**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelos para a análise de dados de contagens longitudinais com
superdispersão: estimação INLA**

Everton Batista da Rocha

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

**Piracicaba
2015**

Everton Batista da Rocha
Bacharel em Estatística

**Modelos para a análise de dados de contagens longitudinais com
superdispersão: estimação INLA**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:

Profa. Dra. **ROSELI APARECIDA LEANDRO**

Co-orientador:

Prof. Dr. **PAULO JUSTINIANO RIBEIRO JUNIOR**

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

**Piracicaba
2015**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Rocha, Everton Batista da

Modelos para a análise de dados de contagens longitudinais com superdispersão:
estimação INLA / Everton Batista da Rocha. - - versão revisada de acordo com a resolução
CoPGr 6018 de 2011. - - Piracicaba, 2015.

155 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Contagens 2. Superdispersão 3. Análise de dados longitudinais 4. Inferência Bayesiana
I. Título

CDD 519.532
R672m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

*Dedico este trabalho à memória de minha mãe, **Anaides B. da Rocha**, pelo carinho, pelo afeto, pelo exemplo de vida e por ter acreditado no meu futuro, impulsionando-me a buscar meus sonhos, mesmo que estes nos distanciassem.*

Mãe, onde quer que a senhora esteja, esta é a realização do nosso sonho!

AGRADECIMENTOS

Ao meu pai, David R. da Rocha, pelo apoio e compreensão em muitos momentos.

À minha irmã, Ana Paula B. da Rocha, pelo afeto e cumplicidade, fazendo-me acreditar em finais felizes.

À Profa. Dra. Roseli A. Leandro, com muito carinho e afeto, pela dedicação e comprometimento como orientadora e, de maneira muito especial, pela confiança em mim depositada: pela compreensão, torcida e apoio ao longo dos anos que convivemos. Neste período, foi meu exemplo de profissionalismo e ética, mostrando-me a responsabilidade científica e social envolvida na transmissão do saber. A convivência ao seu lado me trouxe, além de conhecimento, ensinamentos que levarei para a vida, bem como uma valorosa amizade.

Ao Prof. Dr. Paulo Justiniano Ribeiro Júnior, pela co-orientação e, em especial, pelos cursos ministrados nas dependências da UFPR, que muito contribuíram para a realização deste trabalho.

A todos os professores do programa de doutoramento, que compartilharam seu precioso conhecimento, pois, sem eles, eu não concluiria mais esta fase.

Ao grupo de estudos GEMMix, pelos momentos compartilhados.

Ao Instituto de Matemática e Estatística da Universidade Federal de Goiás, pela vivência, e, em especial, a todos os professores do Grupo de Estatística, que me apoiaram no término deste curso.

Aos funcionários do LCE/ESALQ/USP, pela dedicação e disponibilidade no transcorrer do curso.

A todos os alunos do curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, com os quais compartilhei essa fase da minha vida.

À CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - pela bolsa concedida.

Aos amigos, André, Terezinha, Shelly, Beto, Fernando, Ana Julia, Mariana, Simone Grego, Davi, Luzia, Fernando, Marta, Renato, Mário, pois vocês são minha segunda família. Um agradecimento muito especial, pois em vocês eu encontrei o conforto, o apoio, o sorriso, a generosidade e o afeto; em vocês eu encontrei motivos para acreditar sempre em

dias melhores.

E por fim, a Deus, pois sem ele nada disso teria acontecido.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 DESENVOLVIMENTO	23
2.1 Revisão Bibliográfica	23
2.1.1 Distribuições de probabilidade para dados na forma de contagens	23
2.1.1.1 Distribuição de Poisson	23
2.1.1.2 Distribuição de Poisson inflacionada de zeros	24
2.1.1.3 Distribuição binomial negativa	26
2.1.1.4 Distribuição binomial negativa inflacionada de zeros	27
2.1.2 Dados Longitudinais	28
2.1.3 Modelos Lineares Generalizados	29
2.1.4 Superdispersão	33
2.1.5 Modelos Lineares Generalizados Mistos	35
2.1.6 Modelos Para Dados de Contagens Longitudinais Com Superdispersão	38
2.1.6.1 Modelo 1: Poisson	39
2.1.6.2 Modelo 2: Poisson-gama	39
2.1.6.3 Modelo 3: Poisson-normal	40
2.1.6.4 Modelo 4: Poisson-normal-gama	40
2.1.7 Inferência Bayesiana	41
2.1.7.1 Aproximação de Laplace e Método de Monte Carlo	46
2.1.7.2 Aproximação de Laplace	46
2.1.7.3 Métodos de Monte Carlo com cadeias de Markov	50
2.1.7.3.1 Algoritmo de Metropolis-Hastings	50
2.1.7.3.2 Amostrador de Gibbs	51
2.1.7.4 Método INLA	52
2.1.7.4.1 Modelos gaussianos latentes	53
2.1.7.4.2 Campos aleatórios gaussianos markovianos	55
2.1.7.4.3 Estimação INLA	60

2.1.7.5 Critérios para a comparação de modelos	63
2.1.7.5.1 Critério da informação da deviance	63
2.1.7.5.2 Critério da divergência simétrica de Kullback-Leibler	65
2.1.7.5.3 Transformação integral de probabilidade	65
2.1.7.5.4 Condicionais preditivas ordenadas	67
2.2 Metodologia	69
2.2.1 Ensaio clínico com pacientes portadores de epilepsia	69
2.2.2 Modelos	70
2.2.2.1 Modelo 1: Poisson	70
2.2.2.2 Modelo 2: Poisson-gama	72
2.2.2.3 Modelo 3: Poisson-normal	74
2.2.2.4 Modelo 4: Poisson-normal-gama ou Combinado	75
2.2.2.5 Modelo 5: Poisson inflacionado de zeros - ZIP	77
2.2.2.6 Modelo 6: Binomial negativa inflacionada de zeros - ZINB	79
2.2.2.7 Modelo 7: ZIP-normal	81
2.2.2.8 Modelo 8: ZINB-normal	83
2.2.3 Implementação computacional	86
2.2.3.1 Estudo de simulação	86
3 RESULTADOS E DISCUSSÃO	89
4 CONSIDERAÇÕES FINAIS	113
REFERÊNCIAS	115
APÊNDICE	123

RESUMO

Modelos para a análise de dados de contagens longitudinais com superdispersão: estimação INLA

Em ensaios clínicos é muito comum a ocorrência de dados longitudinais discretos. Para sua análise é necessário levar em consideração que dados observados na mesma unidade experimental ao longo do tempo possam ser correlacionados. Além dessa correlação inerente aos dados é comum ocorrer o fenômeno de superdispersão (ou sobredispersão), em que, existe uma variabilidade nos dados além daquela captada pelo modelo. Um caso que pode acarretar a superdispersão é o excesso de zeros, podendo também a superdispersão ocorrer em valores não nulos, ou ainda, em ambos os casos. Molenberghs, Verbeke e Demétrio (2007) propuseram uma classe de modelos para acomodar simultaneamente a superdispersão e a correlação em dados de contagens: modelo Poisson, modelo Poisson-gama, modelo Poisson-normal e modelo Poisson-normal-gama (ou modelo combinado). Rizzato (2011) apresentou a abordagem bayesiana para o ajuste desses modelos por meio do Método de Monte Carlo com Cadeias de Markov (MCMC). Este trabalho, para modelar a incerteza relativa aos parâmetros desses modelos, considerou a abordagem bayesiana por meio de um método determinístico para a solução de integrais, INLA (do inglês, *Integrated Nested Laplace Approximations*). Além dessa classe de modelos, como objetivo, foram propostos outros quatro modelos que também consideram a correlação entre medidas longitudinais e a ocorrência de superdispersão, além da ocorrência de zeros estruturais e não estruturais (amostrais): modelo Poisson inflacionado de zeros (ZIP), modelo binomial negativo inflacionado de zeros (ZINB), modelo Poisson inflacionado de zeros - normal (ZIP-normal) e modelo binomial negativo inflacionado de zeros - normal (ZINB-normal). Para ilustrar a metodologia desenvolvida, um conjunto de dados reais referentes à contagens de ataques epiléticos sofridos por pacientes portadores de epilepsia submetidos a dois tratamentos (um placebo e uma nova droga) ao longo de 27 semanas foi considerado. A seleção de modelos foi realizada utilizando-se medidas preditivas baseadas em validação cruzada. Sob essas medidas, o modelo selecionado foi o modelo ZIP-normal, sob o modelo corrente na literatura, modelo combinado. As rotinas computacionais foram implementadas no programa R e são parte deste trabalho.

Palavras-chave: Contagens; Superdispersão; Análise de dados longitudinais, Inferência Bayesiana

ABSTRACT

Models for data analysis of longitudinal counts with overdispersion: INLA estimation

Discrete and longitudinal structures naturally arise in clinical trial data. Such data are usually correlated, particularly when the observations are made within the same experimental unit over time and, thus, statistical analyses must take this situation into account. Besides this typical correlation, overdispersion is another common phenomenon in discrete data, defined as a greater observed variability than that nominated by the statistical model. The causes of overdispersion are usually related to an excess of observed zeros (zero-inflation), or an excess of observed positive specific values or even both. Molenberghs, Verbeke e Demétrio (2007) have developed a class of models that encompasses both overdispersion and correlation in count data: Poisson, Poisson-gama, Poisson-normal, Poisson-normal-gama (combined model) models. A Bayesian approach was presented by Rizzato (2011) to fit these models using the Markov Chain Monte Carlo method (MCMC). In this work, a Bayesian framework was adopted as well and, in order to consider the uncertainty related to the model parameters, the Integrated Nested Laplace Approximations (INLA) method was used. Along with the models considered in Rizzato (2011), another four new models were proposed including longitudinal correlation, overdispersion and zero-inflation by structural and random zeros, namely: zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated Poisson-normal (ZIP-normal) and the zero-inflated negative binomial-normal (ZINB-normal) models. In order to illustrate the developed methodology, the models were fit to a real dataset, in which the response variable was taken to be the number of epileptic events per week in each individual. These individuals were split into two groups, one taking placebo and the other taking an experimental drug, and they observed up to 27 weeks. The model selection criteria were given by different predictive measures based on cross validation. In this setting, the ZIP-normal model was selected instead the usual model in the literature (combined model). The computational routines were implemented in R language and constitute a part of this work.

Keywords: Counts; Overdispersion; Longitudinal data analysis; Bayesian inference

LISTA DE FIGURAS

Figura 1 - Gráfico da aproximação normal ($\tilde{\pi}(\eta y)$) e da posteriori ($\pi(\eta y)$) para $y = 3$, $\mu = 0$, $\kappa = 0,001$ e $\eta_0 = 0, 0,5, 1$ e $1,5$, indicados como um ponto preto	58
Figura 2 - Gráfico de perfis individuais para o número de ataques epilépticos sofridos em cada uma das 27 semanas pelos pacientes submetidos ao tratamento placebo (0) e à nova droga epiléptica (1)	89
Figura 3 - (a) Gráfico dos valores de PIT para o modelo Poisson. (b) Histograma dos valores de PIT para o modelo Poisson	92
Figura 4 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson	92
Figura 5 - (a) Gráfico dos valores de PIT para o modelo Poisson-gama. (b) Histograma dos valores de PIT para o modelo Poisson-gama	93
Figura 6 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama	93
Figura 7 - (a) Gráfico dos valores de PIT para o modelo Poisson-normal. (b) Histograma dos valores de PIT para o modelo Poisson-normal	94
Figura 8 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal	94
Figura 9 - (a) Gráfico dos valores de PIT para o modelo combinado. (b) Histograma dos valores de PIT para o modelo combinado	95

Figura 10 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado	95
Figura 11 - (a) Histograma dos valores de CPO para o modelo Poisson. (b) Histograma dos valores de CPO para o modelo Poisson-gama. (c) Histograma dos valores de CPO para o modelo Poisson-normal. (d) Histograma dos valores de CPO para o modelo combinado	97
Figura 12 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} e β_{11} sob o modelo Poisson	97
Figura 13 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , α e β sob o modelo Poisson-gama	98
Figura 14 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} e σ^2 sob o modelo Poisson-normal	98
Figura 15 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , α , β e σ^2 sob o modelo combinado	99
Figura 16 - (a) Gráfico dos valores de PIT para o modelo ZIP. (b) Histograma dos valores de PIT para o modelo ZIP	103
Figura 17 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP	103
Figura 18 - (a) Gráfico dos valores de PIT para o modelo ZINB. (b) Histograma dos valores de PIT para o modelo ZINB	104
Figura 19 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB	104

Figura 20 - (a) Gráfico dos valores de PIT para o modelo ZIP-normal. (b) Histograma dos valores de PIT para o modelo ZIP-normal	105
Figura 21 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal	105
Figura 22 - (a) Gráfico dos valores de PIT para o modelo ZINB-normal. (b) Histograma dos valores de PIT para o modelo ZINB-normal	106
Figura 23 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal	106
Figura 24 - (a) Histograma dos valores de CPO para o modelo ZIP. (b) Histograma dos valores de CPO para o modelo ZINB. (c) Histograma dos valores de CPO para o modelo ZIP-normal. (d) Histograma dos valores de CPO para o modelo ZINB-normal	107
Figura 25 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} e β_{11} sob o modelo ZIP	108
Figura 26 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω e α sob o modelo ZINB	108
Figura 27 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω e σ^2 sob o modelo ZIP-normal	109
Figura 28 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω , α e σ^2 sob o modelo ZINB-normal	109

LISTA DE TABELAS

Tabela 1 - Interpretação do PsFB, segundo Jeffreys (1961)	69
Tabela 2 - Número de observações tomadas em cada uma das semanas do estudo, em ambos os grupos	90
Tabela 3 - Resumos a posteriori para os parâmetros dos modelos Poisson, Poisson- gama, Poisson-normal e combinado	100
Tabela 4 - Resumos a posteriori para os parâmetros dos modelos Poisson, Poisson- gama, Poisson-normal e combinado apresentadas por Rizzato (2011) . . .	101
Tabela 5 - Resumos a posteriori para os parâmetros dos modelos ZIP, ZINB, ZIP- normal e ZINB-normal	111
Tabela 6 - Resumos a posteriori do estudo de simulação do modelo ZIP-normal . . .	112

1 INTRODUÇÃO

É bastante comum a ocorrência de dados longitudinais em ensaios clínicos, assim como, na experimentação agronômica, farmacêutica, biológica, entre outras. Dados longitudinais são medidas provenientes de planejamentos longitudinais, que se caracterizam pela observação de uma ou mais variáveis respostas em uma mesma unidade experimental, estas obtidas ao longo de uma dimensão específica – a exemplo, o tempo.

A natureza da variável resposta e a correlação entre as medidas tomadas na mesma unidade experimental devem ser levadas em conta quando da análise de dados longitudinais, que podem ser de natureza contínua – o ganho de peso observado ao longo do tempo, ou de natureza discreta – a contagem de ocorrências de um determinado evento de interesse ao longo do tempo.

No caso de dados longitudinais de natureza contínua, a literatura apresenta diversas propostas para a análise, sendo o ajuste de modelos lineares mistos a mais utilizada atualmente, dado o desenvolvimento da teoria estatística e dos recursos computacionais. No ajuste desses modelos, a variável resposta é suposta normal e efeitos aleatórios são inseridos no modelo linear clássico de regressão, objetivando a captura da correlação entre as medidas tomadas da mesma unidade experimental. Em tais modelos, além da indução de correlação entre medidas repetidas, é possível modelar a variabilidade ao longo do tempo por meio de diferentes especificações da matriz de covariâncias.

Nos dados longitudinais de natureza discreta – contagens, por exemplo, considerando-se que a variável resposta pode ser descrita por uma distribuição de probabilidade pertencente à família exponencial de distribuições, é possível o uso dos ajustes dos modelos propostos por Nelder e Weddeburn (1972) – os modelos lineares generalizados.

O modelo linear generalizado, doravante MLG, fica definido ao se considerar um componente aleatório, um sistemático – ou preditor linear – e uma função de ligação. Levando-se em conta a inclusão de um efeito aleatório no preditor linear, tem-se o modelo linear generalizado misto – MLGM, proposto por Breslow e Clayton (1993).

No MLGM, o ajuste baseado em verossimilhança pode apresentar problemas, o que se torna mais notório no caso de pequenas amostras, haja visto que a inferência se baseia na distribuição assintótica dos estimadores de máxima verossimilhança. Como solução para este fato, propõe-se a abordagem bayesiana.

A estimação bayesiana necessita de programação computacional intensiva e os métodos mais difundidos para a obtenção de resumos da distribuição conjunta a posteriori são os Métodos de Monte Carlo com Cadeias de Markov – MCMC, especialmente o algoritmo de Metropolis-Hasting e o amostrados de Gibbs.

A execução do algoritmo de Metropolis-Hasting e do amostrador de Gibbs em problemas complexos pode ser demorada para a obtenção da convergência- a exemplo, no ajuste de modelos hierárquicos. Uma alternativa proposta à restrição de tais métodos estocásticos de estimação bayesiana é o uso da estimação baseada na abordagem INLA - *Integrated Nested Laplace Approximation* -, proposta por Rue, Martino e Chopin (2009). Nesse trabalho, um conjunto de dados de natureza discreta, na forma de contagens observados ao longo do tempo, é analisado por meio da inferência bayesiana considerando a nova abordagem INLA.

No caso de dados de contagem, a distribuição mais comumente usada para descrever o mecanismo gerador do fenômeno é a distribuição de Poisson. Todavia, conforme a definição desta distribuição, há a exigência de que a média e a variância da variável resposta sejam iguais, o que nem sempre é observado em fenômenos com dados de contagem, sobretudo os oriundos de planejamentos longitudinais, o que pode ocasionar falta de ajuste do MLG. A alternativa que se sugere é considerar a distribuição Binomial Negativa, que possui uma função de variância mais flexível.

No caso da inclusão de efeitos aleatórios no modelo, dada a natureza dos dados longitudinais, têm-se os modelos combinados (i) Poisson-Normal, (ii) Poisson-Gama e (iii) Poisson-Normal-Gama, apresentados por Molenberghs, Verbeke e Demétrio (2007). Nesses, considera-se a estimação clássica na qual efeitos aleatórios são inseridos nos modelos para acomodar a superdispersão e a correlação entre as medidas tomadas no mesmo indivíduo; cumpre-se notar que os modelos foram ajustados aos mesmos dados que serão estudados neste texto e que Rizzato (2011), considerando a abordagem bayesiana via MCMC, também apresentou os modelos propostos por Molenberghs, Verbeke e Demétrio (2007).

Todos os modelos apresentados por Molenberghs, Verbeke e Demétrio (2007) e particularizados por Rizzato (2011) aqui apresentados foram ajustados por meio da abordagem INLA considerando-se a inferência bayesiana. Para mais, são apresentados outros modelos para o ajuste de dados desta natureza, os quais, quando comparados aos modelos apresentados por Rizzato (2011), apresentam melhores resultados para as medidas de

qualidade de ajuste consideradas.

Destarte, o objetivo do presente é, utilizando a inferência bayesiana, apresentar modelos que podem servir, de maneira consistente e sem alto custo computacional, para a análise de dados longitudinais na forma de contagens por meio da nova abordagem INLA, bem como suscitar a discussão a respeito de critérios bayesianos para a seleção de modelos.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

2.1.1 Distribuições de probabilidade para dados na forma de contagens

Seja $(\Omega, \mathcal{F}, \mathcal{P})$ um espaço de probabilidade, tal que Ω é o espaço amostral de um dado evento de interesse, $\mathcal{F}$ é uma σ -álgebra e $\mathcal{P}$ é a função de probabilidade definida para subconjuntos de Ω . Na realização de um fenômeno aleatório é comum ter-se interesse em descrever o comportamento de uma ou mais quantidades de maneira probabilística. Conforme Magalhães (2011), dado um fenômeno aleatório qualquer, com um certo espaço de probabilidade, deseja-se estudar a estrutura probabilística de quantidades associadas a esse fenômeno. Sendo essa quantidade uma variável aleatória Y , $Y : \Omega \rightarrow \mathbb{R}$, deseja-se associar à variável aleatória um modelo de probabilidade e descrever o comportamento da distribuição desta variável, por exemplo, por meio de sua função de distribuição de probabilidades.

Considerando que a variável aleatória seja discreta, ou seja, que assuma somente um número enumerável de valores (finito ou infinito), uma série de distribuições na literatura são propostas para descrever o comportamento desta variável. No caso de dados de contagens, a distribuição comumente usada é a distribuição de Poisson. Levando-se em conta características inerentes do fenômeno, outras distribuições, como a binomial negativa e a Poisson inflacionada de zeros são correntes na literatura.

2.1.1.1 Distribuição de Poisson

A distribuição de Poisson é de relevante importância tanto teórica como prática dentro da Estatística, tendo sido introduzida na literatura por Siméon Denis Poisson e publicada juntamente com sua teoria de probabilidade em 1837 no trabalho *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*.

Considere que Y seja uma variável aleatória com distribuição de Poisson de parâmetro λ , $Y \sim \text{Poisson}(\lambda)$, em que $\lambda > 0$. A função de probabilidade de Y pode ser parametrizada da seguinte forma:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, \dots, \quad (1)$$

em que y é uma observação de uma contagem em geral feita em um intervalo de tempo.

O valor esperado ou esperança matemática da variável aleatória Y e sua variância são dados por,

$$E(Y) = \lambda \text{ e } Var(Y) = \lambda.$$

A obtenção desses valores encontram-se no Apêndice A deste texto.

A distribuição de Poisson descreve a probabilidade de que ocorra $Y = y$ eventos de interesse em um dado intervalo, geralmente de tempo, considerando que os eventos ocorram independentemente com uma taxa constante λ , a qual representa simultaneamente a média e variância dessa variável aleatória Y . Essa restrição de igualdade entre média e variância deve ser levada em consideração ao supor que o mecanismo gerador de um processo aleatório seja descrito pela distribuição de Poisson.

2.1.1.2 Distribuição de Poisson inflacionada de zeros

O modelo de Poisson é frequentemente usado e deve ser o primeiro a ser considerado quando se deseja descrever um processo de contagens, todavia, na prática, existem situações em que os dados podem estar superdispersados (ou sobredispersos). De acordo com Ridout, Demétrio, Hinde (1998), a superdispersão (ou sobredispersão) em um conjunto de dados pode ocorrer quando a incidência de contagens nulas é maior que o esperado pela distribuição de Poisson, de maneira que esses zeros devem ser tratados de maneira especial. Nesses casos, um modelo que leve em consideração o excesso de zeros, ou seja, a inflação de zeros é mais indicado. Segundo Lambert (1992), o modelo adotado para descrever o comportamento dos dados deve levar em consideração: (i) as contagens nulas, e (ii) as contagens não nulas.

Nos modelos inflacionados de zeros, a ocorrência dos valores nulos são de duas naturezas, os zeros estruturais e os zeros amostrais. Os zeros estruturais ocorrem quando há ausência de determinada característica na população, ocasionando a ocorrência de zeros. Os zeros amostrais são aqueles decorrentes da não observação de uma determinada característica na população. Por exemplo, segundo Ridout, Demétrio e Hinde (1998), em contagens de lesões em plantas ocasionadas por uma determinada doença, uma planta pode não apresentar a lesão por ser resistente à doença, ou simplesmente porque os esporos

causadores da doença ainda não entraram em contato com a planta. No primeiro caso, tem-se um zero estrutural e no segundo caso um zero amostral. Poston e McKibben (2003) consideraram os zeros estruturais e amostrais em um estudo sobre fertilidade, levando em consideração o número de filhos nascidos em uma determinada região. Neste estudo, entre as mulheres que não são mães, foram consideradas duas situações: as mulheres que não são mães devido a causas do organismo (zero estrutural) e aquelas que não são mães por ainda não terem engravidado (zero amostral). Observe que nos dois casos, a contagem é nula, todavia de naturezas distintas. Devido à natureza dos dados (contagens), o modelo Poisson inflacionado de zeros (ZIP) foi considerado nos estudos de Ridout, Demétrio e Hinde (1998) e de Poston e McKibben (2003).

Uma vez que os zeros estruturais e amostrais devem ser levados em consideração no modelo probabilístico, para a paramentrização do modelo ZIP, considere como sendo ω a probabilidade de zeros estruturais e $(1 - \omega)$ a probabilidade de zeros amostrais. Então, diz-se que a variável aleatória Y tem distribuição ZIP com parâmetro λ , probabilidade ω de zeros estruturais e $(1 - \omega)$ de zeros amostrais se sua distribuição de probabilidade pode ser parametrizada da seguinte forma,

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)P(\cdot), & \text{se } y = 0 \\ (1 - \omega)P(\cdot), & \text{se } y > 0, \end{cases}$$

em que $P(\cdot)$ representa a função de probabilidade da distribuição de Poisson de parâmetro λ , podendo, então, a expressão anterior ser reescrita como,

$$P(Y = y) = \begin{cases} \omega + (1 - \omega) \exp(-\lambda), & \text{se } y = 0 \\ (1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y!}, & \text{se } y > 0, \end{cases} \quad (2)$$

isto é, a modelagem da resposta nula leva em consideração a probabilidade da resposta ser igual a zero para o modelo Poisson. Para as contagens não nulas, tem-se a probabilidade de valores diferentes de zero da distribuição de Poisson.

O valor esperado da variável aleatória Y é dado por

$$E(Y) = (1 - \omega)\lambda = \mu,$$

e sua variância é expressa por,

$$Var(Y) = \mu + \left(\frac{\omega}{1-\omega} \right) (1-\omega)^2 \lambda^2 = \mu + \left(\frac{\omega}{1-\omega} \right) \mu^2.$$

A obtenção desses valores encontram-se no Apêndice A deste texto.

2.1.1.3 Distribuição binomial negativa

Considerando dados de contagens com variabilidade extra-Poisson é comum associar misturas de distribuições de Poisson para o comportamento destes dados, como apresentado por Hinde e Demétrio (1998), Ridout, Demétrio e Hinde (1998) e Molenberghs, Verbeke e Demétrio (2007). Essas misturas resultam, em casos particulares de distribuições, tais como a distribuição binomial negativa, que é um modelo de dois estágios.

Conforme Hinde e Demétrio (1998), considere em um primeiro estágio que a variável discreta Y condicionalmente à variável aleatória Z , $Y|Z$, representa um processo de contagem, e que $Y|Z = z \sim \text{Poisson}(z)$,

$$P(Y = y|Z = z) = \frac{z^y \exp(-z)}{y!}, \quad y = 0, 1, \dots,$$

e no segundo estágio do modelo que $Z \sim \Gamma(k, \lambda)$, $z > 0$, k fixo, ou seja, que

$$f_Z(z) = \frac{\left(\frac{\lambda}{k} \right)^\lambda}{\Gamma(\lambda)} z^{\lambda-1} \exp \left(-\frac{z\lambda}{k} \right). \quad (3)$$

Dessa forma, tem-se que a distribuição da variável aleatória Y é dada por

$$P(Y = y) = \frac{\Gamma(y + \lambda)}{y! \Gamma(\lambda)} \frac{\lambda^\lambda k^y}{(k + y)^{y+\lambda}},$$

de maneira que Y tem distribuição binomial negativa com parâmetros k e λ dada por

$$P(Y = y) = \frac{\Gamma(y + \lambda)}{y! \Gamma(\lambda)} \frac{\lambda^\lambda k^y}{(k + y)^{y+\lambda}}.$$

A esperança da variável aleatória Y é dada por

$$E(Y) = \frac{k}{\lambda} = \mu$$

e sua variância,

$$Var(Y) = \frac{k}{\lambda} + \frac{\left(\frac{k}{\lambda}\right)^2}{k} = \mu + \frac{\mu^2}{k}.$$

A obtenção desses valores encontram-se no Apêndice A deste texto.

2.1.1.4 Distribuição binomial negativa inflacionada de zeros

De acordo com Moghimbeigi et al. (2008), em situações práticas, além da superdispersão ocasionada pela inflação de zeros pode ocorrer uma superdispersão também na parte não nula dos dados. Nesse caso, a distribuição binomial negativa inflacionada de zeros (ZINB) é mais apropriada para descrever o comportamento de dados dessa natureza.

Seguindo a parametrização apresentada por Ridout, Demétrio e Hinde (1998) e Greene (1994), considere que Y é uma variável aleatória na forma de contagem com excessos de zeros, seu comportamento distribucional pode ser descrito pela distribuição binomial inflacionadas de zeros, com função de probabilidade expressa por

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)(1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}}, & \text{se } y = 0 \\ (1 - \omega) \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{-c}}{\alpha}\right)^{-y}, & \text{se } y > 0, \end{cases}$$

em que λ e α são parâmetros, sendo α o parâmetro de dispersão, ω é a probabilidade de zeros estruturais e c é uma constante arbitrária. De maneira análoga à distribuição ZIP, ao se utilizar a distribuição ZINB, considera-se uma modelagem para os zeros estruturais e para os zeros amostrais. A obtenção do valor esperado e da variância da variável aleatória Y , neste caso, também deve levar em conta essa estrutura de zeros. Para o valor esperado,

$$E(Y) = (1 - \omega)\lambda,$$

e, para a variância tem-se,

$$Var(Y) = (1 - \omega)\lambda(\lambda\omega + 1 + \alpha\lambda^c).$$

2.1.2 Dados Longitudinais

Dados longitudinais são observações provenientes de planejamentos longitudinais em que medidas são repetidas na mesma unidade experimental em diversas ocasiões ou condições de avaliações, em geral, ao longo do tempo. Uma vez que as medidas são tomadas de maneira sistemática em uma mesma unidade observacional, espera-se que exista uma correlação não nula entre essas medidas. De acordo com Verbeke e Molenberghs (2000), a ocorrência de dados correlacionados é muito comum nas ciências aplicadas, podendo também ocorrer em outras estruturas de dados, tais como observações multivariadas e dados agrupados.

A estrutura de dados longitudinais será considerada regular, em relação ao tempo, se o intervalo entre duas medidas consecutivas quaisquer for constante durante todo o estudo, será considerada balanceada se todas as observações forem feitas no mesmo instante em todas as unidades e completa, se não apresentar observações perdidas (HELMS, 1992). Uma vez que esses dados são obtidos de estudos que podem durar dias, semanas, meses e até mesmo anos, é frequente a ocorrência de dados longitudinais irregulares e desbalanceados, em relação ao tempo, assim como incompletos e tais características devem ser levadas em conta na análise.

Segundo Lima (1996), considerando uma estrutura de dados longitudinais, a cada tratamento (ou grupo) de interesse observado ao longo do tempo está associado um perfil médio de respostas que deve evidenciar o efeito do tratamento e o seu comportamento. Belle et al. (2004) comentam que estudos longitudinais são considerados uma ferramenta robusta em estudos prospectivos, pois possibilitam a identificação de mudanças de padrões individuais, o que é de interesse, por exemplo, em estudos epidemiológicos, pesquisas clínicas e avaliação de terapias. Assim, na análise de dados longitudinais, em geral, o interesse está em estudar o comportamento de uma ou mais variáveis respostas ao longo do tempo.

Para que a análise de dados longitudinais leve em consideração a correlação entre medidas tomadas na mesma unidade, pode-se utilizar o ajuste de modelos de regressão com a inclusão de efeitos aleatórios. Os efeitos aleatórios no modelo são responsáveis por

acomodar a correlação entre as observações que é descrita por meio de uma estrutura de covariâncias que deve ser especificada de acordo com as características inerentes ao estudo.

