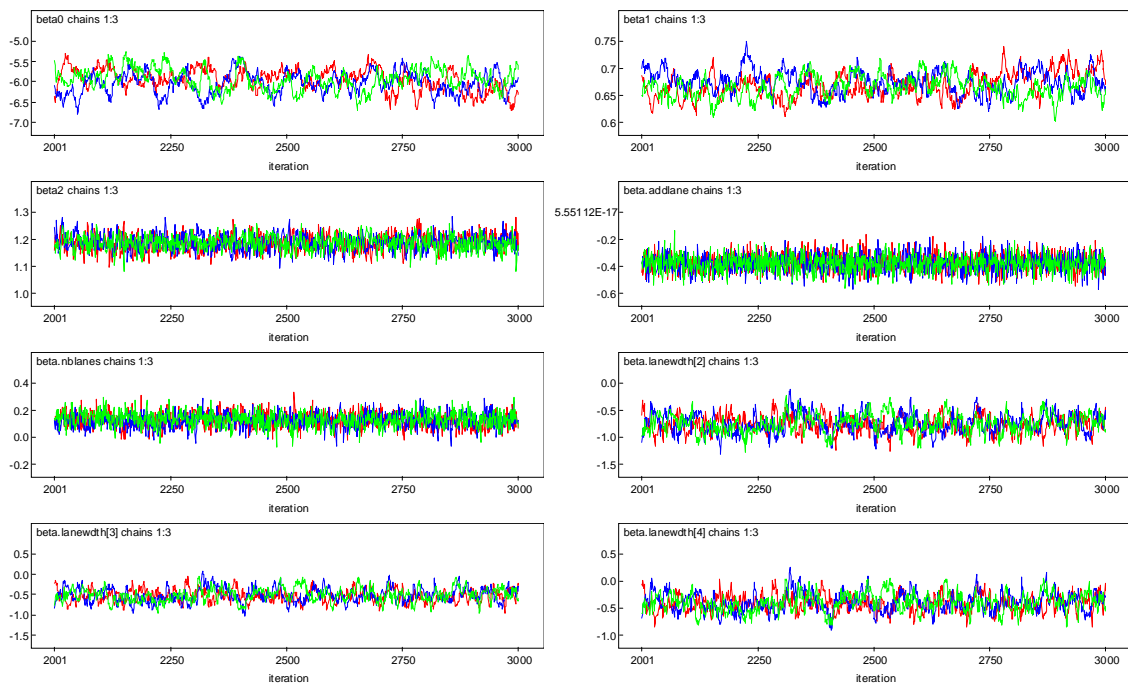


Figura 24| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Assim também como as observações dos gráficos de autocorrelação na Figura 25, embora haja evidência que tanto que as estimativas relativas a β_0 e a β_1 converjam mais lentamente que as restantes.

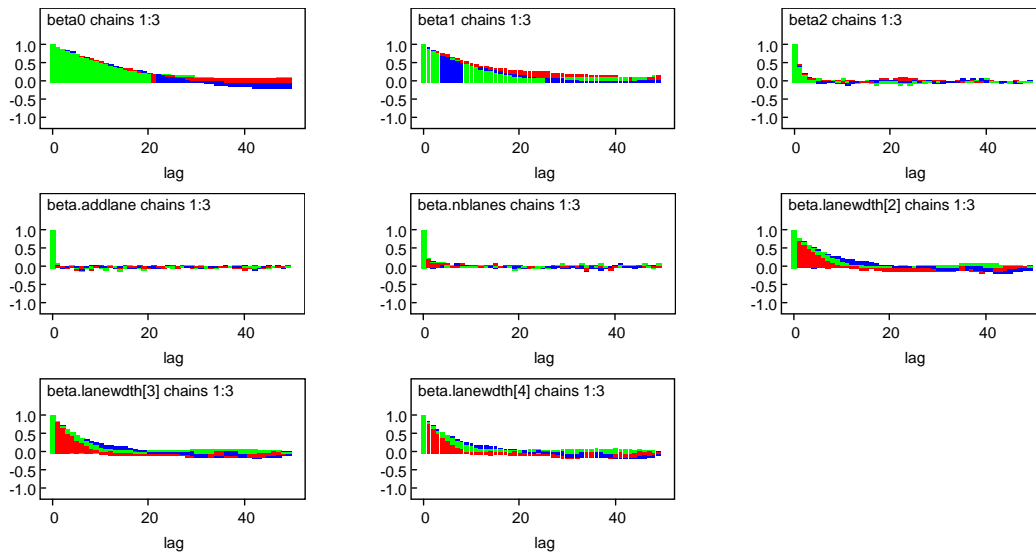


Figura 25] Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *NKS* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos das estatísticas de Gelman-Rubin também sugerem a convergência dos parâmetros como se pode observar na Figura 26, pois os valores de R tendem para 1.

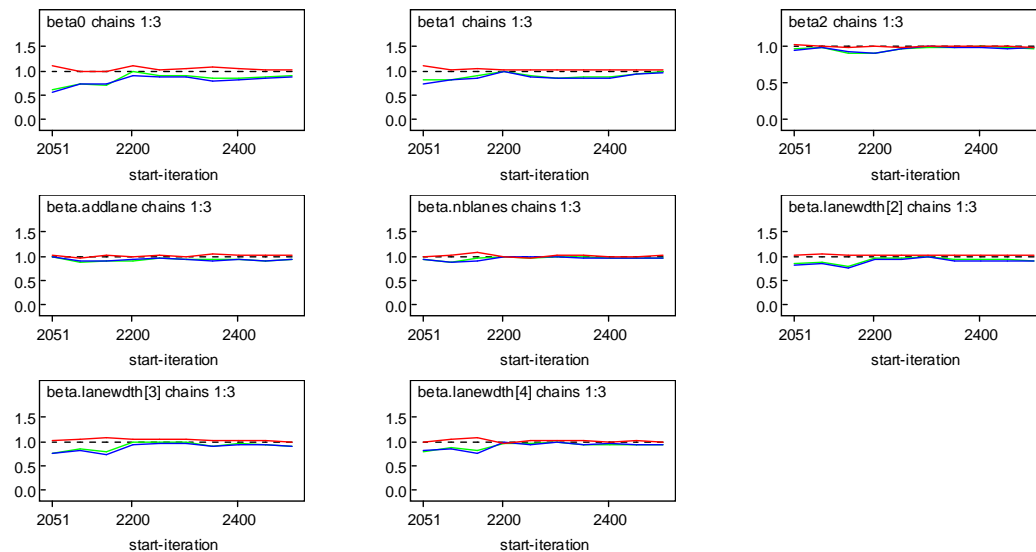


Figura 26] Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a *NKS* obtido do conjunto de dados excluindo as variáveis com valores omissos.

As equações para os números esperados de mortos e feridos graves encontram-se no Quadro 11. Os números destes eventos são mais elevados quando não existe via adicional e quando existem 3 ou 4 vias. Trechos por ordem crescente de número de mortos e feridos graves têm largura de via $B=[3.0, 3.50]$, $C=[3.50, 3.75]$, $D=[3.75, \infty[$ e $A=[0, 3.0]$.

Quadro 11| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo as variáveis com valores omissos.

	Números Esperados de Mortos e Feridos Graves
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='A'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.547 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'B'	$\hat{\lambda}_i = 1.201 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'C'	$\hat{\lambda}_i = 1.517 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'D'	$\hat{\lambda}_i = 1.710 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.890 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'B'	$\hat{\lambda}_i = 1.363 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'C'	$\hat{\lambda}_i = 1.722 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'D'	$\hat{\lambda}_i = 1.941 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='A'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 1.748 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'B'	$\hat{\lambda}_i = 8.241 \times 10^{-4} \times AADT_i^{0.669} \times Length_i^{1.186}$
'C'	$\hat{\lambda}_i = 1.041 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'D'	$\hat{\lambda}_i = 1.174 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 1.983 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'B'	$\hat{\lambda}_i = 9.353 \times 10^{-4} \times AADT_i^{0.669} \times Length_i^{1.186}$
'C'	$\hat{\lambda}_i = 1.181 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$
'D'	$\hat{\lambda}_i = 1.332 \times 10^{-3} \times AADT_i^{0.669} \times Length_i^{1.186}$

O Quadro 12 mostra as estimativas pontuais obtidas pelo modelo ajustado aos dados *NKSI*. Nenhum intervalo de credibilidade de 95% contém o valor zero, o que indica que as variáveis consideradas são relevantes para o modelo.

Os valores médios das densidades *a posteriori* das estimativas dos parâmetros afastaram-se consideravelmente de zero, tal como também se pode observar nos gráficos das densidades na Figura 27.

Quadro 12| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 11.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-5.973	0.268	2.100×10^{-2}	-6.506	-5.463
β_1 (log(AADT))	0.669	0.022	1.630×10^{-3}	0.629	0.714
β_2 (log(Length))	1.186	0.029	9.035×10^{-4}	1.131	1.242
β_3 (Additional_Lane='Yes')	-0.377	0.062	1.069×10^{-3}	-0.501	-0.254
β_4 (Lane_Width='B')	-0.752	0.170	9.146×10^{-3}	-1.074	-0.410
β_5 (Lane_Width='C')	-0.518	0.165	9.014×10^{-3}	-0.836	-0.172
β_6 (Lane_Width='D')	-0.398	0.167	9.222×10^{-3}	-0.719	-0.047
β_7 (Number_of_Lanes='B')	0.127	0.054	1.683×10^{-3}	0.022	0.229

A leitura da média das estimativas dos parâmetros β_j no Quadro 12 mostra que um trecho com via adicional reduz o número esperado de mortos e feridos graves em cerca de 31% em relação a um trecho sem via adicional, mas em iguais circunstâncias no que diz respeito às outras variáveis. Um trecho com três ou quatro vias (*Number_of_Lanes='B'*) tem o número esperado de mortos e feridos graves maior em aproximadamente 14% do que um trecho em circunstâncias semelhantes, mas com duas vias (*Number_of_Lanes='A'*).

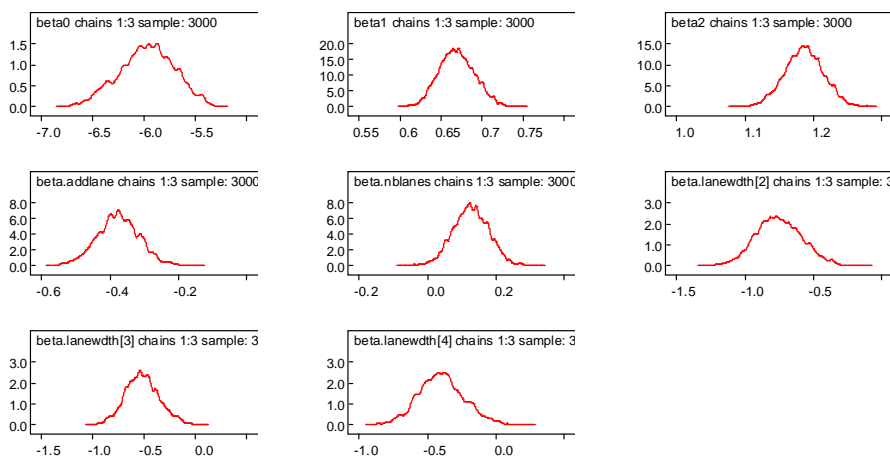


Figura 27| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NKSI* (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Um trecho com largura de via entre 3.00 e 3.50 metros (*Lane_Width='B'*) tem um número de mortos e feridos graves (*KSI*) cerca de 53% menor do que um trecho com largura de via menor que 3.00 metros (*Lane_Width='A'*). Trechos com largura de via entre 3.50 e 3.75 metros (*Lane_Width='C'*) e superior a 3.75 metros (*Lane_Width='D'*) têm, respectivamente, menores valores de *KSI* em cerca de 40% e 33% do que em relação a trechos com largura de via menor que 3.00 metros (*Lane_Width='A'*).

O Quadro 13 mostra os resultados dos números esperados de mortos e feridos graves, num período de cinco anos, para trechos de auto-estrada com uma representação média de *AADT* e *Length*. Pode constatar-se que o número esperado de mortos e feridos graves mais elevado (11.54) ocorre em

trechos sem via adicional, constituídos por 3 ou 4 vias, e cujas larguras são menores que 3 metros. Por outro lado, os trechos com menor número esperado de mortos e feridos graves são os trechos com duas vias, com larguras entre 3 e 3.50 metros e com via adicional. Nestes trechos são esperados, para um período de cinco anos, 3.26 mortos e feridos graves.

Quadro 13| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 11.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	10.17 (1.633)	'A'	11.54 (1.813)
'B'	4.74 (0.314)	'B'	5.39 (0.370)
'C'	5.98 (0.255)	'C'	6.80 (0.401)
'D'	6.75 (0.324)	'D'	7.66 (0.445)
<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	6.99 (1.177)	'A'	7.93 (1.335)
'B'	3.26 (0.279)	'B'	3.73 (0.346)
'C'	4.11 (0.268)	'C'	4.67 (0.390)
'D'	4.64 (0.342)	'D'	5.27 (0.459)

4.2.1.1 Validação do Modelo

A Figura 28 contém os histogramas de dezanove réplicas da distribuição *a posteriori* preditiva obtida pelo modelo de Poisson. Alguns histogramas dessas réplicas, nomeadamente *NKSI.rep* 6, 7, 10 e 18, diferem (visualmente) do histograma de *NKSI* (a cinzento no canto superior esquerdo) o que leva a admitir que as inferências obtidas com este modelo possam não ser confiáveis.

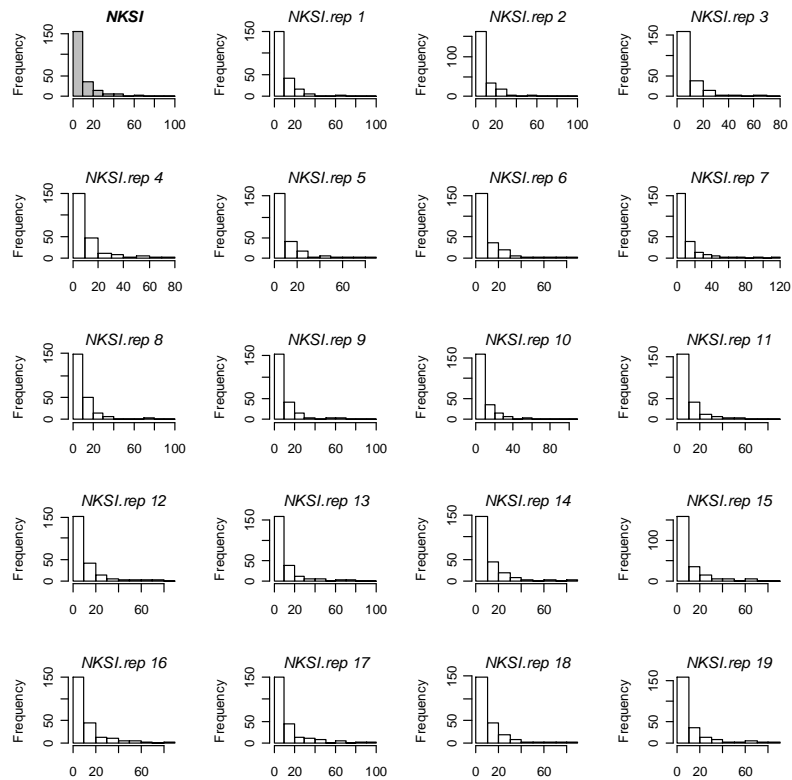


Figura 28| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

Os histogramas das distribuições *a posteriori* das quatro medidas *T* consideradas encontram-se na Figura 29. Tal como se pode observar, o modelo de Poisson ajustado a *NKSI* não produz valores preditivos semelhantes aos dados observados, nomeadamente no que respeita às medidas relativas aos valores máximo e desvio padrão.

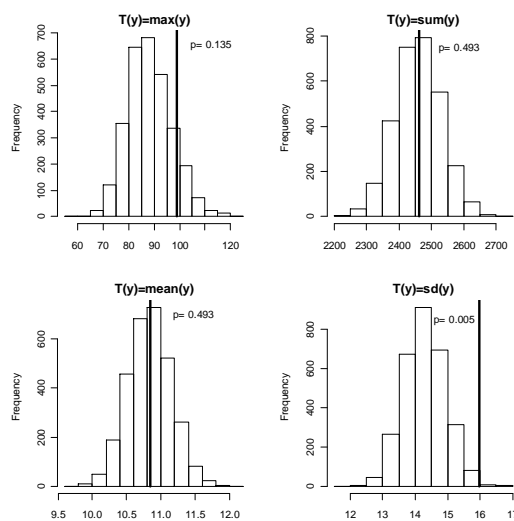


Figura 29| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O *p* é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

A razão entre a variância e a média observada nos dados também não é reflectida nos valores obtidos pelo modelo, como se pode constatar pela observação da Figura 30. A probabilidade de que a razão entre a variância e a média nos dados replicados seja maior do que a mesma quantidade obtida pelos dados observados é de apenas 0.002. Este facto leva a concluir que o modelo de Poisson não gera valores que reflectam a sobredispersão existente nos dados *NKSI*.

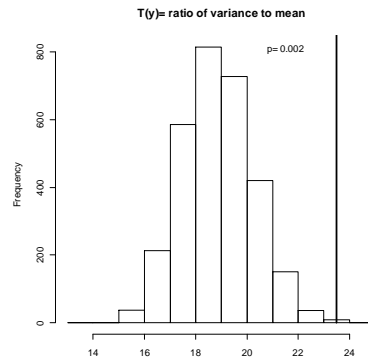


Figura 30| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais $\bar{D} = 992.2$, valor elevado quando comparado a $n-p=222$, o que não indica que o modelo esteja bem ajustado aos dados. O valor de DIC obtido é igual a 1828.77.

4.2.2 Modelo de Poisson-Gama

O modelo Poisson-Gama que produz o menor valor de DIC quando ajustado (ver equações 3.27) a *NKSI* envolve as covariáveis $\log(AADT)$, $\log(Length)$ e *Additional_Lane*. O número esperado de mortos e feridos graves, $\hat{\lambda}_i$, é dado pela equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i \quad (4.5)$$

Os resultados apresentados foram obtidos de uma amostra com dimensão 3000, após simulações de três cadeias de Markov com 30000 iterações das quais 20000 foram descartadas com um *thinning rate* igual a 10.

Os gráficos *trace plots* e das autocorrelações (ver Figura 31 e Figura 32) mostram que a densidade *a posteriori* da estimativa do parâmetro correspondente a *Additional_Lane* foi a que melhor atingiu a distribuição alvo.

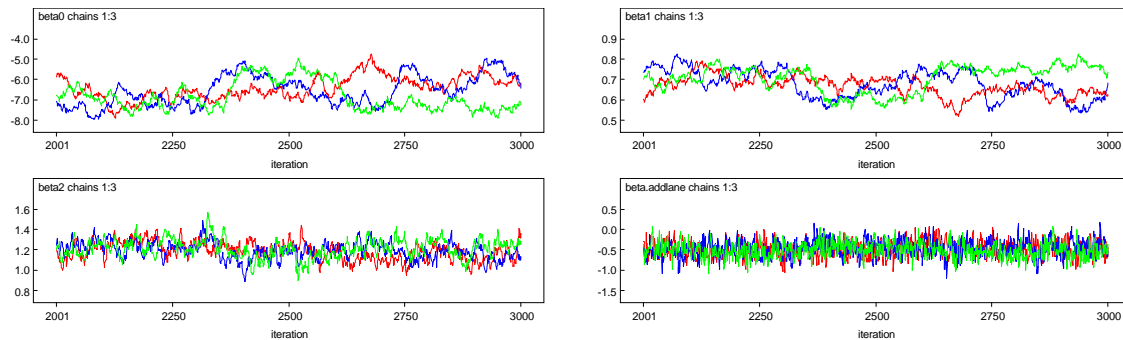


Figura 31| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Pela leitura dos gráficos de autocorrelação na Figura 32 pode-se observar que a convergência de β_0 , β_1 e β_2 é bastante lenta.

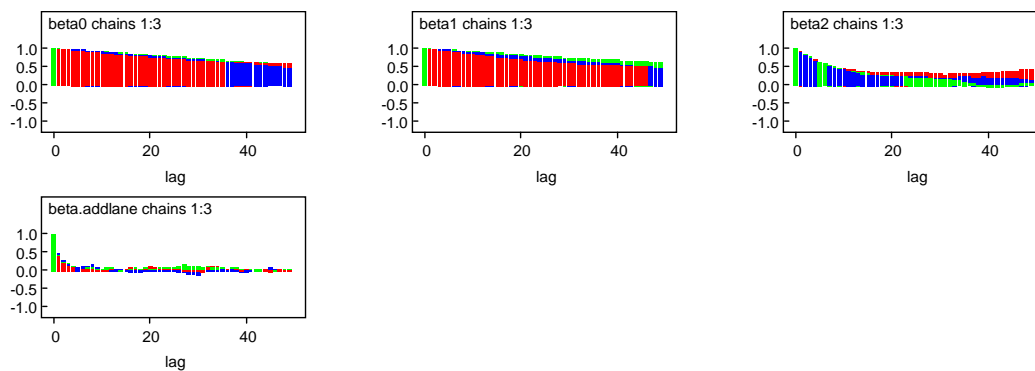


Figura 32| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos das estatísticas de Gelman-Rubin na Figura 33 mostram que os valores da estatística R converge para 1 em todos os parâmetros, o que leva a crer que se obtém a convergência desejada.

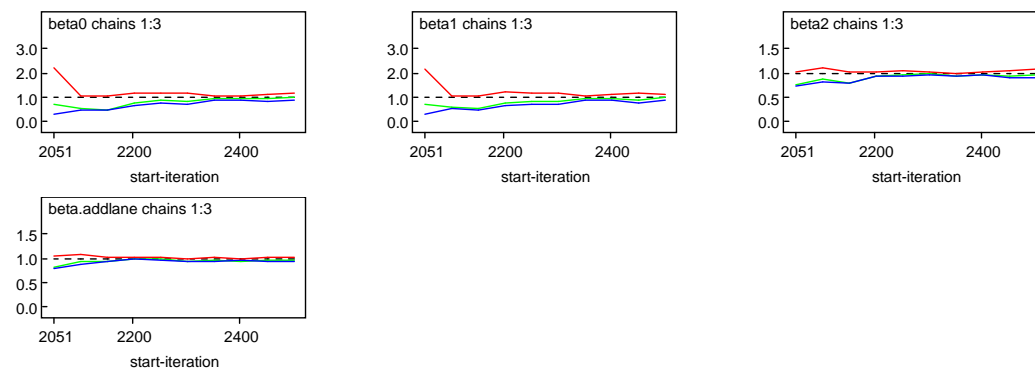


Figura 33| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

As equações dos números esperados de mortos e feridos graves por cada nível da covariável *Additional_Lane* estão expostos no Quadro 14.

Quadro 14| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo as variáveis com valores omissos.

Números Esperados de Mortos e Feridos Graves	
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 1.359 \times 10^{-3} \times AADT_i^{0.685} \times Length_i^{1.195}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 8.348 \times 10^{-4} \times AADT_i^{0.685} \times Length_i^{1.195}$

As estimativas pontuais para o modelo são dadas no Quadro 15. Pode afirmar-se que um trecho com via adicional tem menor número esperado de mortos e feridos graves em cerca de 39%, do que um trecho em iguais condições, mas sem via adicional.

Quadro 15| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 14.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-6.601	0.691	6.790×10^{-2}	-7.661	-5.243
β_1 (log(AADT))	0.685	0.063	6.157×10^{-3}	0.563	0.787
β_2 (log(Length))	1.195	0.089	6.642×10^{-3}	1.024	1.364
β_3 (<i>Additional_Lane</i> ='Yes')	-0.488	0.197	7.526×10^{-3}	-0.873	-0.097

Os gráficos das densidades *a posteriori* para cada parâmetro β_j na Figura 34 mostram que os correspondentes valores médios se afastaram de zero (a média considerada *a priori*).

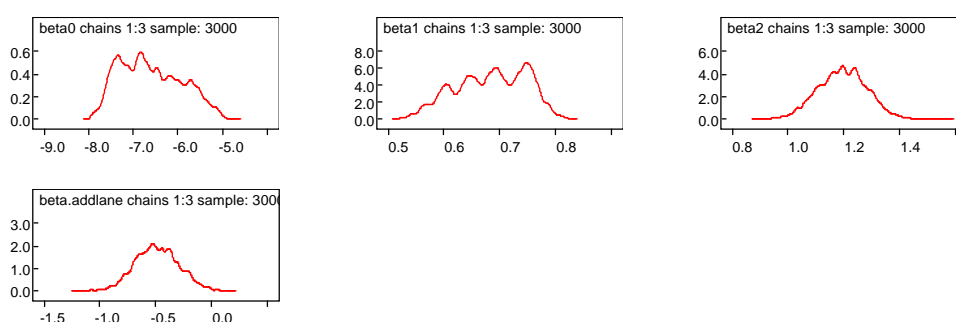


Figura 34| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a NKS/ (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Para um trecho de auto-estrada com uma representação média de *AADT* e *Length* e com via adicional, espera-se a ocorrência de cerca de 4 mortos e feridos graves num período de cinco anos. Tal como se pode observar no Quadro 16.

Quadro 16| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 14.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	6.315 (0.444)
<i>Additional_Lane='Yes'</i>	3.933 (0.759)

Para trechos nas mesmas condições, mas sem via adicional, esperam-se 6.3 mortos e feridos graves.

4.2.2.1 Validação do Modelo

A amostra de dezanove exemplos de dados preditivos replicados pelo modelo Poisson-Gama cujos histogramas podem ser observados na Figura 35 deixam crer que as inferências resultantes deste modelo são confiáveis, pois em geral os histogramas dos dados replicados se assemelham ao histograma dos dados observados.

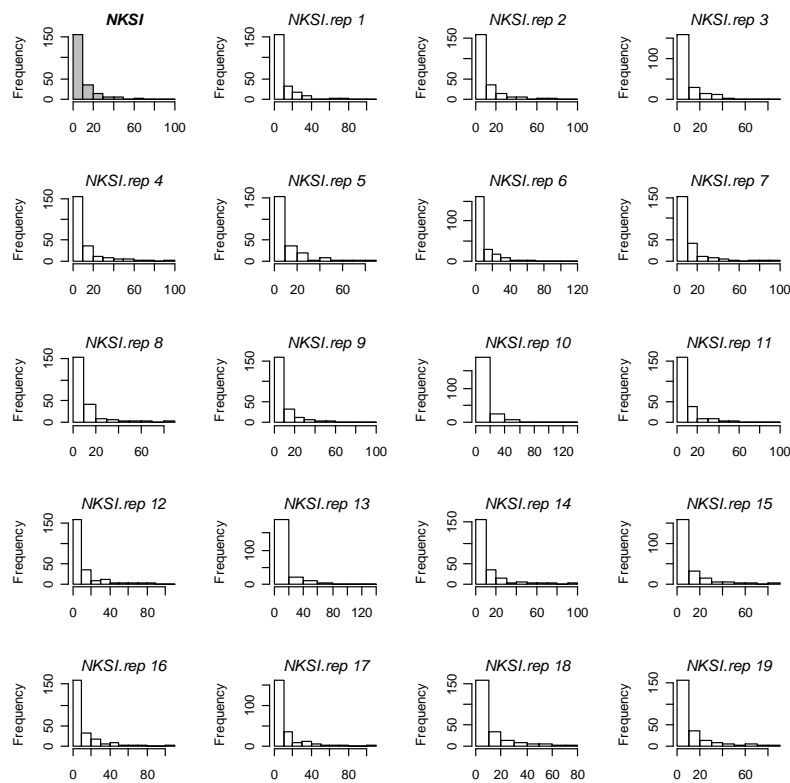


Figura 35| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

As quantidades escolhidas para medir possíveis discrepâncias entre os dados replicados e os dados observados, que se podem observar nos histogramas da Figura 36 e da Figura 37, sugerem que o modelo de Poisson-Gama gera dados que são representativos dos dados observados.

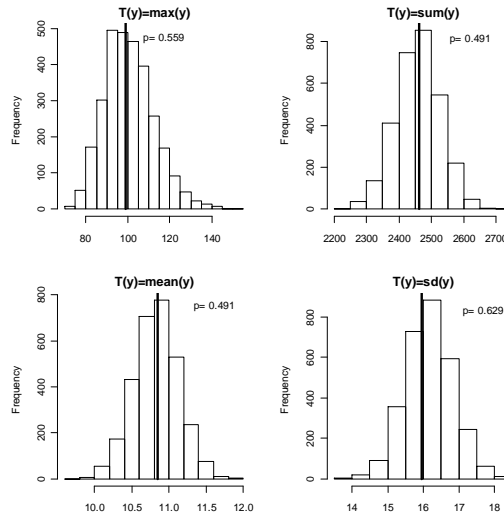


Figura 36| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A sobredispersão dos dados também é tida em consideração por este modelo (ver Figura 37) onde o valor de p_B é igual a 0.648. A estimativa do parâmetro de dispersão é igual a 0.668.

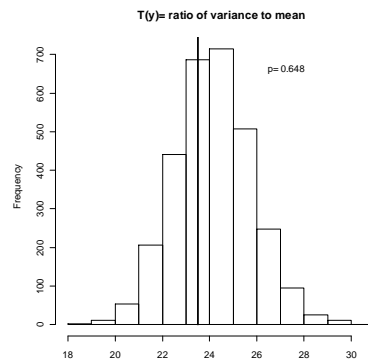


Figura 37| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais é igual a $\bar{D}=198.6$, valor consideravelmente mais baixo do que o obtido pela distribuição de Poisson e mais aproximado a 224 ($n-p$). O valor de DIC é igual a 1094.01.

4.2.3 Modelo de Poisson Log-Normal

O modelo Poisson Log-Normal ajustado aos dados referentes ao número de mortos e feridos graves contém, além da exposição e do comprimento do trecho também a variável *Additional_Lane*. Assim, o número esperado de mortos e feridos graves em cada trecho i , $\hat{\lambda}_i$, é dado por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (4.6)$$

Os resultados apresentados foram obtidos de uma amostra de dimensão 3000 resultante de uma simulação simultânea de 3 cadeias de Markov com 30000 iterações das quais 20000 foram descartadas com *thinning rate* igual a 10.

Os gráficos dos *trace plots* na Figura 38 sugerem uma convergência não adequada dos parâmetros β_0 e β_1 .

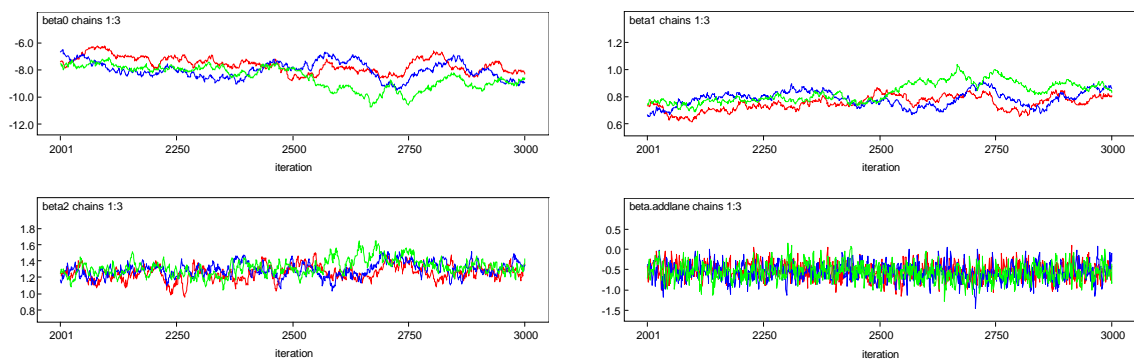


Figura 38| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações na Figura 39 mostram que os parâmetros β_0 , β_1 e β_2 convergem lentamente.

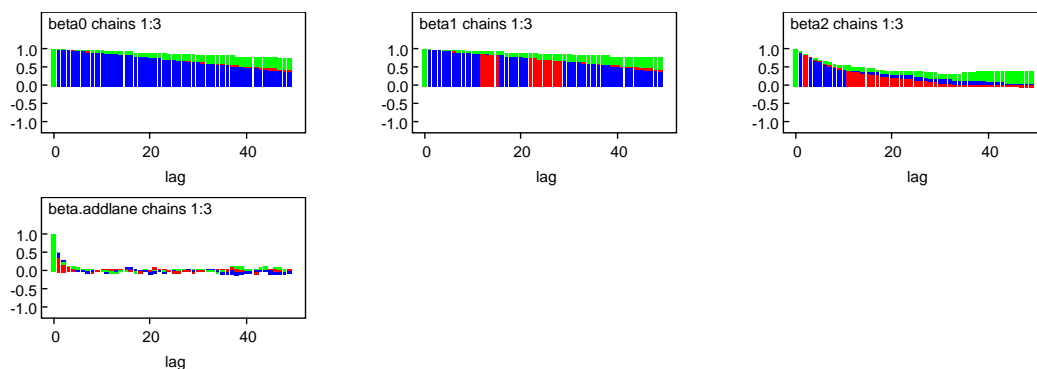


Figura 39| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos se encontram na Figura 40, mostram que a convergência não é atingida muito adequadamente para β_0 e β_1 . Possivelmente ter-se-ia que deixar o algoritmo correr por mais um considerável número de iterações.

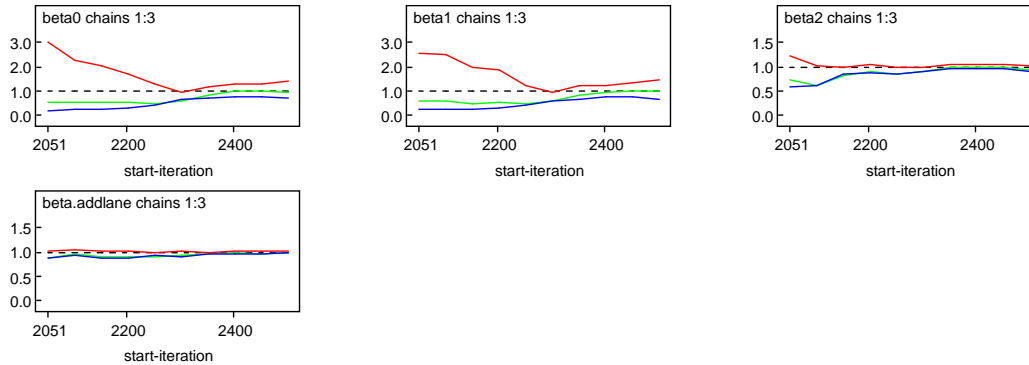


Figura 40| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKS* obtido do conjunto de dados excluindo as variáveis com valores omissos.

As equações que dão os números esperados de mortos e feridos graves resultantes do modelo Poisson Log-Normal, para cada nível de *Additional_Lane*, encontram-se no Quadro 17.

Quadro 17| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo as variáveis com valores omissos.

	Números Esperados de Mortos e Feridos Graves
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 3.148 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.291}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 1.816 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.291}$

As estimativas pontuais deste modelo são dadas no Quadro 18. Pela interpretação da estimativa do parâmetro β_3 pode afirmar-se que um trecho com via adicional tem número esperado de *KSI* 42% menor do que um trecho em iguais condições mas sem via adicional.

Quadro 18| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 17.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.064	0.781	7.795×10^{-2}	-9.862	-6.753
β_1 (log(AADT))	0.789	0.070	6.977×10^{-3}	0.662	0.941
β_2 (log(Length))	1.291	0.094	7.380×10^{-3}	1.111	1.493
β_3 (<i>Additional_Lane</i> ='Yes')	-0.550	0.206	6.700×10^{-3}	-0.964	-0.139

Os gráficos das densidades *a posteriori* das estimativas dos parâmetros β_j podem ser observados na Figura 41.

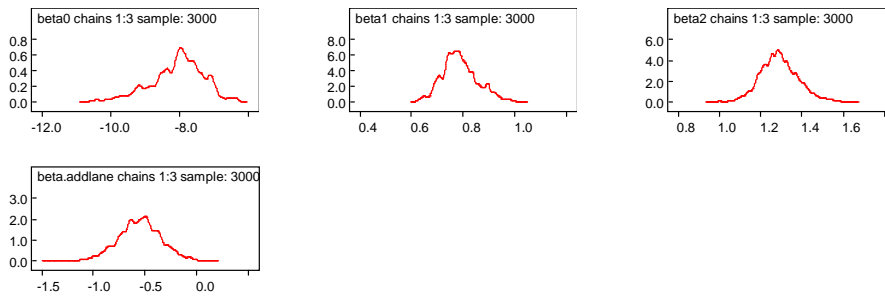


Figura 41| Densidades *a posteriori* dos parâmetros β_i do modelo Poisson Log-Normal ajustado a *NKSI* (obtido do conjunto de dados excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de mortos e feridos graves, para um período de cinco anos, para representações médias de *AADT* e *Length* é de aproximadamente 4.6 e 2.7, para trechos sem e com via adicional, respectivamente, como pode ser observado no Quadro 19.

Quadro 19| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 17.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	4.566 (0.350)
<i>Additional_Lane='Yes'</i>	2.680 (0.538)

4.2.3.1 Validação do Modelo

Pela observação dos histogramas de Figura 42 não há razões para suspeitar da adequabilidade do modelo Poisson Log-Normal para fazer inferências sobre o número esperado de mortos e feridos graves.

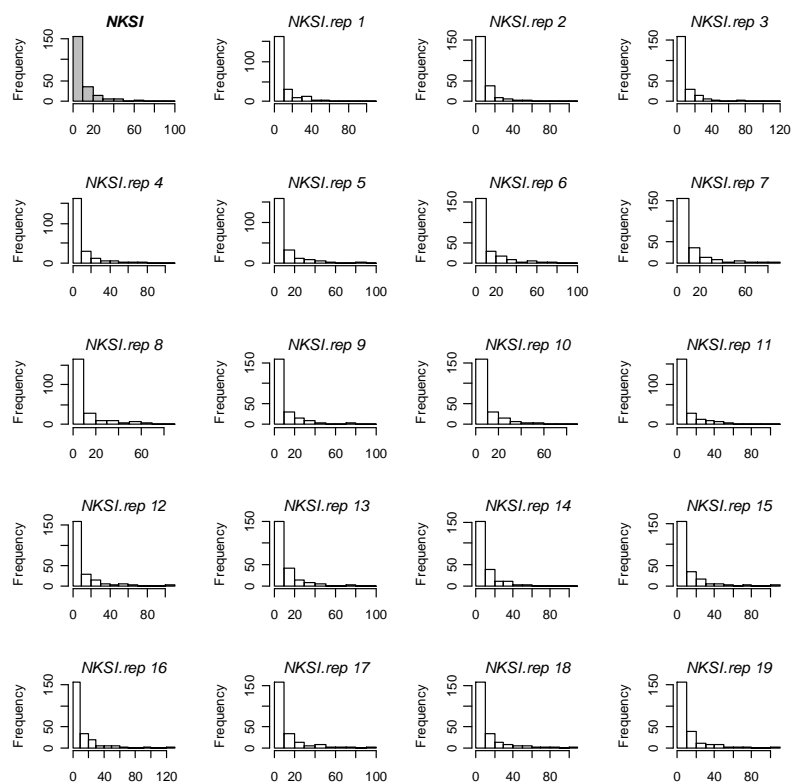


Figura 42| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados onde se excluíram as variáveis com valores omissos.

As quantidades escolhidas para medir possíveis discrepâncias entre dados replicados pelo modelo e os dados observados foram obtidas das distribuições preditivas *a posteriori* e os correspondentes histogramas encontram-se exibidos na Figura 43. Uma vez que todos os histogramas “englobam” a recta vertical pode concluir-se que o modelo Poisson Log-Normal gera dados representativos dos dados observados.

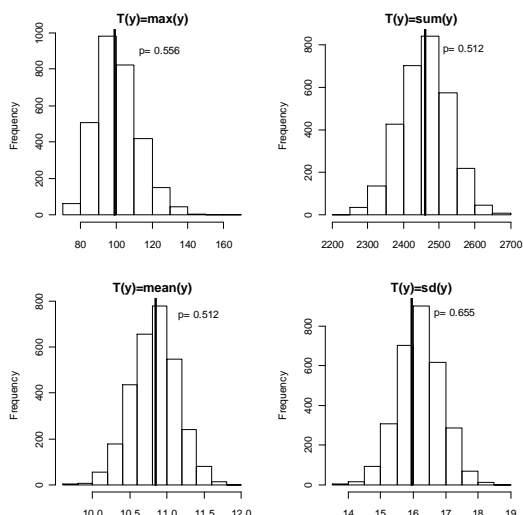


Figura 43| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A sobredispersão existente nos dados também é tida em conta pelos dados replicados como se pode constatar pela observação da Figura 44. A estimativa do parâmetro de dispersão $1/\sigma$ na equação (3.31) é igual a 0.656.

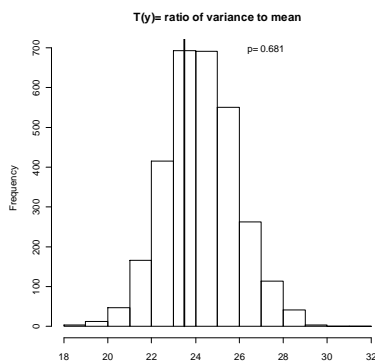


Figura 44| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais produz um valor $\bar{D} = 205.1$, razoável quando comparado com $n-p=224$, o que indica que a sobredispersão é adequadamente modelada. O valor de DIC é igual a 1117.35.

4.3 Discussão

Neste capítulo descreveram-se ajustes de modelos de regressão de Poisson e de dois modelos hierárquicos baseados na distribuição Poisson-Gama e Poisson Log-Normal, empregando uma abordagem Bayesiana. Os dados analisados foram referentes a números de acidentes (*Nac*) e a números de mortos e feridos graves (*NKSI*), onde as variáveis com valores omissos não foram consideradas para inserção nos respectivos modelos. Somente *Lane_Width*, *Number_of_Lanes* e *Additional_Lane* foram tidas em conta como possíveis covariáveis.

O ajuste dos modelos de Poisson a ambos *Nac* e *NKSI* resultou na selecção de $\log(AADT)$, $\log(Length)$, *Additional_Lane* e *Number_of_Lanes* como covariáveis influentes.

Os modelos Poisson-Gama e Poisson Log-Normal, para ambos *Nac* e *NKSI* contêm $\log(AADT)$, $\log(Length)$ e *Additional_Lane* como covariáveis.

O modelo de regressão de Poisson mostrou não ser apropriado para modelar ambos *Nac* e *NKSI*, visto os dados replicados por estes modelos, e referentes à distribuição preditiva *a posteriori*, não serem representativos dos dados observados. O exame gráfico e os *p*-values Bayesianos de quantidades $T(p_B)$, nomeadamente os valores máximos e desvios padrões, assim como a razão entre a variância e a média, mostraram que o modelo de Poisson não capta a sobredispersão nos dados. Este facto é também corroborado pelos valores das médias das desviâncias amostrais obtidas (\bar{D}).

A sobredispersão é tida em conta por ambos os modelos Poisson-Gama e Poisson Log-Normal; no entanto, em *Nac* e *NKSI* o modelo Poisson-Gama mostrou ser o mais adequado visto ter apresentado os menores valores de DIC e os mais apropriados valores das médias das observações amostrais.

No que respeita a *Nac*, e tendo em conta os resultados obtidos pelo ajuste do modelo Poisson-Gama, pode afirmar-se que um trecho de auto-estrada com uma via adicional tem menor número esperado de acidentes (cerca de 33%) do que um trecho em iguais condições, mas sem via adicional. Uma via adicional está também associada a um menor número esperado de mortos e feridos graves (*NKSI*), em cerca de 39%.

5 | MODELAÇÃO DOS DADOS EXCLUINDO OS TRECHOS COM VALORES OMISSOS

A análise estatística descrita no presente capítulo consiste na modelação, selecção e ajuste de modelos a dados obtidos em auto-estradas portuguesas, com exclusão dos trechos com um ou mais valores omissos. Acrescem às variáveis tidas em consideração como possíveis covariáveis no Capítulo 4 as variáveis *Left_Shoulder_Width*, *Right_Shoulder_Width*, *Median_Type* e *Median_Width*. Os dados analisados consistem em 198 trechos de auto-estrada. Tal como no Capítulo 5 em Azeredo Lopes e Cardoso (2009), foi decidido modelar os dados tendo como variáveis resposta o número de acidentes (*Nac*) e o número de mortos e feridos graves (*NKSI*).

5.1 Modelação do Número de Acidentes

A descrição detalhada das características do número de acidentes (*Nac*), *AADT*, *Length* e outras variáveis tidas em consideração como possíveis covariáveis encontra-se na secção 5.1 de Azeredo Lopes e Cardoso (2009).

5.1.1 Modelo de Poisson

Uma vez ajustado o modelo completo (*full model*) ao conjunto das equações (3.22) verificou-se que as variáveis cujo intervalo de credibilidade de 95% para as correspondentes estimativas dos parâmetros β_j continham o valor zero eram: *Lane_Width*, *Nblanes* e *Right_Shoulder_Width*. Um novo ajuste foi feito ignorando estas variáveis. O processo de remover passo a passo as variáveis restantes (*stepwise*), registando os valores de DIC obtidos, resultou no modelo com as variáveis *Additional_Lane* e *Median_Type*, além evidentemente de $\log(AADT)$ e $\log(Length)$ como obtendo o valor de DIC mais baixo. A equação na linha do meio em (3.22) pode agora ser representada por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Median_Type_i \quad (5.1)$$

O algoritmo de MCMC converge após 30000 iterações com 20000 iterações *burn-in*, isto é, inicialmente rejeitadas e com um *thinning rate* igual a 10, resultando em amostras com dimensão igual a 3000. A observação dos *trace plots* da Figura 45 para os parâmetros β_j mostra que todos os valores obtidos para cada parâmetro se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências, facto que corrobora a convergência do algoritmo.

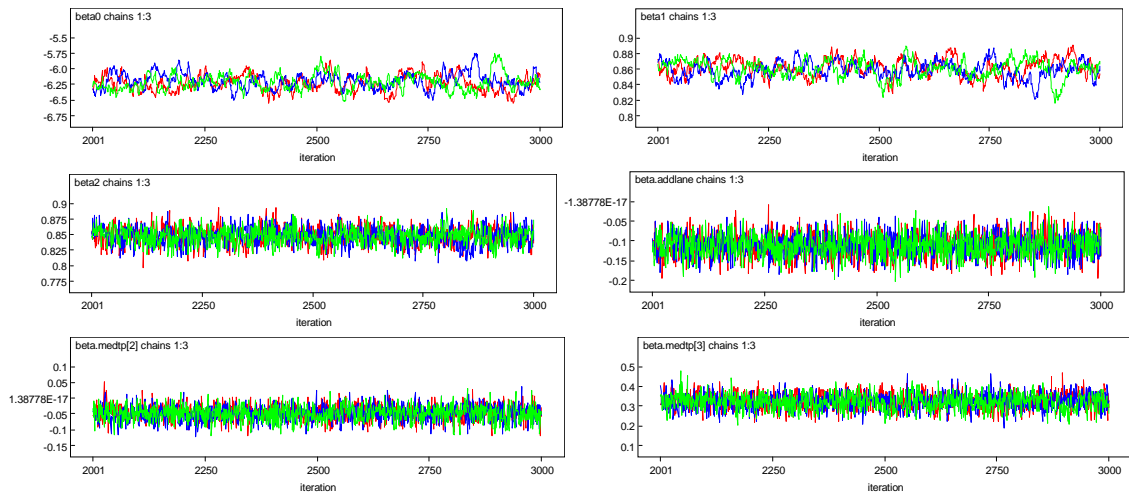


Figura 45] Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos das autocorrelações das estimativas dos parâmetros β_j , reproduzidos na Figura 46 indicam uma convergência lenta de β_0 e β_1 e mais rápida das restantes estimativas.

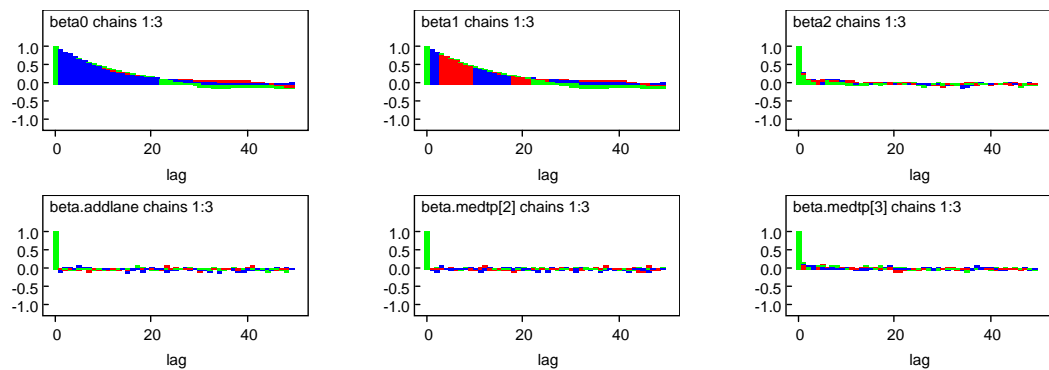


Figura 46] Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 47. O algoritmo converge uma vez que R (ver equação 3.12) tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam ao aumentar o número de iterações.

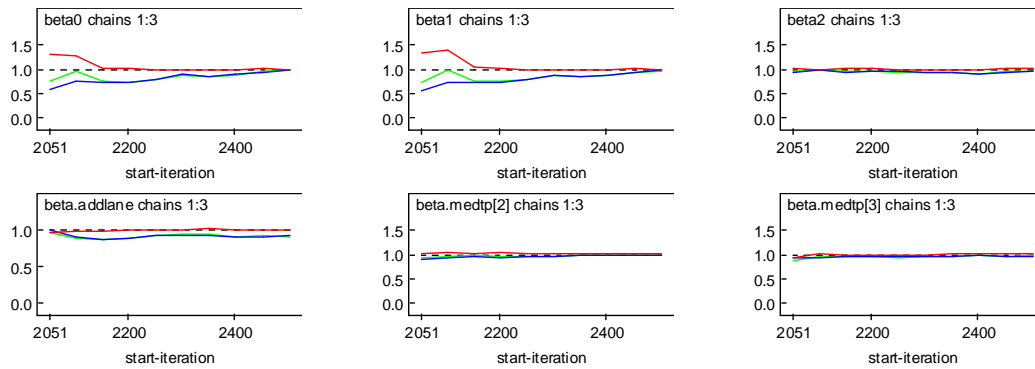


Figura 47| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a *Nac* obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações do modelo, estimado *a posteriori*, para os números esperados de acidentes, num período de cinco anos, de acordo com os níveis das variáveis categóricas *Additional_Lane* e *Median_Type* encontram-se no Quadro 20. Pode observar-se que o número esperado de acidentes num determinado trecho de estradas é menor quando existe uma via adicional e quando o separador é do tipo “GM” (ligeiramente menor do que quando o separador é do tipo “DNJ;NJ”) o que é contraditório com resultados obtidos em Azeredo Lopes e Cardoso (2009) onde o tipo de separador com um menor número de acidentes era o tipo “DNJ;NJ” (ver secção 5.1 do referido artigo).

Quadro 20| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo os trechos com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 2.000 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$
'GM'	$\hat{\lambda}_i = 1.907 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$
'Vala'	$\hat{\lambda}_i = 2.762 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$
<i>Additional_Lane</i> ='Yes'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 1.789 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$
'GM'	$\hat{\lambda}_i = 1.705 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$
'Vala'	$\hat{\lambda}_i = 2.469 \times 10^{-3} \times AADT_i^{0.860} \times Length_i^{0.848}$

O Quadro 21 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% das estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pela equação (5.1).

Quadro 21| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 20.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-6.214	0.129	9.413×10^{-3}	-6.440	-5.951
β_1 (log(AADT))	0.860	0.011	8.179×10^{-4}	0.837	0.880
β_2 (log(Length))	0.848	0.014	4.510×10^{-4}	0.819	0.875
β_3 (Additional_Lane='Yes')	-0.112	0.029	5.816×10^{-4}	-0.168	-0.054
β_4 (Median_Type='GM')	-0.048	0.024	5.282×10^{-4}	-0.097	-8.271×10^{-4}
β_5 (Median_Type='Vala')	0.322	0.041	1.229×10^{-3}	0.242	0.400

Os intervalos de credibilidade de 95% para (todas as estimativas dos parâmetros) β_j não incluem o valor zero, o que indica que todas as covariáveis consideradas têm um efeito significativo no número de acidentes, como se pode observar também nos gráficos da densidade *a posteriori* das estimativas dos parâmetros β_j da Figura 48. Pela leitura dos valores do Quadro 21 pode também depreender-se que uma via adicional reduz o número de acidentes em cerca de 11%. Num trecho com separador do tipo 'GM' é *a posteriori* esperado um número de acidentes inferior em 5% do que um trecho com separador do tipo 'DNJ;NJ'. Num trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 38% mais acidentes do que um trecho com separador do tipo 'DNJ;NJ'.

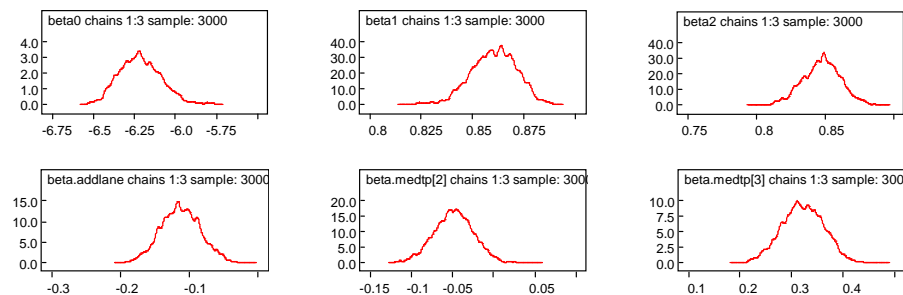


Figura 48| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *Nac* (obtido do conjunto de dados excluindo os trechos com valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de acidentes para os diversos tipos de separador e existência, ou não, de vias adicionais para uma representação média de um trecho de auto-estrada, foi calculado e é dado pelos valores indicados no Quadro 22.

Quadro 22| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 20.

Additional_Lane='No'	média (desvio padrão)	Additional_Lane='Yes'	média (desvio padrão)
Median_Type		Median_Type	
'DNJ;NJ'	25.65 (0.615)	'DNJ;NJ'	22.94 (0.817)
'GM'	24.44 (0.509)	'GM'	21.85 (0.634)
'Vala'	35.41 (1.184)	'Vala'	31.67 (1.377)

Para um trecho típico de auto-estrada sem via adicional e com um separador do tipo 'DNJ;NJ', são esperados, para um período de cinco anos, 25.65 acidentes, enquanto que para um trecho com separadores do tipo 'GM' e 'Vala' são esperados 24.44 e 35.41 acidentes, respectivamente. Para trechos com uma via adicional são esperados 22.94, 21.85 e 31.67 acidentes quando esses trechos possuírem separadores do tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente.

5.1.1.1 Validação do Modelo

Segundo Gelman *et al.* (2004) se o modelo está bem ajustado, então dados obtidos através do modelo, isto é, dados replicados devem ser semelhantes aos dados observados. A Figura 49 mostra vinte histogramas, estando no canto superior esquerdo o histograma dos dados observados (a cinzento) e todos os outros são histogramas de dados replicados da distribuição *a posteriori* preditiva. Observação desta figura mostra que vários histogramas das réplicas diferem consideravelmente do histograma dos dados observados, em particular os histogramas *Nac.rep 2, 8, 9 e 16*.

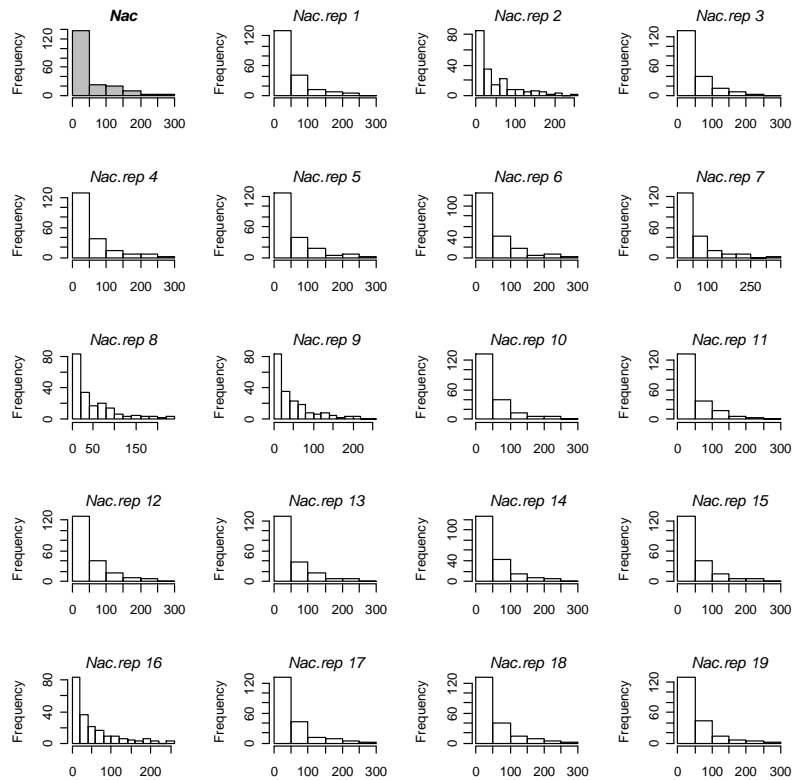


Figura 49| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas ("*Nac.rep*") dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados excluindo os trechos com valores omissos.

Para medir a discrepância foram comparados o maior valor, o valor médio, o desvio padrão e a soma de todos os valores em cada conjunto de dados replicados com os valores correspondentes dos dados observados.

Cada gráfico na Figura 50 representa o histograma de cada um destes valores e as rectas verticais indicam os valores obtidos pelos dados observados. Embora o modelo de Poisson capture a variação que os dados observados aparentam (em particular a soma e o valor médio), não captura para as outras quantidades, nomeadamente o valor máximo, $\max(y)$, e o desvio padrão, $sd(y)$, dos valores observados.

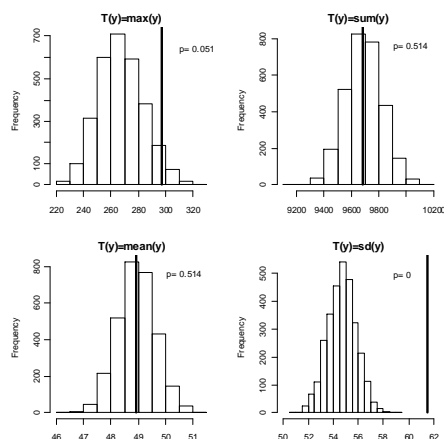


Figura 50| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson (3000 simulações). As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Segundo Congdon (2005), para além da selecção de variáveis, existem outros aspectos dos dados que necessitam de ser tidos em conta, como é o caso da adequação do modelo de Poisson na modelação da sobredispersão dos dados. A variância dos números de acidentes observados, y_i , é igual a 3788.32 sendo a média, 48.92, o que indica sobredispersão. As predições de um modelo que seja adequado a estes dados deverá reflectir esta sobredispersão.

Como se pode observar na Figura 51, a razão entre a variância e a média (igual a 77.44 nos dados observados e representada pela recta vertical) também não é capturada pelo modelo de regressão de Poisson, cujos valores obtidos para a razão entre a variância e o valor médio de vários conjuntos de dados replicados pelo modelo são representados pelo histograma da Figura 51.

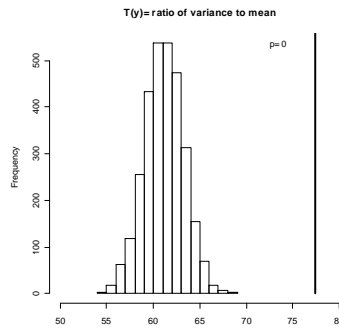


Figura 51| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Em média, a probabilidade de que a razão T entre a variância e a média dos valores replicados pelo modelo seja superior à mesma razão, obtida pelos dados (77.44) é igual a zero, isto é, p_B dado pela equação (3.14) é igual a zero, o que indica a existência de uma enorme discrepância entre os dados observados e os dados replicados pelo modelo. Consequentemente, há evidência de que a sobredispersão não está a ser adequadamente modelada pelo modelo de regressão de Poisson. A média das desviâncias amostrais segundo a equação (3.15) resulta em $\bar{D} = 2495$, um valor bastante elevado quando comparado com os graus de liberdade $n-p$ (194), evidenciando também a não captação da sobredispersão pelo modelo considerado.

