



**MARCEL IRVING PEREIRA MELO**

**MODELAGEM DE PRÊMIOS DE SEGUROS DE  
AUTOMÓVEIS VIA APROXIMAÇÃO DE LAPLACE  
ANINHADA INTEGRADA (INLA)**

**LAVRAS - MG**

**2016**

**MARCEL IRVING PEREIRA MÉLO**

**MODELAGEM DE PRÊMIOS DE SEGUROS DE AUTOMÓVEIS VIA  
APROXIMAÇÃO DE LAPLACE ANINHADA INTEGRADA (INLA)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador  
Dr. Júlio Sílvio de Sousa Bueno Filho

**LAVRAS - MG  
2016**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha  
Catalográfica da Biblioteca Universitária da UFLA, com dados  
informados pelo(a) próprio(a) autor(a).**

Mélo, Marcel Irving Pereira.

Modelagem de prêmios de seguros de automóveis via Aproximação  
de Laplace Aninhada Integrada (INLA) / Marcel Irving Pereira Mélo. –  
Lavras : UFLA, 2016.

100 p.

Dissertação (mestrado acadêmico) – Universidade Federal de Lavras,  
2016.

Orientador: Júlio Sílvio de Sousa Bueno Filho.

Bibliografia.

1. Distribuição Gama 2. Modelos Espaciais 3. Modelos Gaussianos  
Latentes 4. Seguros de Automóveis . I. Universidade Federal de Lavras.  
II. Título.

**MARCEL IRVING PEREIRA MÉLO**

**MODELAGEM DE PRÊMIOS DE SEGUROS DE AUTOMÓVEIS VIA  
APROXIMAÇÃO DE LAPLACE ANINHADA INTEGRADA (INLA)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADO em 01 de abril de 2016.

Dra. Izabela Regina Cardoso de Oliveira

UFLA

Dr. Leandro Ferreira

UNIFAL-MG

Dr. Júlio Sílvio de Sousa Bueno Filho  
Orientador

**LAVRAS - MG**

**2016**

## AGRADECIMENTOS

Ao professor Júlio Sílvio de Sousa Bueno Filho, pela orientação, pelos ensinamentos, incentivos, confiança e por ser um profissional exemplar e consciente do seu papel na construção da ciência.

Ao professor Reinaldo Antônio Gomes Marques pelos ensinamentos, sugestões, críticas, incentivos, confiança e todo auxílio oferecido para a realização do trabalho, sempre disposto a ajudar.

Aos professores Danilo Machado Pires, Leandro Ferreira, Luiz Alberto Beijo e Izabela Regina Cardoso de Oliveira pelos ensinamentos, contribuições, sugestões e críticas.

À minha mãe Neusa e meu pai Osvaldo, sempre dedicados na minha educação e formação humana, serei eternamente grato pelo amor, carinho e esforço que tiveram por mim.

Aos meus irmãos Roger e Rafael pela amizade e respeito.

À minha noiva Renata, pelo amor, carinho, dedicação, companheirismo, respeito e por todos esses anos que compartilhamos juntos. Pelas sugestões, críticas e auxílios para a realização do trabalho.

À minha segunda família, meu sogros Sebastião e Maria de Fátima e meus cunhados Renan e Rosana, pela hospitalidade, pela amizade, respeito e apoio.

Aos professores do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, em especial, aos professores Daniel, Devanil, Lucas, Augusto, Márcio e Thelma.

Aos meus colegas da Pós, Guilherme, Kelly, Janaína, LÍlian, Carlos, Henrique.

Aos funcionários do DEX, sempre dispostos a ajudar e auxiliar no que for necessário, em particular, à Nádia e à Jose.

Aos professores do ICISA da UNIFAL-MG, em especial, os professores Léo, Marçal, Deive, Gislene, Luciene, Letícia e Pamila.

Aos meus professores do ensino fundamental Néilson, João Grilo, Afonso e Maurício que me ensinaram e sempre me incentivaram a estudar.

À CAPES, pelo apoio financeiro e por contribuir com o desenvolvimento da pesquisa e conhecimento.

À todas as pessoas que de algum modo me ajudaram para que eu pudesse concluir o mestrado, o meu muito obrigado!

## RESUMO

Nessa dissertação de mestrado, é realizada a análise bayesiana para dados de seguro de automóveis utilizando modelos que pertencem a uma classe denominada de modelos gaussianos latentes. Tais modelos englobam os modelos lineares generalizados e mistos e podem assumir diversas estruturas, como por exemplo, temporais, espaciais e espaço-temporais. Para a estimação das distribuições marginais a posteriori utiliza-se a Aproximação de Laplace Aninhada Integrada (INLA), um veloz algoritmo determinístico para inferência bayesiana com aplicação nestes modelos. A seleção de modelos foi baseado no *Deviance Information Criterion* (DIC) e na pseudo log-verossimilhança marginal (LPML) que mede a qualidade preditiva do modelo e é baseada na ordenada da densidade preditiva condicional (CPO). Para verificar a calibração do modelo, um histograma da transformação integral de probabilidade (PIT) é construído, a fim de verificar a uniformidade dos dados. À luz dos resultados, conclui-se que condutores do sexo masculino, em média, pagam um prêmio maior que condutores do sexo feminino. De forma análoga, condutores mais jovens pagam, em média, um prêmio maior que condutores mais experientes. Destaca-se ainda que, os modelos que melhor descrevem os dados contém efeitos agrupados (aleatórios) e que, dentre estes, os mais precisos incluem efeitos espaciais.

Palavras-chave: Distribuição Gama, Modelos Espaciais, Modelos Gaussianos Latentes, Seguro de Automóveis.

## ABSTRACT

In this dissertation we carried out Bayesian analysis for auto insurance data using latent Gaussian models. Such class of models include generalized linear (mixed) models and can take various structures, such as temporal, spatial and spatiotemporal. To evaluate the marginal posterior distributions we use the Integrated Nested Laplace Approximation (INLA). This is a fast deterministic algorithm for Bayesian inference with direct application to latent Gaussian models. Model selection was based on Deviance Information Criterion (DIC) as well as the log pseudo marginal likelihood (LPML) that measures the predictive quality of the model and is based on the conditional predictive density ordinate (CPO). To check the model calibration, a histogram of probability integral transform (PIT) was drawn in which we verify the uniformity of the data. In view of the results to drivers premium, we conclude that males pay a higher premium than females, on average. Similarly, youngers pay on average a higher premium than more experienced drivers. We also emphasize that the model that best adjust to data, includes grouping (random) effects, and those are best modelled including spatial effects. Keywords: Automobile Insurance, Gamma Distribution, Gaussian Latent Models, Spatial Models.



## Lista de Figuras

1	Evolução do Prêmio Direto Anual . . . . .	22
2	Principais segmentos do seguro - 2013 . . . . .	23
3	Box-Plot do Prêmio Médio Estratificado por Sexo do Condutor . . . . .	70
4	Proporção do número de sinistros (à esquerda) e do valor total pago em indenizações (à direita) por sexo do condutor, entre os anos de 2007 a 2014. . . . .	71
5	Boxplot do prêmio médio estratificado por faixa etária . . . . .	72
6	Dispersão entre os valores médio do Prêmio de 2007 a 2014 . . . . .	73
7	Evolução do prêmio médio de seguros de automóveis entre os anos de 2007 a 2014 . . . . .	74
8	Histograma da transformação integral de probabilidade . . . . .	79
9	Dispersão entre os valores médios dos prêmios observados e os valores médios dos prêmios preditos . . . . .	80
10	Média a posteriori para os efeitos espaciais para cada estado do Brasil ( $\zeta = \exp(\mathbf{u} + \mathbf{v})$ ) referente aos anos de 2007 a 2014 . . . . .	81
11	Probabilidade a posteriori $p(\zeta_k > 1 \mathbf{y})$ para os efeitos espaciais para cada estado do Brasil referente aos anos de 2007 a 2014 . . . . .	82
12	Tendência temporal da média a posteriori do valor médio do prêmio de seguro de automóveis no Brasil para os anos de 2007 a 2014 . . . . .	83
13	Média a posteriori da interação espaço-tempo $\delta_{kt}$ para o valor médio do prêmio de seguro de automóveis no Brasil para os anos de 2007 a 2014 . . . . .	84

## Lista de Tabelas

1	Elementos de algumas distribuições da família exponencial . . . .	32
2	Função Desvio para os principais Modelos Lineares Generalizados	38
3	Dados Seguros de Automóveis . . . . .	64
4	Componentes aleatórios presentes no modelo . . . . .	68
5	Medidas para a comparação e seleção de modelos . . . . .	75
6	Resumo dos efeitos do modelo VI . . . . .	76
7	Resumo dos efeitos do modelo VI* . . . . .	77
8	Resumo dos efeitos fixos (escala natural) do Modelo VI* . . . . .	77

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>2</b>	<b>O SEGURO</b>	<b>15</b>
<b>2.1</b>	<b>Conceito e Finalidade</b>	<b>15</b>
<b>2.2</b>	<b>Evolução, princípios e estrutura do sistema de seguros no Brasil</b>	<b>17</b>
<b>2.3</b>	<b>Espécies de seguro</b>	<b>19</b>
<b>2.4</b>	<b>Seguro de automóvel no Brasil</b>	<b>20</b>
<b>2.4.1</b>	<b>Cálculo de Prêmio</b>	<b>23</b>
<b>3</b>	<b>MODELO LINEAR GENERALIZADO</b>	<b>25</b>
<b>3.1</b>	<b>Vantagens do Modelo Linear Generalizado Sobre Dados Trans-</b> <b>formados</b>	<b>28</b>
<b>3.2</b>	<b>A Média e a Variância de um Modelo Linear Generalizado</b>	<b>28</b>
<b>3.3</b>	<b>Estimação via Máxima Verossimilhança</b>	<b>32</b>
<b>3.4</b>	<b>Função Desvio</b>	<b>37</b>
<b>4</b>	<b>MODELO LINEAR MISTO</b>	<b>38</b>
<b>4.1</b>	<b>Modelo Geral</b>	<b>39</b>
<b>5</b>	<b>MODELO LINEAR GENERALIZADO MISTO</b>	<b>41</b>
<b>5.1</b>	<b>Modelo Geral</b>	<b>42</b>
<b>5.2</b>	<b>CrITÉrios de Informação para Seleção de Modelos</b>	<b>44</b>
<b>6</b>	<b>ESTATÍSTICA BAYESIANA</b>	<b>45</b>
<b>6.1</b>	<b>Inferência Bayesiana Aproximada</b>	<b>49</b>
<b>6.1.1</b>	<b>Amostrador de Gibbs</b>	<b>50</b>
<b>6.1.2</b>	<b>Metropolis-Hastings</b>	<b>51</b>
<b>6.1.3</b>	<b>INLA: Aproximação de Laplace Aninhada Integrada</b>	<b>52</b>
<b>6.1.3.1</b>	<b>Aproximação de Laplace</b>	<b>53</b>
<b>6.1.3.2</b>	<b>Modelos Gaussianos Latentes</b>	<b>54</b>

6.1.3.3	Metodologia INLA . . . . .	56
6.2	Verificação e Critérios de Seleção do Modelo . . . . .	60
6.2.1	Métodos Baseados na Distribuição Preditiva . . . . .	61
6.2.2	<i>Deviance Information Criterion (DIC)</i> . . . . .	62
7	<b>MATERIAL E MÉTODOS</b> . . . . .	64
8	<b>RESULTADOS</b> . . . . .	70
8.1	Discussões . . . . .	85
9	<b>CONCLUSÃO</b> . . . . .	86
	<b>REFERÊNCIAS</b> . . . . .	88
	<b>APÊNDICE</b> . . . . .	95

## 1 INTRODUÇÃO

O seguro é um instrumento financeiro que permite ao segurado, mediante o pagamento de uma quantia fixa (prêmio), transferir as consequências financeiras de um possível evento danoso (risco). No entanto, ainda que o seguro no Brasil tem alcançado números cada vez maiores- em 2013 a participação do seguro no PIB foi de 3%, enquanto que em 2003 esse número era de 1,81%- ainda não atingiu patamares como em países desenvolvidos, por exemplo os Estados Unidos, em 2013, tiveram a participação dos seguros em 8% do seu PIB.

O seguro de automóveis é um dos ramos mais relevantes do seguro e teve seu desenvolvimento principalmente, depois que a Superintendência de Seguros Privados (SUSEP) editou a Circular n.27 em 1986, conferindo mais liberdade para as seguradoras, com a liberação da tarifa de automóveis. Nesse período, ocorreram quatro evoluções importantes no ramo: correção monetária da importância segurada; implantação do código de defesa do consumidor; serviços de assistência aos veículo e passageiros; especificação das características dos condutores na precificação (PITA; DOMINGUEZ, 2011).

Modelos para precificação do prêmio de seguros de automóveis tem sido propostos e a ferramenta estatística mais comum para este tipo de análise de dados foi por muitos anos a de modelos lineares generalizados (GLM) (ANTONIO; BEIRLANT, 2007). O grande mérito dessa abordagem é a generalização da variável resposta, que pode ser estendida para distribuições que pertençam à família exponencial. E o uso de GLM dentro do contexto atuarial se deve ao fato de que, existe uma vasta aplicação dos modelos da família exponencial na análise de risco, como por exemplo, a modelagem da frequência de sinistros através da distribuições Binomial, Poisson, Binomial Negativa e severidade como a distribuição Gama e Normal Inversa (DICKSON, 2010), (KAAS et al., 2008), (TSE, 2009), (EMBRECHTS; KLÜPPELBERG; MIKOSCH, 1997). Todavia, a análise por modelos lineares generalizados requer a suposição de uma amostra de variáveis aleatórias independentes, o que geralmente não é satisfeito para dados de seguros (DENUIT et al., 2006), (SHI, 2016).

Os modelos lineares generalizados mistos (GLMM), que são extensões dos modelos lineares generalizados com a inclusão de efeitos aleatórios, surgem como alternativa quando não há independência, pois determinam a estrutura de correlação entre as observações e podem modelar dados binários, de contagem, assim como dados agrupados e dados longitudinais.

Uma forma de modelo mais geral pode ser ainda definida. Os modelos gaussianos latentes são uma subclasse de modelos de regressão de estrutura aditiva e englobam os modelos lineares generalizados (mistos), modelos com efeitos espaciais, temporais, espaço-temporais, de maneira que tais modelos permitem uma vasta quantidade de aplicações.

Dentro do paradigma bayesiano, a informação trazida pelos dados, descrita pela função de verossimilhança, é combinada com toda informação disponível *a priori*, especificada em uma distribuição *a priori* e resumida, via Regra de Bayes, em uma distribuição *a posteriori*, que detém toda informação descrita pelo parâmetro de interesse. No entanto, na maioria das situações, a distribuição *a posteriori* não possui forma fechada, de forma que métodos numéricos são necessários para a aproximação dessas distribuições. Os métodos mais tradicionais para a aproximação das densidades marginais *a posteriori* são os Métodos de Monte Carlo via Cadeia de Markov (MCMC). Entretanto, no contexto de modelos gaussianos latentes, tais métodos podem apresentar problemas, no que diz respeito, a tempo de execução e convergência.

Como alternativa aos métodos MCMC, Rue, Martino e Chopin (2009) desenvolveram a metodologia INLA, nomeada como Aproximação de Laplace Aninhada Integrada, um algoritmo determinístico para inferência bayesiana com sua aplicação voltada para os modelos gaussianos latentes e possuindo um custo computacional bem inferior aos métodos MCMC.

Dessa forma, o objetivo geral desse trabalho é apresentar uma análise bayesiana de modelos gaussianos latentes para dados de seguro utilizando a metodologia INLA. Os objetivos específicos são:

- Apresentar a metodologia INLA, bem como, modelos gaussianos latentes e

alguns dos seus casos particulares;

- Comparar algumas das possíveis estruturas dos modelos gaussianos latentes, a fim de selecionar o modelo que melhor descreve os dados de seguro de automóveis.
- Mostrar que a análise prática de modelos complexos com dependência espacial e muitos fatores pode ser feita de forma rápida utilizando o INLA dentro do *Software R* (R Core Team, 2014).

## 2 O SEGURO

### 2.1 Conceito e Finalidade

Segundo Bowers et al. (1997) o seguro tem como fim, proteger o segurado contra sérios danos financeiros resultantes de eventos aleatórios que poderiam atrapalhar seus planos.

O conceito de seguro, segundo Martins e Martins (2008), pode ser entendido como a transferência das consequências derivadas da ocorrência de um risco da qual uma parte, o segurado, transfere a probabilidade de perda financeira para outra parte, a seguradora.

Cavaliere Filho (2014) o conceitua como o contrato pelo qual o segurador, mediante um recebimento de um prêmio, assume perante o segurado a obrigação de pagar-lhe uma determinada indenização, prevista no contrato, caso o risco que está sujeito se concretize em um sinistro.

Já Souza (2007) define o seguro como operação que toma forma jurídica de um contrato, em que uma das partes (segurador) garante obrigação para com a outra (segurado), mediante o recebimento de uma quantia estipulada (prêmio), a compensá-la (indenização) ou compensar outra pessoa que tenha o direito (beneficiário) por um prejuízo (sinistro), resultante de um evento futuro, possível e incerto (risco), indicado no contrato (apólice).

Pimenta (2010) o define como instituto, pelo qual, por meio de cálculos atuariais, pode-se calcular probabilidades de perda, para que, mediante o pagamento de prêmio, as consequências econômicas do risco sejam garantidas por uma terceira pessoa (segurador).

Apesar dos conceitos sobre seguro apresentados anteriormente, cabe ressaltar que o seguro não transfere o risco do segurado para o segurador, pois esse é intransferível. Na verdade o seguro transfere as consequências econômicas do risco caso ele venha a se concretizar em sinistro (CAVALIERE FILHO, 2014).