2.1.3 Modelos Lineares Generalizados

Nos experimentos científicos vários tipos de respostas podem ser observadas. Dentre elas destacam-se as respostas discretas e as contínuas. A essa variedade de respostas (ou dados) tenta-se ajustar uma distribuição de probabilidade, como, por exemplo, as distribuições normal, binomial, Poisson, multinomial, binomial negativa, normal inversa, beta, gama. As distribuições normal, Poisson e multinomial são distribuições especiais que pertencem à família exponencial. Essa é uma classe de distribuições que desfruta de propriedades especiais. Algumas distribuições pertencem à família exponencial uniparamétrica, apenas, quando alguns parâmetros são considerados conhecidos, como por exemplo, a distribuição binomial negativa é considerada um membro da família exponencial quando o número de falhas for fixo, a distribuição multinomial quando o número de tentativas for conhecido e o mesmo acontece para a distribuição binomial. Tal classe de distribuições foi proposta independentemente por B. O. Koopman (KOOPMAN, 1936), E. J. G. Pitman (PITMAN e WISHART, 1936) e G. Darrois (DARMOIS, 1935), quando pesquisavam as propriedades de suficiência estatística, e a partir de então, vários outros aspectos desta família foram temas de estudo, vindo a tornar-se uma importante área da Estatística.

Se uma distribuição de probabilidade ou densidade pertence à família exponencial uniparamétrica ela pode ser escrita na forma,

$$f(y; \theta) = h(y) \exp[\eta(\theta)t(y) - b(\theta)] \quad (4)$$

em que $\eta(\theta)$, $b(\theta)$, $t(y)$ e $h(y)$ assumem valores em subconjuntos dos números reais. De acordo com Cordeiro e Demétrio (2009), as funções $\eta(\theta)$, $b(\theta)$, $t(y)$ não são únicas, por exemplo, $\eta(\theta)$ pode ser multiplicada por uma constante k e $t(y)$ pode ser dividida pela mesma constante. Cordeiro et al. (1995) apresentam 24 distribuições na parametrização (4).

Verificar se uma função pertence, ou não, à família exponencial é muito importante na teoria de modelos lineares generalizados. Considere, como exemplo, a distribuição de Poisson, $P(\theta)$, de parâmetro $\theta > 0$, utilizada para a análise de dados na forma de

contagens. Observe que sua função de probabilidade pode ser escrita da seguinte forma

$$f(y; \theta) = \frac{e^{-\theta} \theta^y}{y!} = \frac{1}{y!} \exp[y \ln(\theta) - \theta],$$

indicando que a distribuição de Poisson pertence à família exponencial de distriuições (4), em que $\eta(\theta) = \ln(\theta)$, $b(\theta) = \theta$, $t(y) = y$ e $h(y) = \frac{1}{y!}$. Parametrizou-se a distribuição de Poisson com parâmetro θ ao invés de λ , como em (1), por ser usual representar a distribuição de probabilidade na forma da família exponencial.

O suporte das distribuições pertencentes à família exponencial não pode depender de θ , sendo assim, a distribuição de probabilidade uniforme em $(0, \theta)$ não pertence à família exponencial.

No caso em que as funções $\eta(\theta)$ e $t(y)$ forem iguais à função identidade, a expressão (4) pode ser reescrita como,

$$f(y; \theta) = h(y) \exp[\theta y - b(\theta)], \quad (5)$$

definindo a família exponencial na forma canônica. Na parametrização (5), θ é chamado de parâmetro canônico. Dessa forma, o logaritmo da função de verossimilhança correspondente a uma única observação é dado por,

$$l(\theta) = \theta y - b(\theta) + \ln[h(y)] \quad (6)$$

de maneira que a função escore $U(\theta) = \frac{dl(\theta)}{d\theta}$ resulta em $U(\theta) = y - b'(\theta)$ e consequentemente, tem-se que $E[U(\theta)] = 0$ e $Var(U(\theta)) = -E \left[\frac{d^2 l(\theta)}{d\theta^2} \right]$ (informação de Fisher), e assim, $E[Y] = b'(\theta)$ e $Var[Y] = b''(\theta)$.

Um outro resultado importante decorrente da expressão (4) é que, se $Y_1, \dots, Y_n$ são n variáveis aleatórias independentes e identicamente distribuídas, a distribuição conjunta de $Y_1, \dots, Y_n$ é dada por,

$$f(y_1, \dots, y_n; \theta) = \left[\prod_{i=1}^n h(y_i) \right] \exp \left[\eta(\theta) \sum_{i=1}^n t(y_i) - nb(\theta) \right]. \quad (7)$$

Sendo assim, a distribuição conjunta de $Y_1, \dots, Y_n$ pertence à família exponencial. Ainda,

a estatística $\sum_{i=1}^n T(Y_i)$ é uma estatística suficiente para o parâmetro θ . Observe ainda que a função de verossimilhança é facilmente deduzida de (7). De acordo com Cordeiro e Demétrio (2009), é, geralmente, verdadeiro que a estatística suficiente de um modelo da família exponencial tem, também, distribuição pertencente à família exponencial. Os autores exemplificam que se $Y_1, \dots, Y_n$ são variáveis aleatórias independentes e identicamente distribuídas, com distribuição de Poisson, $P(\theta)$, então a estatística suficiente $\sum_{i=1}^n T(Y_i)$ pertence a mesma família de distribuição Poisson, $P(n\theta)$ e, assim, é um modelo exponencial uniparamétrico.

A seguir, serão definidos os modelos lineares generalizados, cujo componente aleatório é definido a partir da família exponencial uniparamétrica na forma canônica dada em (5) com a introdução de um parâmetro ϕ ($\phi > 0$, de perturbação), o qual refere-se a uma medida de dispersão da distribuição. Considerando-se a introdução desse parâmetro, Nelder e Wedderburn (1972) conseguiram atribuir distribuições biparamétricas ao componente aleatório do modelo, definindo,

$$f(y; \theta, \phi) = \exp\{\phi^{-1}[y\theta - b(\theta)] + c(y, \phi)\} \quad (8)$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas. Quando ϕ é conhecido, a distribuição de probabilidade (ou de densidade) (8) é equivalente à distribuição na forma canônica (5). Ainda, pode-se mostrar que o valor esperado e a variância de Y com distribuição na família (8) são,

$$\begin{aligned} E[Y] &= b'(\theta) \\ Var[Y] &= \phi b''(\theta). \end{aligned} \quad (9)$$

Observe que em (9), a partir da expressão de variância, ϕ é um parâmetro de dispersão do modelo e o seu inverso (ϕ^{-1}) é um parâmetro de precisão.

No mesmo artigo, Nelder e Wedderburn (1972) mostraram que um conjunto de técnicas estatísticas, em geral estudadas isoladamente, podem ser formuladas de forma unificada, como uma classe de modelos de regressão. A essa teoria unificada de modelagem estatística, que é uma classe de modelos de regressão, denominaram modelos lineares generalizados, também chamados modelos exponenciais lineares. Essa classe de modelos

envolve uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de n observações independentes.

A variável resposta, ou componente aleatório do modelo, como definido anteriormente, tem uma distribuição pertencente à família exponencial de distribuições dada em (8), que engloba uma série de distribuições, como por exemplo, a normal, a gama e a normal inversa para dados contínuos, as distribuições Poisson e binomial negativa para dados de contagens e a distribuição binomial para proporções. Desta forma, fica claro que o componente aleatório é definido assim que são estabelecidas as medidas que serão feitas/observadas. Dobson (2002) menciona ainda que o componente aleatório pode ser medido em variáveis contínuas, binárias, nominais (com mais de 2 categorias), ordinais e contagens, todas elas com distribuições de probabilidade pertencentes à família exponencial. McCullagh e Nelder (1989) representam o componente aleatório como um vetor de observações $\mathbf{y}$, contendo n componentes, que é assumido ser a realização de uma variável aleatória Y cujos componentes são independentemente distribuídos com média μ .

As variáveis explanatórias, ou componente sistemático do modelo, entram na forma de uma estrutura linear e são definidas durante o planejamento do experimento, o que é fundamental para a obtenção de conclusões confiáveis a respeito do estudo em questão. Esse componente resulta em modelos de regressão, de análise de variância e de análise de covariância. Dessa forma, a parte sistemática do modelo é especificada por um conjunto de p variáveis explicativas, isto é, $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$, para $i = 1, \dots, n$ que entram no modelo na forma de uma soma linear de seus efeitos, constituindo o vetor de preditores lineares, $\boldsymbol{\eta}$, de dimensão $n \times 1$, com elementos:

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j. \quad (10)$$

ou em notação matricial (em que $\mathbf{X}$ é a matriz modelo $n \times p$ e $\boldsymbol{\beta}$ é um vetor de p parâmetros desconhecidos), pode-se escrever

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

A ligação entre a variável resposta e as variáveis explanatórias, ou ainda, a ligação entre o componente aleatório e o componente sistemático, se dá por meio de uma

função adequada que recebe o nome de função de ligação, a qual será escolhida de acordo com a escala dos dados e conseqüentemente da distribuição considerada, de tal forma que,

$$\eta_i = g(\mu_i)$$

em que $g(\cdot)$ é a função de ligação, que é monótona e derivável e η é o símbolo introduzido para o preditor linear. Quando a função de ligação é escolhida de tal forma que $g(\mu_i) = \theta_i = \eta_i$, tem-se uma função de ligação canônica, em que o preditor linear modela diretamente o parâmetro canônico θ_i . Nesses casos, segundo Cordeiro e Demétrio (2009), os modelos correspondentes são denominados modelos canônicos, o que resulta, frequentemente, em uma escala adequada para a modelagem com interpretação prática para os parâmetros de regressão, além de vantagens teóricas em termos da existência de um conjunto de estatísticas suficientes para o vetor de parâmetros e alguma simplificação no algoritmo de estimação.

Por exemplo, no caso de dados distribuídos segundo uma distribuição Poisson, a função de ligação canônica é a logarítmica e o modelo de regressão usual é o modelo de regressão de Poisson, dado por:

$$\eta_i = \ln(\theta_i) = \beta_1 + \beta_2 x_1 + \dots + \beta_p x_p$$

Deste modo, um modelo linear generalizado fica definido por uma variável resposta cuja distribuição é um membro da família exponencial, um conjunto de variáveis explicativas descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear.

2.1.4 Superdispersão

Em algumas situações reais, no ajuste de um MLG, é comum a variabilidade descrita pelo modelo ser menor do que aquela observada no fenômeno aleatório. Por exemplo, no ajuste de um MLG, considerando que a distribuição da variável resposta seja Poisson com parâmetro λ , $P(\lambda)$, pode ocorrer da variância da variável aleatória observada Y ser maior que a média. Neste caso, diz-se que há uma variabilidade extra-Poisson inerente ao fenômeno estudado, que é também chamada de superdispersão, ou seja, a variação predita pelo modelo é inferior à variação observada na amostra (HINDE E DEMÉTRIO, 1998). Neste texto, será tratada a superdispersão em dados na forma de contagens, ou seja,

considerando uma distribuição de Poisson para descrever o comportamento da variável observada, tem-se $Var[Y_i] > \lambda_i$, todavia, pode também ocorrer superdispersão para dados na forma de proporção, em que supondo que o comportamento da variável resposta observada seja descrito pela distribuição Binomial de parâmetros m e π , o fenômeno da superdispersão é caracterizado por $Var[Y_i] > m_i\pi_i(1 - \pi_i)$.

De acordo com Hinde e Demétrio (1998), a superdispersão pode ocorrer devido a vários motivos, dentre eles:

- i - variabilidade do material experimental, isto é, presença de alguma variabilidade individual da unidade experimental que não foi considerada no modelo como um componente de variabilidade;
- ii - correlação entre indivíduos, por exemplo, em estudos para testes de fármacos com ninhadas de ratos, pode-se esperar que haja uma correlação entre ratos de uma mesma ninhada;
- iii - agrupamentos amostrais;
- iv - dados de nível agregado, em que, o processo de agregação pode levar a distribuições compostas;
- v - variáveis não observadas, isto é, a não observação e/ou inclusão de variáveis explanatórias no modelo que poderiam contribuir na descrição do comportamento da variável resposta.

Identificar a causa exata do fenômeno da superdispersão em geral não é trivial, todavia, detectada a sua ocorrência, modelos apropriados devem ser especificados, pois caso não seja levada em conta a variabilidade extra à explicada pelo modelo, as estimativas dos parâmetros do modelo podem apresentar erros padrões incorretos e subestimados, o que afeta as conclusões obtidas a partir do modelo de regressão ajustado (HINDE E DEMÉTRIO, 1998). Dados com superdispersão na forma de contagens são discutidos mais profundamente por Breslow (1984) e Lawless (1987), enquanto que Vieira, Hinde e Demétrio (2000) apresentam um estudo para a superdispersão ocasionada pela inflação de zeros em dados obtidos na forma de proporção, além de dados na forma de contagens.

Detectada a superdispersão, Hinde e Demétrio (1998) sugerem duas formas de se tratar a superdispersão:

- (i) considerar uma forma mais geral para a função de variância, possivelmente com a inclusão de novos parâmetros;
- (ii) considerar o ajuste de um modelo de dois estágios para a variável resposta.

Dado que a variável resposta representa contagens com média λ_i , a forma geral para a variância pode ser expressa por

$$Var(Y_i) = \lambda_i \{1 + \phi \lambda_i^\delta\}, \quad (11)$$

em que diferentes valores de δ e ϕ em (11) definem diferentes funções de variância:

- se $\phi = 0$, tem-se a função de variância do modelo de Poisson padrão, ou seja, $Var(Y_i) = \lambda_i$;
- se $\delta = 0$, define-se o modelo de superdispersão constante reparametrizado, em que $Var(Y_i) = \lambda_i \{1 + \phi\}$;
- se $\delta = 1$ define a função de variância do tipo da distribuição binomial negativa, ou seja, $Var(Y_i) = \lambda_i \{1 + \phi \lambda_i\}$.

Rizzato (2011) observa que esses modelos, geralmente, não correspondem à uma distribuição de probabilidade, sendo vistos como uma extensão do modelo básico. Entretanto, modelos em dois estágios, conforme Hinde e Demétrio (1998) e Rizzato (2011), levam a modelos de probabilidade composta para a resposta.

Como definido por Hinde e Demétrio (1998) e citado por Rizzato (2011), um modelo de dois estágios assume que $Y_i | \lambda_i \sim \text{Poisson}(\lambda_i)$ em que λ_i é uma variável aleatória com $E(\lambda_i) = \mu_i$ e $Var(\lambda_i) = \sigma_i^2$. Como consequência, tem-se que $E(Y_i) = \mu_i$ e $Var(Y_i) = \mu_i + \sigma_i^2$, o que representa um modelo com superdispersão. Um caso específico é considerar que $\lambda_i \sim \Gamma(k, \theta_i)$, resultando em Y_i ter distribuição binomial negativa com $E(Y_i) = \frac{k}{\theta_i} = \mu_i$ e $Var(Y_i) = \mu_i + \frac{\mu_i^2}{k}$. Para valores fixos de k essa distribuição pertence à família exponencial.

2.1.5 Modelos Lineares Generalizados Mistos

A teoria dos modelos lineares generalizados considera apenas o caso em que todos os efeitos do modelo são fixos. Uma extensão desse modelo é considerar a inclusão

de termo(s) aleatório(s) no preditor linear. Essa extensão recebe o nome de modelo linear generalizado misto.

Os modelos lineares generalizados mistos, de acordo com Breslow e Clayton (1993), são úteis para acomodar a superdispersão frequentemente observada em dados cujas distribuições sejam binomial (WILLIAMS, 1982) ou Poisson (BRESLOW, 1984), e provenientes de estudos com dados longitudinais ou medidas repetidas (STIRATELLI, LAIRD e WARE, 1984; ZEGER, LIANG e ALBERT, 1988).

Segundo Hinde e Demétrio (1998) após o ajuste de um modelo linear generalizado a um determinado conjunto de dados, considerando a inclusão de todas as variáveis significativas no estudo, é possível que a variação residual seja maior do que a variação admitida pelo modelo. Isso pode ser o indício de uma superdispersão, por exemplo, no ajuste de um modelo Poisson. Nessas situações, devem-se procurar variáveis explanatórias adicionais, ou incluir no modelo uma segunda fonte de variação que acomode esta variabilidade extra.

A classificação de um efeito como fixo ou aleatório está associada aos objetivos da análise. Segundo Barbin (1998), entende-se como efeitos aleatórios aqueles representativos de tratamentos (ou fatores) de uma amostra oriunda de uma determinada população. Efeitos fixos ou parâmetros são aqueles cujas considerações são limitadas, isto é, são restritas aos tratamentos (ou níveis) ensaiados. Assim, os modelos podem ser classificados da seguinte maneira:

- Modelos de efeitos fixos: quando todos os componentes do modelo, exceto o erro, são considerados fixos, constituindo-se um modelo de efeitos fixos ou, *Modelo I*, segundo Eisenhart (1947) *apud* Cordeiro e Demétrio (2009);
- Modelos aleatórios: quando todos os componentes do modelo, exceto a média, são considerados efeitos aleatórios, definindo-se um modelo aleatório ou *Modelo II*.
- Modelos mistos: quando têm-se efeitos fixos que não sejam representados apenas pela média e efeitos aleatórios que não sejam constituídos apenas pelos erros, determinando-se um modelo de efeitos fixos e aleatórios, chamado *Modelo Misto*.

Breslow e Clayton (1993) propuseram a adição de efeitos aleatórios no preditor linear do modelo linear generalizado, definindo o modelo linear generalizado misto, considerando que esses efeitos são normalmente distribuídos com componentes de variância

a serem estimados a partir dos dados. Exemplos do uso de modelos lineares generalizados mistos podem ser vistos em Diggle et al. (2002), Verbeke e Molenberghs (2000, 2005), Molenberghs et al. (2007, 2010), McCulloch et al. (2008), Urbano (2012) e Rizzato (2011).

Considere que $\mathbf{Y}$ seja um vetor de variáveis aleatórias condicionalmente independentes, e suponha a distribuição de $\mathbf{Y}$ dado $\mathbf{b}$ com função densidade de probabilidade pertencente à família exponencial, tal que,

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b} &\sim f_{Y_i | \mathbf{b}}(y_i | \mathbf{b}) \\ f_{Y_i | \mathbf{b}}(y_i | \mathbf{b}) &= \exp \left\{ \phi^{-1} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\}. \end{aligned} \quad (12)$$

De (9) observa-se que $\mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i}$, de forma que, $E(\mathbf{Y} | \mathbf{b}) = \boldsymbol{\mu}$ e ainda,

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b},$$

é a função de ligação em (10), em que $\mathbf{X}$ é a matriz de delineamento (ou matriz modelo), $\boldsymbol{\beta}$ é o vetor de efeitos fixos, $\mathbf{Z}$ é a matriz de delineamento para os efeitos aleatórios e $\mathbf{b}$ é o vetor de efeitos aleatórios. Sob estas condições, μ_i é a média condicional de Y_i dado $\mathbf{b}$. O modelo fica totalmente especificado atribuindo-se uma distribuição de probabilidades aos efeitos aleatórios, $\mathbf{b} \sim f_B(\mathbf{b})$, que conforme Breslow e Clayton (1993), é usual ser normal. Lee e Nelder (1996, 2001) expandiram os modelos propostos por Breslow e Clayton (1993) definindo os modelos lineares generalizados hierárquicos em que o preditor linear pode ser composto por efeitos fixos e efeitos aleatórios não normais. Fabio, Paula e Castro (2012) apresentam um modelo Poisson com intercepto aleatório assumindo que os efeitos aleatórios são descritos por uma distribuição log-gama, com o objetivo de acomodar a superdispersão inerente a dados de contagens e induzir a correlação intra-classe. Uma discussão sobre efeitos aleatórios não normais também é apresentada por Nelson et al. (2006). Considerando em (12) que $\mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, decorre que,

$$E(\mathbf{Y}) = E[E(\mathbf{Y} | \mathbf{b})] = E(\boldsymbol{\mu}) = E[g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})].$$

Observe que a variância marginal de $\mathbf{Y}$ fica expressa por:

$$\begin{aligned} \text{Var}(\mathbf{Y}) &= \text{Var}[E(\mathbf{Y}|\mathbf{b})] + E[\text{Var}(\mathbf{Y})] \\ &= \text{Var}(\boldsymbol{\mu}) + E[\phi^{-1}\text{Var}(\boldsymbol{\mu})] \\ &= \text{Var}[g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})] + E\{\phi^{-1}\text{Var}[g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})]\}. \end{aligned}$$

Assumindo-se independência condicional dos elementos de $\mathbf{Y}$, tem-se,

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \text{Cov}[E(\mathbf{Y}_i|\mathbf{b}), E(\mathbf{Y}_j|\mathbf{b})] + E[\text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j|\mathbf{b})] \\ &= \text{Cov}(\mu_i, \mu_j) + E(0) \\ &= \text{Cov}[g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}), g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})]. \end{aligned}$$

2.1.6 Modelos Para Dados de Contagens Longitudinais Com Superdispersão

Considerando dados na forma de contagem e o objetivo em ajustar um modelo que descreva o comportamento dessa variável resposta em função de outra(s) explicatória(s), pode-se considerar o ajuste de um MLG que leve em conta a natureza da variável resposta, no caso, um modelo de regressão de Poisson. Na ocorrência de superdispersão, pode-se optar ainda pelo ajuste de um MLG, considerando agora a distribuição Binomial Negativa, por exemplo. Todavia, estes modelos, não levam em consideração uma possível estrutura longitudinal nos dados. No caso de dados de contagens longitudinais, pode-se considerar o ajuste de um MLGM, como proposto por Breslow e Clayton (1993), em que a estrutura temporal é modelada pela inserção de efeitos aleatórios no modelo que captam a correlação entre medidas tomadas no mesmo indivíduo. Nos modelos citados anteriormente, a estrutura longitudinal e a superdispersão são tratadas isoladamente. Uma vez que tais características podem ocorrer simultaneamente, surge a necessidade em criar modelos que comportem ambas.

Para a análise de dados de contagens longitudinais com superdispersão, Booth et al. (2003) sugerem o ajuste do modelo Poisson log-linear com efeitos aleatórios, enquanto que Molenberghs, Verbeke e Demétrio (2007) e Vangeneugden et al. (2009) considerando a mesma estrutura, apresentam as expressões do vetor de médias e da matriz de covariâncias para o modelo. Rizzato (2011), apresenta a aproximação bayesiana para a proposta de Molenberghs, Verbeke e Demétrio (2007), que serão particularizados, a seguir.

Conforme Rizzato (2011), seja Y_{ij} uma variável aleatória, representando a j -ésima medida longitudinal do i -ésimo indivíduo, $i = 1, \dots, N$, $j = 1, \dots, n_i$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ o vetor de todas as medidas da unidade i e $\mathbf{x}_{ij}$ um vetor de covariáveis associado a Y_{ij} . Os quatros modelos, a seguir, podem ser considerados para a análise de dados longitudinais na forma de contagens.

2.1.6.1 Modelo 1: Poisson

Considere que Y_{ij} são independentes, com distribuição de Poisson com parâmetro λ_{ij} , isto é, $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$. Então,

$$P(Y_{ij} = y_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}, \quad \text{para } y_{ij} = 0, 1, 2, \dots$$

e admite-se que

$$\ln(\lambda_{ij}) = \mathbf{x}_{ij}'\boldsymbol{\beta},$$

ou ainda,

$$\lambda_{ij} = \exp\left(\mathbf{x}_{ij}'\boldsymbol{\beta}\right).$$

2.1.6.2 Modelo 2: Poisson-gama

Considere Y_{ij} com distribuição de Poisson com parâmetro λ_{ij} , e para acomodar a superdispersão considere um efeito aleatório θ_{ij} com distribuição gama, isto é,

$$\begin{aligned} Y_{ij} | \lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \theta_{ij} \exp(\mathbf{x}_{ij}'\boldsymbol{\beta}) \\ \theta_{ij} &\sim \text{gama}(\alpha_{1j}, \alpha_{2j}) \end{aligned}$$

sendo α_{1j} e α_{2j} parâmetros da distribuição gama.

2.1.6.3 Modelo 3: Poisson-normal

Considere Y_{ij} com distribuição de Poisson com parâmetro λ_{ij} , e que o efeito aleatório $\mathbf{b}_i$, para acomodar a correlação entre medidas longitudinais, seja normalmente distribuído. Então,

$$\begin{aligned} Y_{ij} | \lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \ln(\lambda_{ij}) &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}), \end{aligned}$$

em que $\mathbf{D}$ é a matriz de covariâncias de $\mathbf{b}_i$.

2.1.6.4 Modelo 4: Poisson-normal-gama

Considere que Y_{ij} são descritos por uma distribuição de Poisson com parâmetro λ_{ij} . Diferentes tipos de efeitos aleatórios serão incluídos para acomodar a correlação entre medidas feitas no mesmo indivíduo e a superdispersão, conforme Molenberghs, Verbeke e Demétrio (2007).

Supondo que,

$$\begin{aligned} \lambda_{ij} &= \theta_{ij} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \end{aligned}$$

e que θ_{ij} tem distribuição com

$$E[\boldsymbol{\theta}_i] = E[(\theta_{i1}, \dots, \theta_{in_i})'] = \boldsymbol{\Psi}_i \quad \text{e} \quad \text{Var}[\boldsymbol{\theta}_i] = \boldsymbol{\Sigma}_i.$$

Considere que $\mathbf{Y}_i$ é o vetor de tamanho n_i com as medidas do i -ésimo indivíduo e $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{in_i})'$ com $E(\boldsymbol{\lambda}_i) = \boldsymbol{\mu}_i$ e $\text{Var}(\boldsymbol{\lambda}_i) = \boldsymbol{\Sigma}_i$. Seguindo Molenberghs, Verbeke e Demétrio (2007), o vetor de médias $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$ tem componentes

$$\mu_{ij} = \phi_{ij} \exp \left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2} \mathbf{z}'_{ij} \mathbf{D} \mathbf{z}_{ij} \right)$$

e a matriz de variâncias-covariâncias de $\mathbf{Y}_i$ é dada por,

$$Var(\mathbf{Y}_i) = M_i + M_i(P_i - J_{n_i}M_i),$$

em que M_i é a matriz diagonal que tem como elementos o vetor $\boldsymbol{\mu}_i$, J_{n_i} é uma matriz unitária de dimensão $n_i \times n_i$ e P_i é uma matriz que tem como elementos

$$p_{i,jk} = \exp\left(\frac{1}{2}\mathbf{z}_{ij}'\mathbf{D}\mathbf{z}_{ij}\right) \frac{\sigma_{i,jk} + \phi_{ij}\phi_{ik}}{\phi_j\phi_{ik}} \exp\left(\frac{1}{2}\mathbf{z}_{ij}'\mathbf{D}\mathbf{z}_{ij}\right).$$

Como caso particular, assumindo-se a distribuição gama para o efeito aleatório θ_{ij} para acomodar a superdispersão, tem-se,

$$\begin{aligned} Y_{ij}|\mathbf{b}_i, \theta_{ij} &\sim Poisson(\lambda_{ij}) \\ \lambda_{ij} &= \theta_{ij} \exp(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{b}) \\ \mathbf{b}_i &\sim \mathbf{N}(\mathbf{0}, \mathbf{D}) \\ \theta_{ij} &\sim gama(\alpha_{1j}, \alpha_{2j}) \end{aligned}$$

em que $\mathbf{D}$ a matriz de variâncias-covariâncias referentes a $\mathbf{b}_i$ e α_{1j} e α_{2j} parâmetros da distribuição gama, conforme Molenberghs, Verbeke e Demétrio (2007) e Rizzato (2011).

2.1.7 Inferência Bayesiana

Segundo O'Hagan (1996), o problema fundamental da teoria estatística é o processo de inferência. A partir de alguns dados observados, deseja-se fazer declarações, inferências, sobre uma ou mais características desconhecidas a respeito do mecanismo gerador desses dados. Esse processo inferencial pode ser conduzido utilizando a Estatística Bayesiana, que é uma teoria alternativa completa de estatística, diferente da estatística frequentista.

De acordo com Paulino et al. (2003), o Rev. Thomas Bayes, em 1763, em sua obra póstuma “*An essay towards solving a problem in the doctrine of chances*” comunicada à Royal Statistical Society por Richard Price (Phil. Trans. Roy. Soc. **53**, 370-418), apresentou uma metodologia estatística moderna que leva seu nome: a Inferência Estatística Bayesiana. O resultado central do seu trabalho é conhecido como “Teorema de Bayes” que

em sua forma mais simples, conforme Leandro (2001), pode ser dado por: se H denota uma hipótese e D denota os dados, então o teorema estabelece que

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D)}$$

em que $P(H)$ é considerada como representando o conhecimento sobre H antes da observação dos dados e $P(H|D)$ representando o conhecimento sobre H atualizado pelos dados. Desta forma, especificado $P(D|H)$ e $P(H)$, o teorema fornece uma solução do problema de como aprender/atualizar a informação por meio dos dados.

Para se definir formalmente o Teorema de Bayes, considere o espaço de probabilidade $(\Omega, \mathcal{A}, P)$ e seja uma partição finita ou infinita de Ω , dada por

$$A_1, A_2, \dots, A_m, P(A_i) > 0, A_i \cap A_j = \emptyset, i \neq j, \cup_i A_i = \Omega.$$

Suponha um outro evento B qualquer, tal que $P(B) > 0$. Pode-se verificar a decomposição de B na união de conjuntos disjuntos da seguinte forma,

$$B = \cup_i (A_i \cap B).$$

Assim, tem-se que,

$$P(B) = \sum_i P(A_i \cap B),$$

e,

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B),$$

ou, equivalentemente,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)},$$

que é chamado de Teorema de Bayes.

Uma forma de interpretar o Teorema de Bayes, de acordo com Paulino (2003), como citado por Rocha (2010), consiste em considerar os eventos $A_i = 1, 2, \dots, m$, como

“antecedentes”, “causas”, “hipóteses” ou “estados” a que o investigador atribui graus de credibilidade (incorpora incerteza) ou probabilidades a priori $P(A_i)$, $i = 1, 2, \dots, m$, de natureza subjetiva. Depois da informação adicional que consiste em saber que o evento B se realizou (o evento B pode ser a observação de um conjunto de dados), o pesquisador revê as suas probabilidades a priori usando a fórmula de Bayes e passa a atribuir aos A_i , $i = 1, 2, \dots, m$ as probabilidades a posteriori $P(A_i|B)$, $i = 1, 2, \dots, m$.

Seja θ uma quantidade desconhecida de interesse. Considere que a informação que se dispõe a respeito de θ , a priori de se observar um experimento, possa ser resumida probabilisticamente por $\pi(\theta)$, a qual pode ser atualizada observando-se uma quantidade de interesse aleatória Y , cuja distribuição de probabilidade está relacionada com θ . Da teoria de verossimilhança, sabe-se que essa relação pode ser descrita por meio da função de verossimilhança, $L(\theta|y)$. Segundo Gelman et al. (2004), após se observar $Y = \mathbf{y}$ é possível atualizar a informação que se tem a respeito de θ , considerando como regra de atualização o Teorema de Bayes,

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\theta, \mathbf{y})}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\theta)\pi(\theta)}{\pi(\mathbf{y})} = \frac{L(\theta|\mathbf{y})\pi(\theta)}{\pi(\mathbf{y})} \quad (13)$$

em que $\pi(\theta; \mathbf{y}) = L(\theta|\mathbf{y})$ é a função de verossimilhança, $\pi(\theta)$ é chamada de distribuição a priori e $\pi(\theta|\mathbf{y})$ é a informação atualizada a respeito de θ , chamada de distribuição a posteriori.