Este modelo produz um valor de DIC igual a 3520.26.

5.1.2 Modelo de Poisson-Gama

Uma vez ajustado o modelo completo (*full model*) ao conjunto das equações (3.27) verificou-se que as variáveis cujo intervalo de credibilidade de 95% para as correspondentes estimativas dos parâmetros β_j continham o valor zero eram: *Lane_Width* e *Right_Shoulder_Width*. Um novo ajuste foi feito ignorando estas variáveis, e o processo de remover passo a passo as variáveis restantes registando o menor valor de DIC obtido resultou no modelo com as variáveis *Additional_Lane* e *Median_Type*, além evidentemente de *log(AADT)* e *log(Length)*. A equação genérica que dá o número de acidentes é:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(\text{AADT}_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i + \beta_4 \text{Median_Type}_i \quad (5.2)$$

O algoritmo de MCMC converge após 40000 iterações com 10000 iterações *burn-in*, isto é, inicialmente rejeitadas e com *thinning rate* igual a 5, resultando em amostras com dimensão igual a 18000. A observação dos *trace plots* da Figura 52 para as estimativas dos parâmetros β_j , mostra que todos os valores obtidos para cada parâmetro se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências.

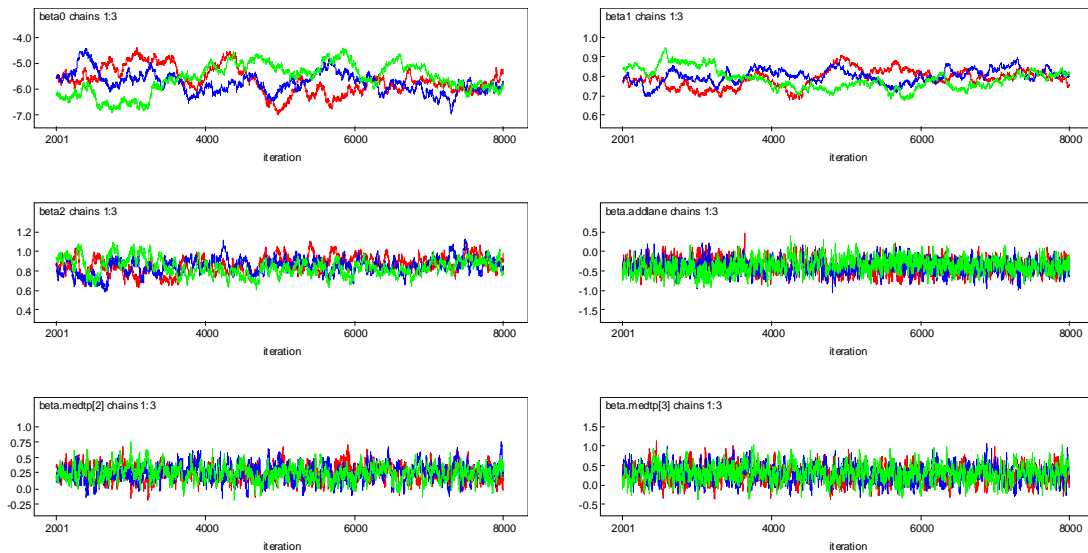


Figura 52| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos das autocorrelações das estimativas dos parâmetros β_j , reproduzidos na Figura 53 indicam uma convergência mais lenta de β_0 , β_1 e β_2 .

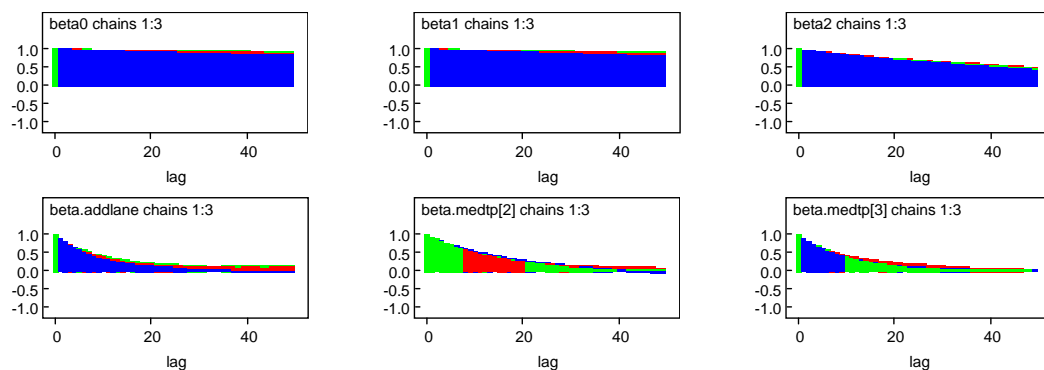


Figura 53| Gráficos das autocorrelações dos seis parâmetros β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos..

Os gráficos dos valores das estatísticas de Gelman-Rubin para as estimativas dos parâmetros β_j estão representados na Figura 54. O algoritmo converge uma vez que R (ver equação (3.12)) tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam à medida que o número de iterações aumenta.

Uma vez que os factores de redução de escala de Gelman-Rubin (ver Figura 54) mostram que a convergência foi atingida, não houve necessidade de repetir a simulação do algoritmo com um maior número de iterações.

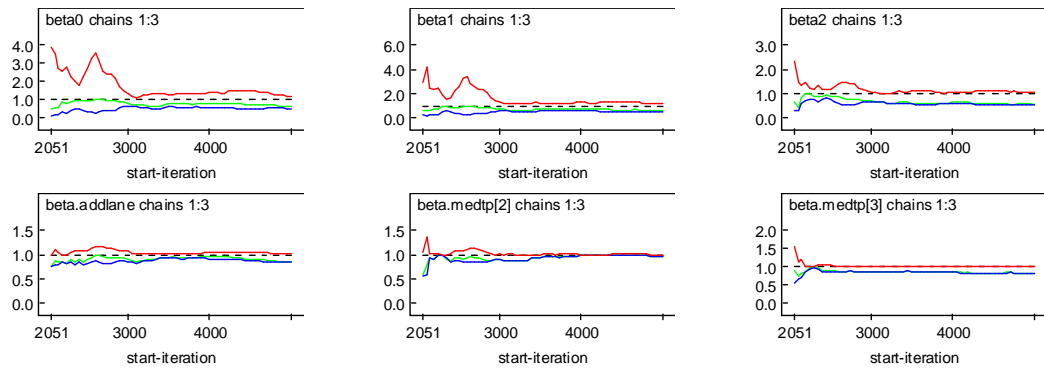


Figura 54| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações do modelo, estimado *a posteriori*, para os números esperados de acidentes, num período de cinco anos, e de acordo com os níveis das variáveis categóricas *Additional_Lane* e *Median_Type*, encontram-se no Quadro 23. Pode-se observar que o número esperado de acidentes num trecho de estradas é menor quando existe uma via adicional (*Additional_Lane*="Yes") e quando o separador é do tipo "DNJ;NJ". O que, ao contrário do que foi obtido quando o modelo de Poisson foi usado, está de acordo com os resultados apresentados em Azeredo Lopes e Cardoso (2009) onde trechos com separadores do tipo "DNJ;NJ" tinham menor número esperado de acidentes em relação aos trechos com outro tipo de separadores (ver secção 5.1 da referida publicação).

Quadro 23| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo os trechos com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\nu}_i = 3.476 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$
'GM'	$\hat{\nu}_i = 4.473 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$
'Vala'	$\hat{\nu}_i = 4.643 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$
<i>Additional_Lane</i> ='Yes'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\nu}_i = 2.379 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$
'GM'	$\hat{\nu}_i = 3.061 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$
'Vala'	$\hat{\nu}_i = 3.177 \times 10^{-3} \times AADT_i^{0.792} \times Length_i^{0.851}$

O Quadro 24 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% das estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pelas equações representadas no Quadro 23.

Quadro 24| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 23.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-5.662	0.504	3.323×10^{-2}	-6.646	-4.702
β_1 (log(AADT))	0.792	0.045	2.856×10^{-3}	0.705	0.877
β_2 (log(Length))	0.851	0.083	4.589×10^{-3}	0.681	1.018
β_3 (Additional_Lane='Yes')	-0.379	0.180	6.365×10^{-2}	-0.726	-8.658×10^{-3}
β_4 (Median_Type='GM')	0.252	0.125	4.662×10^{-3}	0.016	0.505
β_5 (Median_Type='Vala')	0.290	0.206	6.807×10^{-3}	0.016	0.700

Os intervalos de credibilidade de 95% para as estimativas obtidas *a posteriori* dos parâmetros β_j não incluem o valor zero, o que indica que todas as covariáveis consideradas têm um efeito significativo no número de acidentes, como se pode observar também nos gráficos da densidade *a posteriori* dos parâmetros β_j da Figura 55. Pela leitura dos valores do Quadro 24 também se pode depreender que uma via adicional reduz o número de acidentes em cerca de 32%. Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter aproximadamente mais 29% de acidentes do que um trecho (com o mesmo valor de AADT e Length, assim como Additional_Lane) com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 34% mais acidentes do que um trecho com os mesmos comprimento e valores de AADT e Additional_Lane, mas com separador do tipo 'DNJ;NJ'.

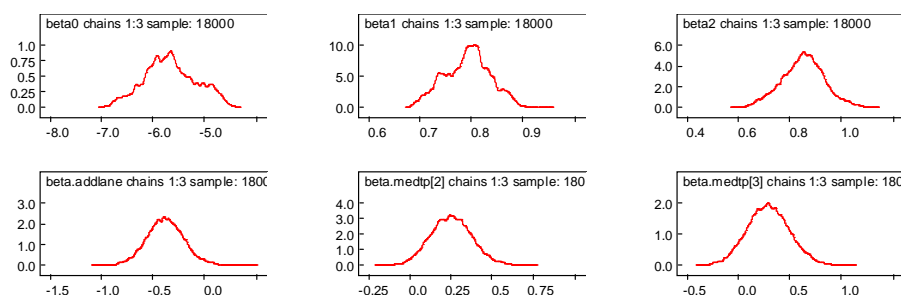


Figura 55| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson-Gama ajustado a Nac (obtido do conjunto de dados excluindo os trechos com valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de acidentes para os diversos tipos de separador e casos de existência ou não de vias adicionais para uma representação média de AADT e do comprimento foi calculado e é dado pelos valores indicados no Quadro 25.

Quadro 25| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 23.

Additional_Lane='No'	média (desvio padrão)	Additional_Lane='Yes'	média (desvio padrão)
Median_Type		Median_Type	
'DNJ;NJ'	23.97 (2.169)	'DNJ;NJ'	16.66 (3.369)
'GM'	30.83 (2.655)	'GM'	21.30 (3.538)
'Vala'	32.42 (6.051)	'Vala'	22.57 (6.054)

Para um trecho típico de auto-estrada sem via adicional e com um separador do tipo 'DNJ;NJ', são esperados, para um período de cinco anos, 23.97 acidentes, enquanto que para um trecho com separadores do tipo 'GM' e 'Vala' são esperados 30.83 e 32.42 acidentes, respectivamente. Para trechos com uma via adicional são esperados 16.66, 21.30 e 22.57 acidentes quando esses trechos possuírem separadores do tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente. Contudo, note-se os elevados valores dos correspondentes desvios padrões.

5.1.2.1 Validação do Modelo

A Figura 56 mostra vinte histogramas, estando no canto superior esquerdo o histograma (a cinzento) dos dados observados. Os restantes são histogramas de dados replicados da distribuição *a posteriori* preditiva obtida quando se assume um modelo Poisson-Gama. A observação desta figura mostra que, em geral, os vários histogramas das réplicas são bastante semelhantes ao histograma dos dados observados, o que indica que este modelo está bem ajustado aos dados.

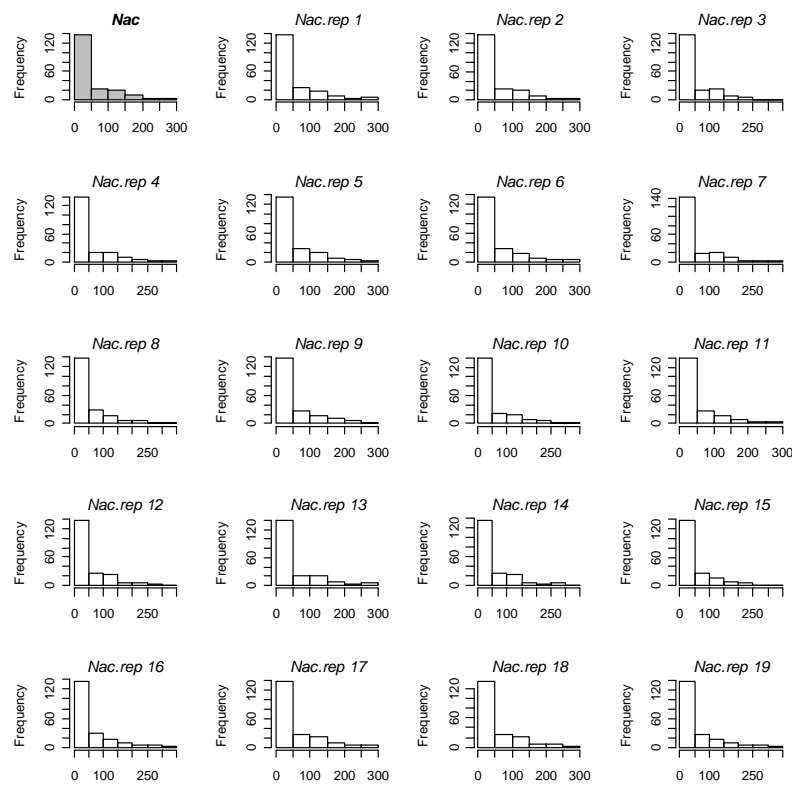


Figura 56] Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas ("*Nac.rep*") dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson-Gama ajustado ao conjunto de dados excluindo os trechos com valores omissos.

Cada histograma da Figura 57 representa o histograma dos dados replicados de cada uma das medidas de discrepâncias consideradas onde as rectas verticais indicam os valores das mesmas discrepâncias obtidos pelos dados observados.

As probabilidades das discrepâncias obtidas pelos dados replicados serem superiores às discrepâncias obtidas pelos dados observados são aproximadamente iguais a 0.50, o que indica que, pelo menos em relação às medidas de discrepância consideradas, os dados são comparáveis.

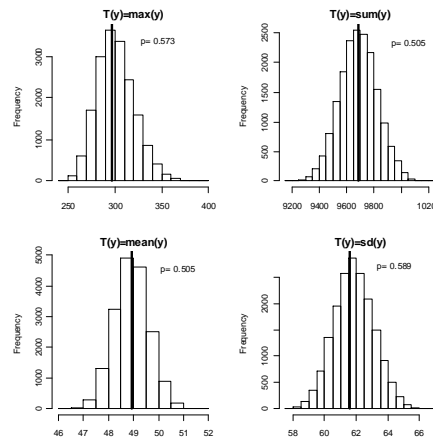


Figura 57| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O histograma representado na Figura 58 refere-se à aproximação da distribuição preditiva *a posteriori* do teste T indicado pela razão entre a variância e o valor médio dos valores replicados; a recta vertical indica o valor da mesma razão nos dados observados (isto é, 77.44).

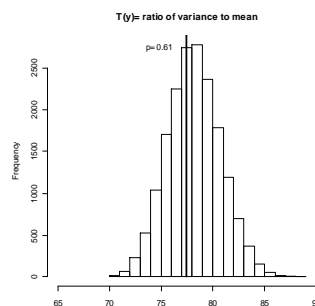


Figura 58| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 18000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida dos dados replicados do modelo Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Em média, a probabilidade de que a razão entre a variância e a média dos valores replicados pelo modelo Poisson-Gama seja superior à mesma razão obtida pelos dados é igual a 0.61 o que indica que os dados observados e os dados replicados pelo modelo Poisson-Gama são comparáveis em termos da referida razão.

A média das desviâncias amostrais, ver equação (3.15) resulta num valor igual a $\bar{D} = 197.7$ o que, comparado com os graus de liberdade, 194, indica um bom ajuste do modelo Poisson-Gama e uma boa modelação da sobredispersão e, conseqüentemente, uma melhoria em relação ao modelo de Poisson. A estimativa do parâmetro de dispersão é igual a 0.523. Este modelo produz um valor de DIC igual a 1320.06, consideravelmente menor que o valor obtido (3520.26) quando assumindo um modelo de Poisson para os números de acidentes.

Todas as discrepâncias investigadas sugerem que o modelo Poisson-Gama gera resultados preditivos semelhantes aos dados observados, ou melhor, os dados observados são valores plausíveis das observações preditivas geradas pelo modelo Poisson-Gama.

5.1.3 Modelo de Poisson Log-Normal

Ao ajustar o modelo completo verificou-se que somente as variáveis $\log(AADT)$, $\log(Length)$, $Additional_Lane$ e $Median_Type$ possuíam estimativas dos coeficientes β_j cujos intervalos de credibilidade de 95% não incluíam o valor zero. O mesmo modelo foi ajustado incluindo estas quatro variáveis. Seguidamente, e com o intuito de verificar se a introdução de alguma variável rejeitada inicialmente, assim como a remoção de $Additional_Lane$ ou de $Median_Type$ melhoravam o modelo, procedeu-se ao ajuste de várias combinações de modelos. Como medida de ajuste foi usado o valor de DIC para cada conjunto de variáveis considerado. O grupo de variáveis que produziu o valor DIC mais baixo foi escolhido como o modelo final e inclui, além da intercepção, $\log(AADT)$, $\log(Length)$, $Additional_Lane$ e $Median_Type$. A equação genérica é dada por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Median_Type_i \quad (5.3)$$

O algoritmo de MCMC convergiu após 40000 iterações com 10000 iterações *burn-in* e com um *thinning rate* igual a 5. Os resultados foram obtidos de uma amostra com dimensão 18000.

A observação dos *trace plots* das estimativas dos coeficientes β_j na Figura 59 mostra que estes valores tendem a “manter-se” na mesma “zona”, não se verificando nenhuma tendência especial. É também óbvio constatar que os parâmetros da intercepção e de $\log(AADT)$ não convergem com tanta precisão como os restantes.

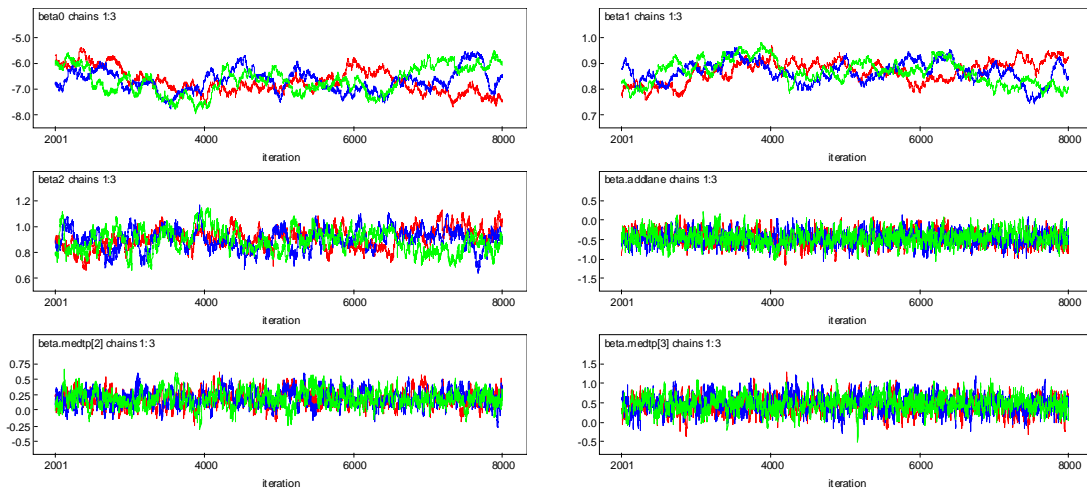


Figura 59| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos das autocorrelações dos valores dos parâmetros acima referidos estão reproduzidos na Figura 60. As estimativas dos parâmetros β_0 , β_1 e β_2 convergem mais lentamente que as restantes.

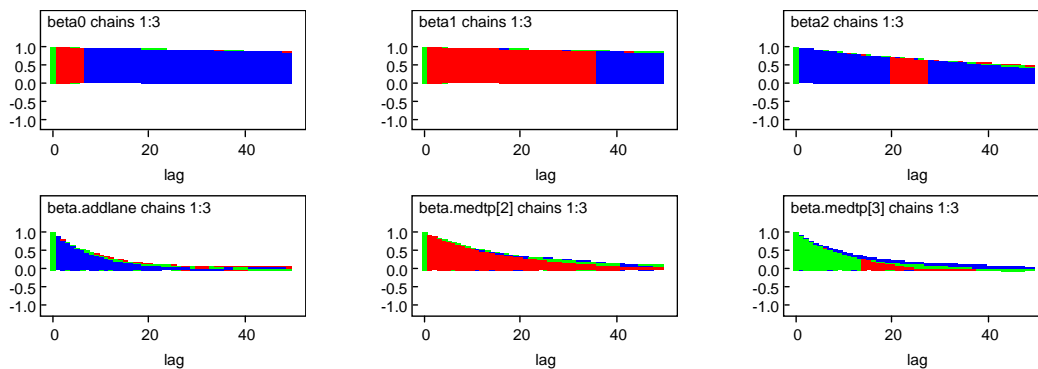


Figura 60| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 61. O algoritmo converge uma vez que R tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam à medida que o número de iterações aumenta.

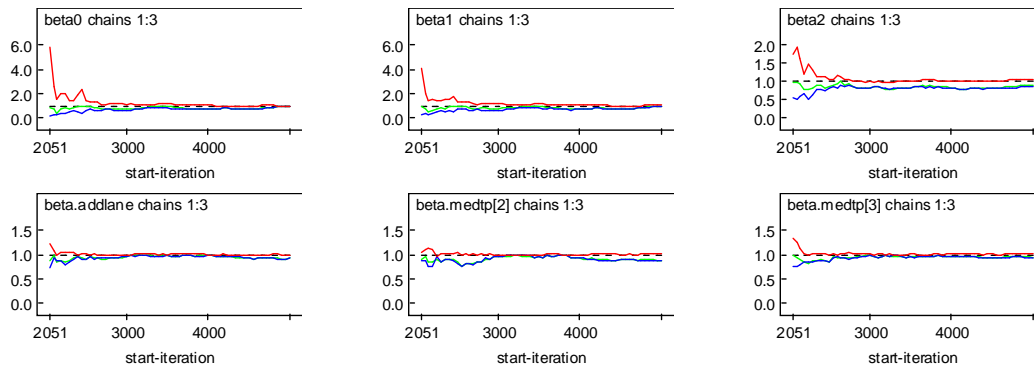


Figura 61| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações para os números esperados de acidentes num período de cinco anos obtidos pelo modelo Poisson Log-Normal encontram-se no Quadro 26.

Quadro 26| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo os trechos com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane='No'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 1.255 \times 10^{-3} \times AADT_i^{0.868} \times Length_i^{0.901}$
'GM'	$\hat{\lambda}_i = 1.519 \times 10^{-3} \times AADT_i^{0.868} \times Length_i^{0.901}$
'Vala'	$\hat{\lambda}_i = 1.974 \times 10^{-3} \times AADT_i^{0.868} \times Length_i^{0.901}$
<i>Additional_Lane='Yes'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 7.900 \times 10^{-4} \times AADT_i^{0.868} \times Length_i^{0.901}$
'GM'	$\hat{\lambda}_i = 9.567 \times 10^{-4} \times AADT_i^{0.868} \times Length_i^{0.901}$
'Vala'	$\hat{\lambda}_i = 1.243 \times 10^{-3} \times AADT_i^{0.868} \times Length_i^{0.901}$

Da observação das equações do Quadro 26 depreende-se que o número esperado de acidentes é menor em trechos de auto-estradas com uma via adicional, qualquer que seja o tipo de separador empregue. Os tipos de separadores por ordem crescente de números esperados de acidentes são, tal como para o modelo Poisson-Gama, 'DNJ;NJ', 'GM' e 'Vala'. Note-se também que o número esperado de acidentes em trechos com uma via adicional e separador do tipo 'Vala' é muito semelhante ao de trechos sem via adicional e com separador do tipo 'DNJ;NJ', o que evidencia a diferença encontrada em trechos com ou sem vias adicionais.

O Quadro 27 contém as estimativas pontuais, assim como intervalos de credibilidade de 95% para as estimativas dos parâmetros das equações no Quadro 26. Depreende-se que os intervalos de credibilidade de 95% correspondentes aos parâmetros β_j não incluem o valor zero, à excepção da

variável *Median_Type* referente a *Median_Type*='GM', no entanto, pela observação dos gráficos das densidades *a posteriori* dos parâmetros β_j na Figura 62, pode constatar-se que a média *a posteriori* para a estimativa de *Median_Type*='GM' (na segunda linha ao meio) se deslocou consideravelmente de zero (o valor médio *a priori*) mas que o intervalo de credibilidade de 95% ainda o inclui. Este facto faz com que haja cuidado ao considerar a variável *Median_Type* como não significativa.

Quadro 27| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 26

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-6.681	0.466	2.985×10^{-2}	-7.488	-5.770
β_1 (log(AADT))	0.868	0.043	2.712×10^{-3}	0.783	0.943
β_2 (log(Length))	0.901	0.081	4.491×10^{-3}	0.739	1.057
β_3 (Additional_Lane='Yes')	-0.463	0.178	5.610×10^{-3}	-0.808	-0.113
β_4 (Median_Type='GM')	0.191	0.126	4.913×10^{-3}	-0.057	0.442
β_5 (Median_Type='Vala')	0.453	0.209	6.911×10^{-3}	0.040	0.861

Os valores das estimativas indicadas no Quadro 27 indicam que, para trechos com os mesmos valores das covariáveis *AADT*, *Length* e *Median_Type*, uma via adicional reduz o número esperado de acidentes em cerca de 37%. Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter aproximadamente mais 21% de acidentes do que um trecho nas mesmas condições mas com separador do tipo 'DNJ;NJ'. Por sua vez, um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 57% mais acidentes do que um trecho em iguais condições mas com separador do tipo 'DNJ;NJ'.

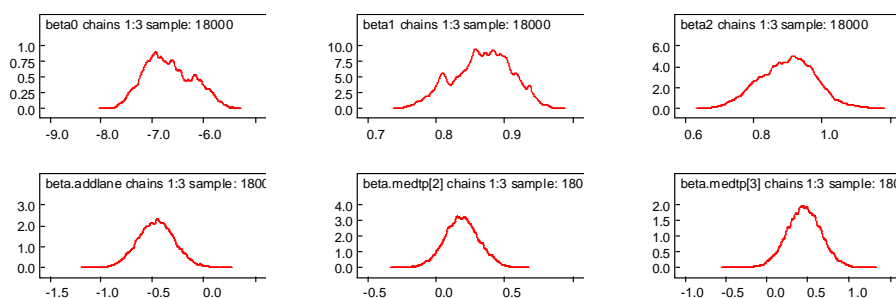


Figura 62| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson Log-Normal ajustado a *Nac* (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.

O número esperado de acidentes para valores médios de *log(AADT)* e *log(Length)* e para os diversos tipos de separador e existência, ou não, de vias adicionais, são dados pelos valores indicados no Quadro 28.

Quadro 28| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 26.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	19.03 (1.841)	'DNJ;NJ'	12.15 (2.336)
'GM'	23.02 (1.942)	'GM'	14.65 (2.468)
'Vala'	30.30 (5.565)	'Vala'	19.38 (4.972)

Num período de cinco anos, e para um trecho típico de auto-estrada sem via adicional e com um separador do tipo 'DNJ;NJ' são esperados 19.03 acidentes; para trechos com separadores do tipo 'GM' e 'Vala' são esperados 23.02 e 30.30 acidentes, respectivamente. Para trechos com uma via adicional são esperados 12.15, 14.65 e 19.38 acidentes para trechos com separadores do tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente.

5.1.3.1 Validação do Modelo

A Figura 63 contém o histograma do número de acidentes observado (a cinzento no canto superior esquerdo) e os dezanove histogramas dos números de acidentes replicados da distribuição *a posteriori* preditiva obtida quando se assume o modelo Poisson Log-Normal. A observação desta figura mostra que, em geral, os vários histogramas dos dados replicados são bastante semelhantes ao histograma dos dados observados, o que indica que o modelo Poisson Log-Normal está bem ajustado.

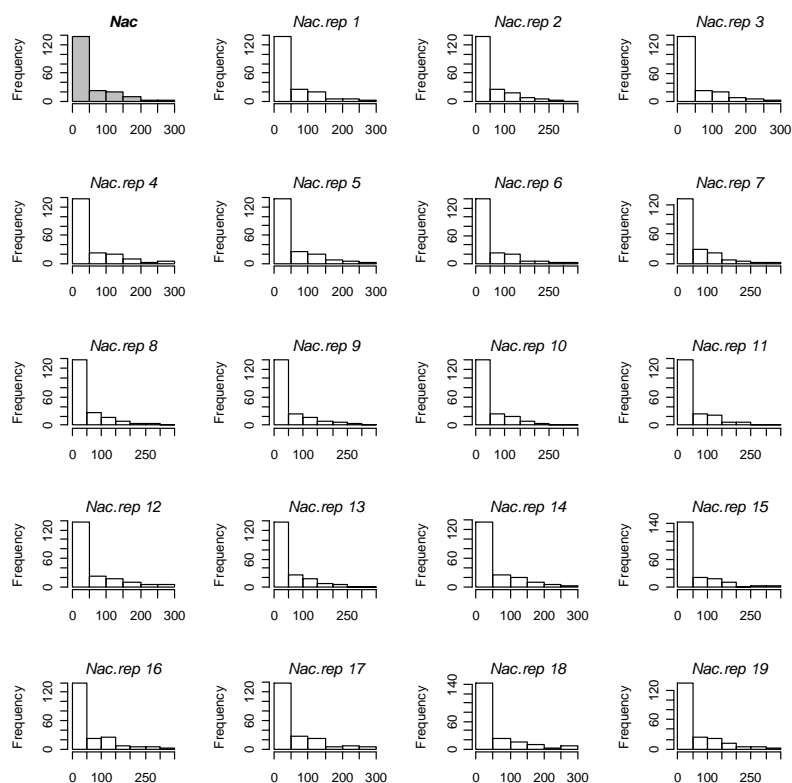


Figura 63] Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados excluindo os trechos com valores omissos.

Os histogramas representados na Figura 64 mostram a maior observação, soma, valor médio e desvio padrão das observações em cada de 198 replicações hipotéticas; todas estas medidas são representativas das respectivas medidas obtidas pelos dados observados (linhas verticais nos histogramas). Claramente, o modelo Poisson Log-Normal capta a variação dos dados observados.

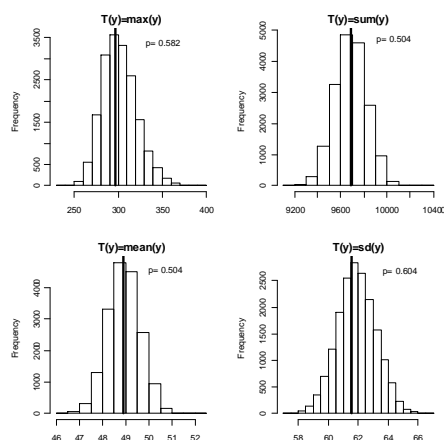


Figura 64] Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O histograma na Figura 65 refere-se à aproximação da distribuição preditiva *a posteriori* do teste $T = \text{variância}/\text{média}$ dos valores replicados; a recta vertical indica o valor da mesma razão T obtida nos dados observados.

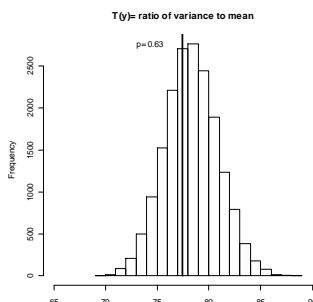


Figura 65] Razão entre a variância e a média de Nac (recta vertical) comparada com 18000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida dos dados replicados do modelo Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Devido ao facto do histograma envolver a recta vertical, pode concluir-se que o modelo Poisson Log-Normal capta a razão entre a variância e o valor médio dos dados observados. Em média, a probabilidade de que as razões T , acima referida e referente aos dados replicados pelo modelo Poisson Log-Normal, sejam superiores à mesma razão obtida nos dados observados, isto é, no número de acidentes observados, é igual a 0.63, o que indica que os dados observados e os replicados são comparáveis em termos da referida razão.

A média das desviâncias amostrais (equação (3.15)) resulta num valor $\bar{D} = 213.5$ (maior do que o valor obtido aquando do ajuste do modelo Poisson-Gama), indicando um razoável ajuste do modelo Poisson Log-Normal ($n-p=194$). A estimativa do parâmetro de dispersão é igual a 0.556.

Este modelo produz um valor de DIC igual a 1346.59, valor menor do que o valor obtido quando assumindo um modelo de Poisson (3520.26), mas ligeiramente maior do que o valor de DIC obtido com um modelo Poisson-Gama (1320.06), o que indica o modelo de Poisson-Gama como o mais bem ajustado a estes dados.

Todas as discrepâncias investigadas sugerem que o modelo Poisson Log-Normal gera resultados preditivos semelhantes aos dados observados. No entanto, os valores de DIC, as desviâncias amostrais e os p -values das medidas de discrepâncias sugerem o modelo de Poisson-Gama como sendo o mais bem ajustado aos dados relativos aos números de acidentes.

5.2 Modelação do Número de Mortos e Feridos Graves

Os gráficos descritivos referentes aos dados dos números de mortos e feridos graves ($NKSI$) podem ser examinados na secção 5.2 de Azeredo Lopes e Cardoso (2009).

5.2.1 Modelo de Poisson

Seguindo o método de selecção de variáveis, sugerido por Ntzoufras (2009), e baseado no valor de DIC chegou-se a um modelo final que inclui as covariáveis $\log(AADT)$, $\log(\text{Length})$, Additional_Lane e Median_Type . A equação genérica é a seguinte:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i + \beta_4 \text{Median_Type}_i \quad (5.4)$$

O algoritmo de MCMC converge com 10000 simulações, das quais 5000 são rejeitadas (*burn-in*) e com um *thinning rate* igual a 5. Os resultados foram baseados numa amostra de dimensão 3000.

Os gráficos da Figura 66 mostram que se obteve convergência das estimativas dos parâmetros β_j nas equações (3.22). Note-se, porém, que os parâmetros das estimativas de β_0 e β_1 (os dois gráficos no topo da Figura 66) não convergem tão bem como os parâmetros seguintes pois há uma maior oscilação entre as iterações.

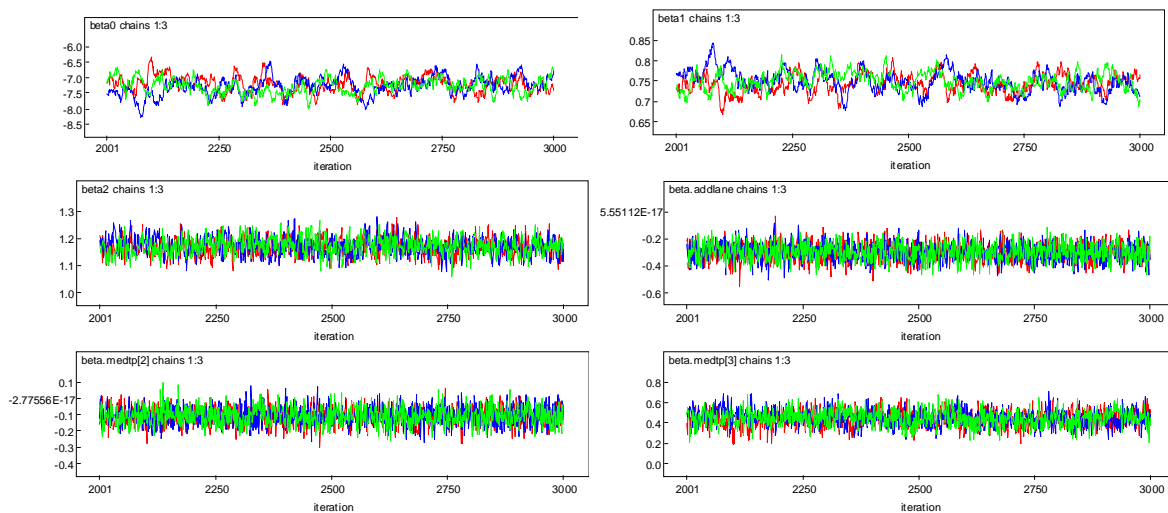


Figura 66] *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos das autocorrelações na Figura 67 corroboram o facto de as estimativas de β_0 e β_1 (dois gráficos a contar da esquerda na primeira linha da Figura 67) não convergirem tão rapidamente como as restantes.

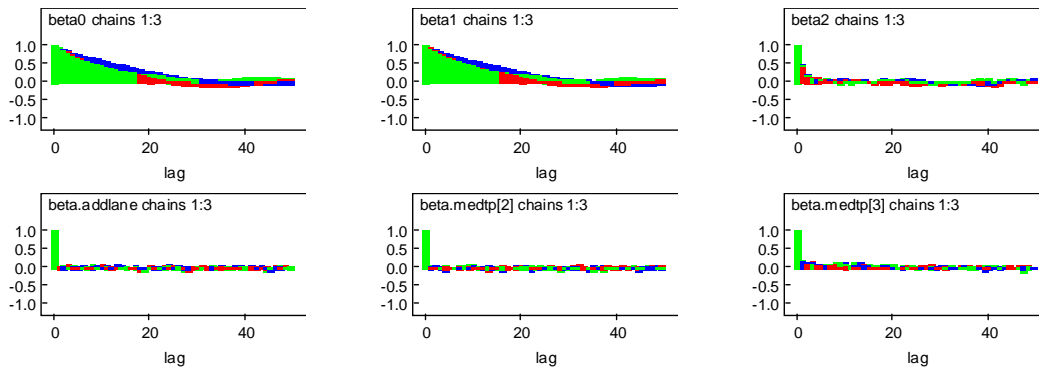


Figura 67| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

As mesmas conclusões são obtidas pela observação das estatísticas de Gelman-Rubin cujos gráficos estão exibidos na Figura 68.

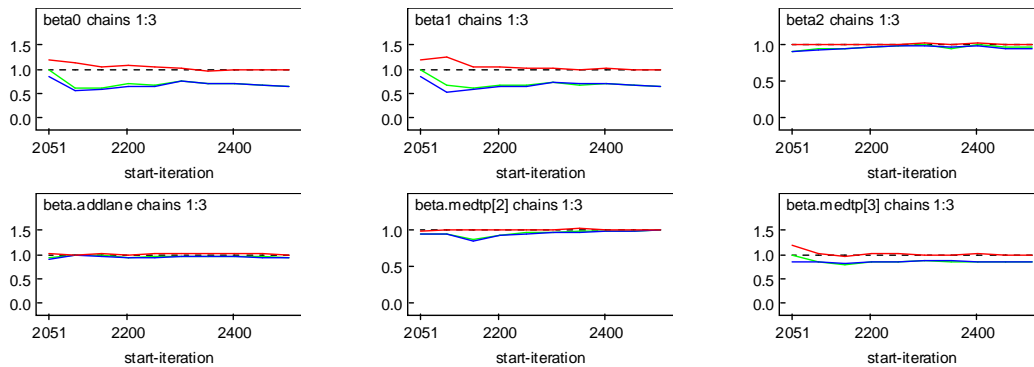


Figura 68| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações dos números esperados de mortos e feridos graves, para um período de cinco anos, obtidas após ajuste de um modelo de Poisson, são apresentadas no Quadro 29.

O número esperado de mortos e feridos graves é mais elevado em trechos sem via adicional. O tipo de separador relacionado com o maior número esperado dos eventos em questão é o tipo 'Vala'. Um ajuste do modelo de Poisson a estes dados resulta num número esperado de mortos e feridos graves ligeiramente maior em trechos com separador do tipo 'DNJ;NJ' do que em trechos com separador 'GM'.

Quadro 29| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados excluindo os trechos com valores omissos.

Números Esperados de Mortos e Feridos Graves	
<i>Additional_Lane='No'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 7.137 \times 10^{-4} \times AADT_i^{0.747} \times Length_i^{1.171}$
'GM'	$\hat{\lambda}_i = 6.405 \times 10^{-4} \times AADT_i^{0.747} \times Length_i^{1.171}$
'Vala'	$\hat{\lambda}_i = 1.113 \times 10^{-3} \times AADT_i^{0.747} \times Length_i^{1.171}$
<i>Additional_Lane='Yes'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 5.290 \times 10^{-4} \times AADT_i^{0.747} \times Length_i^{1.171}$
'GM'	$\hat{\lambda}_i = 4.748 \times 10^{-4} \times AADT_i^{0.747} \times Length_i^{1.171}$
'Vala'	$\hat{\lambda}_i = 8.253 \times 10^{-4} \times AADT_i^{0.747} \times Length_i^{1.171}$

O Quadro 30 mostra as estimativas pontuais do modelo de Poisson ajustado aos dados, tendo *NKSI* como variável resposta.

Quadro 30| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 29.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.245	0.271	2.074×10^{-2}	-7.783	-6.738
β_1 (log(AADT))	0.747	0.024	1.818×10^{-3}	0.702	0.795
β_2 (log(Length))	1.171	0.032	1.115×10^{-3}	1.110	1.237
β_3 (Additional_Lane='Yes')	-0.299	0.065	1.306×10^{-3}	-0.430	-0.172
β_4 (Median_Type='GM')	-0.108	0.055	1.048×10^{-3}	-0.216	2.021×10^{-3}
β_5 (Median_Type='Vala')	0.445	0.079	2.631×10^{-3}	0.286	0.604

Pode-se observar que, ao contrário dos intervalos para as restantes estimativas dos outros parâmetros, o intervalo de credibilidade de 95% para a estimativa do parâmetro β_4 inclui o valor zero. No entanto, pela observação do gráfico da densidade *a posteriori* das estimativas do parâmetro β_4 (representada no gráfico central da segunda fila na Figura 69) pode-se constatar que o valor médio da distribuição *a posteriori* do mesmo parâmetro se “desviou” consideravelmente de zero, o que significa que se pode considerar a variável *Median_Type* como relevante no modelo.

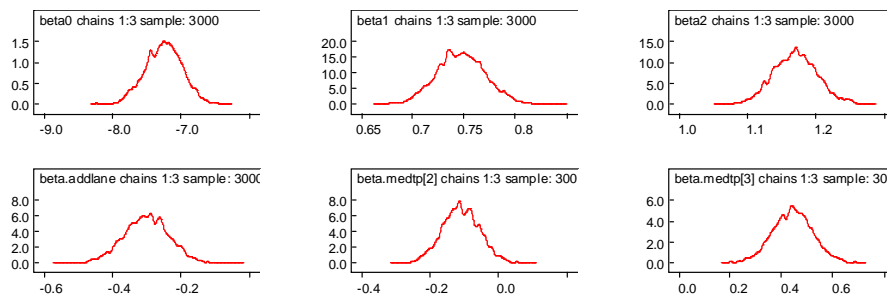


Figura 69| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NKSI* (obtido do conjunto de dados excluindo os trechos com valores omissos) resultantes da simulação de três cadeias de Markov.

A leitura das estimativas médias dos parâmetros no Quadro 30 mostra que com uma via adicional o número de mortos e feridos graves é menor em cerca de 25%. Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter aproximadamente menos 10% de ocorrência de mortos e feridos graves do que um trecho em iguais condições, mas com separador do tipo 'DNJ;NJ'. Por outro lado, em iguais condições, um trecho com separador do tipo 'Vala', é *a posteriori* esperado ter mais 56% de ocorrências de mortos e feridos graves do que um trecho com separador do tipo 'DNJ;NJ'.

Quadro 31| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 29.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	5.637 (0.285)	'DNJ;NJ'	4.213 (0.336)
'GM'	5.091 (0.241)	'GM'	3.778 (0.256)
'Vala'	8.856 (0.552)	'Vala'	6.577 (0.579)

O Quadro 31 dá os resultados dos números esperados de mortos e feridos graves, para um período de cinco anos, para uma representação de valores médios de *AADT* e *Length*. Pode observar-se que uma via adicional reduz consideravelmente esse número esperado, sendo os trechos com separadores de tipo 'Vala' onde são esperados os mais elevados números destes eventos.

5.2.1.1 Validação do Modelo

A Figura 70 mostra o histograma dos dados observados conjuntamente com dezanove histogramas de dados replicados da distribuição preditiva *a posteriori* obtida assumindo o modelo de Poisson. Em geral, os vários histogramas dos dados replicados não se assemelham ao histograma dos dados observados, em particular os histogramas correspondentes a *NKSI.rep 26, 30, 32 e 33*. O que sugere que o modelo de Poisson não esteja bem ajustado aos dados.

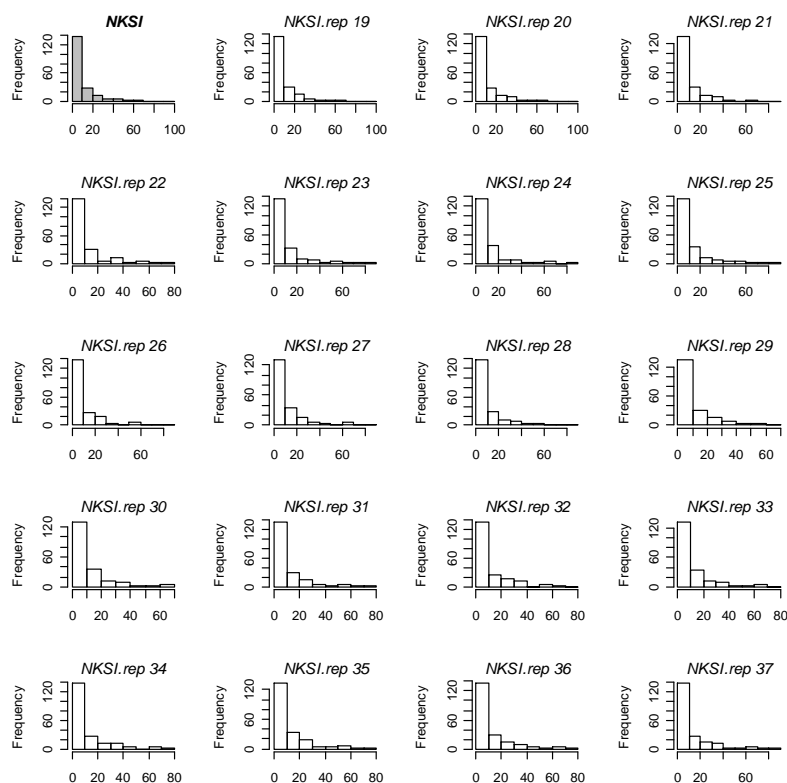


Figura 70| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados excluindo os trechos com valores omissos.

Quando se investigam possíveis discrepâncias, nota-se que este modelo não é congruente com os dados observados no que respeita ao valor máximo e ao desvio padrão, tal como pode ser observado no histograma no canto superior esquerdo e no histograma no canto inferior direito, respectivamente, na Figura 71.

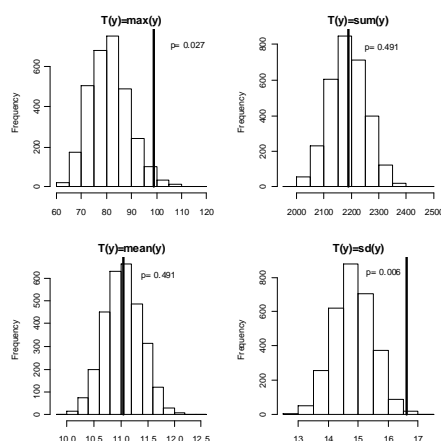


Figura 71| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo de Poisson também não capta a sobredispersão existente nos dados observados pois a razão entre a variância e o valor médio observados é consideravelmente maior do que a mesma razão obtida por dados replicados da distribuição preditiva *a posteriori*, como pode ser observado na Figura 72. A validação preditiva mostra uma probabilidade média $p_B=0.001$ de que a razão entre a variância e o valor médio dos dados replicados exceda a mesma razão dos dados observados.

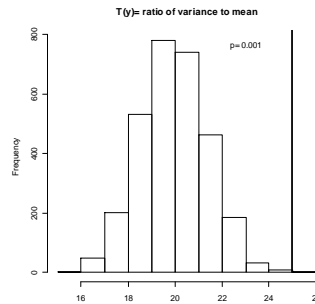


Figura 72| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais (ver equação (3.15)) é igual a $\bar{D} = 773.8$, valor consideravelmente maior do que o número de graus de liberdade, 194, o que indica um ajuste inadequado do modelo de Poisson aos dados referentes ao número de mortos e feridos graves.

Este modelo produz um valor de DIC igual a 1500.36.

5.2.2 Modelo de Poisson-Gama

Foi usado um processo de selecção de variáveis que se iniciou pelo modelo completo e no qual, passo a passo, se foram removendo, ou inserindo, variáveis de maneira a que o modelo final incluisse somente as variáveis relevantes; e de modo a obter o menor valor de DIC. As covariáveis encontradas quando se ajustou um modelo Poisson-Gama a *NKSI* com distribuições *a priori* dadas pelas equações (3.27) são $\log(AADT)$, $\log(Length)$, $Additional_Lane$ e $Median_Type$. A equação genérica é:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Median_Type_i \quad (5.5)$$

Foram calculadas 40000 iterações, das quais 10000 faziam parte do *burn-in*, que, com um *thinning rate* igual a 5 resultaram numa amostra de 18000 observações de onde os resultados aqui apresentados são baseados.

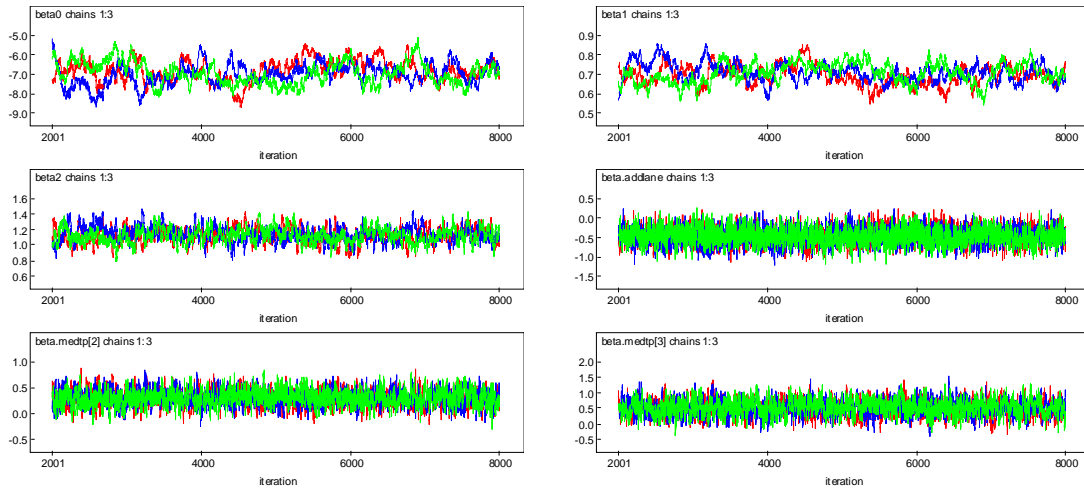


Figura 73| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

Para avaliar a convergência do algoritmo MCMC recorre-se à visualização de gráficos das diversas quantidades registadas. Pela observação da Figura 73 pode constatar-se que as três cadeias de Markov usadas convergem simultaneamente para os mesmos valores de cada parâmetro β_j considerado.

As autocorrelações na Figura 74 mostram que os parâmetros β_1 , β_2 e β_3 (na primeira linha da figura) convergem mais lentamente que os restantes parâmetros.

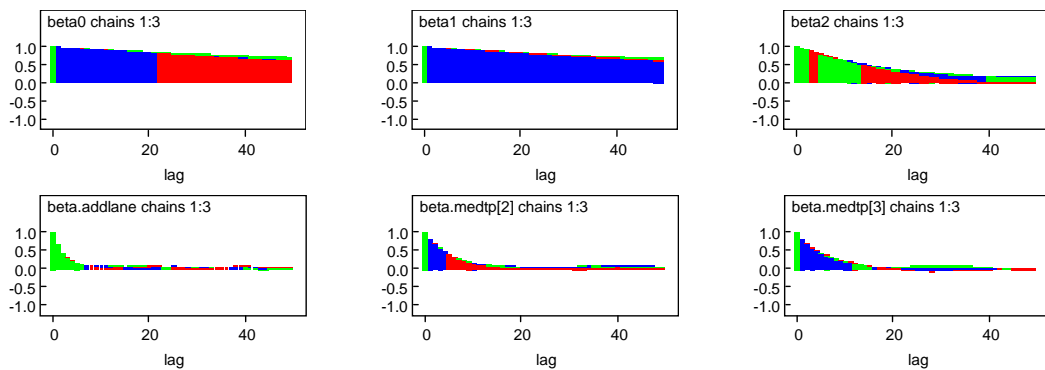


Figura 74| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

As estatísticas de Gelman-Rubin cujos valores estão exibidos na Figura 75 mostram que o algoritmo converge em todos os parâmetros.

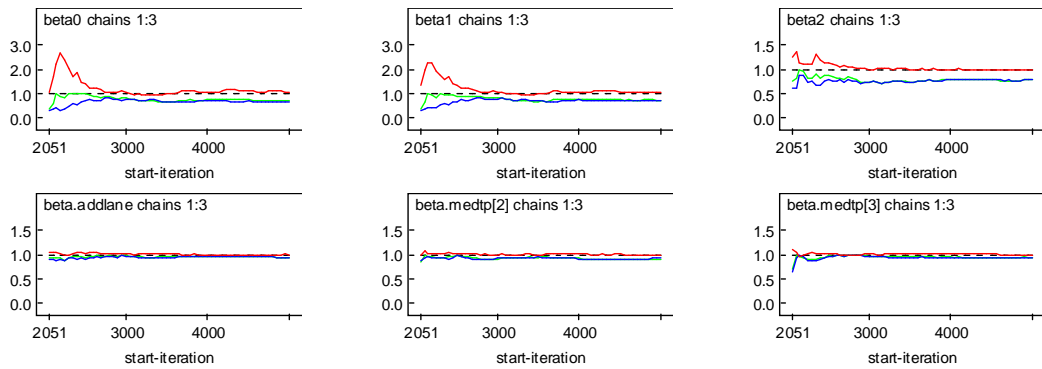


Figura 75| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_i do modelo de Poisson-Gama ajustado a NKS/ obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações que representam o número esperado de mortos e feridos graves, para um período de cinco anos, encontram-se no Quadro 32. O número esperado de mortos e feridos graves é mais elevado quando não existe via adicional. Os trechos de auto-estrada por ordem crescente de números esperados de mortos e feridos graves têm separadores de tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente.

Quadro 32| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados excluindo os trechos com valores omissos.

Números Esperados de Mortos e Feridos Graves	
<i>Additional_Lane='No'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 9.938 \times 10^{-4} \times AADT_i^{0.703} \times Length_i^{1.130}$
'GM'	$\hat{\lambda}_i = 1.340 \times 10^{-3} \times AADT_i^{0.703} \times Length_i^{1.130}$
'Vala'	$\hat{\lambda}_i = 1.668 \times 10^{-3} \times AADT_i^{0.703} \times Length_i^{1.130}$
<i>Additional_Lane='Yes'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 6.206 \times 10^{-4} \times AADT_i^{0.703} \times Length_i^{1.130}$
'GM'	$\hat{\lambda}_i = 8.370 \times 10^{-4} \times AADT_i^{0.703} \times Length_i^{1.130}$
'Vala'	$\hat{\lambda}_i = 1.042 \times 10^{-3} \times AADT_i^{0.703} \times Length_i^{1.130}$

O Quadro 33 contém as estimativas pontuais para o modelo Poisson-Gama ajustado aos dados.

Quadro 33| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 32.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-6.914	0.563	3.369×10^{-2}	-8.019	-5.802
β_1 (log(AADT))	0.703	0.050	2.998×10^{-3}	0.605	0.802
β_2 (log(Length))	1.130	0.096	4.055×10^{-3}	0.941	1.321
β_3 (Additional_Lane='Yes')	-0.471	0.201	4.244×10^{-3}	-0.862	-6.645×10^{-2}
β_4 (Median_Type='GM')	0.299	0.150	3.836×10^{-3}	7.514×10^{-3}	0.595
β_5 (Median_Type='Vala')	0.518	0.246	6.010×10^{-3}	4.292×10^{-2}	1.003

Pode observar-se que os intervalos de credibilidade de 95% não contêm o valor zero para nenhum dos parâmetros o que significa que as correspondentes variáveis são significantes. A Figura 76 mostra as densidades *a posteriori* dos parâmetros β_j . Pode observar-se que as médias *a posteriori* se afastaram bastante do valor zero (o valor médio das distribuições *a priori*).

Os coeficientes do modelo mostram que uma via adicional tem menor número de mortos e feridos graves em cerca de 38%. Um trecho com separadores dos tipos 'GM' e 'Vala' são *a posteriori* esperados ter mais 35% e 68%, respectivamente, de ocorrências de mortos e feridos graves do que trechos, em iguais circunstâncias, mas com separador do tipo 'DNJ;NJ'.

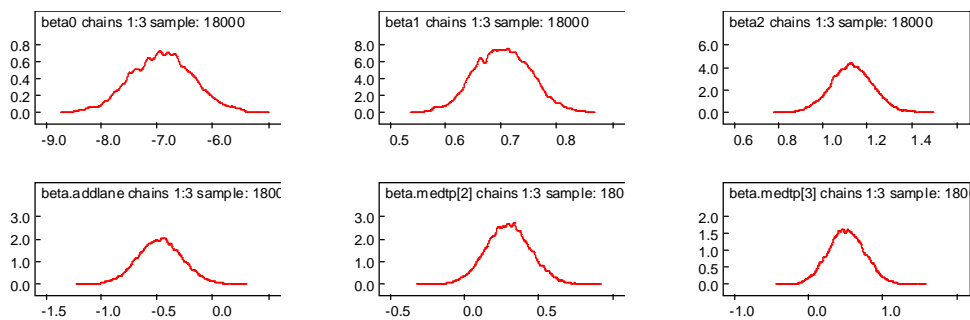


Figura 76| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson-Gama ajustado a NKS/ (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves para uma representação média de $\log(AADT)$ e $\log(Length)$ são dados no Quadro 34 para os diversos tipos de separador e existência, ou não de uma via adicional.

Quadro 34| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 32.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	4.901 (0.570)	'DNJ;NJ'	3.119 (0.719)
'GM'	6.597 (0.652)	'GM'	4.171 (0.782)
'Vala'	8.357 (1.794)	'Vala'	5.336 (1.626)

5.2.2.1 Validação do Modelo

Os histogramas dos dados replicados exibidos na Figura 77 são semelhantes ao histograma dos dados observados (no canto superior esquerdo, a cinzento) o que significa que o modelo Poisson-Gama está bem ajustado aos dados.

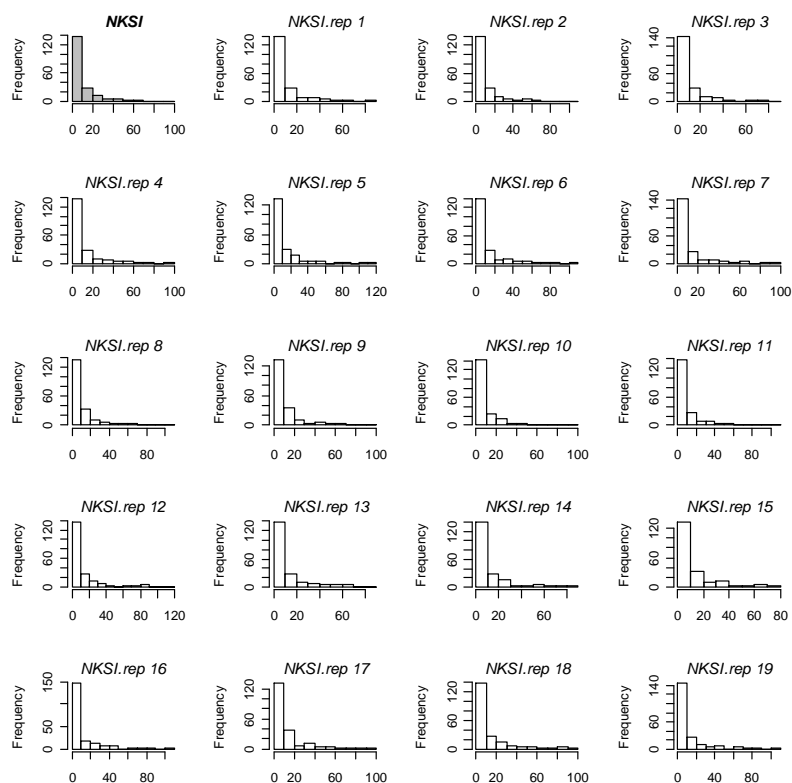


Figura 77| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados excluindo os trechos com valores omissos.