Nas palavras de Bowers et al. (1997) o seguro possui duas limitações básicas:

- Embora os outros tipos de perda são importantes, o seguro não é capaz de amenizá-las. O seguro está restrito em reduzir as consequências de eventos aleatórios que podem ser mensurados em termos monetários.
- O seguro não é capaz de reduzir, pelo menos diretamente, a probabilidade da ocorrência do evento aleatório danoso. Isto é, o ato de contratar um seguro não implica na redução da probabilidade de ocorrer uma perda. No entanto, um sistema de seguros bem planejado frequentemente oferece incentivos financeiros para atividades que previnam a ocorrência de tal perda.

No artigo 757 do Código Civil (BRASIL, 2002) pode-se encontrar o conceito de seguro disposto como: “Pelo contrato de seguro, o segurador se obriga, mediante o pagamento do prêmio, a garantir interesse legítimo do segurado, relativo a pessoa ou coisa, contra riscos predeterminados” (BRASIL, 2012, p.228).

Há três elementos do seguro que são considerados como fundamentais: o risco, a mutualidade e a boa-fé (CAVALIERI FILHO, 2014). Por risco, entende-se como qualquer situação em que exista a incerteza em um acontecimento futuro danoso (HARRINGTON; NIEHAUS; HARRINGTON, 1999). Dado este aspecto, a ocorrência deste acontecimento torna-se independente do anseio das partes e é o grande elo para existência do seguro. Não há motivo para um indivíduo contratar um seguro cuja ocorrência futura de um evento danoso não seja possível.

O mutualismo consiste em um determinado grupo de pessoas, que estejam expostas aos mesmos riscos, contribuírem para a formação de um fundo para reparar as consequências de sinistros que afetem qualquer membro desse grupo. Portanto, a mutualidade representa o suporte econômico em toda operação de seguro, conferindo ao segurador o papel de administrador de fundos mutuais, que recebe os prêmios e paga as indenizações, cobrando por isso um percentual (CAVALIERI FILHO, 2014).

Finalmente, o que Cavalieri Filho (2014) considera como o elemento mais

importante do seguro: a boa-fé. Esse é o elemento jurídico do seguro que se define como a intenção isenta de dolo ou malícia de ambas as partes envolvidas, o segurado e o segurador. A boa-fé oferece viabilidade ao seguro, pois sem ela o seguro se torna dependente de um processo burocrático de fiscalização.

Por outro lado, nas palavras de Souza (2007) os elementos essenciais do seguro são mutualismo, incerteza e previdência. O mutualismo já foi descrito. A incerteza constitui um aspecto importante por estar presente na possibilidade de ocorrência do evento e no momento. E a previdência representa a necessidade que as pessoas tem de proteger a si próprias e a seus bens.

## **2.2 Evolução, princípios e estrutura do sistema de seguros no Brasil**

Na metade do século XIX, foi inaugurado, com a Lei n.556 de 25 de junho de 1850, o Código Comercial Brasileiro. Alguns artigos dessa legislação foram revogados pela lei n. 10.406/2002, o Código Civil, mas os artigos 666 a 730 permaneceram em vigor. Esses artigos disciplinam o seguro marítimo e datam as primeiras normas a respeito de seguro da história do Brasil. Antes disso, os contratos seguiam normas gerais, particularmente, de origem romana. Embora a legislação, nesse período, tratasse apenas de seguros marítimos, os seguros terrestres eram utilizados, ainda que sem normas que os regulassem (SILVA, 2012).

Silva (2012) relata que, durante o século XVIII, surgiram as primeiras companhias seguradoras brasileiras: Boa Fé (1808), Conceito Público (1808) e Indemnidade (1810). Os seguros terrestres passaram a ter legislação própria somente com a promulgação do Código Civil de 1916. Atualmente, no Direito do Seguro, os princípios mais importantes são:

- Princípio da previdência: relativo à necessidade do homem em se preservar, preservar seu patrimônio e as pessoas.
- Princípio do mutualismo: a união em grupos que cooperam mutuamente buscando a garantia de recomposição do patrimônio de um membro que

tenha sido vitimado.

- Princípio da pulverização dos riscos: fracionamento dos riscos entre diversas pessoas para que o evento possa ser suportado.
- Princípio da garantia: aqueles que assumem o risco apenas garantem a recomposição dos danos para a vítima do dano.
- Princípio da função socioeconômica: visa proteger a sociedade e a economia, buscando preservar famílias da ruína e manter empresas.
- Princípio da licitude do interesse segurado: confere legitimidade somente para operações securitárias lícitas.

Em 21 de novembro de 1966, foi instaurado o Decreto-Lei n.73, chamado Lei do Seguro, que dispõe sobre o Sistema Nacional de Seguros Privados (SNSP) e regula as operações de seguros e resseguros. Essa lei é considerada lei complementar por ter sido recepcionada na Constituição de 1988 (MARTINS, 2003).

Para a composição do SNSP, tem-se os seguintes órgãos, cada um com sua função específica (SILVA, 2012):

- **Conselho Nacional de Seguros Privados (CNSP)**: disciplina as regras gerais do SNSP, regula a organização, a composição e o funcionamento das empresas seguradoras e fiscaliza as características gerais dos contratos de seguros. Também, disciplina sobre as sanções legais aplicáveis, julga e processa em segunda instância os recursos das decisões provenientes da Superintendência de Seguros Privados.
- **Superintendência de Seguros Privados (SUSEP)**: fiscaliza a constituição, organização, funcionamento e as operações das empresas de seguros privados.
- **Resseguradores**: sociedades empresárias que realizam operações de resseguro e retrocessão.

- **Sociedades seguradoras autorizadas a operar em seguros privados:** aquelas que se obrigam, perante pagamento do prêmio, a garantir interesse legítimo do segurado, relativo a pessoa ou a coisa, contra riscos predeterminados.
- **Corretores de seguros habilitados:** exercem atividade de intermediação entre os segurados e as seguradoras.

### 2.3 Espécies de seguro

Há riscos em diversos tipos de atividades que podem ser objetos de cobertura, fornecendo diversas espécies de seguro. Dentre essas variadas espécies de seguro existentes, Cavalieri Filho (2014) descreve que as mais comuns são:

- **Seguro de coisas:** também é chamado de seguro de danos e objetiva cobrir os riscos de um bem que faz parte do patrimônio do segurado, por exemplo: seu automóvel, sua casa, entre outros. Esse seguro está descrito nos artigos 778 a 788 do Código Civil, que determina que nos seguros de dano a indenização deve ser suficiente somente para cobrir o dano sofrido. De modo que todo pagamento a mais caracteriza enriquecimento sem causa e estimula a fraude e a especulação (BRASIL, 2002).
- **Seguro de pessoas:** a cobertura é sobre riscos inerentes à pessoa. Tem-se, por exemplo, o seguro de acidentes pessoais e de invalidez por enfermidade, em que o beneficiário é o próprio segurado. Em contrapartida, há o seguro de vida, individual ou coletivo, em que, caso ocorra a morte do segurado, o beneficiário será um membro da família. Ao contrário do seguro de danos, o valor da indenização do seguro de pessoas não tem limitações, exceto o limite estabelecido pelo segurador, o que não impede que o segurado faça outros contratos sobre o mesmo risco com diferentes seguradoras. O Código Civil dispõe sobre o seguro de pessoa nos artigos 789 a 802 (BRASIL, 2002).

- **Seguro de saúde:** é uma espécie de seguro de pessoas que visa a cobertura de serviços de assistência médica na iniciativa privada, constituindo os planos privados de saúde. É importante destacar que além das companhias seguradoras, os planos de saúde também são operados por cooperativas de serviços médicos e por Medicina de Grupo. O que diferencia o seguro de saúde dos planos de saúde em geral “é o fato de ser operado por companhia de seguro mediante regime de livre escolha de médicos e hospitais e reembolso das despesas médico-hospitalares nos limites da apólice”(CAVALIERI FILHO, 2014, p.447).
- **Seguro de responsabilidade civil:** cobre os riscos de perdas e danos causados pelo segurado a terceiros, constituindo uma subespécie do seguro de danos.

## 2.4 Seguro de automóvel no Brasil

De acordo com Pita e Dominguez (2011), o desenvolvimento do seguro de automóveis no Brasil ocorreu, principalmente, depois que a SUSEP editou a Circular n.27 em 1986, o que conferiu mais liberdade para as seguradoras, com a liberação da tarifa de automóveis. Os autores apontam quatro evoluções importantes no ramo do seguro de automóveis. A primeira se relaciona à indenização no caso de perda total, que se baseava no valor de mercado do veículo, limitando-se à importância segurada. Com o aumento da inflação, o valor de mercado superava a importância segurada e a indenização não cobria o prejuízo. Então, incluiu-se uma cláusula de correção monetária da importância segurada. A segunda evolução ocorreu com a implantação do Código de Defesa do Consumidor (CDC). As seguradoras passaram a entregar aos consumidores manuais de mais fácil entendimento do que os contratos de produtos, facilitando o acesso aos produtos.

O terceiro avanço, apontado por Pita e Dominguez (2011), foi a inclusão do serviço de assistência ao veículo e aos passageiros, no início da década de 1990. Um serviço de baixo custo que ampliou a relação da seguradora com o

cliente. Finalmente, o quarto elemento na evolução do seguro de automóveis foi a especificação das características dos condutores na precificação. Isso permitiu que novos segurados pudessem adquirir o produto e ampliou a competição entre as seguradoras.

A cobertura dos seguros de automóveis se estende aos veículos automotores terrestres, com duas ou mais rodas, circulantes em vias públicas. Porém, reboques e tratores emplacados que rodem em vias públicas, também, podem ser segurados. De acordo com Pita e Dominguez (2011) o seguro de automóveis cobre os seguintes riscos: .

- Automóveis (casco): colisão, capotagem, queda de objetos sobre o veículo, queda sobre o veículo da carga transportada por ele, queda acidental em precipícios, de pontes ou viadutos, explosão, queda de raio, roubo ou furto qualificado do veículo ou de partes dele, submersão, queda de granizo e incêndio.
- Responsabilidade Civil Facultativa de Veículos (danos a terceiros): dano material, corporal ou moral causado a terceiros pelo veículo segurado ou reboque ou semirreboque atrelado ao veículo segurado.
- Acidentes pessoais a passageiros: morte, invalidez permanente ou despesas médicas ou hospitalares.

A contratação do seguro diante dos riscos apresentados pode se realizar com cobertura básica e também podem ser contratadas coberturas adicionais. Em qualquer cobertura, perdas parciais estão sujeitas à franquia, um valor predeterminado, que se o prejuízo for inferior, não há indenização. Se, por outro lado, o valor da reparação do dano for superior, o segurado paga o valor da franquia e a seguradora indeniza o valor restante (PITA; DOMINGUEZ, 2011).

Em 1974, a Lei 6.194 criou o Seguro de Danos Pessoais Causados por Veículos Automotores Terrestres (DPVAT). Esse seguro objetiva garantir que, em caso de acidentes causados por veículos ou por suas cargas, as vítimas recebam

uma indenização, quando há morte ou invalidez permanente, ou reembolso de despesas médicas.

O gráfico da Figura 1 mostra a evolução do prêmio direto ao longo dos anos de 2003 a 2014 para o ramo seguro de automóvel (casco), no Brasil. O valor de prêmio direto que em 2003 era de aproximadamente 7 bilhões de reais, em 2014 aumentou para 21,8 bilhões. Observa-se um crescimento de ano para ano, com exceção do ano de 2010 para 2011, em que houve uma queda de 15,2 bilhões de reais para 14,9 bilhões.

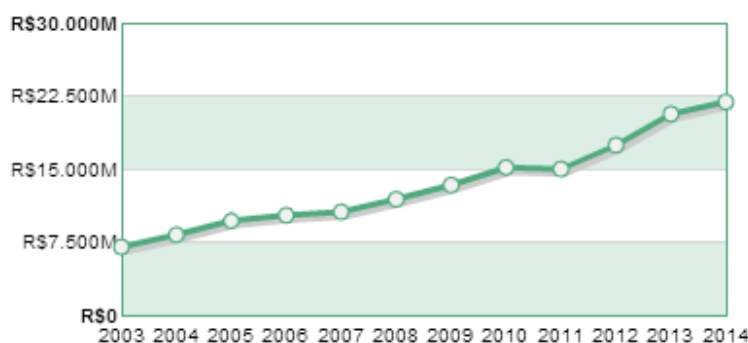


Figura 1 Evolução do Prêmio Direto Anual

A Figura 2 apresenta a distribuição do volume de prêmios diretos entre os principais segmentos de seguro no ano de 2013. Observa-se que o segmento de seguros de automóveis apresentava o segundo maior volume de prêmios, representando 20,2% do mercado.

O seguro de automóveis perde apenas para o segmento Vida Gerador de Benefício Livre (VGBL), que é um produto de acumulação de recursos, um plano previdenciário, inserido no âmbito dos seguros de pessoas pela SUSEP. Esse representava 42,8% do mercado, em 2013.

Na Figura 2 também se pode observar a participação do seguro obrigatório DPVAT, que representava, no ano de 2013, apenas 2,9% do mercado de seguros.

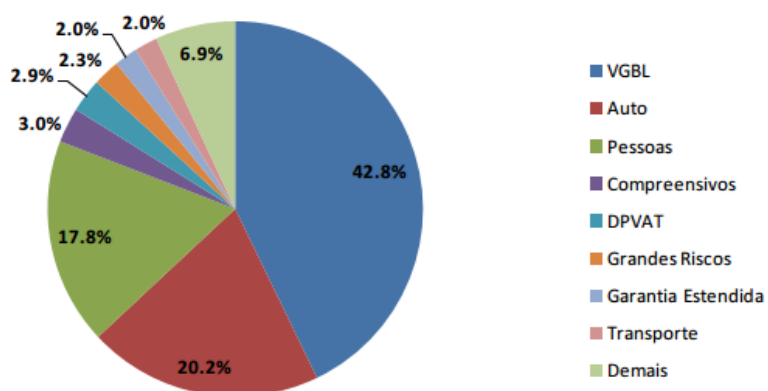


Figura 2 Principais segmentos do seguro - 2013

#### 2.4.1 Cálculo de Prêmio

Segundo Pita e Dominguez (2011), o cálculo do prêmio é baseado no risco ao qual o bem segurado está exposto, tais como suas características, coberturas e o período de cobertura. Os autores ainda salientam que as variáveis mais utilizadas para explicar as características do risco e auxiliar na precificação do prêmio de seguro de automóveis são:

- região de circulação habitual do veículo;
- modelo do veículo;
- ano do veículo;
- gênero e idade do condutor;
- bônus.

Por meio do tratamento estatístico dessas variáveis as seguradoras elaboraram a tarifação dos seguros de automóveis. São diversos os conceitos e metodologias envolvidos no cálculo do prêmio pago pelo segurado. No processo de precificação atuarial do custo de um seguro, destaca-se o método de tarifação baseado



na estimativa da perda esperada com indenizações. Em outras palavras, o valor do prêmio a ser calculado pela seguradora é uma função da variável aleatória da perda, isto é, se  $X$  for a variável aleatória risco da ocorrência de um determinado sinistro, então, um princípio de prêmio  $P$  é calculado de modo que,  $P = g(X)$ , em que  $g(\cdot)$  é alguma função (DICKSON, 2010).

Dickson (2010), Kaas et al. (2008), Denuit et al. (2006), apresentam vários princípios para o cálculo do prêmio, bem como, as propriedades desejáveis para que se possa classificar tais princípios.

A seguir, apresentam-se modelos que, frequentemente, auxiliam na precificação dos seguros de automóveis.

### 3 MODELO LINEAR GENERALIZADO

Os Modelos Lineares Generalizados (GLM) são uma classe de modelos estatísticos que descrevem uma relação entre a média de uma variável resposta e suas variáveis explicativas. Tais modelos, como o próprio nome sugere, são naturalmente uma generalização dos modelos lineares clássicos (MCCULLAGH; NELDER, 1989).

Será aqui apresentado apenas as principais ideias de um modelo linear generalizado. Ao leitor interessado em estudar tais modelos de uma forma mais abrangente e completa, recomendam-se as obras de McCullagh e Nelder (1989), Agresti (2013), Dobson e Barnett (2008) e Cordeiro e Demétrio (2010). Para aqueles que buscam estudar tais modelos dentro do contexto de seguros e atuária, recomendam-se as obras de Jong, Heller et al. (2008), Frees, Derrig e Meyers (2014), Kaas et al. (2008) e Ohlsson e Johansson (2010).

Um modelo linear generalizado é constituído por três componentes: o componente aleatório, o componente sistemático e uma função de ligação.

1. **Componente Aleatório:** é constituído por um vetor de observações  $\mathbf{y} = [y_1, y_2, \dots, y_n]^\top$  assumido ser uma realização do vetor aleatório  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]^\top$ , cujas componentes são independentemente distribuídas e tendo como média  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^\top$ . Ademais, a distribuição das variáveis aleatórias do vetor  $\mathbf{Y}$  pertence a família exponencial de distribuições. Resumidamente, temos

$$Y_i \sim \text{indep. } f_{Y_i}(y_i), \quad i = 1, \dots, n.$$

$$f_{Y_i}(y_i|\gamma, \tau) = \exp \left\{ [y_i\gamma_i - b(\gamma_i)]/\tau^2 - c(y_i, \tau) \right\}. \quad (1)$$

A distribuição dos  $Y_i$ , dada pela equação (1), é conhecida como a forma canônica da família exponencial, com as funções  $b(\cdot)$ ,  $c(\cdot, \cdot)$  conhecidas,  $\tau$  é o parâmetro de escala e  $\gamma_i$  denominado o parâmetro canônico ou natural da

distribuição de  $Y_i$ .

2. **Componente Sistemático:** relaciona um vetor  $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_n]^\top$  de variáveis explicativas através de um modelo linear. Seja  $x_{ij}$ ,  $i = 1, \dots, n$  e  $j = 1, \dots, p$  a  $j$ -ésima variável explicativa da  $i$ -ésima observação. Então,

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}. \quad (2)$$

Essa combinação linear de variáveis explicativas é nomeada de preditor linear e os  $\beta_j$  são os parâmetros cujos valores são desconhecidos e devem ser estimados através dos dados.