Em (13), o termo $\pi(\mathbf{y})$ pode ser substituído por $\pi(\mathbf{y}) = \sum_{\theta} \pi(\theta)L(\theta|\mathbf{y})$, quando θ for um parâmetro discreto, ou por $\pi(\mathbf{y}) = \int \pi(\theta)L(\theta|\mathbf{y})d\theta$, quando θ for um parâmetro contínuo. Segundo Gelman et al. (2004), uma forma equivalente de (13), comumente utilizada na inferência bayesiana até os dias atuais, consiste em omitir o fator $\pi(\mathbf{y})$, uma vez que não depende de θ , e que fixado $\mathbf{y}$ pode ser considerado como uma constante, produzindo uma distribuição não normalizada a posteriori, expressa por,

$$\pi(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y})\pi(\theta). \quad (14)$$

A forma simplificada do Teorema de Bayes (14) é bastante útil e utilizada em problemas que envolvam estimação de parâmetros, um vez que o denominador é apenas uma constante normalizadora. Essa constante normalizadora (k) da posteriori pode ser

calculada,

$$k^{-1} = \pi(\mathbf{y}) = \int \pi(\mathbf{y}, \theta) d\theta = \int L(\theta|\mathbf{y}) \pi(\theta) d\theta = E_{\theta}[p(Y|\theta)],$$

a qual é chamada de distribuição marginal de Y , mas segundo Gelman et al. (2004), um nome mais informativo é distribuição preditiva a priori. Os autores ainda justificam que é priori porque ela não está condicionada a uma observação prévia do processo, e é preditiva porque ela é a distribuição esperada para a observação y dado θ .

Até o momento, pode-se concluir que a função de verossimilhança $L(\theta|\mathbf{y})$ desempenha um papel muito importante na fórmula de Bayes, e portanto, na inferência bayesiana, pois é por meio desta função que o conhecimento a priori sobre o parâmetro desconhecido, θ , é atualizado/modificado. A distribuição a priori não precisa ser uma densidade própria, isto é, não necessita integrar um, mas quando utilizada é necessário cautela, pois a escolha de uma distribuição a priori imprópria pode conduzir a uma distribuição a posteriori imprópria, o que seria incorreto, uma vez que a distribuição a posteriori é uma função densidade de probabilidade. Assim, fica claro que a distribuição a priori desempenha um papel importante na inferência bayesiana.

A utilização da informação a priori no contexto da inferência bayesiana requer a especificação de uma distribuição de probabilidades a priori que descreva a informação que se tem a respeito do parâmetro θ . Neste contexto, alguns métodos de obtenção de priori são discutidos na literatura, todavia o uso de distribuição a priori conjugada é muito comum e desejável, pois, em geral, facilita os cálculos.

Segundo Gelman et al. (2004), se $\mathcal{F}$ é uma classe de distribuições amostrais $\pi(\theta|\mathbf{y})$, e $\mathcal{P}$ é uma classe de distribuições a priori para θ , então a classe $\mathcal{P}$ é conjugada para $\mathcal{F}$ se

$$\pi(\theta|\mathbf{y}) \in \mathcal{P} \quad \forall \quad \pi(\cdot|\theta) \in \mathcal{F} \quad \text{e} \quad \pi(\cdot) \in \mathcal{P}.$$

O interesse está em prioris conjugadas naturais, que surgem definindo $\mathcal{P}$ como sendo o conjunto de todas as densidades que possuem a mesma forma funcional da verossimilhança. Distribuições a priori conjugadas têm vantagens práticas, além de computacionais, pois uma vez que as distribuições a posteriori e priori pertencem à mesma família, a atualização do conhecimento prévio que se tem de θ , envolve apenas uma modificação nos hiperparâmetros.

Como exemplificação, seja $\mathbf{Y} = \{y_1, \dots, y_n\}$ uma amostra aleatória de uma distribuição de Poisson com parâmetro θ . Tem-se que,

$$f(y_i; \theta) = \frac{e^{-\theta} \theta^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots, \quad \theta > 0.$$

A função de probabilidade conjunta pode ser expressa por:

$$f(\mathbf{y}; \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{y_i}}{y_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \propto e^{-n\theta} \theta^{\sum_{i=1}^n y_i}.$$

Observe que o núcleo da verossimilhança é da forma $\theta^{c_1} e^{-c_2 \theta}$, que caracteriza a família de distribuições gama. Desta forma, a distribuição conjugada natural a priori é gama com parâmetros a_1 e a_2 , $G(a_1, a_2)$, com $a_1, a_2 > 0$ e conhecidos.

$$\pi(\theta) = \frac{a_2^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-a_2 \theta} \propto \theta^{a_1-1} e^{-a_2 \theta}, \quad \theta > 0.$$

Portanto, a distribuição a posteriori,

$$\pi(\theta|\mathbf{y}) = \frac{L(\theta|\mathbf{y})\pi(\theta)}{\int_0^\infty L(\theta|\mathbf{y})\pi(\theta)d\theta}$$

será expressa por:

$$\pi(\theta|\mathbf{y}) \propto \theta^{a_1+n\bar{y}-1} e^{-(n+a_2)\theta}$$

ou seja,

$$\theta|\mathbf{y} \sim G(a_1 + n\bar{y}; n + a_2).$$

Outras informações a respeito da utilização de outros tipos de distribuições a priori podem ser obtidas de forma mais detalhada em Berger (1985), Box e Tiao (1992), Bernardo e Smith (2000) e Paulino et al. (2003).

Fica claro que as inferências sobre as quantidades desconhecidas no modelo, considerando o paradigma bayesiano, são obtidas por meio da distribuição a posteriori, que no caso de modelagem, em geral, dado o número de parâmetros, será baseada na distribuição marginal a posteriori de interesse. Por meio da distribuição marginal a posteriori é possível

a obtenção dos resumos a posteriori, tais como média, mediana, percentis, e com estes fazer conclusões pontuais e testar hipóteses. Todavia, estas distribuições, em geral, não são matematicamente tratáveis, e para contornar este problema, métodos computacionalmente intensivos e aproximações determinísticas podem ser utilizados. Existem vários métodos específicos para a obtenção da distribuição a posteriori, sendo os métodos de Monte Carlo com cadeias de Markov os mais utilizados, até os dias de hoje, sobretudo depois da criação do programa WinBUGS (LUNN et al., 2000), e sua atual versão *open source*, OpenBUGS (LUNN et al., 2009).

2.1.7.1 Aproximação de Laplace e Método de Monte Carlo

Considerando a abordagem bayesiana, todas as inferências são realizadas por meio da distribuição conjunta a posteriori, pois ela contém toda a informação a respeito das quantidades desconhecidas. Resumos a posteriori como média, moda, mediana, desvio-padrão e percentis podem ser obtidos por meio da distribuição a posteriori. A obtenção dessas medidas resumo a posteriori estão, em geral, associadas ao cálculo de integrais, denominada por alguns autores, como integrais bayesianas, as quais, em sua grande maioria, não apresentam solução analítica.

Uma vasta literatura está disponível a respeito de métodos para a resolução destas integrais. A seguir, será descrita a aproximação de Laplace, que é um método determinístico, em seguida, serão descritos dois métodos estocásticos.

2.1.7.2 Aproximação de Laplace

Para apresentar a aproximação de Laplace é necessário primeiramente apresentar a definição de expansão em série de Taylor. A expansão em série de Taylor surgiu do interesse em encontrar uma série infinita que converge para uma dada função, de tal forma que utilizando essa informação se possa calcular um valor aproximado da função por meio de somas parciais dessa série.

Considere que f seja uma função arbitrária, contínua e suave em seu domínio de definição. O objetivo é estudar o comportamento da função em torno de um determinado ponto fixo, arbitrário, $x = x_0$, sendo que o valor da função no ponto $x = x_0$ é $f(x_0)$. Dessa forma, tem-se uma aproximação da função $f(x)$ em torno de x_0 . Sob estas condições,

pode-se definir:

$$\begin{aligned} f(x) &= f(x)\Big|_{x=x_0} + \frac{f'(x)}{1!}\Big|_{x=x_0} (x-x_0) + \frac{f''(x)}{2!}\Big|_{x=x_0} (x-x_0)^2 + \\ &\dots + \frac{f^{(n)}(x)}{n!}\Big|_{x=x_0} (x-x_0)^n + \dots \end{aligned}$$

Suponha que se deseje aproximar $p(x) = \frac{f(x)}{k}$, $f(x) \geq 0$ e k é uma constante de normalização. Considere inicialmente, a expansão em série de Taylor de $\ln f(x)$ em torno de $x = x_0$,

$$\begin{aligned} \ln f(x) &= \ln f(x)\Big|_{x=x_0} + \underbrace{\frac{d}{dx} \ln f(x)\Big|_{x=x_0}}_{\text{segundo termo}} (x-x_0) \\ &+ \frac{1}{2} \frac{d^2}{dx^2} \ln f(x)\Big|_{x=x_0} (x-x_0)^2 + \dots \end{aligned} \quad (15)$$

Considerando que os termos de mais alta ordem são negligenciáveis, como em Guerra (2008), e tomando o segundo termo da expansão (15) para $\ln f(x)$ pode-se observar:

$$\frac{d}{dx} \ln f(x)\Big|_{x=x_0} (x-x_0) = \frac{1}{f(x)} \underbrace{\frac{d}{dx} f(x)\Big|_{x=x_0}}_{(*)} (x-x_0).$$

Tomando $f(x)$ como sendo uma função densidade de probabilidade, nota-se que em (*), o valor que anula a derivada determina o máximo local da função. Se for possível determinar esse máximo local e expandir a série de Taylor em torno desse ponto x_{max} , garante-se que o segundo termo da expansão em série de Taylor de $\ln f(x)$ seja sempre zero. Isto é obtido resolvendo $\frac{d}{dx} f(x) = 0$.

Conforme Guerra (2008), tomando os três primeiros termos da expansão em série de Taylor em torno de $x = x_{max}$, tem-se:

$$\begin{aligned} \ln f(x) &= \ln f(x_{max}) + 0 + \frac{1}{2} \frac{d^2}{dx^2} \ln f(x)\Big|_{x=x_{max}} (x-x_{max})^2 \\ e^{\ln f(x)} &= \exp \left[\ln f(x_{max}) + \frac{1}{2} \frac{d^2}{dx^2} \ln f(x)\Big|_{x=x_{max}} (x-x_{max})^2 \right]. \end{aligned} \quad (16)$$

Integrando em relação a x , ambos os lados da igualdade em (16), tem-se,

$$\int e^{\ln f(x_{max})} dy = \underbrace{e^{\ln f(x_{max})}}_{\text{constante}} \int \exp \left[\underbrace{\frac{1}{2} \frac{d^2}{dx^2} \ln f(x) \Big|_{x=x_{max}}}_{\text{constante}} (x - x_{max})^2 \right] dx$$

restando algumas constantes e um termo quadrático em x .

Suponha ainda que $\ln f(x) = L(x)$ e tome $\sigma^2 = -\frac{1}{L''(x_{max})}$, tem-se o resultado a seguir que assemelha-se a uma função densidade de probabilidade normal:

$$\int e^{L(x_{max})} dx \approx e^{L(x_{max})} \int \exp \left[-\frac{(x - x_{max})^2}{2\sigma^2} \right] dx.$$

Este resultado é utilizado por Tierney e Kadane (1986), para a determinação da aproximação de Laplace para integrais, considerando $x_{max} = x^*$, da seguinte forma:

$$\int e^{nL(x^*)} dx \approx e^{nL(x^*)} \int \exp \left[-\frac{n(x - x^*)^2}{2\sigma^2} \right] dx = \sqrt{2\pi\sigma} e^{nL(x^*)}.$$

Retornando ao contexto bayesiano, sabe-se que todas as inferências são baseadas na distribuição a posteriori do parâmetro de interesse, θ , ou $\boldsymbol{\theta}$, no caso deste ser um vetor de parâmetros a serem estimados. Desta forma, dada a verossimilhança $L(\theta|\mathbf{y})$ e a densidade a priori $\pi(\theta)$, o início da inferência bayesiana está baseada na densidade a posteriori definida por:

$$\pi(\theta|\mathbf{y}) = \frac{L(\theta|\mathbf{y})\pi(\theta)}{\int L(\theta|\mathbf{y})\pi(\theta)d\theta}. \quad (17)$$

Suponha o interesse em calcular momentos a posteriori da forma,

$$E[g(\theta)|Y] = \int g(\theta)\pi(\theta|\mathbf{y})d\theta \quad (18)$$

em que $g(\theta)$ é uma função de interesse. Uma vez que a distribuição a posteriori de θ é expressa por (17), a expressão (18) pode ser escrita da seguinte forma,

$$E[g(\theta)|Y] = \frac{\int g(\theta)L(\theta|\mathbf{y})\pi(\theta)d(\theta)}{\int L(\theta|\mathbf{y})\pi(\theta)d\theta} \quad (19)$$

em que θ é o parâmetro desconhecido, podendo, também, ser um vetor de parâmetros

desconhecidos, e $\mathbf{y}$ é o vetor de dados.

Supondo que $g(\theta)$ seja uma função positiva, (19) pode ser reescrita por,

$$E[g(\theta)|Y] = \frac{\int e^{-nh^*(\theta)} d\theta}{\int e^{-nh(\theta)} d\theta} \quad (20)$$

em que $-nh(\theta) = \ln \pi(\theta) + \ln L(\theta|\mathbf{y})$ e $-nh^*(\theta) = \ln g(\theta) + \ln \pi(\theta) + \ln L(\theta|\mathbf{y})$. Sendo $\hat{\theta}$ o valor que maximiza $-h(\theta)$ e $\hat{\theta}^*$ o valor que maximiza $-h^*(\theta)$, definindo $\hat{\sigma} = [h''(\hat{\theta})]^{1/2}$ e $\hat{\sigma}^* = [h^{*''}(\hat{\theta}^*)]^{1/2}$, as aproximações de Laplace para o numerador e denominador de (20) são expressas respectivamente por,

$$\int e^{-nh^*(\theta)} d\theta \approx \sqrt{2\pi\hat{\sigma}^*} n^{-1/2} e^{-nh^*(\hat{\theta}^*)} \quad (21)$$

$$\int e^{-nh(\theta)} d\theta \approx \sqrt{2\pi\hat{\sigma}} n^{-1/2} e^{-nh(\hat{\theta})}. \quad (22)$$

Observe que a aproximação de Laplace consiste de aproximação normal para os integrandos, de forma que, segundo Tierney e Kadane (1986), esta aproximação irá produzir resultados razoáveis desde que a distribuição a posteriori seja unimodal ou pelo menos o denominador possua uma moda única. Assume-se que o tamanho da amostra seja grande, além de que essa expressão seja suave e não nula no espaço paramétrico, exigindo que esta função seja positiva.

Tomando a razão de (21) e (22), tem-se a seguinte aproximação para (19)

$$E[g(\hat{\theta})|\mathbf{y}] \approx \left(\frac{\hat{\sigma}^*}{\hat{\sigma}} \right) e^{-n[h^*(\hat{\theta}^*) - h(\hat{\theta})]}.$$

Uma aproximação similar pode ser feita no caso em que $\boldsymbol{\theta}$ é um vetor de parâmetros de interesse, obtendo, de acordo com Tierney e Kadane (1986):

$$E[g(\hat{\boldsymbol{\theta}})|\mathbf{y}] \approx \left(\frac{\det \boldsymbol{\Sigma}^*}{\det \boldsymbol{\Sigma}} \right)^{1/2} e^{-n[h^*(\hat{\boldsymbol{\theta}}^*) - h(\hat{\boldsymbol{\theta}})]},$$

em que $\hat{\boldsymbol{\theta}}^*$ e $\hat{\boldsymbol{\theta}}$ são os vetores que maximizam h^* e h , respectivamente, e $\boldsymbol{\Sigma}^*$ e $\boldsymbol{\Sigma}$ são menos o inverso da matriz hessiana de h^* e h em relação a $\hat{\boldsymbol{\theta}}^*$ e $\hat{\boldsymbol{\theta}}$, respectivamente, conforme descrito em Tierney e Kadane (1986).

2.1.7.3 Métodos de Monte Carlo com cadeias de Markov

Os métodos de Monte Carlo com cadeias de Markov (ou MCMC), são algoritmos não determinísticos baseados em simulações estocásticas, ou seja, requerem a simulação de números (pseudo)aleatórios de alguma distribuição de probabilidade.

Os métodos objetivam a obtenção de uma amostra da distribuição conjunta a posteriori para a obtenção de estimativas amostrais das características desta. O processo considera a utilização de simulação iterativa, baseada em cadeias de Markov, de maneira que os valores gerados serão dependentes.

Considere que se deseje amostrar valores de uma distribuição $\pi(y), y \in \mathbb{R}^d$, a qual é conhecida a menos de uma constante multiplicativa. Considere as ocorrências $Y^t, t = 0, 1, \dots$ de uma cadeia de Markov que tenha π como distribuição de equilíbrio. Sabe-se que apesar de a cadeia de Markov ser dependente, pela definição de Processos Estocásticos, a média aritmética dos valores gerados dessa cadeia não o é, de maneira que esta é um estimador consistente da média teórica, isto é,

$$Y^{(t)} \rightarrow \pi(y) \quad \text{quando } t \rightarrow \infty \quad \text{e} \quad \frac{1}{n} \sum_{t=1}^n g(Y_i^t) \rightarrow E_{\pi}(g(Y_i)) \quad \text{quando } n \rightarrow \infty.$$

Desta forma, para um número finito de interações,

$$\hat{g} = \frac{1}{n} \sum_{t=1}^n g(Y_i^{(t)})$$

é uma estimativa de $E(g(Y_i))$. Dados os valores gerados, pode-se ainda considerar,

$$Var(\hat{g}) = \frac{s^2}{n} \left[1 + 2 \sum_{k=1}^n \left(1 - \frac{k}{n} \rho_k \right) \right]$$

em que s^2 é a variância amostral e ρ_k é a autocorrelação amostral de ordem k .

Dentre os métodos de Monte Carlo, via Cadeias de Markov, os dois mais conhecidos são o algoritmo de Metropolis-Hastings e o amostrador de Gibbs.

2.1.7.3.1 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings é um método de Monte Carlo com Cadeias de Markov que baseia-se em gerar um determinado valor de uma distribuição auxiliar o qual

deverá ser aceito (ou não) baseado em uma determinada probabilidade. Tal mecanismo de aceitação e rejeição culmina na convergência da cadeia para a distribuição de equilíbrio, isto é, a distribuição de interesse da qual deseja-se obter resumos a posteriori de suas características estatísticas.

Suponha uma cadeia de Markov e que o mecanismo gerador desta esteja em um estado θ e que um valor θ' seja gerado de uma distribuição, chamada distribuição proposta, $q(\cdot|\theta)$. O novo valor de θ , θ' , será aceito (ou não) com a seguinte probabilidade,

$$p = \min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right) \quad (23)$$

sendo π é a distribuição de interesse.

Pode-se especificar o Algoritmo de Metropolis-Hastings nos seguintes passos:

1. Iniciar o contador de iterações $t = 0$;
2. Especificar um valor inicial $\theta^{(0)}$, que eventualmente pode ser um vetor;
3. Gerar um novo valor θ' da distribuição $q(\cdot|\theta)$;
4. Determinar a probabilidade de aceitação do novo valor θ' , p ;
5. Gerar o valor de v , em que $v \sim U(0, 1)$;
6. Se $v \leq p$, aceita-se o novo valor e considera-se $\theta^{(t+1)} = \theta'$. Caso contrário, rejeita-se o novo valor e considera-se $\theta^{(t+1)} = \theta$;
7. Incrementar o contador de t para o estado seguinte e voltar ao segundo passo.

2.1.7.3.2 Amostrador de Gibbs

No amostrador de Gibbs não há mecanismo de aceitação-rejeição, de maneira que a transição de estado para outro da cadeia sempre irá ocorrer e é feita de acordo com as distribuições condicionais completas $\pi(\theta_i|\theta_{-i})$, em que $\theta_{-i}^T = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$. Observe então que, a distribuição condicional completa é a distribuição do i -ésimo componente de θ condicionada em todos os outros componentes, e ela é obtida a partir da distribuição conjunta como,

$$\pi(\theta_i|\theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}.$$

Considerando as distribuições condicionais completas, o Amostrador de Gibbs pode ser definido pelo seguinte esquema,

1. Iniciar o contador de iterações $t = 1$;
2. Especificar valores iniciais para o vetor $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$;
3. Obter um novo valor $\theta^{(t)}$ a partir de $\theta^{(t-1)}$ considerando ageração sucessivas de valores,

$$\begin{aligned}\theta_1^{(t)} &\sim \pi(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}) \\ \theta_2^{(t)} &\sim \pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}) \\ &\vdots \\ \theta_d^{(t)} &\sim \pi(\theta_d | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{d-1}^{(t)})\end{aligned}$$

4. Incrementar o contador de t para o passo seguinte e retornar ao passo 2 até obter a obtenção de convergência.

Mais detalhes a respeito desses dois métodos de Monte Carlo, inclusive recursos computacionais para sua implementação podem ser vistas em Robert e Casella (2010).

2.1.7.4 Método INLA

A utilização de modelos lineares generalizados em análises estatísticas é cada vez mais comum, devido à possibilidade de permitir descrever o comportamento de dados provenientes de diferentes distribuições de probabilidades, desde que estas pertençam a família exponencial. Com o uso de modelos lineares generalizados mistos o cenário é ampliado, visto que, além de ser possível descrever o comportamento de dados provenientes de diferentes distribuições, é possível também acomodar a variabilidade residual que não está sendo descrita pelo modelo de efeitos fixos. A inclusão de efeitos aleatórios permite, neste trabalho, acomodar superdispersão.

Segundo Fong, Rue e Wakefield (2009), no caso de pequenas amostras, a estimação dos parâmetros como é usualmente feita, na abordagem clássica, pode apresentar problemas, especialmente na estimação dos componentes de variância, pois a inferência baseia-se na distribuição assintótica dos estimadores de máxima verossimilhança. Para solucionar esses problemas, é comum propor a utilização da abordagem bayesiana, a qual

é baseada nos dados observados e permite a atualização de informação. Todavia, essa abordagem tem sido prejudicada pela necessidade de programação computacional intensiva.

Dada a necessidade de ajuste de modelos hierárquicos, além dos modelos lineares generalizados mistos, de maneira eficiente e rápida, Rue, Martino e Chopin (2009) propuseram a estimação INLA, do inglês, “*Integrated Nested Laplace Approximation*”. O método INLA é uma aproximação determinística para a inferência bayesiana para a classe dos modelos gaussianos latentes. Nesses modelos, como descrito por Rue, Martino e Chopin (2009), o campo latente é gaussiano, controlado por hiperparâmetros e com variáveis respostas não gaussianas, de modo que as distribuições marginais a posteriori não apresentam forma fechada. Para compreensão desse método, é necessário primeiramente definir o que são modelos gaussianos latentes.

2.1.7.4.1 Modelos gaussianos latentes

Os modelos gaussianos latentes são uma classe dos modelos de regressão de estrutura aditiva, bastante flexíveis. Nesses modelos, considera-se que a distribuição da variável resposta Y_i pertence a uma determinada família de distribuições (não necessariamente à exponencial, porém a inclui), em que a média μ_i está ligada a uma estrutura aditiva do preditor linear η_i por meio de uma função de ligação $g(\cdot)$, tal que $g(\mu_i) = \eta_i$. Rue, Martino e Chopin (2009) mostram que a estrutura aditiva no preditor linear η_i leva em conta o efeito de várias covariáveis da seguinte forma,

$$\eta_i = \alpha + \sum_{j=1}^{n_h} h^{(j)}(\mathbf{u}_{ji}) + \sum_{k=1}^{n_\beta} \beta_k \mathbf{z}_{ki} + \epsilon_i \quad (24)$$

em que $\{h^{(j)}(\cdot)\}$ são funções não conhecidas de covariáveis $\mathbf{u}$, $\{\beta_k\}$ representam o efeito linear de covariáveis $\mathbf{z}$ e os ϵ_i são termos não estruturados, utilizados para acomodar a correlação espacial, temporal e/ou espaço-temporal, caso exista. Essa classe de modelos é muito flexível devido às diferentes formas que as funções desconhecidas $\{h^{(j)}(\cdot)\}$ podem assumir.

Definem-se os modelos gaussianos latentes, segundo Rue, Martino e Chopin (2009) e Rasmussen e Williams (2006), como sendo modelos que assumem a forma aditiva em (24), considerando-se distribuições a priori gaussianas para α , $\{h^{(j)}(\cdot)\}$, $\{\beta_k\}$ e $\{\epsilon_i\}$. Segundo Rasmussen e Williams (2006), na literatura esses modelos também são tratados

como “*gaussian process models*”. Essa classe de modelos tem ampla aplicação, todavia neste trabalho será considerada apenas aplicações em modelos lineares generalizados mistos.

Segundo Dey et al. (2000), o preditor linear dos modelos lineares generalizados pode ser escrito por,

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k \mathbf{z}_{ki}$$

que é um caso particular da expressão (24).

Os termos $h^{(j)}(\cdot)$ presentes na expressão (24) são utilizados para introduzir efeitos aleatórios nesses modelos, definindo assim os modelos lineares generalizados mistos (FAHRMEIR e LANG, 2001). Estes efeitos aleatórios tornam possível acomodar a dispersão causada por heterogeneidade não observada, ou por correlação em dados longitudinais. O preditor linear, sob esta parametrização é dado por,

$$\eta_i = \alpha + \sum_{j=1}^{n_h} h^{(j)}(\mathbf{u}_{ji}) + \sum_{k=1}^{n_\beta} \beta_k \mathbf{z}_{ki}. \quad (25)$$

Denote $\pi(\cdot|\cdot)$, como sendo a distribuição condicional, e considere $\mathbf{x}$ como sendo o vetor de parâmetros de efeitos fixos a serem estimados e efeitos aleatórios a serem preditos. A distribuição $\pi(\mathbf{x}|\boldsymbol{\theta})$ é assumida ser normal com vetor de médias zero e matriz de precisão $\mathbf{Q}(\boldsymbol{\theta})$, com componentes de variância $\boldsymbol{\theta}$ a serem estimados.

Seja $\mathbf{y}$ o vetor de observações y_i com $\{y_i : i \in \mathcal{I}\}$. Sua distribuição é denotada por $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ e assume-se que $\{y_i : i \in \mathcal{I}\}$ são condicionalmente independentes dados $\mathbf{x}$ e $\boldsymbol{\theta}$. A distribuição conjunta a posteriori para os parâmetros do modelo linear generalizado misto, pode ser expressa por,

$$\begin{aligned} \pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp \left[-\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_{i \in \mathcal{I}} \ln \{ \pi(y_i|x_i, \boldsymbol{\theta}) \} \right]. \end{aligned} \quad (26)$$

Considerando-se a expressão (26), para o processo de estimação bayesiano utilizando o método INLA, as seguintes propriedades são assumidas:

- i) O campo latente $\mathbf{x}$ satisfaz a propriedade de independência condicional, e será consi-

derado um GMRF, com matriz de precisão $\mathbf{Q}(\boldsymbol{\theta})$.

- ii) O número de hiperparâmetros deve ser baixo.

Essas propriedades afetam, principalmente, a performance do algoritmo, reduzindo o tempo computacional, pois permitem o uso de métodos numéricos para matrizes esparsas.

2.1.7.4.2 Campos aleatórios gaussianos markovianos

Prates (2011) define um campo aleatório gaussiano (GRF) como sendo um processo estocástico de alguma distribuição marginal de dimensão finita que é distribuído segundo uma normal multivariada. Uma vez que este GRF satisfaça a propriedade de Markov, ele é considerado um campo aleatório gaussiano markoviano (GMRF). O autor ainda aponta que o GMRF também é chamado na literatura de *Gaussian Bayesian network*.

Rue e Held (2005) definem GMRF como sendo um vetor aleatório finito-dimensional que siga uma distribuição normal multivariada e que satisfaz a propriedade de independência condicional (propriedade de Markov). Para ilustrar a independência condicional, considere que $\mathbf{x} = (x_1, x_2, x_3)^T$ seja um vetor aleatório, então, diz-se que x_1 e x_2 são condicionalmente independentes dado x_3 ($x_1 \perp x_2 | x_3$) se, para um valor conhecido de x_3 , a observação do valor de x_2 não acrescenta nenhuma informação sobre a distribuição de x_1 . Sob estas condições, de acordo com Rue e Held (2005), a distribuição conjunta $\pi(\mathbf{x})$ pode ser expressa por:

$$\begin{aligned}\pi(\mathbf{x}) &= \pi(x_1 | x_2, x_3) \pi(x_2 | x_3) \pi(x_3) \\ &= \pi(x_1 | x_3) \pi(x_2 | x_3) \pi(x_3).\end{aligned}$$

Na abordagem INLA, considere que o vetor $\mathbf{x} = (x_1, \dots, x_g)^T$, com $g = 1, \dots, G$, é formado pelos parâmetros de efeitos fixos e pelos efeitos aleatórios do modelo, sendo G o total de parâmetros de efeitos fixos e efeitos aleatórios a serem estimados e preditos, respectivamente. Da propriedade de Markov tem-se que x_g e $x_{g'}$ são condicionalmente independentes em $\mathbf{x}_{-gg'}$, isto é, no vetor $\mathbf{x}$ sem os parâmetros x_g e $x_{g'}$. Sob esta propriedade, para a matriz de precisão deste GMRF, tem-se que $\mathbf{Q} : Q_{gg'} = 0$ se e somente se x_g e $x_{g'}$ são condicionalmente independentes em $\mathbf{x}_{-gg'}$. Consequentemente, $\mathbf{Q} : Q_{gg'} \neq 0$ sempre que houver alguma relação entre x_g e $x_{g'}$ em $\mathbf{x}_{-gg'}$.