Os dados observados são representativos das observações preditivas geradas pelo modelo Poisson-Gama visto todos os histogramas obtidos pelas discrepâncias investigadas, ver Figura 78 e Figura 79, enquadrarem os valores observados das discrepâncias (representadas pelas rectas verticais em cada figura).

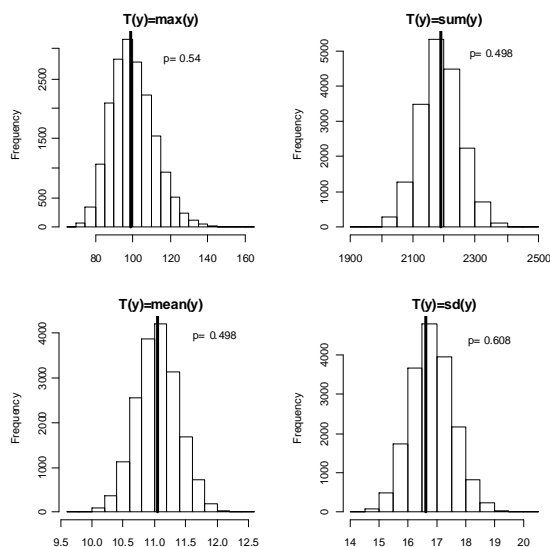


Figura 78| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Pela observação da Figura 79 também se depreende que o modelo Poisson-Gama capta a sobredispersão dos dados, ao contrário do verificado com o modelo de Poisson (comparar com a Figura 72). A probabilidade p_B obtida é igual a 0.634, valor que se encontra nos limites aceitáveis sugeridos por Congdon (2005).

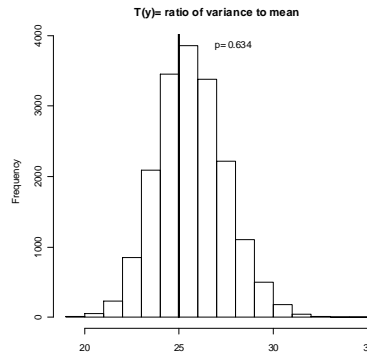


Figura 79| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais resulta num valor igual a $\bar{D} = 174.6$ ($n-p=194$) o que indica um bom ajuste do modelo Poisson-Gama. A estimativa do parâmetro de dispersão é igual a 0.617.

O valor de DIC é igual a 950.26, bastante menor do que o valor obtido pelo modelo de regressão de Poisson (1500.36), o que indica uma melhoria no ajuste do modelo Poisson-Gama.

5.2.3 Modelo de Poisson Log-Normal

Ao ajustar o modelo completo segundo as equações (3.31) verificou-se que as variáveis cujo intervalo de credibilidade de 95% para as correspondentes estimativas dos parâmetros β_j continham o valor zero eram *Lane_Width* e *Right_Shoulder_Width*, que foram conseqüentemente extraídas do modelo. Após sucessivos ajustes do modelo (acrescentando e retirando covariáveis), e sempre registando o valor de DIC, chegou-se ao modelo final que inclui as variáveis *log(AADT)*, *log(Length)*, *Additional_Lane* e *Median_Type*, modelo esse que obtém o menor valor de DIC. A equação genérica que dá o número de mortos e feridos graves é:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Median_Type_i \quad (5.6)$$

O algoritmo de MCMC converge após 40000 iterações com 10000 iterações *burn-in* e com um *thinning rate* igual a 5. Os resultados foram obtidos de uma amostra com dimensão 18000.

A Figura 80 mostra os *trace plots* correspondentes às estimativas dos parâmetros β_j no conjunto de equações em (3.31). Pode depreender-se, dado o facto de não haver grandes oscilações, que estes

parâmetros convergem para os mesmos valores, em cada uma das três cadeias simuladas. Nota-se, porém, que os parâmetros das variáveis categóricas *Additional_Lane* e *Median_Type* (gráfico à direita no meio da Figura 80 e os dois gráficos na linha inferior) convergem “mais rapidamente” do que os restantes.

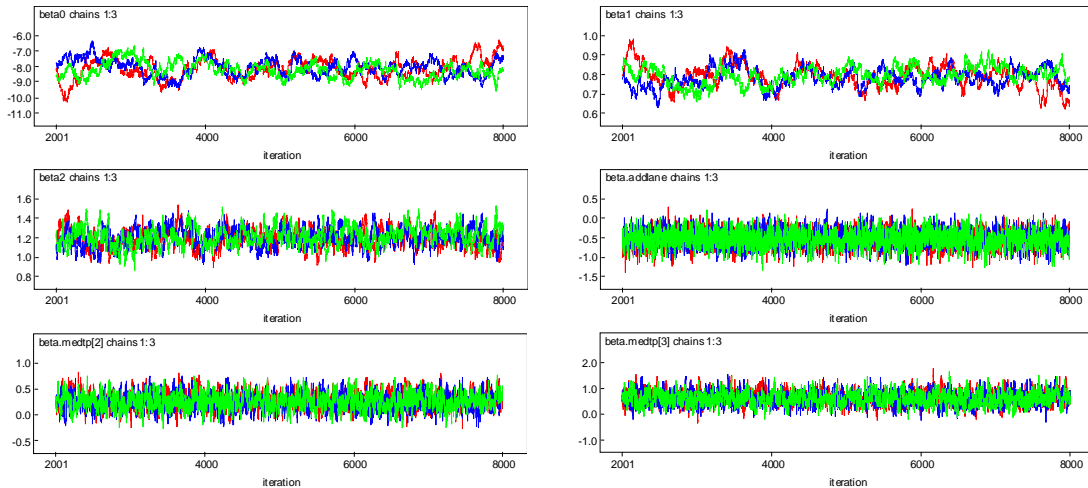


Figura 80| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a *NKS* obtido do conjunto de dados excluindo os trechos com valores omissos.

A Figura 81 contém os gráficos das autocorrelações para as estimativas de cada parâmetro β_j . Visto as autocorrelações das estimativas dos parâmetros das variáveis categóricas serem mais baixas depreende-se que as estimativas para estes parâmetros convergem mais rapidamente do que as estimativas da intercepção assim como de $\log(AADT)$ e $\log(Length)$.

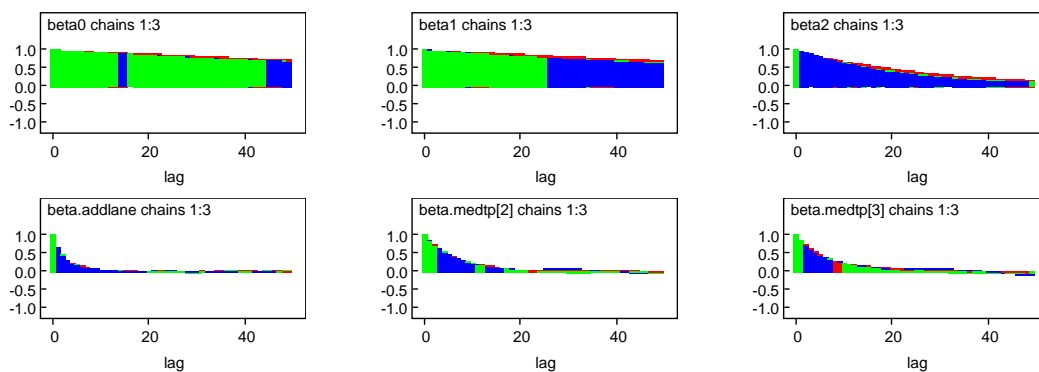


Figura 81| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKS* obtido do conjunto de dados excluindo os trechos com valores omissos.

Os gráficos da Figura 82 mostram que os valores de R (dado pela equação (3.12)) convergem para 1 o que indica convergência do algoritmo.

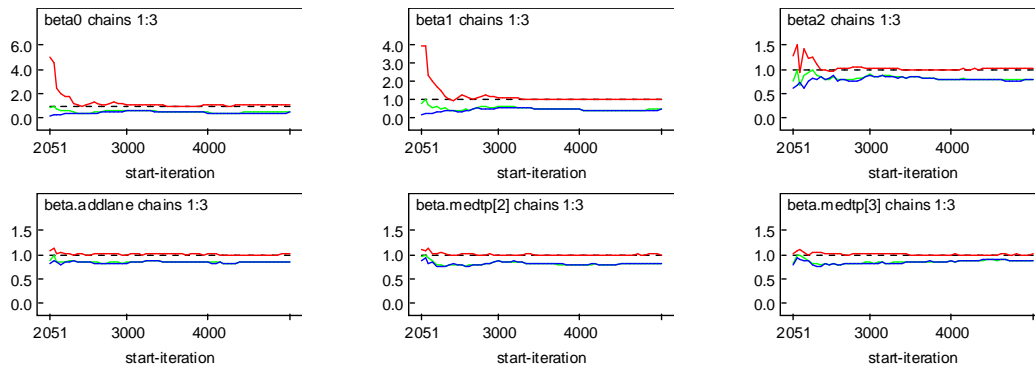


Figura 82| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados excluindo os trechos com valores omissos.

As equações do Quadro 35, cujas estimativas dos parâmetros foram estimadas *a posteriori* usando um modelo Poisson Log-Normal, dão os números esperados de mortos e feridos graves (*NKSI*), para um período de cinco anos, de acordo com os níveis das variáveis categóricas *Additional_Lane* e *Median_Type*.

Quadro 35| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados excluindo os trechos com valores omissos.

	<i>Números Esperados de Mortos e Feridos Graves</i>
<i>Additional_Lane</i> ='No'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 2.943 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$
'GM'	$\hat{\lambda}_i = 3.816 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$
'Vala'	$\hat{\lambda}_i = 5.492 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$
<i>Additional_Lane</i> ='Yes'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 1.731 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$
'GM'	$\hat{\lambda}_i = 2.245 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$
'Vala'	$\hat{\lambda}_i = 3.231 \times 10^{-4} \times AADT_i^{0.789} \times Length_i^{1.201}$

O número de mortos e feridos graves é mais elevado em trechos de auto-estrada onde não haja uma via adicional e onde o separador seja do tipo 'Vala'. Por sua vez, este número é mais baixo em trechos com via adicional e com separador do tipo 'DNJ;NJ'.

O Quadro 36 dá as estimativas pontuais dos parâmetros do modelo obtido. Os intervalos de credibilidade de 95%, à exceção de β_4 , não incluem o valor zero o que indica que as correspondentes variáveis têm um efeito significativo no número de mortos e feridos graves. Visto a densidade *a posteriori* de β_4 se não centrar em zero (como se pode ver no gráfico do meio da

segunda linha na Figura 83, onde o zero aparece na cauda da distribuição), não se deverá rejeitar a variável *Median_Type*.

Quadro 36| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 35.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.131	0.568	3.434×10^{-2}	-9.205	-7.045
β_1 (log(AADT))	0.789	0.051	3.030×10^{-3}	0.690	0.885
β_2 (log(Length))	1.201	0.095	4.081×10^{-3}	1.017	1.391
β_3 (Additional_Lane='Yes')	-0.531	0.204	4.015×10^{-3}	-0.944	-0.135
β_4 (Median_Type='GM')	0.259	0.156	4.371×10^{-3}	-0.040	0.578
β_5 (Median_Type='Vala')	0.624	0.258	6.945×10^{-3}	0.132	1.137

Pela leitura do Quadro 36 depreende-se que com uma via adicional é menor o número de mortos e feridos graves em 41%. Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter aproximadamente mais 30% de acidentes envolvendo mortos e feridos graves do que um trecho com separador do tipo 'DNJ;NJ', nas mesmas condições. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter aproximadamente mais 87% de acidentes envolvendo mortos e feridos graves do que um trecho em condições idênticas de volume de tráfego, comprimento e existência, ou não, de via adicional mas com separador do tipo 'DNJ;NJ'.

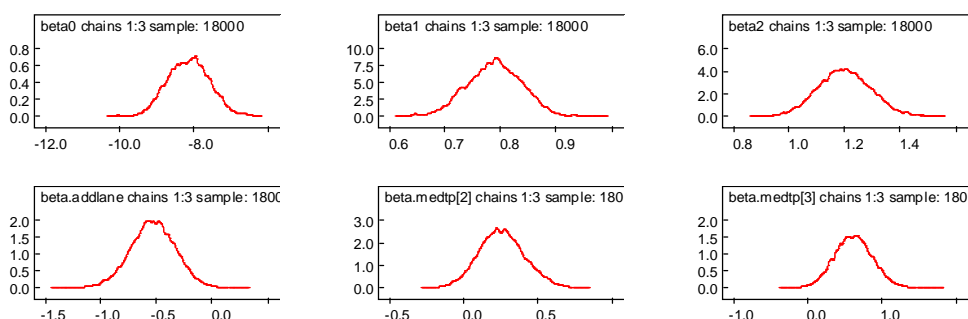


Figura 83| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson Log-Normal ajustado a NKSJ (obtido do conjunto de dados excluindo os trechos com valores omissos)resultantes da simulação de três cadeias de Markov.

O número esperado de mortos e feridos graves em trechos com três tipos de separador e com inclusão, ou não, de vias adicionais, para trechos com uma representação média de *AADT* e *Length*, é dado no Quadro 37.

Quadro 37| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 35.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	3.663 (0.463)	'DNJ;NJ'	2.197 (0.515)
'GM'	4.737 (0.487)	'GM'	2.824 (0.547)
'Vala'	6.939 (1.510)	'Vala'	4.169 (1.258)

Para um trecho típico de auto-estrada sem via adicional e com separador do tipo 'DNJ;NJ' são esperados em média, para um período de cinco anos, 3.663 mortos e feridos graves. Para um trecho também sem via adicional mas com separadores 'GM' e 'Vala', são esperados em média no mesmo período, 4.737 e 6.939 mortos e feridos graves, respectivamente.

Trechos pertencentes a auto-estradas com vias adicionais têm uma representação média do número de mortos e feridos graves mais baixo do que os trechos sem via adicional.

5.2.3.1 Validação do Modelo

Os histogramas de dezanove conjuntos de dados replicados, assim como os dados originais estão exibidos na Figura 84. Uma vez que os histogramas dos dados replicados são bastante semelhantes aos dos dados actuais pode admitir-se que o modelo Poisson Log-Normal considerado está bem ajustado aos dados.

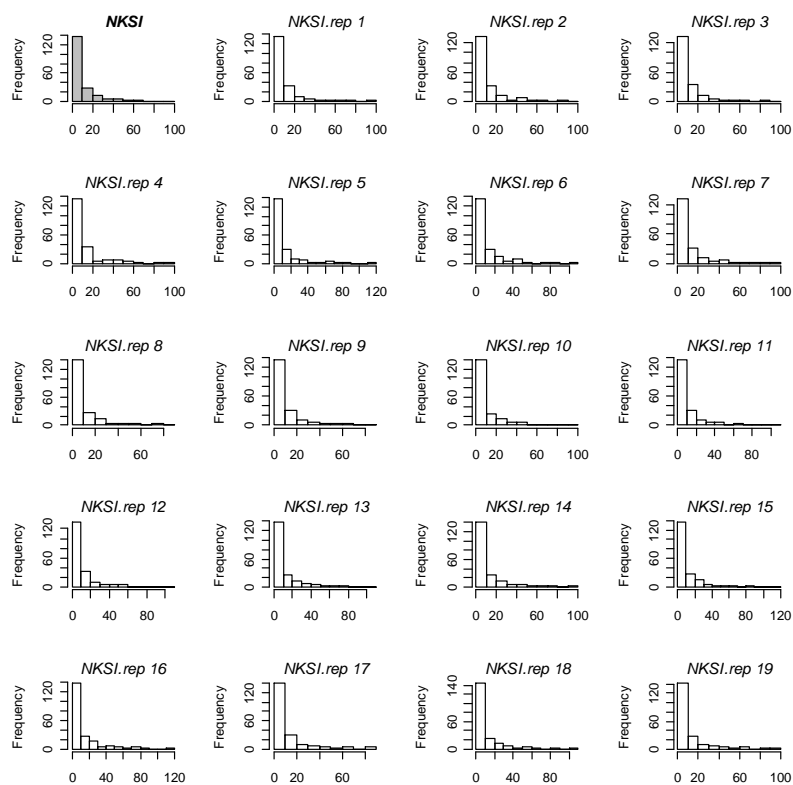


Figura 84| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados excluindo os trechos com valores omissos.

A observação dos histogramas das Figura 85 e Figura 86 mostra que o modelo Poisson Log-Normal capta as variações que os dados observados possuem, uma vez que as rectas verticais se situam aproximadamente no meio dos histogramas dos dados replicados. Além disso, os valores dos p -values são próximos de 0.5.

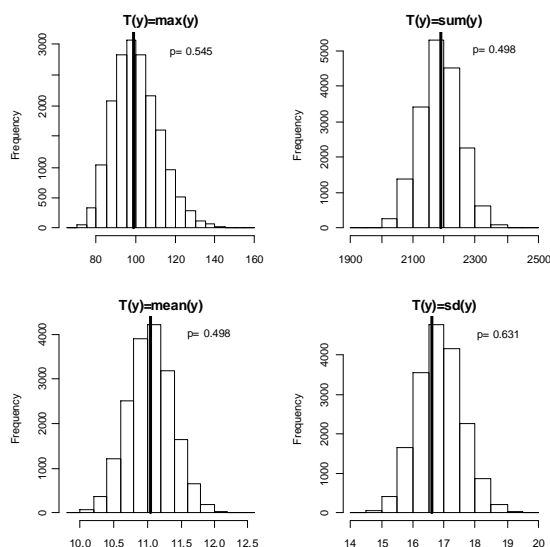


Figura 85] Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Pela observação da Figura 86 também se deprende que o modelo Poisson Log-Normal capta a sobredispersão dos dados observados, repetindo-a também nos dados replicados. O valor de p_B obtido é igual a 0.662, semelhante ao valor equivalente resultante do modelo Poisson-Gama (0.634).

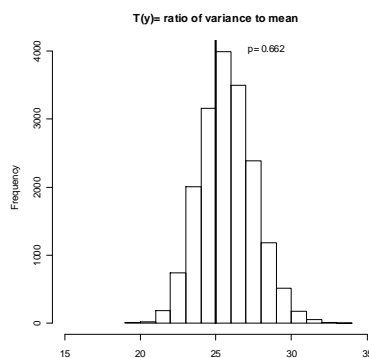


Figura 86] Razão entre a variância e a média de $NKSI$ (recta vertical) comparada com 18000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida dos dados replicados do modelo Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais dada pela equação (3.15) é igual a $\bar{D} = 177.2$ ($n-p=194$), o que, comparado com os graus de liberdade, indica um razoável ajuste do modelo Poisson Log-Normal, assim como uma boa captação da sobredispersão. A estimativa do parâmetro de dispersão é igual a 0.601.

Este modelo produz um valor de DIC igual a 965.16, consideravelmente menor do que o valor obtido (1500.36) quando assumindo um modelo de Poisson, mas ligeiramente maior do que o DIC obtido pelo modelo Poisson-Gama (950.26).

5.3 Discussão

Neste quinto capítulo procedeu-se à descrição do ajuste de três modelos de regressão a dados referentes a números de acidentes (*Nac*) e a números de mortos e feridos graves (*NKSI*), obtidos após exclusão dos trechos de auto-estrada com valores omissos.

O ajuste dos modelos de Poisson, Poisson-Gama e Poisson Log-Normal (sendo os últimos dois, modelos hierárquicos) a *Nac* e a *NKSI*, assim como a selecção de variáveis resultou, para os três modelos, na escolha de *Additional_Lane* e *Median_Type*, além de $\log(AADT)$ e $\log(Length)$, como covariáveis. Embora os modelos Poisson-Gama e Poisson Log-Normal tivessem apresentado resultados muito semelhantes, o modelo Poisson-Gama mostrou ser o mais adequado na modelação de ambos *Nac* e *NKSI*, visto ter apresentado os menores valores de DIC e os mais apropriados valores das médias das desviâncias amostrais (\bar{D}). Tanto o modelo Poisson-Gama como Poisson Log-Normal têm em consideração a sobredispersão existente nestes dados, o que não acontece com o modelo de Poisson.

No que respeita a *Nac*, e tendo em conta os resultados obtidos pelo ajuste do modelo Poisson-Gama, constatou-se que num trecho de auto-estrada com uma via adicional o número esperado de acidentes é cerca de 32% menor quando comparado com o número esperado de acidentes para um trecho em iguais condições (isto é assumindo valores de *AADT*, *Length* e *Median_Type* iguais em ambos os trechos) mas sem vias adicionais. Trechos com separadores do tipo 'GM' e 'Vala' têm um número esperado de acidentes cerca de 29% e 34%, respectivamente, superiores ao número dos mesmos eventos em trechos com separadores do tipo 'DNJ;NJ', obviamente em condições idênticas de *AADT*, *Length* e *Additional_Lane*.

Para *NKSI*, e também de acordo com os resultados obtidos pelo ajuste do modelo Poisson-Gama, pode afirmar-se que num trecho com a presença de via adicional o número esperado de mortos e feridos graves é cerca de 38% inferior ao número dos mesmos eventos esperado para um trecho, em iguais condições de exposição e comprimento, mas sem via adicional. Trechos com separadores do tipo 'GM' e 'Vala' têm, respectivamente, um incremento de cerca de 35% e 68% no número esperado de mortos e feridos graves em relação a segmentos de auto-estrada com separadores do tipo 'DNJ;NJ'.

6 | MODELAÇÃO DOS DADOS CUJOS TRECHOS TENHAM VALORES DE AADT SUPERIORES A 5000 E EXCLUINDO AS VARIÁVEIS COM VALORES OMISSOS

O presente capítulo refere-se à análise dos dados compostos por variáveis que não apresentam valores omissos, e por trechos rodoviários com valores de *AADT* superiores a 5000. Obtiveram-se modelos de previsão de acidentes (*Nac*), de mortos (*NF*), de feridos graves (*NSI*) e da soma dos últimos dois (*NKSI*).

6.1 Modelação do Número de Acidentes

Para uma descrição detalhada dos dados analisados nesta secção, nomeadamente relativos a *Nac*, ver secção 6.1 de Azeredo Lopes e Cardoso (2009).

Os resultados referentes aos modelos descritos no presente capítulo foram obtidos de uma amostra de dimensão 3000 após 30000 iterações de cada uma de três cadeias de Markov das quais 20000 foram descartadas com *thinning rate* igual a 10.

6.1.1 Modelo de Poisson

O modelo que obteve o menor valor de DIC quando se ajustou o modelo de Poisson contém as variáveis $\log(AADT)$, $\log(Length)$, *Additional_Lane*, *Lane_Width* e *Number_of_Lanes*. A equação que dá o valor esperado dos números de acidentes é:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Lane_Width_i + \beta_5 Number_of_Lanes_i \quad (6.1)$$

Os gráficos da Figura 87 mostram não haver razões para duvidar da convergência dos parâmetros β_j da equação (6.1). No entanto pode reparar-se que os valores das iterações das três cadeias de Markov não se “interligam” tão eficazmente para β_0 e β_1 .

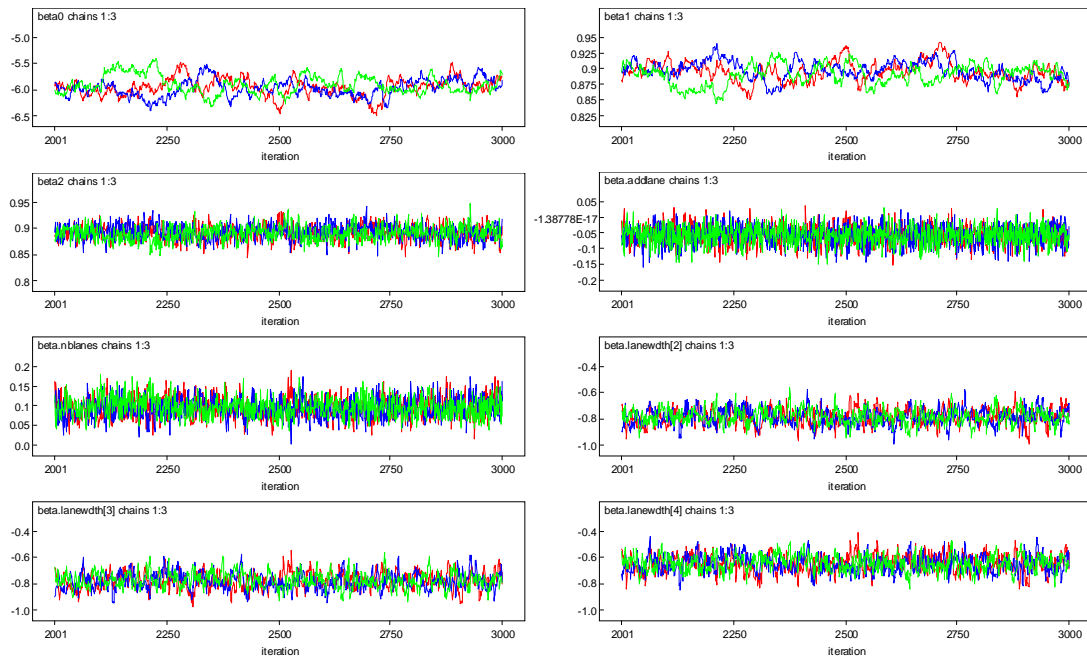


Figura 87| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *Nac* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações na Figura 88 mostram que β_0 e β_1 convergem lentamente.

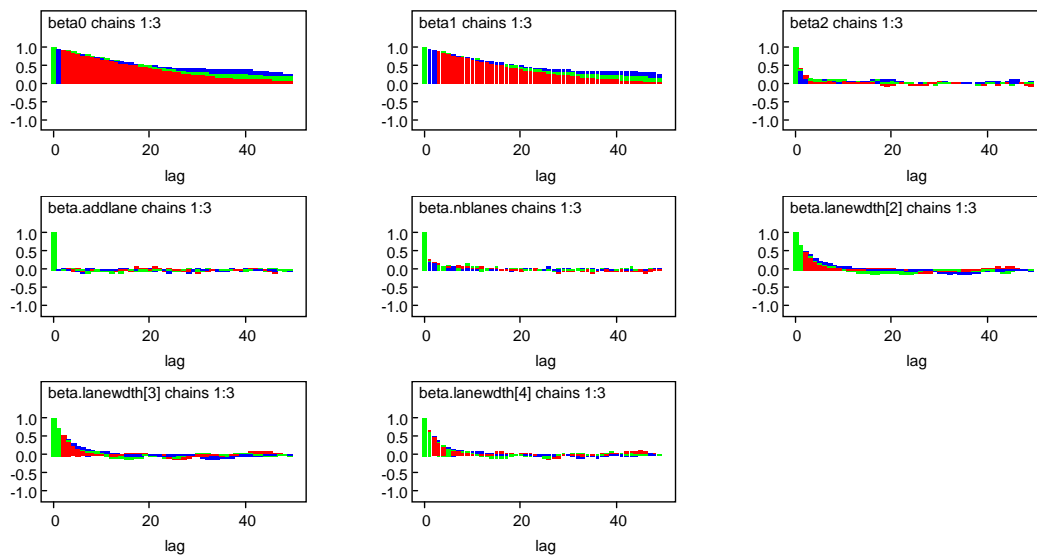


Figura 88| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *Nac* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser observados na Figura 89, estabilizam em certos valores, e em particular para 1 no que diz respeito à estatística *R* (a vermelho na figura), o que

indica que as estimativas dos parâmetros, dadas pelas médias *a posteriori*, convergem para os mesmos valores em cada uma das três cadeias.

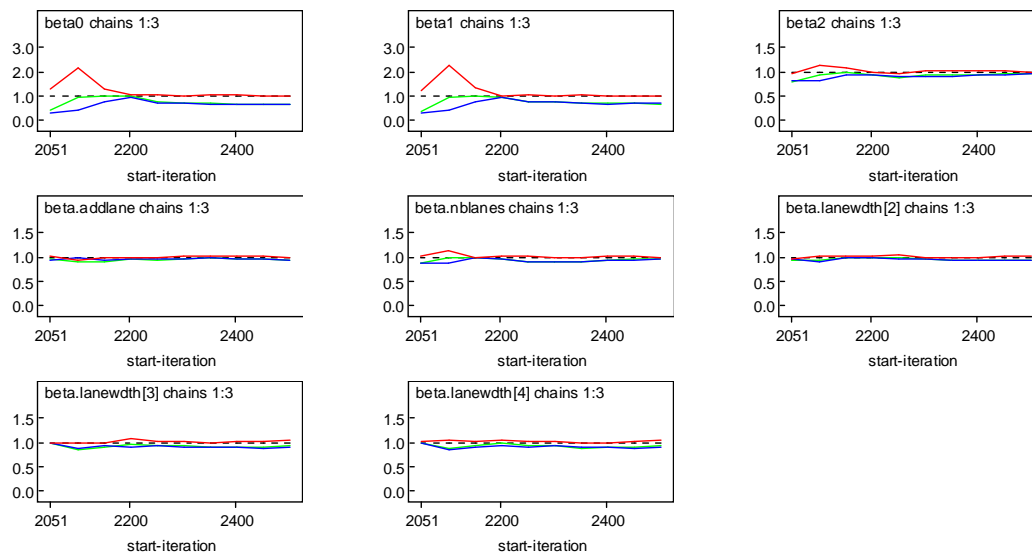


Figura 89] Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As equações que dão os números esperados de acidentes para os diversos níveis das variáveis categóricas obtidas pelo modelo de Poisson podem ser observadas no Quadro 38.

Quadro 38| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

Números Esperados de Acidentes	
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='A'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.616 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'B'	$\hat{\lambda}_i = 1.193 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'C'	$\hat{\lambda}_i = 1.202 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'D'	$\hat{\lambda}_i = 1.370 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
<i>Additional_Lane='No'</i>	
<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.877 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'B'	$\hat{\lambda}_i = 1.312 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'C'	$\hat{\lambda}_i = 1.322 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'D'	$\hat{\lambda}_i = 1.506 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='A'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.469 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'B'	$\hat{\lambda}_i = 1.126 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'C'	$\hat{\lambda}_i = 1.135 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'D'	$\hat{\lambda}_i = 1.293 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
<i>Additional_Lane='Yes'</i>	
<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>	
'A'	$\hat{\lambda}_i = 2.716 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'B'	$\hat{\lambda}_i = 1.238 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'C'	$\hat{\lambda}_i = 1.248 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$
'D'	$\hat{\lambda}_i = 1.422 \times 10^{-3} \times AADT_i^{0.894} \times Length_i^{0.891}$

As estimativas pontuais e os intervalos de credibilidade de 95% para o valor médio das densidades *a posteriori* dos parâmetros β_j podem ser observados no Quadro 39. Pela leitura deste quadro pode afirmar-se que um trecho com via adicional tem um número esperado de acidentes cerca de 6% menor do que um trecho em idênticas circunstâncias, mas sem via adicional. Se o trecho tiver três ou quatro vias (*Number_of_Lanes='B'*) é esperado ter um número de acidentes cerca de 10% mais elevado do que um trecho nas mesmas condições, mas com somente 2 vias (*Number_of_Lanes='A'*).

Quadro 39| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 38.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-5.946	0.173	1.510×10^{-2}	-6.298	-5.587
β_1 (log(AADT))	0.894	0.016	1.363×10^{-3}	0.861	0.926
β_2 (log(Length))	0.891	0.014	5.280×10^{-4}	0.864	0.920
β_3 (Additional_Lane='Yes')	-0.058	0.030	6.289×10^{-4}	-0.118	-1.990×10^{-5}
β_4 (Lane_Width='B')	-0.786	0.060	2.493×10^{-3}	-0.906	-0.669
β_5 (Lane_Width='C')	-0.778	0.062	2.536×10^{-3}	-0.897	-0.655
β_6 (Lane_Width='D')	-0.647	0.063	2.629×10^{-3}	-0.768	-0.519
β_7 (Number_of_Lanes='B')	0.095	0.025	8.310×10^{-4}	0.047	0.148

Um trecho com largura de via no intervalo]3.00, 3.50] tem uma redução no número esperado de acidentes em cerca de 54% em relação a trechos com largura de via menor que 3 metros. Trechos com largura de vias nos intervalos]3.50, 3.75] e]3.75, ∞[têm uma redução de cerca de 54% e 48%, respectivamente, em relação a trechos em iguais circunstâncias, mas com largura de via menor que 3 metros.

As densidades *a posteriori* dos parâmetros β_j podem ser observadas na Figura 90.

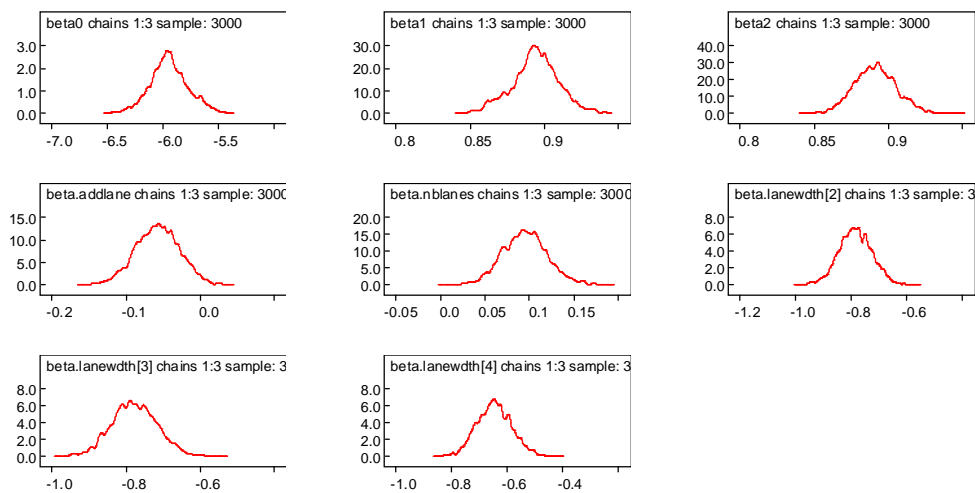


Figura 90| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *Nac* (obtido do conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Foram calculados os números esperados de acidentes para os vários níveis das variáveis categóricas para uma representação média de $\log(AADT)$ e $\log(Length)$. As médias *a posteriori* e os correspondentes desvios padrões das representações médias encontram-se no Quadro 40. Para um trecho representativo de auto-estrada com três ou quatro vias, cujas larguras de via são menores que 3 metros, e sem via adicional, esperam-se 102.4 acidentes em cinco anos. Enquanto que para um trecho com duas vias, com larguras entre 3.00 e 3.50 metros, e com via adicional espera-se, para o mesmo período de tempo, 40.03 acidentes.

Quadro 40] Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 38.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	93.13 (5.580)	'A'	102.40 (5.690)
'B'	42.39 (1.161)	'B'	46.62 (1.230)
'C'	42.73 (0.845)	'C'	47.01 (1.275)
'D'	48.69 (1.272)	'D'	53.55 (1.477)
<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Number_of_Lanes='A'</i>		<i>Number_of_Lanes='B'</i>	
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	87.94 (5.790)	'A'	96.70 (6.144)
'B'	40.03 (1.473)	'B'	44.02 (1.709)
'C'	40.34 (1.177)	'C'	44.38 (1.662)
'D'	45.97 (1.725)	'D'	50.57 (2.081)

6.1.1.1 Validação do Modelo

Os dezanove histogramas resultantes de extracções (*draws*) da distribuição preditiva *a posteriori* dos valores dos números de acidentes, assim como o histograma de *Nac* encontram-se representados na Figura 91. Os dezanove histogramas são comparados com o histograma de *Nac* para investigar se é credível que os números de acidentes provêm de uma distribuição de Poisson (que é a assumida pelo modelo).

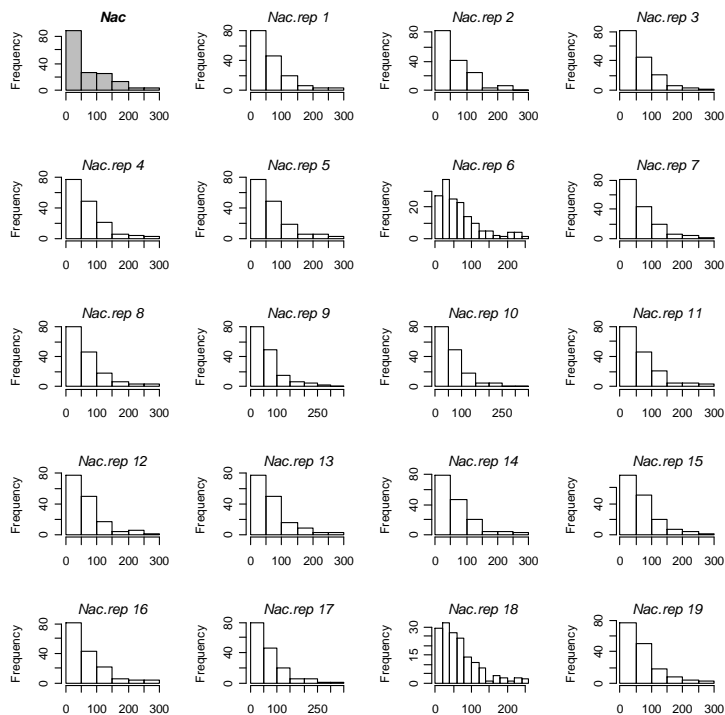


Figura 91| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Alguns dos histogramas dos dados replicados diferem bastante do histograma de *Nac*, nomeadamente *Nac.rep 6* e *Nac.rep 18*. Um método para medir a discrepância entre os dados observados e os que foram obtidos pela distribuição *a posteriori* é comparar quantidades *T* resultantes de 3000 replicações hipotéticas com as quantidades equivalentes observadas em *Nac*.

Os histogramas na Figura 92, e na Figura 93, mostram os valores das diversas quantidades *T* em cada uma das 3000 replicações hipotéticas (na Figura 91 somente 19 das 3000 réplicas foram exibidas); em particular as quantidades referentes aos valores máximos, desvio padrão (Figura 92) e razão entre variância e média (Figura 93) são muito inferiores aos correspondentes valores de *T* obtidos dos dados *Nac*, que são indicados por rectas verticais. Claramente, o modelo de Poisson não capta as variações que foram observadas pelos dados.

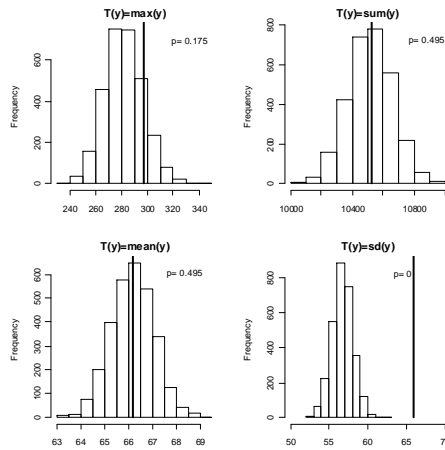


Figura 92| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

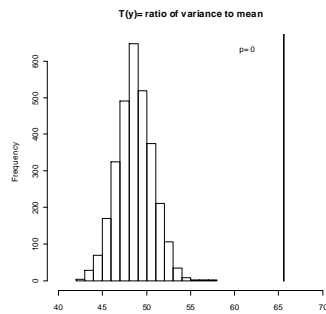


Figura 93| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais é igual a $\bar{D}=3083$, valor extremamente elevado quando comparado com $n-p=154$, o que indica um mau ajuste do modelo de Poisson ao número de acidentes devido à não captação da sobredispersão dos dados. O valor de DIC obtido foi 3989.

6.1.2 Modelo de Poisson-Gama

O modelo de Poisson-Gama foi ajustado aos dados no *software* WinBUGS segundo as equações (3.27). O modelo que produziu o menor valor de DIC inclui a variáveis $\log(AADT)$, $\log(Length)$ e $Additional_Lane$. O número esperado de acidentes, $\hat{\lambda}_i$, é dado pela equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i, \quad (6.2)$$

Os gráficos das iterações vs valores gerados apresentados na Figura 94 mostram que ambos os parâmetros β_0 e β_1 não parecem convergir para esta simulação em particular. No entanto, não há razões para crer que β_2 e β_3 não convirjam.

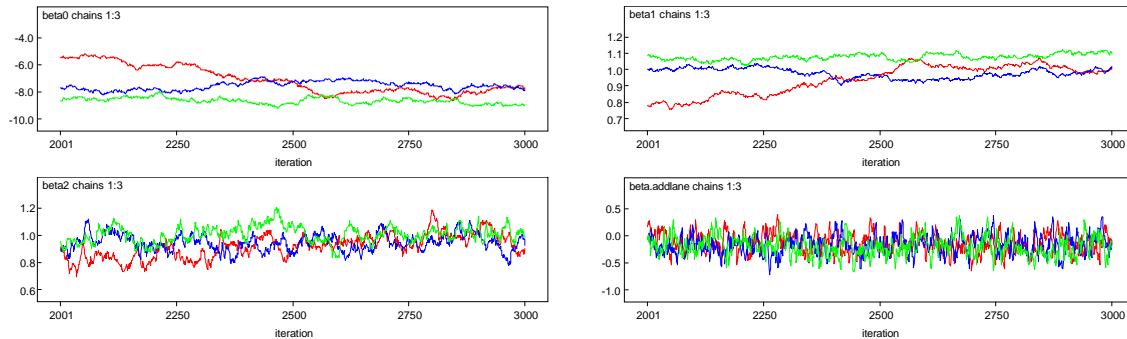


Figura 94| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a Nac obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos da autocorrelação na Figura 95 também mostram evidência de não convergência para β_0 e β_1 e convergência mais lenta para β_2 , embora nenhuma evidência de não convergência para β_3 .

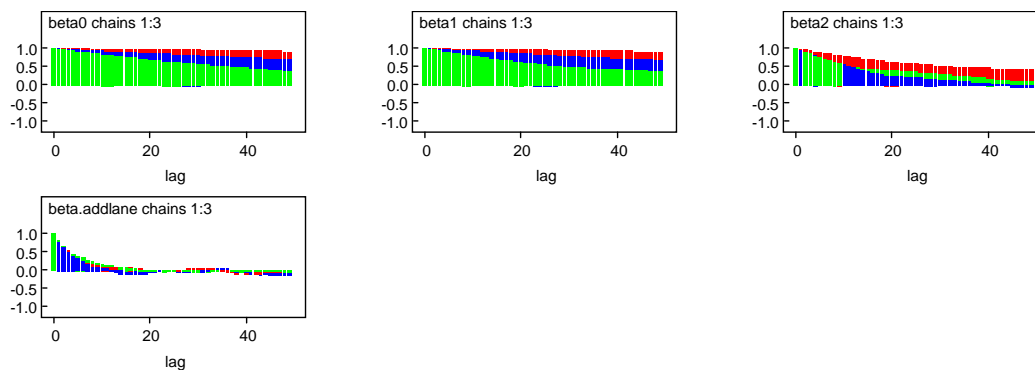


Figura 95| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a Nac obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Uma vez que as estatísticas de Gelman-Rubin convergem e estabilizam (em particular R converge para 1 em todos os parâmetros), como se pode ver na Figura 96, pode assumir-se que o número de iterações escolhido é suficiente para obter distribuições *a posteriori* confiáveis.

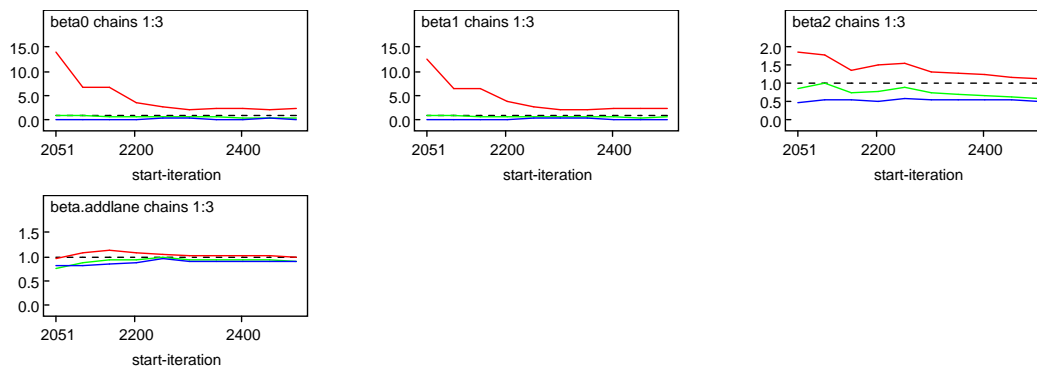


Figura 96| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

As equações para os números esperados de acidentes obtidas através do modelo Poisson-Gama podem ser observadas no Quadro 41 para os dois níveis da variável categórica *Additional_Lane*.

Quadro 41| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 4.145 \times 10^{-4} \times AADT_i^{0.998} \times Length_i^{0.955}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 3.502 \times 10^{-4} \times AADT_i^{0.998} \times Length_i^{0.955}$

As estimativas para os coeficientes do modelo Poisson-Gama ajustado a N_{ac} estão expostas no Quadro 42. Embora o intervalo de credibilidade de 95% para a (média da) estimativa do parâmetro correspondente a *Additional_Lane* contenha o valor zero, pode observar-se pela leitura do gráfico da densidade *a posteriori*, na Figura 97, que o valor médio dessa estimativa *a posteriori* se afastou visivelmente de zero.

Quadro 42| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 41.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.778	0.895	9.211×10^{-2}	-9.010	-5.493
β_1 (log(AADT))	0.998	0.080	8.199×10^{-3}	0.793	1.106
β_2 (log(Length))	0.955	0.080	7.211×10^{-3}	0.786	1.102
β_3 (<i>Additional_Lane</i> ='No')	-0.169	0.183	9.831×10^{-3}	-0.523	0.210

Pela leitura da estimativa do parâmetro referente a *Additional_Lane* conclui-se que o número esperado de acidentes num trecho com via adicional é cerca de 16% menor do que um trecho em iguais circunstâncias, mas sem via adicional.

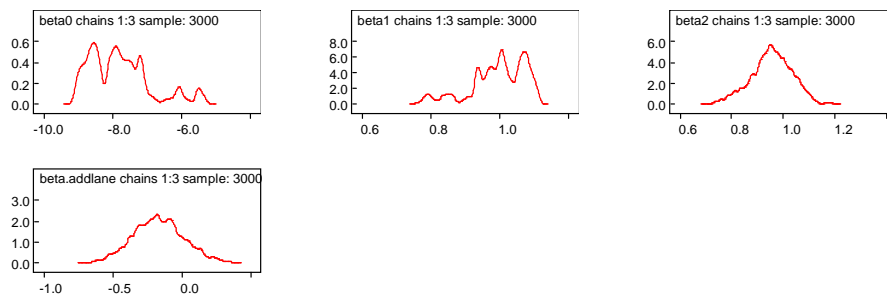


Figura 97| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson-Gama ajustado a *Nac* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

As médias *a posteriori*, e correspondentes desvios padrões, para uma representação média de $\log(AADT)$ e $\log(Length)$ estão exibidos no Quadro 43.

Quadro 43| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 41.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	46.760 (2.845)
<i>Additional_Lane='Yes'</i>	40.030 (7.028)

6.1.2.1 Validação do Modelo

Pela comparação do histograma de *Nac* (a cinzento) com os restantes histogramas na Figura 98, pode-se afirmar que o modelo Poisson-Gama (ver equações no Quadro 41) representa adequadamente os números de acidentes dos dados considerados no presente capítulo.

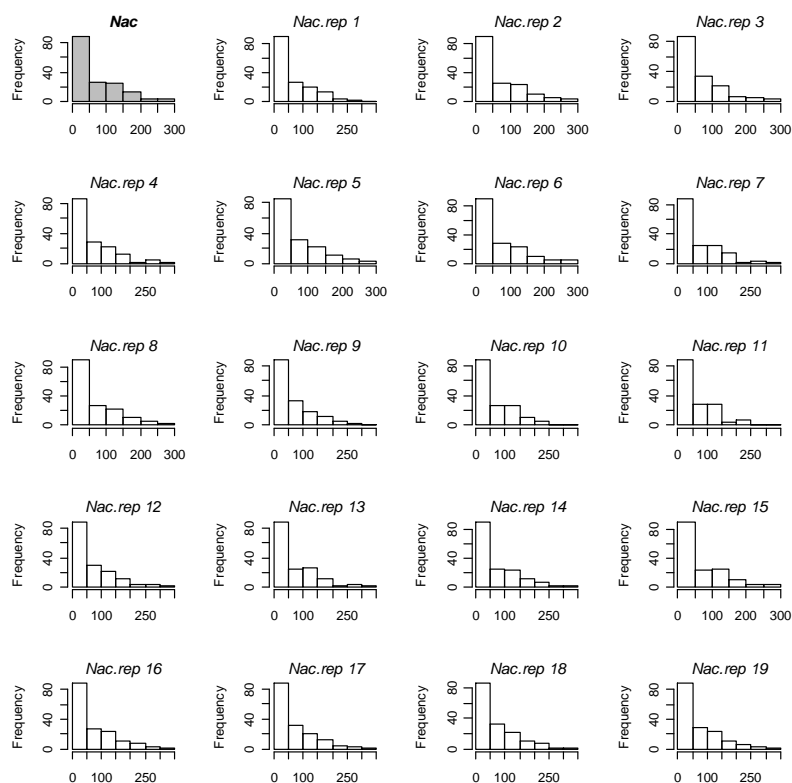


Figura 98] Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As medidas de discrepância mostram que o modelo Poisson-Gama captura as variações ocorridas nos dados observados, como se pode constatar pela observação dos gráficos na Figura 99.

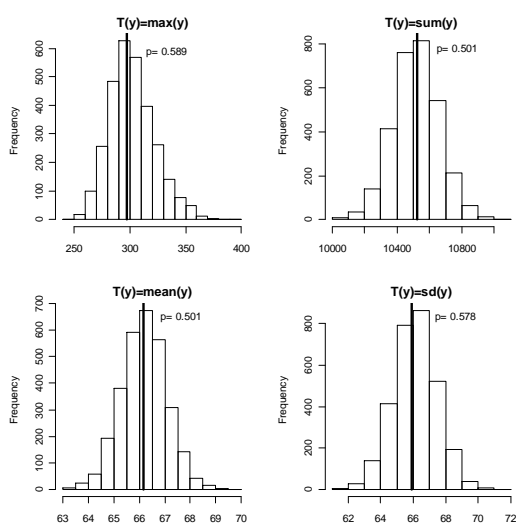


Figura 99] Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O *p* é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

Pela observação da Figura 100, referente à medida de discrepância relativa à razão entre a variância e a média, pode ver-se que o histograma dos valores dessa razão obtidos dos valores da distribuição *a posteriori* preditiva “englobam” o valor da razão obtido dos dados observados (recta vertical). O que significa que os valores replicados pelo modelo e os dados observados possuem idênticas razões entre a variância e a média e conseqüentemente os mesmos “níveis” de sobredispersão.

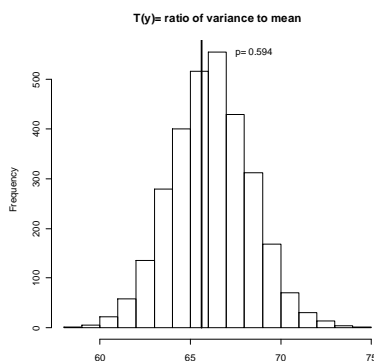


Figura 100| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Este modelo obteve uma média das desviâncias amostrais igual a $\bar{D}=159$, valor mais adequado quando comparado a $n-p$ (156) do que o valor equivalente obtido pelo ajuste do modelo de Poisson, o que indica que o modelo Poisson-Gama captura a sobredispersão ocorrente nos dados. A estimativa do parâmetro de dispersão é igual a 0.502 e o valor de DIC é 1156.4.

6.1.3 Modelo de Poisson Log-Normal

O modelo Poisson Log-Normal ajustado aos dados de acordo com as equações (3.31) resultou no modelo com as seguintes covariáveis:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (6.3)$$

Este modelo reúne o conjunto de variáveis que obtêm o menor valor de DIC.

Os gráficos respeitantes aos *trace plots* das estimativas dos quatro parâmetros β_j , na Figura 101, deixam ver que os parâmetros β_0 e β_1 não parecem convergir devido ao facto das iterações correspondentes às três cadeias de Markov não se interligarem e convergirem para o mesmo valor. No entanto não parece haver razões para suspeitar da não convergência referente aos parâmetros β_1 e β_2 (na segunda linha da Figura 101).

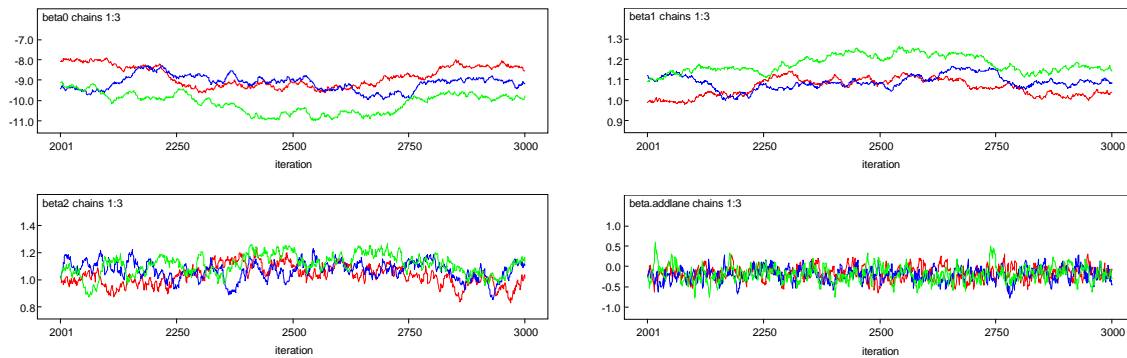


Figura 101| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações exibidos na Figura 102 deixam ver que as estimativas de β_0 e β_1 não parecem convergir e que a estimativa de β_2 converge muito lentamente.

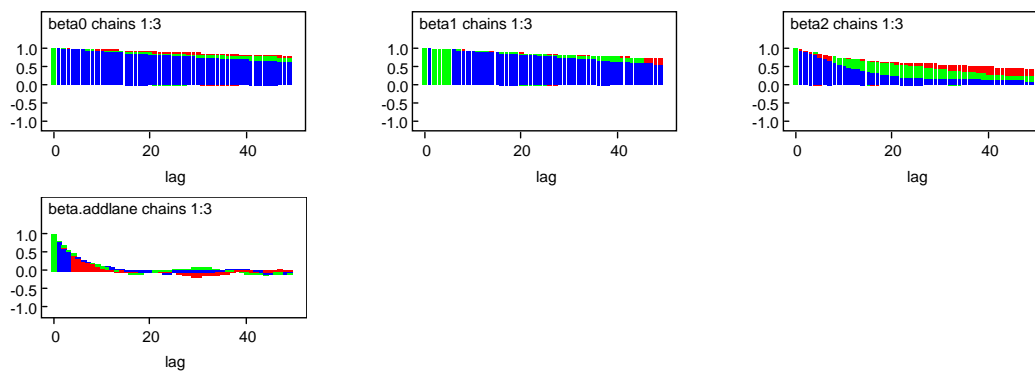


Figura 102| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser observados na Figura 103, mostram que \hat{V} e W estabilizem à medida que o número de iterações aumenta. O valor de R (a vermelho na Figura 103) converge para 1, embora mais lentamente para os parâmetros β_0 e β_1 .

No entanto, seria preferível ter repetido a simulação com um número mais elevado de iterações.

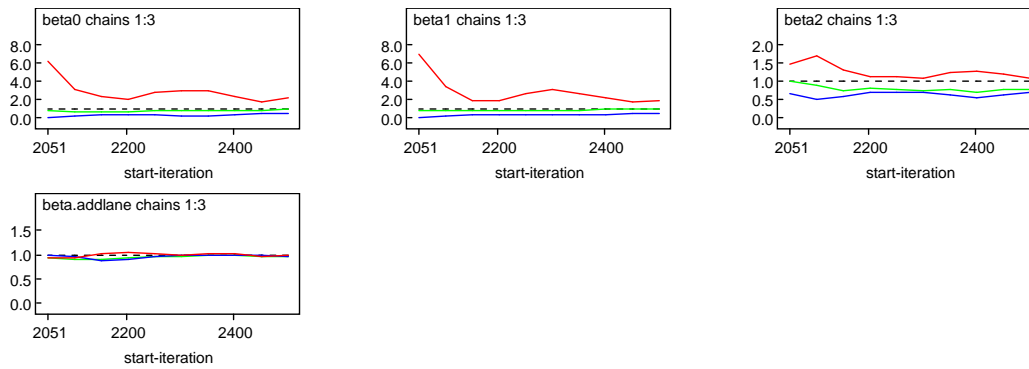


Figura 103| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As equações para o número esperado de acidentes obtido através do modelo Poisson Log-Normal encontram-se no Quadro 44 para os dois níveis da variável categórica *Additional_Lane*.

Quadro 44| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Acidentes
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 8.670 \times 10^{-5} \times AADT_i^{1.109} \times Length_i^{1.070}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 7.211 \times 10^{-5} \times AADT_i^{1.109} \times Length_i^{1.070}$

As estimativas pontuais e intervalos de credibilidade são dados no Quadro 45. A interpretação da estimativa relativa a *Additional_Lane* indica que um trecho com a presença de uma via adicional tem um número de acidentes cerca de 17% menor do que um trecho em condições idênticas, mas sem via adicional.

Quadro 45| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 44.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.353	0.717	7.326×10^{-2}	-10.840	-8.060
β_1 (log(AADT))	1.109	0.063	6.421×10^{-3}	0.997	1.239
β_2 (log(Length))	1.070	0.078	5.280×10^{-4}	0.907	1.218
β_3 (<i>Additional_Lane</i> ='Yes')	-0.184	0.176	6.973×10^{-3}	-0.524	0.177

Os gráficos das densidades das distribuições *a posteriori* (na Figura 104) obtidas para as estimativas dos parâmetros β_j mostram vários picos para as densidades de β_0 e β_1 o que evidencia a não convergência desses parâmetros.

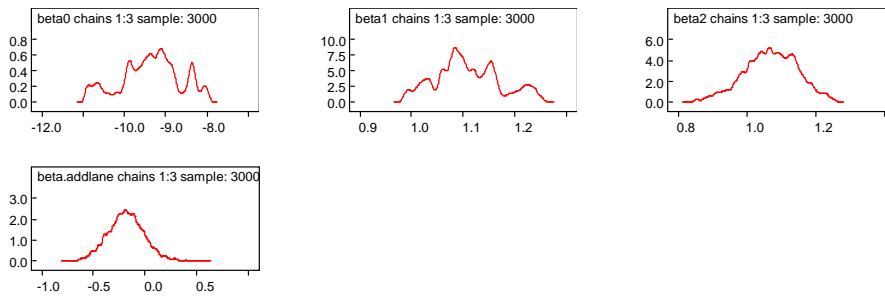


Figura 104| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *Nac* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

As médias *a posteriori* e os correspondentes desvios padrões das representações médias de $\log(AADT)$ e $\log(Length)$ para os dois níveis de *Additional_Lane* obtidos pelo modelo Poisson Log-Normal são dados no Quadro 46. Para um trecho representativo de auto-estrada sem via adicional esperam-se cerca de 36.45 acidentes, enquanto que para um trecho com via adicional se esperam à volta de 30.65 acidentes.

Quadro 46| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 44.

	média (desvio padrão)
<i>Additional_Lane</i> ='No'	36.450 (2.448)
<i>Additional_Lane</i> ='Yes'	30.650 (5.070)

6.1.3.1 Validação do Modelo

Os dezanove histogramas na Figura 105 são exemplos de conjuntos de dados que se esperam observar sob a distribuição *a posteriori* obtida pelo modelo Poisson Log-Normal ajustado aos dados considerados neste capítulo. Como não se verificam grandes diferenças entre o histograma de *Nac* (no canto superior esquerdo) e os restantes histogramas, pode afirmar-se que os dados observados são representativos do modelo Poisson Log-Normal considerado.