Matricialmente, escrevemos

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad (3)$$

em que,

$$\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

A matriz  $\mathbf{X}$  é conhecida e denominada de matriz modelo, cujo número de linhas e colunas é dado pelas  $n$  observações e pelas  $p$  variáveis explicativas, respectivamente. Enquanto que  $\boldsymbol{\beta}$  é o vetor dos  $p$  parâmetros desconhecidos.

3. **Função de Ligação:** como o próprio nome diz, faz a ligação entre os componentes aleatório e sistemático por meio de uma função  $g$  monótona e diferenciável. Seja  $E[Y_i] = \mu_i$ ,  $i = 1, \dots, n$  e  $\eta_i$  como definido em (2) a função de ligação  $g$  especifica que  $\mu_i$  e  $\eta_i$  são, de fato, iguais se escrevermos

$$\eta_i = g(\mu_i). \quad (4)$$

Em suma, temos que um modelo linear generalizado pode ser descrito como

$$Y_i \sim \text{indep. } f_{Y_i}(y_i)$$

$$E[Y_i] = \mu_i = g^{-1}(\eta_i),$$

ou, equivalentemente

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

No modelo linear normal tem-se que  $E[\mathbf{Y}] = \boldsymbol{\mu} = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ . Traçando esse para o contexto de modelos lineares generalizados, tem-se, nesse caso, uma função de ligação  $g(\cdot)$  identidade. Para este modelo, uma função de ligação identidade é plausível, porque tanto a média  $\boldsymbol{\mu}$ , quanto o preditor linear  $\boldsymbol{\eta}$  podem assumir quaisquer valores na reta real. No entanto, para dados de contagem cuja distribuição é de Poisson, necessariamente, tem-se que a média  $\boldsymbol{\mu} > 0$ , logo, a função de ligação identidade não é interessante para esse modelo, uma vez que o preditor linear  $\boldsymbol{\eta}$  pode assumir valores negativos, enquanto que a média  $\boldsymbol{\mu}$  não. Uma função de ligação logarítmica para o modelo Poisson é de maior interesse, pois  $\log(\boldsymbol{\mu}) = \boldsymbol{\eta}$  que aplicando a função inversa tem-se  $\boldsymbol{\mu} = e^{\boldsymbol{\eta}}$ , isto é, o preditor linear pode ser negativo, enquanto que a média sempre será positiva. Algumas vantagens do modelo linear generalizado são apresentadas na seção 3.1.

Cabe ressaltar que, a escolha da função de ligação está vinculada ao tipo da resposta  $Y$ , e varia de modelo para modelo. Em especial, tem-se o interesse na função de ligação que coincide com o parâmetro canônico, pois esta apresenta algumas vantagens teóricas como: simplificação do algoritmo de estimação e os parâmetros não necessitam de transformação (AGRESTI, 2013).

### 3.1 Vantagens do Modelo Linear Generalizado Sobre Dados Transformados

Uma maneira tradicional para modelar os dados é aplicar uma transformação em  $Y_i$  de tal modo que a variável resposta siga aproximadamente distribuição normal com variância constante para que, então, seja possível aplicar regressão via quadrados mínimos ordinários (AGRESTI, 2013). Em modelos lineares generalizados, uma das principais ideias é evitar a transformação dos dados, tendo como alternativa a aplicação de uma função de ligação na média de uma resposta para se ajustar o modelo resultante via máxima verossimilhança (MCCULLOCH; SEARLE; NEUHAUS, 2008). Desta forma, como o processo maximiza a verossimilhança para a escolha da distribuição de  $Y_i$ , não é imprescindível que se tenha normalidade e variância constante.

Uma outra vantagem da formulação por meio de modelos lineares generalizados é que os parâmetros do modelo descrevem  $g(E[Y_i])$ , em vez de  $E[g(Y_i)]$  como acontece quando se aplica alguma transformação nos dados. Consequentemente, quando se aplica a função inversa de  $g$ , os parâmetros descrevem os efeitos das variáveis explicativas na  $E[Y_i]$ , o que é muito mais relevante que os efeitos das variáveis explicativas na  $E[g(Y_i)]$ , que ocorre quando se transforma os dados (AGRESTI, 2013).

### 3.2 A Média e a Variância de um Modelo Linear Generalizado

Para a caracterização da média e a variância de um modelo linear generalizado com variável resposta  $y_i$  é necessária a apresentação do teorema a seguir.

**Teorema 1.** *Seja  $f(y|\theta)$  uma função de densidade probabilidade que satisfaz a seguinte igualdade*

$$\frac{d}{d\theta} E_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] = \int \frac{\partial}{\partial \theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right) f(y|\theta) \right] dy,$$

(também verdade para a família exponencial), então

$$E_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] = 0 \quad (5)$$

e

$$\text{Var}_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] = E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right] = -E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(y|\theta) \right]. \quad (6)$$

*Demonstração.* Por definição é sabido que

$$\begin{aligned} E_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] &= \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right) f(y|\theta) dy \\ &= \int_{-\infty}^{\infty} \left( \frac{\frac{\partial}{\partial \theta} f(y|\theta)}{f(y|\theta)} \right) f(y|\theta) dy \\ &= \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial \theta} f(y|\theta) \right) dy \\ &= \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(y|\theta) dy \\ &= \frac{\partial}{\partial \theta} [1] \\ &= 0. \end{aligned}$$

Também é conhecido que

$$\begin{aligned} \text{Var}_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] &= E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right] - \left( E_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(y|\theta) \right] \right)^2 \\ &= E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right] - 0 \\ &= E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right]. \end{aligned}$$

Por fim, tem-se que

$$\begin{aligned}
E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f(y|\theta) \right] &= \\
&= \int_{-\infty}^{\infty} \left( \frac{\partial^2}{\partial \theta^2} \log f(y|\theta) \right) f(y|\theta) dy \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right) f(y|\theta) dy \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \left( \frac{\frac{\partial}{\partial \theta} f(y|\theta)}{f(y|\theta)} \right) f(y|\theta) dy \\
&= \int_{-\infty}^{\infty} \left\{ \frac{\left[ \frac{\partial^2}{\partial \theta^2} f(y|\theta) \right] f(y|\theta) - \frac{\partial}{\partial \theta} f(y|\theta) \frac{\partial}{\partial \theta} f(y|\theta)}{(f(y|\theta))^2} \right\} f(y|\theta) dy \\
&= \int_{-\infty}^{\infty} \frac{\left[ \frac{\partial^2}{\partial \theta^2} f(y|\theta) \right] f(y|\theta)}{(f(y|\theta))^2} f(y|\theta) dy - \int_{-\infty}^{\infty} \frac{\left[ \frac{\partial}{\partial \theta} f(y|\theta) \right]^2}{(f(y|\theta))^2} f(y|\theta) dy \\
&= \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} f(y|\theta) dy - \int_{-\infty}^{\infty} \left\{ \frac{\left[ \frac{\partial}{\partial \theta} f(y|\theta) \right]}{(f(y|\theta))} \right\}^2 f(y|\theta) dy \\
&= \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(y|\theta) dy - \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 f(y|\theta) dy \\
&= \frac{\partial^2}{\partial \theta^2} [1] - \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 f(y|\theta) dy \\
&= -E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right].
\end{aligned}$$

□

Com o auxílio do Teorema 1, obtemos a  $E[Y_i]$  e  $Var[Y_i]$ . Veja que, partindo da equação (5), temos:

$$\begin{aligned}
E \left[ \frac{\partial}{\partial \gamma_i} \log f_{Y_i}(y_i | \gamma_i, \tau) \right] &= E \left[ \frac{\partial}{\partial \gamma_i} \log (\exp \{ [y_i \gamma_i - b(\gamma_i)] / \tau^2 - c(y_i, \tau) \}) \right] \\
&= E \left[ \frac{\partial}{\partial \gamma_i} \{ [y_i \gamma_i - b(\gamma_i)] / \tau^2 - c(y_i, \tau) \} \right] \\
&= E \left[ \left\{ y_i - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right\} / \tau^2 \right] \\
&= \frac{1}{\tau^2} E \left[ \left\{ y_i - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right\} \right].
\end{aligned}$$

Consequentemente,

$$\begin{aligned}
\frac{1}{\tau^2} E \left[ \left\{ y_i - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right\} \right] &= 0 \\
\Rightarrow E \left[ \left\{ y_i - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right\} \right] &= 0 \\
\therefore E[y_i] = \mu_i &= \frac{\partial b(\gamma_i)}{\partial \gamma_i}. \tag{7}
\end{aligned}$$

A variância é obtida através da equação (6), que nos permite igualar

$$\begin{aligned}
Var \left[ \left\{ y_i - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right\} / \tau^2 \right] &= -E \left[ -\frac{1}{\tau^2} \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} \right] \\
\Rightarrow Var \left[ \frac{y_i - \mu_i}{\tau^2} \right] &= \frac{1}{\tau^2} \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} \\
\therefore Var[y_i] &= \tau^2 \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} = \tau^2 v(\mu_i). \tag{8}
\end{aligned}$$

A função  $v(\mu_i) = \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2}$  é nomeada função de variância, uma vez que indica como a variância de  $y_i$  depende da esperança de  $y_i$ .



A Tabela 1 resume as principais distribuições que pertencem à família exponencial.

Tabela 1 Elementos de algumas distribuições da família exponencial

	Normal	Poisson	Binomial	Gama
Notação	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)/m$	$G(\mu, \nu)$
$\tau$	$\sigma^2$	1	$1/m$	$\nu^{-1}$
$b(\gamma)$	$\gamma^2/2$	$\exp(\gamma)$	$\log(1 + e^\gamma)$	$-\log(-\gamma)$
$c(y; \tau)$	$-\frac{1}{2} \left( \frac{y}{\tau} + \log(2\pi\tau) \right)$	$-\log(y!)$	$\log \binom{m}{my}$	$\nu \log(\nu y) - \log(y) - \log(\Gamma(\nu))$
$E_\gamma[Y]$	$\gamma$	$\exp(\gamma)$	$e^\gamma / (1 + e^\gamma)$	$-1/\gamma$
ligação canônica $g(\mu)$	$\mu$	$\log(\mu)$	$\log \left( \frac{\mu}{1-\mu} \right)$	$1/\mu$
$v(\mu)$	1	$\mu$	$\mu(1-\mu)$	$\mu^2$

Fonte: (MCCULLAGH; NELDER, 1989)

### 3.3 Estimação via Máxima Verossimilhança

Antes de ser apresentada a função de verossimilhança, são estabelecidas duas identidades úteis que nos auxiliam no processo de estimação.

As duas identidades úteis são

$$\begin{aligned}
 \frac{\partial \gamma_i}{\partial \mu_i} &= \left( \frac{\partial \mu_i}{\partial \gamma_i} \right)^{-1} \\
 &= \left[ \frac{\partial}{\partial \gamma_i} \left( \frac{\partial b(\gamma_i)}{\partial \gamma_i} \right) \right]^{-1} \\
 &= \left[ \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} \right]^{-1} \\
 &= \frac{1}{v(\mu_i)}
 \end{aligned} \tag{9}$$

e

$$\begin{aligned}
\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} &= \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \frac{\partial g(\mu_i)}{\partial g(\mu_i)} \\
&= \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \boldsymbol{\beta}} \\
&= \left( \frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} \left( \mathbf{x}_i^\top \boldsymbol{\beta} \right) \\
&= \left( \frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \mathbf{x}_i^\top.
\end{aligned} \tag{10}$$

A função de verossimilhança é dada por

$$\begin{aligned}
L(\boldsymbol{\gamma}_i | y_i) &= \prod_{i=1}^n f_{Y_i}(y_i | \boldsymbol{\gamma}_i, \tau) \\
&= \prod_{i=1}^n \exp \{ [y_i \boldsymbol{\gamma}_i - b(\boldsymbol{\gamma}_i)] / \tau^2 - c(y_i, \tau) \} \\
&= \exp \left\{ \sum_{i=1}^n [y_i \boldsymbol{\gamma}_i - b(\boldsymbol{\gamma}_i)] / \tau^2 - c(y_i, \tau) \right\}.
\end{aligned}$$

Para a obtenção das equações de verossimilhança, pode-se utilizar a função log-verossimilhança  $l = \log(L(\boldsymbol{\gamma}_i | y_i))$ , de forma que

$$l = \sum_{i=1}^n [y_i \boldsymbol{\gamma}_i - b(\boldsymbol{\gamma}_i)] / \tau^2 - \sum_{i=1}^n c(y_i, \tau).$$

Derivando  $l$  em relação a  $\boldsymbol{\beta}$ , tem-se

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\tau^2} \sum_{i=1}^n \left[ y_i \frac{\partial \boldsymbol{\gamma}_i}{\partial \boldsymbol{\beta}} - \frac{\partial b(\boldsymbol{\gamma}_i)}{\partial \boldsymbol{\gamma}_i} \frac{\partial \boldsymbol{\gamma}_i}{\partial \boldsymbol{\beta}} \right]$$

Utilizando o resultado (7)

$$\begin{aligned}
 &= \frac{1}{\tau^2} \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \\
 &= \frac{1}{\tau^2} \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \frac{\partial \mu_i}{\partial \mu_i} \\
 &= \frac{1}{\tau^2} \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}
 \end{aligned}$$

utilizando os resultados (9) e (10)

$$\begin{aligned}
 &= \frac{1}{\tau^2} \sum_{i=1}^n (y_i - \mu_i) \frac{1}{v(\mu_i)} \frac{1}{g_\mu(\mu_i)} \mathbf{x}_i^\top \\
 &= \frac{1}{\tau^2} \sum_{i=1}^n (y_i - \mu_i) w_i g_\mu(\mu_i) \mathbf{x}_i^\top, \tag{11}
 \end{aligned}$$

em que  $w_i = [v(\mu_i)g_\mu^2(\mu_i)]^{-1}$ .

Matricialmente,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu})$$

com  $\mathbf{W} = \text{Diag}\{w_i\}$  e  $\boldsymbol{\Delta} = \text{Diag}\{g_\mu(\mu_i)\}$ ,  $i = 1, \dots, n$ .

Dessa forma as equações de máxima verossimilhança são dadas por

$$\mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} \mathbf{y} = \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} \boldsymbol{\mu}, \tag{12}$$

em que  $\mathbf{W}$ ,  $\boldsymbol{\Delta}$  e  $\boldsymbol{\mu}$  envolvem o vetor  $\boldsymbol{\beta}$  desconhecido. Como essas funções são não lineares em  $\boldsymbol{\beta}$ , a igualdade em (12) não tem solução analítica.

A solução das equações de verossimilhança (12) para  $\boldsymbol{\beta}$  são, frequentemente, resolvidas através do Método Iterativo de Quadrados Mínimos Ponderados. Isso pode ser conseguido com a utilização do Método de Fisher (do inglês *Fisher Scoring Method*) (JENNRICH; SAMPSON, 1976), que é um método iterativo

para maximizar a função de verossimilhança e assume a forma

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \left[ \mathbf{I}(\boldsymbol{\theta}^{(m)}) \right]^{-1} \left. \frac{\partial l}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}}, \quad (13)$$

nas quais  $(m)$  indica a  $m$ -ésima iteração,  $\mathbf{I}(\boldsymbol{\theta})$  denota a matriz da informação de Fisher e  $\boldsymbol{\theta}$  é o vetor de parâmetros. A matriz da informação  $\mathbf{I}(\boldsymbol{\theta})$  é definida como

$$\mathbf{I}(\boldsymbol{\theta}) = E \left[ \frac{\partial l}{\partial \boldsymbol{\theta}} \frac{\partial l}{\partial \boldsymbol{\theta}^\top} \right].$$

Outra forma muito útil de  $\mathbf{I}(\boldsymbol{\theta})$  é, sob certas condições de regularidade (SEARLE; CASELLA; MCCULLOCH, 2009), dada por

$$\mathbf{I}(\boldsymbol{\theta}) = -E \left[ \frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right]. \quad (14)$$

Aplicando este método para  $\boldsymbol{\beta}$  temos o seguinte resultado

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} \left[ \mathbf{I}(\boldsymbol{\beta}^{(m)}) \right]^{-1} \left. \frac{\partial l}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}} \\ &= \boldsymbol{\beta}^{(m)} + (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu}), \end{aligned} \quad (15)$$

pois,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu})$$

e

$$\begin{aligned}
\mathbf{I}(\boldsymbol{\beta}) &= -E \left[ \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right] \\
&= -E \left[ -\frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^\top} + \frac{1}{\tau^2} \mathbf{X}^\top \frac{\partial \mathbf{W} \boldsymbol{\Delta}}{\partial \boldsymbol{\beta}^\top} (\mathbf{y} - \boldsymbol{\mu}) \right] \\
&= \frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^\top} + \mathbf{0} \\
&= \frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} \boldsymbol{\Delta}^{-1} \mathbf{X}, \quad \text{usando o resultado (10)} \\
&= \frac{1}{\tau^2} \mathbf{X}^\top \mathbf{W} \mathbf{X}.
\end{aligned}$$

Para a equação (15), entende-se que  $\mathbf{W}$ ,  $\boldsymbol{\Delta}$  e  $\boldsymbol{\mu}$  são avaliados em  $\boldsymbol{\beta}^{(m)}$ .

Um outro resultado interessante é a relação entre a inversa da matriz da informação de Fisher e a variância assintótica de  $\boldsymbol{\beta}$ . Temos que,

$$Var_\infty[\hat{\boldsymbol{\beta}}] \approx [I(\hat{\boldsymbol{\beta}})]^{-1} = \tau(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}, \quad (16)$$

em que  $Var_\infty$  denota a variância assintótica.

Ainda, tem-se que o estimador de máxima verossimilhança  $\hat{\boldsymbol{\beta}}$  é assintoticamente não viesado e, para amostras grandes, tem distribuição aproximada normal  $p$ -variada. Desta forma,

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1}). \quad (17)$$

Os resultados (16) e (17) são a base do teste de significância de Wald, que poderia ser utilizado no caso de inferência fiducial, no entanto não vai ser o objetivo desse trabalho.

### 3.4 Função Desvio

Quando se considera a escolha de algum modelo para um conjunto de dados, deve-se definir uma amplitude de possibilidades. O *modelo nulo* é o mais simples dos modelos, enquanto que o mais complexo é frequentemente conhecido por *modelo saturado* ou *completo* (FARAWAY, 2005).