Considerando que a média de $\mathbf{x}$ é $\boldsymbol{\mu}$, a distribuição de $\mathbf{x}$ pode ser expressa por,

$$\pi(\mathbf{x}) = (2\pi)^{-g/2} |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

Outro resultado importante, associado aos GMRF, é a aproximação GMRF, a qual é utilizada no processo de estimação INLA. Rue e Held (2005) exemplificam esta aproximação com o seguinte exemplo univariado: suponha uma única observação y de uma distribuição de Poisson com média λ , e suponha também que $\eta = \ln \lambda \sim N(\mu, \kappa^{-1})$, com μ e κ conhecidos. Assim:

$$\begin{aligned} \pi(\eta) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{(\kappa^{-1})}} e^{-\frac{1}{2} \frac{1}{(\kappa^{-1})} (\eta - \mu)^2} \\ &\propto e^{-\frac{1}{2} \frac{1}{(\kappa^{-1})} (\eta - \mu)^2} \\ &= e^{-\frac{1}{2} \kappa (\eta - \mu)^2}, \end{aligned}$$

em que κ é o parâmetro de precisão e $\kappa^{-1} = 1/\kappa$ é a variância. A verossimilhança, $L(\eta|y)$ pode ser expressa por,

$$L(\eta|y) = \frac{e^{-\lambda} \lambda^y}{y!} \propto e^{-\lambda} \lambda^y = e^{-\lambda} e^{y \ln(\lambda)} = e^{-\lambda + y \ln(\lambda)} = e^{-e^\eta + y\eta}.$$

A distribuição a posteriori fica expressa por,

$$\begin{aligned} \pi(\eta|y) &\propto L(\eta|y) \pi(\eta) \\ &= e^{-\frac{1}{2} \kappa (\eta - \mu)^2 - e^\eta + y\eta} = e^{f(\eta)}. \end{aligned}$$

Expandindo a função $f(\eta)$ em série de Taylor até a 2ª ordem, em torno do valor η_0 , tem-se:

$$\begin{aligned} f(\eta) &\approx f(\eta_0) + f'(\eta_0)(\eta - \eta_0) + \frac{1}{2} f''(\eta_0)(\eta - \eta_0)^2 \\ &= f(\eta_0) + f'(\eta_0)\eta - f'(\eta_0)\eta_0 - \frac{1}{2} \left(-f''(\eta_0)(\eta - \eta_0)^2 \right) \end{aligned}$$

$$\begin{aligned}
&= f(\eta_0) + f'(\eta_0)\eta - f'(\eta_0)\eta_0 - \frac{1}{2} \left(-f''(\eta_0)(\eta^2 - 2\eta\eta_0 + \eta_0^2) \right) \\
&= f(\eta_0) + f'(\eta_0)\eta - f'(\eta_0)\eta_0 - \frac{1}{2} \left(-\eta^2 f''(\eta_0) + 2\eta\eta_0 f''(\eta_0) - \eta_0^2 f''(\eta_0) \right) \\
&= f(\eta_0) + f'(\eta_0)\eta - f'(\eta_0)\eta_0 - \frac{1}{2} \left(-\eta^2 f''(\eta_0) \right) - \eta\eta_0 f''(\eta_0) + \frac{1}{2}\eta_0^2 f''(\eta_0) \\
&= f(\eta_0) + \frac{1}{2}\eta_0^2 f''(\eta_0) - f'(\eta_0)\eta_0 - \frac{1}{2} \left(-\eta^2 f''(\eta_0) \right) + f'(\eta_0)\eta - \eta\eta_0 f''(\eta_0) \\
&= \left(f(\eta_0) + \frac{1}{2}\eta_0^2 f''(\eta_0) - f'(\eta_0)\eta_0 \right) - \frac{1}{2} \left(-\eta^2 f''(\eta_0) \right) + \left(f'(\eta_0) - \eta_0 f''(\eta_0) \right) \eta \\
&= \left(f(\eta_0) + \frac{1}{2}\eta_0^2 f''(\eta_0) - f'(\eta_0)\eta_0 \right) + \left(f'(\eta_0) - \eta_0 f''(\eta_0) \right) \eta - \frac{1}{2} \left(-f''(\eta_0) \right) \eta^2 \\
&= a + b\eta - \frac{1}{2}c\eta^2,
\end{aligned}$$

em que $a = f(\eta_0) + \frac{1}{2}\eta_0^2 f''(\eta_0) - f'(\eta_0)\eta_0$ é constante, $b = f'(\eta_0) - \eta_0 f''(\eta_0)$ e $c = -f''(\eta_0)$. Desta forma, $\pi(\eta|y)$ pode ser aproximada por $\tilde{\pi}(\eta|y)$, em que,

$$\pi(\eta|y) \propto e^{-\frac{1}{2}c\eta^2 + b\eta} = \tilde{\pi}(\eta|y). \quad (27)$$

Observe que se uma variável aleatória $Y \sim N(\mu, \sigma^2)$ então,

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y - \mu)^2}$$

ou seja,

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}y^2 + \frac{1}{\sigma^2}\mu y - \frac{1}{2\sigma^2}\mu^2}. \quad (28)$$

Comparando-se (27) e (28) verifica-se que $\tilde{\pi}(\eta|y)$ é o núcleo da função densidade de probabilidade de uma variável aleatória Y com distribuição normal de média $\mu(\eta_0) = b/c$ e precisão $\kappa(\eta_0) = c$.

Observe que a média e a variância de $\tilde{\pi}(\eta|y)$ são expressas em função do valor η_0 . Isto implica que a qualidade da aproximação depende da escolha do valor de η_0 , de forma que a aproximação será tão melhor quanto mais próximo o valor de η_0 estiver do valor que maximiza $\pi(\eta|y)$, ou seja, quanto mais próximo η_0 estiver da moda de $\pi(\eta|y)$.

Exemplificando, na Figura 1, conforme Rue e Held (2005), são apresentados

os gráficos para $\pi(\eta|y)$ e $\tilde{\pi}(\eta|y)$, considerando $y = 3$, $\mu = 0$, $\kappa = 0,001$ e $\eta_0 = 0, 0,5, 1$ e $1,5$. Fica claro que a aproximação ($\tilde{\pi}(\eta|y)$) é melhor quanto mais próximo o valor de η_0 for da moda de $\pi(\eta|y)$.

Figura 1 - Gráfico da aproximação normal ($\tilde{\pi}(\eta|y)$) e da posteriori ($\pi(\eta|y)$) para $y = 3$, $\mu = 0$, $\kappa = 0,001$ e $\eta_0 = 0, 0,5, 1$ e $1,5$, indicados como um ponto preto

De acordo com Rue e Held (2005), esta idéia pode ser generalizada para o caso multivariado. Suponha que $y_1, \dots, y_n$ sejam n observações condicionalmente independentes de uma distribuição não normal e que o mecanismo gerador dessas observações seja descrito por um modelo linear generalizado misto, por exemplo. Sob estas condições, $\mathbf{x}$ é um GMRF, composto pelos parâmetros de efeitos fixos e pelos efeitos aleatórios do modelo, com matriz de precisão $\mathbf{Q}$ e média $\boldsymbol{\mu}$, que possivelmente dependa de alguns hiperparâmetros. A distribuição condicional $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ é expressa por,

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) + \sum_{i=1}^n \ln \pi(y_i|x_i) \right\}. \quad (29)$$

De maneira similar ao caso univariado, utiliza-se a expansão de Taylor de segunda ordem de $\sum_{i=1}^n \ln \pi(y_i|x_i)$ em torno de $\boldsymbol{\mu}_0$, obtendo uma densidade proposta para o GMRF, $\tilde{\pi}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$.

Considerando-se a expressão (29), tem-se,

$$\begin{aligned}
\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) + \sum_{i=1}^n \ln \pi(y_i|x_i) \right\} \\
&= \left\{ -\frac{1}{2}(\mathbf{x}^T - \boldsymbol{\mu}^T) \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) + \sum_{i=1}^n \left(a_i + b_i x_i - \frac{1}{2} c_i x_i^2 \right) \right\} \\
&= \left\{ -\frac{1}{2} (\mathbf{x}^T \mathbf{Q} - \boldsymbol{\mu}^T \mathbf{Q}) (\mathbf{x} - \boldsymbol{\mu}) + \mathbf{b}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \text{diag}(\mathbf{c}) \mathbf{x} \right\} \\
&= \left\{ -\frac{1}{2} (\mathbf{x}^T \mathbf{Q} \mathbf{x} - \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} - \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} + \boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\mu}) + \mathbf{b}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \text{diag}(\mathbf{c}) \mathbf{x} \right\} \\
&\propto \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} + \mathbf{b}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \text{diag}(\mathbf{c}) \mathbf{x} \right\} \\
&= \left\{ -\frac{1}{2} (\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{x}^T \text{diag}(\mathbf{c}) \mathbf{x}) + \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} + \mathbf{b}^T \mathbf{x} \right\} \\
&= \left\{ -\frac{1}{2} \mathbf{x}^T (\mathbf{Q} + \text{diag}(\mathbf{c})) \mathbf{x} + (\mathbf{Q} \boldsymbol{\mu} + \mathbf{b})^T \mathbf{x} \right\}.
\end{aligned}$$

Usando-se essa expansão, a equação (29) pode ser expressa por,

$$\begin{aligned}
\tilde{\pi}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} + \sum_{i \in \mathcal{I}} \left(a_i + b_i x_i - \frac{1}{2} c_i x_i^2 \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\mathbf{Q} + \text{diag}(\mathbf{c})) \mathbf{x} + (\mathbf{Q} \boldsymbol{\mu} + \mathbf{b})^T \mathbf{x} \right\} \tag{30}
\end{aligned}$$

em que c_i é definido como zero, caso seu valor seja negativo. A parametrização (30) é o núcleo de uma $N(\mathbf{Q} \boldsymbol{\mu} + \mathbf{b}, \mathbf{Q} + \text{diag}(\mathbf{c}))$. Assim como no caso univariado, essa aproximação depende de $\boldsymbol{\mu}_0$, de forma que a sua acurácia depende do quão próximo este valor estiver do valor que maximiza $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. A determinação desse valor pode ser feita utilizando o método de Newton-Raphson multivariado. De acordo com Rue e Held (2005), após m iterações do método de Newton-Raphson multivariado, $\boldsymbol{\mu}_m$ é igual ao valor que maximiza $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, e denomina-se a aproximação $\tilde{\pi}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ como aproximação GMRF (*GMRF approximation*).

2.1.7.4.3 Estimação INLA

O método de estimação INLA foi proposto por Rue, Martino e Chopin (2009) para a classe de modelos gaussianos latentes, em que a variável observada ou resposta Y_i , com média μ_i , é ligada à estrutura aditiva do preditor linear η_i por meio de uma função de ligação $g(\cdot)$, tal que $g(\mu_i) = \eta_i$. A estrutura aditiva do preditor linear η_i leva em consideração o efeito de várias covariáveis e pode ser escrita como em (24), e para o caso do modelo linear generalizado misto, dada por (25). Segundo Martins et al. (2013), o modelo descrito em (25) pode ser escrito utilizando uma estrutura hierárquica, em que o primeiro estágio é formado pela função de verossimilhança com propriedades de independência condicional dada pelo GMRF $\mathbf{x}$ e possíveis hiperparâmetros $\boldsymbol{\theta}$ (associados aos componentes de variância). Assim, para o primeiro estágio da hierarquia tem-se:

$$\textbf{Estágio 1. } \mathbf{y}|\mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = L(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) = \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta}).$$

No segundo estágio da hierarquia, é definida a distribuição condicional do GMRF $\mathbf{x}$ para todos (ou alguns) valores de $\boldsymbol{\theta}$,

$$\textbf{Estágio 2. } \mathbf{x}|\boldsymbol{\theta}, \mathbf{y} \sim \pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$$

a qual é expressa por (30).

O terceiro e último passo do modelo hierárquico é dado pela especificação de uma distribuição a priori apropriada para cada um dos hiperparâmetros do modelo,

$$\textbf{Estágio 3. } \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}).$$

Conforme Rue, Martino e Chopin (2009), a distribuição conjunta a posteriori pode ser expressa por (26) e as distribuições marginais a posteriori de interesse podem ser expressas por,

$$\begin{aligned} \pi(x_g|\mathbf{y}) &= \int \pi(x_g|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\ \pi(\theta_\nu|\mathbf{y}) &= \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-\nu}. \end{aligned} \tag{31}$$

em que $g = 1, \dots, G$, sendo G o número de parâmetros de efeitos fixos e efeitos aleatórios do modelo, e $\nu = 1, \dots, \mathcal{V}$, sendo $\mathcal{V}$ o número de componentes de variância.

Rue, Martino e Chopin (2009) utilizaram as expressões em (31) para construir as aproximações aninhadas (*nested*),

$$\begin{aligned}\tilde{\pi}(x_g|\mathbf{y}) &= \int \tilde{\pi}(x_g|\boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ \tilde{\pi}(\theta_\nu|\mathbf{y}) &= \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-\nu}.\end{aligned}\tag{32}$$

Sob a parametrização (32), $\pi(\cdot|\cdot)$ são densidades condicionais aproximadas. As aproximações para $\tilde{\pi}(x_g|\mathbf{y})$ são obtidas pela aproximação de $\pi(\boldsymbol{\theta}|\mathbf{y})$, $\pi(x_g|\boldsymbol{\theta}, \mathbf{y})$ e por meio de integral numérica em $\boldsymbol{\theta}$.

Rue, Martino e Chopin (2009) dividem, então, a aproximação INLA em três passos. No primeiro passo aproxima-se a distribuição marginal a posteriori de $\boldsymbol{\theta}$ por meio da aproximação de Laplace. O segundo passo consiste na obtenção da aproximação de Laplace ou da aproximação de Laplace simplificada de $\pi(x_g|\mathbf{y}, \boldsymbol{\theta})$ para valores de $\boldsymbol{\theta}$. O terceiro passo combina os dois anteriores pelo uso de integração numérica.

O passo 1 da aproximação INLA consiste em explorar a distribuição marginal a posteriori $\pi(\boldsymbol{\theta}|\mathbf{y})$. A aproximação utilizada para a obtenção da distribuição marginal a posteriori dos hiperparâmetros, $\pi(\boldsymbol{\theta}|\mathbf{y})$, segundo Martino e Rue (2007) é expressa por,

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left. \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x} = \mathbf{x}^*(\boldsymbol{\theta})}\tag{33}$$

em que $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ é a aproximação gaussiana para a distribuição condicional do GMRF, definida por Rue e Held (2005), como aproximação GMRF (30). $\mathbf{x}^*(\boldsymbol{\theta})$ é o valor que maximiza a densidade condicional de $\mathbf{x}$, para um dado $\boldsymbol{\theta}$; em que o máximo da expressão (30) é obtido iterativamente pela solução de $\{\mathbf{Q} + \text{diag}(\mathbf{c})\}\boldsymbol{\mu}^{(1)} = \mathbf{b}$. Essa aproximação, segundo Martino e Rue (2007) e Rue, Martino e Chopin (2009), é equivalente à aproximação de Laplace de Tierney e Kadane (1986) para uma distribuição marginal a posteriori.

Obtida a aproximação $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, supondo $\boldsymbol{\theta} = (\theta_1, \dots, \theta_\nu) \in \mathbb{R}^\nu$, o qual pode ser obtido por reparametrizações, Rue, Martino e Chopin (2009) indicam os passos seguintes para explorar $\pi(\boldsymbol{\theta}|\mathbf{y})$.

- i) Determinar o valor que maximiza $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, pela otimização de $\ln \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ em relação a

$\boldsymbol{\theta}$, utilizando por exemplo, o método quase-Newton.

- ii) Considerar $\boldsymbol{\theta}^*$ o valor que maximiza $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, determinar a matriz hessiana $\mathbf{H}$. Sendo $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$, considerar a decomposição $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ e defina $\boldsymbol{\theta}$ como função de $\mathbf{z}$,

$$\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}.$$

- iii) Explorar $\ln \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ pelo uso da parametrização $\mathbf{z}$.
- iv) Aproximar $\pi(\boldsymbol{\theta}_\nu|\mathbf{y})$ diretamente de $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ por integração numérica.

No segundo passo, para aproximar a distribuição marginal a posteriori de cada um dos elementos de $\mathbf{x}$, fixado valores de $\boldsymbol{\theta}$, três métodos podem ser considerados: aproximação gaussiana (que no contexto dos GMRF, é chamada de aproximação GMRF), aproximação de Laplace e aproximação simplificada de Laplace.

A aproximação de Laplace para obtenção de $\pi(x_g|\boldsymbol{\theta}, \mathbf{y})$ é obtida por meio da seguinte expressão,

$$\tilde{\pi}_{LA}(x_g|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-g}|x_g, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-g}=\mathbf{x}_{-g}^*(x_g, \boldsymbol{\theta})} \quad (34)$$

em que $\tilde{\pi}_{GG}$ é a aproximação gaussiana para $\mathbf{x}_{-g}|x_g, \boldsymbol{\theta}, \mathbf{y}$, sendo $\mathbf{x}_{-g}^*(x_g, \boldsymbol{\theta})$ o valor que maximiza a expressão.

Observe que a aproximação (34) implica no cálculo de $\tilde{\pi}_{GG}$ para cada um dos x_g e $\boldsymbol{\theta}$, o que torna o custo computacional alto. Para contornar esse problema, foram propostas duas modificações,

- i) Para otimizar o cálculo de $\tilde{\pi}_{GG}(\mathbf{x}_{-g}|x_g, \boldsymbol{\theta}, \mathbf{y})$, considerar,

$$\mathbf{x}_{-g}^*(x_g, \boldsymbol{\theta}) \approx E_{\tilde{\pi}_G}(\mathbf{x}_{-g}|x_g). \quad (35)$$

- ii) Considerando que apenas os elementos $x_{g'}$ próximos de x_g tem um efeito marginal em x_g , e que a dependência condicional entre x_g e $x_{g'}$ decai com o crescimento da distância entre os nós g e g' , apenas uma região de interesse em torno de g , $R_g(\boldsymbol{\theta})$, determina a distribuição marginal a posteriori de x_g . Assim, a esperança condicional

na aproximação (35) implica que,

$$\frac{E_{\tilde{\pi}_G}(\mathbf{x}_{g'}|x_g) - \mu_{g'}(\boldsymbol{\theta})}{\sigma_{g'}(\boldsymbol{\theta})} = a_{gg'}(\boldsymbol{\theta}) \frac{x_g - \mu_g(\boldsymbol{\theta})}{\sigma_g(\boldsymbol{\theta})}$$

para alguns valores de $a_{gg'}(\boldsymbol{\theta})$ em que $g \neq g'$. Uma simples regra para construção da região $R_{g'}(\boldsymbol{\theta})$ é

$$R_{g'}(\boldsymbol{\theta}) = \{g' : |a_{gg'}(\boldsymbol{\theta})| > 0,0001\}.$$

A aproximação de Laplace simplificada para $\pi(x_g|\boldsymbol{\theta}, \mathbf{y})$, denotada por $\tilde{\pi}_{SLA}(x_g|\boldsymbol{\theta}, \mathbf{y})$, é obtida por meio de uma série de expansões de $\tilde{\pi}_{LA}(x_g|\boldsymbol{\theta}, \mathbf{y})$ em torno de $x_g = \mu_g(\boldsymbol{\theta})$. Essa expansão corrige a aproximação $\tilde{\pi}_G(x_g|\boldsymbol{\theta}, \mathbf{y})$ para locação e assimetria. De acordo com Rue, Martino e Chopin (2009), para os casos em que o comportamento distribucional da variável observada seja descrito por meio de uma distribuição de probabilidade binomial ou Poisson, estas correções da aproximação de Laplace simplificada são suficientes para obtenção de distribuições marginais a posteriori com acurácia para x_g . O benefício do uso dessa aproximação em relação à aproximação de Laplace resulta em menor custo computacional.

2.1.7.5 Critérios para a comparação de modelos

Após o ajuste de modelos é necessário utilizar critérios que permitam a seleção de um modelo que atenda as necessidades do pesquisador, isto é, um modelo que descreva bem o comportamento dos dados e/ou faça boas previsões.

No contexto bayesiano várias medidas de seleção já foram propostas, sendo elas baseadas em medidas de ajustes de diferentes modelos. A seguir, serão expostos os critérios correntes na seleção de modelos ajustados, por meio de inferência bayesiana, baseada na aproximação INLA.

2.1.7.5.1 Critério da informação da deviance

O critério da informação da deviance (DIC) para a seleção de modelos foi proposto por Spiegelhalter et al. (2000) como uma generalização do critério da informação de Akaike (AKAIKE, 1973). Os autores indicam também o uso do critério para a seleção

de variáveis explicativas a serem consideradas no modelo.

O critério leva em consideração a adequabilidade do modelo e a sua complexidade, baseando-se na esperança a posteriori do logaritmo da distribuição condicional dos dados. Sendo a adequabilidade do modelo representada por $\bar{D}$ e sua complexidade por p_D , o cálculo de DIC é baseado no valor esperado da distribuição a posteriori de

$$D(\boldsymbol{\theta}) = -2 \log[p(\mathbf{y}|\boldsymbol{\theta})] + 2 \log[f(\mathbf{y})], \quad (36)$$

isto é, no valor da média a posteriori do desvio, assim como na estatística p_D , que representa a complexidade do modelo medida em função do número efetivo de parâmetros do modelo, dada pela diferença entre a média a posteriori da deviance e a deviance das médias a posteriori,

$$\begin{aligned} p_D &= E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(E_{\boldsymbol{\theta}|\mathbf{y}}[p(\mathbf{y}|\boldsymbol{\theta})]) \\ &= \overline{D(\boldsymbol{\theta})} - \bar{D}(\boldsymbol{\theta}). \end{aligned} \quad (37)$$

Combinando (36) e (37), como segue, tem-se a medida de DIC,

$$DIC = \bar{D}(\boldsymbol{\theta}) + p_D. \quad (38)$$

Uma vez que (38) representa uma medida de complexidade do modelo, os modelos que apresentarem menores valores de DIC podem ser considerados mais adequados por apresentarem um melhor ajuste ponderado pela complexidade.

O DIC, no contexto bayesiano, é um dos critérios de seleção mais utilizados, todavia, segundo Plummer (2008), ele pode penalizar modelos complexos, sobretudo, com muitos efeitos aleatórios. Para contornar esse problema, sugere-se o uso de medidas preditivas como critérios de seleção de modelo. Gelman et al. (2004), Gelfand (1996), Carlin e Louis (1996) e Rao (2003) sugerem o uso de medidas preditivas tanto para seleção de modelos como para validação. Pettit e Young (1990) ainda sugerem o uso de medidas preditivas para a detecção de valores *outliers* e discrepantes. Os autores sugerem o uso de ordenadas preditivas condicionais (CPO) e das transformações uniformizantes (PIT) que serão discutidas posteriormente.

2.1.7.5.2 Critério da divergência simétrica de Kullback-Leibler

Segundo Kullback e Leibler (1951), com base na teoria de Probabilidade e teoria da informação, o critério da divergência simétrica de Kullback-Leibler (KLD), ou simplesmente, divergência de Kullback-Leibler, é uma medida de dissimilaridade entre duas distribuições de probabilidade. Se p_1 e p_0 representam duas densidades de probabilidade, a KLD pode ser definida como,

$$\text{KLD} = \int p_1(x) \log \left(\frac{p_1(x)}{p_0(x)} \right) dx.$$

Para o caso em que se tenham distribuições discretas, tem-se,

$$\text{KLD} = \sum_x p_1(x) \log \left(\frac{p_1(x)}{p_0(x)} \right).$$

Utilizando a desigualdade de Jensen, Kullback e Leibler (1951) mostram que $\text{KLD} \geq 0$, para qualquer par de funções de probabilidade, com igualdade se e só se $p_1(x) = p_0(x)$, $\forall x$.

No contexto bayesiano, considerando a aproximação INLA, Rue, Martino e Chopin (2009) sugerem o uso do KLD para mensurar a qualidade de aproximação, comparando a aproximação obtida via INLA com a aproximação gaussiana. Se o critério apresenta um baixo valor, isto é, próximo de zero, conclui-se por uma boa aproximação.

2.1.7.5.3 Transformação integral de probabilidade

Rosenblatt (1952) propôs uma medida de qualidade de ajuste baseado na transformação integral de probabilidade (do inglês, *probability integral transform*), traduzido para o português por Ribeiro (2012) como transformações uniformizantes e de agora em diante, citado como PIT.

Suponha $\mathbf{X} = (X_1, X_2, \dots, X_k)$ um vetor com funções de distribuições absolutamente contínuas $F_{\mathbf{X}}(x_1, x_2, \dots, x_k)$. Seja ainda $F_{X_i} = P(X_i \leq x_i)$ a função de distribuição marginal univariada de X_i , com $i = 1, 2, \dots, k$. Considere k transformações do

tipo,

$$\begin{aligned}
T(x_1) &= P(X_1 \leq x_1) = F_{X_1}(x_1), \\
T(x_2) &= P(X_2 \leq x_2 | X_1 = x_1) = F_{X_2|X_1}(x_2|x_1), \\
&\dots \\
T(x_k) &= P(X_k \leq x_k | X_1 = x_1, \dots, X_{k-1} = x_{k-1}) \\
&= F_{X_k|X_1, \dots, X_{k-1}}(x_k|x_1, \dots, x_{k-1}).
\end{aligned}$$

Sendo $Z_i = T(X_i)$, para $i = 1, \dots, k$ variáveis aleatórias, diz-se que Z_i são uniformemente e independentemente distribuídas no intervalo $[0, 1]$. Rosenblatt (1952) mostra este resultado usando a estatística de Kolmogorv-Smirnov e a estatística de Von Misses.

Gelfand (1996) considera o uso das medidas de PIT para medir a capacidade preditiva de modelos sob o seguinte cenário: suponha que um modelo M atribua densidade preditiva um passo-a-frente $P(y_t | \mathbf{y}_{t-1}, M)$. Denote a sequência de funções de distribuição acumulada por,

$$F(c | \mathbf{y}_{t-1}, M) = P(y_t \leq c | \mathbf{y}_{t-1}, M).$$

Assim, a PIT correspondente ao modelo M e a sequência $\{y_t\}$ é $P(t; \mathbf{y}_{t-1}, M) = F(y_t | \mathbf{y}_{t-1}, M)$. Desta forma, se a suposição distribucional feita para os dados está correta, sob os dados observados $\{y_t\}$, então, $P(t; \mathbf{y}_{t-1}, M)$ é uniformemente distribuída $[0, 1]$. Sob este contexto, Rue, Martino e Chopin (2009) sugerem o uso das medidas de PIT para seleção do modelo, por meio da aproximação INLA, para o processo bayesiano de inferência. Nesse caso, o critério é considerado como uma medida preditiva de validação cruzada, em que, valores muito pequenos ou muito grandes para PIT indicam um possível *outlier*. Ainda, um histograma dos valores de PIT muito diferente do comportamento de um gráfico similar de uma variável uniformemente distribuída indica um modelo questionável, sob a distribuição considerada para descrever o comportamento da variável resposta. Então, sendo y_i a variável resposta observada, segundo Gneiting e Raftery (2007), os valores de PIT são dados por,

$$PIT_i = P(Y_i^{new} \leq y_i | \mathbf{y}_{-i}).$$

Conforme Ribeiro (2012), no caso de dados de contagem, a distribuição preditiva é discreta e os valores de PIT não seguem uma distribuição uniforme, sob a hipótese nula de que o modelo ajustado é o correto. Como correção, conforme Czado, Gneiting e Held (2009) e Ribeiro (2012), sugere-se o uso de uma versão não aleatória dos valores de PIT,

$$PIT_i = P(Y_i^{new} \leq y_i | \mathbf{y}_{-i}) - 0,5P(Y_i^{new} \leq y_i | \mathbf{y}_{-i}).$$

A interpretação é similar ao caso de dados contínuos.

2.1.7.5.4 Condicionais preditivas ordenadas

De acordo com Hall (2012), a distribuição preditiva a posteriori pode ser escrita como a replicação de $\mathbf{y}$ dado o modelo (usualmente representada por $\mathbf{y}^{rep}$), ou a predição de um novo e não observado $\mathbf{y}$ (usualmente representado por $\mathbf{y}^{new}$), dado o modelo. Dessa forma, a verossimilhança sobre dados replicados ou preditos pode ser expressa, respectivamente, por

$$p(\mathbf{y}^{rep} | \mathbf{y}) = \int p(\mathbf{y}^{rep} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}$$

e

$$p(\mathbf{y}^{new} | \mathbf{y}) = \int p(\mathbf{y}^{new} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}.$$

Como forma de verificar o ajuste, é usual comparar a distribuição preditiva de $\mathbf{y}^{rep}$ com os dados observados $\mathbf{y}$, o que é nomeado, no inglês, como *posterior predictive check*. Segundo Hall (2012), esse critério inclui a incerteza associada aos parâmetros, descrita pelas distribuições a priori, diferente da estatística frequentista, o que de acordo com Meng (1994), viola o princípio da verossimilhança, pois envolve um uso duplo dos dados.

De maneira a não violar o princípio da verossimilhança, Geisser e Eddy (1979) propuseram uma densidade preditiva para a validação cruzada, difundida por Gelfand (1996) e conhecida como condicionais preditivas ordenadas (CPO), do inglês, *conditional predictive*

ordinate. Sua forma pode ser expressa por

$$p(\mathbf{y}_i|\mathbf{y}_{-i}) = \int p(\mathbf{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{-i})d\boldsymbol{\theta},$$

em que $\mathbf{y}_i$ é cada uma das observações $\mathbf{y}$ e $\mathbf{y}_{-i}$ é $\mathbf{y}$ sem a observação i . Considerando o inverso da verossimilhança, a CPO para cada observação i após T interações é, então, dada por

$$CPO_i = \frac{1}{T^{-1} \sum_{t=1}^T p(\mathbf{y}_i|\boldsymbol{\theta}_t)^{-1}}. \quad (39)$$

Conforme (39), o valor de CPO_i é a inversa da média a posteriori da inversa da verossimilhança de y_i . Assim, CPO expressa a probabilidade a posteriori de se observar o valor (ou um conjunto de valores) de y_i quando o modelo é ajustado para todos os valores, exceto y_i . Segundo Gelfand (1996), um alto valor de CPO indica um melhor ajuste do modelo para y_i , e valores muito baixos de CPO sugerem que y_i é um *outlier* ou observação influente. Ntzoufras (2009), considerando o inverso do valor de CPO, ICPO, salienta que valores maiores do que 40 podem ser considerados como pontos *outliers* e maiores do que 70 como pontos influentes. Sob as condições anteriores, a CPO pode ser utilizada como critério para identificar valores outliers, valores influentes e seleção de modelos não aninhados (não encaixados).

Rue, Martino e Chopin (2009) sugerem a inspeção gráfica de CPO para a seleção de modelo, além da comparação dos valores da soma do logaritmo de CPO, que segundo Hall (2012), é um estimador do logaritmo da função de verossimilhança marginal, nomeado no contexto bayesiano como logaritmo da pseudo-verossimilhança marginal (LPsVM). Ao se considerar o quociente da pseudo-verossimilhança marginal (PVM) de dois modelos tem-se uma aproximação do fator de Bayes, o qual Hall (2012) nomeia como pseudo fator de Bayes (PsFB), sendo que, de maneira similar ao fator de Bayes pode-se observar o quociente dos LPsVM.

A interpretação do PsFB é similar à do fator de Bayes, introduzido por Jeffreys (1961), sendo uma alternativa ao teste de hipótese frequentista da razão de verossimilhanças para a seleção de modelos. No caso do PsFB, sob o teste frequentista, podem-se fazer testes para modelos não encaixados, além de esse teste ser baseado em um mecanismo de validação cruzada. Supondo então que M_1 e M_2 são dois modelos a serem comparados, baseados no

valor de $PsFB_{12}$, denotado por $PsFB_{12}$, podem-se selecionar modelos segundo o critério de Jeffreys (1961), vide Tabela 1.

Tabela 1 - Interpretação do $PsFB$, segundo Jeffreys (1961)

$\log(PsFB_{12})$	$PsFB_{12}$	Evidência a favor de M_1
0 - 0,5	1 - 3,2	Insignificante
0,5 - 1	3,2 - 10	Significativa
1 - 2	10 - 100	Forte
> 2	> 100	Decisiva

2.2 Metodologia

2.2.1 Ensaio clínico com pacientes portadores de epilepsia

Para aplicação da metodologia proposta neste trabalho, considerou-se um conjunto de dados de contagens com estrutura longitudinal e caracterizado pela presença de superdispersão por Molenberghs, Verbeke e Demétrio (2007) e por Rizzato (2011).