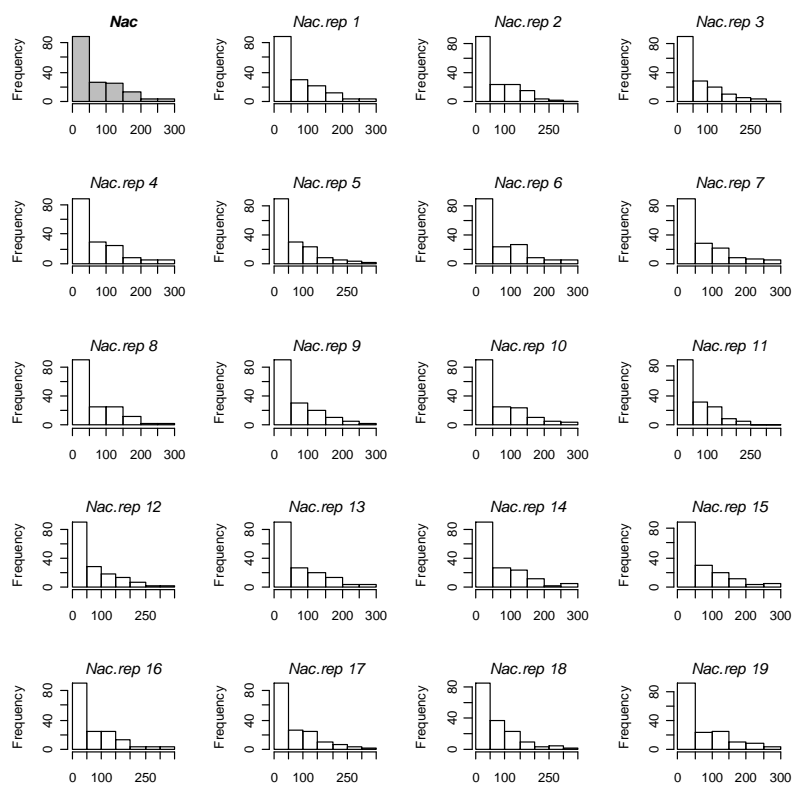


Figura 105| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As medidas de discrepância calculadas, ver Figura 106, mostram que o modelo Poisson Log-Normal capta as variações dos dados observados.

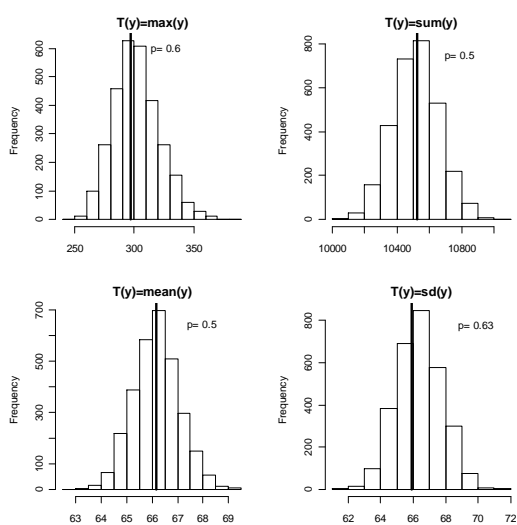


Figura 106| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O *p* é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

O modelo considerado também capta a sobredispersão dos dados, pois a razão entre a variância e a média nos dados observados (valor representado pela recta vertical na Figura 107) é representativa de dados obtidos do modelo sob a distribuição preditiva *a posteriori*.

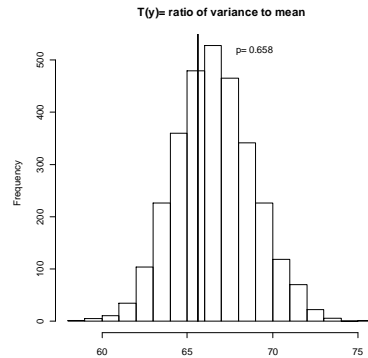


Figura 107| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais resultou num valor $\bar{D}=175$ ($n-p=156$), uma estimativa para o parâmetro de dispersão igual a 0.511 e DIC=1178.7.

6.2 Modelação do Número de Mortos

Descrições e detalhes relativos aos números de mortos (*NF*) podem ser encontrados na secção 6.2 de Azeredo Lopes e Cardoso (2009).

6.2.1 Modelo de Poisson

O processo de selecção de variáveis baseado no menor valor de DIC obtido resultou no modelo com $\log(AADT)$, $\log(Length)$ e $Lane_Width$ como covariáveis quando o modelo de Poisson foi ajustado (ver equações (3.22)) tendo como variável resposta o número de mortos. O número esperado de mortos, $\hat{\lambda}_i$, é dado pela seguinte equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Lane_Width_i \quad (6.4)$$

Os gráficos das autocorrelações dos parâmetros β_j referentes à equação (6.4) estão exibidos na Figura 108. A interpretação destes gráficos não sugere que haja razões para duvidar da convergência dos parâmetros considerados.

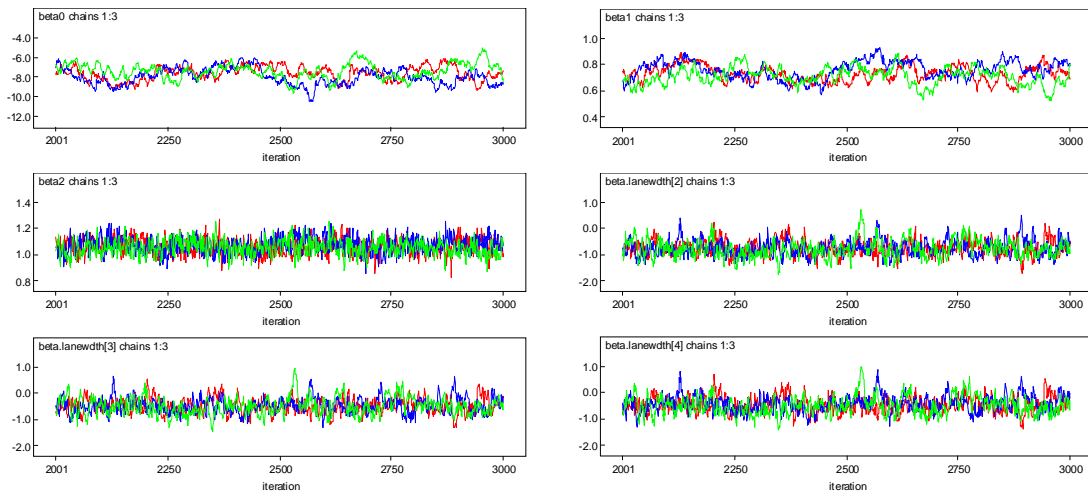


Figura 108] Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As autocorrelações na Figura 109 mostram que as convergências dos parâmetros β_0 e β_1 são mais lentas do que a dos restantes parâmetros.

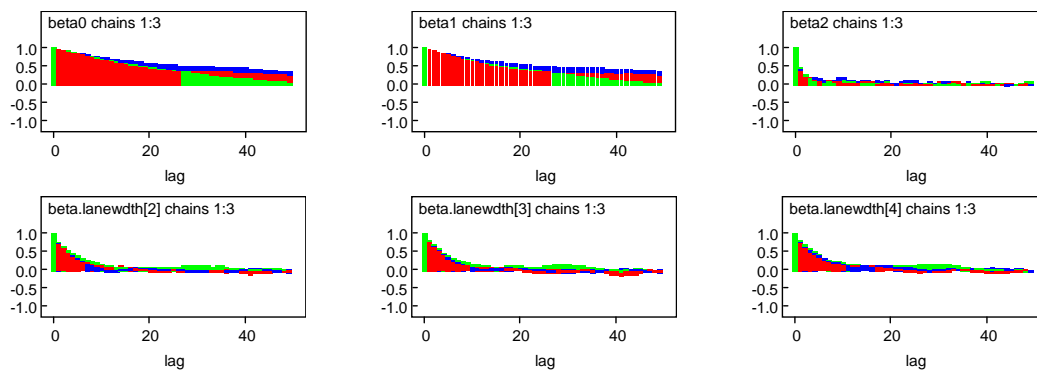


Figura 109] Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos se podem observar na Figura 110, mostram que R convergiu para o valor 1 e aí se estabilizou à medida que o número de iterações aumentou.

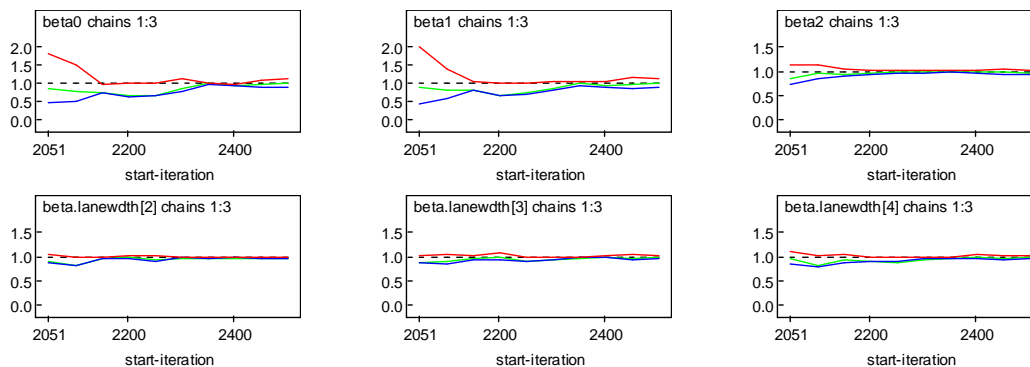


Figura 110| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As equações referentes ao número esperado de mortos, para os quatro níveis da covariável $Lane_Width$ e para os valores de $AADT$ e $Length$ são dadas no Quadro 47. Pode depreender-se que o maior número de mortos é esperado em trechos com uma largura de via menor que três metros ($Lane_Width='A'$).

Quadro 47| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Mortos
$Lane_Width$	
'A'	$\hat{\lambda}_i = 4.560 \times 10^{-4} \times AADT_i^{0.731} \times Length_i^{1.059}$
'B'	$\hat{\lambda}_i = 2.114 \times 10^{-4} \times AADT_i^{0.731} \times Length_i^{1.059}$
'C'	$\hat{\lambda}_i = 2.790 \times 10^{-4} \times AADT_i^{0.731} \times Length_i^{1.059}$
'D'	$\hat{\lambda}_i = 2.858 \times 10^{-4} \times AADT_i^{0.731} \times Length_i^{1.059}$

O Quadro 48 contém as estimativas pontuais e os intervalos de credibilidade de 95% para os parâmetros β_j da equação (6.4). As densidades das distribuições *a posteriori* para os parâmetros β_j estão apresentadas na Figura 111.

Quadro 48| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 47.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.693	0.811	7.319×10^{-2}	-9.189	-6.163
β_1 ($\log(AADT)$)	0.731	0.068	6.021×10^{-3}	0.603	0.860
β_2 ($\log(Length)$)	1.059	0.062	2.587×10^{-3}	0.939	1.180
β_3 ($Lane_Width='B'$)	-0.769	0.312	1.660×10^{-2}	-1.355	-0.101
β_4 ($Lane_Width='C'$)	-0.491	0.305	1.712×10^{-2}	-1.050	0.175
β_5 ($Lane_Width='D'$)	-0.467	0.321	1.868×10^{-2}	-1.042	0.229

Pela leitura das estimativas relativas à variável *Lane_Width* depreende-se que um trecho com largura de via entre 3.00 e 3.50 metros tem uma redução no número de mortos em cerca de 54%, quando comparado com um trecho em idêntica situação, mas com largura de via menor que 3.00 metros. Trechos com larguras de via entre 3.50 e 3.75 metros, assim como larguras de via maiores que 3.75 metros, têm menores números de mortos em cerca de 39% e 37%, respectivamente, relativamente a um trecho em iguais circunstâncias, mas com via de largura menor que 3.00 metros.

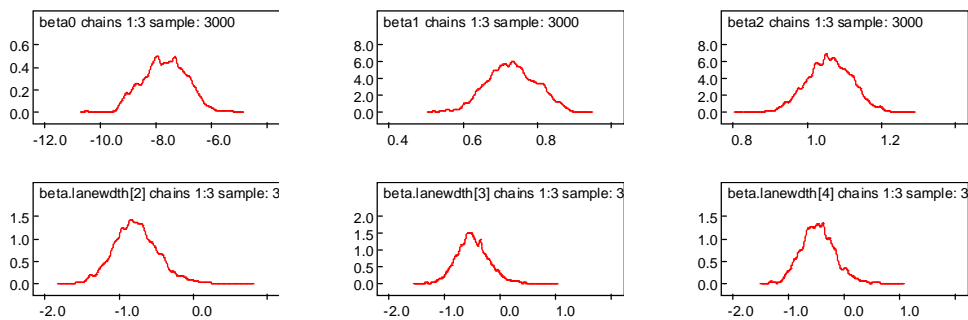


Figura 111| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NF* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos em trechos segundo os vários tipos de largura de via, para uma representação média de *AADT* e *Length* são dados no Quadro 49. Para um trecho representativo de auto-estrada com largura de via menor que 3.00 metros são esperados 4.174 mortos num período de cinco anos. O menor número esperado de mortos está associado a trechos com largura de via entre 3.00 e 3.50 metros.

Quadro 49| Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 47.

<i>Lane_Width</i>	
'A'	4.174 (1.185)
'B'	1.870 (0.225)
'C'	2.458 (0.190)
'D'	2.524 (0.269)

6.2.1.1 Validação do Modelo

Se o modelo está bem ajustado, dados replicados por esse modelo deverão ser semelhantes aos dados observados (Gelman *et al.* (2004)). O exame dos histogramas na Figura 112 mostra que os dados replicados produzem histogramas em nada semelhantes aos dos dados observados.

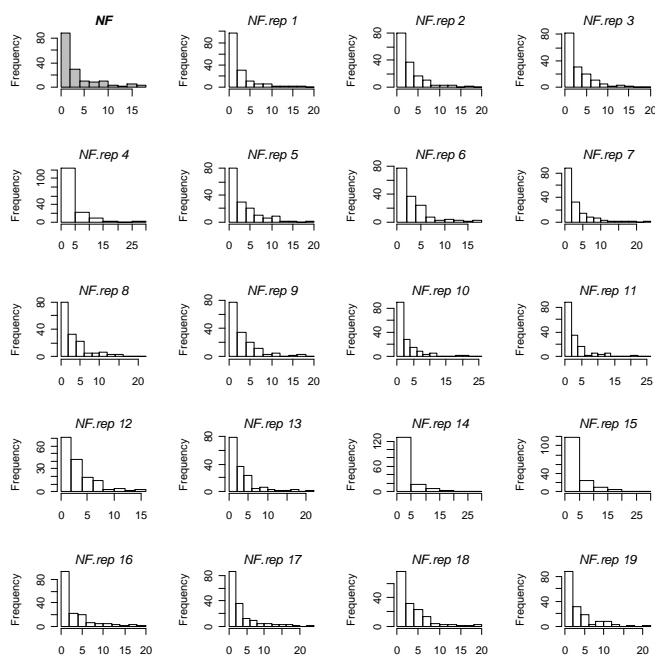


Figura 112| Histograma de *NF* (no canto superior esquerdo) e dezanove réplicas (“*NF.rep*”) dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As medidas de discrepância mostram que o modelo de Poisson ajustado aos dados em causa não captura a variação que os dados observados possuem (ver Figura 113).

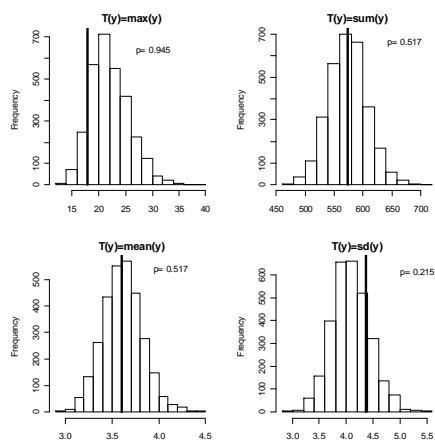


Figura 113| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

A sobredispersão também não é captada pelo modelo como pode ser observado na Figura 114. O valor de p_B é igual a 0.173, valor bastante baixo (idealmente seria aproximadamente igual a 0.5).

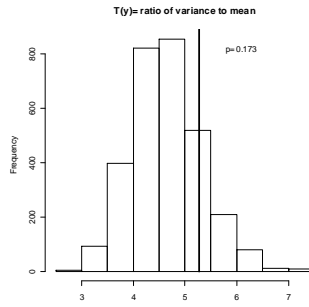


Figura 114| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a posteriori preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo de Poisson ajustado a NF obteve um valor $\bar{D}=243.5$, bastante elevado quando comparado a $n-p=156$ (mais uma vez indicando que o modelo não capturou a sobredispersão nos dados) e um valor DIC igual a 689.62.

6.2.2 Modelo de Poisson-Gama

O modelo de Poisson-Gama resultante de um ajuste aos dados de acordo com as equações (3.27) tem a seguinte forma genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) \quad (6.5)$$

Onde $\hat{\lambda}_i$ dá o número esperado de mortos, para cada trecho, num período de cinco anos. O modelo com o menor valor de DIC é obtido com as variáveis $\log(AADT)$ e $\log(Length)$, isto é, a introdução de quaisquer outras variáveis vai aumentar o valor de DIC.

A Figura 115 mostra os gráficos respeitantes aos *trace plots* dos três parâmetros β_j considerados. Não parece haver razões para suspeitar da não convergência dos parâmetros, embora as iterações dos parâmetros β_0 e β_1 não se interliguem tanto para as três cadeias de Markov como por exemplo as iterações de β_2 .

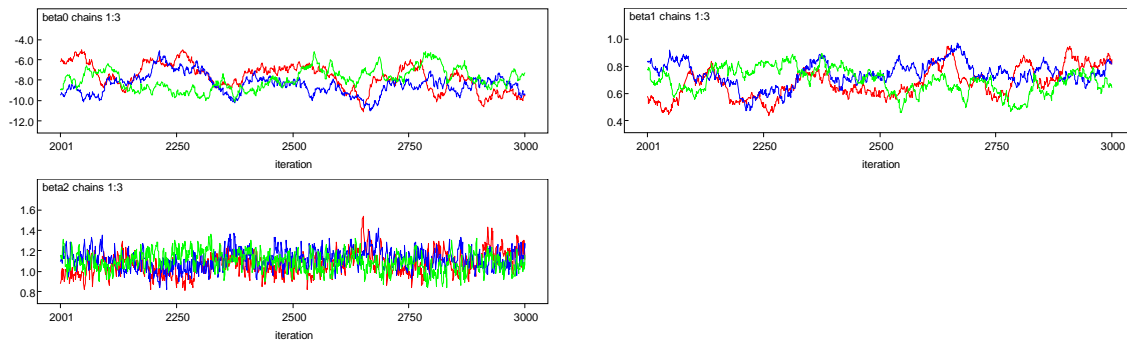


Figura 115| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações na Figura 116 sugerem que os parâmetros β_0 e β_1 convergem mais lentamente que β_2 .

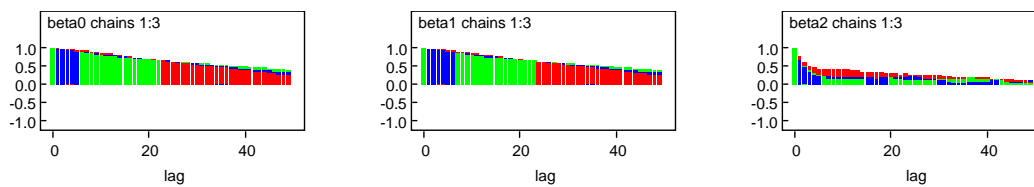


Figura 116| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser observados na Figura 117, deixam ver que R converge para 1, para cada estimativa dos parâmetros β_j e que, tanto \hat{V} como W estabilizam à medida que o número de iterações aumenta.

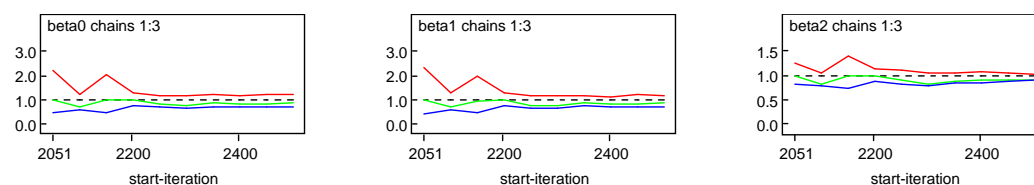


Figura 117| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

A equação no Quadro 50 dá o número esperado de mortos num período de cinco anos.

Quadro 50| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

Números Esperados de Mortos	
	$\hat{\lambda}_i = 3.355 \times 10^{-4} \times AADT_i^{0.704} \times Length_i^{1.093}$

As estimativas pontuais e respectivos intervalos de credibilidade de 95% estão apresentadas no Quadro 51.

Quadro 51| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 50.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.000	1.147	1.087×10^{-1}	-10.010	-5.648
β_1 (log(AADT))	0.704	0.101	9.502×10^{-3}	0.498	0.882
β_2 (log(Length))	1.093	0.099	6.260×10^{-3}	0.906	1.291

As densidades das distribuições *a posteriori* das estimativas dos parâmetros β_j são dadas pelos gráficos da Figura 118.

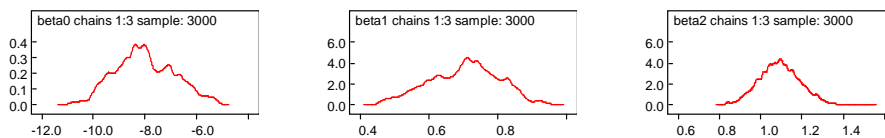


Figura 118| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a NF (obtido do conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

A média *a posteriori* e o correspondente desvio padrão para uma representação média de $\log(AADT)$ e $\log(Length)$ são apresentados no Quadro 52. Conclui-se que, para um trecho representativo de auto-estrada se esperam cerca de 2.3 mortos e feridos graves num período de cinco anos.

Quadro 52| Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 50.

média (desvio padrão)	2.364 (0.1906)
-----------------------	----------------

6.2.2.1 Validação do Modelo

Quando comparando o histograma de *NF*, no canto superior esquerdo, com os restantes histogramas na Figura 119 pode observar-se que existem alguns conjuntos de dados replicados que não se assemelham a dados observados, o que leva a crer que os dados observados não parecem ser plausíveis do modelo considerado.

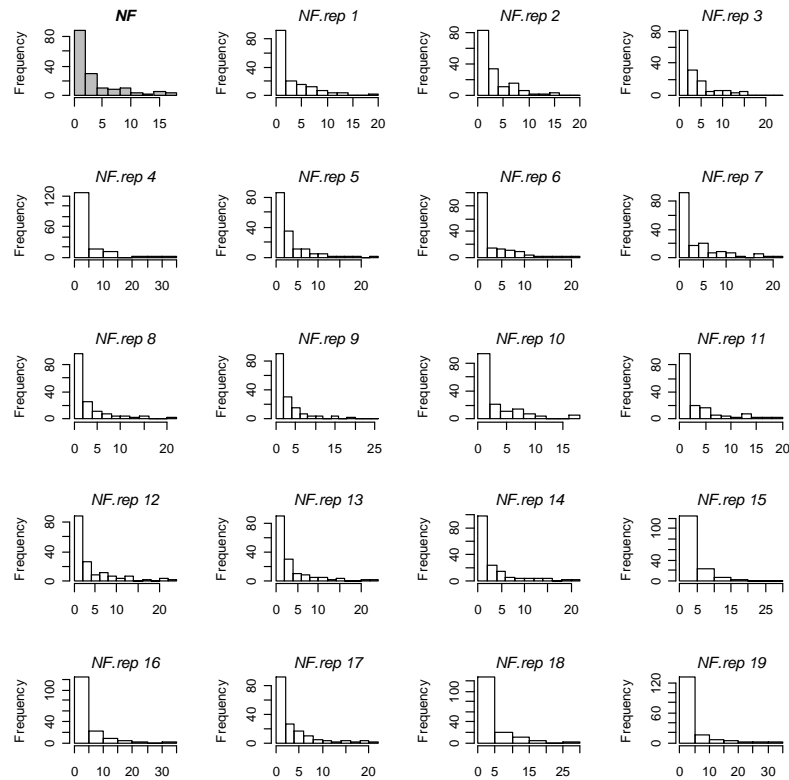


Figura 119| Histograma de *NF* (no canto superior esquerdo) e dezanove réplicas (“*NF.rep*”) dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os histogramas das medidas de discrepância exibidos na Figura 120 mostram que o modelo descrito pela equação no Quadro 50 obtém valores máximos superiores ao valor máximo obtido pelos dados observados. A probabilidade de que os valores máximos obtidos pelos dados replicados seja maior que o valor máximo observado é igual a 0.918 (p_B), valor bastante elevado.

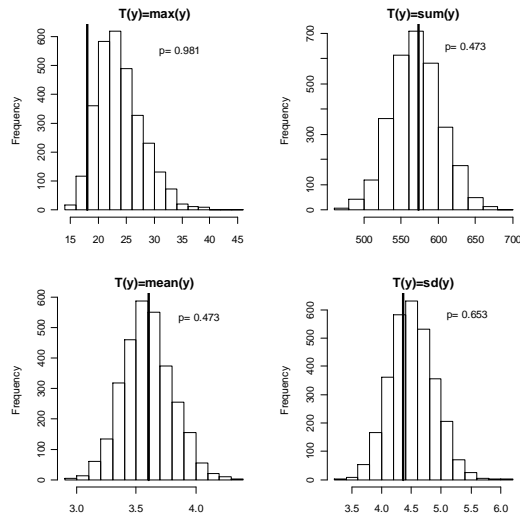


Figura 120| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo também não parece captar adequadamente a sobredispersão dos dados, como pode ser observado pelo histograma e recta vertical na Figura 121, onde $p_B=0.702$, valor superior a 0.5.

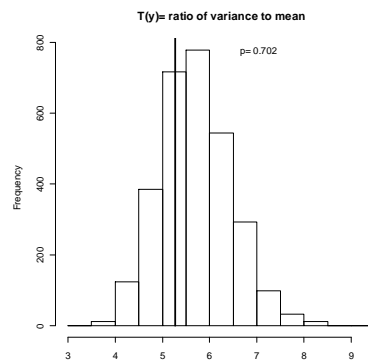


Figura 121| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo de Poisson-Gama ajustado aos dados obteve uma média das desviâncias amostrais igual a $\bar{D}=125.9$ ($n-p=157$). O valor de DIC obtido é igual a 607.3 e a estimativa para o parâmetro de dispersão é igual a 0.364.

6.2.3 Modelo de Poisson Log-Normal

O modelo hierárquico Poisson Log-Normal tendo como variável resposta o número de mortos (NF) foi ajustado aos dados e o modelo genérico dado pela equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) \quad (6.6)$$

obteve o menor valor de DIC, onde $\hat{\lambda}_i$ dá o número esperado de mortos num período de cinco anos.

As iterações das estimativas dos parâmetros β_0 e β_1 na Figura 122 sugerem não haver evidência de não convergência. No entanto pela observação dos gráficos das iterações de β_0 e β_1 nota-se que as iterações referentes às três cadeias não se interligam.

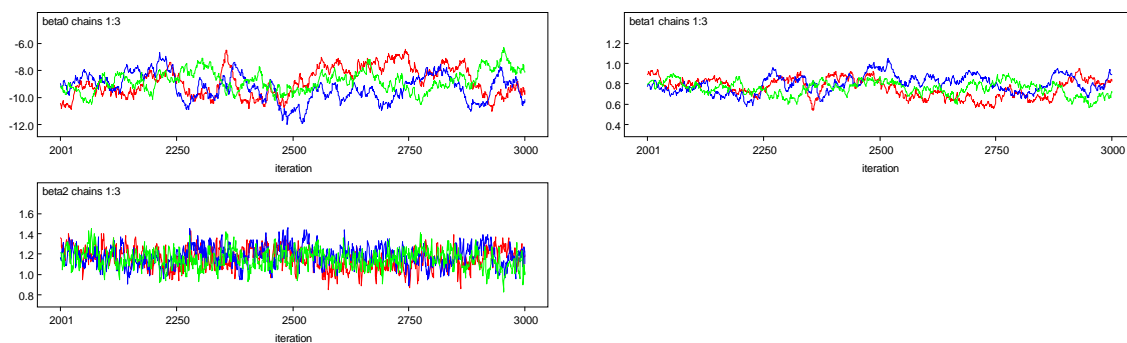


Figura 122| Trace plots (iterrações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações (na Figura 123) mostram que os parâmetros β_0 e β_1 convergem lentamente.

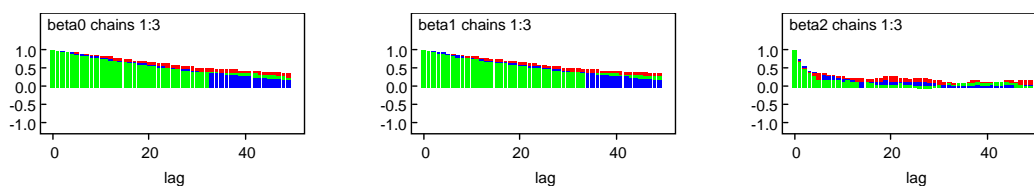


Figura 123| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das estatísticas de Gelman-Rubin na Figura 124 mostram que, embora atinja valores próximos de 1, a estatística R (a encarnado na figura) não parece convergir para 1 para os parâmetros β_0 e β_1 .

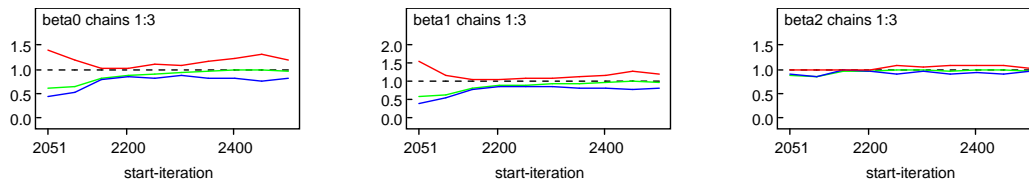


Figura 124| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

A equação que dá o número esperado de mortos, para um período de cinco anos, encontra-se no Quadro 53. O Quadro 54 contém as estimativas pontuais e intervalos de credibilidade de 95%.

Quadro 53| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos.

<i>Números Esperados de Mortos</i>	
$\hat{\lambda}_i = 1.273 \times 10^{-4} \times AADT_i^{0.770} \times Length_i^{1.162}$	

Pela observação dos gráficos das densidades *a posteriori* dos parâmetros β_j (ver Figura 125) constata-se que as médias se desviaram bastante do valor zero que foi admitido para a distribuição *a priori*.

Quadro 54| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 53.

<i>Parâmetros</i>	<i>média</i>	<i>desvio padrão</i>	<i>erros de MC</i>	<i>2.5%</i>	<i>97.5%</i>
β_0 (<i>intersecção</i>)	-8.969	0.969	9.100×10^{-2}	-10.770	-7.165
β_1 (<i>log(AADT)</i>)	0.770	0.085	7.907×10^{-3}	0.609	0.928
β_2 (<i>log(Length)</i>)	1.162	0.097	5.622×10^{-3}	0.971	1.350

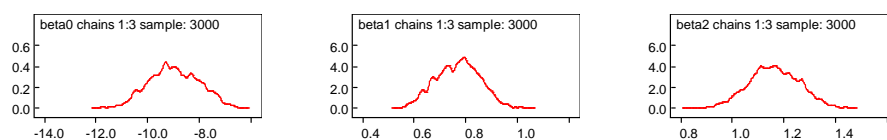


Figura 125| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a NF (obtido do conjunto de dados cujos trechos possuem valores de $AADT$ maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Esperam-se cerca de 2 mortos (em cinco anos) para um trecho com uma representação média de $AADT$ e $Length$, como pode ser observado no Quadro 55.

Quadro 55] Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 53.

<i>média (desvio padrão)</i>	1.961 (0.173)

6.2.3.1 Validação do Modelo

Os histogramas dos dados replicados são, na maioria, diferentes do histograma dos dados observados (ver Figura 126). Isto leva a crer que os dados observados não parecem ser muito bem representados pelo modelo obtido, o que põe em dúvida o ajuste deste modelo aos dados.

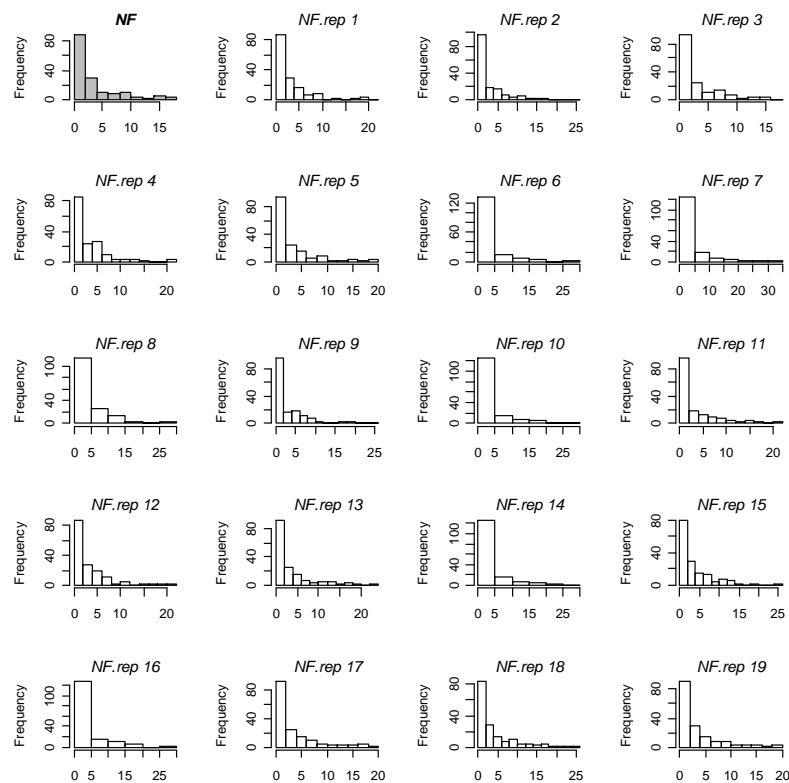


Figura 126] Histograma de *NF* (no canto superior esquerdo) e dezanove réplicas ("*NF.rep*") dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As discrepâncias entre os dados observados e os dados replicados pelo modelo dado pela equação no Quadro 53 podem ser observadas no histograma no canto superior esquerdo e no canto inferior direito na Figura 127 correspondendo a $T=max$ (valor máximo) e $T=sd$ (desvio padrão). Os valores observados correspondentes a estas quantidades não são representativos dos dados replicados.

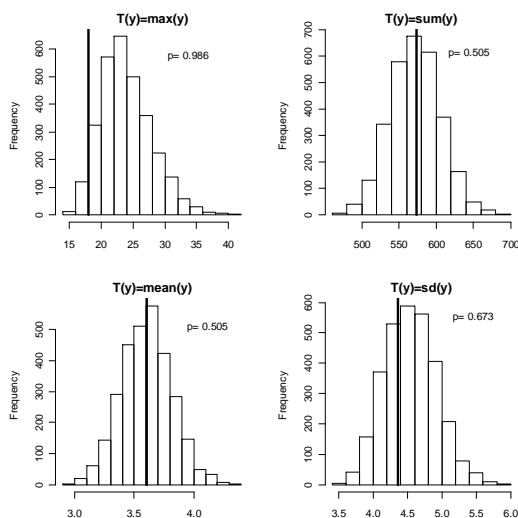


Figura 127| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A sobredispersão também não é bem modelada pois, como se pode verificar na Figura 128, o valor observado (recta vertical) não se encontra “envolvido” pelo histograma das razões entre a variância e a média obtidos pelos dados da distribuição *a posteriori* preditiva.

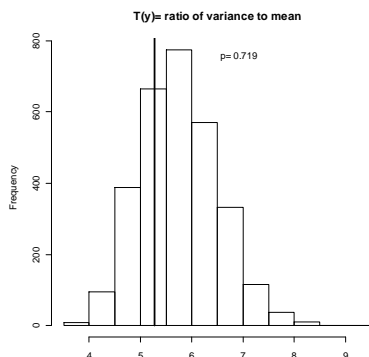


Figura 128| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson Log-Normal tendo o número de mortos como variável resposta obtém uma média das desviâncias amostrais, $\bar{D}=127$, o qual é razoavelmente inferior a $n-p=157$. O valor de DIC obtido é 612.49 e a estimativa para o parâmetro de dispersão é igual a 0.346.

6.3 Modelação do Número de Feridos Graves

A descrição dos dados empregues nesta secção, nomeadamente de *NSI* (número de feridos graves) assim também como gráficos descritivos encontram-se na secção 6.3 de Azeredo Lopes e Cardoso (2009).

6.3.1 Modelo de Poisson

O modelo de Poisson ajustado aos dados obteve um número esperado de feridos graves, $\hat{\lambda}_i$, dado pela seguinte equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i + \beta_4 Lane_Width_i \quad (6.7)$$

O conjunto de covariáveis obtido produziu o menor valor de DIC. Os gráficos de um conjunto de iterações das estimativas dos parâmetros β_j na Figura 129 levam a crer não haver razões para duvidar da não convergência dos parâmetros.

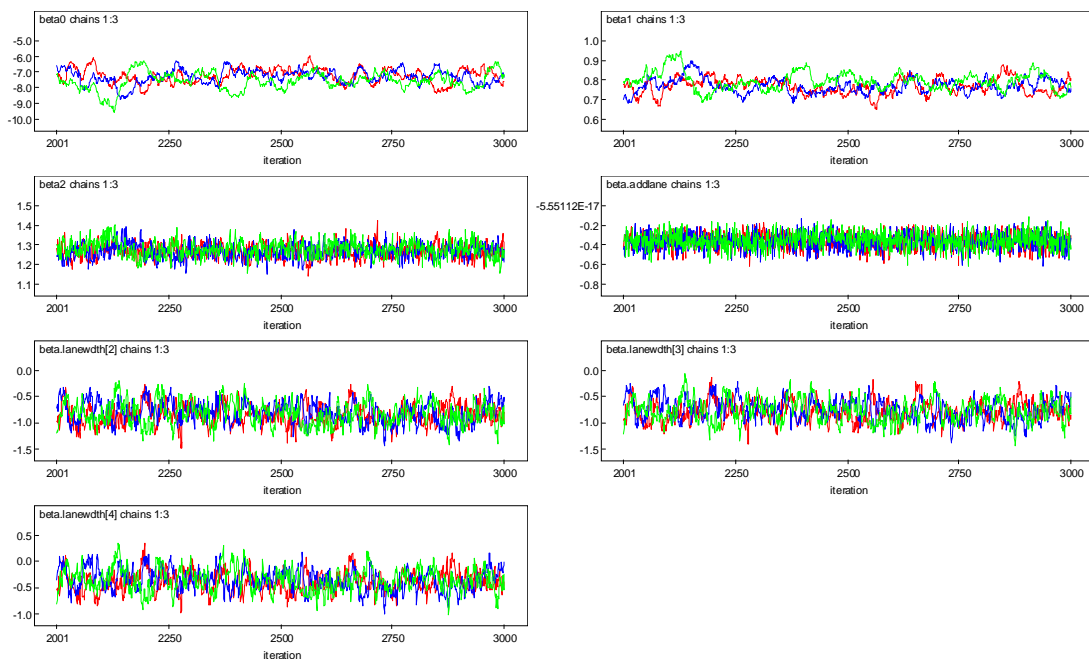


Figura 129] *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações na Figura 130 deixam ver que as estimativas de β_0 e β_1 convergem mais lentamente do que as dos restantes parâmetros.

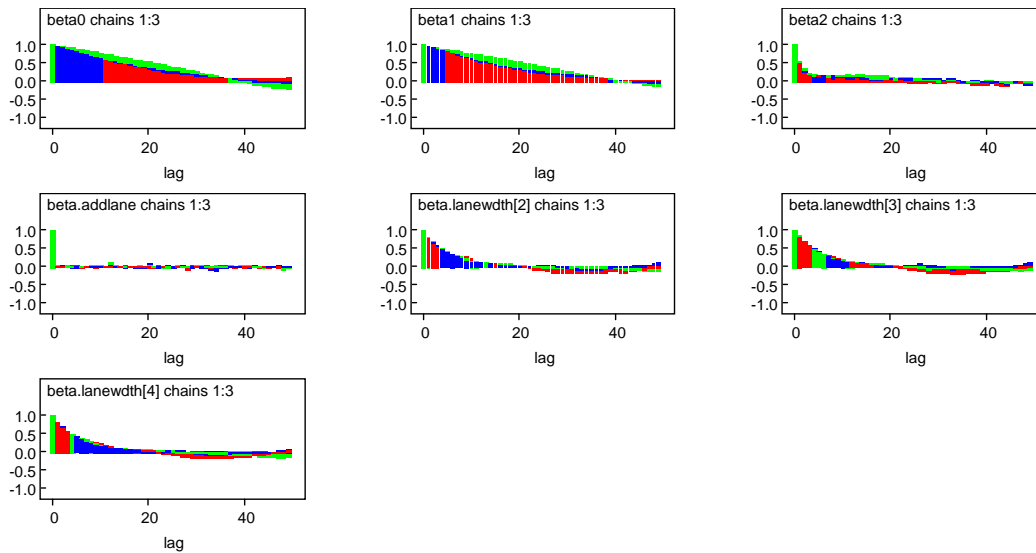


Figura 130| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das estatísticas de Gelman-Rubin (na Figura 131) mostram que todas as estatísticas estabilizam à medida que aumenta o número de iterações e que *R* converge para 1.

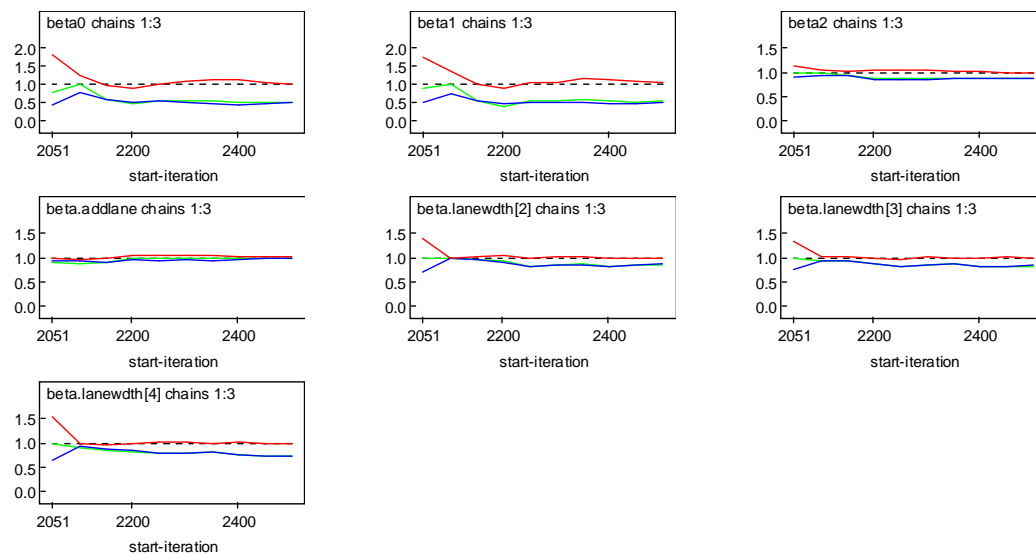


Figura 131| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As equações dos números esperados de feridos graves, para os diversos níveis das variáveis categóricas *Additional_Lane* e *Lane_Width*, estão exibidas no Quadro 56.

Quadro 56| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

		Números Esperados de Feridos Graves
Additional_Lane='No'		
Lane_Width		
'A'		$\hat{\lambda}_i = 6.209 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'B'		$\hat{\lambda}_i = 2.712 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'C'		$\hat{\lambda}_i = 2.875 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'D'		$\hat{\lambda}_i = 4.301 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
Additional_Lane='Yes'		
Lane_Width		
'A'		$\hat{\lambda}_i = 4.315 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'B'		$\hat{\lambda}_i = 1.884 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'C'		$\hat{\lambda}_i = 1.998 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$
'D'		$\hat{\lambda}_i = 2.989 \times 10^{-4} \times AADT_i^{0.777} \times Length_i^{1.270}$

O Quadro 57 contém as estimativas pontuais e os intervalos de credibilidade de 95% para os parâmetros β_j na equação (6.7). Pela leitura dos valores das estimativas depreende-se que um trecho de auto-estrada com via adicional tem um número de feridos cerca de 31% inferior ao de um trecho em idênticas condições, mas sem via adicional. Quanto à largura da via, constata-se que um trecho com largura de via tipo 'B' (3.00, 3.50] tem um número de feridos menor em aproximadamente 56% do que um trecho em iguais condições mas com largura de via tipo 'A' ([0, 3.00]). Trechos com largura de vias dos tipos 'C' (3.50, 3.75] e 'D' (3.75, ∞) têm um menor número de feridos, cerca de 54% e 31%, respectivamente, em relação a trechos com largura de via de tipo 'A'.

Quadro 57| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 56.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.384	0.505	4.364×10^{-2}	-8.515	-6.488
β_1 (log(AADT))	0.777	0.042	3.635×10^{-3}	0.701	0.867
β_2 (log(Length))	1.270	0.038	1.724×10^{-3}	1.195	1.346
β_3 (Additional_Lane='Yes')	-0.364	0.078	4.015×10^{-3}	-0.522	-0.210
β_4 (Lane_Width='B')	-0.828	0.208	1.852×10^{-3}	-1.215	-0.417
β_5 (Lane_Width='C')	-0.770	0.208	1.210×10^{-2}	-1.155	-0.344
β_6 (Lane_Width='D')	-0.367	0.211	1.219×10^{-2}	-0.762	0.066

Os gráficos das densidades *a posteriori* das estimativas dos parâmetros β_j encontram-se na Figura 132. Em todos eles se observa um afastamento de zero dos valores médios.

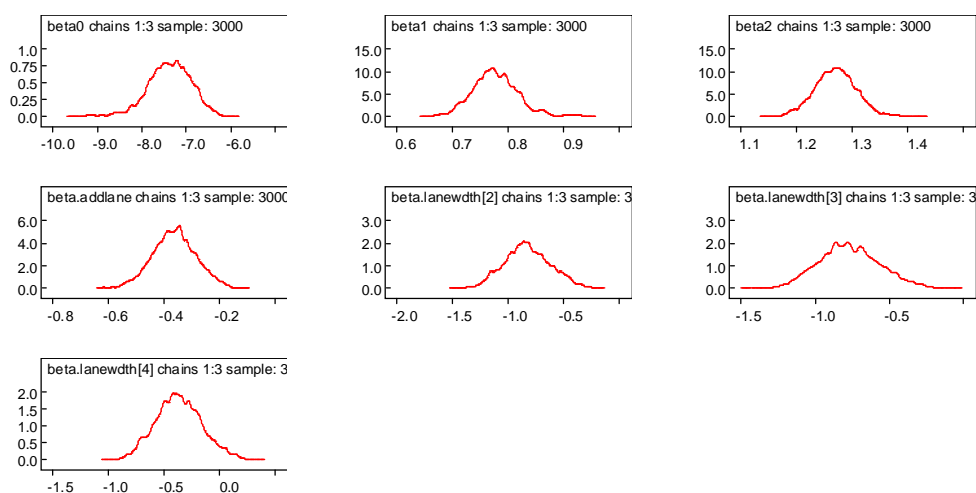


Figura 132| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Os valores médios *a posteriori*, e correspondentes desvios padrões, para uma representação média de *AADT* e *Length* estão exibidos no Quadro 58. Pode depreender-se que, para um trecho típico de auto-estrada sem via adicional e com largura de via do tipo 'A' ([0, 3.00]), se esperam 12.310 feridos graves, num período de cinco anos.

Quadro 58| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 56.

<i>Additional_Lane</i> ='No'	média (desvio padrão)	<i>Additional_Lane</i> ='Yes'	média (desvio padrão)
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	12.310 (2.418)	'A'	8.578 (1.803)
'B'	5.286 (0.377)	'B'	3.683 (0.373)
'C'	5.599 (0.302)	'C'	3.898 (0.312)
'D'	8.376 (0.474)	'D'	5.837 (0.555)

O menor número de feridos graves (3.683) é esperado em trechos com via adicional e com largura de via do tipo 'B', ou seja, entre 3 e 3.50 metros.

6.3.1.1 Validação do Modelo

Dezanove histogramas obtidos da distribuição *a posteriori* preditiva estão expostas na Figura 133. Alguns destes histogramas, nomeadamente *NSI.rep 3, 8 e 18*, não se assemelham ao histograma dos dados observados (no canto superior esquerdo) o que leva a crer a existência de falhas no modelo considerado.

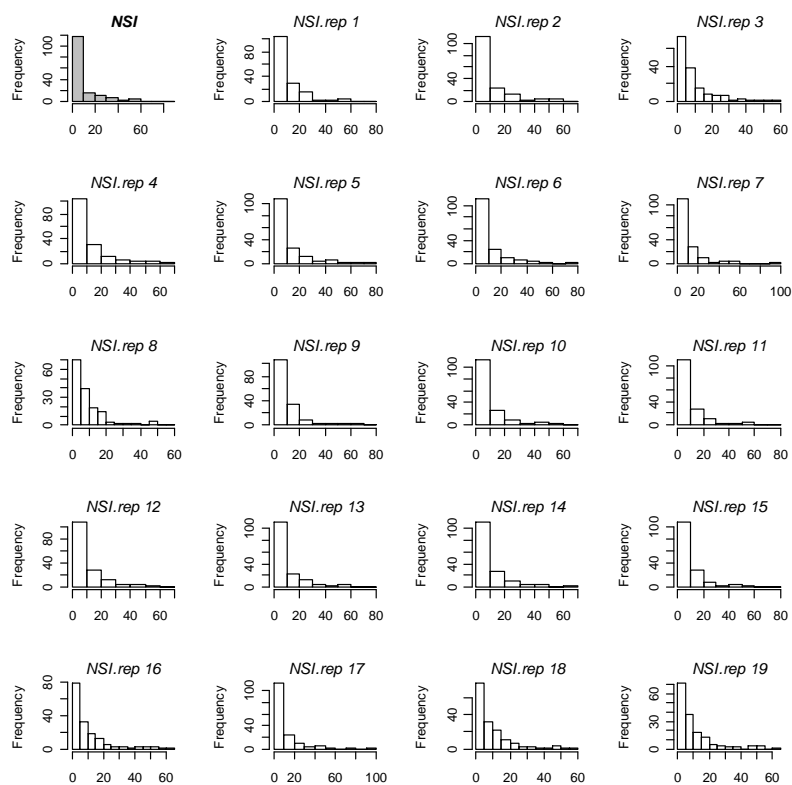


Figura 133| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas (“*NSI.rep*”) dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As medidas empregues para avaliar a discrepância entre os dados replicados pelo modelo e os dados observados mostram que o modelo não captura o valor máximo ($T=max$) e desvio padrão ($T=sd$), como se pode ver na Figura 134. O modelo também não captura a dispersão existente nos dados observados como se pode observar na Figura 135.

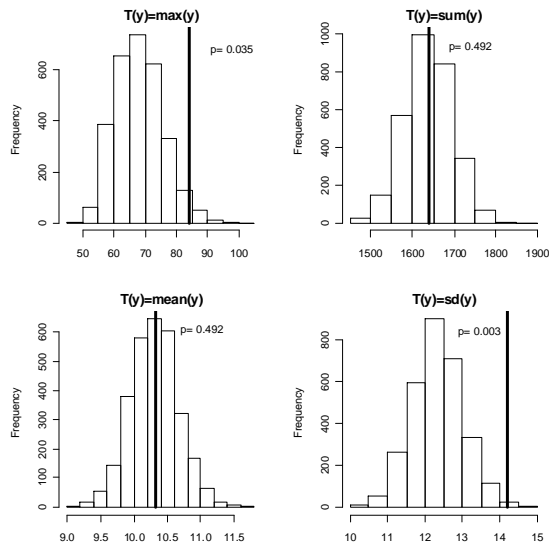


Figura 134| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O p -value Bayesiano aplicado à razão entre a variância e a média (ver Figura 135), isto é, a probabilidade de que a mesma razão obtida pelos dados replicados possa ser mais extrema que a obtida pelos dados observados, é igual a $p_B=0.001$, valor bastante baixo o que faz duvidar do modelo de Poisson ajustado aos dados referentes aos números de feridos graves.

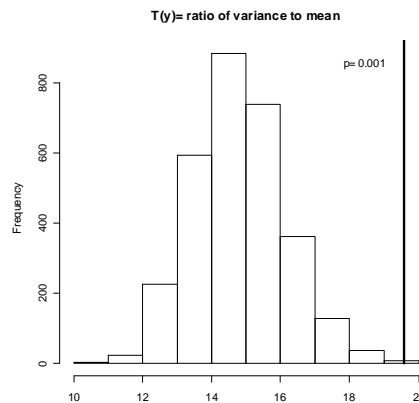


Figura 135| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo de Poisson obteve uma média das desviâncias amostrais $\bar{D}=640$, valor consideravelmente elevado quando comparado com $n-p=155$, indicando que o modelo não captura a sobredispersão existente nos dados. O valor de DIC obtido é igual a 1228.45.

6.3.2 Modelo de Poisson-Gama

O modelo hierárquico Poisson-Gama que obteve o menor valor de DIC quando ajustado aos dados referentes ao número de feridos graves inclui como covariáveis $\log(AADT)$, $\log(Length)$ e $Additional_Lane$. O número esperado de feridos graves, $\hat{\lambda}_i$, é dado pela seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i, \quad (6.8)$$

Os gráficos das iterações das estimativas dos parâmetros β_j relativos às três cadeias de Markov, representados na Figura 136, mostram que as distribuições *a posteriori* para os parâmetros β_0 e β_1 não parecem convergir.

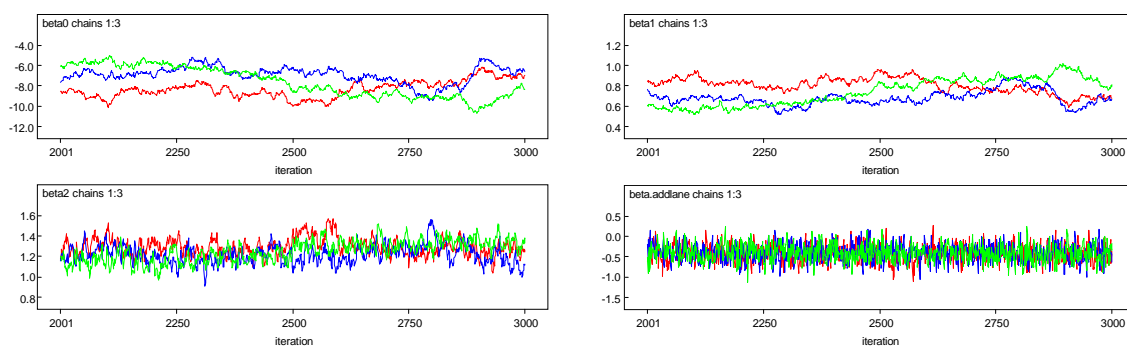


Figura 136| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações exibidas na Figura 137 mostram autocorrelações muito elevadas para β_0 e β_1 .

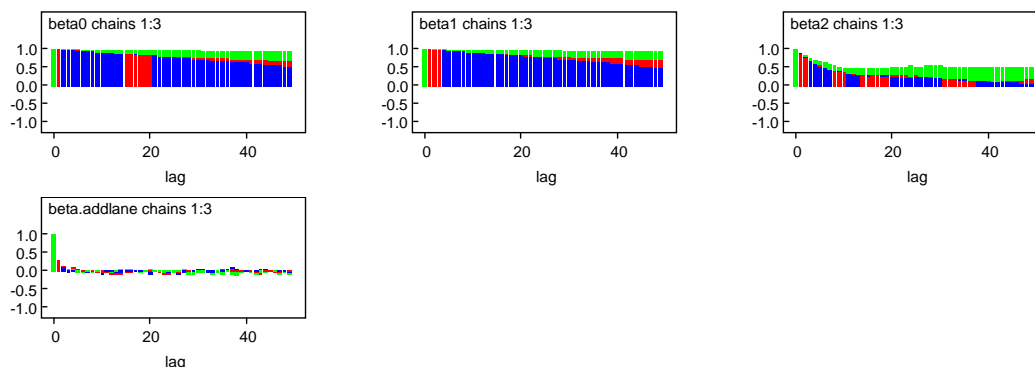


Figura 137| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser examinados na Figura 138, mostram que R (a vermelho) tende para 1 para todos os parâmetros β_j e que \hat{V} e W estabilizam à medida que aumentam as iterações, o que leva a crer que a convergência das cadeias de Markov é atingida.

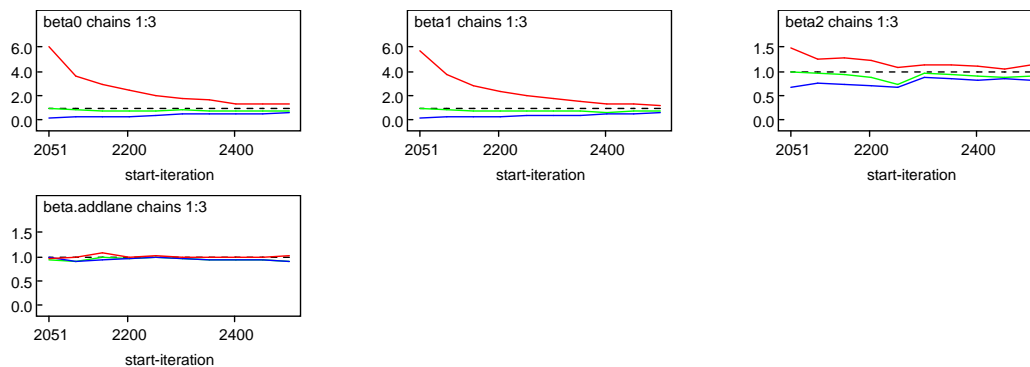


Figura 138| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os números esperados de feridos graves, para um período de cinco anos, após ajuste do modelo Poisson-Gama, são dados pelas equações do Quadro 59 referentes a trechos sem e com via adicional.

Quadro 59| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Feridos Graves
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 4.891 \times 10^{-4} \times AADT_i^{0.741} \times Length_i^{1.252}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 3.237 \times 10^{-4} \times AADT_i^{0.741} \times Length_i^{1.252}$

As estimativas obtidas pelo modelo estão expostas no Quadro 60. Pela estimativa do parâmetro β_3 (relativo à covariável *Additional_Lane*) pode afirmar-se que a presença de uma via adicional reduz o número de feridos graves em cerca de 34%.

Quadro 60| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 59.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.623	1.237	1.251×10^{-1}	-9.858	-5.515
β_1 (log(<i>AADT</i>))	0.741	0.110	1.114×10^{-2}	0.554	0.941
β_2 (log(<i>Length</i>))	1.252	0.098	7.550×10^{-3}	1.065	1.453
β_3 (<i>Additional_Lane</i> ='Yes')	-0.413	0.207	4.929×10^{-3}	-0.819	3.040×10^{-3}

As densidades *a posteriori* dos parâmetros β_j , ver Figura 139, mostram que as respectivas médias se afastaram de zero. No entanto verifica-se a existência de vários “picos” nas densidades de β_0 e β_1 .

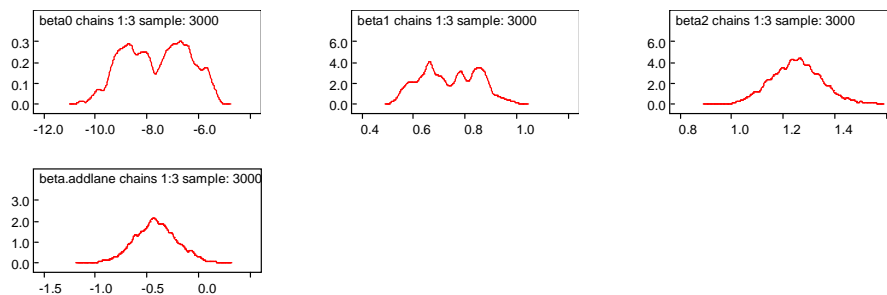


Figura 139| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a *NSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Para um trecho com uma representação média de *AADT* e *Length* esperam-se, em cinco anos, 6.5 e 4.3 feridos graves em trechos sem e com via adicional, respectivamente, como pode ser observado no Quadro 61.

Quadro 61| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 59.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	6.458 (0.495)
<i>Additional_Lane='Yes'</i>	4.344 (0.863)

6.3.2.1 Validação do Modelo

Pela comparação dos histogramas na Figura 140 constata-se que os dados observados são plausíveis sob a distribuição preditiva *a posteriori*.