O modelo nulo possui apenas um parâmetro, sendo representado por um único  $\mu$  para todos os  $y_i$ . Tal modelo designa toda a variação entre os  $y_i$  para o componente aleatório. Já o modelo saturado possui  $n$  parâmetros – um para cada uma das  $n$  respostas  $y_i$  – e designa toda variação nos  $y_i$  para o componente sistemático não remanescendo nada para o componente aleatório.

Segundo McCullagh e Nelder (1989), o modelo saturado é não informativo, pois não fornece nenhum resumo dos dados. E não parcimonioso, porque possui muitos parâmetros. Todavia, tem grande utilidade no momento da escolha do modelo, pois fornece um indicativo de como deve ser o ajuste de um modelo com  $p < n$  parâmetros, assim sendo, utilizado para comparação com outros modelos ajustados.

A qualidade do ajuste de um modelo linear generalizado é medida através da função desvio ou simplesmente *deviance* (em inglês), que consiste em uma distância entre o modelo saturado (com  $n$  parâmetros) e o modelo sob investigação (com  $p$  parâmetros) avaliado na estimativa de máxima verossimilhança  $\hat{\beta}$  (AGRESTI, 2013).

Suponha um modelo linear generalizado com vetor de observações  $\mathbf{y} = [y_1, y_2, \dots, y_n]^\top$ , denote por  $l(\boldsymbol{\mu}|\mathbf{y})$  a função de log-verossimilhança especificada em termos das médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^\top$ . Denote ainda por  $l(\hat{\boldsymbol{\mu}}|\mathbf{y})$  a função de log-verossimilhança maximizada sobre  $\boldsymbol{\beta}$  para o modelo em observação. O modelo saturado é denotado por  $l(\mathbf{y}|\mathbf{y})$ . Desta forma, tem-se que

$$-2 \log \frac{L(\mathbf{y}|\hat{\boldsymbol{\mu}})}{L(\mathbf{y}|\mathbf{y})} = -2 [l(\hat{\boldsymbol{\mu}}|\mathbf{y}) - l(\mathbf{y}|\mathbf{y})] = \tau D(\mathbf{y}|\hat{\boldsymbol{\mu}}). \quad (18)$$

$D(\mathbf{y}|\hat{\boldsymbol{\mu}})$  é definida como função desvio. Um pequeno valor da função desvio significa que, com menos parâmetros, obtém-se um ajuste tão bom quanto o do modelo completo. Definam-se  $\tilde{\gamma}_i = \gamma_i(\tilde{\boldsymbol{\mu}})$  e  $\hat{\gamma}_i = \gamma_i(\hat{\boldsymbol{\mu}})$ , as estimativas de máxima verossimilhança para os modelos completo ( $p = n$ ) e incompleto ( $p < n$ ), respectivamente, tem-se que a função desvio é dada por

$$D(\mathbf{y}|\hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i[\tilde{\gamma}_i - \hat{\gamma}_i] + [b(\hat{\gamma}_i) - b(\tilde{\gamma}_i)]\}.$$

Na Tabela 2, apresenta-se a função desvio para os modelos mais conhecidos.

Tabela 2 Função Desvio para os principais Modelos Lineares Generalizados

GLM	Função Desvio
Normal	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Poisson	$2 \sum_{i=1}^n [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]$
Binomial	$2 \sum_{i=1}^n \{y_i \log(y_i/\hat{\mu}_i) + (m - y_i) \log[(m - y_i)/(m - \hat{\mu}_i)]\}$
Gama	$2 \sum_{i=1}^n \{-\log(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}$
Normal Inversa	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (\hat{\mu}_i^2 y_i)$

#### 4 MODELO LINEAR MISTO

A maioria dos estudos práticos, nos mais distintos campos da ciência, envolve algum tipo de modelo de regressão, com a finalidade de descrever comportamentos, identificar fatores que afetam/explicam algum fenômeno. No entanto, uma grande gama desses modelos utilizam suposição de que as observações são in-

dependentes e é comum em muitas áreas da ciência, como por exemplo, economia, saúde, genética, agrária, a ocorrência de dados que possuem uma dependência entre si, nomeadamente, dados agrupados. Sendo assim, os modelos lineares mistos tratam de uma estrutura mais complexa. As observações entre grupos ou classes são independentes, mas as observações dentro de cada grupo são dependentes, pois elas pertencem a mesma subpopulação. Dessa forma, tem-se dois tipos de variação: entre classes e intra-classes (DEMIDENKO, 2013).

Os modelos lineares mistos, ou modelos de efeitos mistos, ou, simplesmente, modelos mistos são modelos que apresentam tanto efeitos fixos quanto efeitos aleatórios e são usados, essencialmente, para descrever relações entre uma variável resposta e algumas variáveis explicativas em dados que são agrupados de acordo com um ou mais fatores de classificação (PINHEIRO; BATES, 2006). Exemplos de tipos de dados agrupados são: dados longitudinais, dados de medidas repetidas, dados multiníveis e planejamento em blocos.

Será apresentado, um breve resumo sobre modelos mistos. Ao leitor interessado em maiores detalhes sobre esses modelos, recomenda-se os trabalhos de McCulloch, Searle e Neuhaus (2008), Pinheiro e Bates (2006), Rencher e Schaalje (2008), Searle, Casella e McCulloch (2009) e Demidenko (2013).

#### 4.1 Modelo Geral

Em modelos mistos, alguns parâmetros não são tratados como constantes, mas sim, como variáveis aleatórias. Os efeitos fixos são usados para modelar a média de  $\mathbf{Y}$ , enquanto que efeitos aleatórios determinam a estrutura da variância-covariância de  $\mathbf{Y}$ . De fato, uma primeira razão para a inclusão de efeitos aleatórios é simplificar a árdua tarefa de especificar os  $n(n + 1)/2$  elementos distintos de  $Var[\mathbf{y}_{n \times 1}]$  (MCCULLOCH; SEARLE; NEUHAUS, 2008).

O modelo linear misto é especificado, de maneira mais conveniente, condicionalmente e é matricialmente definido por



$$E[\mathbf{Y}|\mathbf{U} = \mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \quad (19)$$

tal que

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1q} \\ z_{21} & z_{22} & \dots & z_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nq} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_q \end{bmatrix}.$$

Tem-se no modelo (19), o vetor  $\mathbf{y}$  referente à variável resposta;  $\boldsymbol{\beta}$  o vetor de parâmetros para os efeitos fixos, assim como  $\mathbf{u}$  é o vetor de parâmetros dos efeitos aleatórios. As matrizes  $\mathbf{X}$  e  $\mathbf{Z}$  denotam as matrizes modelo para os efeitos fixos e aleatórios, respectivamente. Costuma-se adotar letras romanas para os parâmetros dos efeitos aleatórios e letras gregas para os parâmetros dos efeitos fixos.

Por questão de simplicidade, define-se,  $E[\mathbf{Y}|\mathbf{U} = \mathbf{u}] = E[\mathbf{Y}|\mathbf{u}]$ .

Para obter  $E[\mathbf{Y}]$  e  $Var[\mathbf{Y}]$  é necessário especificar  $\mathbf{u}$ , de modo que

$$\mathbf{U} \sim (\mathbf{0}, \mathbf{D}), \quad (20)$$

isto é,  $\mathbf{U}$  tem distribuição cuja  $E[\mathbf{U}] = \mathbf{0}$  e  $Var[\mathbf{U}] = \mathbf{D}$ . Portanto,

$$\begin{aligned}
E[\mathbf{Y}] &= E[E[\mathbf{Y}|\mathbf{u}]] \\
&= E[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}] \\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}E[\mathbf{u}] \\
&= \mathbf{X}\boldsymbol{\beta}.
\end{aligned} \tag{21}$$

Para especificar  $Var[\mathbf{Y}]$ , é necessário que se defina também

$$Var[\mathbf{Y}|\mathbf{u}] = \mathbf{R}. \tag{22}$$

Consequentemente,

$$\begin{aligned}
Var[\mathbf{Y}] &= E[Var[\mathbf{Y}|\mathbf{u}]] + Var[E[\mathbf{Y}|\mathbf{u}]] \\
&= E[\mathbf{R}] + Var[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}] \\
&= \mathbf{R} + \mathbf{Z}Var[\mathbf{u}]\mathbf{Z}^\top \\
&= \mathbf{R} + \mathbf{Z}\mathbf{D}\mathbf{Z}^\top.
\end{aligned} \tag{23}$$

Dos resultados (21) e (23), confirma-se que efeitos fixos modelam a média, enquanto que efeitos aleatórios determinam a variância de  $\mathbf{Y}$ .

## 5 MODELO LINEAR GENERALIZADO MISTO

Os modelos lineares generalizados mistos são extensões dos modelos lineares generalizados com a inclusão de efeitos aleatórios. O uso desses modelos tem aumentado significativamente, principalmente, por estes terem a capacidade de modelar dados binários, de contagem, assim como dados agrupados e longitu-

dinais (ANTONIO; BEIRLANT, 2007).

## 5.1 Modelo Geral

Para especificar o modelo, inicia-se descrevendo a distribuição condicional do vetor resposta  $\mathbf{Y}$  dado os efeitos aleatórios  $\mathbf{u}$ . Assim como (1) e (4), assume-se, usualmente, mas não necessariamente, que os elementos de  $\mathbf{Y}$  sejam, condicionalmente independentes, cada um distribuído com densidade da forma da família exponencial ou aproximada. Portanto,

$$Y_i|\mathbf{u} \sim indep. f_{Y_i|\mathcal{U}}(y_i|\mathbf{u})$$

$$f_{Y_i|\mathcal{U}}(y_i|\mathbf{u}) = \exp\{[y_i\gamma_i - b(\gamma_i)]/\tau^2 - c(y_i, \tau)\}. \quad (24)$$

Como salienta McCulloch, Searle e Neuhaus (2008), por meio de (5) é sabido que a média condicional de  $Y_i$  é relacionada com  $\gamma_i$  em (24) através da identidade  $\mu_i = \partial b(\gamma_i)/\partial \gamma_i$ . É em uma transformação desta média que se tem o interesse em modelar linearmente, contendo tanto fatores fixos quanto aleatórios. Dessa forma,

$$E[Y_i|\mathbf{u}] = \mu_i$$

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u}. \quad (25)$$

em que,

- $g(\cdot)$  é uma função de ligação conhecida que conecta a média condicional de  $Y_i$  à forma linear dos preditores;
- $\mathbf{x}_i^\top$  é a  $i$ -ésima linha da matriz modelo para os efeitos fixos;
- $\boldsymbol{\beta}$  é o vetor de parâmetros dos efeitos fixos;

- $\mathbf{z}_i^\top$  que é a  $i$ -ésima linha da matriz modelo para os efeitos aleatórios;
- $\mathbf{u}$  é o vetor de efeitos aleatórios.

Cabe ressaltar que  $\mu_i$  denota a média condicional de  $Y_i|\mathbf{u}$ , diferentemente do caso dos modelos lineares generalizados, em que  $\mu_i$  representa a média incondicional de  $Y_i$ . Na mesma lógica, (8) representa, nesse contexto, a variância de  $Y_i|\mathbf{u}$ .

Para completar a especificação, assume-se uma distribuição para os efeitos aleatórios. Portanto,

$$U \sim f_U(\mathbf{u}). \quad (26)$$

Uma vez que a especificação do modelo em (24) e (25) é realizada condicionalmente ao valor de  $\mathbf{u}$ , é também interessante derivar aspectos da distribuição marginal de  $Y_i$ , com o propósito de entender as suposições assumidas para os dados observados.

A esperança de  $\mathbf{Y}$  pode ser obtida pela lei da esperança iterada:

$$\begin{aligned} E[Y_i] &= E[E[Y_i|\mathbf{u}]] \\ &= E[\mu_i] \\ &= E[g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u})]. \end{aligned} \quad (27)$$

Para a derivação da variância de  $\mathbf{Y}$ , utiliza-se o teorema da variância total (CASELLA; BERGER, 2002). Tem-se:

$$\begin{aligned}
\text{Var}[Y_i] &= \text{Var}[E[Y_i|\mathbf{u}]] + E[\text{Var}[Y_i|\mathbf{u}]] \\
&= \text{Var}[\mu_i] + E[\tau^2 v(\mu_i)] \\
&= \text{Var}[g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u})] + E[\tau^2 v(g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u}))]. \quad (28)
\end{aligned}$$

O uso de efeitos aleatórios introduz uma correlação entre as observações que possuam qualquer efeito aleatório em comum. Desta forma, a covariância entre duas observações é dada por:

$$\begin{aligned}
\text{Cov}(Y_i, Y_j) &= \text{Cov}(E[Y_i|\mathbf{u}], E[Y_j|\mathbf{u}]) + E[\text{Cov}(Y_i, Y_j|\mathbf{u})] \\
&= \text{Cov}(\mu_i, \mu_j) + E[0] \\
&= \text{Cov}(g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u}), g^{-1}(\mathbf{x}_j^\top \boldsymbol{\beta} + \mathbf{z}_j^\top \mathbf{u})). \quad (29)
\end{aligned}$$

## 5.2 Critérios de Informação para Seleção de Modelos

Dentre os critérios de informação utilizados para verificar a adequação dos modelos aos dados e também para a comparação entre modelos destacam-se o *Critério de Informação de Akaike* (AIC) (AKAIKE, 1974) e o *Critério Bayesiano de Schwarz* (BIC) (SCHWARZ et al., 1978). Tais critérios se baseiam no valor da log-verossimilhança avaliada no ponto de máxima verossimilhança e também contam com uma penalização para o número de parâmetros  $p$  estimados do modelo em questão. A ideia é obter um modelo parcimonioso que se ajuste bem aos dados. As estatísticas dos critérios são:

$$AIC = -2l(\hat{\theta}_{mle}|\mathbf{y}) + 2p \quad (30)$$

e

$$BIC = -2l(\hat{\theta}_{mle}|\mathbf{y}) + p \log n, \quad (31)$$

em que  $n$  define o tamanho da amostra e  $\hat{\theta}_{mle}$  é o estimador de máxima verossimilhança para  $\theta$ .

Em termos de seleção de modelos, diz-se que um modelo  $m_1$  é preferível ao modelo  $m_2$  pelo Critério de Akaike se  $AIC_{m_1} < AIC_{m_2}$ . O mesmo raciocínio vale para o BIC.

## 6 ESTATÍSTICA BAYESIANA

De acordo com Christensen et al. (2011), a análise bayesiana baseia-se na premissa de que toda incerteza deve ser modelada usando probabilidades e que a inferência estatística deve tirar conclusões lógicas baseadas nas leis de probabilidade.

Segundo Gelman et al. (2013), a característica essencial de métodos bayesianos é seu uso explícito da probabilidade para quantificar a incerteza em inferências baseadas na análise estatística dos dados.

A abordagem bayesiana da estatística é, fundamentalmente, diferente da abordagem clássica ou frequentista, que até então tem sido apresentada. Na abordagem frequentista, o parâmetro  $\theta$  é considerado desconhecido, porém sendo uma quantidade fixa. Resumidamente, uma amostra aleatória é retirada da população indexada por  $\theta$  e, baseado nos valores observados da amostra, obtém-se todo o conhecimento sobre o valor de  $\theta$ .

Já no paradigma bayesiano, o parâmetro  $\theta$  é considerado uma quantidade cuja variação pode ser descrita por uma distribuição de probabilidade, denominada *distribuição a priori*, ou simplesmente, *a priori*. A distribuição *a priori* é uma distribuição subjetiva, baseada na crença do experimentador. Uma amostra é coletada de uma população indexada por  $\theta$  e a distribuição *a priori* é atualizada

com a informação contida na amostra. A essa atualização, dá-se o nome de distribuição *a posteriori*, ou apenas *a posteriori* (CASELLA; BERGER, 2002). Através da distribuição *a posteriori* tiram-se conclusões sobre  $\theta$ .

De acordo com Paulino, Turkman e Murteira (2003), a possibilidade de se incorporar informações passadas no modelo é uma das vantagens em se utilizar do paradigma bayesiano.

O nome da corrente Estatística Bayesiana se deve ao fato de que o intermédio da atualização da priori para a posteriori é dado com o uso do Regra de Bayes. Tal regra é descrita a seguir.

Denote por  $p(\theta)$  a distribuição *a priori* e por  $p(\mathbf{y}|\theta)$  a distribuição amostral (ou distribuição dos dados), então a distribuição *a posteriori*, a distribuição condicional de  $\theta$  dado a amostra,  $\mathbf{y}$ , é

$$p(\theta|\mathbf{y}) = \frac{p(\theta, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\theta)p(\mathbf{y}|\theta)}{p(\mathbf{y})}, \quad (32)$$

em que,  $p(\mathbf{y}) = \sum_{\theta} p(\theta)p(\mathbf{y}|\theta)$  com a soma sobre todos os valores possíveis de  $\theta$  ou  $p(\mathbf{y}) = \int p(\theta)p(\mathbf{y}|\theta)d\theta$  no caso de  $\theta$  contínuo.

Outra forma equivalente de (32) omite  $p(\mathbf{y})$ , que não depende de  $\theta$  e, fixado  $\mathbf{y}$ , pode ser considerada uma constante normalizadora e é bastante útil em problemas que envolvem estimação de parâmetros. Desse modo, tem-se a *densidade a posteriori não normalizada*, dada pelo lado direito da seguinte expressão

$$p(\theta|\mathbf{y}) \propto p(\theta)p(\mathbf{y}|\theta). \quad (33)$$

Cabe salientar que a distribuição  $p(\mathbf{y}|\theta)$  é uma função de  $\theta$  ao invés de  $\mathbf{y}$  e é nomeada *função de verossimilhança*, para  $\mathbf{y}$  fixo, que, por sua vez, fornece a *verossimilhança* ou *plausibilidade* de cada um dos possíveis valores de  $\theta$  (GELMAN et al., 2013).