Os dados são provenientes de um ensaio clínico aleatorizado com dupla ocultação, conduzido para comparar um composto placebo com uma nova droga anti-epiléptica (AED), ambas utilizadas para o tratamento adjunto de pacientes que sofrem de ataques epiléticos, em combinação com uma ou duas outras AEDs em doses terapêuticas. O principal objetivo do estudo, inicialmente conduzido por Faught et al. (1996), é avaliar se o novo tratamento adjunto reduz, ou não, o número de ataques epiléticos sofridos na última semana de acompanhamento. Os dados foram analisados por Molenberghs, Verbeke e Demétrio (2007) considerando-se a abordagem clássica e por Rizzato (2011), por meio da abordagem bayesiana baseada em MCMC.

Os pacientes selecionados para comporem a amostra foram todos diagnosticados com epilepsia refratária de início parcial e foram aleatorizados aos dois tratamentos (placebo e nova AED) após 12 semanas, período que segundo Faught et al. (1996), foi considerado para a estabilização das AEDs. Depois desse período, um grupo composto por 45 pacientes passaram a receber o placebo, enquanto que outro grupo formado por 44 pacientes, iniciaram o tratamento com a nova AED. O período máximo de observação foi de 27 semanas, sendo que nem todos os indivíduos foram observados até o término do estudo, de maneira que a estrutura de dados longitudinal, neste caso, é incompleta.

2.2.2 Modelos

Para a análise de dados de contagens na ocorrência de superdispersão, que pode ser ocasionada pelo excesso de zeros, de acordo com Hall (2000) pode-se considerar o ajuste de modelos inflacionados de zeros. Sob os modelos inflacionados de zeros duas partes compõem a sua estrutura, a saber: (i) a contagem nula, podendo ser dividida em zeros estruturais e zeros amostrais; e (ii) a contagem não nula, em que a distribuição amostral dos dados é descrita segundo uma distribuição de Poisson ou binomial negativa.

Seja Y_{ij} uma variável aleatória, representando a j -ésima medida longitudinal do i -ésimo indivíduo, $i = 1, \dots, N$, $j = 1, \dots, n_i$ e $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ o vetor de todas as medidas do indivíduo i , $\mathbf{x}_{ij}$ um vetor de covariáveis associado a Y_{ij} .

A seguir, serão apresentados oito modelos que serão considerados neste trabalho. Os primeiros quatro modelos foram propostos por Molenberghs, Verbeke e Demétrio (2007) e ajustados aos dados que são considerados neste texto, por meio da abordagem clássica. Rizzato (2011) apresentou a abordagem bayesiana desses modelos baseada nos métodos estocásticos MCMC. Os modelos são apresentados na seção 2.1.6 e as distribuições a posteriori, conforme Rizzato (2011), são apresentadas a seguir.

Os demais modelos são propostas deste trabalho, os quais consideram que a superdispersão detectada por Molenberghs, Verbeke e Demétrio (2007) e Rizzato (2011) é ocasionada pela inflação de zeros. Os oito modelos foram ajustados aos dados por meio do método INLA.

2.2.2.1 Modelo 1: Poisson

Considere que Y_{ij} são independentes, com distribuição de Poisson com parâmetro λ_{ij} , isto é, $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$. Então,

$$P(Y_{ij} = y_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}, \quad \text{para } y_{ij} = 0, 1, 2, \dots,$$

e admite-se que

$$\lambda_{ij} = \exp(\mathbf{x}_{ij}'\boldsymbol{\beta}).$$

Neste caso, considerando os dados do ensaio clínico, o parâmetro do modelo

pode ser expresso por,

$$\lambda_{ij} = \begin{cases} \exp\{\beta_{00} + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \exp\{\beta_{10} + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento.} \end{cases}$$

Considerando o processo de estimação bayesiano é necessário obter a distribuição conjunta a posteriori para os parâmetros do modelo de interesse. Assim,

$$\pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11} | \mathbf{Y}) \propto L(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11} | \mathbf{Y}) \pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}), \quad (40)$$

em que $\pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11} | \mathbf{Y})$ é a distribuição conjunta a posteriori, $L(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11} | \mathbf{Y})$ é a função de verossimilhança e $\pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$ é a distribuição conjunta a priori. Para completar a especificação do modelo, é necessário modelar a incerteza relativa aos parâmetros. Considerando independência entre os parâmetros,

$$\begin{aligned} \pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}) &= \pi(\beta_{00})\pi(\beta_{01})\pi(\beta_{10})\pi(\beta_{11}) \\ &= \prod_{k,l} \pi(\beta_{kl}), \end{aligned} \quad (41)$$

em que $k = 0, 1$, $l = 0, 1$.

Utilizando (41), pode-se reescrever (40) da seguinte forma,

$$\pi(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11} | \mathbf{Y}) = \pi(\beta_{kl} | \mathbf{Y}) \propto L(\beta_{kl} | \mathbf{Y}) \pi(\beta_{kl}).$$

Para a função de verossimilhança,

$$L(\beta_{kl} | \mathbf{Y}) \propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}}.$$

Assumindo-se independência entre os parâmetros $\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}$, a incerteza relativa aos mesmos é incorporada ao modelo considerando-se as seguintes distribuições a priori:

$$\beta_{kl} \sim N(a, b^2)$$

em que a , b , são constantes conhecidas, $k = 0, 1$ e $l = 0, 1$.

Então,

$$\pi(\beta_{kl}) = \prod_{k,l} \frac{1}{\sqrt{2\pi\alpha^2}} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_{kl} - a}{b} \right)^2 \right\} \propto \prod_{k,l} \exp \left\{ -\frac{1}{2b^2} (\beta_{kl} - a)^2 \right\}. \quad (42)$$

Assim, a distribuição conjunta a posteriori é proporcional a

$$\pi(\beta_{kl}|\mathbf{Y}) \propto \exp \left(-\sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ -\frac{1}{2b^2} (\beta_{kl} - a)^2 \right\}. \quad (43)$$

Os resumos a posteriori são obtidos por meio das distribuições condicionais completas a posteriori para todas as quantidades desconhecidas do modelo, conforme (43).

2.2.2.2 Modelo 2: Poisson-gama

Considere que Y_{ij} com distribuição de Poisson com parâmetro λ_{ij} , e para acomodar a superdispersão considere um efeito aleatório θ_{ij} com distribuição gama, isto é,

$$\begin{aligned} Y_{ij}|\lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \theta_{ij} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) \\ \theta_{ij} &\sim \text{gama}(\alpha_{1j}, \alpha_{2j}) \end{aligned}$$

sendo α_{1j} e α_{2j} parâmetros da distribuição gama.

O modelo hierárquico a ser ajustado, impondo a restrição $\alpha_{1j}\alpha_{2j} = 1$, ou seja, $\alpha_{1j} = \alpha_{2j}^{-1}$, conforme Molenberghs, Verbeke e Demétrio (2007) e Rizzato (2011), é dado por,

$$\begin{aligned} Y_{ij}|\theta_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \begin{cases} \theta_{ij} \exp\{\beta_{00} + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \theta_{ij} \exp\{\beta_{10} + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento} \end{cases} \\ \theta_{ij} &\sim \text{gama}(\alpha_{2j}^{-1}, \alpha_{2j}). \end{aligned}$$

Conforme citado por Rizzato (2011), este modelo é uma generalização do modelo binomial negativo.

A função de verossimilhança, nesse caso, pode ser expressa por,

$$L(\beta_{kl}, \alpha_2 | \mathbf{Y}) \propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp \{ -\alpha_2^{-1} \theta_{ij} \}.$$

Para modelar a incerteza relativa aos parâmetros β_{00} , β_{01} , β_{10} e β_{11} , assume-se as seguintes distribuições a priori:

$$\beta_{kl} \sim N(a, b^2),$$

em que a e b são constantes conhecidas, $k = 0, 1$ e $l = 0, 1$. Considera-se também que os θ_{ij} são independentes e a distribuição a priori para o parâmetro α_2 é dada por

$$\alpha_2 \sim \text{Gama}(c, d),$$

em que c e d são constantes conhecidas. A distribuição conjunta a posteriori é proporcional a

$$\begin{aligned} \pi(\beta_{kl}, \alpha_2, \boldsymbol{\theta} | \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ - \frac{1}{2b^2} (\beta_{kl} - a)^2 \right\} \\ &\quad \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp \{ -\alpha_2^{-1} \theta_{ij} \} \alpha_2^{c-1} \exp \{ -\alpha_2/d \}. \end{aligned}$$

Os resumos a posteriori são obtidos por meio das distribuições condicionais completas a posteriori para todas as quantidades desconhecidas do modelo,

$$\begin{aligned} \pi(\beta_{kl} | \alpha_2, \boldsymbol{\theta}, \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ - \frac{1}{2b^2} (\beta_{kl} - a)^2 \right\}, \\ \pi(\alpha_2 | \beta_{kl}, \boldsymbol{\theta}, \mathbf{Y}) &\propto \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp \{ -\alpha_2^{-1} \theta_{ij} \} \alpha_2^{c-1} \exp \{ -\alpha_2/d \}, \\ \pi(\boldsymbol{\theta} | \beta_{kl}, \alpha_2, \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp \{ -\alpha_2^{-1} \theta_{ij} \}. \end{aligned}$$

2.2.2.3 Modelo 3: Poisson-normal

Considere que Y_{ij} com distribuição de Poisson com parâmetro λ_{ij} , e que o efeito aleatório $\mathbf{b}_i$, para acomodar a correlação entre medidas longitudinais, seja normalmente distribuído. Então,

$$\begin{aligned} Y_{ij}|\lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \ln(\lambda_{ij}) &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i \\ b_i &\sim N(\mathbf{0}, \mathbf{D}), \end{aligned}$$

em que $\mathbf{D}$ é a matriz de covariâncias de $\mathbf{b}_i$.

O modelo hierárquico a ser ajustado é,

$$\begin{aligned} Y_{ij}|b_i &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \begin{cases} \exp\{(\beta_{00} + b_i) + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \exp\{(\beta_{10} + b_i) + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento} \end{cases} \\ b_i &\sim N(0, \sigma_b^2). \end{aligned}$$

A função de verossimilhança sob o modelo hierárquico é dada por,

$$L(\beta_{kl}, \mathbf{b}, \sigma_b^2 | \mathbf{Y}) \propto \exp\left(-\sum_{i,j} \lambda_{ij}\right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left\{-\frac{b_i^2}{2\sigma_b^2}\right\},$$

em que $\mathbf{b} = (b_1, \dots, b_N)^T$.

Para modelar a incerteza relativa aos parâmetros β_{00} , β_{01} , β_{10} e β_{11} , assume-se as seguintes distribuições a priori:

$$\beta_{kl} \sim N(a, b^2),$$

em que a e b são constantes conhecidas, $k = 0, 1$ e $l = 0, 1$, e para o parâmetro σ^2 , a distribuição a priori é dada por:

$$\sigma_b^2 \sim \text{Gama Inversa}(g, h),$$

em que g e h são constantes conhecidas, de forma que,

$$\pi(\sigma_b^2) = \frac{h^g}{\Gamma(g)} (\sigma_b^2)^{-(g+1)} \exp \left\{ -\frac{h}{\sigma_b^2} \right\} \propto (\sigma_b^2)^{-(g+1)} \exp \left\{ -\frac{h}{\sigma_b^2} \right\}. \quad (44)$$

Considerando-se, também, independência a priori entre todos os parâmetros, a distribuição conjunta a posteriori é dada por,

$$\begin{aligned} \pi(\beta_{kl}, \mathbf{b}, \sigma_b^2 | \mathbf{Y}) &\propto \exp \left(-\sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ -\frac{1}{2b^2} (\beta_{kl} - a)^2 \right\} \\ &\quad \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ -\frac{b_i^2}{2\sigma_b^2} \right\} (\sigma_b^2)^{-(g+1)} \exp \left\{ -\frac{h}{\sigma_b^2} \right\}. \end{aligned}$$

Os resumos a posteriori para as quantidades desconhecidas do modelo são obtidos por meio das distribuições condicionais completas a posteriori para todas estas quantidades. Desse modo,

$$\begin{aligned} \pi(\beta_{kl} | \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \exp \left(-\sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ -\frac{1}{2b^2} (\beta_{kl} - a)^2 \right\}, \\ \pi(\mathbf{b} | \beta_{kl}, \sigma_b^2, \mathbf{Y}) &\propto \exp \left(-\sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ -\frac{b_i^2}{2\sigma_b^2} \right\}, \\ \pi(\sigma_b^2 | \beta_{kl}, \mathbf{b}, \mathbf{Y}) &\propto \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ -\frac{b_i^2}{2\sigma_b^2} \right\} (\sigma_b^2)^{-(g+1)} \exp \left\{ -\frac{h}{\sigma_b^2} \right\}. \end{aligned}$$

2.2.2.4 Modelo 4: Poisson-normal-gama ou Combinado

Considere que Y_{ij} com distribuição de Poisson com parâmetro λ_{ij} . Diferentes tipos de efeitos aleatórios serão incluídos para levar em conta a correlação entre medidas feitas no mesmo indivíduo e a superdispersão,

$$\begin{aligned} Y_{ij} | \mathbf{b}_i, \theta_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \theta_{ij} \exp(\mathbf{x}_{ij}' \boldsymbol{\beta} + \mathbf{z}_{ij}' \mathbf{b}) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \\ \theta_{ij} &\sim \text{gama}(\alpha_{1j}, \alpha_{2j}) \end{aligned}$$

sendo $\mathbf{D}$ a matriz de variâncias-covariâncias referentes a $\mathbf{b}_i$ e α_{1j} e α_{2j} parâmetros da distribuição gama.

O modelo hierárquico a ser ajustado é a combinação dos dois anteriores,

$$\begin{aligned}
Y_{ij}|b_i, \theta_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \begin{cases} \theta_{ij} \exp\{(\beta_{00} + b_i) + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \theta_{ij} \exp\{(\beta_{10} + b_i) + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento} \end{cases} \\
b_i &\sim N(0, \sigma_b^2) \\
\theta_{ij} &\sim \text{gama}(\alpha_2^{-1}, \alpha_2).
\end{aligned}$$

A função de verossimilhança fica expressa por,

$$\begin{aligned}
L(\beta_{kl}, \boldsymbol{\theta}, \mathbf{b}, \alpha_2, \sigma_b^2 | \mathbf{Y}) &\propto \exp\left(-\sum_{i,j} \lambda_{ij}\right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left\{-\frac{b_i^2}{2\sigma_b^2}\right\} \\
&\quad \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp\{-\alpha_2^{-1}\theta_{ij}\}.
\end{aligned}$$

Para modelar a incerteza relativa aos parâmetros β_{00} , β_{01} , β_{10} e β_{11} , assume-se as seguintes distribuições a priori

$$\beta_{kl} \sim N(a, b^2),$$

em que a e b são constantes conhecidas, $k = 0, 1$ e $l = 0, 1$. Para o parâmetro σ_b^2 , a distribuição a priori é dada por

$$\sigma_b^2 \sim \text{Gama Inversa}(g, h),$$

em que g e h são constantes conhecidas, e para o parâmetro α_2 ,

$$\alpha_2 \sim \text{Gamma}(c, d),$$

em que c e d são constantes conhecidas. Supondo independência a priori entre todos os

parâmetros, a distribuição conjunta a posteriori é dada por,

$$\begin{aligned} \pi(\beta_{kl}, \boldsymbol{\theta}, \mathbf{b}, \alpha_2, \sigma_b^2 | \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \prod_{l=0}^1 \exp \left\{ - \frac{1}{2b^2} (\beta_{kl} - a)^2 \right\} \\ &\quad \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ - \frac{b_i^2}{2\sigma_b^2} \right\} (\sigma_b^2)^{-(g+1)} \exp \left\{ - \frac{h}{\sigma_b^2} \right\} \\ &\quad \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp\{-\alpha_2^{-1}\theta_{ij}\} \alpha_2^{c-1} \exp\{-\alpha_2/d\}. \end{aligned}$$

Os resumos a posteriori são obtidos por meio das distribuições condicionais completas a posteriori para todas as quantidades desconhecidas do modelo,

$$\begin{aligned} \pi(\beta_{kl} | \boldsymbol{\theta}, \mathbf{b}, \alpha_2, \sigma_b^2, \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{k,l} \exp \left\{ - \frac{1}{2b^2} (\beta_{kl} - a)^2 \right\}, \\ \pi(\boldsymbol{\theta} | \beta_{kl}, \mathbf{b}, \alpha_2, \sigma_b^2, \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp\{-\alpha_2^{-1}\theta_{ij}\}, \\ \pi(\mathbf{b} | \beta_{kl}, \boldsymbol{\theta}, \alpha_2, \sigma_b^2, \mathbf{Y}) &\propto \exp \left(- \sum_{i,j} \lambda_{ij} \right) \prod_{i,j} \lambda_{ij}^{y_{ij}} \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ - \frac{b_i^2}{2\sigma_b^2} \right\}, \\ \pi(\alpha_2 | \beta_{kl}, \boldsymbol{\theta}, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{i,j} \frac{\alpha_2^{-\alpha_2^{-1}}}{\Gamma(\alpha_2^{-1})} \theta_{ij}^{\alpha_2^{-1}-1} \exp\{-\alpha_2^{-1}\theta_{ij}\} \alpha_2^{c-1} \exp\{-\alpha_2/d\}, \\ \pi(\sigma_b^2 | \beta_{kl}, \boldsymbol{\theta}, \mathbf{b}, \alpha_2, \mathbf{Y}) &\propto \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left\{ - \frac{b_i^2}{2\sigma_b^2} \right\} (\sigma_b^2)^{-(g+1)} \exp \left\{ - \frac{h}{\sigma_b^2} \right\}. \end{aligned}$$

2.2.2.5 Modelo 5: Poisson inflacionado de zeros - ZIP

Suponha que os Y_{ij} com distribuição de Poisson inflacionado de zeros (ZIP) com parâmetro λ_{ij} e probabilidade de zeros estruturais dada por ω , isto é, $Y_{ij} \sim \text{ZIP}(\lambda_{ij}, \omega)$, $\lambda_{ij} > 0$. Então,

$$P(Y_{ij} = y_{ij}) = \begin{cases} \omega + (1 - \omega)e^{-\lambda_{ij}}, & \text{se } y_{ij} = 0 \\ (1 - \omega) \frac{\lambda_{ij} e^{-\lambda_{ij}}}{y_{ij}!}, & \text{se } y_{ij} > 0. \end{cases}$$

Admitindo-se que

$$\ln(\lambda_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} \quad \text{e} \quad \ln \left(\frac{\omega}{1 - \omega} \right) = \mathbf{z}_{ij}^T \boldsymbol{\gamma}$$

em que $\mathbf{x}_{ij}^T$ é a matriz de covariáveis, $\mathbf{z}_{ij}^T$, neste caso, é um vetor coluna de 1's, de forma que $\boldsymbol{\beta}$ é um vetor de parâmetros e $\boldsymbol{\gamma} = \gamma$.

Para o parâmetro λ_{ij} ,

$$\lambda_{ij} = \begin{cases} \exp\{\beta_{00} + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \exp\{\beta_{10} + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento.} \end{cases}$$

Para a função de verossimilhança,

$$L(\beta_{\{kl\}}, \gamma | \mathbf{Y}) = \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!}. \quad (45)$$

Assumindo-se independência entre os parâmetros β_{00} , β_{01} , β_{10} , β_{11} e ω , a incerteza relativa aos mesmos é incorporada ao modelo considerando-se as seguintes distribuições a priori:

$$\beta_{kl} \sim N(a, b^2) \quad \text{e} \quad \omega \sim \text{Beta}(c, d),$$

em que a , b , c e d são constantes conhecidas, $k = 0, 1$ e $l = 0, 1$. A distribuição a priori para ω pode ser expressa por

$$\begin{aligned} \pi(\omega) &= \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} \omega^{c-1} (1-\omega)^{d-1} \\ &\propto \omega^{c-1} (1-\omega)^{d-1}. \end{aligned} \quad (46)$$

Devido à restrição do GMRF para o ajuste pelo método INLA, Rue, Martino e Chopin (2009) propõem uma reparametrização para modelar a incerteza relativa à ω , modelando $\logit(\omega)$.

Utilizando (45), (42) e (46), a distribuição conjunta a posteriori expressa por,

$$\begin{aligned} \pi(\beta_{kl}, \omega | \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \prod_{k,l} \exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \omega^{c-1} (1-\omega)^{d-1}. \end{aligned}$$

Os resumos a posteriori no processo bayesiano são obtidos por meio das distribuições condicionais completas a posteriori para todas as quantidades desconhecidas do modelo, a saber:

$$\begin{aligned}\pi(\beta_{kl}|\omega, \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \\ \pi(\omega|\beta_{kl}, \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \omega^{c-1}(1 - \omega)^{d-1}.\end{aligned}$$

2.2.2.6 Modelo 6: Binomial negativa inflacionada de zeros - ZINB

Considere que as observações de Y_{ij} são descritas pela distribuição de probabilidade ZINB, expressa por

$$P(Y_{ij} = y_{ij}) = \begin{cases} \omega + (1 - \omega)(1 + \alpha \lambda_{ij}^c)^{-\frac{\lambda_{ij}^{1-c}}{\alpha}} & , \text{ se } y_{ij} = 0 \\ (1 - \omega) \frac{\Gamma\left(y_{ij} + \frac{\lambda_{ij}^{1-c}}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{\lambda_{ij}^{1-c}}{c}\right)} (1 + \alpha \lambda_{ij}^c)^{-\frac{\lambda_{ij}^{1-c}}{\alpha}} \left(1 + \frac{\lambda_{ij}^{1-c}}{\alpha}\right)^{-y} & , \text{ se } y_{ij} > 0, \end{cases}$$

em que ω é a probabilidade de zeros estruturais, λ é a taxa, α é o parâmetro de superdispersão e c uma constante arbitrária. Como em Lawless (1987), Hinde e Demétrio (1998) e Nelder e Wedderburn (1972), considerou-se $c = 1$, obtendo-se,

$$P(Y_{ij} = y_{ij}) = \begin{cases} \omega + (1 - \omega)(1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} & , \text{ se } y_{ij} = 0 \\ (1 - \omega) \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y} & , \text{ se } y_{ij} > 0, \end{cases} \quad (47)$$

como parametrizado por Rue, Martino e Chopin (2009) para o ajuste em INLA.

Assume-se que,

$$\ln(\lambda_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} \quad \text{e} \quad \ln\left(\frac{\omega}{1-\omega}\right) = \mathbf{z}_{ij}^T \boldsymbol{\gamma}$$

e considera-se as mesmas especificações para λ_{ij} no modelo ZIP. A função de verossimilhança sob o modelo de regressão ZINB pode ser expressa por,

$$L(\beta_{kl}, \omega, \alpha | \mathbf{Y}) = \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right].$$

Para modelar a incerteza relativa aos parâmetros, assumindo independência entre eles, considere,

$$\beta_{kl} \sim N(a, b^2), \quad \omega \sim \text{beta}(c, d), \quad \text{e} \quad \alpha \sim \text{gama}(f, g),$$

em que a, b, c, d, f e g são constantes conhecidas e $k = 0, 1, l = 0, 1$. As expressões das distribuições à priori de $\pi(\beta_{kl})$ e $\pi(\omega)$ estão expressas em (42) e (46), respectivamente. Para modelar a incerteza relativa à α , considere,

$$\pi(\alpha) = \frac{g^f}{\Gamma(f)} \alpha^{f-1} \exp(-g\alpha) \propto \alpha^{f-1} \exp(-g\alpha). \quad (48)$$

Considerando independência entre todos os parâmetros, a distribuição conjunta a posteriori é dada por

$$\begin{aligned} \pi(\beta_{kl}, \omega, \alpha | \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\ &\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\ &\prod_{k,l} \exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \omega^{c-1} (1-\omega)^{d-1} \alpha^{f-1} \exp(-g\alpha). \end{aligned}$$

As distribuições condicionais completas a posteriori para as quantidades desconhecidas podem ser expressas por,

$$\begin{aligned}
\pi(\beta_{kl}|\omega, \alpha, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \\
\pi(\omega|\beta_{\{kl\}}, \alpha, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\omega^{c-1}(1-\omega)^{d-1}, \\
\pi(\alpha|\beta_{\{kl\}}, \omega, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\alpha^{f-1} \exp(-g\alpha).
\end{aligned}$$

2.2.2.7 Modelo 7: ZIP-normal

Considere a distribuição ZIP para Y_{ij} e a distribuição normal para o efeito aleatório $\mathbf{b}_i$, para acomodar a correlação entre dados longitudinais, tem-se

$$\begin{aligned}
Y_{ij}|\lambda_{ij}, \omega &\sim \text{ZIP}(\lambda_{ij}, \omega) \\
\ln(\lambda_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i \\
\ln\left(\frac{\omega}{1-\omega}\right) &= \mathbf{z}_{ij}^T \boldsymbol{\gamma} \\
\mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}),
\end{aligned}$$

em que $\mathbf{D}$ é a matriz de covariâncias de $\mathbf{b}_i$. Segue que,

$$\begin{aligned}\lambda_{ij} &= \begin{cases} \exp\{(\beta_{00} + b_i) + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \exp\{(\beta_{10} + b_i) + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento}, \end{cases} \\ b_i &\sim N(0, \sigma_b^2).\end{aligned}$$

A função de verossimilhança fica especificada por

$$\begin{aligned}L(\beta_{kl}, \omega, \mathbf{b}_i, \sigma_b | \mathbf{Y}) &= \prod_{ij} P(Y_{ij} | b_i) f(b_i) \\ &= \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \prod_i \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_b^2}} \exp\left\{-\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2}\right)\right\} \\ &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp\left\{-\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2}\right)\right\}.\end{aligned}\tag{49}$$

Para o ajuste bayesiano hierárquico do modelo 7, inicialmente, considere distribuições a priori normais para os parâmetros β_{kl} , como em (42), e priori beta para ω , como em (46). Para completar a especificação do modelo, considere que b_i são variáveis aleatórias independentes com distribuição $N(0, \sigma^2)$. Assim, a distribuição a priori para σ_b^2 é dada por $\sigma_b^2 \sim \text{Gama Inversa}(g, h)$, expressa em (44).

Utilizando-se (42), (46), (49), (44) e considerando-se, também, independência entre todos os parâmetros de interesse, a distribuição conjunta a posteriori é expressa por

$$\begin{aligned}\pi(\beta_{kl}, \omega, \mathbf{b}, \sigma_b^2 | \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp\left\{-\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2}\right)\right\} \prod_{k,l} \exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \\ &\quad \omega^{c-1} (1 - \omega)^{d-1} (\sigma_b^2)^{-(r+1)} \exp\left\{-\frac{s}{\sigma_b^2}\right\}.\end{aligned}$$

Os resumos a posteriori no processo bayesiano são obtidos por meio das dis-

tribuições condicionais completas a posteriori para todas as quantidades desconhecidas do modelo, a saber:

$$\begin{aligned}
\pi(\beta_{kl}|\omega, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\
&\quad \exp \left\{ -\frac{1}{2b^2} (\beta_{kl} - a)^2 \right\} \\
\pi(\omega|\beta_{kl}, \omega, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\
&\quad \omega^{c-1} (1 - \omega)^{d-1} \\
\pi(\mathbf{b}|\beta_{kl}, \omega, \sigma_b^2|\mathbf{Y}) &\propto \prod_{y_{ij}=0} [\omega + (1 - \omega) \exp(-\lambda_{ij})] \prod_{y_{ij}>0} \frac{(1 - \omega) \exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!} \\
&\quad \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\} \\
\pi(\sigma_b^2|\beta_{kl}, \omega, \mathbf{b}, \mathbf{Y}) &\propto \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\} (\sigma_b^2)^{-(r+1)} \exp \left\{ -\frac{s}{\sigma_b^2} \right\}.
\end{aligned}$$

2.2.2.8 Modelo 8: ZINB-normal

Por fim, considere que Y_{ij} segue distribuição ZINB, com distribuição dada por (47), e a inclusão de uma variável aleatória $\mathbf{b}_i$ para acomodar a correlação entre dados longitudinais. Neste caso, para a modelagem bayesiana hierárquica,

$$\begin{aligned}
Y_{ij}|\lambda_{ij}, \omega, \alpha &\sim ZINB(\lambda_{ij}, \omega) \\
\ln(\lambda_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i \\
\ln \left(\frac{\omega}{1 - \omega} \right) &= \mathbf{z}_{ij}^T \boldsymbol{\gamma} \\
\mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}),
\end{aligned}$$

em que $\mathbf{D}$ é a matriz de covariâncias de $\mathbf{b}_i$. Segue que,

$$\begin{aligned}
\lambda_{ij} &= \begin{cases} \exp\{(\beta_{00} + b_i) + \beta_{01}t_{ij}\} & \text{se } i \in \text{placebo} \\ \exp\{(\beta_{10} + b_i) + \beta_{11}t_{ij}\} & \text{se } i \in \text{tratamento}, \end{cases} \\
b_i &\sim N(0, \sigma_b^2).
\end{aligned}$$

A função de verossimilhança fica especificada por

$$\begin{aligned}
L(\beta_{kl}, \omega, \alpha, \mathbf{b}_i, \sigma_b | \mathbf{Y}) &= \prod_{ij} P(Y_{ij} | b_i) f(b_i) \\
&= \prod_{y_{ij}=0} \left[\omega + (1 - \omega)(1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\quad \prod_{y_{ij}>0} \left[(1 - \omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\quad \prod_i \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\} \\
&\propto \prod_{y_{ij}=0} \left[\omega + (1 - \omega)(1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\quad \prod_{y_{ij}>0} \left[(1 - \omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha \lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\quad \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\}. \tag{50}
\end{aligned}$$

Para completar a especificação do modelo hierárquico, considere as distribuições a priori normais para os parâmetros β_{kl} (42), priori beta para ω (46) e priori inversa gama para o parâmetro σ_b^2 (44). Supondo independência a priori entre todos os parâmetros e que os efeitos aleatórios são normais e independentemente distribuídos, a distribuição conjunta a posteriori é parametrizada por

$$\begin{aligned}
\pi(\beta_{kl}, \omega, \alpha, \mathbf{b}, \sigma_b^2 | \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp\left\{-\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2}\right)\right\} \prod_{k,l} \exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\} \\
&\omega^{c-1} (1-\omega)^{d-1} \alpha^{f-1} \exp(-g\alpha) (\sigma_b^2)^{-(r+1)} \exp\left\{-\frac{s}{\sigma_b^2}\right\}.
\end{aligned}$$

As distribuições condicionais completas a posteriori para as quantidades desconhecidas para a obtenção dos resumos a posteriori podem ser expressas por

$$\begin{aligned}
\pi(\beta_{kl} | \omega, \alpha, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\exp\left\{-\frac{1}{2b^2}(\beta_{kl} - a)^2\right\}, \\
\pi(\omega | \beta_{kl}, \alpha, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\omega^{c-1} (1-\omega)^{d-1}, \\
\pi(\alpha | \beta_{kl}, \omega, \mathbf{b}, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1 + \alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\alpha^{f-1} \exp(-g\alpha),
\end{aligned}$$

$$\begin{aligned}
\pi(\mathbf{b}|\beta_{kl}, \omega, \alpha, \sigma_b^2, \mathbf{Y}) &\propto \prod_{y_{ij}=0} \left[\omega + (1-\omega)(1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \right] \\
&\prod_{y_{ij}>0} \left[(1-\omega) \frac{\Gamma\left(y_{ij} + \frac{1}{\alpha}\right)}{y_{ij}! \Gamma\left(\frac{1}{\alpha}\right)} (1+\alpha\lambda_{ij})^{-\frac{1}{\alpha}} \left(1 + \frac{\lambda_{ij}^{-1}}{\alpha}\right)^{-y_{ij}} \right] \\
&\prod_i \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\}, \\
\pi(\sigma_b^2|\beta_{kl}, \omega, \alpha, \mathbf{b}, \mathbf{Y}) &\propto \prod_i \frac{1}{\sqrt{\sigma_b^2}} \exp \left\{ -\frac{1}{2} \left(\frac{b_i^2}{\sigma_b^2} \right) \right\} (\sigma_b^2)^{-(r+1)} \exp \left\{ -\frac{s}{\sigma_b^2} \right\}.
\end{aligned}$$

2.2.3 Implementação computacional

A implementação computacional foi realizada no programa R, R Core Team (2015), utilizando o pacote INLA (RUE, MARTINO, CHOPIN, 2009), além de rotinas específicas criadas para obtenção dos resumos a posteriori. O pacote INLA encontra-se disponível para download em <http://www.r-inla.org/>, podendo também ser obtido diretamente do R por meio de linhas de comandos específicas. As rotinas de análise e as linhas de comando para obtenção do INLA são apresentadas no apêndice deste trabalho.