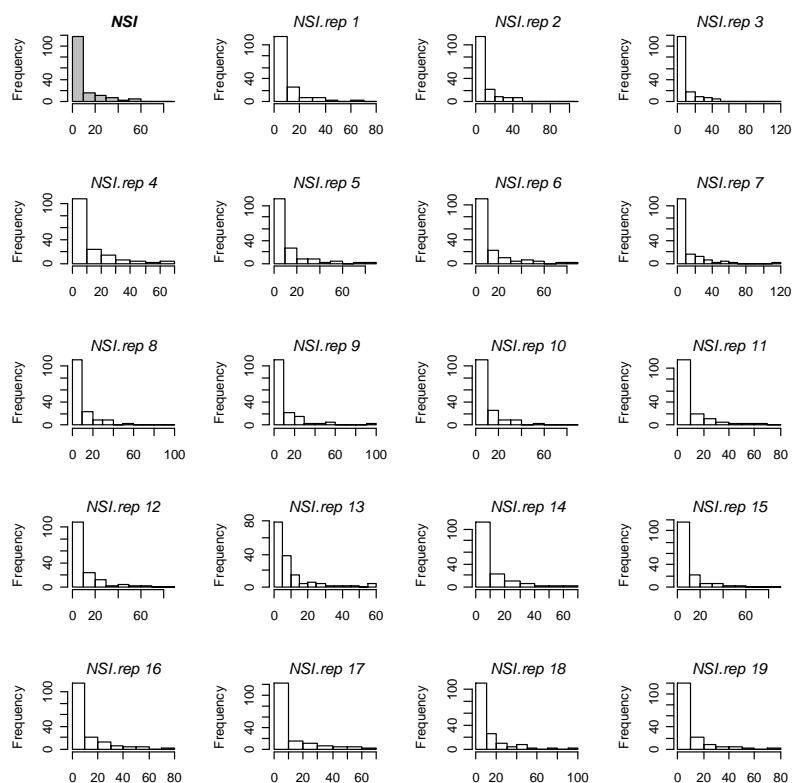


Figura 140| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas (“*NSI.rep*”) dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

A verificação das comparações gráficas das quantidades correspondentes aos dados com as quantidades obtidas de simulações preditivas *a posteriori* (ver Figura 141) leva a crer que o modelo ajustado é adequado.

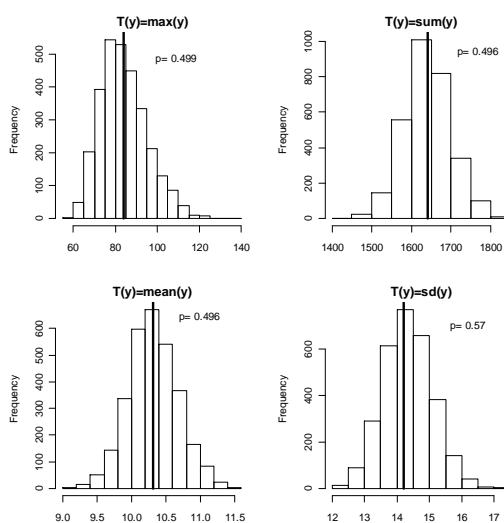


Figura 141| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

A mesma conclusão é obtida pela leitura da Figura 142, que sugere que o modelo ajustado tem em consideração a dispersão existente nos dados observados.

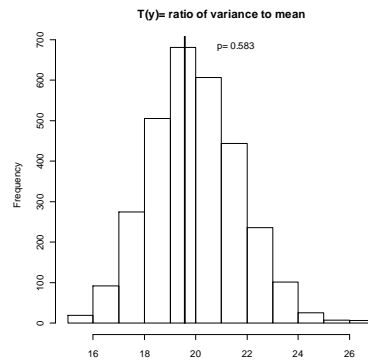


Figura 142| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a posteriori preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O valor da média das desviâncias amostrais obtido pelo modelo é igual a $\bar{D}=140.6$, valor mais próximo de $n-p=156$ do que o do modelo de Poisson, indicando um bom ajuste. O valor de DIC é 782.22 e a estimativa do parâmetro de dispersão é igual a 0.550.

6.3.3 Modelo de Poisson Log-Normal

O modelo hierárquico Poisson Log-Normal que possui o menor valor de DIC após o ajuste aos dados referentes aos números de feridos graves é dado pela equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (6.9)$$

Onde $\hat{\lambda}_i$ representa o número esperado de feridos graves.

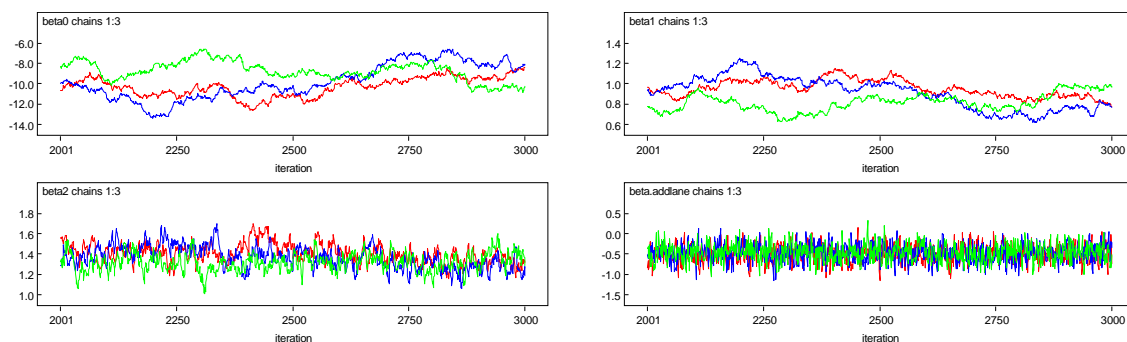


Figura 143| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NSI obtido do conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

O facto de as iterações das estimativas de β_0 e β_1 para as três cadeias de Markov não se interligarem, nem tão pouco parecerem convergir para um valor comum (ver os dois gráficos na primeira linha na Figura 143) leva a acreditar que as estimativas de β_0 e β_1 não parecerem convergir. A mesma interpretação é deduzida pela leitura dos gráficos de autocorrelação exibidos na Figura 144.

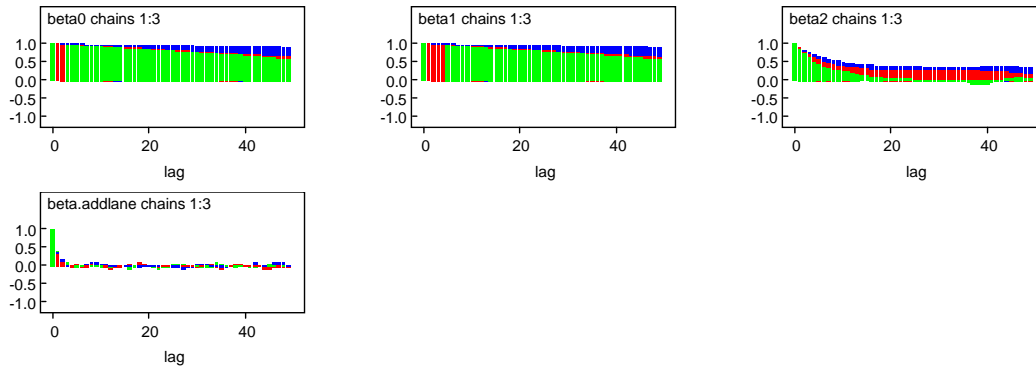


Figura 144| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Pela observação dos gráficos das estatísticas de Gelman-Rubin apresentadas na Figura 145 depreende-se que, uma vez que R tende para 1 e as restantes estatísticas estabilizam à medida que aumenta o número de iterações, há razões para admitir a convergência das estimativas de todos os parâmetros β_j , e por consequência do algoritmo considerado.

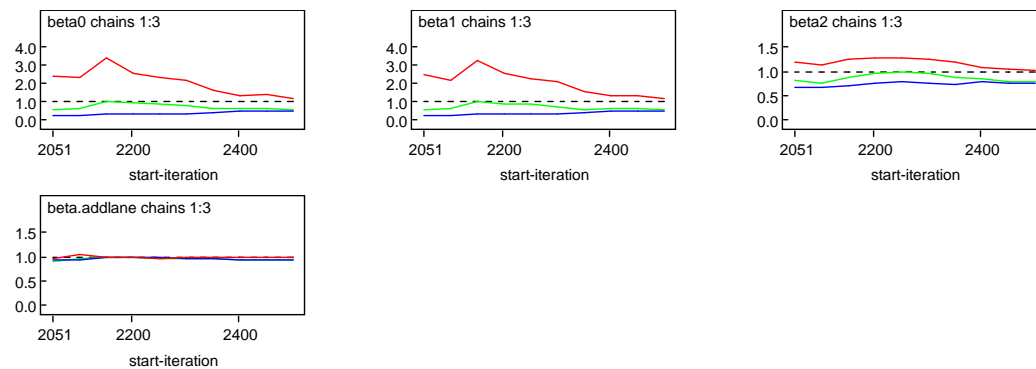


Figura 145| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

No Quadro 62 podem ver-se as equações para estimar os números esperados de feridos graves (em cinco anos) para trechos de auto-estrada com e sem via adicional para o modelo Poisson Log-Normal que apresentou o menor valor de DIC.

Quadro 62| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

	<i>Números Esperados de Feridos Graves</i>
<i>Additional_Lane='No'</i>	$\hat{\lambda}_i = 6.631 \times 10^{-5} \times AADT_i^{0.893} \times Length_i^{1.362}$
<i>Additional_Lane='Yes'</i>	$\hat{\lambda}_i = 4.097 \times 10^{-5} \times AADT_i^{0.893} \times Length_i^{1.362}$

As estimativas obtidas por este modelo podem ser observadas no Quadro 63. A leitura da média *a posteriori* da estimativa de β_3 mostra que a existência de uma via adicional reduz o número esperado de feridos graves em cerca de 38% quando comparado a um trecho em iguais condições, mas sem via adicional.

Quadro 63| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 62.

<i>Parâmetros</i>	<i>média</i>	<i>desvio padrão</i>	<i>erros de MC</i>	<i>2.5%</i>	<i>97.5%</i>
β_0 (<i>intersecção</i>)	-9.621	1.420	1.445×10^{-1}	-12.450	-7.085
β_1 (<i>log(AADT)</i>)	0.893	0.127	1.293×10^{-2}	0.660	1.134
β_2 (<i>log(Length)</i>)	1.362	0.105	8.026×10^{-3}	1.171	1.577
β_3 (<i>Additional_Lane='Yes'</i>)	-0.482	0.213	5.886×10^{-3}	-0.902	5.836×10^{-2}

Os gráficos das densidades *a posteriori* obtidas para as estimativas dos parâmetros β_j são exibidos na Figura 146. As densidades para β_0 e β_1 possuem inúmeros “picos”, o que mostra que estas densidades não foram obtidas com muita precisão.

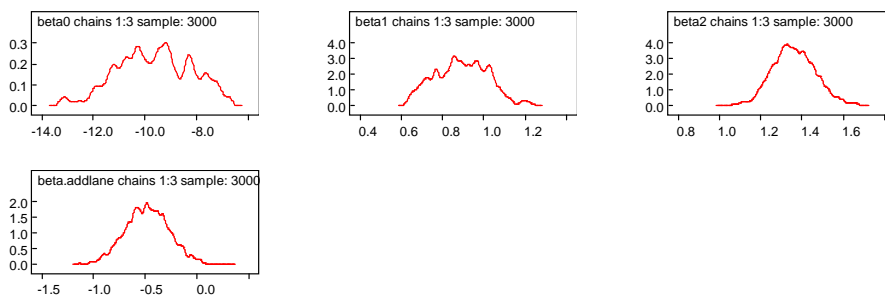


Figura 146| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *NSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Para trechos de auto-estrada com uma representação média de *AADT* e *Length* e com ou sem via adicional são esperados, em cinco anos, respectivamente, 3 ou 5 feridos graves (ver Quadro 64). Valores estes mais elevados dos que obtidos pelo modelo Poisson-Gama (ver valores do Quadro 61).

Quadro 64| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 62.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	4.884 (0.416)
<i>Additional_Lane='Yes'</i>	3.071 (0.637)

6.3.3.1 Validação do Modelo

O modelo Poisson Log-Normal considerado é consistente com os dados, visto não haver diferenças substanciais entre os dados observados e os dados simulados como se pode deduzir pela comparação dos histogramas na Figura 147.

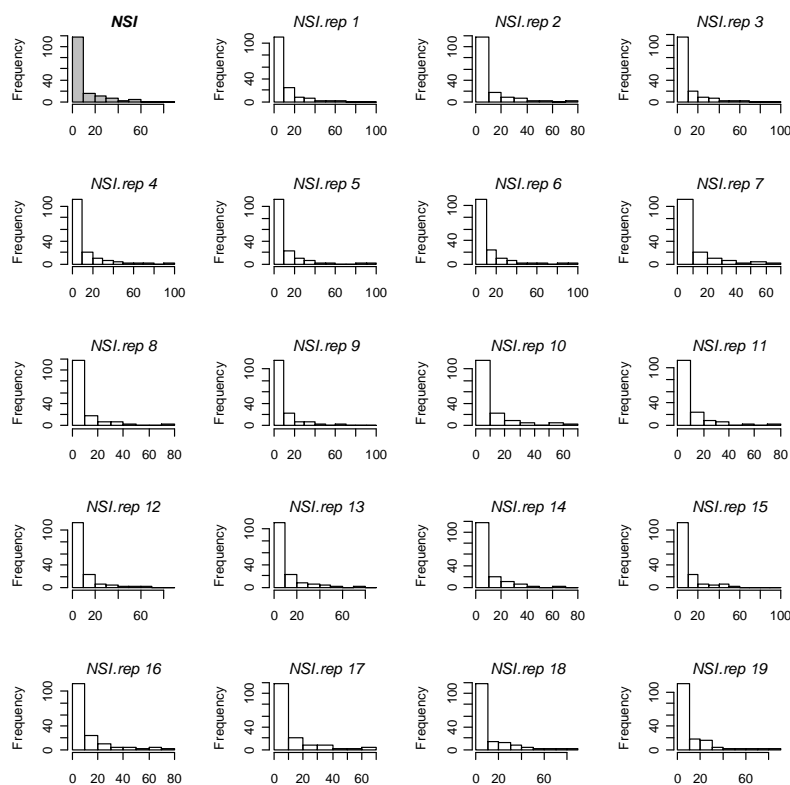


Figura 147| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas ("*NSI.rep*") dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As comparações gráficas das quantidades *T* correspondentes aos dados (rectas verticais na Figura 148) com as quantidades *T* extraídas das simulações preditivas *a posteriori* (histogramas), levam a crer que o modelo obtido é adequado, uma vez que os valores observados (rectas verticais) são representativos dos valores simulados (histogramas).

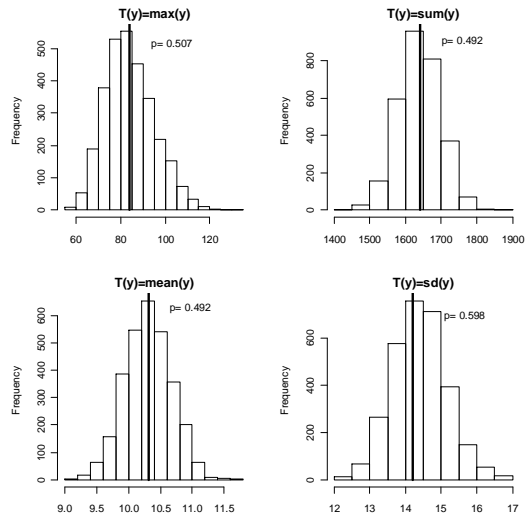


Figura 148| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Idêntica conclusão se retira da leitura da Figura 149 no que refere a $T = \text{razão entre a variância e o valor médio}$. Neste caso depreende-se que o modelo ajustado (Poisson Log-Normal) atende à dispersão dos dados que foram observados.

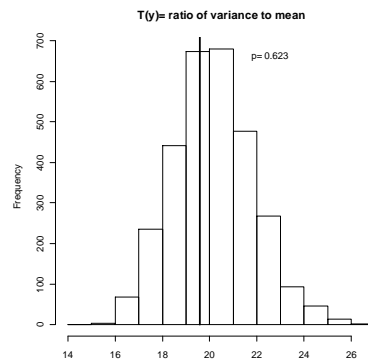


Figura 149| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson Log-Normal obtido produziu um valor $\bar{D} = 143.1$ ($n-p=156$), $\text{DIC} = 792.3$ e uma estimativa do parâmetro de dispersão igual a 0.545.

6.4 Modelação do Número de Mortos e Feridos Graves

A descrição dos dados analisados nesta secção encontra-se no Capítulo 6, secção 6.4 em Azeredo Lopes e Cardoso (2009).

6.4.1 Modelo de Poisson

O modelo que obteve o menor valor de DIC, após uma selecção de variáveis, e de um ajuste do modelo Poisson ao número de mortos e feridos graves tem a seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (6.10)$$

Pela observação dos gráficos na Figura 150 não parece haver razões para duvidar da convergência dos parâmetros β_j correspondentes à equação (6.10), pois as iterações de cada cadeia de Markov interligam-se e convergem para o mesmo valor.

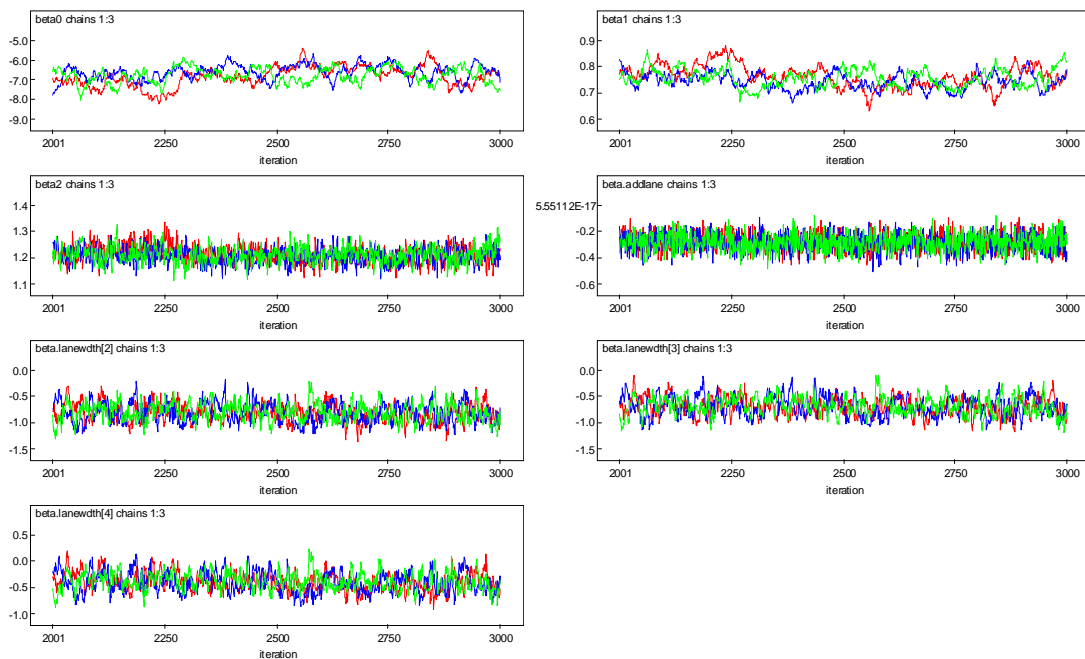


Figura 150| Trace plots (iterrações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações na Figura 151 também não sugerem razões para duvidar da convergência, embora β_0 e β_1 convirjam mais lentamente do que os restantes parâmetros.

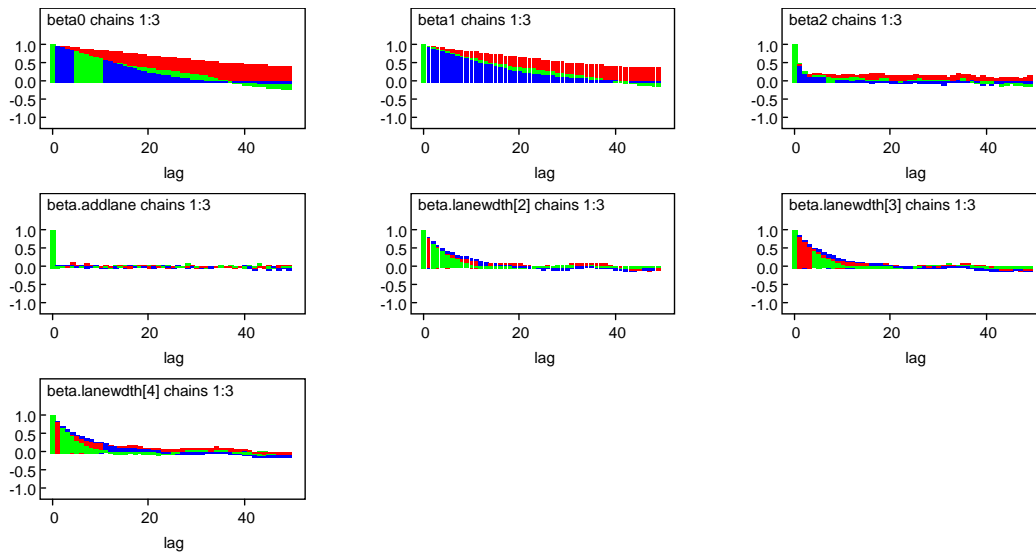


Figura 151| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin exibidas na Figura 152 estabilizam e, em especial R , converge para 1. Não há, assim, razões para duvidar da convergência do algoritmo.

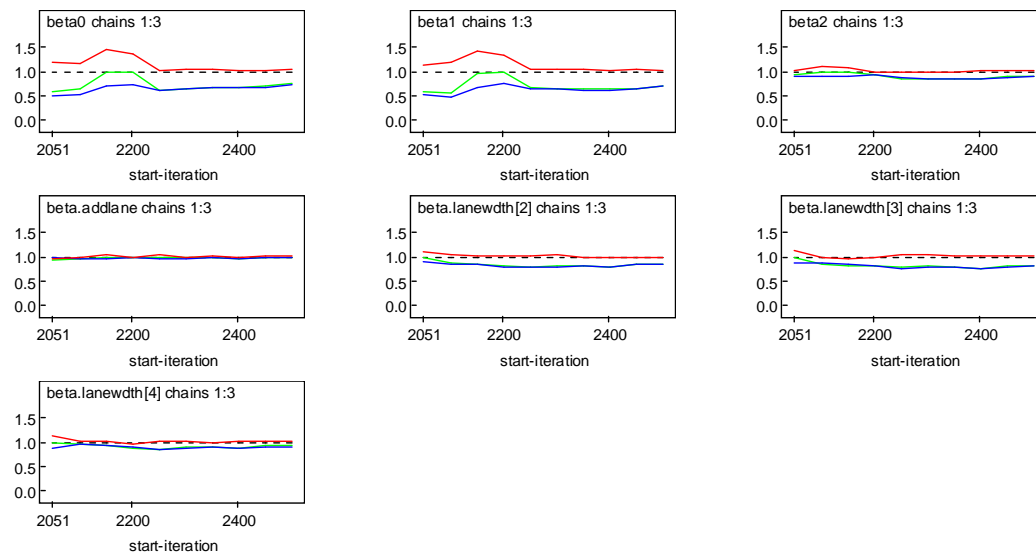


Figura 152| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As equações dos números esperados de mortos e feridos graves são exibidas no Quadro 65 para cada nível das variáveis categóricas.

Quadro 65| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

		Números Esperados de Mortos e Feridos Graves
<i>Additional_Lane='No'</i>		
<i>Lane_Width</i>		
'A'		$\hat{\lambda}_i = 1.161 \times 10^{-3} \times AADT_i^{0.754} \times Length_i^{1.210}$
'B'		$\hat{\lambda}_i = 5.135 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
'C'		$\hat{\lambda}_i = 5.784 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
'D'		$\hat{\lambda}_i = 7.795 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
<i>Additional_Lane='Yes'</i>		
<i>Lane_Width</i>		
'A'		$\hat{\lambda}_i = 8.741 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
'B'		$\hat{\lambda}_i = 3.866 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
'C'		$\hat{\lambda}_i = 4.355 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$
'D'		$\hat{\lambda}_i = 5.868 \times 10^{-4} \times AADT_i^{0.754} \times Length_i^{1.210}$

As estimativas do modelo obtido encontram-se no Quadro 66. O modelo ajustado permite afirmar que a existência de uma via adicional num trecho de auto-estrada reduz o número esperado de mortos e feridos graves em cerca de 25% quando comparado com um trecho sem via adicional e em idênticas situações ao anterior.

Quadro 66| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 65.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-6.758	0.435	3.907×10^{-2}	-7.708	-5.968
β_1 (log(AADT))	0.754	0.036	3.223×10^{-3}	0.689	0.834
β_2 (log(Length))	1.210	0.032	1.460×10^{-3}	1.148	1.272
β_3 (Additional_Lane='Yes')	-0.284	0.066	1.583×10^{-3}	-0.414	-0.156
β_4 (Lane_Width='B')	-0.816	0.179	9.935×10^{-3}	-1.157	-0.458
β_5 (Lane_Width='C')	-0.697	0.180	1.010×10^{-2}	-1.030	-0.342
β_6 (Lane_Width='D')	-0.399	0.182	1.074×10^{-2}	-0.746	-0.035

Trechos com largura de via de tipo 'B' ([3.00, 3.50]), 'C' ([3.50, 3.75]) e 'D' ([3.75, ∞]) têm uma redução de aproximadamente, respectivamente, 56%, 50% e 33%, quando comparadas com trechos com largura de via de tipo 'A' ([0.00, 3.00]).

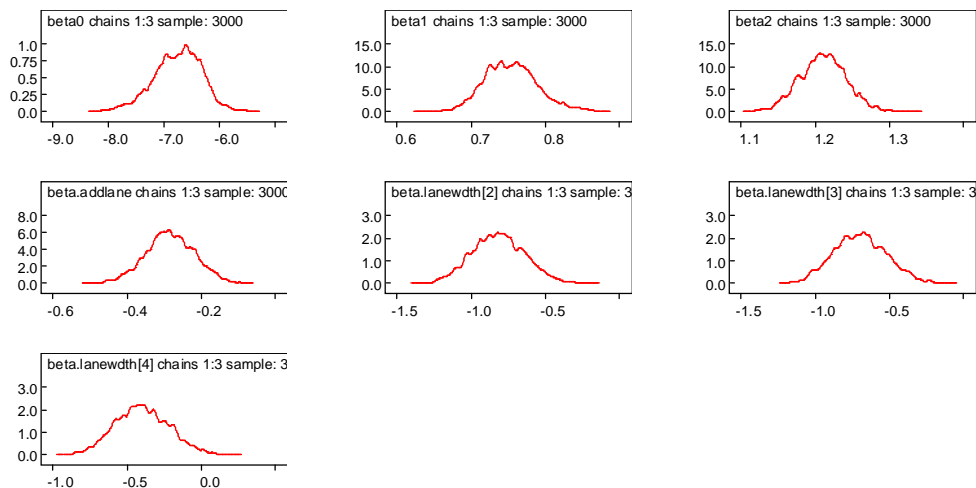


Figura 153| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NKSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

As densidades *a posteriori* dos parâmetros β_j estão representadas na Figura 153. Os valores médios de β_j afastaram-se consideravelmente de zero.

Quadro 67| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 65.

<i>Additional_Lane='No'</i>	<i>média (desvio padrão)</i>	<i>Additional_Lane='Yes'</i>	<i>média (desvio padrão)</i>
<i>Lane_Width</i>		<i>Lane_Width</i>	
'A'	16.570 (2.782)	'A'	12.500 (2.271)
'B'	7.236 (0.441)	'B'	5.457 (0.465)
'C'	8.145 (0.376)	'C'	6.139 (0.407)
'D'	10.980 (0.545)	'D'	8.282 (0.679)

Os números esperados de mortos e feridos graves para uma representação média de *AADT* e *Length*, e para um período de cinco anos, estão apresentados no Quadro 67 para os diversos níveis das covariáveis. Depreende-se que os valores mais elevados de mortos e feridos graves ocorrem em trechos sem via adicional e para larguras de via do tipo 'A' e 'D', isto é, entre 0 e 3 metros e maior que 3.75 metros, respectivamente.

6.4.1.1 Validação do Modelo

A comparação dos histogramas obtidos de dados replicados da distribuição preditiva *a posteriori* com o histograma dos dados observados (ver Figura 154) mostra haver algumas discrepâncias no modelo.

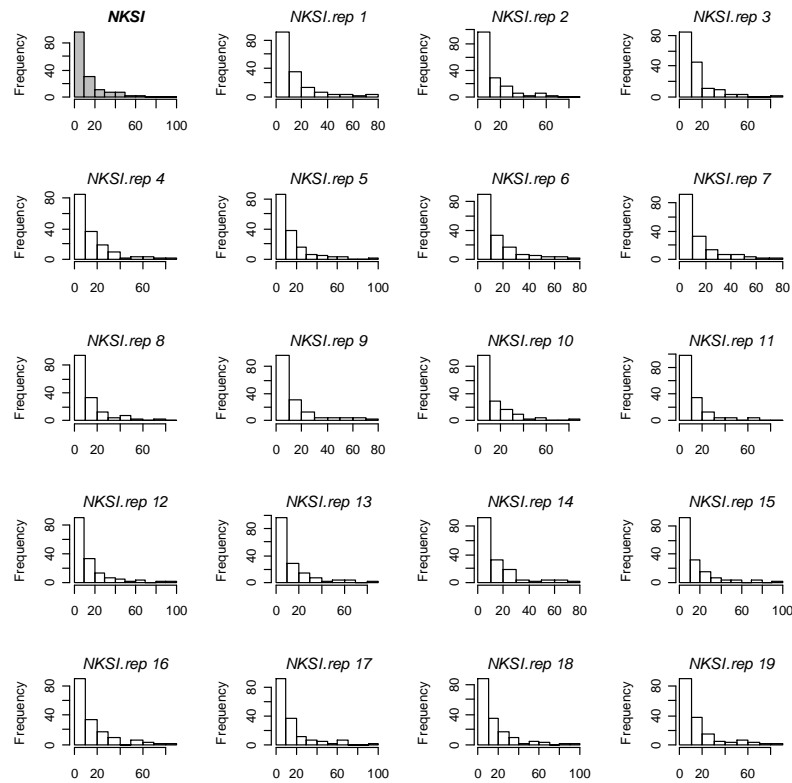


Figura 154| Histograma de *NKSJ* (no canto superior esquerdo) e dezanove réplicas (“*NKSJ.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As quantidades *T* calculadas servem para medir as possíveis discrepâncias entre dados obtidos pelo modelo ajustado e os dados observados.

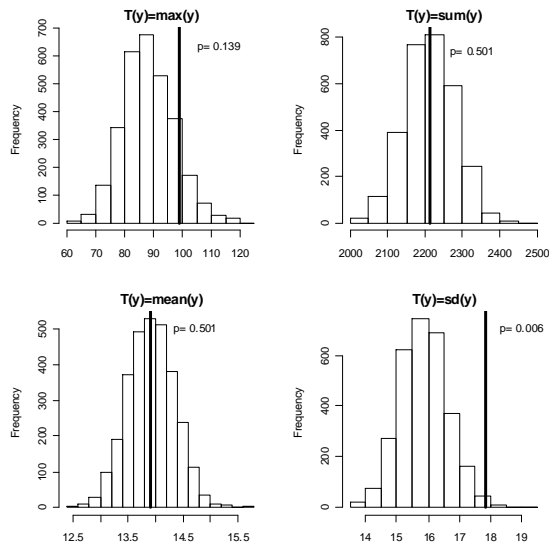


Figura 155| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Pela leitura da Figura 155 verifica-se que o modelo considerado não reproduz convenientemente o valor máximo e o desvio padrão do conjunto formado pelos dados. O modelo também não capta a dispersão dos dados como pode ser observado na Figura 156.

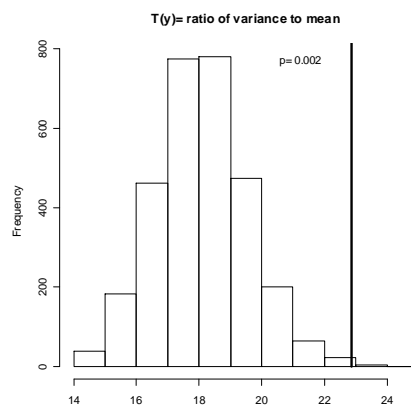


Figura 156| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo de Poisson representado pelas equações no Quadro 65 produz $\bar{D} = 750.3$ (valor bastante elevado quando comparado a $n-p=155$, indicando que o modelo de Poisson não capta a sobredispersão existente nos dados) e DIC igual a 1388.82.

6.4.2 Modelo de Poisson-Gama

O modelo Poisson-Gama cujo ajuste aos dados obteve o menor valor de DIC tem a seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Additional_Lane_i \quad (6.11)$$

Os gráficos das iterações na Figura 157 mostram haver razões para duvidar da convergência das estimativas de β_0 e β_1 .

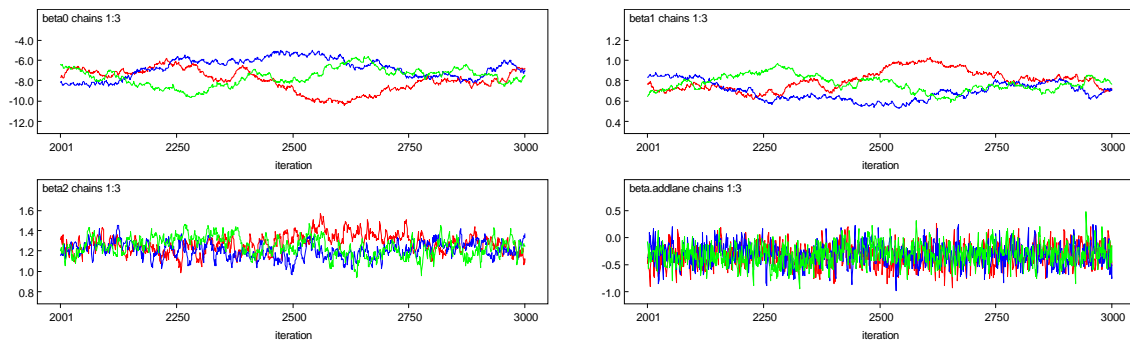


Figura 157| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As autocorrelações (ver Figura 158) evidenciam dificuldades na convergência dos parâmetros β_0 , β_1 e uma convergência mais lenta de β_2 .

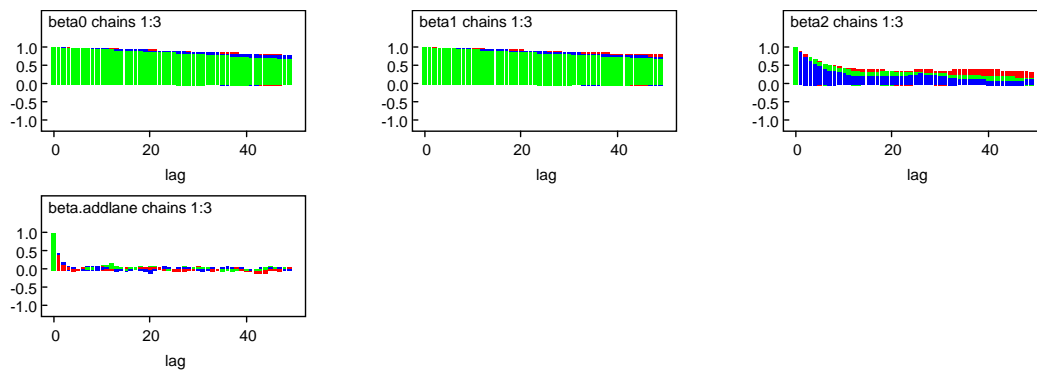


Figura 158| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser observados na Figura 159, mostram que, à medida que aumentam os números de iterações, as três estatísticas estabilizam e R tende para 1.

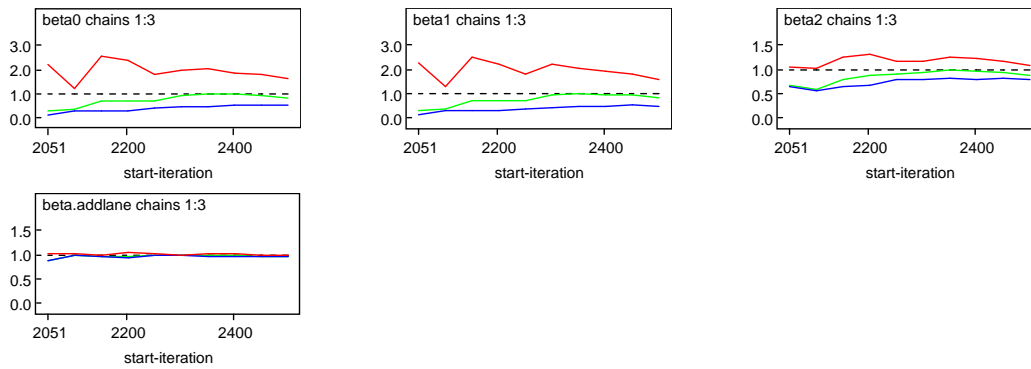


Figura 159| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_i do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As equações no Quadro 68 dão o número esperado de mortos e feridos graves obtidos pelo modelo Poisson-Gama, para um período de cinco anos, e para valores de *AADT*, *Length* e *Additional_Lane*.

Quadro 68| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Mortos e Feridos Graves
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 5.782 \times 10^{-4} \times AADT_i^{0.757} \times Length_i^{1.241}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 4.190 \times 10^{-4} \times AADT_i^{0.757} \times Length_i^{1.241}$

As estimativas para o modelo obtido encontram-se no Quadro 69. Um trecho com via adicional tem uma redução no número esperado de mortos e feridos graves, em cinco anos, de cerca de 28%, em relação a um trecho sem via adicional, mas tendo outras condições idênticas ao trecho com via adicional.

Quadro 69| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 68.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.456	1.105	1.122×10^{-1}	-9.865	-5.429
β_1 ($\log(AADT)$)	0.757	0.099	1.005×10^{-2}	0.568	0.975
β_2 ($\log(Length)$)	1.241	0.091	6.731×10^{-3}	1.066	1.422
β_3 (<i>Additional_Lane</i> ='Yes')	-0.322	0.196	6.558×10^{-3}	-0.710	0.079

As densidades *a posteriori* dos parâmetros β_j encontram-se na Figura 160.

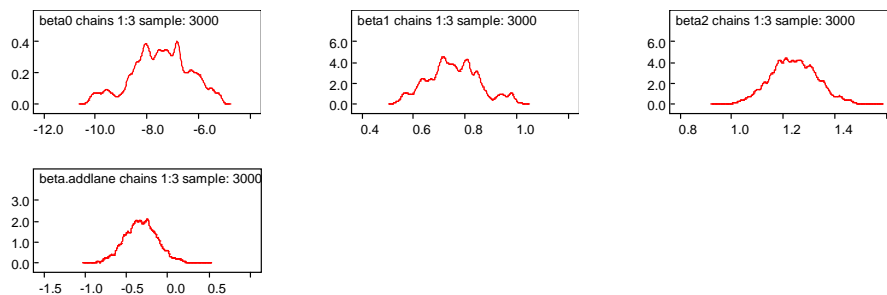


Figura 160| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves, obtidos de uma representação de um trecho com valores médios de *AADT* e *Length*, encontram-se no Quadro 70 para os dois níveis da variável *Additional_Lane*. Mais uma vez se constata um menor número esperado de ocorrências para trechos com presença de via adicional.

Quadro 70| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 68.

	<i>média (desvio padrão)</i>
<i>Additional_Lane='No'</i>	8.777 (0.623)
<i>Additional_Lane='Yes'</i>	6.456 (1.221)

6.4.2.1 Validação do Modelo

O exame dos histogramas da Figura 161 mostra que o histograma dos dados observados (a cinzento na figura) é verosímil sob a distribuição preditiva *a posteriori*.

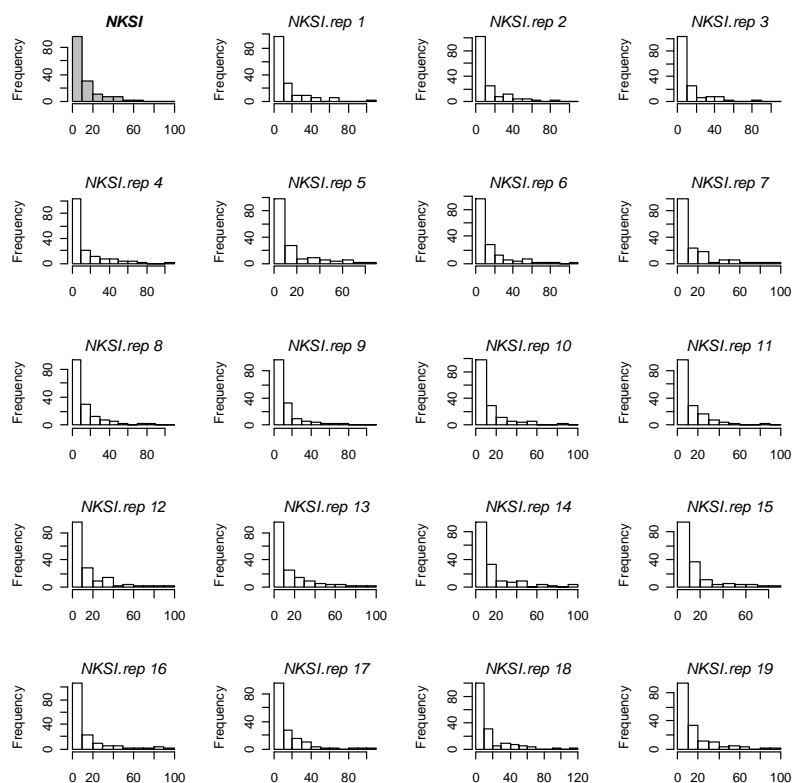


Figura 161| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas ("*NKSI.rep*") dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Pela interpretação dos histogramas na Figura 162 e na Figura 163 pode observar-se que o modelo captura adequadamente as quantidades *T* calculadas.

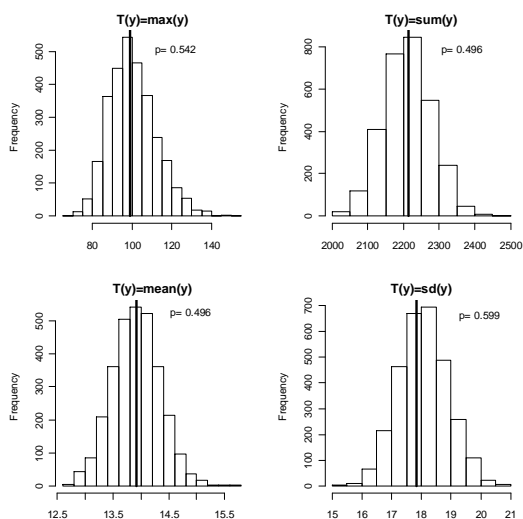


Figura 162| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O *p* é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

Os *p*-values Bayesianos de todas as quantidades são próximos de 0.5.

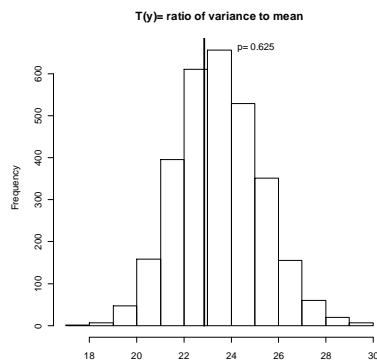


Figura 163| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a posteriori preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson-Gama produz $\bar{D}=146.5$, valor que indica um ajuste razoável uma vez que é próximo de $n-p=156$. O valor de DIC é igual a 846.58 e o parâmetro de dispersão é estimado em 0.502.

6.4.3 Modelo de Poisson Log-Normal

Tal como para a distribuição Poisson-Gama, o modelo baseado na distribuição Poisson Log-Normal ajustado aos dados referentes aos números de mortos e feridos graves que produziu o menor valor de DIC é representado pela equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Additional_Lane}_i \quad (6.12)$$

Os gráficos das iterações de cada parâmetro β_j ($j=1, \dots, 3$) na Figura 164 mostram haver evidência de não convergência de β_0 e β_1 .

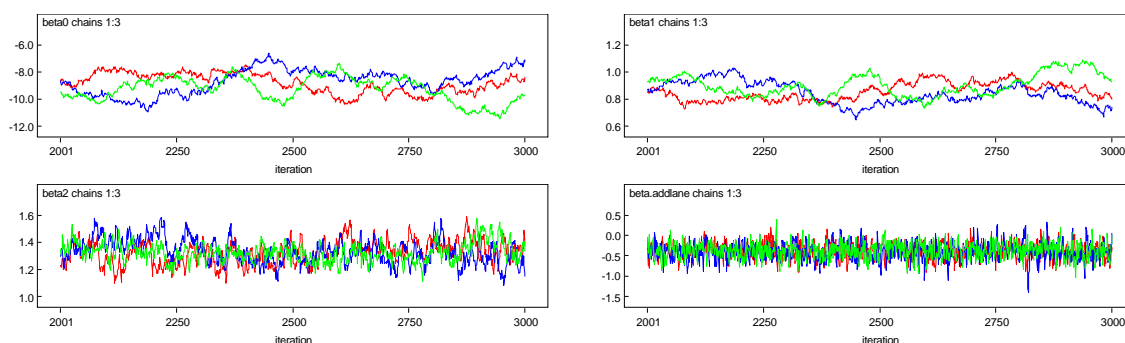


Figura 164| Trace plots (iterrações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NKS1 obtido do conjunto de dados cujos trechos possuem valores de AADT maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos das autocorrelações mostram também evidência de dificuldades na convergência de β_0 e β_1 (ver Figura 165).

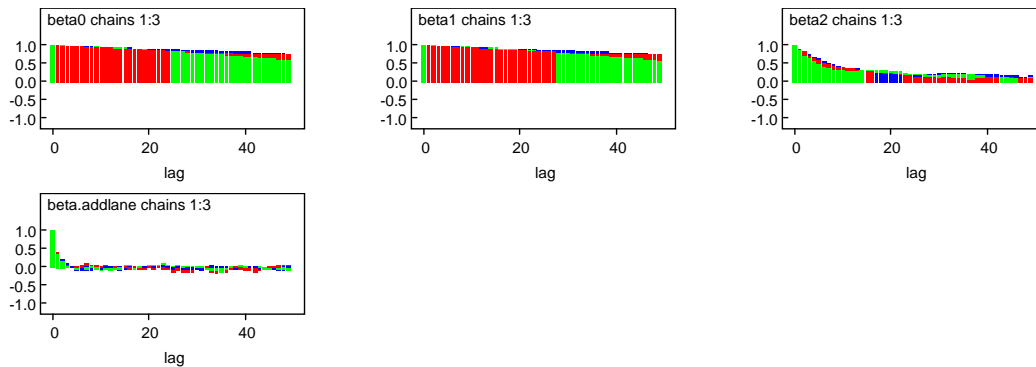


Figura 165| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

No entanto, as estatísticas de Gelman-Rubin mostram que os quatro parâmetros β_j convergem, como pode ser observado nos gráficos da Figura 166.

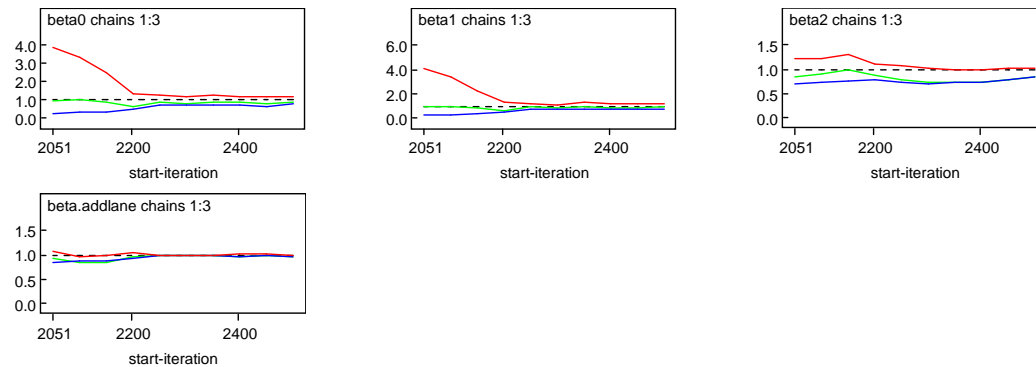


Figura 166| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

As duas equações para o número esperado de mortos e feridos graves para um período de cinco anos e para os dois níveis da variável *Additional_Lane* são apresentadas no Quadro 71.

Quadro 71| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

	Números Esperados de Mortos e Feridos Graves
<i>Additional_Lane</i> ='No'	$\hat{\lambda}_i = 1.309 \times 10^{-4} \times AADT_i^{0.866} \times Length_i^{1.326}$
<i>Additional_Lane</i> ='Yes'	$\hat{\lambda}_i = 8.904 \times 10^{-5} \times AADT_i^{0.866} \times Length_i^{1.326}$

As estimativas do modelo estão incluídas no Quadro 72. A estimativa do parâmetro correspondente a *Additional_Lane* sugere que um trecho de auto-estrada com uma via adicional tem uma redução no número de mortos e feridos graves de cerca de 32% em relação a um trecho em idêntica situação, mas sem via adicional.

Quadro 72| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 71.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.941	0.885	8.832×10^{-2}	-10.900	-7.342
β_1 (log(AADT))	0.866	0.080	7.940×10^{-3}	0.728	1.043
β_2 (log(Length))	1.326	0.082	5.543×10^{-3}	1.171	1.502
β_3 (Additional_Lane='Yes')	-0.385	0.198	5.335×10^{-3}	-0.780	-7.309×10^{-3}

As densidades *a posteriori* das estimativas dos parâmetros β_j encontram-se na Figura 167. É de reparar nas formas peculiares (sem máximos evidentes) das densidades correspondentes a β_0 e β_1 .

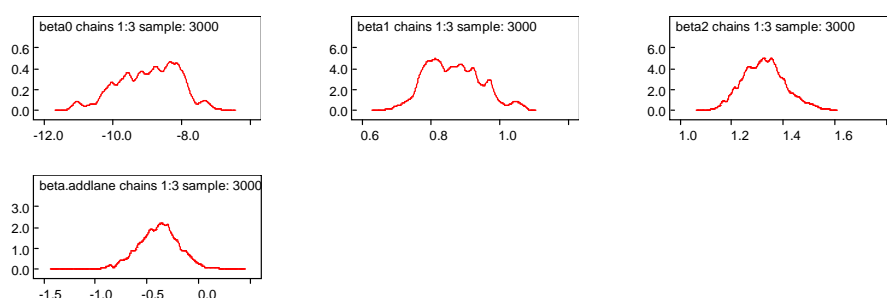


Figura 167| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *NKSI* (obtido do conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves para representações médias de *AADT* e *Length*, para um período de cinco anos, e de acordo com os dois níveis da variável *Additional_Lane*, são apresentados no Quadro 73.

Quadro 73| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 71.

	média (desvio padrão)
<i>Additional_Lane</i> ='No'	6.904 (0.499)
<i>Additional_Lane</i> ='Yes'	4.770 (0.909)

6.4.3.1 Validação do Modelo

A comparação dos histogramas na Figura 168 deixa crer que o histograma dos dados observados é razoável sob a distribuição preditiva *a posteriori* obtida pelo modelo considerado.

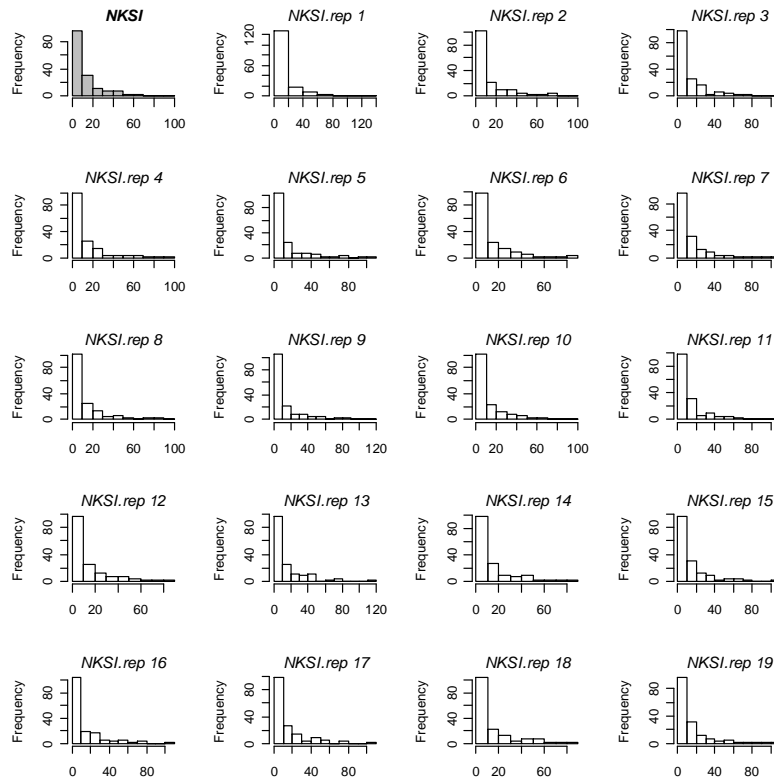


Figura 168| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos possuem valores de *AADT* maiores que 5000 e excluindo as variáveis com valores omissos.

Os gráficos na Figura 169 das medidas de discrepância calculadas mostram que o modelo considerado capta essas quantidades, isto é, os valores máximos, soma, média e desvio padrão.

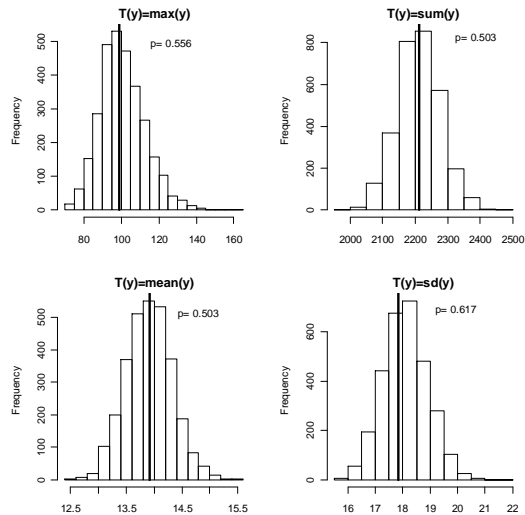


Figura 169| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo Poisson Log-Normal ajustado ao número de mortos e feridos graves também capta a dispersão dos dados, como se pode observar no histograma e recta vertical representados na Figura 170.

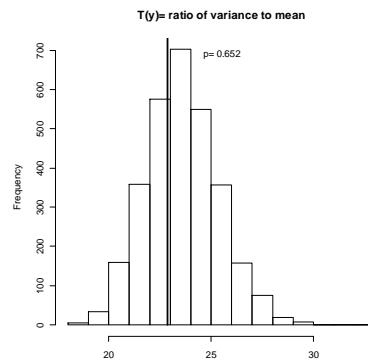


Figura 170| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo obteve uma média das desviâncias amostrais, \bar{D} , igual a 151.7 ($n-p=156$), DIC=858.46 e uma estimativa do parâmetro de dispersão igual a 0.487.

6.5 Discussão

No presente capítulo foram analisados somente os trechos de auto-estrada cujos valores de *AADT* foram superiores a 5000 veículos. As variáveis elegíveis para inserção nos modelos consistiram nas variáveis sem valores omissos, isto é, *Lane_Width*, *Number_of_Lanes*, *Additional_Lane* assim como $\log(AADT)$ e $\log(Length)$. As variáveis de resposta consideradas foram *Nac* (número de acidentes), *NF* (número de mortos), *NSI* (número de feridos graves) e *NKSI* (número de mortos e feridos graves). Concluiu-se que o modelo de regressão de Poisson não é apropriado para modelar nenhum dos eventos considerados pois, tal como foi descrito em Azeredo Lopes e Cardoso (2007), a distribuição de Poisson não tem em conta a sobredispersão existente nos dados.

Os modelos hierárquicos obtidos através da distribuição de Poisson-Gama obtiveram melhores medidas de ajuste, nomeadamente valores de DIC ligeiramente menores, do que os modelos hierárquicos baseados na distribuição Poisson Log-Normal.

Relativamente ao modelo obtido aquando do ajuste da distribuição Poisson-Gama ao número de acidentes, constatou-se que somente a variável categórica *Additional_Lane* foi incluída no modelo. Para um período de cinco anos, a existência de uma via adicional num trecho de auto-estrada reduz em cerca de 16% o número de acidentes, quando comparado com o número de acidentes ocorridos num trecho em idênticas condições, mas sem via adicional.

O número de mortos não parece ser influenciado pelas variáveis categóricas existentes, tendo sido obtido um modelo no qual *NF* depende somente de $\log(AADT)$ e $\log(Length)$. Os baixos números de ocorrências observados poderão explicar a insensibilidade à variação nas restantes potenciais variáveis explicativas. Os modelos referentes aos números de feridos graves e de mortos e feridos graves incluem a variável *Additional_Lane*. Um trecho com via adicional tem uma redução de cerca de 34% e 28% em *NSI* e *NKSI*, respectivamente, em relação a um trecho sem via adicional.

É também de notar que as observações dos números de vítimas não são independentes. Isto porque vão ser relacionadas com a taxa de ocupação dos veículos, informação essa que não foi possível obter.

7 | MODELAÇÃO DOS DADOS CONSTITUÍDOS POR TRECHOS SEM VALORES OMISSOS E COM VALORES DE AADT SUPERIORES 5000

A análise estatística descrita neste capítulo consiste na modelação, selecção e ajuste de modelos a conjuntos de dados obtidos nas auto-estradas portuguesas que registaram valores de *AADT* superiores a 5000 veículos, dos quais foram removidos os trechos com valores omissos. Os dados analisados consistem em 130 trechos de auto-estrada. Tal como no Capítulo 7 em Azeredo Lopes e Cardoso (2009), foi decidido modelar os dados tendo como variáveis resposta o número de acidentes (*Nac*), o número de mortos (*NF*), o número de feridos graves (*NSI*) e o número de mortos e feridos graves (*NKS*).

7.1 Modelação do Número de Acidentes

A descrição detalhada das características do número de acidentes, *AADT*, *Length* e outras variáveis tidas em consideração como possíveis covariáveis encontra-se na secção 7.1 de Azeredo Lopes e Cardoso (2009).

7.1.1 Modelo de Poisson

O processo *stepwise* de adicionar passo a passo as variáveis ao modelo simples, isto é, o modelo contendo $\log(AADT)$ e $\log(Length)$, e registando o valor de DIC resultou no modelo com a inclusão das variáveis *Left_Shoulder_Width* e *Median_Type*, como obtendo o valor de DIC mais baixo. A equação do meio em (3.22) pode agora ser representada por:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Left_Shoulder_Width_i + \beta_4 Median_Type_i \quad (7.1)$$

O algoritmo de MCMC converge após 30000 iterações com 20000 iterações *burn-in*, isto é, inicialmente rejeitadas e com um *thinning rate* igual a 10, resultando em amostras com dimensão igual a 3000. A observação dos *trace plots* da Figura 171 para os parâmetros β_j mostra que todos os valores obtidos para cada parâmetro se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências, facto que corrobora a convergência do algoritmo.

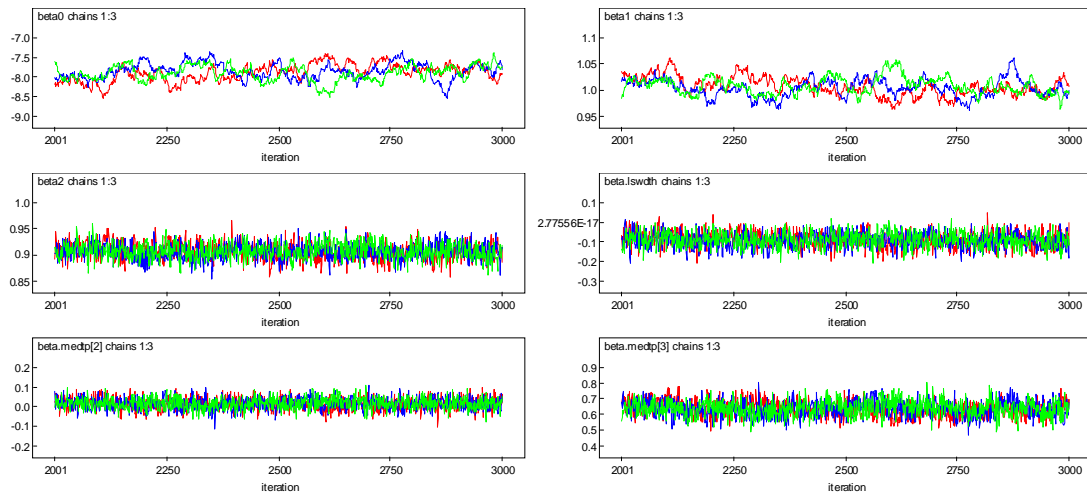


Figura 171| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações dos parâmetros β_j , reproduzidos na Figura 172 indicam uma convergência lenta de β_0 e β_1 e mais rápida das restantes covariáveis.