Desse modo, a função de verossimilhança tem função determinante em (32) e (33), uma vez que representa a maneira pela qual os dados  $\mathbf{y}$  contribuem para a informação sobre  $\theta$  (PAULINO; TURKMAN; MURTEIRA, 2003).

Em suma, a distribuição *a posteriori* é construída via Regra de Bayes, composta pela informação prévia do pesquisador, descrita pela distribuição *a priori* e pela informação fornecida pelos dados através da função de verossimilhança.

A distribuição *a priori* mais simples do ponto de vista matemático da inferência bayesiana é chamada de *priori conjugada* e ocorre quando as distribuições *a priori* e *a posteriori* pertencem a mesma família de distribuições (CASELLA; BERGER, 2002). Quando se usa uma *priori* conjugada, a integração necessária para obter a constante normalizadora na Regra de Bayes é contornada. No entanto, nem todo modelo aleatório possui uma *priori* conjugada associada e, algumas vezes, se utiliza uma *priori* não conjugada mesmo que a *priori* conjugada exista para o modelo, uma vez que a informação prévia pode não ser bem representada pela *priori* conjugada (KELLY; SMITH, 2011).

Denomina-se de *priori informativa*, a distribuição *a priori* construída baseada na informação prévia do pesquisador sobre o parâmetro  $\theta$ . Para a ocasião em que o pesquisador desconhece tal informação prévia sobre  $\theta$ , ou os dados possuem uma grande relevância, é racional que a escolha seja de uma *priori não informativa* (GELMAN et al., 2013). A maneira mais conhecida para a definição de uma *priori* não informativa foi proposta por Jeffreys (1961).

A distribuição *a posteriori* é usada para fazer afirmações sobre  $\theta$ , pois nela contém toda sua informação probabilística. Sendo assim, podem ser utilizados, por exemplo, como estimadores pontuais a média, moda e mediana da *posteriori*, bem como estimadores intervalares, conhecidos como *intervalos de credibilidade*. Define-se  $C$  um intervalo de credibilidade  $100(1 - \alpha)\%$ ,  $0 < \alpha < 1$  para  $\theta$  se  $P[\theta \in C] = 1 - \alpha$ .

Apesar da aparente semelhança com os intervalos de confiança, os intervalos de credibilidade possuem interpretação completamente diferente da abordagem frequentista. Um intervalo de confiança com  $100(1 - \alpha)\%$  sugere que se fosse re-



petido o mesmo experimento, nas mesmas condições, para um grande número  $n$  de vezes, então o verdadeiro valor de  $\theta$  estaria fora do intervalo em apenas  $100\alpha\%$  das vezes. Desta forma, no intervalo de confiança, o intervalo é aleatório, a estimativa intervalar é um dos possíveis valores realizados desse intervalo aleatório e, uma vez que  $\theta$  é tratado como fixo, a estimativa intervalar cobre o verdadeiro valor  $\theta$  com probabilidade 0 ou 1. Quando se declara um intervalo com confiança  $100(1 - \alpha)\%$ , significa que  $100(1 - \alpha)\%$  dos valores amostrados do intervalo aleatório cobrem o verdadeiro valor do parâmetro  $\theta$  (CASELLA; BERGER, 2002).

Diferentemente, um intervalo de credibilidade indica explicitamente a probabilidade *a posteriori* de  $\theta$  encontrar-se dentro de seus limites,  $P(\theta \in C)$ . Isso é possível devido ao parâmetro de interesse  $\theta$  ser tratado como uma variável aleatória, que por sua vez é associada a uma distribuição de probabilidade (BLANGIARDO; CAMELETTI, 2015).

Com o mesmo nível de credibilidade, existe uma infinidade de intervalos. A região com menor amplitude para um mesmo nível de credibilidade é chamada de *intervalo de máxima densidade a posteriori* (HPD), do inglês *highest posterior density*), e consiste em tomar os valores de  $\theta$  que possuem maior densidade *a posteriori*.

Uma outra distribuição importante para a inferência bayesiana é a chamada *distribuição preditiva a posteriori*, pois uma vez que os dados  $\mathbf{y}$  são observados, pode-se ter o interesse em prever  $y_i^*$  desconhecidos, mas observáveis do mesmo processo. A distribuição *a posteriori* preditiva, que será denotada por  $p(\mathbf{y}^*|\mathbf{y})$ , é *a posteriori*, pois é condicional as observações  $\mathbf{y}$  e é preditiva porque é uma predição para um  $\mathbf{y}^*$  observável (GELMAN et al., 2013). A distribuição  $p(\mathbf{y}^*|\mathbf{y})$  pode ser obtida do seguinte modo:

$$\begin{aligned}
p(\mathbf{y}^*|\mathbf{y}) &= \int p(\mathbf{y}^*,\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\
&= \int p(\mathbf{y}^*|\boldsymbol{\theta},\mathbf{y})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\
&= \int p(\mathbf{y}^*|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}.
\end{aligned} \tag{34}$$

Em que na última linha da equação (34) se considera que  $\mathbf{y}^*$  e  $\mathbf{y}$  são condicionalmente independentes dado  $\boldsymbol{\theta}$ .

## 6.1 Inferência Bayesiana Aproximada

Quando se utiliza *priori* conjugada, a forma funcional da distribuição a posteriori é conhecida. Ainda que a distribuição *a posteriori* seja analiticamente tratável, o uso de *priori* conjugada é muito rigoroso e não representa a maioria dos casos na prática (GELMAN et al., 2013). Quando o uso da conjugada não é adequada, na maioria das situações a distribuição *a posteriori* não possui forma fechada e as integrais necessárias para a obtenção das marginais e avaliação dos estimadores bayesianos não possuem solução analítica (ROBERT; CASELLA, 2004). Desse modo, métodos numéricos são requisitados para que esse impasse seja contornado. Na sequência apresentam-se, de maneira sucinta, duas metodologias consagradas da computação bayesiana: os métodos de Monte Carlo via Cadeias de Markov (MCMC) mais precisamente, os algoritmos Amostrador de Gibbs (*Gibbs Sampling*) e o de Metropolis-Hastings. Resumidamente, os métodos MCMC geram valores através de uma distribuição condicional *a posteriori* para cada parâmetro, de modo que se tenham uma sequência de valores que representam uma aproximação da distribuição *a posteriori* e que podem ser utilizados para obter qualquer medida de resumo *a posteriori* de interesse. Para um estudo detalhado, recomenda-se as obras de Press (2002) e Robert e Casella (2004).

### 6.1.1 Amostrador de Gibbs

O Amostrador de Gibbs foi, primeiramente, proposto por Geman e Geman (1984) dentro do contexto de modelos de processamento de imagens e proposto por Gelfand e Smith (1990) como uma abordagem baseada em amostragem para calcular densidades marginais dentro do paradigma da inferência bayesiana.

O interesse maior está na distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$  de um conjunto de parâmetros  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_p\}$ . Para simular valores de  $p(\boldsymbol{\theta}|\mathbf{y})$ , é necessário uma *distribuição condicional completa* para cada parâmetro  $\theta_i$ , ou seja, uma distribuição condicional *a posteriori* de  $\theta_i$  dado todos os outros parâmetros. Para simular um valor de  $p(\boldsymbol{\theta}|\mathbf{y})$ , o algoritmo gera, iterativamente, valores de todas as condicionais completas. Um algoritmo descrito em Press (2002) é apresentado a seguir:

1. escolha valores iniciais para  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$  e inicie o contador da cadeia  $t = 1$ ;
2. obtenha um novo vetor  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$ , a partir de  $\boldsymbol{\theta}^{(t-1)}$ , através de sucessivas geração de valores:
  - $\theta_1^{(t)}$  da distribuição condicional completa  $p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}, \mathbf{y})$
  - $\theta_2^{(t)}$  da distribuição condicional completa  $p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}, \mathbf{y})$  com  $\theta_1^{(t-1)}$  substituído por  $\theta_1^{(t)}$  ;
  - ⋮
  - $\theta_p^{(t)}$  da distribuição condicional completa  $p(\theta_p|\theta_1^{(t)}, \theta_2^{(t-1)}, \dots, \theta_{p-1}^{(t-1)}, \mathbf{y})$  com  $\theta_1^{(t-1)}, \dots, \theta_{p-1}^{(t-1)}$  substituído por  $\theta_1^{(t)}, \dots, \theta_{p-1}^{(t)}$  ;
3. altere o contador  $t$  para  $t + 1$ ;
4. retorne ao passo 2 até atingir convergência.

### 6.1.2 Metropolis-Hastings

O algoritmo de Metropolis-Hastings foi inicialmente proposto por Metropolis et al. (1953) e, então, generalizado por Hastings (1970). Para fins didáticos, é considerado o caso em que  $\theta$  é univariado.

1. escolha um valor inicial para  $\theta^0$  e inicie o contador da cadeia  $t = 1$ ;
2. amostra  $\theta^*$  de uma distribuição geradora de candidatos proposta  $q(\theta^*|\theta^{(t-1)})$ ;
3. calcule a probabilidade de aceitação  $\alpha = \min\left(1, \frac{p(\mathbf{y}|\theta^*)p(\theta^*)}{p(\mathbf{y}|\theta^{(t-1)})p(\theta^{(t-1)})}\right)$  e gere um número aleatório  $u$  de uma distribuição uniforme 0 e 1;
4. se  $\alpha > u$ , então, aceite  $\theta^*$  e faça  $\theta^{(t+1)} = \theta^*$ , caso contrário, rejeite  $\theta^*$  e faça  $\theta^{(t+1)} = \theta^{(t)}$ ;
5. incremente o contador da cadeia de  $t$  para  $t + 1$  e retorne ao passo 2.

Para implementar o algoritmo de Metropolis-Hastings se faz necessário a escolha de uma distribuição de probabilidade auxiliar que gera os valores candidatos a comporem a amostra de  $\theta$ . Frequentemente, uma função de densidade simétrica é preferida, tal que  $q(\theta^*|\theta^{(t-1)}) = q(\theta^{(t-1)}|\theta^*)$ . Para esse caso, Blangiardo e Cameletti (2015) sugerem como possíveis escolhas as distribuições Uniforme e Normal centrada em torno do último valor, como segue:

$$q(\theta^*|\theta^{(t-1)}) \sim \text{Uniforme}(\theta^{(t-1)} - \gamma, \theta^{(t-1)} + \gamma) \quad (35)$$

ou ainda,

$$q(\theta^*|\theta^{(t-1)}) \sim \text{Normal}(\theta^{(t-1)}, \gamma^2), \quad (36)$$

em que  $\gamma$  é um parâmetro de calibração escolhido, apropriadamente, com respeito à eficiência do algoritmo (controle da taxa de aceitação do algoritmo).

### 6.1.3 INLA: Aproximação de Laplace Aninhada Integrada

A *Aproximação de Laplace Aninhada Integrada* (INLA), do inglês *Integrated Nested Laplace Approximation*, é um algoritmo determinístico para inferência bayesiana, proposto por Rue, Martino e Chopin (2009), que foi desenvolvido especialmente para ser aplicado a uma classe de modelos estatísticos, conhecidos como modelos gaussianos latentes. Tais modelos, englobam uma gama de modelos de regressão com estrutura aditiva.

Em palavras, os modelos gaussianos latentes são tais que, o seu campo latente é gaussiano, controlado por alguns hiperparâmetros e também podem assumir não normalidade na variável resposta. Na maioria das situações as distribuições marginais *a posteriori* não possuem forma fechada e métodos computacionais para obtenção dessas distribuições são evocados. Comumente, os métodos de Monte Carlo via Cadeia de Markov são os métodos implementados para a solução desses problemas. No entanto, Rue, Martino e Chopin (2009) salientam que tais métodos não são uma ferramenta apropriada para uma análise rotineira de modelos gaussianos latentes, pois podem apresentar um tempo computacional muito elevado ou até mesmo não atingir convergência. Nesse ponto, o INLA é uma alternativa para contornar essas limitações. Sua eficiência computacional é a sua principal vantagem frente aos métodos MCMC dentro do contexto de modelos gaussianos latentes.

Rue, Martino e Chopin (2009) ainda destacam que o INLA ainda possui virtudes como a de não necessitar a verificação da convergência, possuir simples critérios de comparação de modelos e diversas medidas preditivas de modo que os modelos possam ser comparados e o modelo em estudo ser contestado.

O núcleo da abordagem INLA ocorre via aproximação de Laplace, apresentada a seguir.

### 6.1.3.1 Aproximação de Laplace

Suponha que o interesse é calcular a seguinte integral

$$\int f(y)dy = \int \exp(\log f(y))dy, \quad (37)$$

em que  $f(y)$  é uma função de densidade de uma variável aleatória  $Y$ . Aplicando a aproximação de segunda ordem via Série de Taylor em  $\log f(y)$  avaliado em  $y = y_0$ , tem-se

$$\log f(y) \approx \log f(y_0) + (y - y_0) \left. \frac{\partial \log f(y)}{\partial y} \right|_{y=y_0} + \frac{(y - y_0)^2}{2} \left. \frac{\partial^2 \log f(y)}{\partial y^2} \right|_{y=y_0}.$$

Definindo  $y_0$  como a moda  $y^* = \arg \max_y \log f(y)$ , então  $\left. \frac{\partial \log f(y)}{\partial y} \right|_{y=y^*} = 0$ .

A integral de interesse pode ser aproximada como:

$$\begin{aligned} \int f(y)dy &\approx \int \exp \left( \log f(y^*) + \frac{(y - y^*)^2}{2} \left. \frac{\partial^2 \log f(y)}{\partial y^2} \right|_{y=y^*} \right) dy \\ &= f(y^*) \int \exp \left( \frac{(y - y^*)^2}{2} \left. \frac{\partial^2 \log f(y)}{\partial y^2} \right|_{y=y^*} \right) dy. \end{aligned}$$

Veja que, se for denotado  $\sigma^{2*} = -1 / \left. \frac{\partial^2 \log f(y)}{\partial y^2} \right|_{y=y^*}$ , obtém-se o núcleo de uma distribuição normal

$$\begin{aligned} \int f(y)dy &\approx f(y^*) \int \exp \left( -\frac{(y - y^*)^2}{2\sigma^{2*}} \right) dy \\ &= f(y^*) \sqrt{2\pi\sigma^{2*}} \int \frac{1}{\sqrt{2\pi\sigma^{2*}}} \exp \left( -\frac{(y - y^*)^2}{2\sigma^{2*}} \right) dy. \quad (38) \end{aligned}$$

Mais precisamente, a integral avaliada no intervalo  $(\alpha, \beta)$  pode ser aproximada por

$$\int_{\alpha}^{\beta} f(y) dy \approx f(y^*) \sqrt{2\pi\sigma^{2*}} (\Phi(\beta) - \Phi(\alpha)), \quad (39)$$

em que  $\Phi(\cdot)$  denota a função de distribuição da  $Normal(y^*, \sigma^{2*})$ .

### 6.1.3.2 Modelos Gaussianos Latentes

Como dito anteriormente, o método INLA foi desenvolvido para fazer inferência, em especial, na classe dos modelos gaussianos latentes. Para definir um modelo gaussiano latente dentro do paradigma bayesiano, deve-se primeiro considerar uma distribuição de probabilidade para a variável resposta  $y_i$ , caracterizada por um parâmetro  $E[Y_i] = \mu_i$ , definido como uma função do preditor de estrutura aditiva  $\eta_i$  através de uma função de ligação  $g(\cdot)$ , tal que  $g(\mu_i) = \eta_i$ . De forma que o preditor  $\eta_i$  é dado por:

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m x_{mi} + \sum_{l=1}^L f_l(z_{li}),$$

- $\beta_0$  é um escalar que representa o intercepto;
- $\boldsymbol{\beta} = \{\beta_1, \beta_2, \dots, \beta_M\}$  quantificam o efeito linear das variáveis explicativas  $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_M\}$  na variável resposta;
- $\boldsymbol{f} = \{f_1(\cdot), f_2(\cdot), \dots, f_L(\cdot)\}$  é uma coleção de funções definidas em termos de um conjunto de covariáveis  $\boldsymbol{z} = \{z_1, z_2, \dots, z_L\}$ .

Rue, Martino e Chopin (2009) ressaltam que essa classe de modelos possuem uma riqueza de aplicações, devido a enorme quantidade de formas que as funções  $f_l(\cdot)$  podem tomar. Natario (2013) exemplifica algumas dessas formas:

- Modelos Lineares Generalizados:  $f_l(\cdot) = 0$ .

- Modelos Lineares Generalizados Mistos:  $\mathbf{f} \sim Normal(\mathbf{0}, \mathbf{Q}^{-1})$ .
- Modelos Espaciais e Espaço-Temporais:

$f_i(\cdot) \sim CAR$  (efeitos espaciais estruturados);

$f_i(\cdot) \sim Normal(0, \sigma_{f_i}^2)$  (efeitos não espacialmente estruturados);

$f_i(\cdot) \sim AR(\phi, \sigma_{f_i}^2)$  (efeitos temporais);

$f_i(\cdot) \sim Normal(0, \sigma_{f_i}^2)$  (efeitos não temporalmente estruturados).