Para os modelos propostos e os apresentados por Rizzato (2011), verificou-se a qualidade do ajuste e seleção por meio do valor de KLD e inspecionando os valores de CPO e PIT. A escolha entre dois modelos foi realizada por meio do PsFB. Para poder comparar os tratamentos, criaram-se contrastes para comparar os parâmetros β_{11} e β_{01} em cada ajuste, isto é, o contraste Δ_1 , $\Delta_1 = \beta_{11} - \beta_{01}$, e o contraste Δ_2 dado pela razão entre os dois parâmetros, $\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$. Dessa maneira, se o valor 0 ou 1 pertence ao intervalo de credibilidade associado aos contrastes Δ_1 e Δ_2 , respectivamente, conclui-se que não há diferença entre os dois tratamentos.

2.2.3.1 Estudo de simulação

Neste trabalho, para a seleção entre dois modelos, será considerado o PsFB, apresentado na seção 2.1.7.5.4. Para verificar o comportamento desta medida, considerando o modelo selecionado neste trabalho, assim como, as estimativas dos seus parâmetros, realizou-se um estudo de simulação para verificar o comportamento da medida utilizada para seleção do modelo.

Simulou-se 1000 conjuntos de dados considerando as mesmas especificações dos dados deste trabalho, baseando-se no modelo selecionado e nas suas estimativas. Ressalta-se que neste trabalho, após o ajuste de todos os modelos propostos, dois modelos serão considerados candidatos à descrever o comportamento dos dados. Assim, ajustou-se esses dois modelos aos dados simulados e observou-se qual modelo era selecionado segundo o PsFB. Mensurou-se a porcentagem de acerto e em seguida, apresentou-se a média das estimativas dos parâmetros obtidas em cada ajuste.

3 RESULTADOS E DISCUSSÃO

Inicialmente será feita uma análise exploratória dos dados. Na Figura 2 tem-se os gráficos dos perfis individuais para o número de ataques epiléticos sofridos em cada uma das 27 semanas pelos pacientes submetidos ao tratamento placebo e a nova droga epilética.

Figura 2 - Gráfico de perfis individuais para o número de ataques epiléticos sofridos em cada uma das 27 semanas pelos pacientes submetidos ao tratamento placebo (0) e à nova droga epilética (1)

De acordo com a Figura 2, observa-se que há um comportamento instável e a presença de valores extremos, tanto para o grupo de indivíduos tratados com o placebo, como para aqueles que receberam a nova droga epilética. A presença de valores extremos é um indício de superdispersão, sobretudo com a perda de informação. Nota-se que para o grupo placebo não houve observações a partir da 21^a semana, assim como, na 17^a semana, o número de pacientes em observação para ambos os grupos se reduz a um pouco menos da

metade do valor inicial, como pode ser verificado na Tabela 2, o que também foi mostrado por Rizzato (2011), graficamente e por meio de análise tabular.

Tabela 2 - Número de observações tomadas em cada uma das semanas do estudo, em ambos os grupos

Tempo (semanas)	Número de observações		Tempo (semanas)	Número de observações	
	Placebo	Tratamento		Placebo	Tratamento
1	45	44	16	40	37
5	42	42	17	18	17
10	41	40	20	2	8
15	40	38	27	0	3

Retornando a Figura 2, há indícios de que o comportamento ao longo do tempo, para ambos os grupos sejam similares, graficamente. O ajuste de um modelo de regressão faz-se necessário para evidenciar estatisticamente (ou não) este indício.

Inicialmente, foram considerados os modelos propostos por Molenberghs, Verbeke e Demétrio (2007) e particularizados sob a abordagem bayesiana por Rizzato (2011), modelo Poisson, modelo Poisson-gama, modelo Poisson-normal e modelo Poisson-normal-gama (ou modelo combinado). Para verificar a qualidade do ajuste, as Figuras 3, 5, 7 e 9 apresentam os gráficos dos valores de PIT, bem como o histograma desses.

Para o primeiro modelo (Figura 3), em que foi considerado o ajuste de um modelo de regressão de Poisson, apenas de efeitos fixos, pelas medidas de PIT, verifica-se que o comportamento das mesmas se distancia da distribuição do modelo uniforme, o que pode ser notado tanto na distribuição dos valores (Figura 3 (a)), como no histograma destes (Figura 3 (b)), indicando um ajuste não desejável. Para este mesmo ajuste, separou-se os dados quanto aos tratamentos, ou seja, separou-se as contagens de ataques epiléticos em um grupo formado pelos indivíduos que receberam o tratamento placebo e em outro grupo formado pelos indivíduos que receberam a nova droga anti-epilética. Na Figura 4, tem-se os histogramas e os gráficos quantil-quantil uniforme considerando os dados dos pacientes submetidos ao tratamento placebo (Figura 4(a)), à nova droga (Figura 4(b)) e todos os pacientes, e verifica-se também que em todos os casos há um distanciamento do comportamento de uma uniforme padrão.

Para o modelo Poisson-gama, considerou-se o ajuste de um modelo de regressão de Poisson, com a inserção de um efeito aleatório gama pré-multiplicando o preditor linear, para acomodar a superdispersão. Nota-se que o comportamento das medidas de PIT, Figuras 5 e 6, também se distanciam da distribuição uniforme, não apresentado, inclusive, uma melhoria no ajuste, em comparação com o modelo anterior, indicando que apenas o efeito aleatório gama não está captando toda a variabilidade extra-Poisson existente nos dados.

No modelo Poisson-normal tem-se o ajuste um modelo Poisson com efeitos aleatórios normais, com o intuito de acomodar a correlação existente entre dados longitudinais. Nota-se, por meio das Figuras 7 e 8 uma melhoria no comportamento dos valores de PIT, em comparação com os modelos 1 e 2, todavia, sabe-se que o fenômeno de superdispersão e a existência de correlação entre medidas podem ocorrer simultaneamente e este modelo, acomoda apenas a correlação.

Para o modelo combinado, combinaram-se os modelos Poisson-gama e Poisson-normal, isto é, foi considerado um efeito aleatório gama para acomodar a superdispersão e um efeito aleatório normal para captar a correlação entre medidas tomadas no mesmo indivíduo. De acordo com as Figuras 9 e 10, os valores de PIT apresentam um comportamento razoável, todavia, observando-se o histograma (Figura 9 (b)), notam-se intervalos de classes com baixa frequência, havendo a necessidade de se estudarem esses valores.

Observando-se a Figura 11 nota-se que a melhor capacidade preditiva, entre os modelos propostos, é para o modelo combinado, pois comparando-se o histograma dos valores de CPO deste modelo, com os demais, nota-se uma menor frequência acumulada para baixos valores de CPO. Para esses modelos, considerando a capacidade preditiva e a adequação da distribuição proposta aos dados, o modelo 4 deve ser adotado.

Figura 3 - (a) Gráfico dos valores de PIT para o modelo Poisson. (b) Histograma dos valores de PIT para o modelo Poisson

Figura 4 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson

Figura 5 - (a) Gráfico dos valores de PIT para o modelo Poisson-gama. (b) Histograma dos valores de PIT para o modelo Poisson-gama

Figura 6 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-gama

Figura 7 - (a) Gráfico dos valores de PIT para o modelo Poisson-normal. (b) Histograma dos valores de PIT para o modelo Poisson-normal

Figura 8 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo Poisson-normal

Figura 9 - (a) Gráfico dos valores de PIT para o modelo combinado. (b) Histograma dos valores de PIT para o modelo combinado

Figura 10 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo combinado

Ademais, nas Figuras 12, 13, 14 e 15 apresentam-se os gráficos das densidades marginais a posteriori para os parâmetros β_{00} , β_{01} , β_{10} e β_{11} , para os modelos 1, 2, 3 e

4, respectivamente. Observa-se que em todos os casos o comportamento das densidades é suave, não apresentando bimodalidade, características desejáveis, uma vez que a presença de bimodalidade pode acarretar problemas na aproximação de Laplace, que é baseada na obtenção do máximo de uma função. O mesmo ocorre para as densidades marginais dos parâmetros α e β , associados ao efeito aleatório gama, nos modelos 2 e 4, conforme as Figuras 13 e 15, respectivamente, e para os parâmetros σ^2 associados ao efeito aleatório normal nos modelos 3 e 4 (Figuras 14 e 15, respectivamente).

Na Tabela 3, apresentam-se as medidas resumos a posteriori para os parâmetros dos modelos Poisson, Poisson-gama, Poisson-normal e combinado, assim como, os contrastes criados para a comparação de tratamentos. As medidas obtidas por Rizzato (2011) são apresentadas na Tabela 4. Observa-se que os valores obtidos pela aproximação INLA são muito próximos dos valores obtidos por Rizzato(2011), levando às mesmas conclusões, sendo que aproximação INLA foi exata, pois KLD=0. Se considerado o modelo combinado para descrever o comportamento dos ataques epilépticos dos pacientes ao longo do tempo, conclui-se que não há diferença entre os tratamentos ao nível de 95% de credibilidade, pois os valores zero e um pertencem aos intervalos de credibilidades para os contrastes Δ_1 e Δ_2 , respectivamente. Além disso, verifica-se que para os demais modelos não existe diferença entre os tratamentos, observando-se os intervalos de credibilidade para os dois contrastes, exceto para o modelo Poisson.

Retornando à Figura 2, observa-se que há indivíduos que, ao longo do tempo, os ataques epilépticos reduziram, tanto para o tratamento placebo, como para a nova droga. Por outro lado, existem indivíduos cujas trajetórias dos seus perfis apresenta um comportamento contrário, isto é, os ataques aumentaram ao longo do tempo. Para esses, a ocorrência de zero ataques pode estar associada a zeros estruturais, ou seja, a não ocorrência de ataques está associada a outro efeito que não seja o da nova droga ou do placebo, eventualmente, devido à estabilização anterior ao estudo. Baseado nesse indício gráfico e no comportamento das medidas de PIT e CPO, sugere-se considerar modelos que captem a ocorrência de zeros estruturais e não-estruturais.

Figura 11 - (a) Histograma dos valores de CPO para o modelo Poisson. (b) Histograma dos valores de CPO para o modelo Poisson-gama. (c) Histograma dos valores de CPO para o modelo Poisson-normal. (d) Histograma dos valores de CPO para o modelo combinado

Figura 12 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} e β_{11} sob o modelo Poisson

Figura 13 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , α e β sob o modelo Poisson-gama

Figura 14 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} e σ^2 sob o modelo Poisson-normal

Figura 15 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , α , β e σ^2 sob o modelo combinado

Tabela 3 - Resumos a posteriori para os parâmetros dos modelos Poisson, Poisson-gama, Poisson-normal e combinado

Modelo Poisson					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	KLD
β_{00}	1,2663	0,0424	1,2665	(1,1826 ; 1,3490)	0
β_{01}	-0,0133	0,0043	-0,0133	(-0,0219 ; -0,0048)	0
β_{10}	1,4533	0,0383	1,4534	(1,3777 ; 1,4537)	0
β_{11}	-0,0328	0,0038	-0,0328	(-0,0403 ; -0,0253)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0189	0,0058	-0,0189	(-0,0302 ; -0,0075)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	2,6923	1,3564	2,3689	(1,3402 ; 5,9574)	-
Modelo Poisson-gama					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	1,2587	0,1121	1,2577	(1,0416 ; 1,4814)	0
β_{01}	-0,0126	0,0111	-0,0126	(-0,0343 ; 0,0092)	0
β_{10}	1,4744	0,1095	1,4735	(1,2621 ; 1,6918)	0
β_{11}	-0,0353	0,0102	-0,0353	(-0,0552 ; -0,0153)	0
α	0,5264	0,0255	0,5258	(0,4781 ; 0,5783)	0
β	1,9045	0,0918	1,9023	(1,7308 ; 2,0894)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0222	0,0151	-0,0222	(-0,0518 ; 0,0074)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	3,9448	86,7636	2,1716	(-21,8108 ; 22,6276)	-
Modelo Poisson-normal					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	0,8178	0,1681	0,8181	(0,4865 ; 1,1473)	0
β_{01}	-0,0143	0,0044	-0,0143	(-0,0229 ; -0,0057)	0
β_{10}	0,6475	0,1705	0,6479	(0,3112 ; 0,9816)	0
β_{11}	-0,0120	0,0043	-0,0120	(-0,0205 ; -0,0035)	0
σ^2	1,0838	0,0862	1,0791	(0,9272 ; 1,2638)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	0,0023	0,0062	0,0023	(-0,0099 ; 0,0144)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,9207	0,5124	0,8229	(0,2437 ; 2,4286)	-
Modelo combinado					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	0,9119	0,1767	0,9122	(0,5634 ; 1,2583)	0
β_{01}	-0,0249	0,0077	-0,0249	(-0,0399 ; -0,0099)	0
β_{10}	0,6561	0,1794	0,6567	(0,3015 ; 1,0072)	0
β_{11}	-0,0118	0,0075	-0,0118	(-0,0265 ; 0,0028)	0
α	2,4404	0,2090	2,4273	(2,0656 ; 2,886)	0
β	0,4129	0,0349	0,4122	(0,3470 ; 0,4832)	0
σ^2	1,0727	0,0882	1,0700	(0,9079 ; 1,2522)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	0,0130	0,0107	0,0130	(-0,0008 ; 0,0340)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,5425	0,4468	0,4887	(-0,0979 ; 1,5964)	-

Tabela 4 - Resumos a posteriori para os parâmetros dos modelos Poisson, Poisson-gama, Poisson-normal e combinado apresentadas por Rizzato (2011)

Modelo Poisson				
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade
β_{00}	1,2660	0,0424	1,2650	(1,1820 ; 1,3470)
β_{01}	-0,0133	0,0043	-0,0133	(-0,0217 ; -0,0047)
β_{10}	1,4530	0,0380	1,4530	(1,3780 ; 1,5270)
β_{11}	-0,0328	0,0038	-0,0329	(-0,0402 ; -0,0253)
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0195	0,0057	-0,0195	(-0,0308 ; -0,0082)
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	2,8910	13,8980	2,4580	(1,4110 ; 6,9910)
Modelo Poisson-gama				
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade
β_{00}	1,2650	0,1140	1,2630	(1,0440 ; 1,4920)
β_{01}	-0,0128	0,0112	-0,0128	(-0,0345 ; 0,0093)
β_{10}	1,4750	0,1096	1,4740	(1,2640 ; 1,6920)
β_{11}	-0,0350	0,0102	-0,0349	(-0,0550 ; -0,0150)
α	0,5258	0,0255	0,5248	(0,4773 ; 0,5783)
β	1,9060	0,0925	1,9050	(1,7290 ; 2,0950)
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0222	0,0149	-0,0222	(-0,0516 ; 0,0072)
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	2,5460	64,5300	2,0940	(-21,1600 ; 27,8600)
Modelo Poisson-normal				
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade
β_{00}	0,8083	0,1700	0,8098	(0,4647 ; 1,1370)
β_{01}	-0,0143	0,0044	-0,0143	(-0,0231 ; -0,0056)
β_{10}	0,6414	0,1699	0,6431	(0,3099 ; 0,9728)
β_{11}	-0,0120	0,0043	-0,0120	(-0,0205 ; -0,0035)
σ^2	1,2110	0,2012	1,1890	(0,8767 ; 1,6630)
$\Delta_1 = \beta_{11} - \beta_{01}$	0,0023	0,0062	0,0022	(-0,0099 ; 0,0147)
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,9573	1,5560	0,8414	(0,2208 ; 2,4290)
Modelo combinado				
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade
β_{00}	0,8955	0,1829	0,8937	(0,5365 ; 1,2530)
β_{01}	-0,0248	0,0077	-0,0248	(-0,0401 ; -0,0096)
β_{10}	0,6575	0,1838	0,6590	(0,2927 ; 1,0140)
β_{11}	-0,0119	0,0076	-0,0120	(-0,0271 ; 0,0029)
α	2,4620	0,2113	2,4520	(2,0800 ; 2,9050)
β	0,4092	0,0350	0,4078	(0,3443 ; 0,4808)
σ^2	1,1910	0,2027	1,1680	(0,8549 ; 1,6490)
$\Delta_1 = \beta_{11} - \beta_{01}$	0,0129	0,0108	0,0130	(-0,0083 ; 0,0342)
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,5637	1,7300	0,4800	(-0,1283 ; 1,6330)

Levando em conta uma possível presença de zeros estruturais, a seguir, apresentam-se os ajustes dos modelos ZIP, ZINB, ZIP-normal e ZINB-normal. De maneira similar aos modelos anteriores, inicialmente estuda-se o comportamento das medidas de PIT para estes modelos propostos.

No primeiro caso, modelo ZIP, considera-se o ajuste de um modelo de regressão ZIP, de efeitos fixos, para acomodar a superdispersão. Para este ajuste, o comportamento das medidas de PIT se distanciam da distribuição uniforme, não indicando um bom ajuste, conforme as Figuras 16 e 17. Observe que assim como nos modelos anteriores, obteve-se também os histogramas para os valores de PIT, separando os dados em função dos grupos de tratamentos e para estes, obteve-se o gráfico quantil-quantil uniforme, assim como para todos os dados, no intuito de verificar se ocorre a aproximação dos dados ao comportamento da distribuição uniforme padrão. O mesmo é feito para os demais modelos.

Para o modelo ZINB, considera-se o ajuste de um modelo de regressão ZINB, o qual modela a superdispersão, além de descrever a existência de zeros estruturais. Observando os valores de PIT, nota-se que o comportamento das medidas também se distanciam da distribuição uniforme (Figuras 18 e 19).

Nos dois casos anteriores, os modelos acomodam uma possível superdispersão nos dados, no primeiro devido a inflação de zeros e no segundo caso, além da inflação de zeros, a superdispersão na parte não nula. Todavia, no caso de dados longitudinais deve-se também inserir termos no modelo que considere a correlação entre medidas tomadas no mesmo indivíduo. Nas Figuras 20 e 21, têm-se os valores de PIT para o modelo ZIP-normal, em que se insere um efeito aleatório normal no preditor linear do modelo, para acomodar essa correlação. Observa-se que o comportamento dos valores de PIT são satisfatórios, se aproximando mais da distribuição uniforme, quando comparado com os dois modelos anteriores.

Uma outra possibilidade é também considerar efeitos aleatórios normais no modelo ZINB, para que este, além de acomodar a superdispersão, leve em conta a correlação entre medidas longitudinais. Nas Figuras 22 e 23, têm-se os gráficos das medidas de PIT, todavia, nota-se que há um distanciamento do comportamento da distribuição uniforme.

Entre os novos modelos propostos neste trabalho (modelos 5 à 8), observa-se que, com base nas medidas de PIT, o modelo ZIP-normal apresenta um comportamento mais satisfatório, indicando uma melhor adequação da distribuição de probabilidades da variável resposta aos dados, isto é, o comportamento dos valores de PIT se aproximam do comportamento teórico de uma distribuição uniforme padrão. Aqui, foi considerado o comportamento dos valores de PIT na totalidade dos dados e quando estes foram separados dado o tratamento.

Figura 16 - (a) Gráfico dos valores de PIT para o modelo ZIP. (b) Histograma dos valores de PIT para o modelo ZIP

Figura 17 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP

Figura 18 - (a) Gráfico dos valores de PIT para o modelo ZINB. (b) Histograma dos valores de PIT para o modelo ZINB

Figura 19 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB

Figura 20 - (a) Gráfico dos valores de PIT para o modelo ZIP-normal. (b) Histograma dos valores de PIT para o modelo ZIP-normal

Figura 21 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZIP-normal

Figura 22 - (a) Gráfico dos valores de PIT para o modelo ZINB-normal. (b) Histograma dos valores de PIT para o modelo ZINB-normal

Figura 23 - (a) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal - tratamento placebo. (b) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal - nova droga. (c) Histograma e gráfico quantil-quantil uniforme dos valores de PIT para o modelo ZINB-normal

Na Figura 24 nota-se que a melhor capacidade preditiva, entre os novos modelos propostos, é para o modelo ZIP-normal, pois comparando-se o histograma dos valores de

CPO do modelo ZIP-normal com os demais modelos, este apresenta uma menor frequência acumulada para baixos valores de CPO, isto é, uma quantidade menor de valores com baixa capacidade preditiva. Caso fossem considerados esses modelos, considerando a capacidade preditiva e a adequação da distribuição proposta aos dados, o modelo ZIP-normal deveria ser adotado.

Figura 24 - (a) Histograma dos valores de CPO para o modelo ZIP. (b) Histograma dos valores de CPO para o modelo ZINB. (c) Histograma dos valores de CPO para o modelo ZIP-normal. (d) Histograma dos valores de CPO para o modelo ZINB-normal

Em relação as Figuras 25, 26, 27 e 28, têm-se os gráficos das densidades marginais a posteriori para os parâmetros β_{00} , β_{01} , β_{10} e β_{11} , para os modelos ZIP, ZINB, ZIP-normal e ZINB-normal, respectivamente. Observa-se que em todos os casos o comportamento das densidades são suaves, não apresentando bimodalidade, características desejáveis, pois a aproximação de Laplace é afetada na maximização de funções que não possuam moda única. Para o parâmetro relativo à probabilidade de zeros estruturais, ω , observa-se o mesmo comportamento, para todos os modelos, conforme as Figuras 25, 26, 27 e 28.

Associado aos modelos ZINB e ZINB-normal, tem-se o parâmetro de superdispersão, α , também com densidades marginais a posteriori suaves e unimodias, assim como para o parâmetro σ^2 , nos modelos ZIP-normal e ZINB-normal, conforme as Figuras 26, 28, 27 e 28, respectivamente.

Figura 25 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} e β_{11} sob o modelo ZIP

Figura 26 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω e α sob o modelo ZINB

Figura 27 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω e σ^2 sob o modelo ZIP-normal

Figura 28 - Densidades marginais a posteriori para β_{00} , β_{01} , β_{10} , β_{11} , ω , α e σ^2 sob o modelo ZINB-normal

Na Tabela 5, apresentam-se as medidas resumos a posteriori para os parâmetros dos modelos ZIP, ZINB, ZIP-normal e ZINB-normal, assim como, os contrastes criados para a comparação de tratamentos. Observa-se que os valores obtidos pela aproximação INLA são exatos, pois KLD=0. Se considerado o modelo ZIP-normal para descrever

o comportamento dos ataques epilépticos dos pacientes ao longo do tempo, conclui-se que não há diferença entre os tratamentos ao nível de 95% de credibilidade, pois os valores zero e um pertencem aos intervalos de credibilidades para os contrastes Δ_1 e Δ_2 , respectivamente. Além disso, verifica-se que para os demais modelos não existe diferença entre os tratamentos, observando-se os intervalos de credibilidade para os dois contrastes, exceto para o modelo ZIP, que não é conclusivo.

Sob os modelos apresentados por Rizzato (2011), seleciona-se para o ajuste, o modelo combinado, ou seja, o modelo Poisson-normal-gama. Para os modelos propostos neste trabalho, considera-se para o ajuste, o modelo ZIP-normal. Assim, existe a necessidade de se fazer a seleção entre os dois modelos. Nas Figuras 9, 10, 20 e 21 tem-se os histogramas e os gráficos quantil-quantil uniforme dos valores de PIT, para ambos ajustes. Nota-se que há um indício de que o modelo ZIP-normal apresenta melhor adequação aos dados, isto é, distribuição imposta para a variável resposta está correta, pois seu gráfico mais se aproxima do histograma de uma distribuição uniforme padrão. Além disso, baseado nas Figuras 24 e 11, há indícios de que o modelo ZIP-normal e o modelo combinado apresentem capacidade preditiva similares. Torna-se necessário indicar qual modelo melhor se ajusta ao dados.

Baseando-se nas medidas de CPO para os dois modelos, obteve-se o valor do logaritmo do PsFB, o qual equivale a 1,1300, de maneira que pelo critério de Jeffreys (1961), há fortes evidências à favor do modelo ZIP-normal sob o modelo combinado, devendo esse ser selecionado, como esperado pela interpretação gráfica.

Para verificar a eficiência do PsFB na seleção de modelos, geraram-se 1000 conjuntos de dados sob as mesmas especificações do modelo selecionado, isto é, sob as especificações do modelo ZIP-normal, as mesmas dimensões do conjunto de dados em estudo e considerando as estimativas da Tabela 5. Para cada conjunto de dados, ajustaram-se os dois modelos candidatos, ou seja, o modelo combinado e o modelo ZIP-normal e calculou-se o valor do PsFB, para cada conjunto. Observou-se que o PsFB indica a seleção do modelo ZIP-normal em 96,4% dos ajustes, o que torna a medida favorável para a seleção de modelos.

Apresentam-se na Tabela 6 os valores da média das estimativas pontuais a posteriori dos parâmetros para todos os ajustes, sob o modelo ZIP-normal. Assim, para o parâmetro β_{00} , o valor 0,8954 é a média dos valores a posteriori da média de β_{00} em cada um dos ajustes, e assim sucessivamente. Observa-se que as estimativas dos parâmetros são próximas dos valores reais (Tabela 5), e ainda, pelos intervalos de credibilidade de 95%,

nota-se que os valores fixados para a simulação pertencem a todos eles, de forma que, as estimativas são estatisticamente iguais aos valores fixados, como esperado, exceto para ω .

Tabela 5 - Resumos a posteriori para os parâmetros dos modelos ZIP, ZINB, ZIP-normal e ZINB-normal

Modelo ZIP					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	KLD
β_{00}	1,4117	0,0436	1,4119	(1,3255 ; 1,4967)	0
β_{01}	0,0034	0,0045	0,0034	(-0,0055 ; 0,0122)	0
β_{10}	1,7686	0,0401	1,7687	(1,6895 ; 1,8468)	0
β_{11}	-0,0209	0,0041	-0,0209	(-0,0290 ; -0,0128)	0
ω	0,4173	0,0113	0,4174	(0,3948 ; 0,4391)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0232	0,0061	-0,0232	(-0,0351 ; -0,0112)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	-3,3470	137,6889	-3,5983	(-68,8766 ; 67,6000)	-
Modelo ZINB					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	1,2630	0,1119	1,2620	(1,0461 ; 1,4853)	0
β_{01}	-0,0123	0,0111	-0,0123	(-0,0340 ; 0,0095)	0
β_{10}	1,4820	0,1097	1,4811	(1,2694 ; 1,6998)	0
β_{11}	-0,0351	0,0102	-0,0351	(-0,0551 ; -0,0151)	0
ω	0,0062	0,0057	0,0046	(0,0005 ; 0,0211)	0
α	0,5316	0,0262	0,5311	(0,4815 ; 0,5846)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0223	0,0151	-0,0223	(-0,0519 ; 0,0073)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	-4,2414	121,3955	2,3106	(-22,1227 ; 24,7185)	-
Modelo ZIP-normal					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	0,8957	0,1564	0,8959	(0,5879 ; 1,2022)	0
β_{01}	-0,0081	0,0046	-0,0081	(-0,0171 ; 0,0009)	0
β_{10}	0,7938	0,1591	0,7940	(0,4806 ; 1,1054)	0
β_{11}	-0,0111	0,0045	-0,0111	(-0,0198 ; -0,0023)	0
ω	0,1661	0,0029	0,1659	(0,1609 ; 0,1723)	0
σ^2	0,9707	0,0351	0,9739	(0,8948 ; 1,0297)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	-0,0029	0,0064	-0,0029	(-0,0156 ; 0,0096)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,8062	104,3577	1,3038	(-8,2400 ; 9,5346)	-
Modelo ZINB-normal					
Parâmetro	Média	D.P.	Mediana	Intervalo de Credibilidade	
β_{00}	0,9403	0,1741	0,9407	(0,5967 ; 1,2816)	0
β_{01}	-0,0238	0,0074	-0,0238	(-0,0383 ; -0,0094)	0
β_{10}	0,6956	0,1772	0,6962	(0,3453 ; 1,0423)	0
β_{11}	-0,0120	0,0071	-0,0120	(-0,0260 ; 0,0020)	0
ω	0,0465	0,0103	0,0457	(0,0286 ; 0,0690)	0
α	3,2825	0,3036	3,2624	(2,7413 ; 3,9330)	0
σ^2	1,0648	0,0870	1,0589	(0,9080 ; 1,2478)	0
$\Delta_1 = \beta_{11} - \beta_{01}$	0,0118	0,0103	0,0118	(-0,0083 ; 0,0320)	0
$\Delta_2 = \frac{\beta_{11}}{\beta_{01}}$	0,5709	0,4370	0,5030	(-0,0648 ; 1,5000)	-

Tabela 6 - Resumos a posteriori do estudo de simulação do modelo ZIP-normal

Parâmetro	Modelo ZIP-normal				Valor Real
	Média	D.P.	Intervalo de Credibilidade		
β_{00}	0,8954	0,1492	(0,6003 ; 1,1879)		0,8957
β_{01}	-0,0081	0,0022	(-0,0125 ; -0,0037)		-0,0081
β_{10}	0,7908	0,1519	(0,4906 ; 1,0882)		0,7938
β_{11}	-0,0111	0,0025	(-0,0159 ; -0,0063)		-0,0111
ω	0,2021	0,0111	(0,1808 ; 0,2243)		0,1661
σ^2	0,8916	0,1806	(0,6634 ; 1,2492)		0,9707

4 CONSIDERAÇÕES FINAIS

Em estudos com dados de contagens ao longo do tempo, ou seja, estudos longitudinais, pode ocorrer de além da correlação entre medidas tomadas no mesmo indivíduo, ocorra uma variabilidade além daquela descrita pelo modelo, caracterizando uma superdispersão. Além disso, esses conjuntos de dados podem apresentar inflação de zeros. Quando essas situações ocorrem, é necessário a formulação de modelos estatísticos que englobem tais características, conjuntamente.

Neste trabalho, considerando a existência de inflação de zeros, foram propostos quatro modelos para o ajuste de dados de contagens longitudinais com superdispersão: modelo ZIP, modelo ZINB, modelo ZIP-normal, modelo ZINB-normal. Todos eles consideram uma variância além daquela imposta pelo modelo de regressão de Poisson (média = variância). Os ajustes foram feitos utilizando a abordagem bayesiana por meio da aproximação determinística INLA, que torna a obtenção das medidas a posteriori algo rápido e eficiente.

O modelo selecionado foi o ZIP-normal, que além de acomodar a superdispersão, causada pelo excesso de zeros, também acomoda a correlação entre medidas tomadas no mesmo indivíduo. Sob o ajuste, não se verificou diferença entre os tratamentos, como em estudos anteriores.