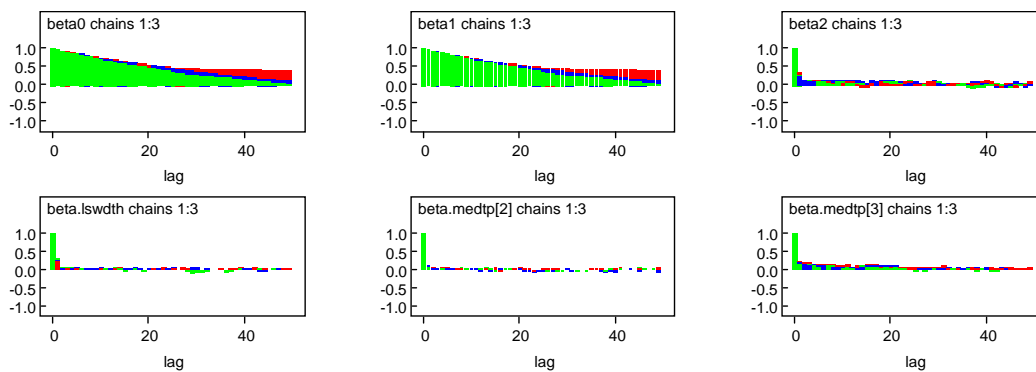


Figura 172| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 173. O algoritmo converge uma vez que R (ver equação 3.12) tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam ao aumentar o número de iterações.

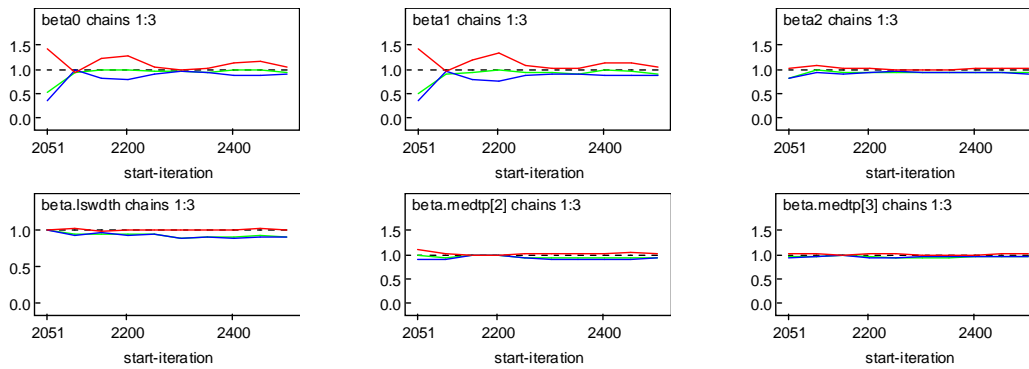


Figura 173| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

As equações do modelo, estimado *a posteriori*, para os números esperados de acidentes, num período de cinco anos, de acordo com os níveis das variáveis categóricas *Left_Shoulder_Width* e *Median_Type* encontram-se no Quadro 74. Pode observar-se que o número esperado de acidentes num determinado trecho de auto-estradas é menor quando a largura da berma esquerda é do tipo 'B' (largura maior do que 0.50 metros) e quando o separador é do tipo "DNJ;NJ".

Quadro 74| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

	Números Esperados de Acidentes
<i>Left_Shoulder_Width</i> ='A'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 3.750 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$
'GM'	$\hat{\lambda}_i = 3.803 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$
'Vala'	$\hat{\lambda}_i = 7.090 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$
<i>Left_Shoulder_Width</i> ='B'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 3.426 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$
'GM'	$\hat{\lambda}_i = 3.474 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$
'Vala'	$\hat{\lambda}_i = 6.477 \times 10^{-4} \times AADT_i^{1.006} \times Length_i^{0.907}$

O Quadro 75 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% das estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pela equação (7.1).

Quadro 75| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 74.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.888	0.205	1.833×10^{-2}	-8.334	-7.508
β_1 (log(AADT))	1.006	0.017	1.532×10^{-3}	0.974	1.043
β_2 (log(Length))	0.907	0.015	5.502×10^{-4}	0.876	0.936
β_3 (Left_Shoulder_Width='B')	-0.091	0.037	1.201×10^{-3}	-0.164	-0.018
β_4 (Median_Type='GM')	0.014	0.029	7.992×10^{-4}	-0.042	0.070
β_5 (Median_Type='Vala')	0.637	0.049	1.953×10^{-3}	0.540	0.734

Os intervalos de credibilidade de 95% para os parâmetros β_j , à exceção de β_4 , não incluem o valor zero, o que indica que todas as covariáveis consideradas têm um efeito significativo no número de acidentes, como se pode observar também nos gráficos da densidade *a posteriori* dos parâmetros β_j da Figura 174. Pela leitura dos valores do Quadro 75 também se pode depreender que um trecho de auto-estrada com largura da berma esquerda de tipo 'B' (maior que 0.50 metros) tem uma redução no número esperado de acidentes em cerca de 91% em relação a trechos com largura de berma esquerda do tipo 'A' (menor que 0.50 metros).

Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter mais cerca de 41% acidentes em relação a um trecho com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 89% mais acidentes do que um trecho com separador do tipo 'DNJ;NJ'.

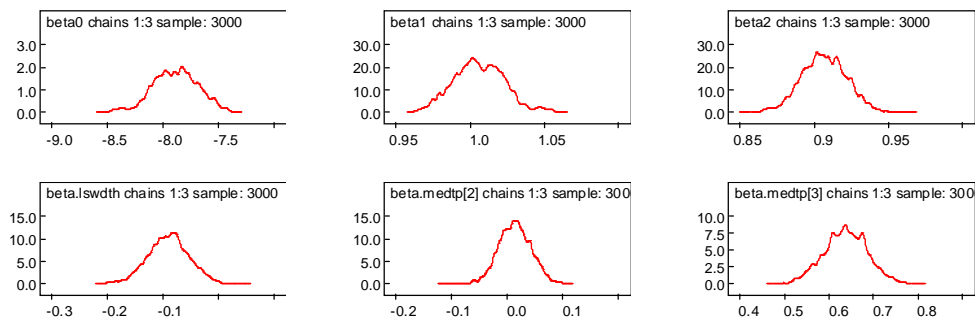


Figura 174| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a N_{ac} (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de acidentes para os diversos tipos de separador e larguras da berma esquerda, para uma representação média foram calculados sendo dados pelos valores indicados no Quadro 76.

Quadro 76] Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 74.

<i>Left_Shoulder_Width='A'</i>	<i>média (desvio padrão)</i>	<i>Left_Shoulder_Width='B'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	44.02 (1.516)	'DNJ;NJ'	40.20 (1.126)
'GM'	44.65 (1.761)	'GM'	40.76 (0.728)
'Vala'	83.27 (3.992)	'Vala'	76.03 (3.060)

Para um trecho típico de auto-estrada com largura de berma esquerda menor que 0.50 metros e com um separador do tipo 'DNJ;NJ', são esperados, para um período de cinco anos, 44.02 acidentes, enquanto que para um trecho com separadores do tipo 'GM' e 'Vala' são esperados, respectivamente, 44.65 e 83.27 acidentes. Para trechos com largura de berma esquerda superior a 0.50 metros são esperados 40.20, 40.76 e 76.03 acidentes quando esses trechos possuem separadores do tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente.

7.1.1.1 Validação do Modelo

Segundo Gelman *et al.* (2004) se o modelo está bem ajustado, então dados obtidos através do modelo, isto é, dados replicados devem ser semelhantes aos dados observados. A Figura 175 mostra vinte histogramas, estando no canto superior esquerdo o histograma dos dados observados (a cinzento) e todos os outros são histogramas de dados replicados da distribuição *a posteriori* preditiva. A observação desta figura mostra que vários histogramas das réplicas diferem consideravelmente do histograma dos dados observados, em particular os histogramas *Nac.rep 13* e *17*.

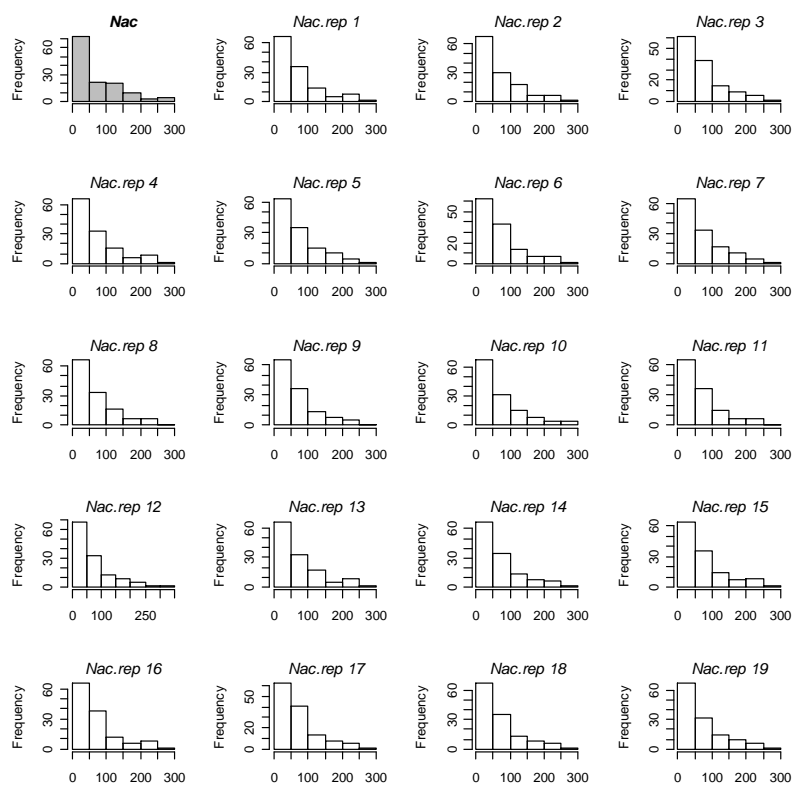


Figura 175| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Cada histograma na Figura 176 representa o histograma de cada uma das quantidades T calculadas. As rectas verticais indicam os valores obtidos pelos dados observados. Embora o modelo de Poisson capture a variação que os dados observados representam (em particular a soma e o valor médio), não captura para as outras quantidades, nomeadamente o valor máximo, $max(y)$ e o desvio padrão, $sd(y)$.

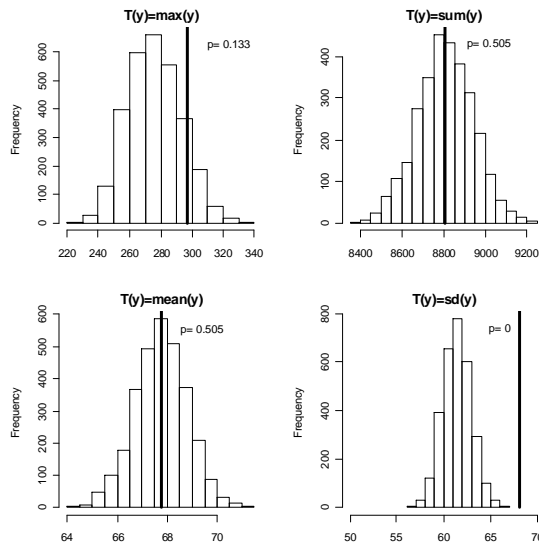


Figura 176| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Como se pode observar na Figura 177, a razão entre a variância e a média também não é capturada pelo modelo de regressão de Poisson, cujos valores para essa razão são representados pelo histograma e o valor da mesma razão obtido pelos dados observados é representado pela recta vertical.

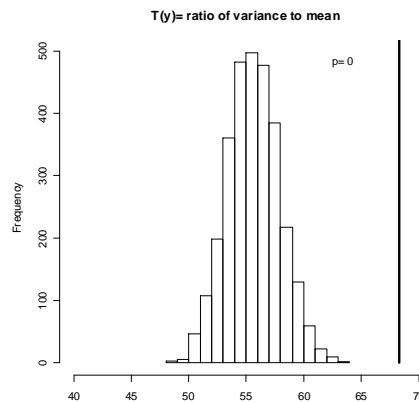


Figura 177| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais segundo a equação (3.15) resulta em $\bar{D} = 1779$, um valor bastante elevado quando comparado com os graus de liberdade $n-p$ (126), revelando que o modelo considerado não captura a sobredispersão existente nos dados. Este modelo produz um valor de DIC igual a 2512.57.

7.1.2 Modelo de Poisson-Gama

O modelo baseado na distribuição Poisson-Gama que, uma vez ajustado aos dados referentes aos números de acidentes, obteve o menor valor de DIC pode ser representado pela seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Median_Type_i \quad (7.2)$$

A observação dos *trace plots* da Figura 178 para os parâmetros β_j , mostra que todos os valores obtidos para os parâmetros β_2 e β_3 se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências. O mesmo não acontece para β_0 e β_1 .

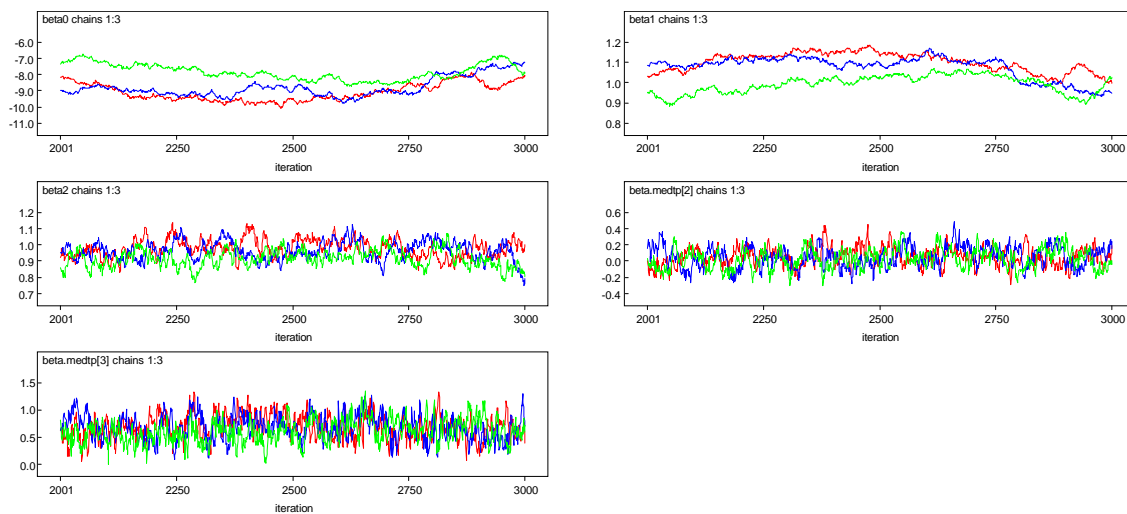


Figura 178| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações dos parâmetros β_j , reproduzidos na Figura 179 indicam uma convergência mais lenta de β_0 , e β_1 .

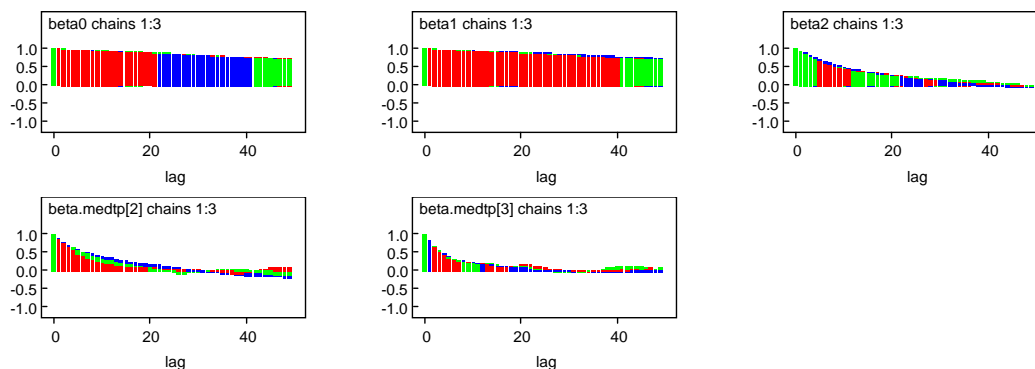


Figura 179| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

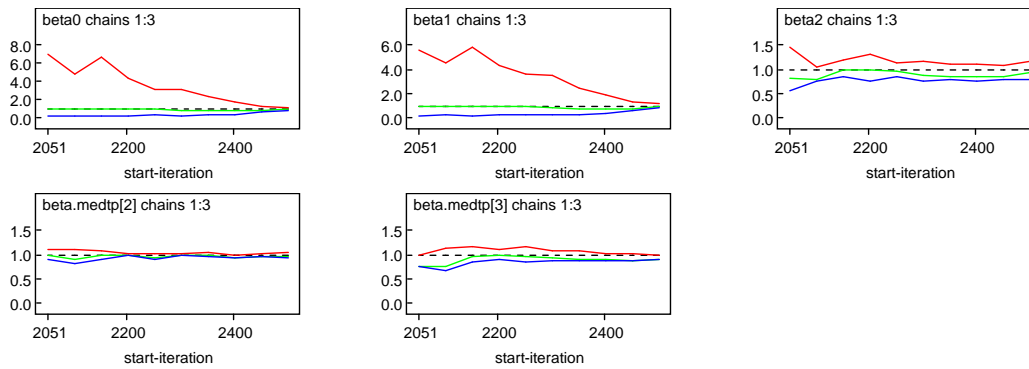


Figura 180| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a N_{ac} obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 180. O algoritmo converge uma vez que R (ver equação (3.12)) tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam à medida que o número de iterações aumenta.

Quadro 77| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	Números Esperados de Acidentes
Median_Type	
'DNJ;NJ'	$\hat{\lambda}_i = 1.883 \times 10^{-4} \times AADT_i^{1.057} \times Length_i^{0.952}$
'GM'	$\hat{\lambda}_i = 1.964 \times 10^{-4} \times AADT_i^{1.057} \times Length_i^{0.952}$
'Vala'	$\hat{\lambda}_i = 3.623 \times 10^{-4} \times AADT_i^{1.057} \times Length_i^{0.952}$

As equações do modelo, estimado *a posteriori*, para os números esperados de acidentes, num período de cinco anos, e de acordo com os níveis das variáveis categóricas *Median_Type*, encontram-se no Quadro 77. Pode observar-se que o número esperado de acidentes num trecho de estradas é menor quando o separador é do tipo "DNJ;NJ".

Quadro 78| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 77.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.578	0.758	7.743×10^{-2}	-9.733	-7.036
β_1 (log(AADT))	1.057	0.067	6.865×10^{-3}	0.923	1.161
β_2 (log(Length))	0.952	0.061	5.022×10^{-3}	0.832	1.069
β_4 (Median_Type='GM')	0.042	0.123	7.908×10^{-3}	-0.182	0.281
β_5 (Median_Type='Vala')	0.655	0.225	1.351×10^{-2}	0.223	1.104

O Quadro 78 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% das estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pelas equações representadas no Quadro 77. Na Figura 181 pode observar-se as densidades *a posteriori* dos parâmetros e reparar nas formas bastante peculiares referentes a β_0 e β_1 , possivelmente devido a não se ter atingido uma convergência conveniente para a distribuição *a posteriori*.

Pela leitura dos valores do Quadro 78 depreende-se que um trecho com separador do tipo 'GM' é *a posteriori* esperado ter aproximadamente mais 4% de acidentes do que um trecho (com o mesmo valor de *AADT* e *Length*, assim como *Median_Type*) com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 92% mais acidentes do que um trecho com os mesmos comprimento e valores de *AADT* e *Additional_Lane*, mas com separador do tipo 'DNJ;NJ'.

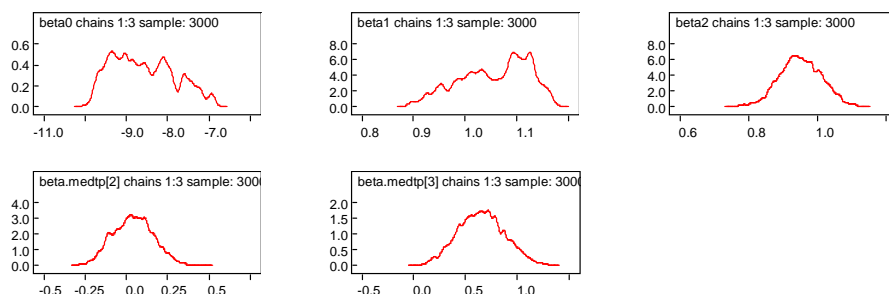


Figura 181| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a *Nac* (obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov

O número esperado de acidentes para os diversos tipos de separador para uma representação média de *AADT* e *Length* foi calculado, e é dado pelos valores indicados no Quadro 79.

Quadro 79| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 77.

<i>Median_Type</i>	<i>média (desvio padrão)</i>
'DNJ;NJ'	39.95 (3.930)
'GM'	41.57 (2.836)
'Vala'	77.95 (15.250)

7.1.2.1 Validação do Modelo

A observação da Figura 182 mostra que, em geral, os vários histogramas das réplicas são bastante semelhantes ao histograma dos dados observados, o que indica que o modelo em causa está bem ajustado aos dados.

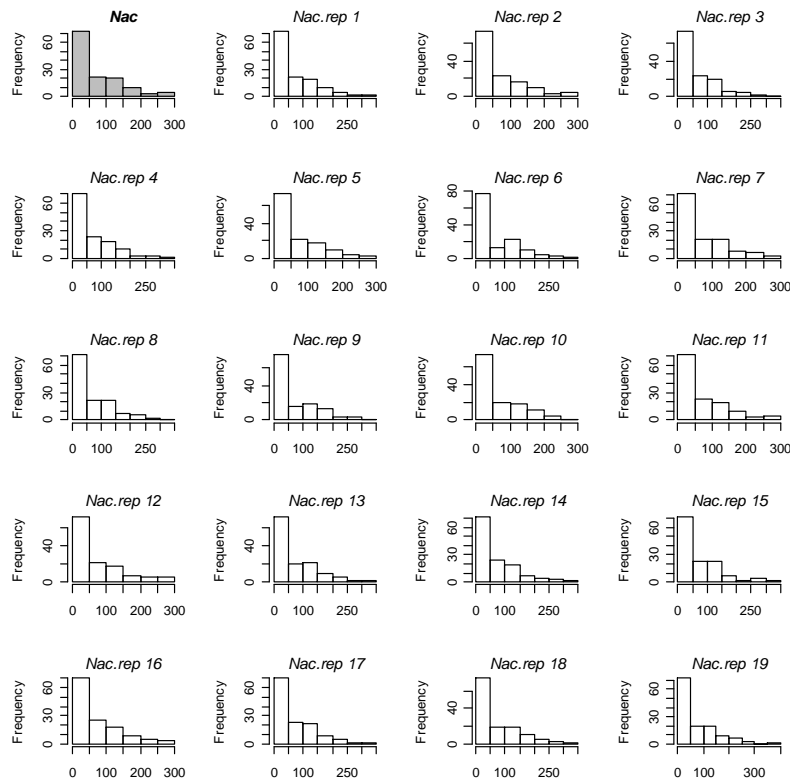


Figura 182| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Cada histograma da Figura 183 representa o histograma dos dados replicados de cada uma das discrepâncias consideradas onde as rectas verticais indicam os valores das mesmas discrepâncias obtidos pelos dados observados.

As probabilidades das discrepâncias obtidas pelos dados replicados serem superiores às discrepâncias obtidas pelos dados observados são aproximadamente iguais a 0.50 o que indica que, em relação às medidas de discrepância consideradas, os dados replicados são consistentes com os dados observados.

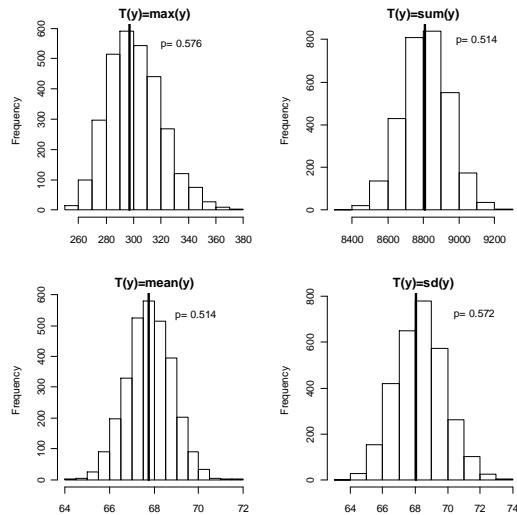


Figura 183| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Em média, a probabilidade de que a razão entre a variância e a média dos valores replicados pelo modelo Poisson-Gama seja superior à mesma razão obtida pelos dados é igual a 0.577 (ver Figura 184) o que indica que os dados observados e os dados replicados pelo modelo Poisson-Gama são comparáveis em termos da referida razão.

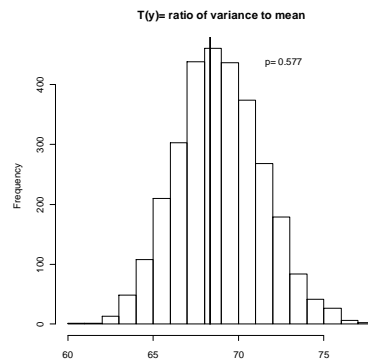


Figura 184| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais, ver equação (3.15) resulta num valor igual a $\bar{D}=133.7$ o que, comparado com os graus de liberdade, 127, indica um bom ajuste do modelo Poisson-Gama e uma melhoria em relação ao modelo de Poisson. A estimativa do parâmetro de dispersão é igual a 0.339. Este modelo produz um valor de DIC igual a 950.40, consideravelmente menor do que o valor obtido (2512.57) quando assumindo um modelo de Poisson para os números de acidentes.

Todas as discrepâncias investigadas sugerem que o modelo Poisson-Gama gera resultados preditivos semelhantes aos dados observados, ou seja, os dados observados são plausíveis de pertencerem a observações preditivas geradas pelo modelo Poisson-Gama.

7.1.3 Modelo de Poisson Log-Normal

O modelo baseado na distribuição Poisson Log-Normal que, uma vez ajustado aos dados referentes aos números de acidentes, obteve o menor valor de DIC é representado pela seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Median_Type}_i \quad (7.2)$$

A observação dos *trace plots* da Figura 185 para os parâmetros β_j , mostra que os valores obtidos para os parâmetros β_2 e β_3 se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências. O mesmo não acontece para β_0 e β_1 , onde se não verifica convergência dos valores das respectivas estimativas.

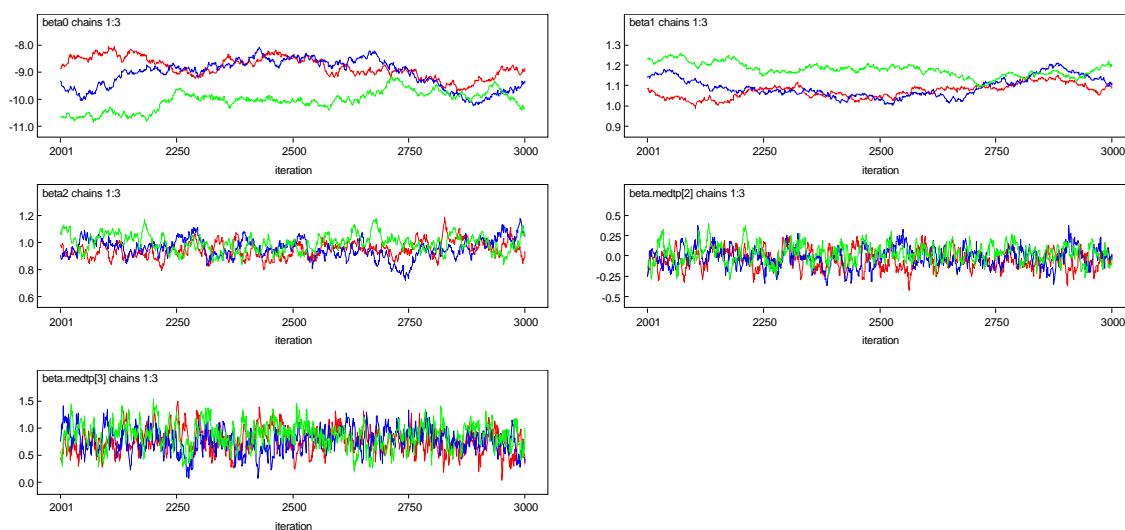


Figura 185| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a *Nac* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações dos valores dos parâmetros acima referidos estão reproduzidos na Figura 186. Os parâmetros β_0 , β_1 e β_2 convergem mais lentamente que os restantes.

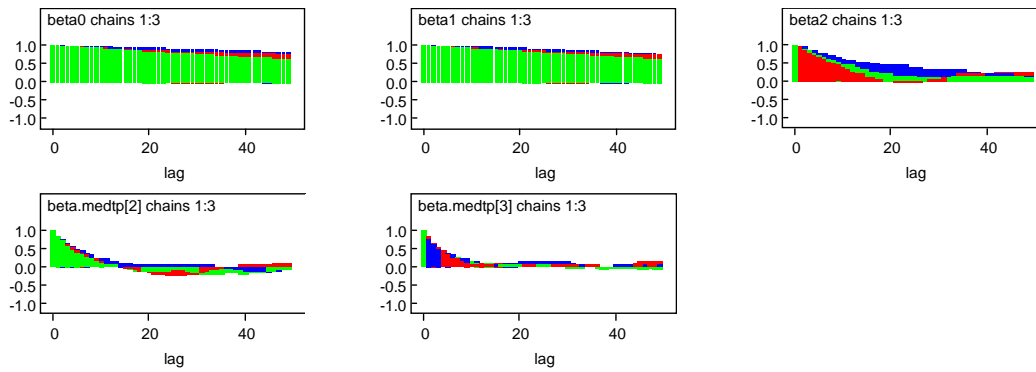


Figura 186| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 187. O algoritmo converge uma vez que R tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam à medida que o número de iterações aumenta.

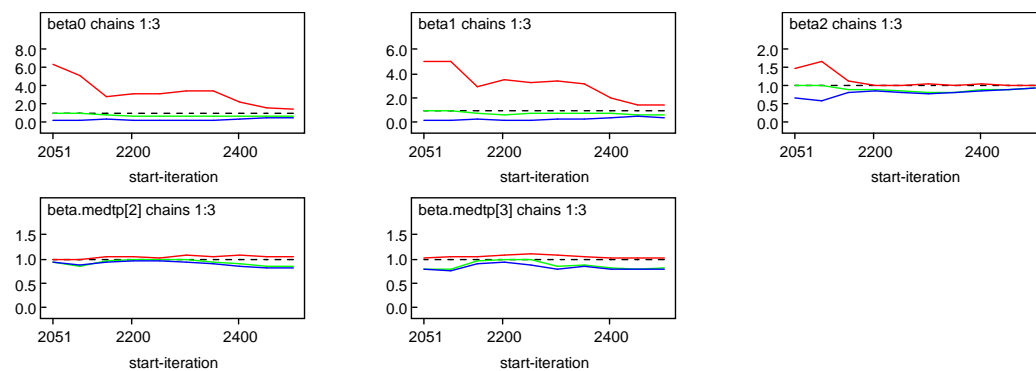


Figura 187| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a Nac obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

As equações para os números esperados de acidentes num período de cinco anos obtidos pelo modelo Poisson Log-Normal encontram-se no Quadro 80.

Quadro 80| Número esperado de acidentes, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

	<i>Números Esperados de Acidentes</i>
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 8.995 \times 10^{-5} \times AADT_i^{1.114} \times Length_i^{0.964}$
'GM'	$\hat{\lambda}_i = 8.897 \times 10^{-5} \times AADT_i^{1.114} \times Length_i^{0.964}$
'Vala'	$\hat{\lambda}_i = 1.995 \times 10^{-4} \times AADT_i^{1.114} \times Length_i^{0.964}$

Da observação das equações do Quadro 80 depreende-se que o número esperado de acidentes é menor, e bastante semelhante entre si, em trechos de auto-estradas com separadores dos tipos 'DNJ;NJ' e 'GM' do que nos separadores do tipo 'Vala'.

O Quadro 81 contém as estimativas pontuais, assim como intervalos de credibilidade de 95% para as estimativas dos parâmetros das equações no Quadro 80.

Os intervalos de credibilidade de 95% para os parâmetros β_j não incluem o valor zero, à excepção da variável *Median_Type* referente a *Median_Type="GM"*; no entanto, pela observação dos gráficos das densidades *a posteriori* dos parâmetros β_j na Figura 188, pode constatar-se que a média *a posteriori* para a estimativa de *Median_Type="GM"* (na segunda linha ao meio) se deslocou consideravelmente de zero (o valor médio *a priori*) mas que o intervalo de credibilidade de 95% ainda o inclui. Este facto faz com que se deva ser cauteloso ao considerar a variável *Median_Type* como não significativa.

Quadro 81| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 80.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.316	0.678	6.935×10^{-2}	-10.650	-8.283
β_1 (log(AADT))	1.114	0.060	6.148×10^{-3}	1.019	1.231
β_2 (log(Length))	0.964	0.067	5.353×10^{-3}	0.838	1.099
β_4 (Median_Type='GM')	-0.011	0.120	7.126×10^{-3}	-0.243	0.218
β_5 (Median_Type='Vala')	0.797	0.228	1.295×10^{-2}	0.353	1.246

Os valores das estimativas indicadas no Quadro 81 indicam que, para trechos com os mesmos valores das covariáveis *AADT* e *Length*, um separador do tipo 'GM' reduz o número esperado de acidentes em cerca de somente 1% quando comparado a um trecho com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter aproximadamente mais 122% de acidentes do que um trecho nas mesmas condições mas com separador do tipo 'DNJ;NJ'.

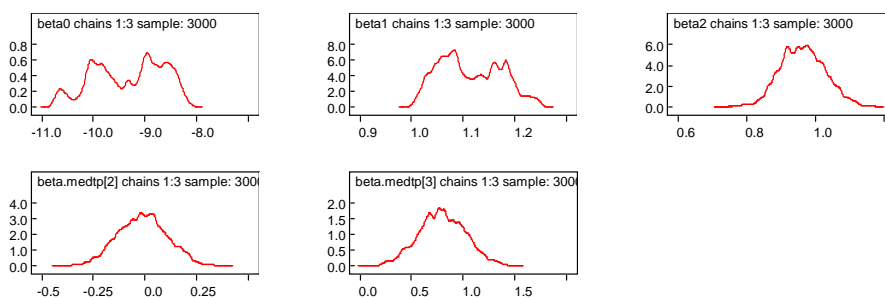


Figura 188| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *Nac* (obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov

O número esperado de acidentes para valores médios de *log(AADT)* e *log(Length)* e para os diversos tipos de separador e existência, ou não, de vias adicionais, é dado pelos valores indicados no Quadro

82. Num período de cinco anos, e para um trecho típico de auto-estrada com um separador do tipo 'DNJ;NJ' são esperados 34.8 acidentes, para trechos com separadores do tipo 'GM' e 'Vala' são esperados 34.4 e 78.5 acidentes, respectivamente.

Quadro 82| Médias *a posteriori* e desvios padrões do número esperado de acidentes, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 80.

<i>Median_Type</i>	<i>média (desvio padrão)</i>
'DNJ;NJ'	34.83 (3.337)
'GM'	34.38 (2.407)
'Vala'	78.45 (15.680)

7.1.3.1 Validação do Modelo

A Figura 189 contém o histograma do número de acidentes observado (a cinzento) e os dezanove histogramas dos números de acidentes replicados da distribuição *a posteriori* preditiva obtida quando se assume o modelo Poisson Log-Normal. A observação desta figura mostra que, em geral, os vários histogramas dos dados replicados são bastante semelhantes ao histograma dos dados observados, o que indica que o modelo Poisson Log-Normal está bem ajustado.

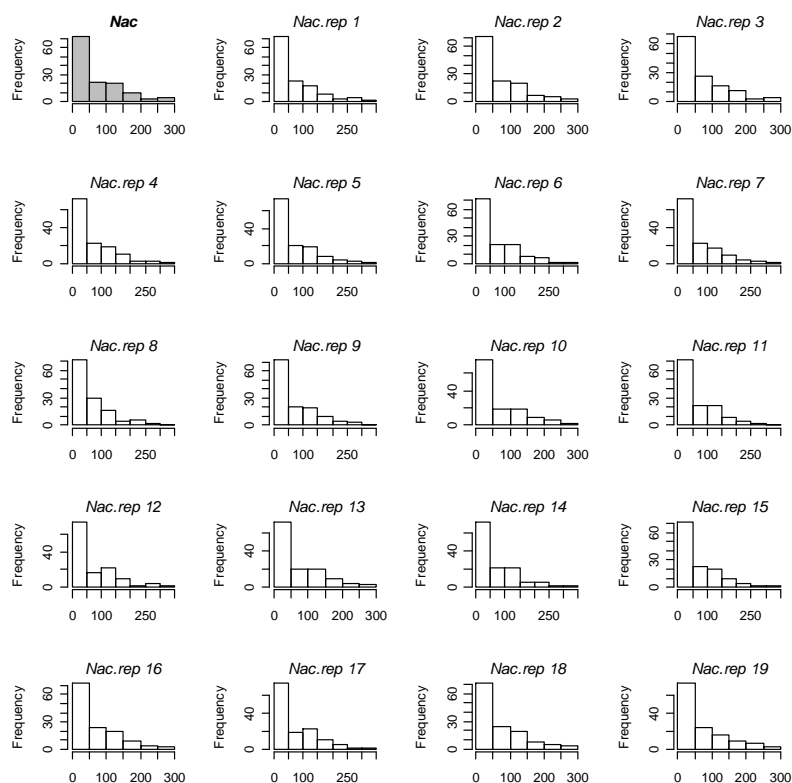


Figura 189| Histograma de *Nac* (no canto superior esquerdo) e dezanove réplicas (“*Nac.rep*”) dos números de acidentes obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Pela observação da Figura 190 deduz-se claramente que o modelo Poisson Log-Normal capta a variação dos dados observados no que respeita às quantidades *T* calculadas.

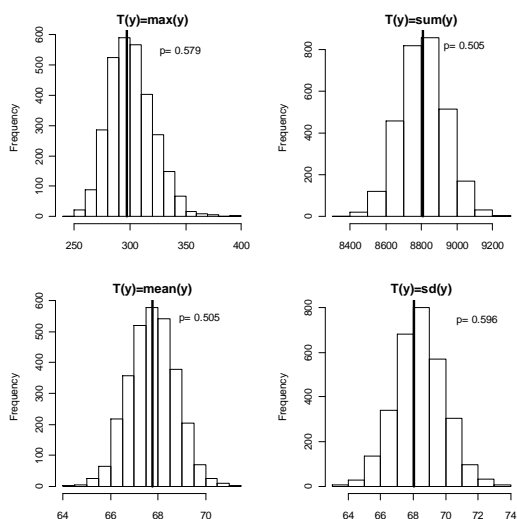


Figura 190| Histogramas dos valores *T*: máximo, soma, média e desvio padrão dos dados replicados referentes ao número de acidentes obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de *T*, obtido pelos dados replicados, ser superior ao *T* obtido pelos dados observados.

O histograma na Figura 191 refere-se à aproximação da distribuição preditiva *a posteriori* do teste $T = \text{variância}/\text{média}$ dos valores replicados; a recta vertical indica o valor da mesma razão T obtida nos dados observados.

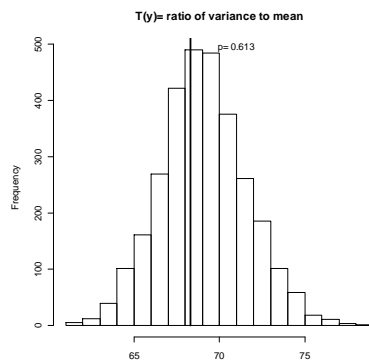


Figura 191| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

Devido ao facto do histograma “envolver” a recta vertical, pode concluir-se que o modelo Poisson Log-Normal capta a razão entre a variância e o valor médio dos dados observados. Em média, a probabilidade de que a razão T , acima referida, dos dados replicados pelo modelo Poisson Log-Normal seja superior à mesma razão obtida nos dados observados, isto é, no número de acidentes observados, é igual a 0.613, o que indica que os dados observados e os replicados são comparáveis em termos da referida razão.

A média das desviâncias amostrais (equação (3.15)) resulta num valor $\bar{D} = 150.70$ (maior do que o valor obtido aquando do ajuste do modelo Poisson-Gama), indicando um razoável ajuste do modelo Poisson Log-Normal ($n-p=127$). A estimativa do parâmetro de dispersão é igual a 0.321.

Este modelo produz um valor de DIC igual a 971.67, valor menor do que o valor obtido quando assumindo um modelo de Poisson (2512.57), mas ligeiramente maior do que o valor de DIC obtido com um modelo Poisson-Gama (950.40), o que indica o modelo de Poisson-Gama como o mais bem ajustado a estes dados.

Todas as discrepâncias investigadas sugerem que o modelo Poisson Log-Normal gera resultados preditivos semelhantes aos dados observados. No entanto os valores de DIC, as desviâncias amostrais e os p -values das medidas de discrepâncias sugerem o modelo de Poisson-Gama como sendo o mais bem ajustado aos dados relativos aos números de acidentes.

7.2 Modelação do Número de Mortos

A descrição dos dados analisados nesta secção encontra-se detalhada no Capítulo 7, secção 7.2 de Azeredo Lopes e Cardoso (2009).

7.2.1 Modelo de Poisson

O modelo baseado na distribuição Poisson que, uma vez ajustado aos dados referentes aos números de mortos, obteve o menor valor de DIC pode ser representado pela seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Median_Type_i \quad (7.2)$$

A observação dos *trace plots* da Figura 192 para os parâmetros β_j , mostra que os valores obtidos para os parâmetros β_j se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências.

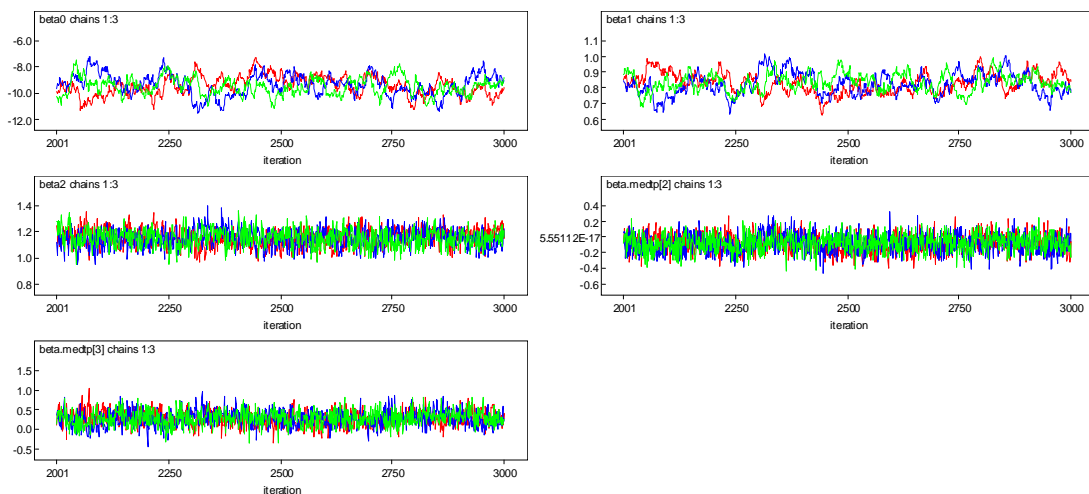


Figura 192| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NF* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações dos valores dos parâmetros acima referidos estão reproduzidos na Figura 193. Os parâmetros β_0 , e β_1 convergem mais lentamente que os restantes.

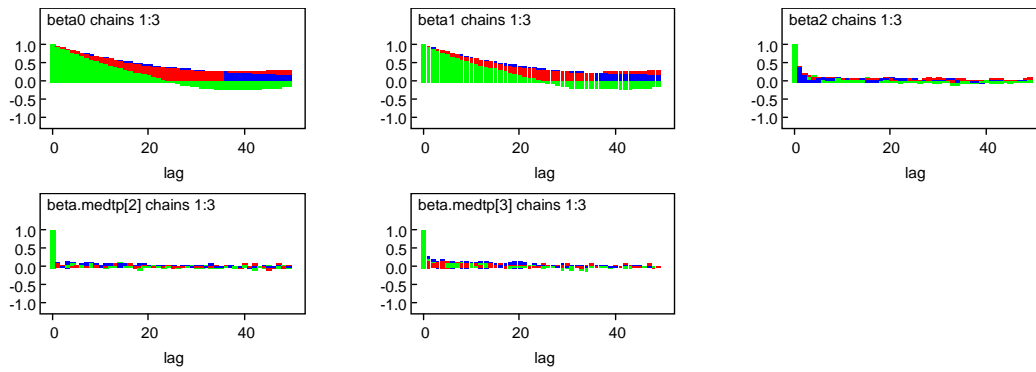


Figura 193| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_i do modelo de Poisson ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_i estão representados na Figura 194. O algoritmo converge uma vez que R tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam à medida que o número de iterações aumenta.

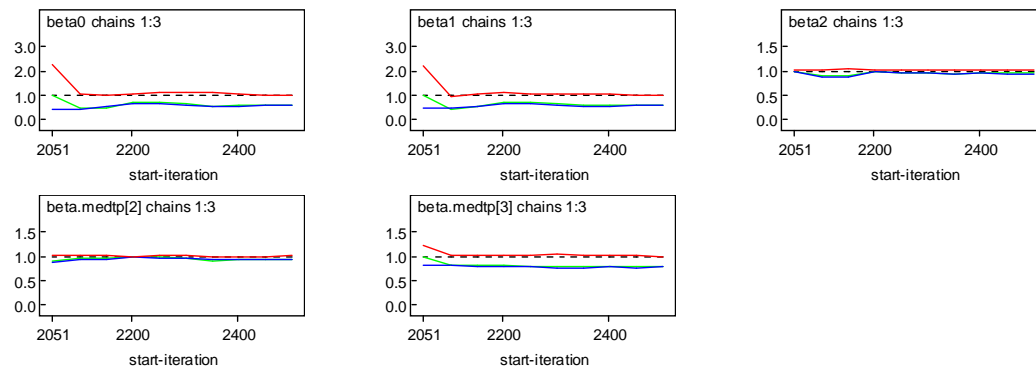


Figura 194| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_i do modelo de Poisson ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

As equações para os números esperados de mortos num período de cinco anos obtidos pelo modelo de Poisson encontram-se no Quadro 83 para os três níveis de $Median_Type$.

Quadro 83| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

	Números Esperados de Mortos
$Median_Type$	
'DNJ;NJ'	$\hat{\lambda}_i = 7.936 \times 10^{-5} \times AADT_i^{0.829} \times Length_i^{1.151}$
'GM'	$\hat{\lambda}_i = 7.306 \times 10^{-5} \times AADT_i^{0.829} \times Length_i^{1.151}$
'Vala'	$\hat{\lambda}_i = 1.047 \times 10^{-4} \times AADT_i^{0.829} \times Length_i^{1.151}$

Da observação das equações do Quadro 83 depreende-se que o número esperado de mortos é menor, e bastante semelhante entre si, em trechos de auto-estradas com separadores dos tipos 'DNJ;NJ' e 'GM' do que dos trechos com separador do tipo 'Vala'.

O Quadro 84 contém as estimativas pontuais, assim como os intervalos de credibilidade de 95% para as estimativas dos parâmetros das equações no Quadro 83.

Os intervalos de credibilidade de 95% para os parâmetros β_j não incluem o valor zero, à excepção da variável *Median_Type*; no entanto, pela observação dos gráficos das densidades *a posteriori* dos parâmetros β_j na Figura 195, pode constatar-se que as médias *a posteriori* para as estimativas de *Median_Type* (na segunda linha) se deslocaram consideravelmente de zero (o valor médio *a priori*) mas que o intervalo de credibilidade de 95% ainda o inclui. Este facto faz com que se seja cauteloso ao considerar a variável *Median_Type* como não significativa.

Quadro 84| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 83.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.441	0.736	6.142×10^{-2}	-10.960	-8.000
β_1 (log(AADT))	0.829	0.063	5.271×10^{-3}	0.704	0.957
β_2 (log(Length))	1.151	0.066	2.586×10^{-3}	1.024	1.279
β_4 (<i>Median_Type</i> ='GM')	-0.083	0.113	3.320×10^{-3}	-0.302	0.142
β_5 (<i>Median_Type</i> ='Vala')	0.277	0.194	7.438×10^{-3}	-0.101	0.651

Os valores das estimativas indicadas no Quadro 84 indicam que, para trechos com os mesmos valores das covariáveis *AADT* e *Length*, um separador do tipo 'GM' reduz o número esperado de mortos em cerca de 8% quando comparado com um trecho com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter aproximadamente mais 32% de ocorrências de mortos do que um trecho nas mesmas condições, mas com separador do tipo 'DNJ;NJ'.

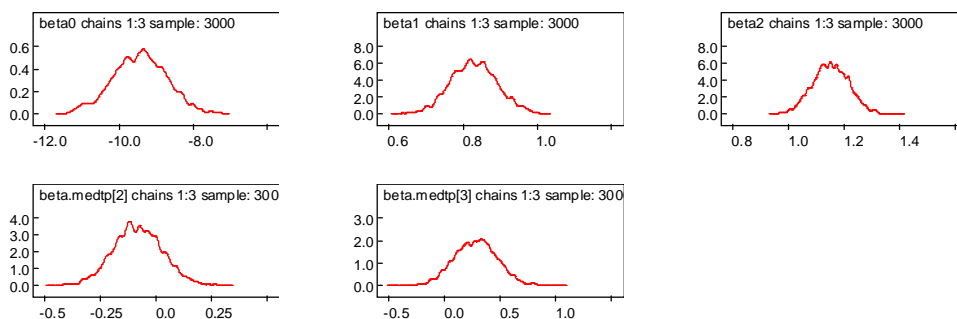


Figura 195| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a *NF* (obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de mortos em trechos de acordo com os três tipos de separador, e para uma representação média de *AADT* e *Length*, é dado no Quadro 85.

Quadro 85| Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 83.

<i>Median_Type</i>	<i>média (desvio padrão)</i>
'DNJ;NJ'	2.35 (0.235)
'GM'	2.16 (0.168)
'Vala'	3.12 (0.502)

Para um trecho típico de auto-estrada com separador do tipo 'DNJ;NJ' são esperados em média, para um período de cinco anos, 2.35 mortos. Para um trecho com separadores de tipo 'GM' e 'Vala', são esperados em média no mesmo período, 2.16 e 3.12 mortos, respectivamente.

7.2.1.1 Validação do Modelo

Os histogramas de dezanove conjuntos de dados replicados, assim como os dados originais, estão exibidos na Figura 196.

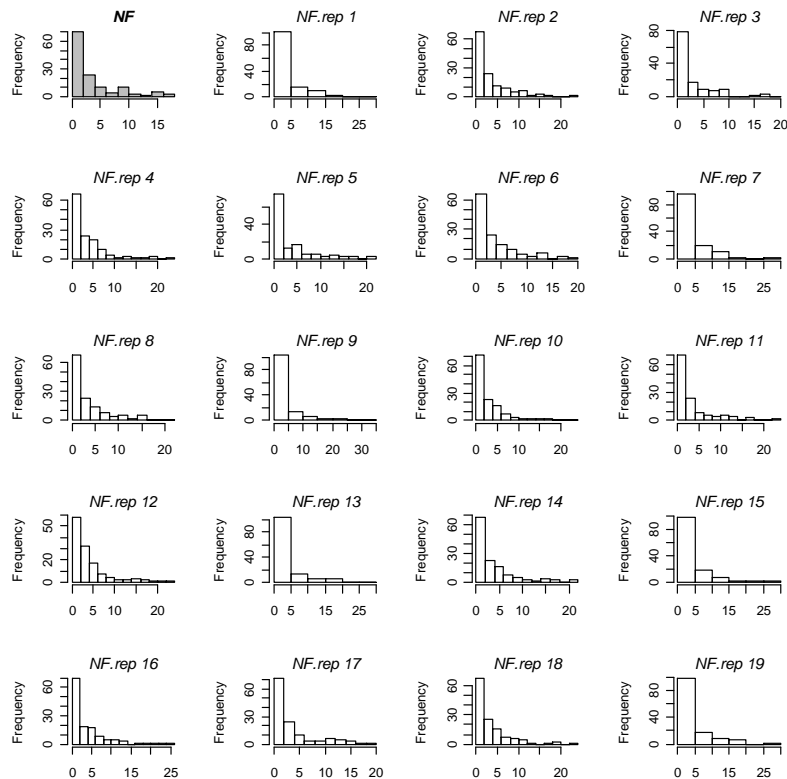


Figura 196| Histograma de *NF* (no canto superior esquerdo) e dezanove réplicas ("*NF.rep*") dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Uma vez que a maioria dos histogramas dos dados replicados diferem do histograma dos dados observados pode admitir-se que o modelo de Poisson considerado não se encontra bem ajustado aos dados.

Quando se investigam possíveis discrepâncias, nota-se que este modelo não é congruente com os dados observados no que respeita ao valor máximo e ao desvio padrão, tal como pode ser observado no histograma no canto superior esquerdo e no histograma no canto inferior direito, respectivamente, na Figura 197.

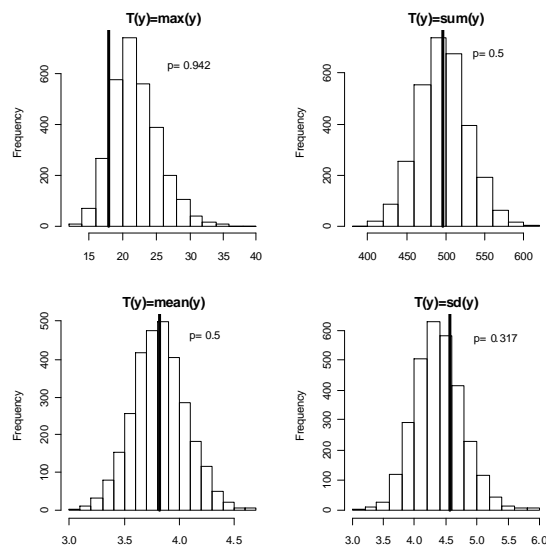


Figura 197| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo de Poisson também não capta a sobredispersão existente nos dados observados pois a razão entre a variância e o valor médio observados é consideravelmente maior do que a mesma razão obtida por dados replicados da distribuição preditiva *a posteriori*, como pode ser observado na Figura 198. A validação preditiva mostra uma probabilidade média $p_B=0.282$ de que a razão entre a variância e o valor médio dos dados replicados exceda a mesma razão dos dados observados.

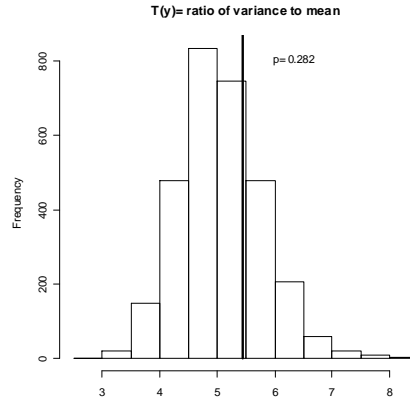


Figura 198| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a posteriori preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais (ver equação (3.15)) é igual a $\bar{D} = 176.1$, valor consideravelmente maior do que o número de graus de liberdade, 127, o que não indica um ajuste adequado do modelo de Poisson aos dados referentes ao número de mortos e feridos graves.

Este modelo produz um valor de DIC igual a 540.74.

7.2.2 Modelo de Poisson-Gama

O modelo de Poisson que, ajustado aos números de mortos (NF), produziu o menor valor de DIC tem a seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) \quad (7.5)$$

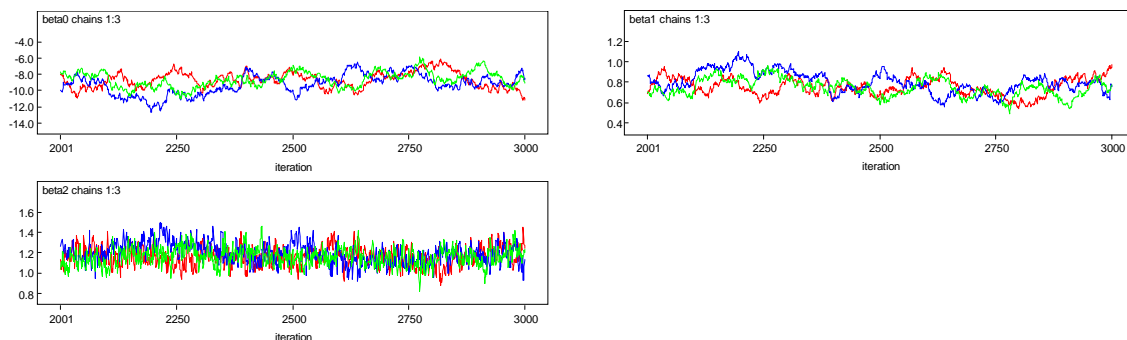


Figura 199| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Os gráficos da Figura 199 mostram que se obteve convergência das estimativas dos parâmetros β_j correspondentes à equação (7.5). Note-se, porém, que os parâmetros das estimativas de β_0 e β_1 (os dois gráficos no topo da Figura 199) não convergem tão bem como o parâmetro β_2 pois há uma maior oscilação entre as iterações.

Os gráficos das autocorrelações na Figura 200 corroboram o facto de as estimativas de β_0 e β_1 (os primeiros dois gráficos a contar da esquerda na Figura 200) não convergirem tão rapidamente como β_2 .

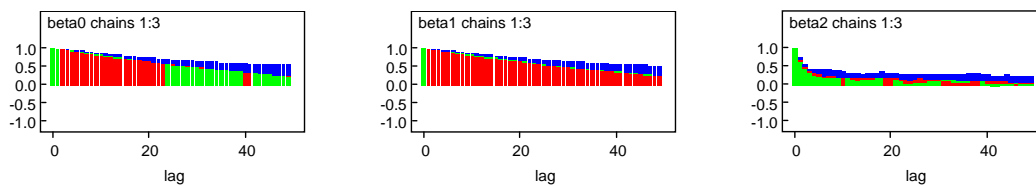


Figura 200| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Pela observação dos gráficos na Figura 201, correspondentes às estatísticas de Gelman-Rubin, depreende-se que a convergência é atingida para as três estimativas dos parâmetros β_j .

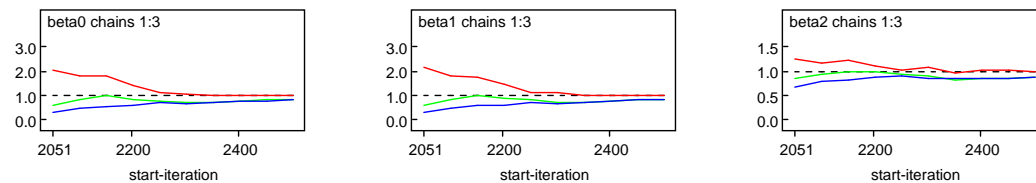


Figura 201| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a NF obtido do conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

O número esperado de mortos é obtido pela equação representada no Quadro 86 para valores de $AADT$ e $Length$.

Quadro 86| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

<i>Números Esperados de Mortos</i>
$\hat{\lambda}_i = 1.423 \times 10^{-4} \times AADT_i^{0.767} \times Length_i^{1.172}$

As estimativas e intervalos de credibilidade de 95% para os parâmetros do modelo Poisson-Gama ajustado encontram-se no Quadro 87.

Quadro 87| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 86.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.858	1.087	1.012×10^{-1}	-11.100	-6.848
β_1 (log(AADT))	0.767	0.095	8.848×10^{-3}	0.590	0.962
β_2 (log(Length))	1.172	0.095	5.567×10^{-3}	0.989	1.367

Os gráficos das densidades *a posteriori* das estimativas dos parâmetros β_j podem ser observados na Figura 202.

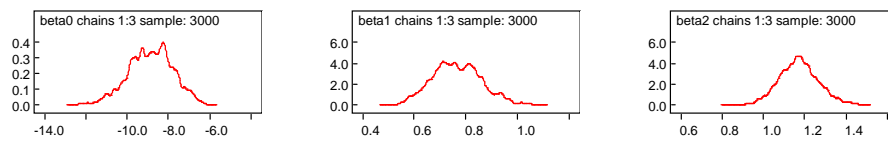


Figura 202| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a *NF* (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Para um trecho típico de auto-estrada são esperadas, num período de cinco anos, 2.327 mortes, dada uma representação média de *AADT* e *Length*. Este valor e o respectivo desvio padrão são apresentados no Quadro 88.