Defina-se  $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\}$  como o conjunto de parâmetros das variáveis latentes de interesse para se fazer inferência. Ademais, seja  $\boldsymbol{\psi} = [\psi_1, \psi_2, \dots, \psi_K]^\top$  um vetor de  $K$  hiperparâmetros. Assumindo independência condicional, a distribuição das  $n$  observações é dada pela seguinte verossimilhança:

$$p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^n p(y_i|\theta_i, \boldsymbol{\psi}),$$

em que cada  $y_i$  se relaciona a somente um elemento  $\theta_i$  no campo latente  $\boldsymbol{\theta}$ . Para o campo latente  $\boldsymbol{\theta}$ , assume-se uma *priori* normal multivariada com vetor de médias  $\mathbf{0}$  e matriz de precisão  $\mathbf{Q}(\boldsymbol{\psi})$ , isto é,  $\boldsymbol{\theta}|\boldsymbol{\psi} \sim Normal(\mathbf{0}, \mathbf{Q}(\boldsymbol{\psi})^{-1})$  com função de densidade dada por

$$p(\boldsymbol{\theta}|\boldsymbol{\psi}) = (2\pi)^{-n/2} |\mathbf{Q}(\boldsymbol{\psi})|^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\theta}^\top \mathbf{Q}(\boldsymbol{\psi}) \boldsymbol{\theta}\right).$$

Os componentes do campo latente  $\boldsymbol{\theta}$  são assumidos condicionalmente independentes, o que implica em uma matriz de precisão  $\mathbf{Q}(\boldsymbol{\psi})$  esparsa. Tal condição permite um ganho computacional no processo de inferência. Dessa forma, a distribuição *a posteriori* de  $\boldsymbol{\theta}$  e  $\boldsymbol{\psi}$  pode ser escrita como:



$$\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) &= \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})}{p(\mathbf{y})} \\
&\propto p(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}) \\
&\propto p(\boldsymbol{\theta}, \boldsymbol{\psi}) \times p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi}) \\
&\propto p(\boldsymbol{\psi}) \times p(\boldsymbol{\theta} | \boldsymbol{\psi}) \times p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi}) \\
&\propto p(\boldsymbol{\psi}) \times p(\boldsymbol{\theta} | \boldsymbol{\psi}) \times \prod_{i=1}^n p(y_i | \theta_i, \boldsymbol{\psi}) \\
&\propto p(\boldsymbol{\psi}) \times |Q(\boldsymbol{\psi})|^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\theta}^\top Q(\boldsymbol{\psi}) \boldsymbol{\theta}\right) \times \\
&\quad \prod_{i=1}^n \exp(\log(p(y_i | \theta_i, \boldsymbol{\psi}))) \\
&\propto p(\boldsymbol{\psi}) \times |Q(\boldsymbol{\psi})|^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\theta}^\top Q(\boldsymbol{\psi}) \boldsymbol{\theta} + \sum_{i=1}^n \log(p(y_i | \theta_i, \boldsymbol{\psi}))\right).
\end{aligned}$$

### 6.1.3.3 Metodologia INLA

Rue, Martino e Chopin (2009) afirmam que para fazer inferência bayesiana tem-se o interesse em obter as distribuições marginais *a posteriori* para cada elemento do vetor de parâmetros e do vetor de hiperparâmetros, pois toda a informação sobre os parâmetros se encontram nessas distribuições. Na metodologia INLA, o interesse é obter aproximações dessas marginais *a posteriori*, tanto das variáveis latentes

$$p(\theta_i | \mathbf{y}) = \int p(\theta_i, \boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi} = \int p(\theta_i | \boldsymbol{\psi}, \mathbf{y}) p(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi},$$

quanto dos hiperparâmetros do modelo gaussiano latente

$$p(\psi_k | \mathbf{y}) = \int p(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}_{-k}.$$

em que  $\boldsymbol{\psi}_{-k}$  é o vetor  $\boldsymbol{\psi}$  sem a  $k$ -ésima componente.

Denote por  $\tilde{p}(\cdot|\cdot)$  uma distribuição condicional aproximada em seus argumentos, a metodologia INLA tem por objetivo construir aproximações aninhadas:

$$\begin{aligned}\tilde{p}(\boldsymbol{\psi}_k|\mathbf{y}) &= \int \tilde{p}(\boldsymbol{\psi}|\mathbf{y})d\boldsymbol{\psi}_{-k}, \\ \tilde{p}(\theta_i|\mathbf{y}) &= \int \tilde{p}(\theta_i|\boldsymbol{\psi},\mathbf{y})\tilde{p}(\boldsymbol{\psi}|\mathbf{y})d\boldsymbol{\psi}.\end{aligned}$$

Para o cálculo da distribuição *a posteriori* conjunta dos hiperparâmetros, utiliza-se o seguinte resultado.

$$\begin{aligned}p(\boldsymbol{\psi}|\mathbf{y}) &= \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})}{p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y})} \frac{1}{p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{p(\mathbf{y})} \frac{1}{p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &\propto \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})}.\end{aligned}$$

A aproximação da distribuição *a posteriori* conjunta dos hiperparâmetros é dada por

$$\tilde{p}(\boldsymbol{\psi}|\mathbf{y}) \propto \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})}, \quad (40)$$

em que  $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})$  é a aproximação gaussiana de  $p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})$  e  $\boldsymbol{\theta}^*(\boldsymbol{\psi})$  é a moda da condicional completa de  $\boldsymbol{\theta}$  para um dado  $\boldsymbol{\psi}$  que é obtida através de algum método de otimização, como por exemplo, o de Newton-Raphson.

Blangiardo e Cameletti (2015) afirmam que a aproximação gaussiana acaba sendo precisa, uma vez que  $p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})$  é quase Normal, pois a priori do campo

latente necessita de normalidade e  $\mathbf{y}$  geralmente é não informativo com distribuição bem comportada.

O próximo passo é calcular  $\tilde{p}(\theta_i|\boldsymbol{\psi},\mathbf{y})$  que é necessário para o cálculo da marginal *a posteriori*  $p(\theta_i|\mathbf{y})$ . Para essa aproximação Rue, Martino e Chopin (2009) sugerem três estratégias distintas: aproximação gaussiana, aproximação de Laplace completa e aproximação de Laplace simplificada. Tais estratégias variam entre eficiência computacional e precisão da aproximação.

A aproximação gaussiana é a mais simples das estratégias, possui maior eficiência computacional e menor precisão. Consiste em aproximar  $p(\theta_i|\boldsymbol{\psi},\mathbf{y})$  diretamente das marginais de  $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\mathbf{y})$ , obtida anteriormente no denominador de (40), para encontrar  $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$ . Isto é,  $\tilde{p}(\theta_i|\boldsymbol{\psi},\mathbf{y}) = N(\mu_i(\boldsymbol{\psi}), \sigma_i^2(\boldsymbol{\psi}))$ , em que  $\boldsymbol{\mu}(\boldsymbol{\psi})$  é o vetor de médias e  $\boldsymbol{\sigma}^2(\boldsymbol{\psi})$  o vetor de variâncias marginais da aproximação gaussiana.

A aproximação de Laplace completa é a mais precisa e a computacionalmente mais custosa dentre as estratégias e consiste em reescrever  $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$  e usar agora a aproximação de Laplace para obter  $\tilde{p}(\theta_i|\boldsymbol{\psi},\mathbf{y})$ . Note primeiro que

$$\begin{aligned} p(\theta_i|\boldsymbol{\psi},\mathbf{y}) &= \frac{p((\theta_i, \boldsymbol{\theta})_{-i}|\boldsymbol{\psi},\mathbf{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y})} \\ &= \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}{p(\boldsymbol{\psi}|\mathbf{y})} \frac{1}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y})} \\ &\propto \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y})}. \end{aligned}$$

Assim a aproximação de Laplace completa de  $p(\theta_i|\boldsymbol{\psi},\mathbf{y})$  é

$$\tilde{p}(\theta_i|\boldsymbol{\psi},\mathbf{y}) \propto \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}{\tilde{p}(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y})} \Bigg|_{\boldsymbol{\theta}_{-i}=\boldsymbol{\theta}_{-i}^*(\theta_i,\boldsymbol{\psi})}, \quad (41)$$

em que  $\tilde{p}(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y})$  é a aproximação gaussiana para  $\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\mathbf{y}$  e  $\boldsymbol{\theta}_{-i}^*(\theta_i,\boldsymbol{\psi})$  é a moda.

A aproximação de Laplace simplificada é baseada na expansão em Série

de Taylor da aproximação de Laplace dada em (41) e para mais detalhes ver (RUE; MARTINO; CHOPIN, 2009).

Uma vez obtidas  $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$  e  $\tilde{p}(\theta_i|\boldsymbol{\psi}, \mathbf{y})$ , a marginal *a posteriori*  $p(\theta_i|\mathbf{y})$  é aproximada por

$$\tilde{p}(\theta_i|\mathbf{y}) = \int \tilde{p}(\theta_i|\boldsymbol{\psi}, \mathbf{y})\tilde{p}(\boldsymbol{\psi}|\mathbf{y})d\boldsymbol{\psi},$$

e a integral pode ser resolvida numericamente através de uma soma finita ponderada, dada por

$$\tilde{p}(\theta_i|\mathbf{y}) = \sum_j \tilde{p}(\theta_i|\boldsymbol{\psi}^{(j)}, \mathbf{y})\tilde{p}(\boldsymbol{\psi}^{(j)}|\mathbf{y})\Delta_j, \quad (42)$$

para alguns relevantes pontos de integração  $\{\boldsymbol{\psi}^{(j)}\}$  com um correspondente conjunto de pesos  $\{\Delta_j\}$ .

Blangiardo e Cameletti (2015) fornecem os passos da metodologia INLA:

1. Explorar a distribuição *a posteriori* conjunta  $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$  a fim de obter pontos  $\{\boldsymbol{\psi}^{(j)}\}$  com alta densidade, necessários para a integração numérica em (42). Para esta etapa, Rue, Martino e Chopin (2009) propõem dois esquemas distintos de exploração que requerem uma reparametrização do espaço de  $\boldsymbol{\psi}$ , a fim de se obter densidades mais regulares. Os passos são descritos a seguir:
  - (a) Localize a moda de  $\boldsymbol{\psi}^*$  de  $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$  através da otimização de  $\log \tilde{p}(\boldsymbol{\psi}|\mathbf{y})$  com respeito a  $\boldsymbol{\psi}$ , por exemplo, utilizando o Método de Newton-Raphson.
  - (b) Calcule a matriz Hessiana negativa  $\mathbf{H}$  da configuração modal.
  - (c) Calcule a autodecomposição  $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ , em que  $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$ .
  - (d) Defina a nova variável  $\mathbf{z}$ , com componentes padronizados e mutuamente ortogonais, tais que

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

O primeiro esquema de exploração constrói uma grade de pontos usando a parametrização  $\mathbf{z}$ , que utiliza a maior parte da densidade de

$\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$ .

O segundo esquema, chamado *planejamento composto central* (do inglês *central composite design*) (CCD), deve ser usado para reduzir custo computacional; nesse esquema, o problema da integração é visto como um problema de planejamento. Utilizando a moda  $\boldsymbol{\psi}^*$  e a matriz hessiana  $\mathbf{H}$  alguns pontos relevantes no espaço de  $\boldsymbol{\psi}$  são selecionados para fazer uma aproximação de segunda ordem para uma variável resposta.

Após realizar a exploração por grade, cada marginal *a posteriori*  $p(\psi_k|\mathbf{y})$  pode ser obtida usando um algoritmo de interpolação, baseado nos valores da densidade de  $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$  avaliada no conjunto de pontos de integração  $\{\psi_k^{(j)}\}$ . O algoritmo da interpolação pode ser encontrado em Martins et al. (2013).

2. Para cada valor  $\{\psi_k^{(j)}\}$ , a condicional *a posteriori*  $\tilde{p}(\theta_i|\psi^{(j)},\mathbf{y})$  são avaliadas em uma grade de valores selecionados para  $\theta_i$  e a marginal *a posteriori*  $\tilde{p}(\theta_i|\mathbf{y})$  é obtida pela integração numérica em (42).

## 6.2 Verificação e Critérios de Seleção do Modelo

Um importante aspecto da modelagem bayesiana se refere à avaliação da plausibilidade e do ajuste do modelo. Por plausibilidade entende-se pela a avaliação de quais variáveis devem ser incluídas no modelo, os tipos de suposições que devem ser assumidas nos parâmetros, as *prioris* que devem ser assumidas para os parâmetros e hiperparâmetros. Para o ajuste, o intuito é encontrar o modelo que melhor se ajusta aos dados disponíveis e comparar distintos modelos para as variáveis incluídas (BLANGIARDO; CAMELETTI, 2015). Uma forma de combinar os conceitos de plausibilidade e ajuste é dada pela análise de sensibilidade (GELMAN et al., 2013). Normalmente são utilizadas metodologias baseadas na distribuição preditiva e na função *deviance*, que são sucintamente apresentadas a seguir.

### 6.2.1 Métodos Baseados na Distribuição Preditiva

A ideia por trás desses métodos é segregar a amostra de observações  $\mathbf{y}$  em dois grupos, tal que  $\mathbf{y} = (\mathbf{y}_f, \mathbf{y}_c)$ , em que  $\mathbf{y}_f$  é utilizado para o ajuste do modelo e para estimar a distribuição *a posteriori* dos parâmetros, e  $\mathbf{y}_c$  é usado para verificar a plausibilidade do modelo (*model criticism*).

A maneira mais comum para segregar  $\mathbf{y}$  é através da validação cruzada (*cross-validation*), que segrega  $\mathbf{y}$  de modo que cada observação pertença somente a um dos dois grupos  $\mathbf{y}_f$  e  $\mathbf{y}_c$ . Para verificar a adequação do modelo, dois índices são usados. São eles:

- ordenada preditiva condicional (*conditional predictive ordinate*):  $CPO_i = p(y_i^* | \mathbf{y}_f)$ ;
- transformação integral de probabilidade (*probability integral transform*)  $PIT_i = p(y_i^* \leq y_i | \mathbf{y}_f)$ .

Gneiting, Balabdaoui e Raftery (2007) sugerem que a distribuição empírica da PIT pode ser usada para a avaliação do desempenho preditivo do modelo; uma distribuição preditiva é considerada coerente se a distribuição empírica da PIT é *Uniforme*(0,1). Uma outra medida de resumo do ajuste do modelo é dada pela *pseudo marginal log verossimilhança* (*log pseudo marginal likelihood*) (LPML) sugerida por, Geisser e Eddy (1979), a LPML deriva considerações preditivas e conduz a pseudo fatores de Bayes para a escolha de modelos. Segundo Carlin e Louis (2008), a LPML pode ser usada para comparar/competir modelos, em termos de predição, e quanto maior for o seu valor, melhor é o ajuste do modelo. A LPML é dada por:

$$LPML = \sum_{i=1}^n \log CPO_i, \quad (43)$$

em que  $CPO_i$  é a  $i$ -ésima ordenada preditiva condicional, isto é, é a densidade preditiva baseada em  $\mathbf{y}_{-i}$  avaliada em  $y_i$ .

### 6.2.2 Deviance Information Criterion (DIC)

Sugerido por Spiegelhalter et al. (2002) o DIC é reconhecido como uma versão bayesiana do AIC em (30) e foi desenvolvido justamente para comparar modelos bayesianos. As diferenças entre o DIC e o AIC, estão no fato de se utilizar o estimador bayesiano  $\hat{\theta}_{bayes} = E[\theta|\mathbf{y}]$  no lugar do estimador de máxima verossimilhança  $\hat{\theta}_{mle}$ ; e no de substituir o número de parâmetros  $p$  por um outro, denotado por  $p_D$ , que representa o número efetivo de parâmetros e faz uma correção de viés baseado nos dados. Analogamente ao AIC, o DIC é definido como:

$$DIC = -2 \log(p(\mathbf{y}|\hat{\theta}_{bayes})) + 2p_D, \quad (44)$$

em que,

$$p_D = E_{\theta|\mathbf{y}}[-2 \log(p(\mathbf{y}|\theta))] + 2 \log(p(\mathbf{y}|\hat{\theta}_{bayes})). \quad (45)$$

Recordando a equação (18) a *Deviance* avaliada em  $\theta$  é dada por  $D(\theta) = -2 \log(p(\mathbf{y}|\theta))$ . Consequentemente, pode-se reescrever (45), como

$$p_D = E_{\theta|\mathbf{y}}[D(\theta)] - D(E_{\theta|\mathbf{y}}[\theta]) = \bar{D} - D(\bar{\theta}). \quad (46)$$

Equivalentemente (44), temos

$$DIC = \bar{D} + p_D. \quad (47)$$

Em palavras, o DIC é a soma da média *a posteriori* da *Deviance* e o número efetivo de parâmetros que é determinado por (46). Análogo ao AIC, diz-se que modelo  $m_1$  é preferível ao modelo  $m_2$  se  $DIC_{m_1} < DIC_{m_2}$ . Cabe ressaltar que o modelo com menor DIC não necessariamente será um modelo adequado se não reproduzir os dados adequadamente, sendo, então, necessário avaliar a consis-

tência das previsões do modelo (CONGDON, 2010). Outro ponto a se destacar é que o DIC tem a tendência de sub penalizar modelos complexos com muitos efeitos aleatórios (PLUMMER, 2008).



## 7 MATERIAL E MÉTODOS

Os dados são referentes ao seguro de automóveis entre os anos de 2007 a 2014 por unidade da federação (UF) e foram obtidos junto ao AUTOSEG - Sistema de Estatísticas de Automóveis da Superintendência de Seguros Privados. O banco de dados contém ao todo 2160 observações e é constituído pelas variáveis: prêmio médio; sexo do condutor; faixa etária do condutor; importância segurada média; número de sinistros por colisão; número de sinistros por incêndio ou roubo; número de sinistro por outras causas; valor total de indenizações por colisão; valor total de indenizações por incêndio ou roubo; valor total de indenizações por outras causas. A Tabela 3 mostra como os dados estão dispostos

Tabela 3 Dados Seguros de Automóveis

Ano	UF	Sexo	Faixa Etária	Importância Segurada Média	...	Prêmio Médio
2007	AC	Feminino	18-25	30299	...	1233
2007	AC	Feminino	26-35	33696	...	1083
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2007	TO	Masculino	>55	32646	...	1617
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2008	AC	Feminino	18-25	32899	...	1268
2008	AC	Feminino	26-35	35860	...	1108
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2014	TO	Masculino	>55	54938	...	2007

Uma análise descritiva dos dados de seguros de automóveis é realizada. Posteriormente, para o ajuste dos modelos, será considerado como variável resposta  $y$  o valor médio do prêmio de seguro de automóvel, tendo como intuito verificar a influência que as variáveis citadas exercem no valor médio do prêmio de seguro de automóveis.

Para o ajuste dos modelos, utiliza-se a metodologia INLA, mais precisamente, utilizando a estratégia da Aproximação de Laplace Completa para as aproximações das distribuições marginais a posteriori e planejamento composto central para exploração da grade de pontos. O critério utilizado para a seleção do modelo será baseado nos critérios,  $DIC$ , que fornece o melhor *trade-off* entre ajuste e

complexidade e *LPML* que mensura a qualidade preditiva do modelo. Para a verificação da calibração do modelo constrói-se um histograma das transformações integral de probabilidade.