A seleção do modelo ZIP-normal em comparação com o corrente na literatura, modelo combinado (RIZZATO, 2011), deu-se por meio da inspeção gráfica dos valores de CPO, PIT, além da verificação dos valores de KLD e do PsFB. Os valores de DIC não foram observados pois este é afetado pelo número de parâmetros a serem estimados e efeitos aleatórios a serem preditos. Sob estas considerações, propõe-se para a seleção de modelos, considerando a abordagem bayesiana, a verificação conjunta dos valores de CPO, PIT, KLD e do PsFB.

Pelo estudo de simulação, o uso do PsFB para a seleção de modelos mostrou-se satisfatório.

Todos os procedimentos foram implementados no programa estatístico R (R DEVELOPMENT CORE TEAM, 2015).

Como extensão desse trabalho será feito um estudo considerando modelos para dados de proporção com estrutura longitudinal e superdispersão, considerando a abor-

dagem INLA, no intuito de propor modelos que levem em conta essas duas características isoladamente e simultaneamente.

REFERÊNCIAS

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: INTERNATIONAL SYMPOSIUM OF INFORMATION THEORY, 2., Budapest, 1973. **Proceedings...** Budapest: Akademiai Kiadó, 1973. p. 267-281.
- BARBIN, D. **Componentes de variância - teoria e aplicações**. 2 ed. Piracicaba: FEALQ, 1998. 120p.
- BELLE, G.V; FISHER, L.D.; HEAGERTY, P.J.; LUMLEY, T. **Biostatistics - a methodology for health sciences**. 2nd ed. New Jersey: Wiley, 2004. 871p.
- BERGER, J.O. **Statistical decision theory and bayesian analysis**. 2nd ed. New York: Springer, 1985. 634p.
- BERNARDO, J.M.; SMITH, A.F.M. **Bayesian theory**. New York: Wiley, 2000. 610p.
- BOOTH, J.G.; CASELLA, G.; FRIEDL, H.; HOBERT, J.P. Negative binomial loglinear mixed models. **Statistical Modelling**, London, v3, p.179-191, 2003.
- BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. New York: Wiley, 1992. 608p.
- BRESLOW, N. E. Extra-Poisson Variation in Log-Linear Models. **Journal of the Royal Statistical Society Series C (Applied Statistics)**, London, v.33, p.38-44, 1984.
- BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, Boston, v.88, p.9-25, 1993.
- CARLIN, B.P.; LOUIS, T.A. **Bayes and empirical Bayes methods for data analysis**. London: Chapman & Hall, 1996, 399p.
- CLAYTON, D.; KALDOR, J. Empirical Bayes estimates of age - standardized relative risks for use in disease mapping. **Biometrics**, Washington, v.43, p.671-681, 1987.
- CORDEIRO, G.M.; CRIBARI-NETO, F.; AUBIN, E.Q.; FERRARI, S.L.P. Bartlett corrections for one-parameter exponential family models. **Journal of Statistical Computation and Simulation**, New York, v.53, p.211-231, 1995.
- CORDEIRO, G.M.; DEMÉTRIO, C.G.B. **Modelos lineares generalizados e extensões**. Piracicaba, 2009. 504p.
- CZADO, C.; GNEITING, T.; HELD, L. Predictive model assessment for count data. **Biometrics**, Washington, v.65, n.4, p.1254-1261, 2009.

DARMOIS, G. Sur les lois de probabilités à estimation exhaustive. **Comptes Rendus de l'Académie des Sciences - Series I - Mathematics**, Paris, v.200, p.1265–1266, 1935.

DAVIDIAN, M.; GILTINAN, D.M. **Nonlinear models for repeated measurement data**. New York: Chapman & Hall, 1995. 359p.

DEY, D.K.; GHOSH, S.K.; MALLICK, B.K. **Generalized linear models - a bayesian perspective**. New York: Marcel Dekker, 2000. 423 p.

DIGGLE, P.; HEAGERTY, P.; LIANG, K.-Y.; ZEGER, S. **Analysis of longitudinal data**. 2nd ed. Oxford: Oxford University Press, 2002. 253p.

DOBSON, A.J. **An introduction to generalized linear models**. London: Chapman and Hall, 2002. 221 p.

EHLERS, R.S. **Introdução à Inferência Bayesiana**, 2007. Disponível em: <<http://leg.ufpr.br/~ehlers/bayes>>. Acesso em: 10 out. 2008.

FABIO, L.C.; PAULA, G.A.; CASTRO, M. A Poisson mixed model with nonnormal random effect distribution. **Computational Statistics & Data Analysis**, New York, v.56, n. 6, p.1499-1510, 2012.

FAHRMEIR, L.; LANG, S. Bayesian inference for generalized additive mixed models based on markov random field priors. **Journal of the Royal Statistical Society Series C (Applied Statistics)**, London, v.50, n. 2, p.201-220, 2001.

FAUGHT, E.; WILDER, B.J.; RAMSAY, R.E.; KRAMER, L.D.; PLEDGER, G.W.; KARIM, R.M. Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. **Neurology**, New York, v.46, p.1684-1690, 1996.

FONG, Y.; RUE, H.; WAKEFIELD, J. Bayesian inference for generalized linear mixed models. **Biostatistics**, Oxford, v.11, p.397-412, 2010.

GEISSER, S.; EDDY, W. A predictive approach to model selection. **Journal of the American Statistical Association**, Boston, v.74, n.365, p.153-160, 1979.

GELFAND, A.E. Model determination using sampling-based methods. In: GILKS, W.; S. RICHARDSON, S.; SPIEGELHALTER, D. (Eds.), **Markov chain Monte Carlo in practice**. Florida: Chapman & Hall/CRC, 1996. chap. 9, p. 145-161.

GELMAN, A.; CARLIN, J.B.; STERN, H.S.; RUBIN, D.B. **Bayesian data analysis**. Florida: Chapman & Hall/CRC, 2004. 668p.

GNEITING, T.; RAFTERY, A.E. Strictly proper scoring rules, prediction, and estimation. **Journal of the American Statistical Association**, Boston, v.102, n.477, p.359-378, 2007.

- GREENE, H.G. **Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models**. 2008. Disponível em: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293115>. Acesso em: 03 jan. 2015.
- GUERRA, R. **Laplace approximation**, 2008. Disponível em: <http://www.ece.rice.edu/vc3/elec633/graphical_models_notes_091108.pdf>. Acesso em: 28 ago. 2012.
- HALL, B. **Bayesian inference**, 2012. Disponível em: <<http://www.icesi.edu.co/CRAN/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf>>. Acesso em: 03 jan. 2015.
- HALL, D.B. Zero-inflated Poisson and binomial regression with random effects: a case study. **Biometrics**, Washington, v.56, p.1030-1039, 2000.
- HELMS, R.W. Intentionally incomplete longitudinal designs: I. Methodology and Comparison of some full span designs. **Statistics in Medicine**, Chichester, v.11, p.1889-1913, 1992.
- HINDE, J., DEMÉTRIO, C.G.B. Overdispersion: models and estimation. **Computational Statistics and Data Analysis**, New York, v.27, p.151-170, 1998.
- JEFFREYS, H. **Theory of Probability**. 3rd ed. Oxford: Oxford University Press, 1961. 459p.
- KOOPMAN, B. On distribution admitting a sufficient statistic. **Transactions of the American Mathematical Society**, Providence, v.39, n.3, p.399-409, 1936.
- KULLBACK, S.; LEIBLER, R.A. On information and sufficiency. **The Annals of Mathematical Statistics**, Cornell, v.22, n.1, p.79-86, 1951.
- LAMBERT, D. Zero-inflated Poisson regression model, with an application to defects in manufacturing. **Technometrics**, Washington, v.34, n.1, p.1-14, 1992.
- LAWLESS, J. Negative binomial and mixed poisson regression. **The Canadian Journal of Statistics**, Ottawa, v.15, p.209-225, 1987.
- LIMA, C.G. **Análise de dados longitudinais provenientes de experimentos em blocos casualizados**. 1996. 126p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1996.
- LEE, Y.; NELDER, J.A. Hierarchical generalized linear models. **Journal of the Royal Statistical Society Series B**, London, v.58, p.619-678, 1996.
- LEANDRO, R. **Introdução à estatística bayesiana**. REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46. **Minicurso**. Piracicaba: ESALQ, 2001. 51p.

LUNN, D.; SPIEGELHALTER, D.; THOMAS, A.; BEST, N. The BUGS project: evolution, critique, and future directions. **Statistics in Medicine**, New York, v.28, p.3049-3067, 2009.

LUNN, D.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS - a bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, London, v.10, p.325-337, 2000.

MANTON, K.G.; WOODBURY, M.A.; RIGGAN, W.B.; CREASON, J.P.; PELLOM, A.C. Empirical Bayes Procedures for Stabilizing Maps of U.S. cancer mortality rates. **Journal of the American Statistical Association**, Boston, v.84, p.637-650, 1989.

MAGALHÃES, M.N. **Probabilidade e variáveis aleatórias**. 3 ed. São Paulo: edUSP, 2011. 411p.

MARTINO, S.; RUE, H. Approximate bayesian inference for hierarchical gaussian Markov random fields models. **Journal of Statistical Planning and Inference**, Amsterdam, v.137, p. 3177-3192, 2007.

MARTINS, T.G.; SIMPSON, D.; LINDGREN, F.; RUE, H. Bayesian computing with INLA: new features. **Computational Statistics and Data Analysis**, Amsterdam, v.67, p.68-93, 2013.

MCCULLAGH, P.; NELDER, J. **Generalized linear models**. 2nd ed. London: Chapman and Hall, 1989. 511p.

MCCULLOCH, C. E.; SEARLE, S. R.; NEUHAUS, J. M. **Generalized, linear, and mixed models**. 2nd ed. New York: John Wiley and Sons, 2008. 384p.

MENG, X. Posterior predictive p-values. **Annals of Statistics**, Hayward, v.22, n.3, p.1142-1160, 1994.

MOGHIMBEIGI, G.; ESHRAGHIAN, M.R; MOHAMMAD, K.; MCARDLE, B. Multilevel zero-inflated negative binomial regression modeling for overdispersed count data with extra zeros. **Journal of Applied Statistics**, Abingdon, v.35, n.10, p.1193-1202, 2008.

MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B.; An extended random-effects approach to modeling repeated, overdispersion count data. **Lifetime Data Analysis**, Hingham, v.13, n.4, p.513-531, 2007.

MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B.; VIEIRA, A.M.C. A family of generalized linear models for repeated measures with normal and conjugate random effects. **Statistical Science**, Hayward, v.25, n.3, p.325-347, 2010.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society, Series A (General)**, London, v.135, n.3, p.370-384, 1972.

NELSON, K.P.; LIPSITZ, S.T.; FITZMAURICE, G.M.; IBRAHIM, J.; PARZEN, M. STRAWDERMAN, R. Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. **Journal of Computational and Graphical Statistics**, Alexandria, v.15, p.39-57, 2006.

NTZOUFRAS, I. **Bayesian modeling using WinBUGS**. New Jersey: Wiley, 2009. 492p.

O'HAGAN, A. Kendall's Advanced Theory of Statistics, Vol. 2b: Bayesian inference. by A. O'Hagan Review by: I. R. Dunsmore. **Biometrics**, Washington, v.52, p.1525-1526, 1996.

PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

PETTIT, L.I.; YOUNG, K.D.S. Measuring the effect of observations on bayes factors. **Biometrika**, London, v.77, n.3, p.455-466, 1990.

PITMAN, E.; WISHART, J. Sufficient statistics and intrinsic accuracy. **Mathematical Proceedings of the Cambridge Philosophical Society**, Cambridge, v.32, n.4, p.567-579, 1936.

PLUMMER, M. Penalized loss functions for Bayesian model comparison. **Biostatistics**, Oxford, v.9, n.3, p.523-539, 2008.

POSTON, D.L.; MCKIBBEN, S.L. Using zero-inflated count regression models to estimated the fertility of U.S. women. **Journal of Modern Applied Statistical Methods**, Detroit, v.2, n.2, p.371-379, 2003.

PRATES, M.O. **Link specification and spatial dependence for generalized linear mixed models**. 2011. 71p. These (Doctor of Philosophy) - University of Connecticut, Connecticut, 2011.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.R-project.org/>>. Acesso em: 25 jan. 2015.

RAO, J.N.K. **Small Area Estimation**. 2nd ed. New York: Wiley Series in Survey Methodology, 2003. 344p.

RASMUSSEN, C.E.; WILLIAMS, K.I. **Gaussian processes for machine learning**. Massachusetts: MIT Press, 2006. 248 p.

RIBEIRO, M.C.R. **Modelação de dados espaço-temporais em segurança rodoviária**. 2012. 124p. Tese (Doutorado em Estatística e Investigação Operacional) - Faculdade de Ciências, Universidade de Lisboa, Lisboa, 2012.

RIDOUT, M.; DEMÉTRIO, C.G.B., HINDE, J. Models for count data with many zeros. In: INTERNATIONAL BIOMETRIC CONFERENCE, 19., 1998, Cape Town. **Proceedings...**, Cape Town: International Biometric Society, 1998. 1 CD-ROM.

RIZZATO, F.B. **Modelos para análise de dados discretos longitudinais com superdispersão**. 2011. 143p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 03 mar. 2015.

ROBERT, C.P; CASELLA, G. **Introducing Monte Carlo methods with R**. New York: Springer, 2010. 283 p.

ROCHA, E.B. **Modelos não-lineares para dados longitudinais provenientes de experimentos em blocos casualizados abordagem bayesiana**. 2010. 205p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.

ROSENBLATT, M. Remarks on multivariate transformation. **The Annals of Mathematical Statistics**, Cornell, v.23, n.3, p.470-472, 1952.

ROSSI, R.M. **Introdução aos métodos bayesianos na análise de dados zootécnicos com o uso do WinBUGS e R**. Maringá: EdUEM, 2011. 191p.

RUE, H.; HELD, L. **Gaussian markov random fields**. New York: Chapman and Hall, 2005. 255p.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). **Journal of the Royal Statistical Society Series B**, London, v.71, p.319-392, 2009.

SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P.; LINDE, A.V.D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society Series B**, London, v.64, n.4, p.583-639, 2000.

STIRATELLI, R.; LAIRD, N.; WARE, J. Random effects models for serial observations with binary responses. **Biometrics**, Washington, v.40, p.961-971, 1984.

TIERNEY, L.; KADANE, J.B. Accurate approximations for posterior moments and marginal densities. **Journal of the American Statistical Association**, Boston, v.81, n.393: p.82-86, 1986.

URBANO, M.R. **Modelagem de mortalidade natural e superdispersão em dados entomológicos**. 2012. 117p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2012.

VANGENEUGDEN, T.; MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B. Marginal correlation from an extended random-effects model for repeated, overdispersed counts. **Journal of Applied Statistics**, Abingdon, v.38, n.2, p.215-232, 2011.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer, 2000. 568p.

----- **Models for discrete longitudinal data**. New York: Springer, 2005. 683p.

VIEIRA, A.M.C.; HINDE, J.P.; DEMÉTRIO, C.G.B. Zero-inflated proportion data models applied to a biological control assay. **Journal of Applied Statistics**, Abingdon, v.27, n.3, p.373-389, 2000.

WILLIAMS, D.A. Extra-Binomial Variation in Logistic Linear Models. **Journal of Applied Statistics**, Abingdon, v.31, p.144-148, 1982.

ZEGER, S.L.; LIANG, K.Y.; ALBERT, P.S. Models for Longitudinal Data: A Generalized Estimating Equation Approach. **Biometrics**, Washington, v.44, p.1049-1060, 1988.

APÊNDICE

APÊNDICE A

A.1 Distribuição de Poisson

Considere que Y seja uma variável aleatória com distribuição de Poisson de parâmetro λ , $Y \sim \text{Poisson}(\lambda)$, em que $\lambda > 0$. A função de probabilidade de Y pode ser parametrizada da seguinte forma:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, \dots,$$

em que y é uma observação de uma contagem em geral feita em um intervalo de tempo.

O valor esperado ou esperança matemática da variável aleatória Y é dado por,

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y P(Y = y) = \sum_{y=0}^{\infty} y \frac{e^{-\lambda} \lambda^y}{y!} \\ &= e^{-\lambda} \sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} = \lambda e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} = \lambda, \end{aligned}$$

uma vez que $\sum_{y=k}^{\infty} \frac{\lambda^{y-k}}{(y-k)!} = e^{\lambda}$, sendo k uma constante.

Para a variância da variável aleatória Y , tem-se

$$\text{Var}(Y) = E(Y^2) - E(Y)^2 = E(Y^2 - Y) + E(Y) - E(Y)^2.$$

Uma vez que,

$$E(Y^2 - Y) = \sum_{y=0}^{\infty} (y^2 - y) P(Y = y) = \lambda^2 e^{-\lambda} \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} = \lambda^2,$$

tem-se,

$$\text{Var}(Y) = \lambda.$$

A.2 Distribuição de Poisson inflacionada de zeros

Diz-se que a variável aleatória Y tem distribuição ZIP com parâmetro λ , probabilidade ω de zeros estruturais e $(1 - \omega)$ de zeros amostrais se sua distribuição de probabilidade pode ser parametrizada da seguinte forma,

$$P(Y = y) = \begin{cases} \omega + (1 - \omega) \exp(-\lambda), & \text{se } y = 0 \\ (1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y!}, & \text{se } y > 0, \end{cases} \quad (51)$$

isto é, a modelagem da resposta nula leva em consideração a probabilidade da resposta ser igual a zero para o modelo Poisson. Para as contagens não nulas, tem-se a probabilidade de valores diferentes de zero da distribuição de Poisson.

O valor esperado da variável aleatória Y para $y = 0$ e dado por

$$E(Y) = yP(Y = y) = y[\omega + (1 - \omega) \exp(-\lambda)] = 0, \quad (52)$$

e para $y > 0$,

$$\begin{aligned} E(Y) &= \sum_{y=1}^{\infty} yP(Y = y) = \sum_{y=1}^{\infty} y \left[(1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y!} \right] \\ &= \sum_{y=1}^{\infty} y \left[(1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y(y-1)!} \right] = \sum_{y=1}^{\infty} (1 - \omega) \frac{\lambda \lambda^{y-1} \exp(-\lambda)}{(y-1)!} \Big|_{y=z+1}^{z=y-1} \\ &= (1 - \omega) \lambda \sum_{z=0}^{\infty} \frac{\lambda^z \exp(-\lambda)}{z!} = (1 - \omega) \lambda, \end{aligned} \quad (53)$$

pois, $\sum_{z=0}^{\infty} \frac{\lambda^z \exp(-\lambda)}{z!} = 1$. Portanto, somando-se (52) e (53)

$$E(Y) = 0 + (1 - \omega) \lambda = (1 - \omega) \lambda = \mu.$$

Para a variância, de maneira análoga à esperança, considerando-se inicialmente $y = 0$, tem-se

$$E(Y^2) = y^2[\omega + (1 - \omega) \exp(-\lambda)] = 0 \Rightarrow \text{Var}(Y) = 0. \quad (54)$$

Quando $y > 0$, tem-se

$$\begin{aligned} E(Y^2) &= \sum_{y=1}^{\infty} y^2 (1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y!} = \sum_{y=1}^{\infty} y^2 (1 - \omega) \frac{\lambda^y \exp(-\lambda)}{y(y-1)!} \\ &= \sum_{y=1}^{\infty} y (1 - \omega) \frac{\lambda^y \exp(-\lambda)}{(y-1)!} \Big|_{y=z+1}^{z=y-1} = (1 - \omega) \sum_{z=0}^{\infty} (z+1) \frac{\lambda^z \lambda \exp(-\lambda)}{z!} \\ &= (1 - \omega) \lambda \left[\sum_{z=0}^{\infty} z \frac{\lambda^z \exp(-\lambda)}{z!} + \sum_{z=0}^{\infty} \frac{\lambda^z \exp(-\lambda)}{z!} \right] \\ &= (1 - \omega) \lambda \sum_{z=0}^{\infty} z \frac{\lambda^z \exp(-\lambda)}{z!} + \lambda \sum_{z=0}^{\infty} \frac{(1 - \omega) \lambda^z \exp(-\lambda)}{z!} = (1 - \omega) \lambda [1 + \lambda], \end{aligned} \quad (55)$$

uma vez que $\sum_{z=0}^{\infty} z \frac{\lambda^z \exp(-\lambda)}{z!} = \lambda$ e $\sum_{z=0}^{\infty} \frac{(1-\omega)\lambda^z \exp(-\lambda)}{z!} = 1$.

De (52), (53), (54) e (55), tem-se que a variância da variável aleatória Y é dada por,

$$\begin{aligned}
 Var(Y) &= E(Y^2) - E(Y)^2 = (1-\omega)\lambda[1+\lambda] - (1-\omega)^2\lambda^2 \\
 &= (1-\omega)\lambda[1+\lambda - (1-\omega)\lambda] = (1-\omega)\lambda[1+\omega\lambda] \\
 &= (1-\omega)\lambda + \omega(1-\omega)\lambda^2 = \mu + \omega \frac{(1-\omega)^2}{(1-\omega)} \lambda^2 \\
 &= \mu + \left(\frac{\omega}{1-\omega} \right) (1-\omega)^2 \lambda^2 = \mu + \left(\frac{\omega}{1-\omega} \right) \mu^2.
 \end{aligned}$$

A.3 Distribuição binomial negativa

Considere que a variável discreta Y condicionalmente à variável aleatória Z , $Y|Z$, representa um processo de contagem, e que $Y|Z = z \sim \text{Poisson}(z)$,

$$P(Y = y|Z = z) = \frac{z^y \exp(-z)}{y!}, \quad y = 0, 1, \dots,$$

e no segundo estágio do modelo que $Z \sim \Gamma(k, \lambda)$, $z > 0$, k fixo, ou seja, que

$$f_Z(z) = \frac{\left(\frac{\lambda}{k}\right)^\lambda}{\Gamma(\lambda)} z^{\lambda-1} \exp\left(-\frac{z\lambda}{k}\right). \quad (56)$$

Dessa forma, tem-se que a distribuição da variável aleatória Y é dada por

$$\begin{aligned} P(Y = y) &= \int_0^{+\infty} P(Y = y|Z = z) f_Z(z) dz = \int_0^{+\infty} \frac{z^y \exp(-z)}{y!} \frac{\left(\frac{\lambda}{k}\right)^\lambda}{\Gamma(\lambda)} z^{\lambda-1} \exp\left(-\frac{z\lambda}{k}\right) \\ &= \frac{\left(\frac{\lambda}{k}\right)^\lambda}{y! \Gamma(\lambda)} \int_0^{+\infty} z^{y+\lambda-1} \exp\left\{-z\left(1 + \frac{\lambda}{k}\right)\right\} dz \quad \left| \begin{array}{l} t=z\left(1 + \frac{\lambda}{k}\right) \Rightarrow dt = \left(1 + \frac{\lambda}{k}\right) dz \\ z \rightarrow 0 \Rightarrow t \rightarrow 0; \quad z \rightarrow +\infty \Rightarrow t \rightarrow +\infty \end{array} \right. \\ &= \frac{\left(\frac{\lambda}{k}\right)^\lambda}{y! \Gamma(\lambda)} \int_0^{+\infty} \left(\frac{t}{\left(1 + \frac{\lambda}{k}\right)}\right)^{y+\lambda-1} \exp(-t) \frac{dt}{\left(1 + \frac{\lambda}{k}\right)} \\ &= \frac{\left(\frac{\lambda}{k}\right)^\lambda}{y! \Gamma(\lambda)} \int_0^{+\infty} t^{y+\lambda-1} \left(\frac{1}{1 + \frac{\lambda}{k}}\right)^{y+\lambda} \left(1 + \frac{\lambda}{k}\right) \exp(-t) \frac{dt}{\left(1 + \frac{\lambda}{k}\right)} \\ &= \frac{\left(\frac{\lambda}{k}\right)^\lambda}{y! \Gamma(\lambda)} \left(\frac{k}{k + \lambda}\right)^{y+\lambda} \int_0^{+\infty} t^{y+\lambda-1} \exp(-t) dt = \frac{\left(\frac{\lambda}{k}\right)^\lambda}{y! \Gamma(\lambda)} \left(\frac{k}{k + \lambda}\right)^{y+\lambda} \Gamma(y + \lambda) \\ &= \frac{1}{y! \Gamma(\lambda)} \frac{\lambda^\lambda}{k^\lambda} \frac{k^\lambda k^y}{(k + y)^{y+\lambda}} \Gamma(y + \lambda) = \frac{\Gamma(y + \lambda)}{y! \Gamma(\lambda)} \frac{\lambda^\lambda k^y}{(k + y)^{y+\lambda}}, \end{aligned}$$

de maneira que Y tem distribuição binomial negativa com parâmetros k e λ dada por

$$P(Y = y) = \frac{\Gamma(y + \lambda)}{y! \Gamma(\lambda)} \frac{\lambda^\lambda k^y}{(k + y)^{y+\lambda}}.$$

A esperança da variável aleatória Y é dada por

$$E(Y) = E[E(Y|Z)] = E[Z] = \frac{k}{\lambda} = \mu$$

e sua variância,

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y|Z)] + \text{Var}[E(Y|Z)] \\ &= E[Z] + \text{Var}[Z] = \frac{k}{\lambda} + \frac{k}{\lambda^2} \\ &= \frac{k}{\lambda} + \frac{\left(\frac{k}{\lambda}\right)^2}{k} = \mu + \frac{\mu^2}{k}. \end{aligned}$$

A.4 Distribuição binomial negativa inflacionada de zeros

Considere que Y é uma variável aleatória na forma de contagem com excessos de zeros, seu comportamento distribucional pode ser descrito pela distribuição binomial inflacionadas de zeros, com função de probabilidade expressa por

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)(1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}}, & \text{se } y = 0 \\ (1 - \omega) \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{-c}}{\alpha}\right)^{-y}, & \text{se } y > 0, \end{cases}$$

em que λ e α são parâmetros, sendo α o parâmetro de dispersão, ω é a probabilidade de zeros estruturais e c é uma constante arbitrária. Para o valor esperado, portanto, considerando-se $y = 0$, tem-se

$$E(Y) = y \left[\omega + (1 - \omega)(1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \right] = 0.$$

Quando $y > 0$,

$$\begin{aligned} E(Y) &= \sum_{i=1}^{\infty} (1 - \omega) y \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{1-c}}{\alpha}\right)^{-y} \\ &= (1 - \omega) \sum_{i=1}^{\infty} y \frac{\Gamma\left(y + \frac{\lambda^{1-c}}{\alpha}\right)}{y(y-1)! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(1 + \frac{\lambda^{1-c}}{\alpha}\right)^{-y}, \end{aligned}$$

considerando $z = y - 1$ e $y = z + 1$,

$$\begin{aligned}
E(Y) &= (1 - \omega) \sum_{i=0}^{\infty} \frac{\Gamma\left(z + \frac{\lambda^{1-c} + 1}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^{z+1} \\
&= (1 - \omega) \sum_{i=0}^{\infty} \frac{\left(z + \frac{\lambda^{1-c}}{\alpha}\right) \Gamma\left(z + \frac{\lambda^{1-c}}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^z \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) \\
&= (1 - \omega) \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left[\sum_{i=0}^{\infty} \frac{z \Gamma\left(z + \frac{\lambda^{1-c}}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^z + \right. \\
&\quad \left. \sum_{i=0}^{\infty} \frac{\lambda^{1-c}}{\alpha} \frac{\Gamma\left(z + \frac{\lambda^{1-c}}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^z \right] \\
&= (1 - \omega) \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) \left[\sum_{i=0}^{\infty} \frac{z \Gamma\left(z + \frac{\lambda^{1-c}}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^z + \right. \\
&\quad \left. \frac{\lambda^{1-c}}{\alpha} \sum_{i=0}^{\infty} \frac{\Gamma\left(z + \frac{\lambda^{1-c}}{\alpha}\right)}{z! \Gamma\left(\frac{\lambda^{1-c}}{\alpha}\right)} (1 + \alpha\lambda^c)^{-\frac{\lambda^{1-c}}{\alpha}} \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right)^z \right] \\
&= (1 - \omega) \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) \left(\lambda + \frac{\lambda^{1-c}}{\alpha}\right) = (1 - \omega) \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) \left(\lambda + \frac{\lambda}{\alpha\lambda^c}\right) \\
&= (1 - \omega) \left(\frac{\alpha\lambda^c}{\alpha\lambda^c + \lambda}\right) \left(\frac{\lambda\alpha\lambda^c + \lambda}{\alpha\lambda^c}\right) = (1 - \omega)\lambda.
\end{aligned}$$

APÊNDICE B

B.1 Rotinas programa R para a exemplificação da Figura 1

```
rm(list=ls())

#definição de valores
mu    <- 0
kappa <- 0.001
y     <- 3
eta   <- seq(-2,3,0.01)
eta0  <- c(0,0.5,1,1.5)

#distribuição a posteriori exata
exata <- function(mu,kappa,y,eta){
  pi.y <- exp( -(kappa/2)*((eta-mu)^2)+y*eta-exp(eta) )
  pi.y <- pi.y/sum(pi.y)
  return(pi.y)
}

#aproximação gaussiana para a distribuição a posteriori
aprox <- function(mu,kappa,y,eta0,eta){
  a <- log(exata(mu,kappa,y,eta0))
  b <- kappa*mu + y - exp(eta0) + exp(eta0)*eta0
  c <- exp(eta0)+kappa
  pi.y <- exp( a+ b*eta - 0.5*c*(eta^2) )
  pi.y <- pi.y/sum(pi.y)
  return(pi.y)
}

#análise gráfica
par(mfrow=c(2,2))

posterior <- exata(mu=mu,kappa=kappa,y=y,eta=eta)
post.apro <- aprox(mu=mu,kappa=kappa,y=y,eta=eta,eta0=eta0[1])
plot(eta, posterior,type='l',ylim=c(0,0.01),lwd=3,
main=expression(paste(eta[0],"=",0)),
ylab="posteriori",xlab=expression(paste(eta)))
lines(eta,post.apro,type='l',lwd=2,lty=2)
x0<-0;y0<-0
points(x0,y0,pch=16,lwd=2)
legend(x=-1,y=0.012,lty=c(1,2),box.lty=0,ncol=2,bty="n",lwd=c(2,2),
      c(expression(paste(pi,"(",eta,"|","y",")")),
        expression(paste(tilde(pi),"(",eta,"|","y",")"))))

posterior <- exata(mu=mu,kappa=kappa,y=y,eta=eta)
post.apro <- aprox(mu=mu,kappa=kappa,y=y,eta=eta,eta0=eta0[2])
plot(eta, posterior,type='l',ylim=c(0,0.01),lwd=3,
main=expression(paste(eta[0],"=",0,5)),
ylab="posteriori",xlab=expression(paste(eta)))
lines(eta,post.apro,type='l',lwd=3,lty=2)
x0<-0.5;y0<-0
```

```

points(x0,y0,pch=16,lwd=3)
legend(x=-1,y=0.012,lty=c(1,2),box.lty=0,ncol=2,bty="n",lwd=c(2,2),
      c(expression(paste(pi,"(",eta,"|","y",")")),
        expression(paste(tilde(pi),"(",eta,"|","y",")"))))