Quadro 88| Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 86.

média (desvio padrão)	2.327 (0.193)

7.2.2.1 Validação do Modelo

De um modo geral os histogramas obtidos de dados replicados pelo modelo Poisson-Gama representados na Figura 203 não se assemelham ao histograma dos dados observados (no canto superior esquerdo). Facto que pode indicar falhas potenciais do modelo em causa como sugere Gelman *et al.* (2004).

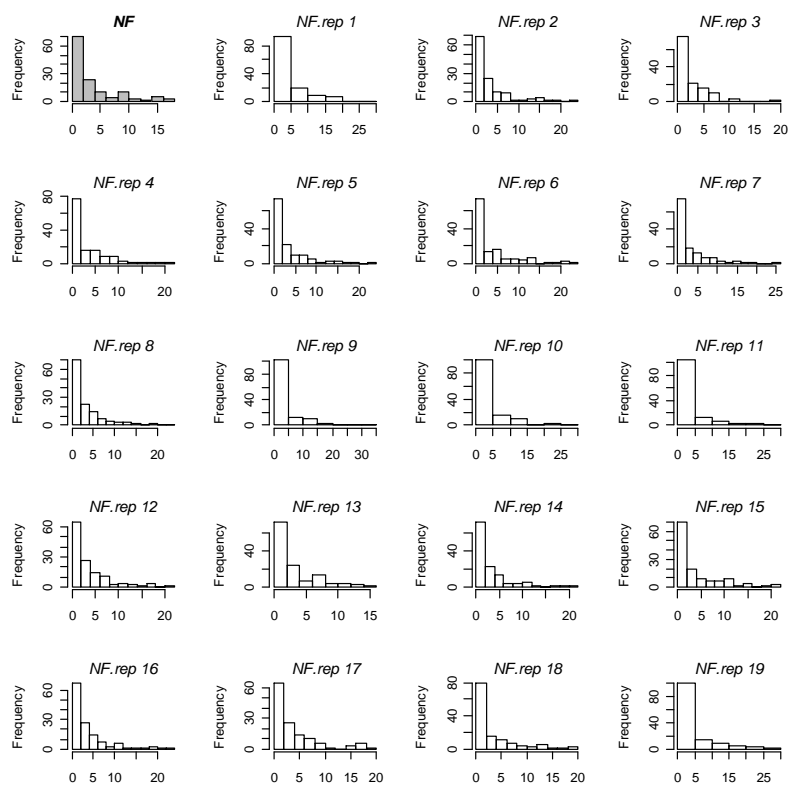


Figura 203| Histograma de NF (no canto superior esquerdo) e dezanove réplicas (" $NF.rep$ ") dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm $AADT$ maior que 5000 e não possuem valores omissos.

Os histogramas das medidas T de discrepância calculadas, assim como os valores dessas medidas observados nos dados (rectas verticais), encontram-se na Figura 204.

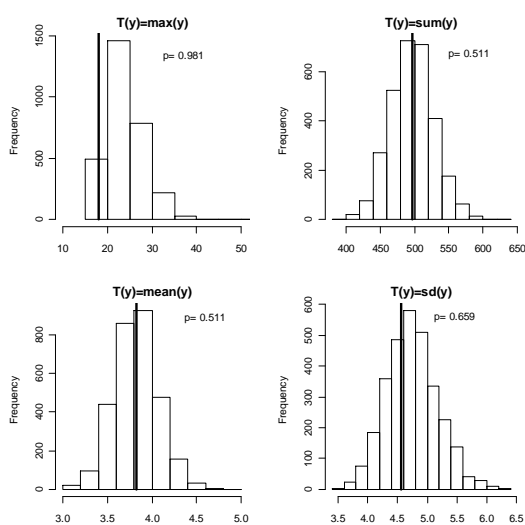


Figura 204| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Pode constatar-se que o modelo considerado não capta devidamente a variação que os dados observados contêm, nomeadamente no que diz respeito aos valores máximos (ver histograma no canto superior esquerdo na Figura 204).

A medida T respeitante à razão entre a variância e o valor médio indica que o modelo ajustado também não parece captar adequadamente a dispersão existente nos dados, tal como se pode constatar pela leitura da Figura 205. O p -value Bayesiano é igual a 0.693, valor consideravelmente superior ao ideal, 0.5.

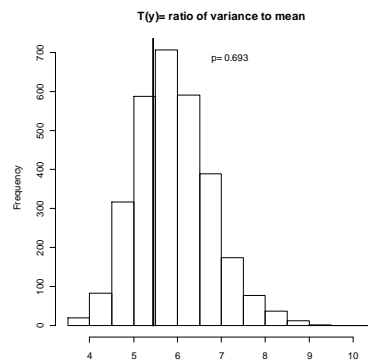


Figura 205| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

A média das desviâncias amostrais (ver equação (3.15)) é igual a $\bar{D}=107.1$, valor consideravelmente maior do que o número de graus de liberdade, 126, o que parece não indicar um ajuste adequado do modelo de Poisson-Gama aos dados referentes ao número de mortos, visto não captar muito bem a sobredispersão nos dados.

Este modelo produz um valor de DIC igual a 499.78 e uma estimativa do parâmetro de dispersão igual a 0.240.

7.2.3 Modelo de Poisson Log-Normal

O modelo hierárquico Poisson Log-Normal tendo como variável resposta o número de mortos (NF) que ajustado aos dados obteve o menor valor de DIC é dado pela equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) \quad (7.6)$$

onde $\hat{\lambda}_i$ diz respeito ao número esperado de mortos num período de cinco anos.

As iterações das estimativas dos parâmetros β_0 e β_1 na Figura 206 sugerem não haver evidência de não convergência. No entanto, pela observação dos gráficos das iterações de β_0 e β_1 nota-se que as

iterrações referentes às trs cadeia n3o se interligam como seria de esperar no caso de converg4ncia.

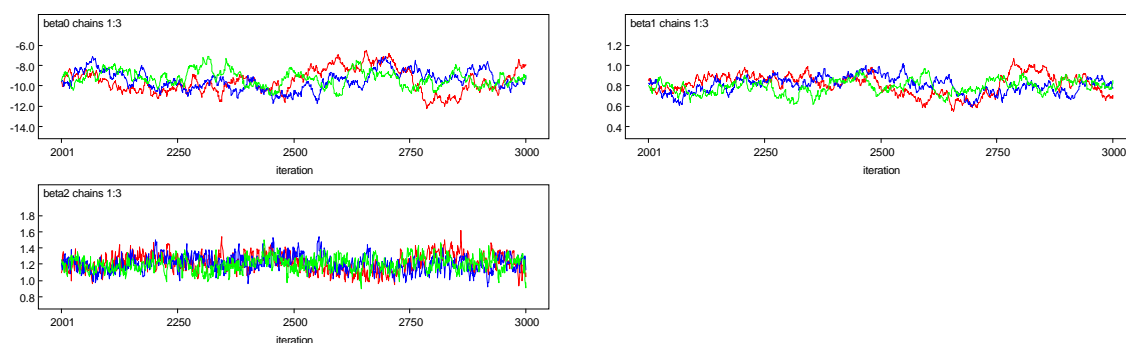


Figura 206| Trace plots (iterrações vs valores gerados) correspondentes a trs cadeia de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos t4em $AADT$ maior que 5000 e n3o possuem valores omissos.

Os gr3aficos das autocorrela3oes (na Figura 207) mostram que todos os parâmetros convergem lentamente.

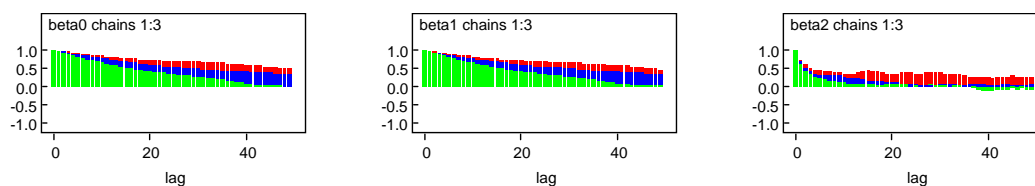


Figura 207| Gr3aficos das autocorrela3oes correspondentes a trs cadeia de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos t4em $AADT$ maior que 5000 e n3o possuem valores omissos.

Os gr3aficos das estatísticas de Gelman-Rubin na Figura 208 mostram que a estatística R (a encarnado na figura) converge para 1 para todos os parâmetros considerados, o que indica a converg4ncia do algoritmo.

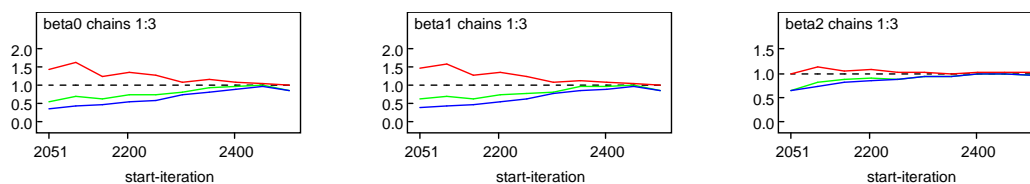


Figura 208| Estatísticas de Gelman-Rubin, correspondentes a trs cadeia de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NF obtido do conjunto de dados cujos trechos t4em $AADT$ maior que 5000 e n3o possuem valores omissos.

A equa33o que d3a o n3mero esperado de mortos, para um per3odo de cinco anos, encontra-se no Quadro 89. O Quadro 90 cont4m as estimativas pontuais e os intervalos de credibilidade de 95%.

Quadro 89| Número esperado de mortos, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Números Esperados de Mortos	
$\hat{\lambda}_i = 8.123 \times 10^{-5} \times AADT_i^{0.804} \times Length_i^{1.209}$	

Quadro 90| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 89.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.418	0.975	8.968×10^{-2}	-11.280	-7.468
β_1 (log(AADT))	0.804	0.084	7.679×10^{-3}	0.633	0.966
β_2 (log(Length))	1.209	0.098	5.916×10^{-3}	1.029	1.400

Pela observação dos gráficos das densidades *a posteriori* dos parâmetros β_j (ver Figura 209) constata-se que as médias se desviaram bastante do valor zero que foi admitido *a priori*.

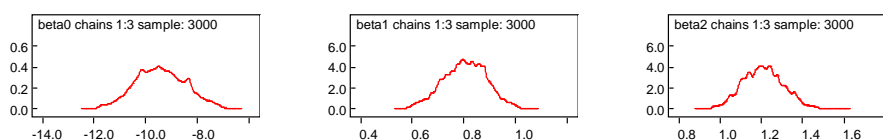


Figura 209| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a NF (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Esperam-se 2.069 mortos (em cinco anos) para uma representação média de AADT e Length, com um desvio padrão de 0.182, como pode ser observado no Quadro 91.

Quadro 91| Médias *a posteriori* e desvios padrões do número esperado de mortos, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 89.

<i>média (desvio padrão)</i>	2.069 (0.182)
------------------------------	---------------

7.2.3.1 Validação do Modelo

Os histogramas dos dados replicados são, na maioria, diferentes do histograma dos dados observados (ver Figura 210). Isto leva a crer que os dados observados não parecem ser muito representativos do modelo obtido, o que põe em dúvida o ajuste deste modelo aos dados.

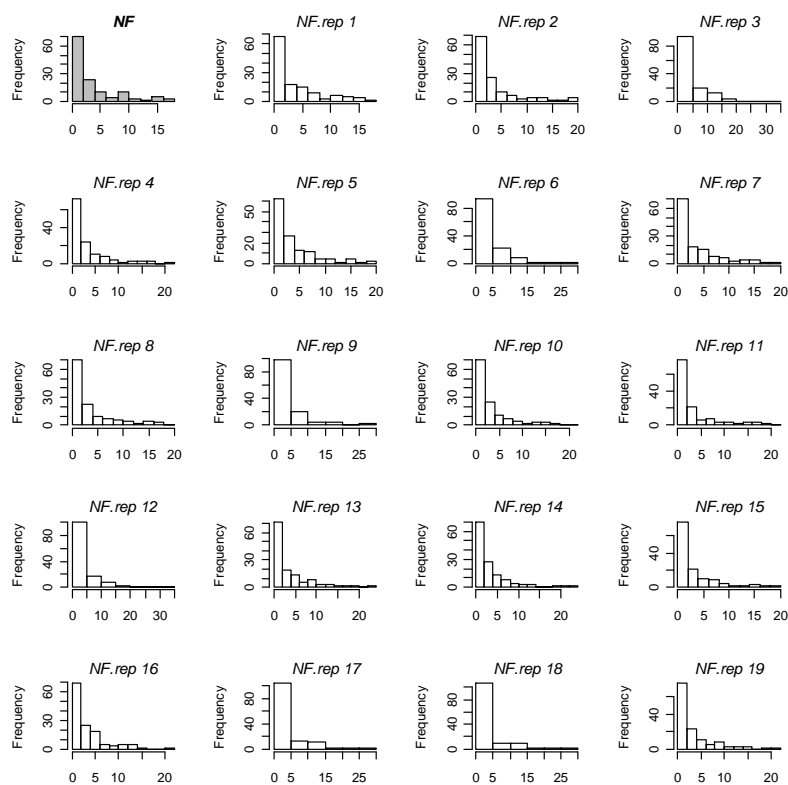


Figura 210| Histograma de NF (no canto superior esquerdo) e dezanove réplicas (“ $NF.rep$ ”) dos números de mortos obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

As discrepâncias entre os dados observados e os dados replicados pelo modelo dado pela equação no Quadro 90 podem ser observados no histograma do canto superior esquerdo da Figura 211 correspondendo a $T=max$. Os valores observados correspondentes a esta quantidade não são representativos dos dados replicados.

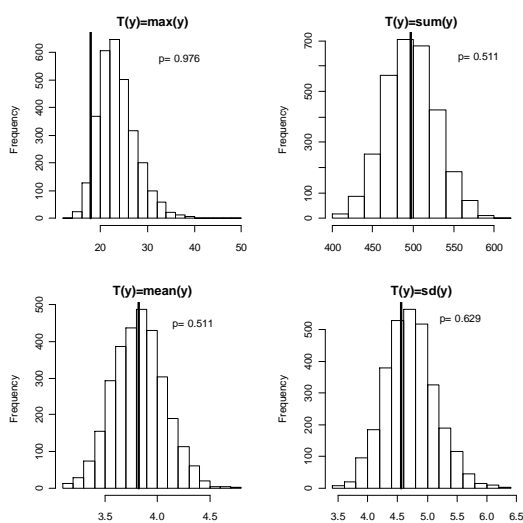


Figura 211| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos obtidos pelo modelo de Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A sobredispersão parece ser bem modelada pois como se pode observar na Figura 212, o valor observado (recta vertical) encontra-se “envolvido” pelo histograma das razões entre a variância e a média obtidos pelos dados da distribuição *a posteriori* preditiva.

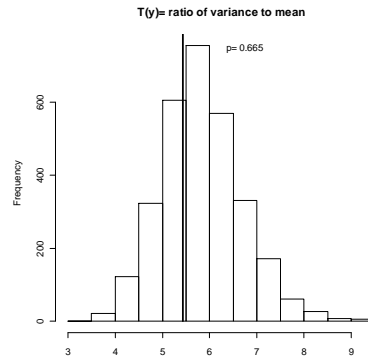


Figura 212| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson Log-Normal tendo o número de mortos como variável resposta obtém uma média das desviâncias amostrais, $\bar{D}=111.1$, o qual é inferior a $n-p=126$. O valor de DIC obtido é 505.39 e a estimativa para o parâmetro de dispersão é igual a 0.216.

7.3 Modelação do Número de Feridos Graves

A descrição dos dados analisados nesta secção encontra-se detalhada no Capítulo 7, secção 7.3 de Azeredo Lopes e Cardoso (2009).

7.3.1 Modelo de Poisson

O processo de adicionar passo a passo as variáveis ao modelo simples, isto é, o modelo contendo $\log(AADT)$ e $\log(Length)$, e registando o valor de DIC resultou no modelo com a inclusão das variáveis $Left_Shoulder_Width$ e $Median_Type$, como obtendo o valor de DIC mais baixo. A equação do meio em (3.22) pode agora ser representada por:

$$\log(\hat{\lambda}_j) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Left_Shoulder_Width_i + \beta_4 Median_Type_i \quad (7.7)$$

A observação dos *trace plots* da Figura 213 para os parâmetros β_j mostra que todos os valores obtidos para cada parâmetro se situam sempre entre “a mesma zona” sem se verificarem periodicidades nem tendências, facto que corrobora a convergência do algoritmo.

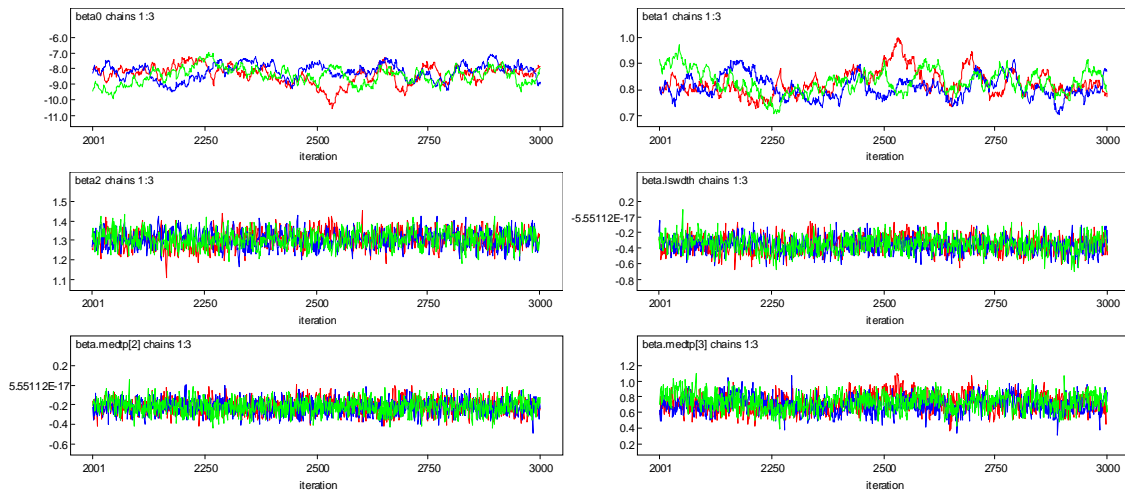


Figura 213| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a *NSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações dos parâmetros β_j , reproduzidos na Figura 214 indicam uma convergência mais lenta de β_0 e β_1 .

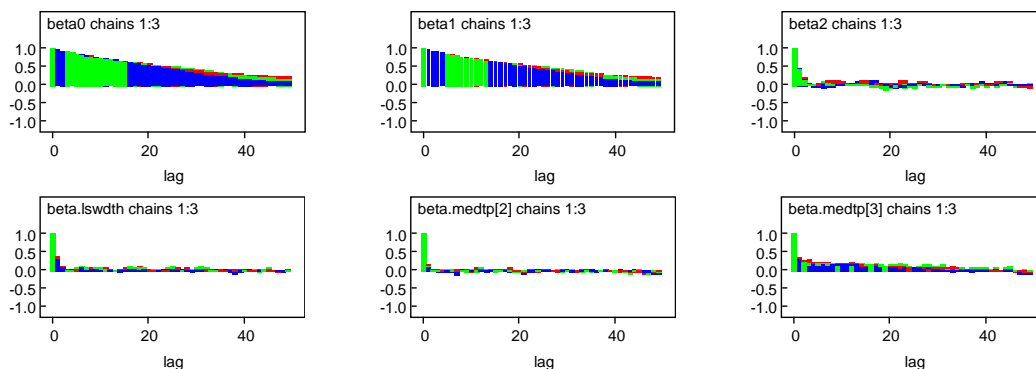


Figura 214| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a *NSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os gráficos dos valores das estatísticas de Gelman-Rubin para os parâmetros β_j estão representados na Figura 215. O algoritmo converge uma vez que R tende para 1 (a vermelho na figura) e que, tanto W (a azul) como \hat{V} (a verde), estabilizam ao aumentar o número de iterações.

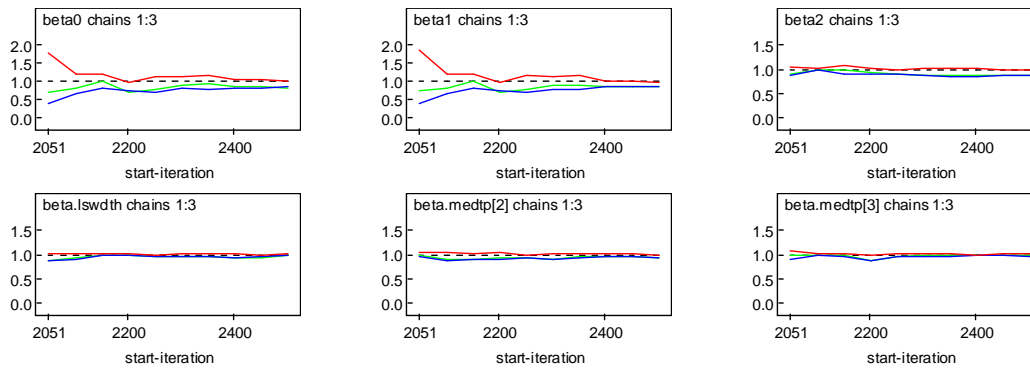


Figura 215| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a NSI obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

As equações do modelo estimado *a posteriori*, para os números esperados de acidentes, num período de cinco anos, de acordo com os níveis das variáveis categóricas *Left_Shoulder_Width* e *Median_Type* encontram-se no Quadro 92. Pode observar-se que o número esperado de feridos graves num determinado trecho de auto-estradas é menor quando a largura da berma esquerda é do tipo 'B' (largura maior do que 0.50 metros) e quando o separador é do tipo "GM".

Quadro 92| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	Números Esperados de Feridos Graves
<i>Left_Shoulder_Width</i> ='A'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 2.378 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$
'GM'	$\hat{\lambda}_i = 1.904 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$
'Vala'	$\hat{\lambda}_i = 4.873 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$
<i>Left_Shoulder_Width</i> ='B'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 1.667 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$
'GM'	$\hat{\lambda}_i = 1.335 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$
'Vala'	$\hat{\lambda}_i = 3.416 \times 10^{-4} \times AADT_i^{0.821} \times Length_i^{1.307}$

O Quadro 93 mostra os valores médios, desvios padrões, erros de Monte Carlo e intervalos de credibilidade de 95% das estimativas da densidade *a posteriori* dos parâmetros β_j do modelo descrito pela equação (7.7).

Quadro 93| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 92.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-8.344	0.538	4.834×10^{-2}	-9.460	-8.283
β_1 (log(AADT))	0.821	0.046	4.112×10^{-3}	0.741	0.815
β_2 (log(Length))	1.307	0.042	1.392×10^{-3}	1.225	1.307
β_3 (Left_Shoulder_Width='B')	-0.355	0.101	3.391×10^{-3}	-0.553	-0.354
β_4 (Median_Type='GM')	-0.222	0.074	1.861×10^{-3}	-0.367	-0.224
β_5 (Median_Type='Vala')	0.717	0.110	5.392×10^{-3}	0.507	0.715

Pela leitura dos valores do Quadro 93 depreende-se também que um trecho de auto-estrada com largura da berma esquerda de tipo 'B' (maior que 0.50 metros) tem um número esperado de feridos graves cerca de 30% menor do que um trecho com largura de berma esquerda do tipo 'A' (menor que 0.50 metros).

Um trecho com separador do tipo 'GM' é *a posteriori* esperado ter um aumento do número de feridos graves cerca de 20% maior do que um trecho com separador do tipo 'DNJ;NJ'. Um trecho com separador do tipo 'Vala' é *a posteriori* esperado ter 105% mais feridos graves do que um trecho com separador do tipo 'DNJ;NJ'.

As densidades *a posteriori* das estimativas dos parâmetros β_j podem ser observadas na Figura 216.

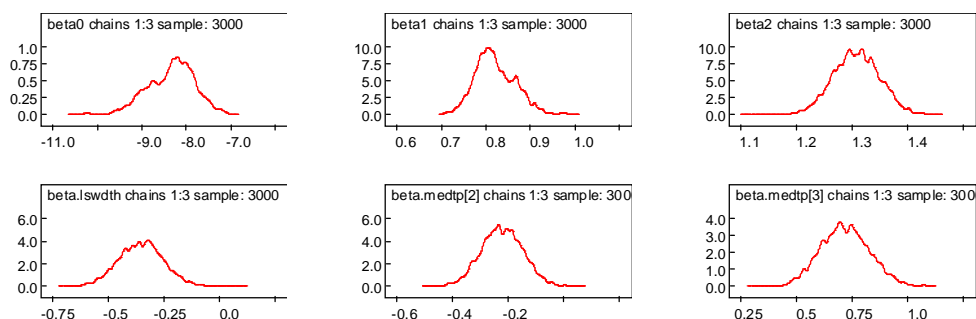


Figura 216| Densidades *a posteriori* dos parâmetros β_i do modelo de Poisson ajustado a NSI (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

O número esperado de feridos graves para os diversos tipos de separador e larguras da berma esquerda, para uma representação média foi calculado e é dado pelos valores indicados no Quadro 94.

Quadro 94| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 92.

<i>Left_Shoulder_Width='A'</i>	<i>média (desvio padrão)</i>	<i>Left_Shoulder_Width='B'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	8.449 (0.765)	'DNJ;NJ'	5.914 (0.422)
'GM'	6.775 (0.720)	'GM'	4.729 (0.246)
'Vala'	17.350 (1.955)	'Vala'	12.130 (1.064)

Para um trecho típico de auto-estrada com largura de berma esquerda menor que 0.50 metros e com um separador do tipo 'DNJ;NJ', são esperados, para um período de cinco anos, 8.5 feridos graves, enquanto que para um trecho com separadores do tipo 'GM' e 'Vala' são esperados, respectivamente, 6.8 e 17.4 feridos graves. Para trechos com largura de berma esquerda superior a 0.50 metros são esperados 5.9, 4.7 e 12.1 feridos graves quando esses trechos possuem separadores do tipo 'DNJ;NJ', 'GM' e 'Vala', respectivamente.

7.3.1.1 Validação do Modelo

Em geral os histogramas dos dados replicados diferem do histograma dos dados observados, como pode ser observado na Figura 217, o que evidencia falhas no modelo considerado.

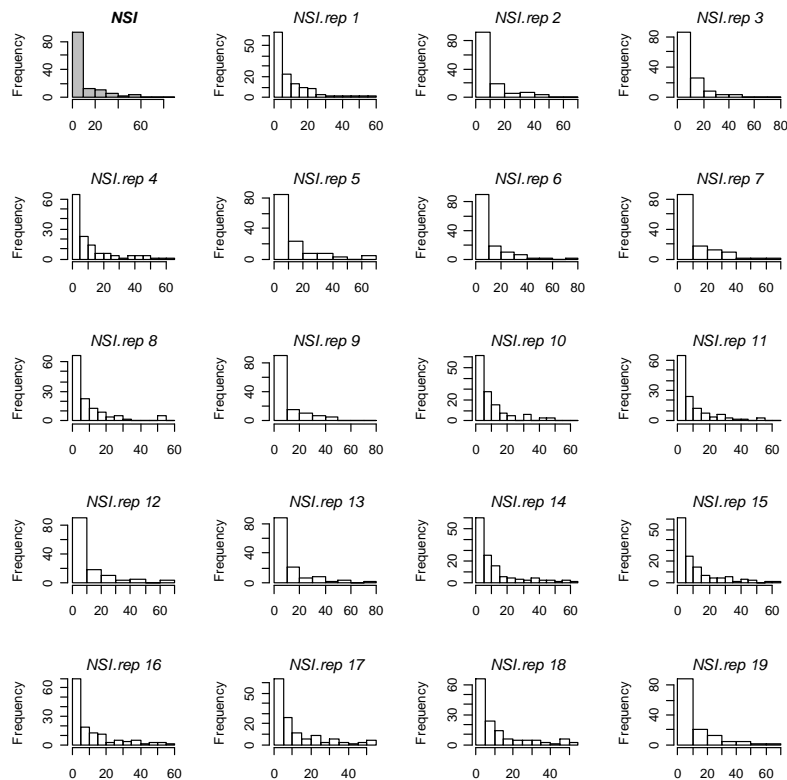


Figura 217| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas ("*NSI.rep*") dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Pela leitura dos histogramas na Figura 218 depreende-se que o modelo de Poisson considerado não capta a variação existente nos dados, em especial no que respeita ao valor máximo e ao desvio padrão.

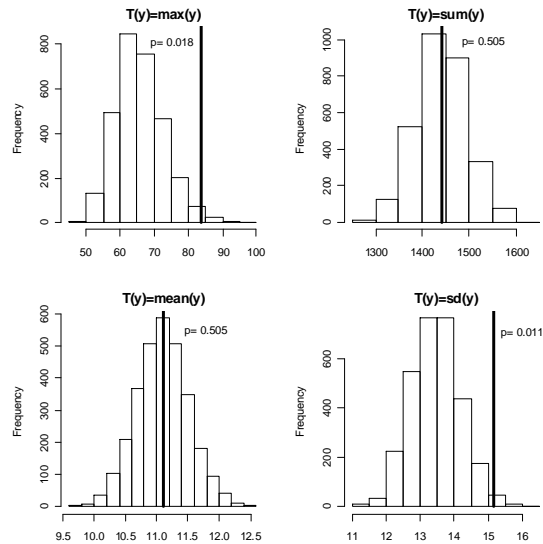


Figura 218| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo também não capta a dispersão existente nos dados, tal como se pode observar na Figura 219, onde $p_B=0.003$.

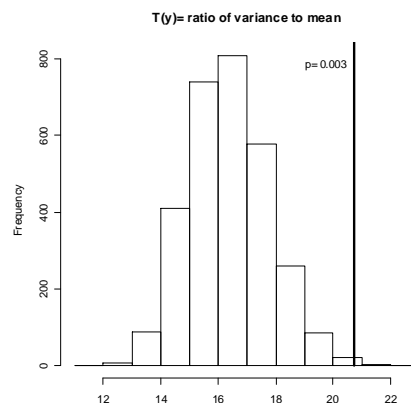


Figura 219| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson, tendo o número de feridos graves como variável resposta, obtém uma média das desviâncias amostrais, $\bar{D}=427.1$, o qual é consideravelmente superior a $n-p=126$. O valor de DIC obtido é 909.75.

7.3.2 Modelo de Poisson-Gama

O modelo hierárquico Poisson-Gama que obteve o menor valor de DIC quando ajustado aos dados referentes ao número de feridos graves inclui como covariáveis $\log(AADT)$, $\log(Length)$, $Left_Shoulder_Width$ e $Median_Type$. O número esperado de feridos graves, $\hat{\lambda}_i$ é dado pela seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Left_Shoulder_Width_i + \beta_4 Median_Type_i \quad (7.8)$$

Os gráficos das iterações das estimativas dos parâmetros β_j relativos às três cadeias de Markov, representados na Figura 220, mostram que as distribuições *a posteriori* para os parâmetros β_0 e β_1 não parecem convergir adequadamente para β_0 e β_1 .

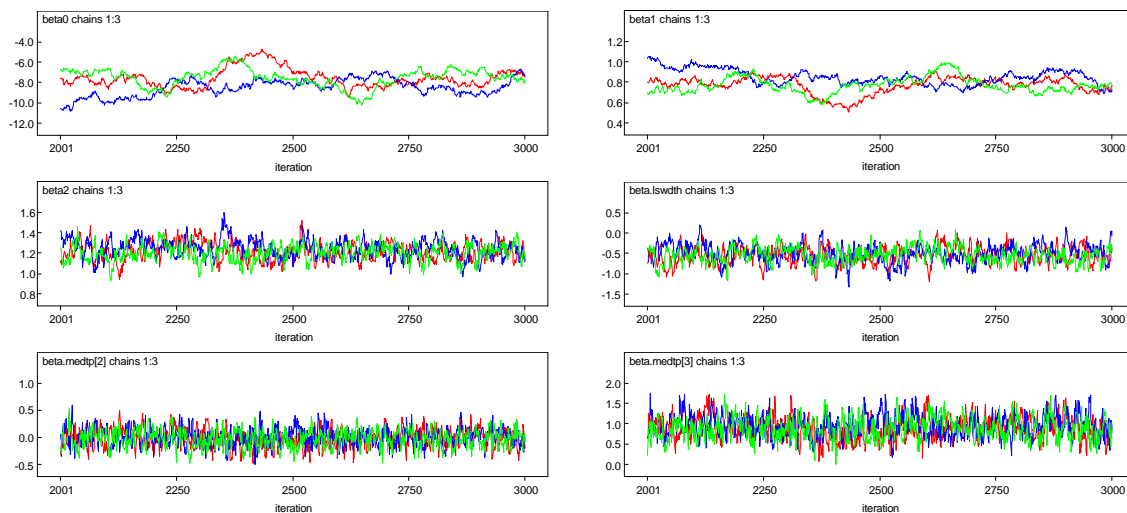


Figura 220| Trace plots (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a NSI obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações exibidas na Figura 221 mostram autocorrelações muito elevadas para as estimativas de β_0 e β_1 e uma convergência mais lenta para estes parâmetros.

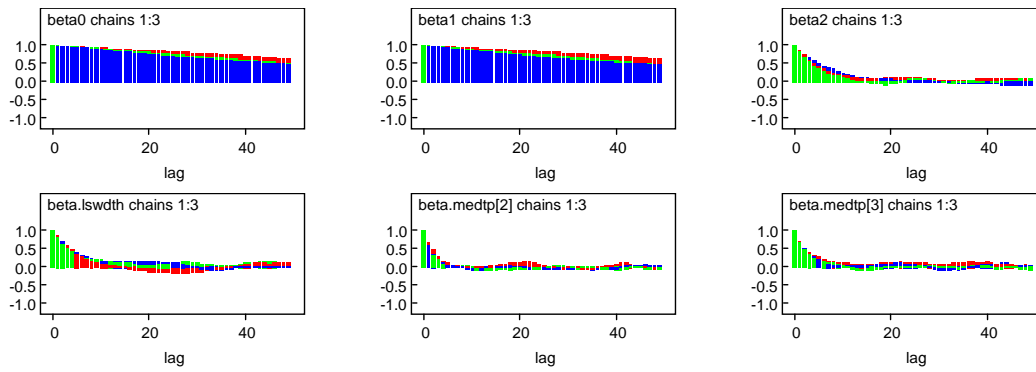


Figura 221| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser examinados na Figura 222, mostram que R (a vermelho) tende para 1 para todos os parâmetros β_j e que \hat{V} e W estabilizam à medida que aumentam as iterações, o que leva a crer que a convergência das cadeias de Markov é atingida.

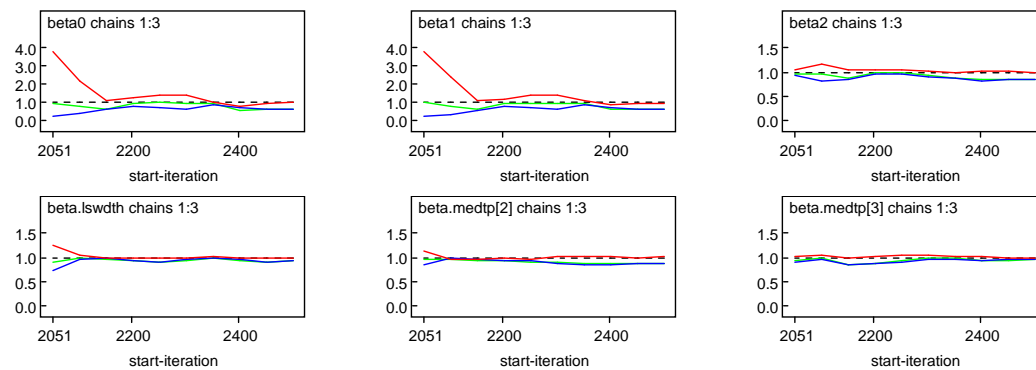


Figura 222| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os números esperados de feridos graves, para um período de cinco anos, após ajuste do modelo Poisson-Gama, são dados pelas equações do Quadro 95 referentes a trechos com os vários níveis das variáveis *Left_Shoulder_Width* e *Median_Type*.

Quadro 95| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	Números Esperados de Feridos Graves
<i>Left_Shoulder_Width='A'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 3.773 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$
'GM'	$\hat{\lambda}_i = 3.717 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$
'Vala'	$\hat{\lambda}_i = 9.350 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$
<i>Left_Shoulder_Width='B'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 2.218 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$
'GM'	$\hat{\lambda}_i = 2.185 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$
'Vala'	$\hat{\lambda}_i = 5.497 \times 10^{-4} \times AADT_i^{0.795} \times Length_i^{1.221}$

As estimativas obtidas pelo modelo são expostas no Quadro 96. Pela estimativa do parâmetro β_3 (relativo à covariável *Left_Shoulder_Width*) pode afirmar-se que a uma largura da berma esquerda de tipo 'B' (maior que 0.50 metros) corresponde um número de feridos graves cerca de 41% menor do que num trecho com uma largura de berma esquerda de tipo 'A' (menor que 0.50 metros). Um trecho com separador do tipo 'GM' ou 'Vala' tem menos cerca de 2% ou mais cerca de 148% no número de feridos graves, respectivamente, em relação a um trecho com separador do tipo 'DNJ;NJ'.

Quadro 96| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 95.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.882	1.003	9.982×10^{-2}	-9.855	-5.611
β_1 (log(AADT))	0.795	0.086	8.564×10^{-3}	0.596	0.964
β_2 (log(Length))	1.221	0.088	5.254×10^{-3}	1.048	1.391
β_3 (<i>Left_Shoulder_Width='B'</i>)	-0.531	0.211	1.211×10^{-2}	-0.950	-0.140
β_4 (<i>Median_Type='GM'</i>)	-0.015	0.163	5.863×10^{-3}	-0.337	0.308
β_5 (<i>Median_Type='Vala'</i>)	0.907	0.278	1.293×10^{-2}	0.390	1.478

As densidades *a posteriori* dos parâmetros β_j , ver Figura 223, mostram que as respectivas médias se afastaram de zero.

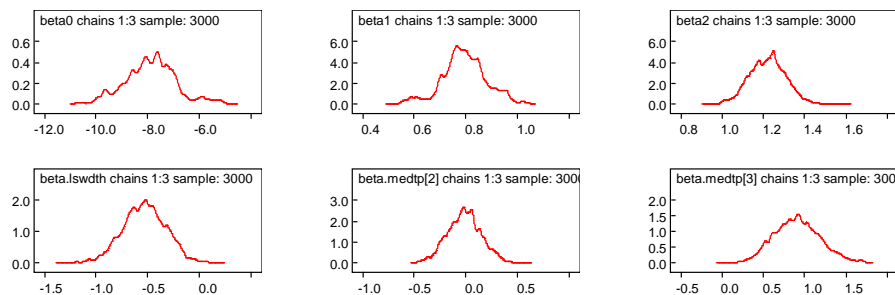


Figura 223| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a *NSI* (obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

O Quadro 97 contém os números esperados de feridos graves (para uma representação de um trecho com valores médios de *AADT* e *Length*) para os diversos níveis das covariáveis existentes no modelo.

Quadro 97| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 95.

<i>Left_Shoulder_Width='A'</i>	<i>média (desvio padrão)</i>	<i>Left_Shoulder_Width='B'</i>	<i>média (desvio padrão)</i>
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	9.083 (1.800)	'DNJ;NJ'	5.293 (0.764)
'GM'	8.978 (1.940)	'GM'	5.183 (0.480)
'Vala'	23.050 (7.066)	'Vala'	13.360 (3.286)

7.3.2.1 Validação do Modelo

Pela comparação dos histogramas na Figura 224 constata-se que a maioria dos dados observados são plausíveis sob a distribuição preditiva *a posteriori*.

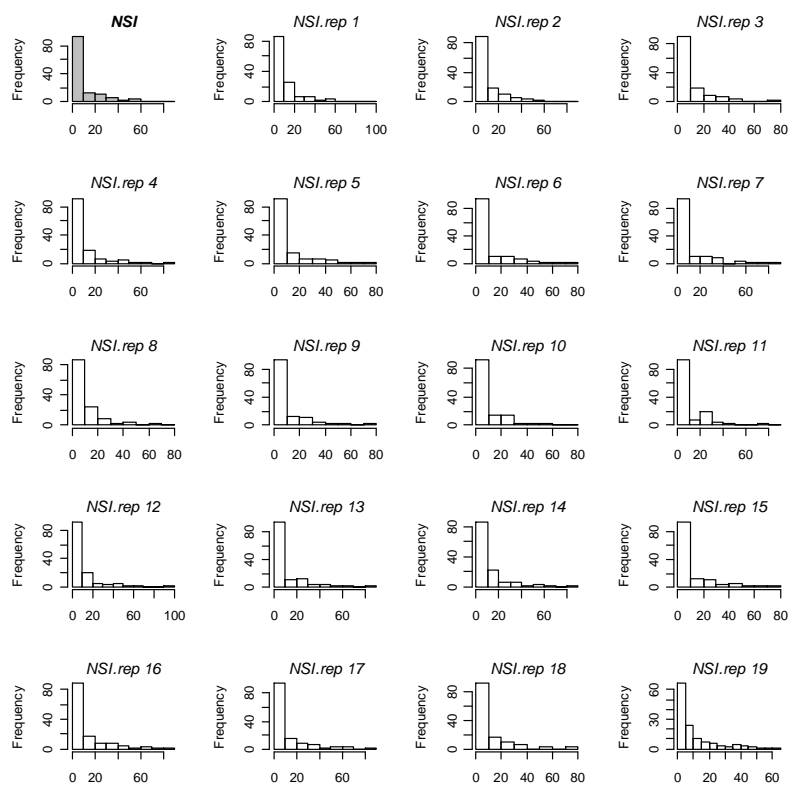


Figura 224| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas (“*NSI.rep*”) dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

A verificação das comparações gráficas das quantidades correspondentes aos dados com as quantidades obtidas de simulações preditivas *a posteriori* (ver Figura 225) leva a crer que o modelo ajustado é adequado.

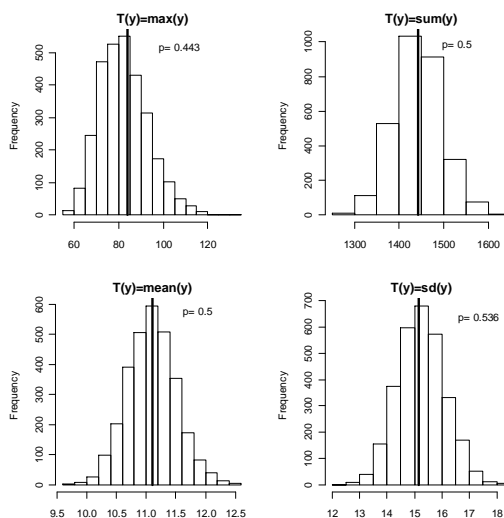


Figura 225| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A mesma conclusão é obtida pela leitura da Figura 226, que sugere que o modelo ajustado tem em consideração a dispersão existente nos dados observados.

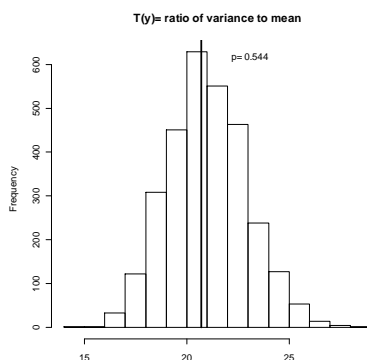


Figura 226| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O valor da média das desviâncias amostrais obtido pelo modelo é igual a $\bar{D}=118.3$, valor mais próximo de $n-p=126$, indicando um bom ajuste. O valor de DIC é 648.60 e a estimativa do parâmetro de dispersão é igual a 0.378.

7.3.3 Modelo de Poisson Log-Normal

O modelo hierárquico Poisson Log-Normal que possui o menor valor de DIC quando do ajuste aos dados referentes aos números de feridos graves é dado pela equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Left_Shoulder_Width}_i + \beta_4 \text{Median_Type}_i \quad (7.9)$$

Onde $\hat{\lambda}_i$ representa o número esperado de feridos graves.

O facto de as iterações das estimativas de β_0 e β_1 para as três cadeias de Markov não se interligarem, nem tão pouco parecerem convergir para um valor comum (ver os dois gráficos na primeira linha na Figura 227) leva a acreditar que as estimativas de β_0 e β_1 não parecerem convergir.

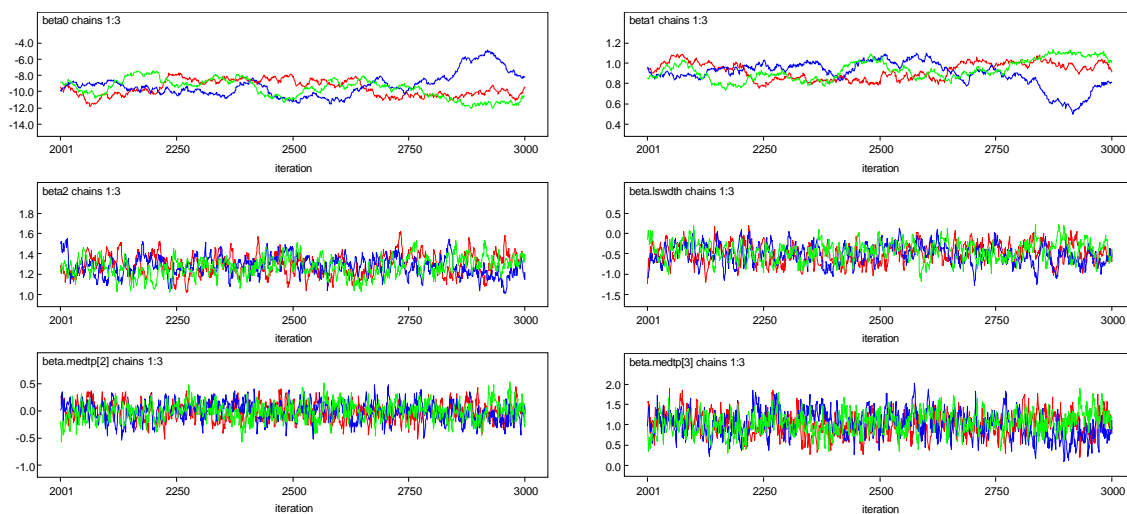


Figura 227| Trace plots (iterrações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a NSI obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

A mesma interpretação é deduzida pela leitura dos gráficos de autocorrelação exibidos na Figura 228.

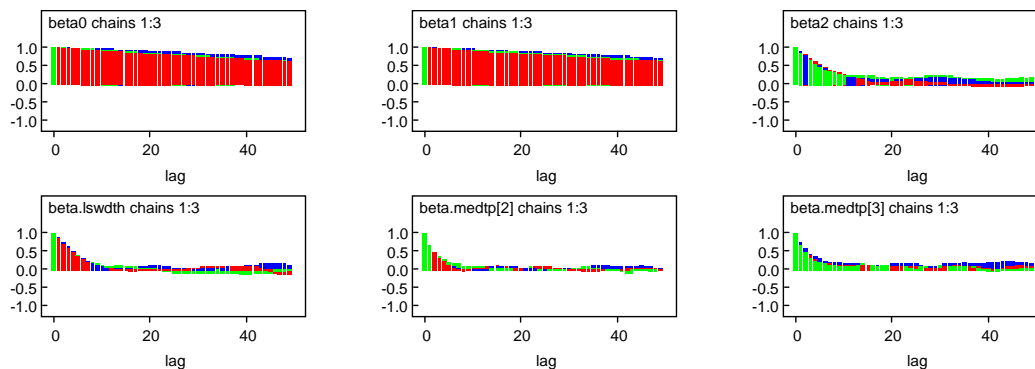


Figura 228| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NSI obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Pela observação dos gráficos das estatísticas de Gelman-Rubin apresentadas na Figura 229 depreende-se que, uma vez que R tende para 1 e as restantes estatísticas estabilizam à medida que aumenta o número de iterações, há razões para admitir a convergência das estimativas de todos os parâmetros β_j e por consequência do algoritmo considerado.

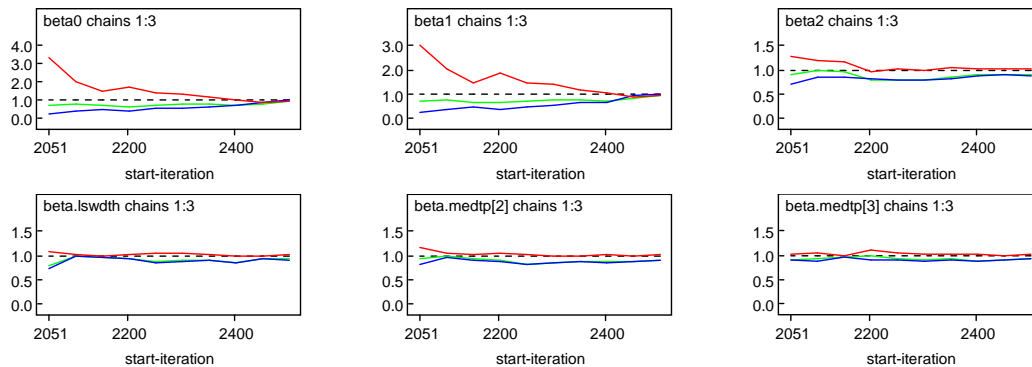


Figura 229| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NSI obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

No Quadro 98 podem ver-se as equações de onde se deduzem os números esperados de feridos graves (em cinco anos) para trechos de auto-estrada com e sem via adicional para o modelo Poisson Log-Normal representado pela equação genérica (7.9).

Quadro 98| Número esperado de feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	Números Esperados de Feridos Graves
<i>Left_Shoulder_Width='A'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 7.638 \times 10^{-5} \times AADT_i^{0.918} \times Length_i^{1.281}$
'GM'	$\hat{\lambda}_i = 7.509 \times 10^{-5} \times AADT_i^{0.918} \times Length_i^{1.281}$
'Vala'	$\hat{\lambda}_i = 2.120 \times 10^{-4} \times AADT_i^{0.918} \times Length_i^{1.281}$
<i>Left_Shoulder_Width='B'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 4.727 \times 10^{-5} \times AADT_i^{0.918} \times Length_i^{1.281}$
'GM'	$\hat{\lambda}_i = 4.647 \times 10^{-5} \times AADT_i^{0.918} \times Length_i^{1.281}$
'Vala'	$\hat{\lambda}_i = 1.312 \times 10^{-4} \times AADT_i^{0.918} \times Length_i^{1.281}$

As estimativas obtidas por este modelo podem ser observadas no Quadro 99. A leitura da média *a posteriori* de β_3 mostra que trechos cuja berma esquerda é superior a 0.50 metros (tipo 'B') têm uma redução no número esperado de feridos graves de cerca de 38% quando comparado a um trecho em iguais condições, mas com largura de berma esquerda inferior a 0.50 metros. Trechos com

separadores do tipo 'GM' têm uma redução no número de feridos graves de cerca de 2% em relação a trechos idênticos, mas com separador do tipo 'DNJ;NJ'. Trechos com separador do tipo 'Vala' têm um aumento do número esperado de feridos graves de cerca de 178% em relação a trechos com separador do tipo 'DNJ;NJ'.

Quadro 99| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 98.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.480	1.204	1.193×10^{-1}	-11.5105	-6.087
β_1 (log(AADT))	0.918	0.103	1.017×10^{-2}	0.634	1.092
β_2 (log(Length))	1.281	0.097	6.326×10^{-3}	1.094	1.472
β_3 (Left_Shoulder_Width='B')	-0.480	0.237	1.309×10^{-2}	-0.952	-0.023
β_4 (Median_Type='GM')	-0.017	0.168	7.401×10^{-3}	-0.356	0.310
β_5 (Median_Type='Vala')	1.021	0.289	1.482×10^{-2}	0.448	1.592

Os gráficos das densidades *a posteriori* obtidas para as estimativas dos parâmetros β_j estão exibidos na Figura 230. As densidades para β_0 e β_1 possuem inúmeros "picos" e caudas longas.

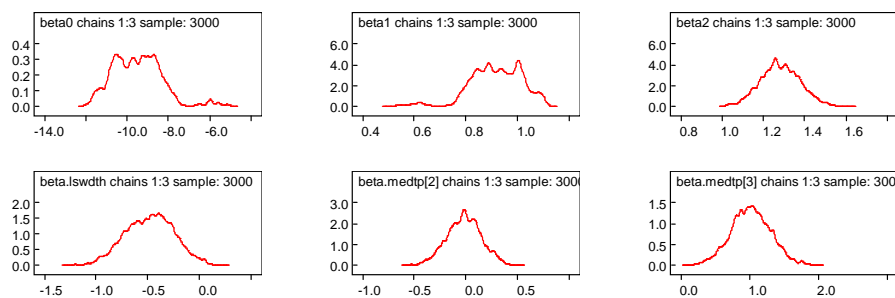


Figura 230| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a NSI (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

O Quadro 100 contém os números esperados de feridos graves para cada nível das variáveis categóricas, para um trecho com representações médias de $\log(\text{AADT})$ e $\log(\text{Length})$.

Quadro 100| Médias *a posteriori* e desvios padrões do número esperado de feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 98.

	média (desvio padrão)		média (desvio padrão)
Left_Shoulder_Width='A'		Left_Shoulder_Width='B'	
Median_Type		Median_Type	
'DNJ;NJ'	7.027 (1.551)	'DNJ;NJ'	4.300 (0.682)
'GM'	6.940 (1.671)	'GM'	4.195 (0.408)
'Vala'	19.920 (6.148)	'Vala'	12.140 (3.020)

7.3.3.1 Validação do Modelo

Pela observação da Figura 231 constata-se que os histogramas de dezanove conjuntos de dados replicados pelo modelo considerado (ver Quadro 98) assemelham-se bastante ao histograma dos dados observados, o que indica que o modelo se encontra bem ajustado.

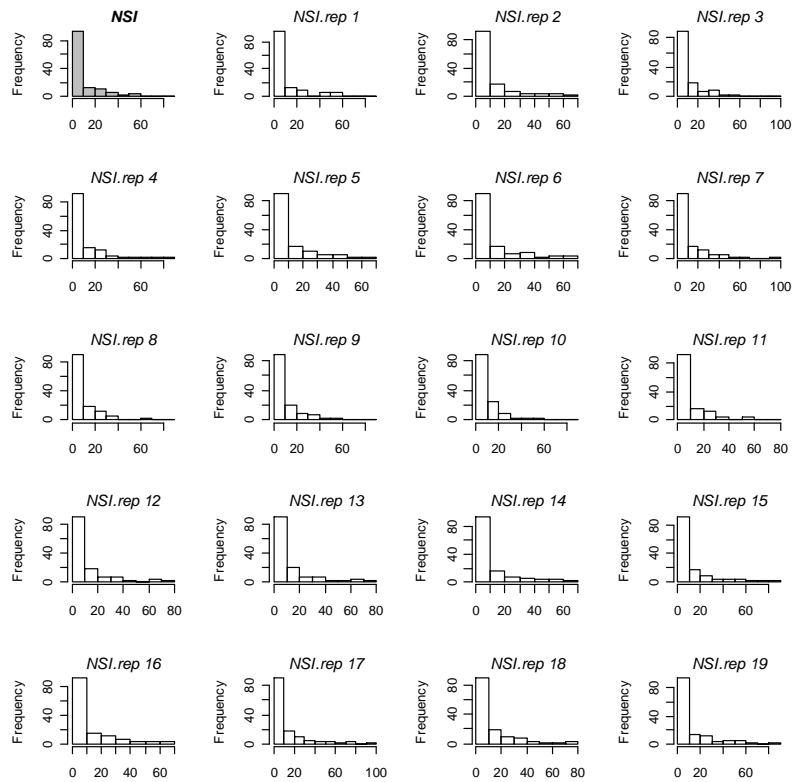


Figura 231| Histograma de *NSI* (no canto superior esquerdo) e dezanove réplicas (“*NSI.rep*”) dos números de feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

O modelo obtido capta as variações referentes às medidas *T* calculadas (ver Figura 232).

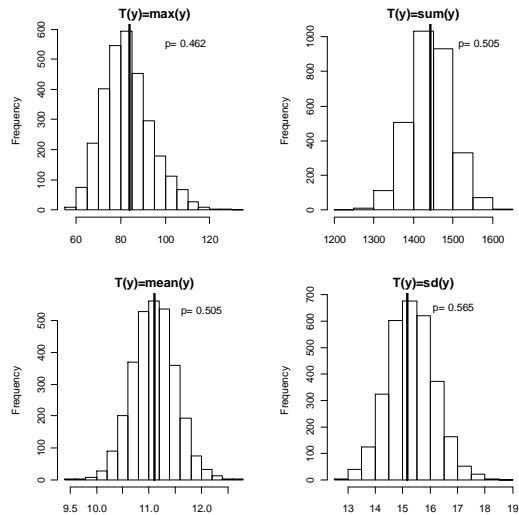


Figura 232| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

A razão entre a variância e o valor médio (ver Figura 233) é bem captada pelo modelo obtido quando se assume a distribuição Poisson Log-Normal.

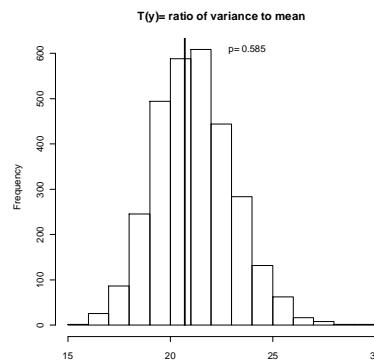


Figura 233| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo descrito nesta secção, e cujas equações para os números esperados de feridos graves estão apresentadas no Quadro 98, obteve um valor $\bar{D}=118.8$ ($n-p=126$), DIC igual a 653.57 e uma estimativa para o parâmetro de dispersão igual a 0.373.

7.4 Modelação do Número de Mortos e Feridos Graves

A descrição dos dados analisados nesta secção encontra-se no Capítulo 7, secção 7.4 em Azeredo Lopes e Cardoso (2009).

7.4.1 Modelo de Poisson

O modelo que obteve o menor valor de DIC, após se ter procedido a uma selecção de variáveis, e de um ajuste da distribuição de Poisson ao número de mortos e feridos graves tem a seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Left_Shoulder_Width_i + \beta_4 Median_Type_i \quad (7.10)$$

Pela observação dos gráficos na Figura 234 não parece haver razões para duvidar da convergência dos parâmetros β_2 a β_4 correspondentes à equação (7.10), pois as iterações de cada cadeia de Markov interligam-se e convergem para o mesmo valor. O mesmo não pode ser afirmado relativamente aos parâmetros β_0 e β_1 .

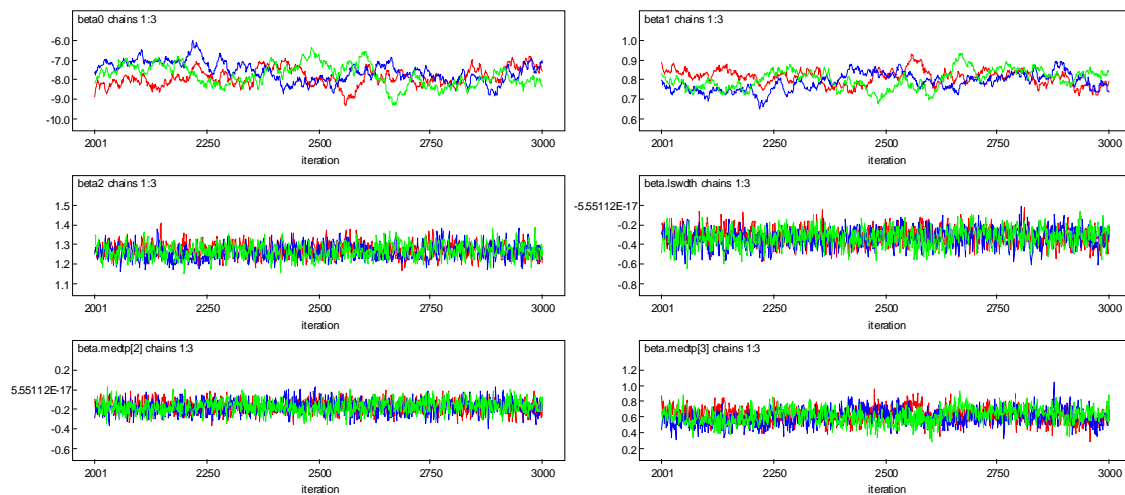


Figura 234| Trace plots (iterrações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson ajustado a NKS1 obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações na Figura 235 também não sugerem razões para duvidar da convergência, embora os parâmetros β_0 e β_1 convirjam mais lentamente do que as restantes variáveis.