A aproximação de Laplace Completa foi escolhida como estratégia de aproximação das marginais a posteriori, pois, segundo Schrödle et al. (2011), o cálculo de medidas preditivas frequentemente requer a aproximação de Laplace Completa para alcançar resultados confiáveis e tais medidas podem ser obtidas na metodologia INLA sem a necessidade de reexecutar o ajuste do modelo.

As variáveis explicativas consideradas no estudo são:

- *Sexo* que representa o sexo do condutor, com dois níveis: masculino e feminino;
- *FaixaEtaria* representando a faixa etária do condutor, dispondo de cinco níveis: entre 18 e 25 anos, entre 26 e 35 anos, entre 36 e 45 anos, entre 46 e 55 anos e maior que 55 anos;
- *ImpSegurada* é o valor médio em reais da importância segurada;
- *SinistrosColisao* é o número de sinistros ocorridos por colisão;
- *SinistrosIncRoubo* é o número de sinistros ocorridos por incêndio ou roubo;
- *SinistrosOut* é o número de sinistros ocorridos por outras causas;
- *IndenizacaoColisao* é o valor em reais pago em indenizações devido a sinistros por colisão;
- *IndenizacaoIncRoubo* é valor em reais pago em indenizações devido a sinistros ocorridos por incêndio ou roubo;
- *IndenizacaoOut* é o valor em reais pago em indenizações devido a sinistros ocorridos por outras causas;

- $UF$  que representa a unidade da federação, com vinte e sete níveis referentes a cada estado do Brasil;
- $Ano$  que representa o ano, contendo 8 níveis de 2007 a 2014.

Para a análise, será descrito um modelo geral que servirá de base para todos os modelos ajustados neste estudo. O modelo geral é similar ao modelo descrito em Bernardinelli et al. (1995), que ainda sugere que efeitos de área e tendência temporal sejam tratados como aleatórios. O modelo geral é apresentado a seguir:

$$y_{ijkt} | \mu_{ijkt}, \tau_g \sim \text{Gama}(\tau_g, \tau_g / \mu_{ijkt}),$$

$$i = 1, 2, j = 1, 2, \dots, 5, k = 1, \dots, 27, t = 1, \dots, 8.$$

$$\begin{aligned} \eta_{ijkt} = \log(\mu_{ijkt}) = & \log(\text{ImpSegurada}_{ijkt}) + \beta_0 + \beta_1 \text{Sexo}_i + \\ & \beta_2 \text{FaixaEtaria}_j + \beta_3 \text{SinistrosColisao}_{ijkt} + \\ & \beta_4 \text{SinistrosIncRoubo}_{ijkt} + \beta_5 \text{SinistrosOut}_{ijkt} + \\ & \beta_6 \text{IndenizacaoColisao}_{ijkt} + \beta_7 \text{IndenizacaoIncRoubo}_{ijkt} + \\ & \beta_8 \text{IndenizacaoOut}_{ijkt} + \\ & u_k + v_k + \gamma_t + \phi_t + \delta_{kt}. \end{aligned} \quad (48)$$

De modo que a priori do hiperparâmetro de precisão das observações gama  $\tau_g$ , é definida para  $\log \tau_g$  de tal modo que  $\log(\tau_g) \sim \text{loggamma}(1, 0.01)$ . A variável “valor médio da importância segurada” é uma medida de exposição, portanto, é incluída no modelo como um *offset*. Para os parâmetros dos efeitos fixos (incluindo um intercepto)  $\beta$ , assume-se que há informação suficiente nos dados, de modo que uma priori normal com grande variância forneça pouca informação *a priori* (FONG; RUE; WAKEFIELD, 2009). Desse modo,  $\beta \sim \text{Normal Multivariada}_{12}(\mathbf{0}, 100^2 \mathbf{I})$ .

O efeito aleatório referente a unidade da federação possui duas componentes:  $u_k$  que é um efeito estruturado no espaço e  $v_k$  não estruturado no espaço. Para o efeito estruturado espacialmente  $u_k$ , aplica-se o Modelo de Besag, uma

estrutura espacial autorregressiva condicional intrínseca (iCAR) (BESAG; KOOPERBERG, 1995). Seja  $\mathbf{u} = [u_1, \dots, u_{27}]^\top$ , o modelo de Besag é definido como:

$$u_k | u_l, k \neq l, \tau_u \sim Normal \left( \frac{1}{n_i} \sum_{k \sim l} u_l, \frac{1}{n_i \tau_u} \right),$$

em que  $n_i$  é o número de vizinhos da área  $k$ ,  $k \sim l$  indica que as unidades da federação  $k$  e  $l$  são vizinhas. A priori do hiperparâmetro de precisão  $\tau_u$  é definida para  $\log \tau_u$  de tal modo que  $\log(\tau_u) \sim \text{loggamma}(1, 5 \times 10^{-5})$ .

Para o efeito espacial não estruturado assume-se  $v_k \sim Normal(0, 1/\tau_v)$ , e a priori do hiperparâmetro de precisão  $\tau_v$  é definida para  $\log \tau_v$ , de tal modo que  $\log \tau_v \sim \text{loggamma}(1, 5 \times 10^{-5})$ .

Da união do Modelo de Besag,  $u_k$ , e o modelo espacial não estruturado  $v_k$  origina-se o modelo de Besag–York–Mollié (BYM) (BESAG; YORK; MOLLÍÉ, 1991).

Já  $\gamma_t$  e  $\phi_t$  são as duas componentes de efeito aleatório da variável ano. Sendo que  $\gamma_t$  é temporalmente estruturada e  $\phi_t$  é não estruturada no tempo. Para o efeito temporal estruturado  $\gamma_t$ , pode ser modelado dinamicamente, por exemplo, através de um passeio aleatório de ordem 1 (RW1) ou de ordem 2 (RW2), definidos como

RW1

$$\gamma_t | \gamma_{t-1} \sim Normal(\gamma_{t-1}, 1/\tau_\gamma)$$

e

RW2

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim Normal(2\gamma_{t-1} + \gamma_{t-2}, 1/\tau_\gamma).$$

A priori do hiperparâmetro de precisão  $\tau_\gamma$  é definida para  $\log(\tau_\gamma)$ , de modo que  $\log(\tau_\gamma) \sim \text{loggamma}(1, 5 \times 10^{-5})$ .

O parâmetro não temporalmente estruturado  $\phi_t \sim Normal(0, 1/\tau_\phi)$  e  $\log(\tau_\phi) \sim loggama(1, 5 \times 10^{-5})$ .

Há ainda o parâmetro  $\delta_{kt}$  que é não estruturalmente modelado e representa a interação entre o espaço e tempo. Dessa forma,  $\delta_{kt} \sim Normal(0, 1/\tau_\delta)$  e  $\log(\tau_\delta) \sim loggama(1, 5 \times 10^{-5})$ .

O modelo que contém interação espaço-tempo se atenta para o fato de que as unidades espaciais podem se comportar diferentemente da tendência temporal principal quando observada no decorrer de um longo período de tempo (SCHRÖDLE et al., 2011). A inclusão da interação espaço-tempo no modelo deve explicar diferenças da tendência temporal para diferentes unidades da federação. Se  $\delta_{kt} < 0$ , então a tendência média do unidade da federação específica é menor que a tendência média, em contrapartida,  $\delta_{kt} > 0$  implica que a tendência da unidade da federação específica é mais acentuada do que a tendência média. Outras especificações para interação de espaço-tempo, que não são tratadas neste trabalho, podem ser encontradas em Knorr-Held (2000).

Knorr-Held (2000) complementa que no modelo de Bernardinelli et al. (1995) várias modificações e extensões são possíveis, no entanto, recomenda omitir apenas efeitos principais em que sua interação não é assumida. Dessa forma, a Tabela 4 elenca algumas variações do modelo geral (48) a serem ajustados.

Tabela 4 Componentes aleatórios presentes no modelo

Modelo	$u_k$	$v_k$	$\gamma_t$	$\phi_t$	$\delta_{kt}$
I	–	–	–	–	–
II	–	sim	–	sim	sim
III	sim	–	–	sim	–
IV	sim	–	sim(rw1)	–	–
V	sim	–	sim (rw1)	sim	–
VI	sim	sim	–	sim	sim
VII	sim	sim	sim (rw1)	sim	sim
VIII	sim	sim	sim (rw2)	sim	sim

O software R (R Core Team, 2014), bem como os pacotes *stats* (R Core Team, 2014), *lattice* (SARKAR, 2008), *maptools* (BIVAND; LEWIN-KOH, 2016), *spdep* (BIVAND; PIRAS, 2015) e *INLA* (RUE; MARTINO; CHOPIN, 2009) serão utilizados para a descrição dos dados e para o ajuste dos modelos apresentados. Enquanto que o mapa do Brasil, bem como, sua estrutura de vizinhança foram obtidos junto ao Ministério Meio Ambiente.

## 8 RESULTADOS

Antes de ser apresentado o ajuste dos modelos descritos na Seção 7, realiza-se uma breve descrição dos dados. Inicia-se tentando comparar condutores do sexo masculino e feminino. Na Figura 3, apresenta-se uma comparação entre o sexo do condutor e seu respectivo valor médio do prêmio, através de um *boxplot* estratificado.

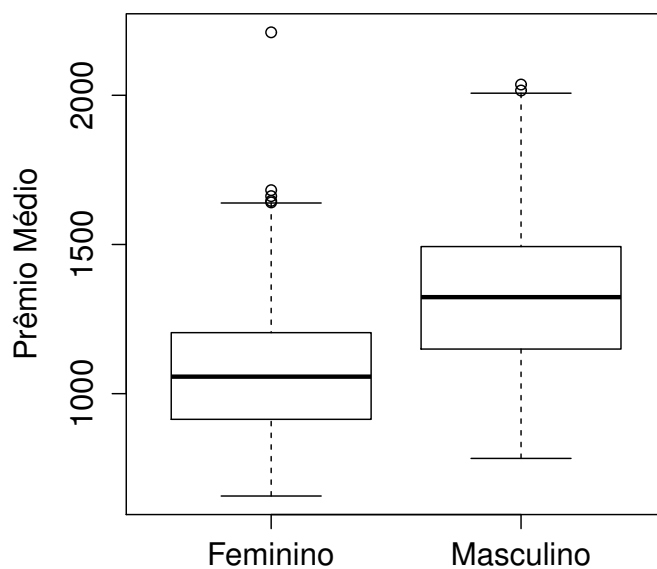


Figura 3 Box-Plot do Prêmio Médio Estratificado por Sexo do Conductor

Pode-se observar que condutores do sexo masculino pagam, em média, um maior prêmio para ter seu automóvel segurado. Esse fato pode ser explicado através da participação dos condutores do sexo masculino no número de sinistros e, sobretudo, no valor total pago em indenizações. Na Figura 4 é apresentada essa

participação.

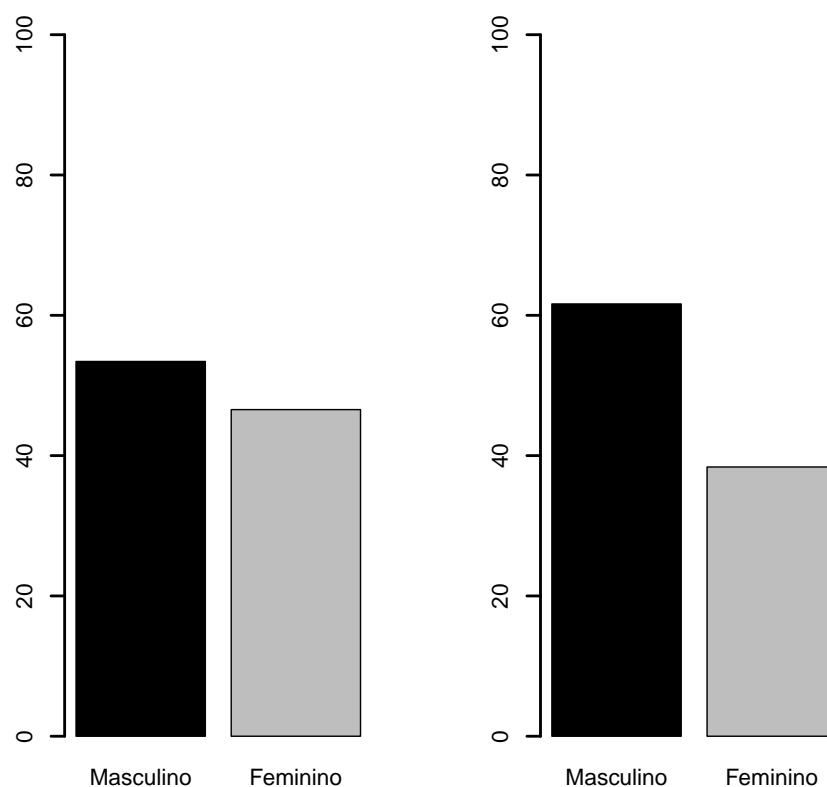


Figura 4 Proporção do número de sinistros (à esquerda) e do valor total pago em indenizações (à direita) por sexo do condutor, entre os anos de 2007 a 2014.

Observa-se que condutores do sexo masculino sofreram mais sinistros que condutores do sexo feminino (53,42% do total de sinistros), mas esse número não é tão expressivo quanto a proporção do valor gasto em indenizações dos sinistros dos condutores do sexo masculino, que foi de 61,62%. Condutores do sexo masculino pagam, em média, um prêmio de seguro de automóvel mais caro, devido à sua



predominante participação no total gasto em indenizações.

No mesmo sentido da comparação entre condutores do sexo masculino e feminino, compara-se condutores de faixas etárias diferentes. A Figura 5 apresentam *boxplots* do valor médio em prêmios estratificado por faixas etárias.

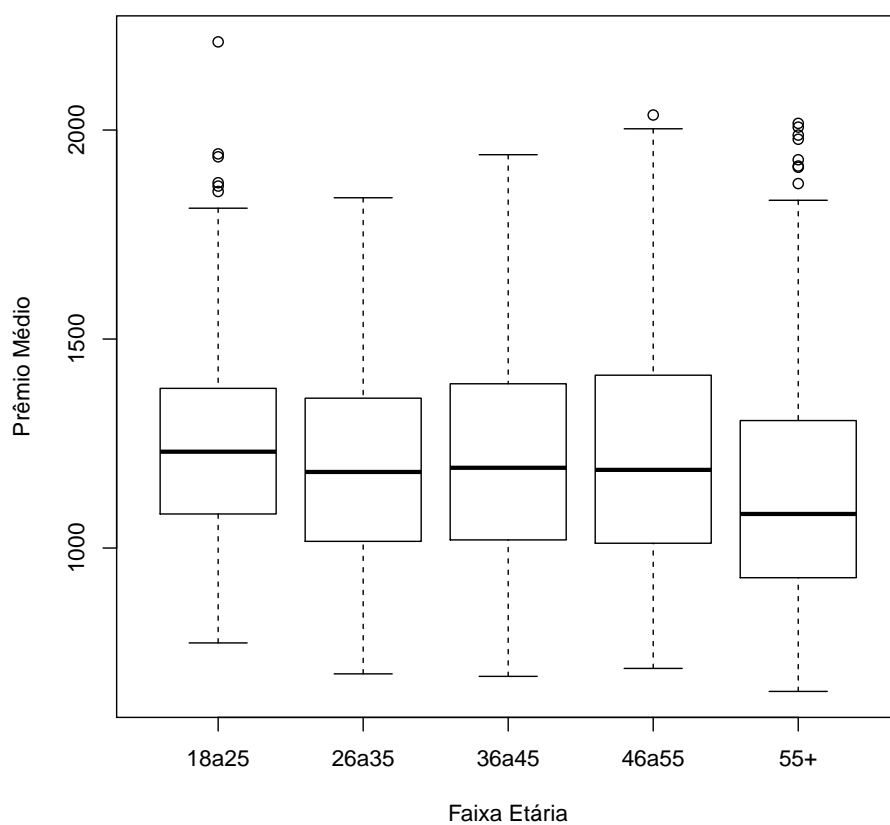


Figura 5 Boxplot do prêmio médio estratificado por faixa etária

Os *boxplots* sugerem uma pequena tendência de condutores mais jovens pagarem, em média, um prêmio um pouco maior que condutores mais experientes.

A Figura 6 apresenta a dispersão entre o valor médio do prêmio dos anos de 2007 a 2014.

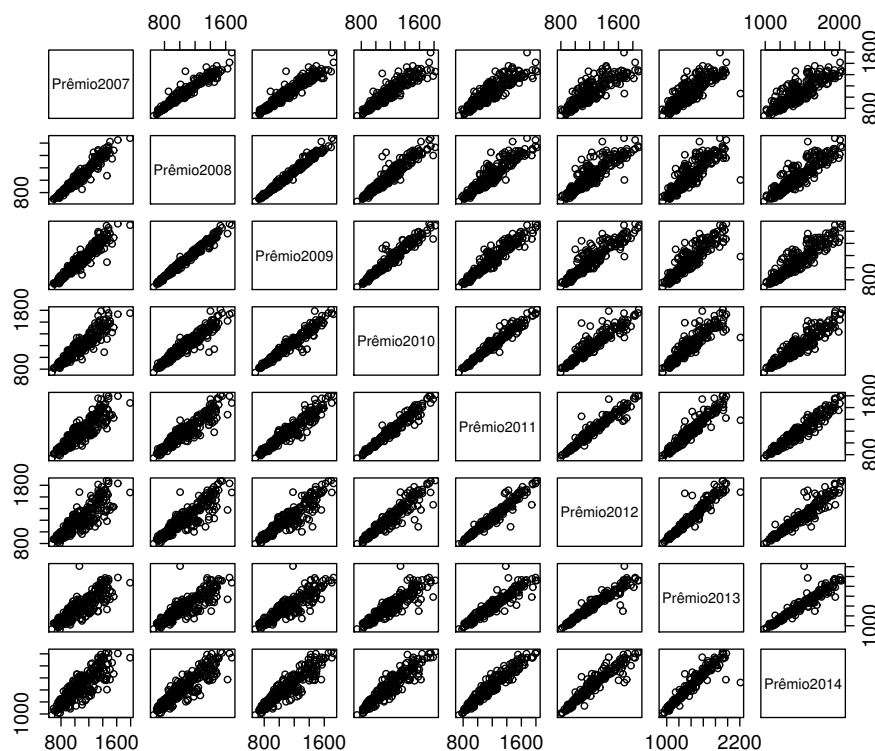


Figura 6 Dispersão entre os valores médio do Prêmio de 2007 a 2014

Pode-se observar uma forte associação linear entre os valores do prêmio médio entre todos anos em estudo. Apesar de ainda forte, essa associação linear tende a diminuir a medida que a diferença entre os anos aumentam. Uma vez que a variável resposta dos modelos a serem ajustados é o prêmio médio, a existência dessa associação entre os valores médios do prêmio e os anos é um indício de que o modelo deve considerar a correlação entre as observações da variável resposta.