```

```

posterior <- exata(mu=mu,kappa=kappa,y=y,eta=eta)
post.apro <- aprox(mu=mu,kappa=kappa,y=y,eta=eta,eta0=eta0[3])
plot(eta, posterior,type='l',ylim=c(0,0.01),lwd=3,
     main=expression(paste(eta[0],"=",1)),
     ylab="posteriori",xlab=expression(paste(eta)))
lines(eta,post.apro,type='l',lwd=3,lty=2)
x0<-1;y0<-0
points(x0,y0,pch=16,lwd=3)
legend(x=-1,y=0.012,lty=c(1,2),box.lty=0,ncol=2,bty="n",lwd=c(2,2),
      c(expression(paste(pi,"(",eta,"|","y",")")),
        expression(paste(tilde(pi),"(",eta,"|","y",")"))))

```

```

posterior <- exata(mu=mu,kappa=kappa,y=y,eta=eta)
post.apro <- aprox(mu=mu,kappa=kappa,y=y,eta=eta,eta0=eta0[4])
plot(eta, posterior,type='l',ylim=c(0,0.01),lwd=3,
     main=expression(paste(eta[0],"=",1,5)),
     ylab="posteriori",xlab=expression(paste(eta)))
lines(eta,post.apro,type='l',lwd=3,lty=2)
x0<-1.5;y0<-0
points(x0,y0,pch=16,lwd=3)
legend(x=-1,y=0.012,lty=c(1,2),box.lty=0,ncol=2,bty="n",lwd=c(2,2),
      c(expression(paste(pi,"(",eta,"|","y",")")),
        expression(paste(tilde(pi),"(",eta,"|","y",")"))))

```

B.2 Rotinas programa R para entrada e análise exploratória

```
rm(list=ls(all=TRUE))

#-----#
#                                     #
#  Entrada de dados                  #
#                                     #
#-----#

#número de ataques epilépticos
y1<-c( contagens )
#variável indicando o tratamento placebo ou a nova droga
trt1<-c( tratamentos )
#tempos, variando de 1 à 27
t1<-c( tempos )

trt1<-as.factor(trt1);is.factor(trt1)
length(y1)
length(t1)
length(trt1)
trt11<-rep(trt1,each=27);is.factor(trt11)
indv<-sort(rep(1:89,27))
tg1<-c(rep(0,45),rep(1,44))
G1<-rep(tg1,each=27)
tg2<-c(rep(1,45),rep(0,44))
G2<-rep(tg2,each=27)
z1<-G1*t1
z2<-G2*t1

#-----#
#                                     #
#  Análise Exploratória             #
#                                     #
#-----#

#obtenção do gráfico de perfis dos indivíduos, por tratamento
teste2 <- data.frame( indv=rep(1:89,each=27),
                      year=t1,
                      grp=as.factor( c(rep(0,27*45),rep(1,27*44) ) ),
                      resp=y1
)

library(nlme)
perfil1<-groupedData(resp ~ year|indv,order.groups=T,
data=teste2,labels=list(x="tempo", y="resposta"))
plot(perfil1,outer=~grp,key=F)
```

B.3 Rotinas programa R para estruturação dos dados e ajuste do modelo Poisson

```
#-----#
#                                     #
#  Análise Bayesiana usando INLA    #
#                                     #
#-----#

#obtenção do pacote INLA e todas as suas dependências
source("http://www.math.ntnu.no/inla/givemeINLA.R")
install.packages("pixmap",dep=T)
install.packages("sp",dep=T)

#análise bayesiana usando INLA
library(INLA)
dat<-data.frame(y1,trt11,t1,indv)
formula1<-y1~trt11/t1-1
formula1.1<-y1~-1+G1+G2+z1+z2

#-----#
#Modelo 1 - Modelo Poisson
#-----#

#ajuste do modelo
mod1<-inla(formula1,data=dat,family="poisson",control.compute =
list(dic = TRUE,cpo=TRUE))
names(mod1)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod1$marginals.fixed)
par(mfrow=c(2,2))
plot(mod1$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod1$marginals.fixed$trt110:t1,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod1$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod1$marginals.fixed$trt111:t1,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod1$cpo$pit,xlab="Índice",ylab="PIT",main="(a)",ylim=c(0,1))
hist(mod1$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)",ylim=c(0,500))
dev.off()

#obtenção dos gráficos dos valores de CPO
hist(mod1$cpo$cpo,xlab="CPO",ylab="Frequência",main="")
```



```

#medidas resumo a posteriori do modelo
summary(mod1)

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod1.1<-inla(formula1.1,data=dat,family="poisson",control.compute = list(dic = TRUE))
summary(mod1.1)
ds1<-inla.make.lincomb(z1=1, z2=-1)
names(ds1) = "ds1"
ds1.mod<-inla(formula1.1, "poisson",
              data = dat,
              lincomb = ds1)
summary(ds1.mod)

#obtenção das medidas resumo a posteriori do contraste para o quociente
mod1.2<-inla(formula1.1,data=dat,family="poisson",control.compute =
list(config=TRUE))
samples1<- inla.posterior.sample(1,mod1.2)
str(samples1)
dim(samples1[[1]]$latent)
names(samples1[[1]]$latent[2403:2407,1])
#z1 está na posição 2406 e z2 na posição 2407
nsim<-1000
z1.p<-2406
z2.p<-2407
samples1<- inla.posterior.sample(nsim,mod1.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples1[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples1[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

```

B.4 Rotinas programa R para ajuste do modelo Poisson-gama

```
#-----
#Modelo 2 - Modelo Poisson-Gamma
#-----

#ajuste do modelo
mod2<-inla(formula1,data=dat,family="nbinomial",control.compute =
list(dic = TRUE,cpo=TRUE))
names(mod2)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod2$marginals.fixed)
par(mfrow=c(2,2))
plot(mod2$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod2$marginals.fixed$trt110:t1,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod2$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod2$marginals.fixed$trt111:t1,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod2$cpo$pit,xlab="Índice",ylab="PIT",main="(a)",ylim=c(0,1))
hist(mod2$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)",ylim=c(0,500))
dev.off()

#obtenção do gráfico dos valores de CPO
hist(mod2$cpo$cpo,xlab="CPO",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod2)

#obtenção das medidas resumo a posteriori e do gráfico para os
#parâmetros associados ao efeito aleatório gama
names(mod2$marginals.hyperpar)
over1<-mod2$marginals.hyperpar[[1]];over1
D1<-inla.tmarginal(function(x) 1/x, over1)
inla.qmarginal(0.025, D1);inla.qmarginal(0.975, D1)
inla.qmarginal(0.5, D1);mean(D1[,1])
sd(D1[,1]);median(D1[,1])

par(mfrow=c(2,1))
plot(mod2$marginals.hyperpar[[1]],xlab=expression(paste(alpha)),
ylab="Densidade",type="l",main="")
plot(inla.smarginal(D1), type="l",xlab=expression(paste(beta)),
ylab="Densidade",main="")
```

```
dev.off()
```

```
#obtenção das medidas resumo a posteriori do contraste para a diferença
mod2.1<-inla(formula1.1,data=dat,family="nbinomial",control.compute =
  list(dic = TRUE))
summary(mod2.1)
ds2<-inla.make.lincomb(z1=1, z2=-1)
names(ds2) = "ds2"
ds2.mod<-inla(formula1.1, "nbinomial",
  data = dat,
  lincomb = ds2)
summary(ds2.mod)
```

```
#obtenção das medidas resumo a posteriori do contraste para o quociente
mod2.2<-inla(formula1.1,data=dat,family="nbinomial",control.compute =
  list(config = TRUE))
samples2<- inla.posterior.sample(1,mod2.2)
str(samples2)
dim(samples2[[1]]$latent)
names(samples2[[1]]$latent[2403:2407,1])
#z1 está na posição 2406 e z2 na posição 2407
nsim<-1000
z1.p<-2406
z2.p<-2407
samples2<- inla.posterior.sample(nsim,mod2.2)
```

```
rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples2[[i]]$latent[z1.p,1]
}
head(rs1)
```

```
rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples2[[i]]$latent[z2.p,1]
}
head(rs2)
```

```
rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic
```

B.5 Rotinas programa R para ajuste do modelo Poisson-normal

```
#-----
#Modelo 3 - Modelo Poisson-Normal
#-----

#ajuste do modelo
formula2<-y1~trt11/t1-1+f(indv,model="iid")
formula2.1<-y1~-1+G1+G2+z1+z2+f(indv,model="iid")
mod3<-inla(formula2,data=dat,family="poisson",control.compute =
list(dic = TRUE,cpo=TRUE))
names(mod3)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod3$marginals.fixed)
par(mfrow=c(2,2))
plot(mod3$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod3$marginals.fixed$trt110:t1,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod3$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod3$marginals.fixed$trt111:t1,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

#obtenção das medidas resumo a posteriori e do gráfico para o
#parâmetro associado ao efeito aleatório normal
names(mod3$marginals.hyperpar)
tau1<-mod3$marginals.hyperpar[[1]];tau1
stdv1<-inla.tmarginal(function(x) 1/sqrt(x), tau1)
inla.qmarginal(0.025, stdv1);inla.qmarginal(0.975, stdv1)
inla.qmarginal(0.5, stdv1);mean(stdv1[,1])
sd(stdv1[,1]);median(stdv1[,1])
plot(inla.smarginal(stdv1), type="l",xlab=expression(paste(sigma^2)),ylab="Densidade")

#obtenção dos gráficos dos valores PIT
par(mfrow=c(2,1))
plot(mod3$cpo$pit,xlab="Índice",ylab="PIT",main="(a)",ylim=c(0,1))
hist(mod3$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)",ylim=c(0,500))
dev.off()

#obtenção do gráfico dos valores CPO
hist(mod3$cpo$cpo,xlab="Probabilidade",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod3)

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod3.1<-inla(formula2.1,data=dat,family="poisson",control.compute = list(dic = TRUE))
summary(mod3.1)
```

```

ds3<-inla.make.lincomb(z1=1, z2=-1)
names(ds3) = "ds3"
ds3.mod<-inla(formula2.1, "poisson",
              data = dat,
              lincomb = ds3)
summary(ds3.mod)

```

```

#obtenção das medidas resumo a posteriori do contraste para o quociente
mod3.2<-inla(formula2.1,data=dat,family="poisson",control.compute =
  list(config = TRUE))
samples2<- inla.posterior.sample(1,mod3.2)
str(samples2)
dim(samples2[[1]]$latent)
names(samples2[[1]]$latent[2492:2496,1])
#z1 está na posição 2495 e z2 na posição 2496
nsim<-1000
z1.p<-2495
z2.p<-2496
samples3<- inla.posterior.sample(nsim,mod3.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples3[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples3[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

```

B.6 Rotinas programa R para ajuste do modelo combinado

```
#-----
#Modelo 4 - Modelo Combinado
#-----

#ajuste do modelo
mod4<-inla(formula2,data=dat,family="nbinomial",control.compute =
list(dic = TRUE,cpo=TRUE))
names(mod4)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod4$marginals.fixed)
par(mfrow=c(2,2))
plot(mod4$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod4$marginals.fixed$'trt110:t1',xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod4$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod4$marginals.fixed$'trt111:t1',xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

#obtenção das medidas resumo a posteriori e do gráfico para o
#parâmetro associado ao efeito aleatório normal
names(mod4$marginals.hyperpar)
tau2<-mod4$marginals.hyperpar[[2]];tau2
stdv2<-inla.tmarginal(function(x) 1/sqrt(x), tau2)
inla.qmarginal(0.025, stdv2);inla.qmarginal(0.975, stdv2)
inla.qmarginal(0.5, stdv2);mean(stdv2[,1])
sd(stdv2[,1]);median(stdv2[,1])
plot(inla.smarginal(stdv2), type="l",xlab=expression(paste(sigma^2)),ylab="Densidade")

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod4$cpo$pit,xlab="Índice",ylab="PIT",main="(a)",ylim=c(0,1))
hist(mod4$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)",ylim=c(0,500))
dev.off()

#obtenção dos gráficos dos valores de CP0
hist(mod4$cpo$cpo,xlab="Probabilidade",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod4)

#obtenção das medidas resumo a posteriori e do gráfico para os
#parâmetros associados ao efeito aleatório gama
over2<-mod4$marginals.hyperpar[[1]];over2
D2<-inla.tmarginal(function(x) 1/x, over2)
```

```

inla.qmarginal(0.025, D2);inla.qmarginal(0.975, D2)
inla.qmarginal(0.5, D2);mean(D2[,1])
sd(D2[,1]);median(D2[,1])

par(mfrow=c(2,1))
plot(mod4$marginals.hyperpar[[1]],xlab=expression(paste(alpha)),
ylab="Densidade",type="l",main="")
plot(inla.smarginal(D2), type="l",xlab=expression(paste(beta)),
ylab="Densidade",main="")
dev.off()

#obtenção das medidas resumo a posteriori para o contraste para a diferença
mod4.1<-inla(formula2.1,data=dat,family="nbinomial",control.compute =
list(dic = TRUE))
summary(mod4.1)

ds4<-inla.make.lincomb(z1=1, z2=-1)
names(ds4) = "ds4"
ds4.mod<-inla(formula2.1, "nbinomial",
              data = dat,
              lincomb = ds4)
summary(ds4.mod)

#obtenção das medidas resumo a posteriori para o contraste para o quociente
mod4.2<-inla(formula2.1,data=dat,family="nbinomial",control.compute =
list(config = TRUE))
samples2<- inla.posterior.sample(1,mod4.2)
str(samples2)
dim(samples2[[1]]$latent)
names(samples2[[1]]$latent[2492:2496,1])
#z1 está na posição 2495 e z2 na posição 2496
nsim<-1000
z1.p<-2495
z2.p<-2496
samples4<- inla.posterior.sample(nsim,mod4.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples4[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples4[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd

```

```
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

#gráfico combinado com os valores de CP0 para todos os modelos
par(mfrow=c(2,2))
hist(mod1$cpo$cpo,xlab="CP0",ylab="Frequência",main="(a)")
hist(mod2$cpo$cpo,xlab="CP0",ylab="Frequência",main="(b)")
hist(mod3$cpo$cpo,xlab="CP0",ylab="Frequência",main="(c)")
hist(mod4$cpo$cpo,xlab="CP0",ylab="Frequência",main="(d)")
dev.off()
```


B.7 Rotinas programa R para estruturação dos dados e ajuste do modelo ZIP

```
#-----
#Modelo 5 - Modelo ZIP
#-----

#ajuste do modelo
mod5<-inla(formula1,data=dat,family="zeroinflatedpoisson1",control.compute =
list(dic = TRUE,cpo=TRUE))
names(mod5)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod5$marginals.fixed)
par(mfrow=c(2,2))
plot(mod5$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod5$marginals.fixed$`trt110:t1`,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod5$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod5$marginals.fixed$`trt111:t1`,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

names(mod5$marginals.hyperpar)
plot(inla.smarginal(mod5$marginals.hyperpar$`zero-probability parameter
for zero-inflated poisson_1`), type="l",xlab=expression(paste(omega)),ylab="Densidade")

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod5$cpo$pit,xlab="Índice",ylab="PIT",main="(a)")
hist(mod5$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)")
dev.off()

#obtenção do gráfico dos valores de CP0
hist(mod5$cpo$cpo,xlab="CP0",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod5)

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod5.1<-inla(formula1.1,data=dat,family="zeroinflatedpoisson1",
control.compute = list(dic = TRUE))
summary(mod5.1)
ds5<-inla.make.lincomb(z1=1, z2=-1)
names(ds5) = "ds5"
ds5.mod<-inla(formula1.1, "zeroinflatedpoisson1",
              data = dat,
              lincomb = ds5)
summary(ds5.mod)
```

```

#obtenção das medidas resumo a posteriori do contraste para o quociente
mod5.2<-inla(formula1.1,data=dat,family="zeroinflatedpoisson1",
control.compute = list(config=TRUE))
samples1<- inla.posterior.sample(1,mod5.2)
str(samples1)
dim(samples1[[1]]$latent)
names(samples1[[1]]$latent[2403:2407,1])
#z1 está na posição 2406 e z2 na posição 207
nsim<-1000
z1.p<-2406
z2.p<-2407
samples5<- inla.posterior.sample(nsim,mod5.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples5[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples5[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

```

B.8 Rotinas programa R para estruturação dos dados e ajuste do modelo ZINB

```
#-----
#Modelo 6 - ZINB
#-----

#ajuste do modelo
mod6<-inla(formula1,data=dat,family="zeroinflatednbinomial1",
control.compute = list(dic = TRUE,cpo=TRUE))
names(mod6)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod6$marginals.fixed)
par(mfrow=c(2,2))
plot(mod6$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod6$marginals.fixed$`trt110:t1`,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod6$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod6$marginals.fixed$`trt111:t1`,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

names(mod6$marginals.hyperpar)
par(mfrow=c(2,1))
plot(inla.smarginal(mod6$marginals.hyperpar$`zero-probability
parameter for zero-inflated nbinomial_1`), type="l",
xlab=expression(paste(omega)),ylab="Densidade")
plot(inla.smarginal(mod6$marginals.hyperpar$`size for
nbinomial zero-inflated observations`), type="l",xlab=expression(paste(alpha)),
ylab="Densidade")
dev.off()

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod6$cpo$pit,xlab="Índice",ylab="PIT",main="(a)")
hist(mod6$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)")
dev.off()

#obtenção do gráfico dos valores de CPO
hist(mod6$cpo$cpo,xlab="CPO",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod6)

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod6.1<-inla(formula1.1,data=dat,family="zeroinflatednbinomial1",
control.compute = list(dic = TRUE))
summary(mod6.1)
```

```

ds6<-inla.make.lincomb(z1=1, z2=-1)
names(ds6) = "ds6"
ds6.mod<-inla(formula1.1, "zeroinflatednbinomial1",
              data = dat,
              lincomb = ds6)
summary(ds6.mod)

#obtenção das medidas resumo a posteriori do contraste para a quociente
mod6.2<-inla(formula1.1,data=dat,family="zeroinflatednbinomial1",
control.compute = list(config=TRUE))
samples1<- inla.posterior.sample(1,mod6.2)
str(samples1)
dim(samples1[[1]]$latent)
names(samples1[[1]]$latent[2403:2407,1])
#z1 está na posição 2406 e z2 na posição 2407
nsim<-1000
z1.p<-2406
z2.p<-2407
samples6<- inla.posterior.sample(nsim,mod6.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples6[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples6[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

```

B.9 Rotinas programa R para estruturação dos dados e ajuste do modelo ZIP-normal

```
#-----
#Modelo 7 - ZIP-Normal
#-----

#ajuste do modelo
formula2<-y1~trt11/t1-1+f(indv,model="iid")
formula2.1<-y1~-1+G1+G2+z1+z2+f(indv,model="iid")
mod7<-inla(formula2,data=dat,family="zeroinflatedpoisson1"
,control.compute = list(dic = TRUE,cpo=TRUE))

names(mod7)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod7$marginals.fixed)
par(mfrow=c(2,2))
plot(mod7$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod7$marginals.fixed$trt110:t1,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod7$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod7$marginals.fixed$trt111:t1,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

names(mod7$marginals.hyperpar)
par(mfrow=c(2,1))
plot(inla.smarginal(mod7$marginals.hyperpar$'zero-probability
parameter for zero-inflated poisson_1'), type="l",xlab=expression(paste(omega)),
ylab="Densidade")

tau3<-mod7$marginals.hyperpar[[2]];tau3
stdv3<-inla.tmarginal(function(x) 1/sqrt(x), tau3)
inla.qmarginal(0.025, stdv3);inla.qmarginal(0.975, stdv3)
inla.qmarginal(0.5, stdv3);mean(stdv3[,1]);sd(stdv3[,1])
plot(inla.smarginal(stdv3), type="l",xlab=expression(paste(sigma^2)),
ylab="Densidade")

#obtenção dos gráficos dos valores de PIT
mod7.1<- inla.cpo(mod7)
par(mfrow=c(2,1))
plot(mod7.1$cpo$pit,xlab="Índice",ylab="PIT",main="(a)")
hist(mod7.1$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)")
dev.off()

#obtenção do gráfico dos valores de CP0
hist(mod7.1$cpo$cpo,xlab="CP0",ylab="Frequência",main="")

#medidas resumo a posteriori
```

```

summary(mod7.1)

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod7.2<-inla(formula2.1,data=dat,family="zeroinflatedpoisson1",
control.compute = list(dic = TRUE))
summary(mod7.2)
ds7<-inla.make.lincomb(z1=1, z2=-1)
names(ds7) = "ds7"
ds7.mod<-inla(formula2.1, "zeroinflatedpoisson1",
              data = dat,
              lincomb = ds7)
summary(ds7.mod)

summary(mod7.2)

#obtenção das medidas resumo a posteriori do contraste para o quociente
mod7.3<-inla(formula2.1,data=dat,family="zeroinflatedpoisson1",
control.compute = list(config=TRUE))
samples1<- inla.posterior.sample(1,mod7.3)
str(samples1)
dim(samples1[[1]]$latent)
names(samples1[[1]]$latent[2493:2496,1])
#z1 está na posição 2495 e z2 na posição 2496
nsim<-1000
z1.p<-2495
z2.p<-2496
samples7<- inla.posterior.sample(nsim,mod7.3)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples7[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples7[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));
print("IC para rs");rs.ic

```

B.10 Rotinas programa R para estruturação dos dados e ajuste do modelo ZIB-normal

```
#-----
#Modelo 8 - ZINB-Normal
#-----

#ajuste do modelo
mod8<-inla(formula2,data=dat,family="zeroinflatednbinomial1",
control.compute = list(dic = TRUE,cpo=TRUE))
names(mod8)

#obtenção dos gráficos das densidades marginais a posteriori dos parâmetros
names(mod8$marginals.fixed)
par(mfrow=c(2,2))
plot(mod8$marginals.fixed$trt110,xlab=expression(paste(beta[0][0])),
ylab="Densidade",type="l",main="")
plot(mod8$marginals.fixed$`trt110:t1`,xlab=expression(paste(beta[0][1])),
ylab="Densidade",type="l",main="")
plot(mod8$marginals.fixed$trt111,xlab=expression(paste(beta[1][0])),
ylab="Densidade",type="l",main="")
plot(mod8$marginals.fixed$`trt111:t1`,xlab=expression(paste(beta[1][1])),
ylab="Densidade",type="l",main="")
dev.off()

names(mod8$marginals.hyperpar)
par(mfrow=c(3,1))
plot(inla.smarginal(mod8$marginals.hyperpar$`zero-probability
parameter for zero-inflated nbinomial_1`), type="l",xlab=expression(paste(omega)),
ylab="Densidade")
plot(inla.smarginal(mod8$marginals.hyperpar$`size for nbinomial
zero-inflated observations`), type="l",xlab=expression(paste(alpha)),ylab="Densidade")

tau4<-mod8$marginals.hyperpar[[3]];tau4
stdv4<-inla.tmarginal(function(x) 1/sqrt(x), tau4)
inla.qmarginal(0.025, stdv4);inla.qmarginal(0.975, stdv4)
inla.qmarginal(0.5, stdv4);mean(stdv4[,1]);sd(stdv4[,1])
plot(inla.smarginal(stdv3), type="l",xlab=expression(paste(sigma^2)),
ylab="Densidade")

#obtenção dos gráficos dos valores de PIT
par(mfrow=c(2,1))
plot(mod8$cpo$pit,xlab="Índice",ylab="PIT",main="(a)")
hist(mod8$cpo$pit,xlab="PIT",ylab="Frequência",main="(b)")
dev.off()

#obtenção dos gráficos dos valores de CPD
hist(mod8$cpo$cpo,xlab="CPD",ylab="Frequência",main="")

#medidas resumo a posteriori
summary(mod8)
```

```

#obtenção das medidas resumo a posteriori do contraste para a diferença
mod8.1<-inla(formula2.1,data=dat,family="zeroinflatednbinomial1",
control.compute = list(dic = TRUE))
summary(mod8.1)
ds8<-inla.make.lincomb(z1=1, z2=-1)
names(ds8) = "ds8"
ds8.mod<-inla(formula2.1, "zeroinflatednbinomial1",
              data = dat,
              lincomb = ds8)
summary(ds8.mod)

#obtenção das medidas resumo a posteriori do contraste para o quociente
mod8.2<-inla(formula2.1,data=dat,family="zeroinflatednbinomial1",
control.compute = list(config=TRUE))
samples1<- inla.posterior.sample(1,mod8.2)
str(samples1)
dim(samples1[[1]]$latent)
names(samples1[[1]]$latent[2493:2496,1])
#z1 está na posição 2495 e z2 na posição 2496
nsim<-1000
z1.p<-2495
z2.p<-2496
samples8<- inla.posterior.sample(nsim,mod8.2)

rs1<-NULL
for(i in 1:nsim){
  rs1[i]<-samples8[[i]]$latent[z1.p,1]
}
head(rs1)

rs2<-NULL
for(i in 1:nsim){
  rs2[i]<-samples8[[i]]$latent[z2.p,1]
}
head(rs2)

rs<-rs1/rs2;
rs.med<-mean(rs);print("Média de rs");rs.med
rs.sd<-sd(rs);print("DP de rs");rs.sd
rs.md<-median(rs);print("Mediana de rs");rs.md
rs.ic<-cbind(quantile(rs,0.025),quantile(rs,0.975));print("IC para rs");rs.ic

#gráfico combinado com os valores de CP0
par(mfrow=c(2,2))
hist(mod5$cpo$cpo,xlab="CP0",ylab="Frequência",main="(a)")
hist(mod6$cpo$cpo,xlab="CP0",ylab="Frequência",main="(b)")
hist(mod7.1$cpo$cpo,xlab="CP0",ylab="Frequência",main="(c)")
hist(mod8$cpo$cpo,xlab="CP0",ylab="Frequência",main="(d)")

```



```
dev.off()
```

```
#gráfico combinado com os valores de CP0
```

```
par(mfrow=c(2,1))
```

```
hist(mod1$cpo$cpo,xlab="CP0",ylab="Frequência",main="(a)")
```

```
hist(mod7.1$cpo$cpo,xlab="CP0",ylab="Frequência",main="(b)")
```

```
dev.off()
```

```
#PsFB
```

```
cpo4<-na.omit(mod4$cpo$cpo)
```

```
cpozipn<-na.omit(mod7.1$cpo$cpo)
```

```
round(sum(log(cpozipn))/sum(log(cpo4)),4)
```

B.11 Rotinas programa R para o estudo de simulação

```
#-----
#Estudo de simulação
#-----
rm(list=ls())
require(INLA)

if( file.exists("out/sim-zip.txt") ){
  B <- read.table("out/sim-zip.txt",head=T)
}else{
  B <- NULL;B$rep <- 0
}

for(i in (max(B$rep)+1):1000){
  ### ZIP-Normal
  #estrutura dos dados
  nis <- 89
  nts <- 27
  nobs <- nis*nts
  tj <- rep(1:nts,nis)
  is <- sort(rep(1:nis,nts))
  dr <- as.factor( c(rep(0,45*nts ),rep(1,44*nts )))

  #distribuições dos parâmetros e variável resposta
  bis <- rep(rnorm(nis,mean=0,sd=sqrt(1/0.9707)),each=nts)
  etas <- ifelse(dr==0,0.8957,0.7938) + ifelse(dr==0,-0.0081,-0.0111)*tj + bis

  y.sim <- rbinom(nobs,size=1,prob=(1-0.1661))*rpois(nobs,lambda=exp(etas))
  dat.sim <- data.frame(dr,is,tj,y.sim)

  #INLA ajuste
  form.sim <- y.sim~dr/tj-1+f(is,model="iid")
  zipsim.pngfit <- suppressWarnings(inla(form.sim,data=dat.sim,family="nbinomial",
  control.compute = list(dic = TRUE,cpo=TRUE)))
  zipsim.pngfit <- suppressWarnings(inla.cpo(zipsim.pngfit,verbose=F))

  zipsim.zipfit <- suppressWarnings(inla(form.sim,data=dat.sim,
  family="zeroinflatedpoisson1",control.compute = list(dic = TRUE,cpo=TRUE)))
  zipsim.zipfit <- suppressWarnings(inla.cpo(zipsim.zipfit,verbose=F))
}
```

```

#inferência
b0.0 <- data.frame(rep=i, zipsim.zipfit$summary.fixed[1,c(1:3,5)])
b1.0 <- data.frame(rep=i, zipsim.zipfit$summary.fixed[3,c(1:3,5)])
b0.1 <- data.frame(rep=i, zipsim.zipfit$summary.fixed[2,c(1:3,5)])
b1.1 <- data.frame(rep=i, zipsim.zipfit$summary.fixed[4,c(1:3,5)])
pzip <- data.frame(rep=i, zipsim.zipfit$summary.hyperpar[1,c(1:3,5)])
tau <- data.frame(rep=i, zipsim.zipfit$summary.hyperpar[2,c(1:3,5)])

#seleção
dic.pngfit <- zipsim.pngfit$dic$dic
dic.zipfit <- zipsim.zipfit$dic$dic
peff.zipfit <- summary(zipsim.zipfit)$dic$p.eff
peff.pngfit <- summary(zipsim.pngfit)$dic$p.eff

if(peff.zipfit > peff.pngfit){
  pbf <- sum(log(zipsim.pngfit$cpo$cpo),na.rm=T)/sum(log(zipsim.zipfit$cpo$cpo),
,na.rm=T)
  pbfl <- "png:zip"
}else if(peff.zipfit < peff.pngfit){
  pbf <- sum(log(zipsim.zipfit$cpo$cpo),na.rm=T)/sum(log(zipsim.pngfit$cpo$cpo),
,na.rm=T)
  pbfl <- "zip:png"
}else{
  pbf <- sum(log(zipsim.pngfit$cpo$cpo),na.rm=T)/sum(log(zipsim.zipfit$cpo$cpo),
,na.rm=T)
  pbfl <- "png=zip"
}
if(dic.pngfit < dic.zipfit){
  dic.dec <- 0
}else if(dic.pngfit > dic.zipfit){
  dic.dec <- 1
}else dic.dec <- -1

#resultados da simulação
A <- data.frame(rep=i,dic.pngfit=dic.pngfit,dic.zipfit=dic.zipfit,dic.dec=dic.dec,
pbf=pbf,pbfl=pbfl)
if(i == 1){
  write.table(A,"out/sim-zip.txt",append=F,row.names=F,col.names=T)
  write.table(b0.0,"out/sim-zip-b00.txt",append=F,row.names=F,col.names=T)
  write.table(b1.0,"out/sim-zip-b10.txt",append=F,row.names=F,col.names=T)
  write.table(b0.1,"out/sim-zip-b01.txt",append=F,row.names=F,col.names=T)
  write.table(b1.1,"out/sim-zip-b11.txt",append=F,row.names=F,col.names=T)
  write.table(pzip,"out/sim-zip-pzip.txt",append=F,row.names=F,col.names=T)
  write.table(tau,"out/sim-zip-tau.txt",append=F,row.names=F,col.names=T)
}else{
  write.table(A,"out/sim-zip.txt",append=T,row.names=F,col.names=F)
  write.table(b0.0,"out/sim-zip-b00.txt",append=T,row.names=F,col.names=F)
  write.table(b1.0,"out/sim-zip-b10.txt",append=T,row.names=F,col.names=F)
  write.table(b0.1,"out/sim-zip-b01.txt",append=T,row.names=F,col.names=F)
  write.table(b1.1,"out/sim-zip-b11.txt",append=T,row.names=F,col.names=F)
  write.table(pzip,"out/sim-zip-pzip.txt",append=T,row.names=F,col.names=F)

```

```
    write.table(tau,"out/sim-zip-tau.txt",append=T,row.names=F,col.names=F)
  }
  print(i)
}
```