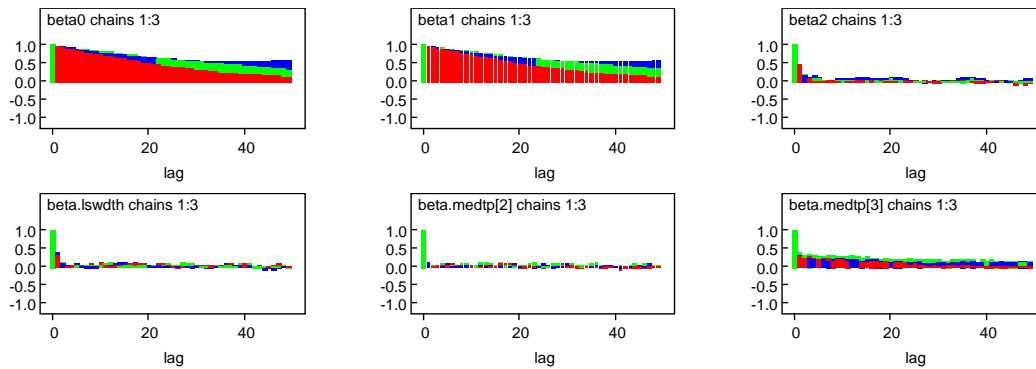


Figura 235| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson ajustado a NKS/ obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

As estatísticas de Gelman-Rubin exibidas na Figura 236 estabilizam e, em especial R , converge para 1. Não havendo assim razões para duvidar da convergência do algoritmo.

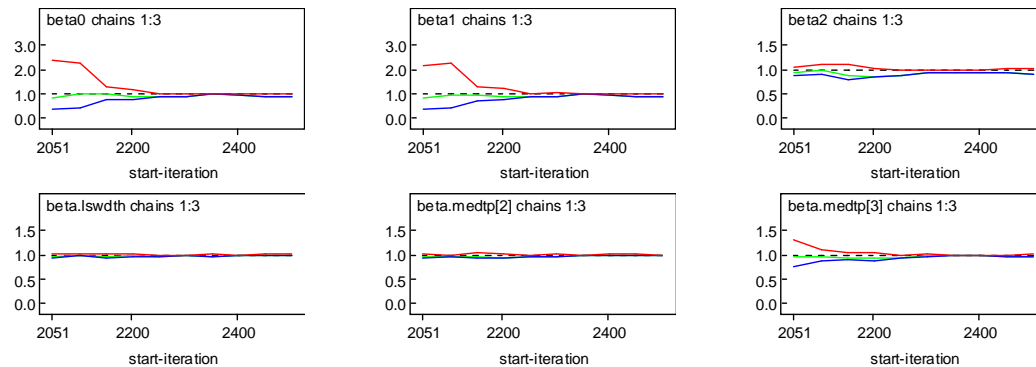


Figura 236| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson ajustado a NKS/ obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Quadro 101| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	<i>Números Esperados de Mortos e Feridos Graves</i>
<i>Left_Shoulder_Width='A'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 4.276 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$
'GM'	$\hat{\lambda}_i = 3.587 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$
'Vala'	$\hat{\lambda}_i = 7.880 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$
<i>Left_Shoulder_Width='B'</i>	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 3.078 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$
'GM'	$\hat{\lambda}_i = 2.582 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$
'Vala'	$\hat{\lambda}_i = 5.672 \times 10^{-4} \times AADT_i^{0.800} \times Length_i^{1.267}$

As equações dos números esperados de mortos e feridos graves são exibidas no Quadro 101 para cada nível das variáveis categóricas.

As estimativas do modelo obtido encontram-se no Quadro 102. O modelo ajustado permite afirmar que um trecho de auto-estrada com largura da berma esquerda superior a 0.50 metros (tipo 'B'), tem um número esperado de mortos e feridos graves reduzido em cerca de 28% quando comparado com um trecho com largura de berma esquerda inferior a 0.50 metros (tipo 'A') mas em idênticas situações ao anterior. Um trecho com separadores do tipo 'GM' e 'Vala' têm redução, e aumento, respectivamente, no número esperado de mortos e feridos graves de cerca de 16% e 84% em relação a um trecho com separador do tipo 'DNJ;NJ'.

Quadro 102| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 101.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.757	0.513	4.768×10^{-2}	-8.714	-6.782
β_1 (log(AADT))	0.800	0.044	4.029×10^{-3}	0.716	0.882
β_2 (log(Length))	1.267	0.037	1.387×10^{-3}	1.198	1.339
β_3 (Left_Shoulder_Width='B')	-0.329	0.088	2.902×10^{-3}	-0.504	-0.157
β_4 (Median_Type='GM')	-0.176	0.064	1.913×10^{-3}	-0.301	-0.050
β_5 (Median_Type='Vala')	0.611	0.098	5.209×10^{-3}	0.419	0.803

As densidades *a posteriori* dos parâmetros β_j estão representadas na Figura 237. Os valores médios de β_j afastaram-se consideravelmente de zero.

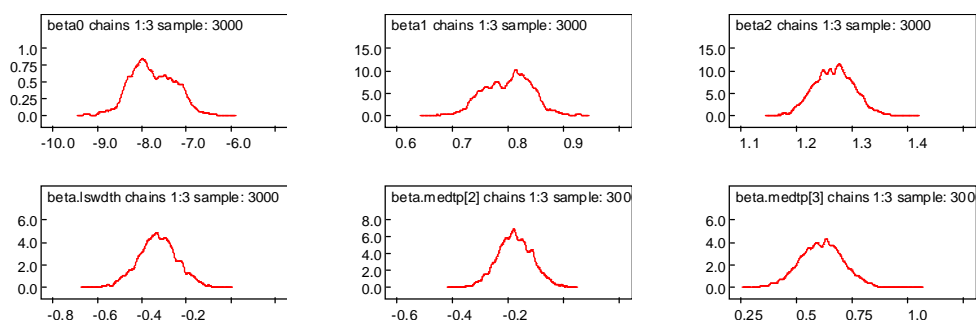


Figura 237| Densidades *a posteriori* dos parâmetros β_j do modelo de Poisson ajustado a NKS/ (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves para uma representação média de AADT e Length, e para um período de cinco anos, estão apresentados no Quadro 103 para os diversos níveis das covariáveis. Depreende-se que os valores mais elevados de mortos e feridos graves são esperados em trechos com largura de berma esquerda inferior a 0.50 metros (tipo 'A'). Trechos com separadores do tipo 'Vala' têm o maior número esperado de mortos e feridos graves para valores médios de AADT

e *Length*. Os trechos com separadores do tipo 'GM' são os que apresentam o menor valor de mortos e feridos graves.

Quadro 103| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 101.

<i>Left_Shoulder_Width</i> ='A'	média (desvio padrão)	<i>Left_Shoulder_Width</i> ='B'	média (desvio padrão)
<i>Median_Type</i>		<i>Median_Type</i>	
'DNJ;NJ'	11.430 (0.933)	'DNJ;NJ'	8.214 (0.532)
'GM'	9.591 (0.873)	'GM'	6.882 (0.307)
'Vala'	21.080 (2.046)	'Vala'	15.150 (1.167)

7.4.1.1 Validação do Modelo

A comparação dos histogramas obtidos de dados replicados da distribuição preditiva *a posteriori* com o histograma dos dados observados (ver Figura 238) mostra haver algumas discrepâncias no modelo.

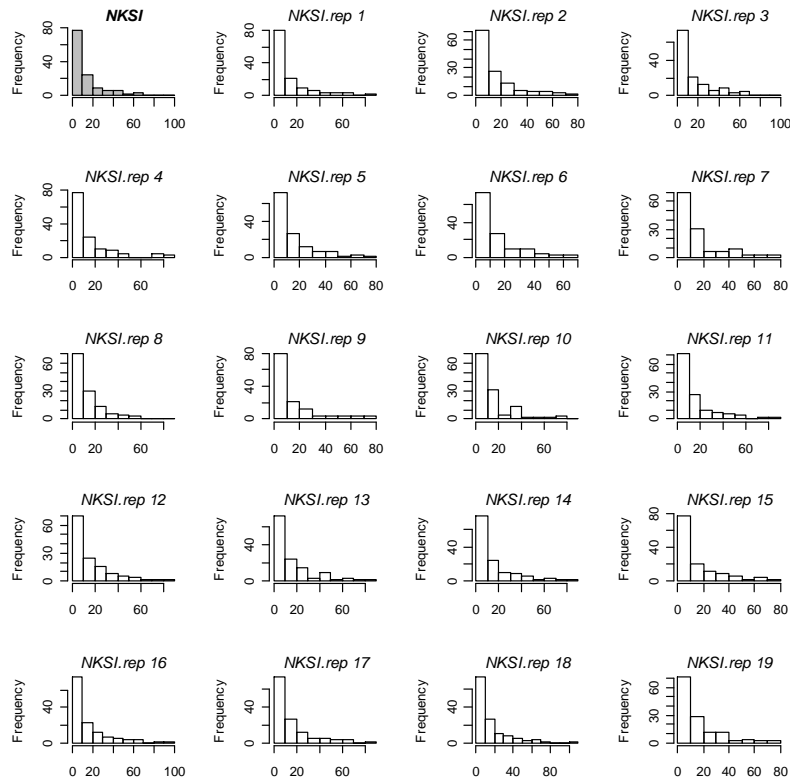


Figura 238| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas ("*NKSI.rep*") dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo de Poisson ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Pela leitura da Figura 239 verifica-se que o modelo considerado não reproduz convenientemente o valor máximo e o desvio padrão do conjunto formado pelos dados. O modelo também não capta a dispersão dos dados como pode ser observado na Figura 240.

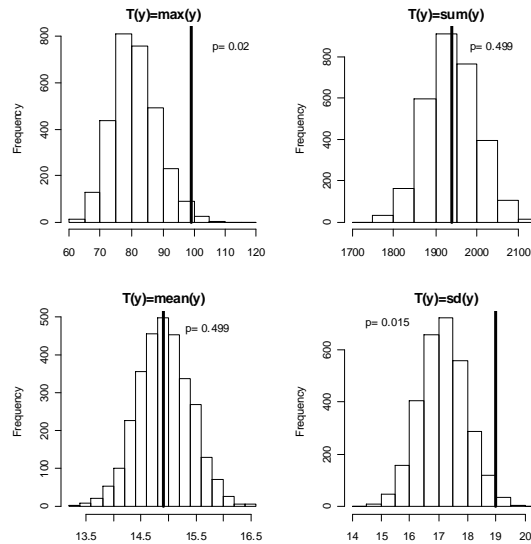


Figura 239| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo de Poisson. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O valor do p -value Bayesiano correspondente à razão entre a variância e a média é de somente 0.002.

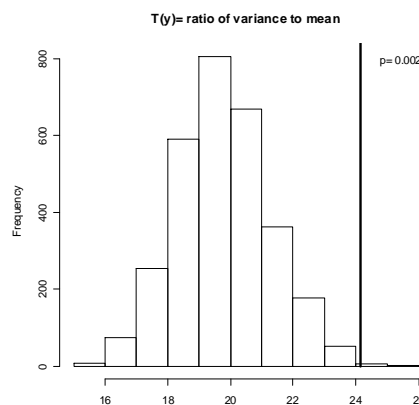


Figura 240| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo de Poisson representado pelas equações no Quadro 101 produz $\bar{D}=482.3$ (valor bastante elevado quando comparado a $n-p=126$, o que indica que este modelo não capta a sobredispersão existente nos dados) e DIC igual a 1005.32.

7.4.2 Modelo de Poisson-Gama

O modelo Poisson-Gama cujo ajuste aos dados obteve o menor valor de DIC tem a seguinte equação genérica:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(Length_i) + \beta_3 Left_Shoulder_Width_i + \beta_4 Median_Type_i \quad (7.11)$$

Os gráficos das iterações na Figura 241 mostram haver razões para duvidar da convergência das estimativas de β_0 e β_1 .

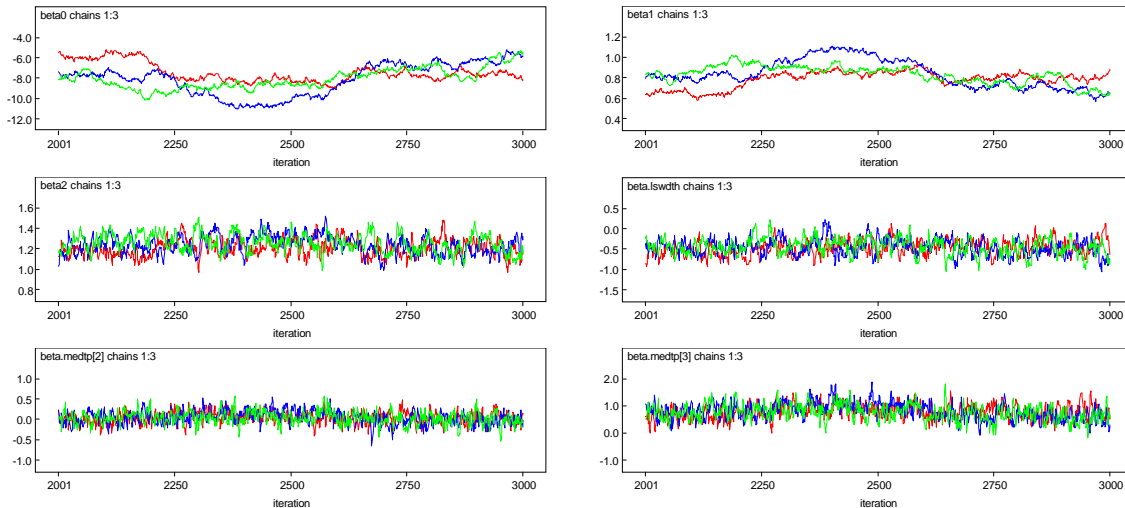


Figura 241| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

As autocorrelações (ver Figura 242) evidenciam dificuldades na convergência dos parâmetros β_0 , β_1 .

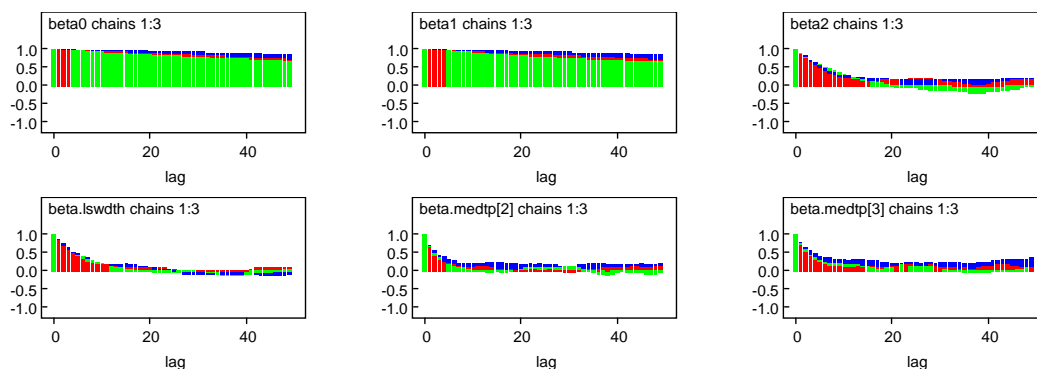


Figura 242| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson-Gama ajustado a *NKSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

As estatísticas de Gelman-Rubin, cujos gráficos podem ser observados na Figura 243, mostram que, à medida que aumentam os números de iterações, as três estatísticas estabilizam e R tende para 1.

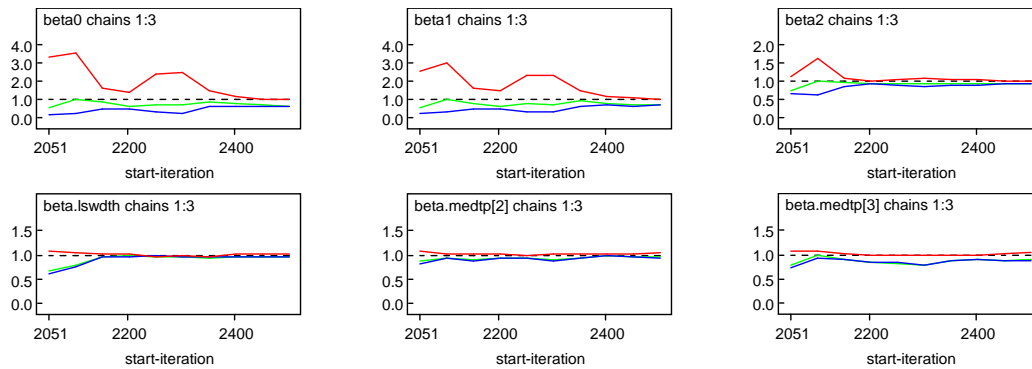


Figura 243| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson-Gama ajustado a NKS/ obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

As equações no Quadro 104 dão o número esperado de mortos e feridos graves obtidos pelo modelo Poisson-Gama, para um período de cinco anos, e para valores de *AADT*, *Length*, *Left_Shoulder_Width* e *Median_Type*.

Quadro 104| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson-Gama ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	<i>Números Esperados de Mortos e Feridos Graves</i>
<i>Left_Shoulder_Width</i> ='A'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 3.741 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$
'GM'	$\hat{\lambda}_i = 3.870 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$
'Vala'	$\hat{\lambda}_i = 8.074 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$
<i>Left_Shoulder_Width</i> ='B'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 2.362 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$
'GM'	$\hat{\lambda}_i = 2.443 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$
'Vala'	$\hat{\lambda}_i = 5.098 \times 10^{-4} \times AADT_i^{0.818} \times Length_i^{1.232}$

As estimativas para o modelo obtido encontram-se no Quadro 105. Um trecho com largura da berma esquerda superior a 0.50 metros (tipo 'B') tem um número esperado de mortos e feridos graves (em cinco anos), cerca de 37% menor do que um trecho em iguais circunstâncias, mas com largura da berma esquerda inferior a 0.50 metros (tipo 'A').

Um trecho com separador do tipo 'GM' tem um número esperado de mortos e feridos graves cerca de 3% menor do que um trecho com separador do tipo 'DNJ;NJ'. Por outro lado, num trecho de auto-estrada com separador do tipo 'Vala' existe uma ocorrência de mortos e feridos graves cerca de 116% superior a um trecho com separador do tipo 'DNJ;NJ'.

Quadro 105| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 104.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-7.891	1.245	1.263×10^{-1}	-10.670	-5.594
β_1 (log(AADT))	0.818	0.107	1.079×10^{-2}	0.627	1.066
β_2 (log(Length))	1.232	0.089	6.132×10^{-3}	1.073	1.409
β_3 (Left_Shoulder_Width='B')	-0.460	0.194	1.162×10^{-2}	-0.834	-0.078
β_4 (Median_Type='GM')	-0.034	0.157	7.839×10^{-3}	-0.283	0.341
β_5 (Median_Type='Vala')	0.769	0.289	1.728×10^{-2}	0.217	1.346

As densidades *a posteriori* dos parâmetros β_j encontram-se na Figura 244. Observa-se em particular as formas peculiares das densidades correspondentes a β_0 e β_1 .

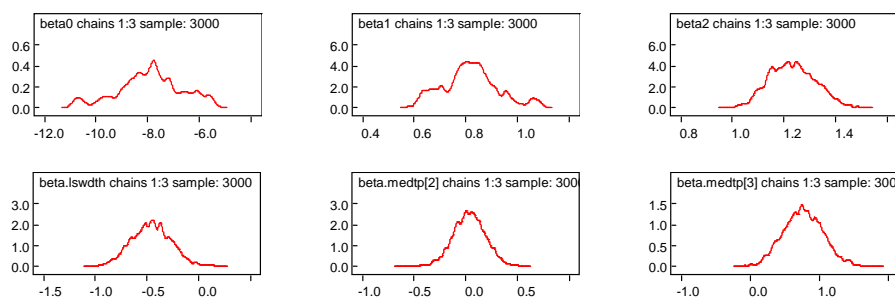


Figura 244| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson-Gama ajustado a NKS (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves obtidos de valores médios de AADT e Length encontram-se no Quadro 70 para os dois níveis da variável Left_Shoulder_Width e para os três níveis da variável Median_Type. Mais uma vez se constata um menor número esperado de mortos e feridos graves em trechos com largura de berma esquerda superior a 0.50 metros (tipo 'B').

Quadro 106| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 104.

	média (desvio padrão)	Left_Shoulder_Width='B'	média (desvio padrão)
Left_Shoulder_Width='A'			
Median_Type		Median_Type	
'DNJ;NJ'	11.580 (2.309)	'DNJ;NJ'	7.240 (1.028)
'GM'	11.970 (2.344)	'GM'	7.443 (0.638)
'Vala'	25.510 (7.510)	'Vala'	15.940 (3.963)

7.4.2.1 Validação do Modelo

De acordo com a comparação sugerida por Gelman *et al.* (2004), o exame dos histogramas da Figura 245 mostra que o histograma dos dados observados (a cinzento na figura) é verosímil sob a distribuição preditiva *a posteriori*.

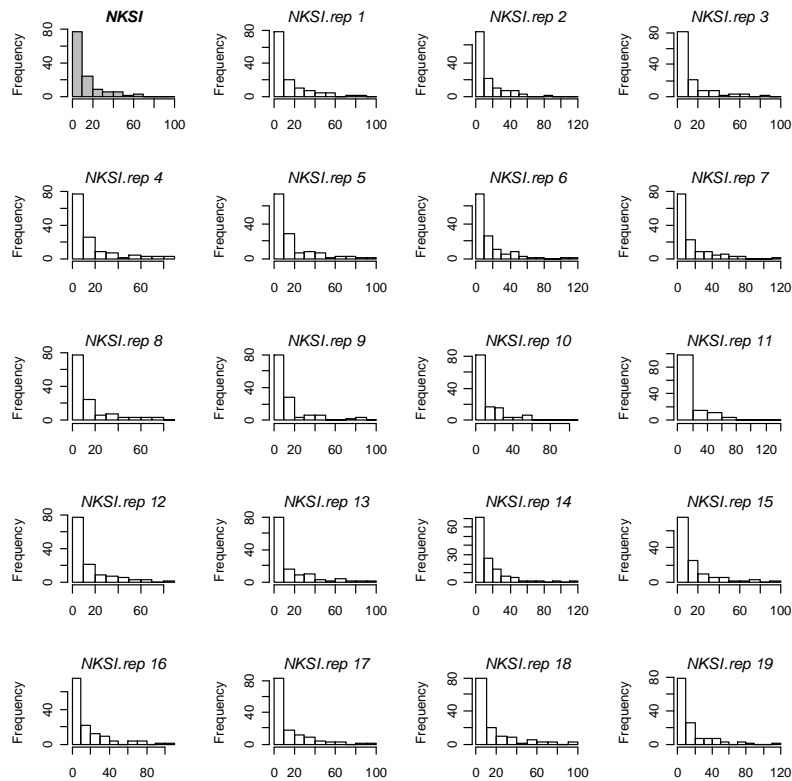


Figura 245| Histograma de *NKSI* (no canto superior esquerdo) e dezanove réplicas (“*NKSI.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson-Gama ajustado ao conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Pela interpretação dos histogramas na Figura 246 e na Figura 247 pode observar-se que o modelo captura adequadamente as quantidades *T* calculadas.

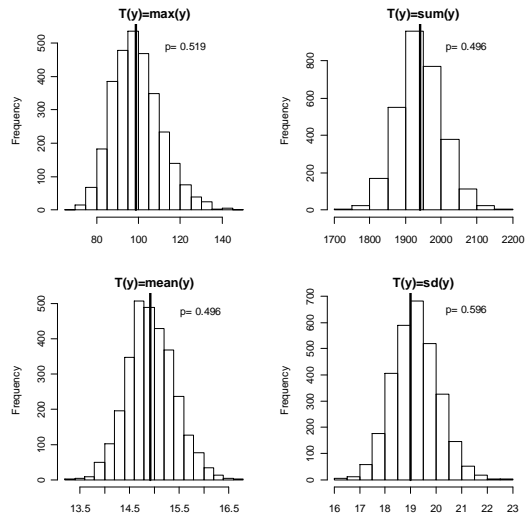


Figura 246| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson-Gama. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

Os p -values Bayesianos de todas as quantidades são próximos de 0.5.

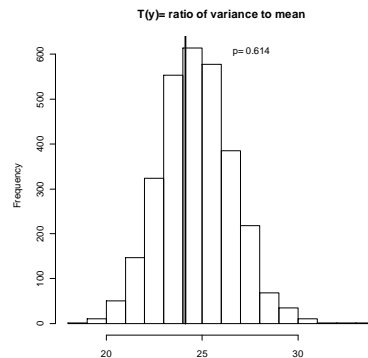


Figura 247| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição *a posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson-Gama e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo Poisson-Gama produz $\bar{D}=124.8$, valor que indica um ajuste razoável uma vez que é próximo de $n-p=126$. O valor de DIC é igual a 702.90 e o parâmetro de dispersão é estimado em 0.350.

7.4.3 Modelo de Poisson Log-Normal

Tal como para a distribuição Poisson-Gama, o modelo baseado na distribuição Poisson Log-Normal ajustado aos dados referentes aos números de mortos e feridos graves que produziu o menor valor de DIC é representado pela equação:

$$\log(\hat{\lambda}_i) = \beta_0 + \beta_1 \log(AADT_i) + \beta_2 \log(\text{Length}_i) + \beta_3 \text{Left_Shoulder_Width}_i + \beta_4 \text{Median_Type}_i \quad (7.12)$$

Os gráficos das iterações de cada parâmetro β_j ($j=1, \dots, 3$) na Figura 248 mostra haver evidência de não convergência de β_0 e β_1 .

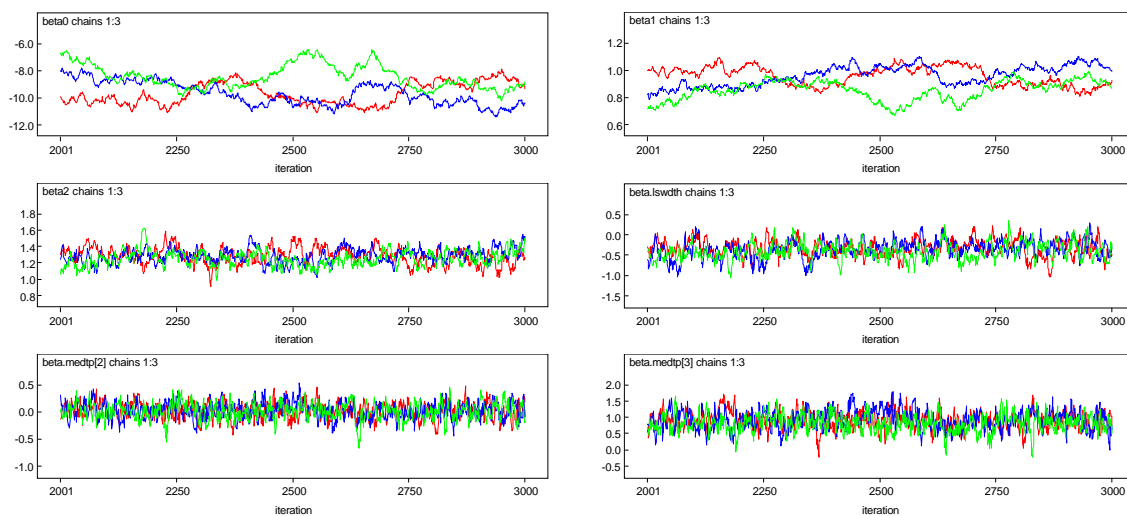


Figura 248| *Trace plots* (iterações vs valores gerados) correspondentes a três cadeias de Markov para cada parâmetro β_j do modelo Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

Os gráficos das autocorrelações mostram também evidência de dificuldades na convergência de β_0 e β_1 (ver Figura 249).

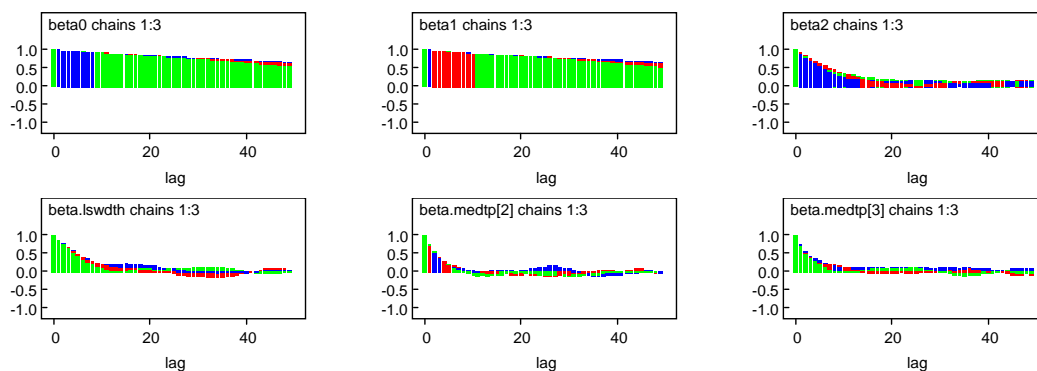


Figura 249| Gráficos das autocorrelações correspondentes a três cadeias de Markov para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a *NKSI* obtido do conjunto de dados cujos trechos têm *AADT* maior que 5000 e não possuem valores omissos.

No entanto, as estatísticas de Gelman-Rubin mostram que os quatro parâmetros β_j convergem, como pode ser observado nos gráficos na Figura 250.

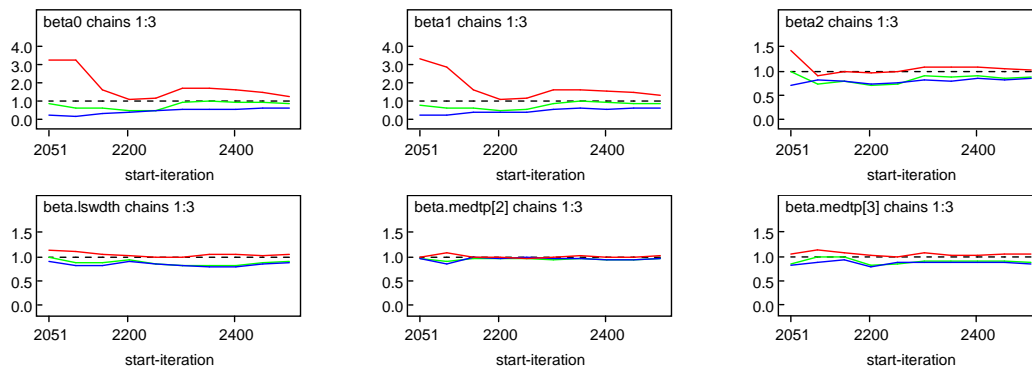


Figura 250| Estatísticas de Gelman-Rubin, correspondentes a três cadeias de Markov, para os parâmetros β_j do modelo de Poisson Log-Normal ajustado a NKSJ obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

As equações para o número esperado de mortos e feridos graves para um período de cinco anos e para os níveis das variáveis *Left_Shoulder_Width* e *Median_Type* são apresentadas no Quadro 107.

Quadro 107| Número esperado de mortos e feridos graves, para um período de cinco anos, obtido após ajuste de um modelo de regressão de Poisson Log-Normal ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

	Números Esperados de Mortos e Feridos Graves
<i>Left_Shoulder_Width</i> ='A'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 9.349 \times 10^{-5} \times AADT_i^{0.923} \times Length_i^{1.275}$
'GM'	$\hat{\lambda}_i = 9.515 \times 10^{-5} \times AADT_i^{0.923} \times Length_i^{1.275}$
'Vala'	$\hat{\lambda}_i = 2.250 \times 10^{-4} \times AADT_i^{0.923} \times Length_i^{1.275}$
<i>Left_Shoulder_Width</i> ='B'	
<i>Median_Type</i>	
'DNJ;NJ'	$\hat{\lambda}_i = 6.481 \times 10^{-5} \times AADT_i^{0.923} \times Length_i^{1.275}$
'GM'	$\hat{\lambda}_i = 6.596 \times 10^{-5} \times AADT_i^{0.923} \times Length_i^{1.275}$
'Vala'	$\hat{\lambda}_i = 1.559 \times 10^{-4} \times AADT_i^{0.923} \times Length_i^{1.275}$

As estimativas do modelo estão incluídas no Quadro 108. A estimativa do parâmetro correspondente a *Left_Shoulder_Lane* sugere que um trecho de auto-estrada com largura de berma esquerda superior a 0.50 metros (tipo 'B') tem uma redução no número de mortos e feridos graves de cerca de 31% em relação a um trecho em idêntica situação, mas com largura da berma esquerda inferior a 0.50 metros (tipo 'A').

Quadro 108| Estimativas pontuais e intervalos de credibilidade de 95% para os parâmetros do modelo exibido no Quadro 107.

Parâmetros	média	desvio padrão	erros de MC	2.5%	97.5%
β_0 (intersecção)	-9.278	1.025	1.032×10^{-1}	-10.9270	-6.941
β_1 (log(AADT))	0.923	0.087	8.720×10^{-3}	0.729	1.066
β_2 (log(Length))	1.275	0.096	6.828×10^{-3}	1.093	1.472
β_3 (Left_Shoulder_Width='B')	-0.366	0.214	1.313×10^{-2}	-0.803	0.047
β_4 (Median_Type='GM')	0.018	0.155	6.214×10^{-3}	-0.299	0.312
β_5 (Median_Type='Vala')	0.878	0.283	1.460×10^{-2}	0.322	1.451

Trechos com separadores de tipos 'GM' e 'Vala' têm aumentos no número de mortos e feridos graves de cerca de 2% e 141%, respectivamente, em relação a trechos em circunstâncias idênticas, mas com separador do tipo 'DNJ;NJ'.

As densidades *a posteriori* das estimativas dos parâmetros β_j encontram-se na Figura 251. É de reparar nas formas peculiares das densidades correspondentes a β_0 e β_1 .

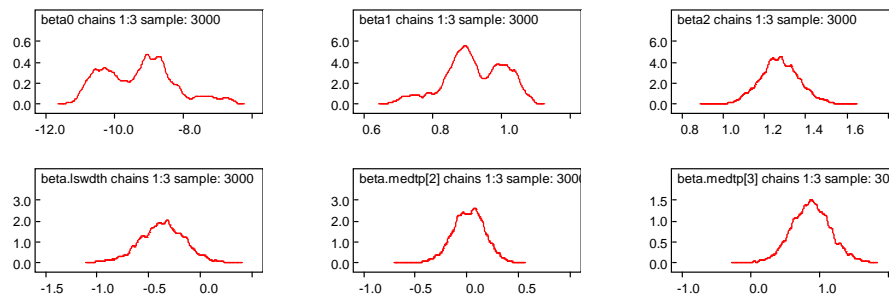


Figura 251| Densidades *a posteriori* dos parâmetros β_j do modelo Poisson Log-Normal ajustado a *NKSI* (obtido do conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos) resultantes da simulação de três cadeias de Markov.

Os números esperados de mortos e feridos graves para representações médias de *AADT* e *Length*, para um período de cinco anos e de acordo com os níveis das covariáveis, são apresentados no Quadro 109.

Quadro 109| Médias *a posteriori* e desvios padrões do número esperado de mortos e feridos graves, em cinco anos, para uma representação média de um trecho de auto-estrada, obtidas do modelo exibido no Quadro 107.

	média (desvio padrão)	Left_Shoulder_Width='B'	média (desvio padrão)
Left_Shoulder_Width='A'			
Median_Type		Median_Type	
'DNJ;NJ'	8.926 (1.714)	'DNJ;NJ'	6.139 (0.895)
'GM'	9.135 (2.017)	'GM'	6.209 (0.549)
'Vala'	22.030 (6.644)	'Vala'	15.050 (3.670)

7.4.3.1 Validação do Modelo

A comparação dos histogramas na Figura 252 deixa crer que o histograma dos dados observados é razoável sob a distribuição preditiva *a posteriori* obtida pelo modelo considerado.

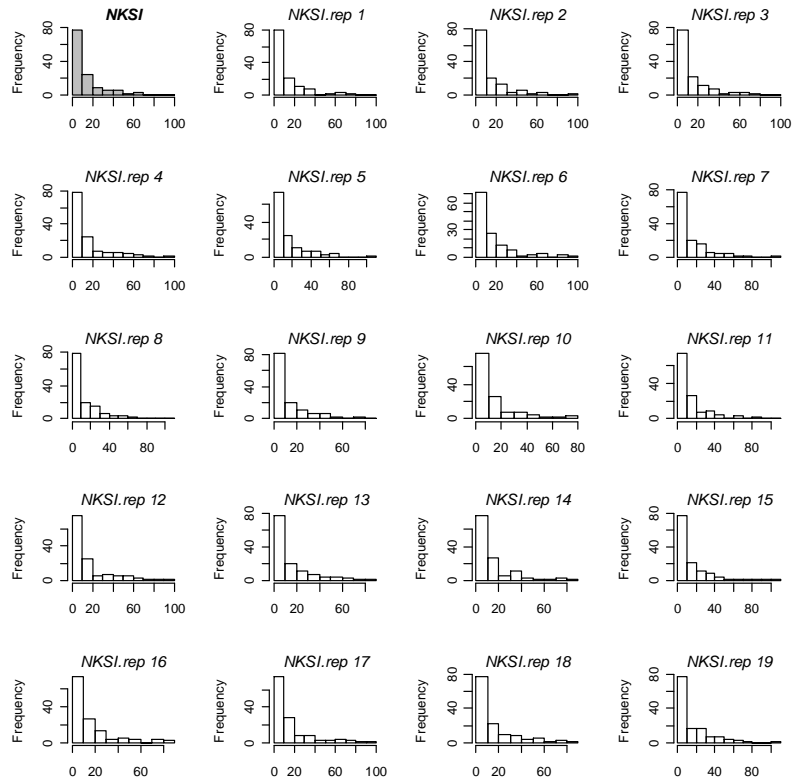


Figura 252| Histograma de *NKSÍ* (no canto superior esquerdo) e dezanove réplicas (“*NKSÍ.rep*”) dos números de mortos e feridos graves obtidos da distribuição preditiva *a posteriori* segundo o modelo Poisson Log-Normal ajustado ao conjunto de dados cujos trechos têm AADT maior que 5000 e não possuem valores omissos.

Os gráficos na Figura 253 das medidas de discrepância calculadas mostram que o modelo considerado capta essas quantidades, isto é, os valores máximos, soma, média e desvio padrão. Em todas as quantidades o *p*-value Bayesiano, p_B é aproximadamente igual a 0.5.

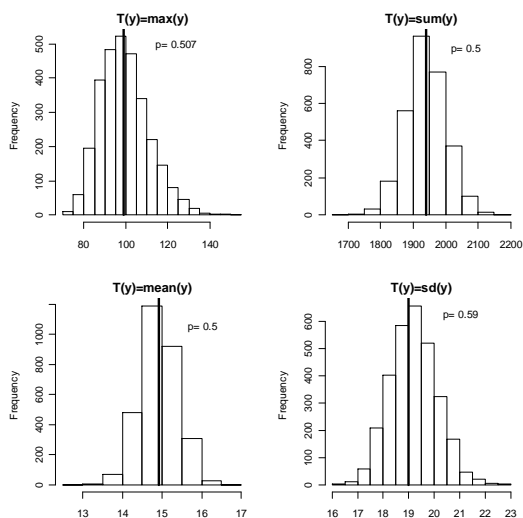


Figura 253| Histogramas dos valores T : máximo, soma, média e desvio padrão dos dados replicados referentes ao número de mortos e feridos graves obtidos pelo modelo Poisson Log-Normal. As rectas verticais representam as mesmas quantidades obtidas pelos dados observados. O p é a probabilidade de T , obtido pelos dados replicados, ser superior ao T obtido pelos dados observados.

O modelo Poisson Log-Normal ajustado ao número de mortos e feridos graves também capta a dispersão dos dados, como se pode observar no histograma e na recta vertical representados na Figura 254.

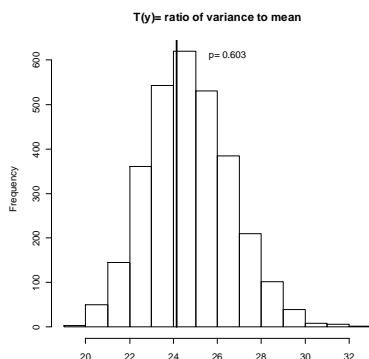


Figura 254| Razão entre a variância e a média dos dados observados (recta vertical) comparada com 3000 simulações da distribuição a *posteriori* preditiva da mesma razão obtida nos dados replicados do modelo de Poisson Log-Normal e a probabilidade de que a razão referente aos dados replicados seja superior à razão dos dados observados.

O modelo obteve uma média das desviâncias amostrais, \bar{D} , igual a 128.10 ($n-p=126$), DIC=709.92 e uma estimativa do parâmetro de dispersão igual a 0.335.

7.5 Discussão

No presente capítulo analisaram-se dados referentes a trechos de auto-estradas que registaram valores de *AADT* superiores a 5000 veículos e que não possuíam valores omissos em quaisquer das variáveis registadas. Como variável de resposta foram considerados *Nac*, *NF*, *NSI* e *NKSI*. O modelo de regressão de Poisson e modelos hierárquicos baseados nas distribuições Poisson-Gama e Poisson Log-Normal foram ajustados aos dados usando o *software* WinBUGS. A selecção das variáveis foi feita com base nos valores de DIC, sendo escolhido o modelo que apresentasse o menor valor desta quantidade. Foram feitas também verificações da convergência dos algoritmos correspondentes às cadeias de Markov para cada um dos modelos finais.

O modelo de Poisson mostrou não ser apropriado para modelar estes tipos de eventos pois assume que não haja sobredispersão nos dados. Os modelos hierárquicos baseados nas distribuições Poisson-Gama e Poisson Log-Normal reflectem mais adequadamente a distribuição dos dados observados. Estes dois modelos apresentaram resultados muito semelhantes, no entanto, os valores de DIC obtidos foram ligeiramente menores para os modelos baseados na distribuição Poisson-Gama. As conclusões apresentadas na presente secção foram baseadas nesta distribuição.

O número de acidentes é influenciado por $\log(AADT)$ e $\log(Length)$ assim como o tipo de separador existente no trecho de auto-estrada, *Median_Type*. Pode afirmar-se que um trecho com separadores dos tipos 'GM' (guarda metálica) e 'Vala' têm maior número de acidentes, de 4% e 92%, respectivamente, em relação a trechos em circunstâncias idênticas, mas com separador do tipo 'DNJ;NJ' (duplo New Jersey e New Jersey).

Relativamente ao número de mortos pôde constatar-se que não houve nenhuma variável categórica que fosse incluída no modelo final, o que leva a afirmar que este tipo de evento depende somente de $\log(AADT)$ e $\log(Length)$. O modelo final relativo ao número de feridos graves inclui as variáveis categóricas *Left_Shoulder_Width* e *Median_Type*. Constatou-se que os trechos com largura da berma esquerda superior a 0.50 metros têm uma redução no número de feridos graves em cerca de 41% em relação a trechos, em iguais condições, mas com largura da berma esquerda inferior a 0.50 metros. Trechos com separador do tipo 'GM' têm uma redução no número de feridos graves de cerca de 2% em relação a trechos em iguais condições mas com separador do tipo 'DNJ;NJ'. Trechos com separadores do tipo 'Vala' têm um aumento de aproximadamente 148% em relação a trechos em iguais condições mas com separador 'DNJ;NJ'.

Mais uma vez, também se faz notar que o número de vítimas não é uma variável independente pois depende da taxa de ocupação dos veículos.

O objectivo do estudo descrito no presente documento foi a obtenção de modelos de previsão de vários eventos relacionados com acidentes de tráfego, através de métodos estatísticos Bayesianos, por oposição aos métodos *clássicos* já usados em Azeredo Lopes e Cardoso (2009) para o mesmo conjunto de dados.

O conjunto de dados analisados consistiu em medições bidireccionais registadas durante um período de cinco anos (1999 a 2003) em trechos de auto-estradas da rede portuguesa. As medições incluíram, para cada trecho, o número de acidentes, de mortos e de feridos graves, assim como o tráfego médio diário anual, comprimento, número de vias, existência de vias adicionais, largura de via e tipos e larguras de ambas as bermas e dos separadores centrais. Os dados foram reagrupados em quatro subconjuntos que foram analisados separadamente. O primeiro conjunto de dados analisado consistiu em todos os trechos de auto-estrada, mas considerando apenas as variáveis sem valores omissos, isto é, as variáveis com valores omissos foram removidas da análise (o número de possíveis covariáveis foi, assim, reduzido). O segundo conjunto de dados foi constituído apenas por trechos que não apresentassem valores omissos. Os dois últimos conjuntos de dados analisados foram baseados nos mesmos requisitos que os dados anteriormente descritos, mas cujos valores de tráfego médio diário anual (*AADT*) fossem superior a 5000 veículos. As variáveis resposta consideradas foram os números de acidentes (*Nac*), números de mortos (*NF*), números de feridos graves (*NSI*) e números de mortos e feridos graves (*NKSI*).

Para cada uma das variáveis resposta foram considerados três tipos de modelos lineares generalizados; modelos de regressão de Poisson e modelos hierárquicos baseados nas distribuições de mistura Poisson-Gama e Poisson Log-Normal. Em todos os modelos ajustados foram assumidas distribuições *a priori* não informativas para os coeficientes dos parâmetros dos modelos que se pretendiam estimar. Consequentemente, foram obtidos resultados, para as estimativas dos coeficientes dos modelos, muito semelhantes aos obtidos quando empregando a abordagem *clássica* ao mesmo conjunto de dados (ver Azeredo Lopes e Cardoso (2009)). Contudo, mesmo nos casos não-informativos, as simulações *a posteriori* obtidas mostraram ser úteis para se fazer inferência preditiva e validações dos modelos.

Os modelos lineares generalizados foram ajustados aos quatro conjuntos de dados usando procedimentos *stepwise* baseados no critério de desviância da informação (DIC) e incluíram técnicas de selecção de variáveis dos tipos *forward* e *backward*.

As distribuições *a posteriori* foram obtidas através do *WinBUGS*, empregando o método de Monte Carlo via cadeias de Markov (MCMC). Este método baseia-se na recolha de valores dos parâmetros do modelo que interessam estimar, e que sejam oriundos de distribuições aproximadas. Os valores recolhidos são depois corrigidos de maneira a melhor aproximar a distribuição alvo, isto é, a distribuição *a posteriori*. As amostras são recolhidas sequencialmente, sendo a distribuição destas amostras dependente do último valor recolhido. Consequentemente, as amostras recolhidas formam

uma cadeia de Markov. Contudo, a chave para o sucesso deste método não é a propriedade de Markov, mas sim o facto de as distribuições aproximadas serem aperfeiçoadas a cada passo da simulação, no sentido de convergirem para a distribuição alvo. É necessário ter em conta que, se as iterações não prosseguirem por tempo suficiente, as simulações obtidas poderão não ser representativas da distribuição alvo. Por outro lado, mesmo que as simulações tenham atingido uma convergência aproximada, as iterações continuam a ser influenciadas pelas aproximações tomadas inicialmente (*a priori*) ao invés de serem influenciadas pela distribuição alvo (*a posteriori*).

Para assegurar que as iterações passassem a ser influenciadas pela distribuição alvo procedeu-se à leitura e interpretação dos *trace plots* para cada modelo considerado no presente estudo. Outro problema que ocorre com recolhas amostrais obtidas através de simulações iterativas é a correlação intra-sequencial. Estas foram verificadas através dos gráficos das autocorrelações apresentadas neste documento.

As simulações efectuadas no presente estudo constaram de três sequências, isto é, cadeias de Markov, com valores iniciais dispersos pelo espaço dos parâmetros. A convergência foi monitorizada através da comparação da variação “entre” e “intra” dos valores de simulações sequenciais até a variação “intra” ser aproximadamente igual à variação “entre”. Estes valores foram verificados pelas observações dos gráficos das estatísticas de Gelman-Rubin.

Os modelos foram validados quanto à qualidade dos ajustes através de verificações preditivas *a posteriori* e cálculo de quantidades (medidas de discrepância) que reflectem características dos dados que deveriam, posteriormente, também ser reflectidas pelos dados replicados pelo modelo. As medidas consideradas foram os valores máximos, os valores médios, os desvios padrões, a soma e a razão entre a variância e o valor médio (usado para avaliar a dispersão existente nos dados).

Após a análise dos diferentes subconjuntos de dados (descritos nos Capítulos 4 a 7) para os quatro tipos de eventos (variáveis resposta) concluiu-se que, de todos os modelos ajustados, o modelo de regressão de Poisson não é apropriado para a modelação de qualquer dos tipos de eventos considerados, visto o modelo de Poisson não ter captado algumas das características dos dados, nomeadamente, no que se refere aos valores máximos, desvios padrões e dispersão existente nestes dados de contagem. Os modelos hierárquicos baseados nas distribuições de mistura Poisson-Gama e Poisson Log-Normal podem considerar-se adequados para modelar este tipo de dados. Os resultados obtidos por estes modelos foram bastante semelhantes, tendo o modelo Poisson-Gama obtido ligeiramente melhores resultados no que se refere ao critério DIC, do que o modelo Poisson Log-Normal, sendo embora necessária uma análise mais aprofundada para verificar com mais rigor diferenças e possíveis supremacias entres estes dois tipos de modelos.

Os modelos obtidos pelos conjuntos de dados contendo todos os trechos, indiferentemente dos valores de *AADT* registados, tiveram como variáveis resposta o número de acidentes (*Nac*) e o número de mortos e feridos graves (*NKSI*). Os modelos obtidos dos dados constituídos por variáveis sem valores omissos (descritos no Capítulo 4) consideram, para ambos *Nac* e *NKSI*, o *AADT*, comprimento de trecho (*Length*) e existência ou não de via adicional (*Additional_Lane*) como variáveis explanatórias. A análise descrita no Capítulo 5 (constituído por trechos sem valores omissos) produziu

modelos para Nac e $NKSI$ com as covariáveis $AADT$, $Length$, $Additional_Lane$ e tipo de separador ($Median_Type$). Por conseguinte, pode concluir-se que o número de acidentes e o número de mortos e feridos graves são influenciados por estas variáveis, sendo os números de eventos esperados menores quando o trecho de auto-estrada apresentar uma via adicional e tiver um separador do tipo 'DNJ;NJ'.

No que se refere à análise de trechos com valores de $AADT$ superiores a 5000 veículos (descritos nos Capítulos 6 e 7) os eventos constituídos pelos números de mortos (NF) e pelos números de feridos graves (NSI) foram também considerados como variáveis resposta. Tal como em Azeredo Lopes e Cardoso (2009), quando se analisaram estes trechos e se ignoraram as variáveis com valores omissos (os dados analisados no Capítulo 6), os números esperados de todos os eventos considerados dependeram somente do comprimento do trecho ($Length$) e do valor de $AADT$.

No entanto, quando se excluíram os trechos com valores omissos, mas se incluíram as variáveis relativas aos tipos e larguras das bermas e dos separadores (os dados analisados no Capítulo 7), constatou-se que o número esperado de acidentes ocorridos nesses trechos dependem, não só de $AADT$ e $Length$, como também do tipo de separador (sendo os números esperados de acidentes mais elevados em trechos com separador do tipo 'Vala' e menor em trechos com separadores dos tipos 'DNJ;NJ' e 'GM'). O número de mortes depende unicamente de $AADT$ e $Length$. O número de mortos e feridos graves, assim como o número de feridos graves, dependem de $AADT$ e $Length$ e também da largura da berma esquerda ($Left_Shoulder_Width$) e do tipo de separador ($Median_Type$).

As vantagens de se terem usado técnicas Bayesianas consistiram em se ter podido escolher (*a priori*) a distribuição para os coeficientes dos modelos considerados. Estas distribuições constaram, para os modelos de Poisson, da distribuição Normal (com valor médio zero e desvio padrão elevado, para evidenciar a falta de conhecimentos prévios sobre os coeficientes) e, para os modelos Poisson-Gama e Poisson Log-Normal, de distribuições que partilham parâmetros desconhecidos, formando assim modelos hierárquicos. Em todos os modelos foram obtidos intervalos de credibilidade de 95% e distribuições *a posteriori* para as estimativas dos respectivos coeficientes.

Os modelos apresentados – em especial os do tipo Poisson-Gama – estão aptos a serem usados na avaliação do nível de sinistralidade na rede portuguesa de auto-estradas, nos casos em que não seja necessária a desagregação por sentido de circulação na ligação em causa.

Como trabalho futuro é essencial a execução de uma análise de *sensibilidade*, que consiste em ajustar vários modelos de probabilidade, nomeadamente, várias distribuições *a priori*, ao mesmo problema. Existe também necessidade de ajustar modelos desagregados para despistes e resto dos acidentes.

Lisboa, LNEC, Dezembro de 2009

VISTOS

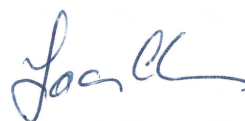
AUTORIAS



Doutora Sofia Mucharreira
de Azeredo Lopes
Investigadora Auxiliar



Eng.º António Lemonde de Macedo
Investigador Coordenador
Director do Departamento de Transportes



Eng.º João Lourenço Cardoso
Investigador Principal Habilitado
Chefe do Núcleo de Planeamento, Tráfego
e Segurança

- Amaral Turkman, M.A. e Silva, G.L. (2000) – *Modelos Lineares Generalizados – da teoria à prática*. Sociedade Portuguesa de Estatística (SPE), VIII Congresso Anual da SPE, Lisboa.
- Azeredo Lopes, S. e Cardoso, J.L. (2007) – *Accident Prediction Models for Portuguese Motorways*. ICT Informação Científica, Vias de Comunicação – INCVC 2, Laboratório Nacional de Engenharia Civil, Lisboa, Portugal.
- Azeredo Lopes, S. e Cardoso, J.L. (2009) – *Accident Prediction Models for Bidirectional Data on Portuguese Motorways*. ICT Informação Científica, Vias de Comunicação – INCVC 4, Laboratório Nacional de Engenharia Civil, Lisboa, Portugal.
- Cameron, A.C. and Trivedi, P.K. (1998) – *Regression Analysis of Count Data*. Cambridge University Press, Econometric Society Monographs, No.30, Cambridge, U.K.
- Carriquiry, A. and Pawlovich, M.D. (2005) – *From Empirical Bayes to Full Bayes: Methods For Analyzing Traffic Safety Data*. http://www.dot.state.ia.us/crashanalysis/pdfs/eb_fb_comparison_whitepaper_october2004.pdf
- Congdon, P. (2003) – *Applied Bayesian Modelling*. Wiley Series in Probability and Statistics, Wiley, Chichester, U.K.
- Congdon, P. (2005) – *Bayesian Models for Categorical Data*. Wiley Series in Probability and Statistics, Wiley, Chichester, U.K.
- Congdon, P. (2006) – *Bayesian Statistical Modelling*. Wiley Series in Probability and Statistics, 2nd edition, Wiley, Chichester, U.K.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) – *Bayesian Data Analysis*. Texts in Statistical Science, 2nd edition, Chapman & Hall, London, U.K.
- Lan, B., Persaud, B., Lyon, C. and Bhim, R. (2009) – *Validation of a Full Bayes methodology for observational before-after road safety studies and application to evaluation of rural signal conversions*. Accident Analysis and Prevention, vol. 41, Issue 3, pp.574-580.
- Li, W., Carriquiry, A., Pawlovich, M. and Welch, T. (2008) – *The Choice of Statistical Models in Road Safety Countermeasure Effectiveness Studies in Iowa* Accident Analysis and Prevention, vol. 40, pp.1531-1542.

- 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*, vol. 38, No. 4, pp.751-766.
- Lord, D., Washington, P.S. and Ivan, J.N. (2004) – *Poisson, Poisson-Gamma and Zero-Inflated Regression Models for Motor Vehicle Crashes: Balancing Statistical Fit and Theory* *Accident Analysis and Prevention*, vol. 37, No. 1, pp.35-46.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) – *WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility*. *Statistics and Computing*, **10**: 325-337.
- Miaou, S. and Lord, D. (2003) – *Modeling Traffic crash-flow relationships for intersections: dispersion parameter, functional form and Bayes versus empirical Bayes methods*. *Transportation Research Record*, 1840, pp.31-40.
- Ntzoufras, I. (2009) – *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics, Wiley, New Jersey.
- Paulino, C.D., Amaral Turkman, M.A. e Murteira, B. (2003) – *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- R Development Core Team (2009) – *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Schlüter, P.J., Deely, J.J. and Nicholson, A.J. (1997) – *Ranking and selecting motor model*. *Journal of the Royal Statistical Society, Series D (The Statistician)* **46**, No. 3, pp 293-316.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) – *Bayesian Measures of Model Complexity and Fit*. *Journal of the Royal Statistical Society*, **B**, 64, Part 4, pp 583-639.
- Spiegelhalter, D.J., Thomas, D., Best, N.G. and Lunn, D. (2003) – *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Cambridge, U.K. <http://www.mrc-cam.ac.uk/bugs>.
- Spiegelhalter, D.J. and Rice, K. (2009) – *Bayesian Statistics*. *Scholarpedia*, 4(8):5230, revision #60276
- Song, J.J., Ghosh, M., Miaou, S. and Mallick, B. (2006) – *Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping*. *Journal of Multivariate Analysis*, **97**, pp 246-273.

Sturtz, S., Ligges, U., and Gelman, A. (2005) – *R2WinBUGS: A Package for Running WinBUGS from R*. Journal of Statistical Software, **12**(3), 1-16.

Tunaru, R. (2002) – *Hierarchical Bayesian Models for Multiple Count Data*. Austrian Journal of Statistics, Volume 31, Number 2&3, pp 221-229.

ANEXOS

ANEXO I – Excerto do código de *WinBUGS* onde a distribuição dos eventos foi assumida como a distribuição de Poisson

```

model {

  for (i in 1:N){
    y[i] ~ dpois(mu[i])
    y.rep[i] ~ dpois(mu[i])

    log(mu[i])<- beta0+beta1*log(AADT[i])+beta2*log(Length[i])
      + beta.addlane*Additional_Lane[i]
      + beta.nblanes*Number_of_Lanes[i]
      + beta.lanewidth*Lane_Width[i]
  }

  # Priors

  beta0 ~ dnorm(0.0, 0.0001)
  beta1 ~ dnorm(0.0, 0.0001)
  beta2 ~ dnorm(0.0, 0.0001)
  beta.addlane ~ dnorm(0.0, 0.0001)
  beta.nblanes ~ dnorm(0.0, 0.0001)
  beta.lanewidth ~ dnorm(0.0, 0.0001)

}

```

ANEXO II – Excerto do código de *WinBUGS* onde a distribuição dos eventos foi assumida como a distribuição Poisson-Gama

```

model {

  for (i in 1:N) {
    y[i] ~ dpois(mu[i])
    y.rep[i]~dpois(mu[i])
    mu[i]<- exp( beta0+beta1*log(AADT[i])+beta2*log(Length[i])
               + beta.addlane*Additional_Lane[i]
               + beta.nblanes*Number_of_Lanes[i]
               + beta.lanewidth*Lane_Width[i] )
               * e[i]
    e[i] ~ dgamma(phi, phi)
  }

# Priors

  phi ~ dgamma(0.1, 0.1)

  beta0 ~ dnorm(0.0, 0.0001)
  beta1 ~ dnorm(0.0, 0.0001)
  beta2 ~ dnorm(0.0, 0.0001)
  beta.addlane ~ dnorm(0.0, 0.0001)
  beta.nblanes ~ dnorm(0.0, 0.0001)

}

```

ANEXO III – Excerto do código de *WinBUGS* onde a distribuição dos eventos foi assumida como a distribuição Poisson Log-Normal

```

model {

  for (i in 1:N) {
    y[i] ~ dpois(mu[i])
    y.rep[i]~dpois(mu[i])
    log(mu[i])<- beta0+beta1*log(AADT[i])+beta2*log(Length[i])
                + beta.addlane*Additional_Lane[i]
                + beta.nblanes*Number_of_Lanes[i]
                + beta.lanewidth*Lane_Width[i]
                + e[i]
#
    e[i] ~ dnorm(0.0, tau)
  }
#
# Priors

  tau ~ dgamma(0.001,0.001)

  beta0 ~ dnorm(0.0, 0.0001)
  beta1 ~ dnorm(0.0, 0.0001)
  beta2 ~ dnorm(0.0, 0.0001)
  beta.addlane ~ dnorm(0.0, 0.0001)
  beta.nblanes ~ dnorm(0.0, 0.0001)

}

```