A Figura 7 apresenta a evolução do valor médio do prêmio do seguro de automóveis para os anos de 2007 a 2014. Pode-se constatar que houve um cresci-

mento do valor médio do prêmio no período.

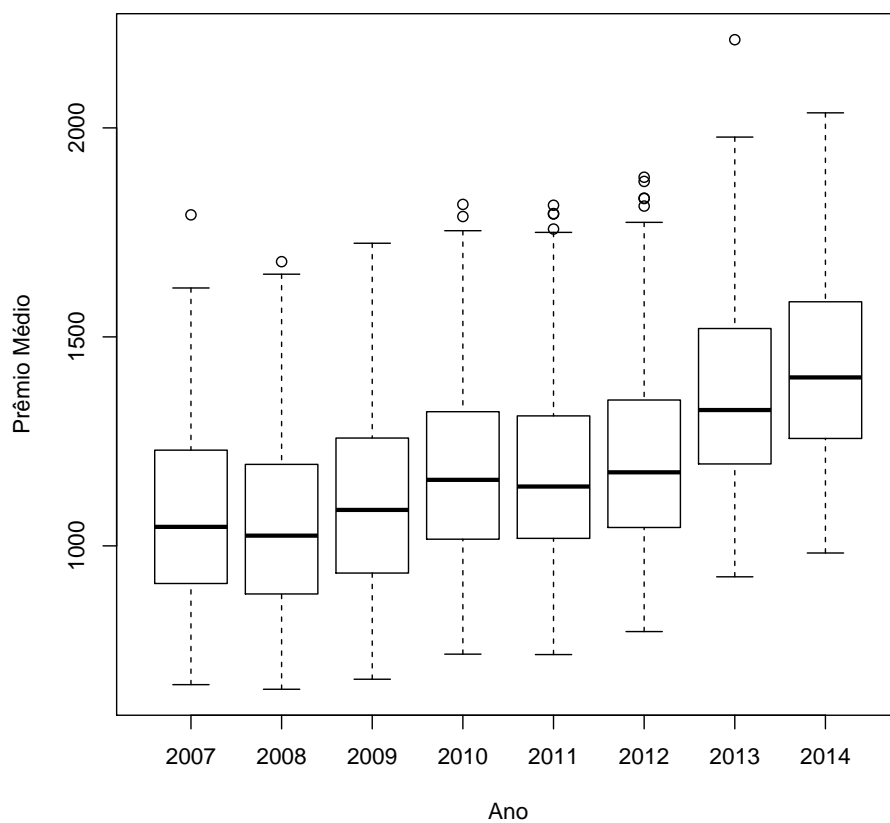


Figura 7 Evolução do prêmio médio de seguros de automóveis entre os anos de 2007 a 2014

Com a finalidade de selecionar o melhor modelo, apresenta-se na Tabela 5 os valores do Critério de Informação de *Deviance*, bem como a pseudo marginal log verossimilhança.

Tabela 5 Medidas para a comparação e seleção de modelos

Modelo	$DIC$	$p_D$	$\bar{D}$	$LPML$
I	27293,76	15,43	27278,33	-13660,11
II	22515,78	232,97	22282,81	-11266,41
III	23458,00	62,10	23295,89	-11732,53
IV	23457,99	62,09	23395,90	-11732,53
V	23457,91	61,96	23395,95	-11732,53
VI	22515,76	232,64	22283,08	-11266,37
VII	22516,86	232,72	22283,14	-11266,45
VIII	22516,03	232,75	22283,28	-11266,46

A Tabela 5 mostra que o modelo melhor ajustado foi o modelo II, uma vez que detém o menor  $\bar{D} = 22282,81$ , no entanto, sua penalização, medida pelo número efetivo de parâmetros  $p_D$  é grande o suficiente para que o modelo não forneça o menor DIC. Sendo assim, o modelo VI, que detém o menor DIC, é o modelo que oferece o melhor *trade - off* entre ajuste e complexidade e, portanto, é o modelo selecionado para o estudo. Para corroborar com a escolha, o modelo VI possui a maior LPML, o que indica que é o modelo com maior capacidade preditiva dentre os modelos ajustados.

Dessa forma, apresenta-se um resumo dos efeitos do modelo VI na Tabela 6.

Tabela 6 Resumo dos efeitos do modelo VI

Parâmetro	Média <i>a posteriori</i>	IC 95%	<i>sd</i>
Intercepto*	-3,208	(-3,263; -3,153)	0,028
Sexo Masculino	0,121	(+0,117; +0,124)	0,002
FaixaEtaria26-35	-0,217	(-0,222; -0,212)	0,003
FaixaEtaria36-45	-0,297	(-0,302; -0,291)	0,003
FaixaEtaria46-55	-0,296	(-0,301; -0,290)	0,003
FaixaEtaria>55	-0,347	(-0,351; -0,341)	0,003
Nº Colisão	-0,012	(-0,032; +0,008)	0,010
Nº Inc. ou Roubo	0,050	(+0,015; +0,085)	0,011
Nº Outras Causas	-0,004	(-0,011; +0,003)	0,004
Ind. Colisão	-0,038	(-0,061; -0,016)	0,011
Ind. Inc. ou Roubo	-0,015	(-0,052; +0,021)	0,018
Ind. Outras Causas	0,003	(-0,007; +0,013)	0,005
Hiperparâmetro	Média <i>a posteriori</i>	IC 95%	<i>sd</i>
$\tau_g$	776,378	(727,741; 827,265)	25,321
$\tau_u$	101,155	(53,025; 179,627)	32,669
$\tau_v$	1774,501	(278,940; 5566,120)	1427,7
$\tau_\phi$	204,347	(67,397; 441,230)	97,447
$\tau_\delta$	943,357	(740,072; 1184,296)	113,072

\* Intercepto confundido com Sexo Feminino e Faixa Etária de 18 a 25 anos

Observa-se que na Tabela 6, as variáveis "número de sinistros ocorridos por colisão", "número de sinistros ocorridos por outras causas", "indenização de sinistros ocorridos por incêndio ou roubo" e "indenização de sinistros ocorridos por outras causas" não possuem efeitos significativos, uma vez que o valor 0 pertence ao intervalo de 95% de credibilidade dessas variáveis. Dessa forma, tais variáveis foram retiradas do modelo. Posteriormente, um novo modelo (VI\*) foi ajustado sem essas variáveis o que ocasionou um melhor ajuste ( $DIC = 22512,92$ ) e uma melhor capacidade preditiva ( $LPML = -11263,38$ ). Um resumo do modelo é descrito na Tabela 7

Tabela 7 Resumo dos efeitos do modelo VI\*

Parâmetro	Média <i>a posteriori</i>	IC 95%	sd
Intercepto*	-3,2080	(-3,2638; -3,1522)	0,0279
Sexo Masculino	0,1216	(+0.1184; +0.1249)	0,0017
FaixaEtaria26-35	-0,2170	(-0.2221; -0.2119)	0,0026
FaixaEtaria36-45	-0,2977	(-0.3030; -0.2925)	0,0027
FaixaEtaria46-55	-0,2964	(-0.3016; -0.2912)	0,0026
FaixaEtaria>55	-0,3476	(-0.3527; -0.3424)	0,0026
Nº Inc. ou Roubo	0,0324	(+0,0249; +0,0398)	0,0038
Ind. Colisão	-0,0481	(-0.0557; -0.0405)	0,0039
Hiperparâmetro	Média <i>a posteriori</i>	IC 95%	sd
$\tau_g$	776,0178	(727,5512; 826,7817)	25,2492
$\tau_u$	97,2469	(49,6866; 175,6810)	32,5086
$\tau_v$	1848,2822	(265,1595; 5923,9904)	1539,5079
$\tau_\phi$	201,2357	(67,2357; 438,8471)	96,9871
$\tau_\delta$	938,7998	(737,7329; 1170,2523)	110,2751

\* Intercepto confundido com Sexo Feminino e Faixa Etária de 18 a 25 anos

Como os parâmetros estão na escala logarítmica, por uma questão de facilidade de interpretação, é aconselhável trazê-los para sua escala natural, através de sua função inversa  $\exp(\cdot)$ . A Tabela 8 apresenta os efeitos das variáveis em sua escala natural.

Tabela 8 Resumo dos efeitos fixos (escala natural) do Modelo VI\*

Parâmetro	Média <i>a posteriori</i>	LI <sub>IC95%</sub>	LS <sub>IC95%</sub>
Intercepto*	0,0405	0,0383	0,0427
Sexo Masculino	1,1294	1,1257	1,1330
FaixaEtaria26-35	0,8050	0,8009	0,8090
FaixaEtaria36-45	0,7425	0,7386	0,7464
FaixaEtaria46-55	0,7435	0,7397	0,7473
FaixaEtaria>55	0,7064	0,7028	0,7100
Nº Inc. ou Roubo	1,0329	1,0253	1,0406
Ind. Colisão	0,9530	0,9458	0,9602

\* Intercepto confundido com Sexo Feminino e Faixa Etária de 18 a 25 anos

Os efeitos que antes eram aditivos são, após a transformação, multiplicativos e estão descritos na Tabela 8. Para uma interpretação, o efeito do intercepto, que está confundido com condutores do sexo feminino e faixa etária de 18 a 25 anos, é de 0,0405, que significa um prêmio médio no valor de 4,05% do valor da importância segurada média. Para condutores do sexo masculino entre 18 e 25 anos há um aumento de 12,94% a mais em relação ao intercepto, isto é,  $1,1294 \times 0,0405 = 0,0457$ , que representa 4,57% do valor da importância segurada média. Para um condutor do sexo feminino entre 36 a 45 anos, por exemplo, equivale a 74,25% do valor em relação ao intercepto, isto é,  $0,7425 \times 0,0405 = 0,0301$  3,01% do valor da importância segurada média.

Para checar a calibração do modelo, Czado, Gneiting e Held (2009) sugerem a construção do histograma da transformação integral de probabilidade. Um modelo bem calibrado deve ter um histograma próximo da uniformidade. A Figura 8 apresenta o histograma da transformação integral de probabilidade que mostra estar próximo da uniforme, indicando que o modelo VI\* está bem calibrado.

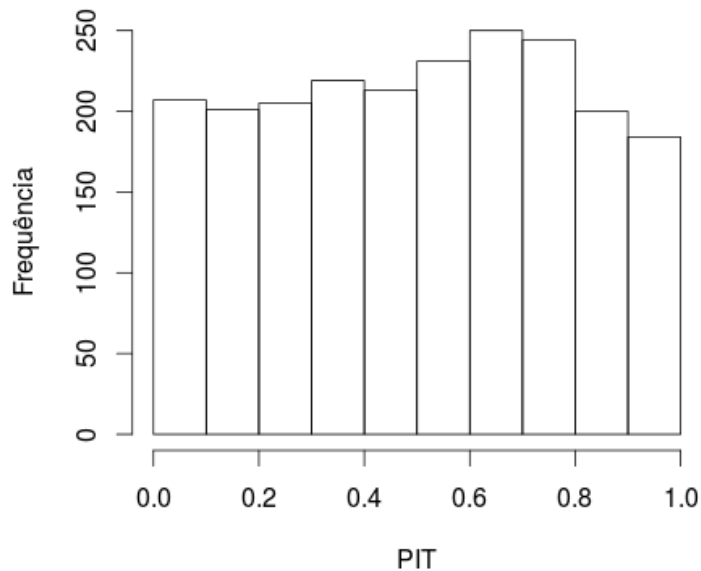


Figura 8 Histograma da transformação integral de probabilidade

A Figura 9 apresenta um gráfico de dispersão da média *a posteriori* da distribuição preditiva contra os valores médio do prêmio de seguro de automóveis. Salvo algumas exceções, pode ser observado que os valores preditos do prêmio médio estão bem próximos dos valores observados, o que é um bom indicativo do desempenho da distribuição preditiva *a posteriori*. Tais exceções, tratam-se das observações 166, 1111, 1112, 1247, 1258, 1895 que pertencem a classes de condutores que tiveram um valor médio do prêmio baixo, no entanto, possuem perfis em que o valor do prêmio médio é mais elevado. As observações destacadas foram retiradas do conjunto de dados e um novo ajuste do modelo foi realizado, porém não houve alterações inferenciais.



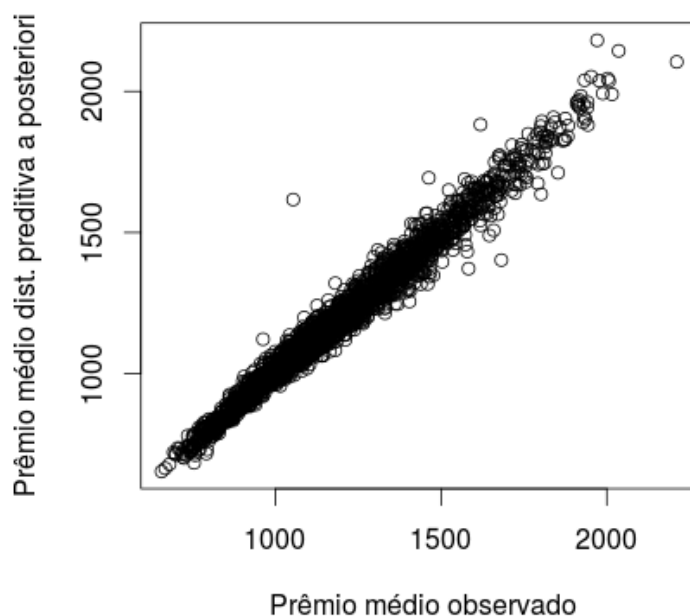


Figura 9 Dispersão entre os valores médios dos prêmios observados e os valores médios dos prêmios preditos

A Figura 10 apresenta a média a posteriori do efeito espacial para cada estado brasileiro. Destaques para os estados Rio de Janeiro e Pará (áreas mais escuras) que obtiveram os maiores efeitos dentre os estados, o que implica em uma maior influência desses estados no aumento do valor médio do prêmio de seguro de automóveis. Por outro lado, Santa Catarina, Sergipe e Paraíba (áreas em branco) foram os estados que tiveram efeitos espaciais menores do que 1, de modo que contribuíram para uma diminuição do valor médio do prêmio em seus respectivos estados. Ademais, todos os estados da região sul e a maioria da região nordeste, tiveram efeitos espaciais menores que 1 e de uma maneira geral, tem-se que as regiões norte, centro-oeste e sudeste apresentaram efeitos maiores que 1, exceto Minas Gerais, Distrito Federal e Mato Grosso do Sul em que tais efeitos

provocam uma redução no valor do prêmio médio.

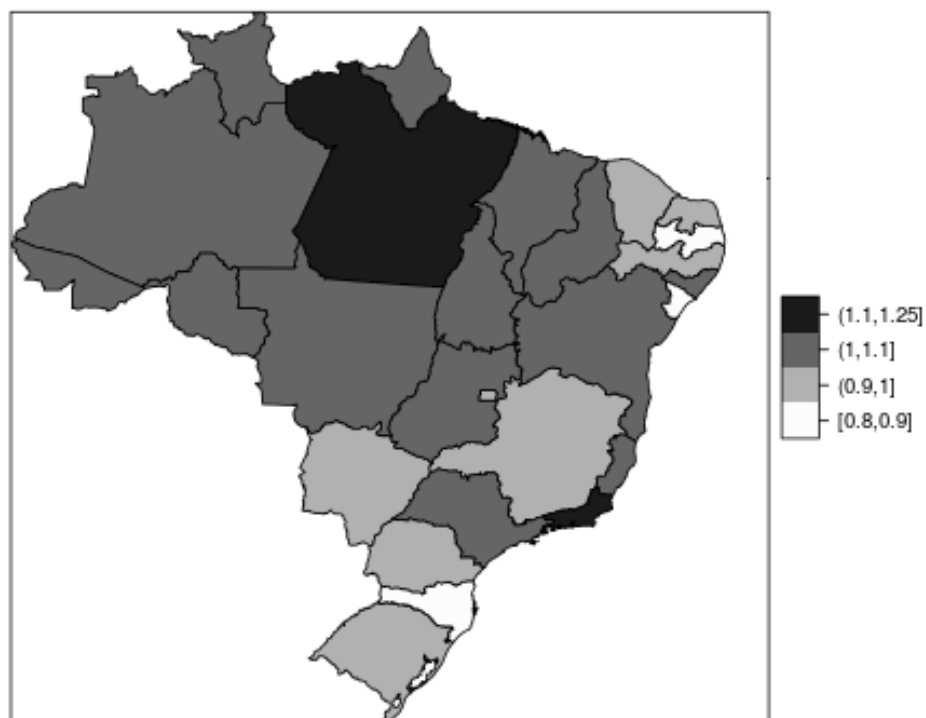


Figura 10 Média a posteriori para os efeitos espaciais para cada estado do Brasil ( $\zeta = \exp(\mathbf{u} + \mathbf{v})$ ) referente aos anos de 2007 a 2014

Com o auxílio da Figura 11, tem-se a probabilidade a posteriori de que o efeito espacial seja maior que 1, que auxilia no mapeamento da incerteza associada à média posteriori apresentada na Figura 10.

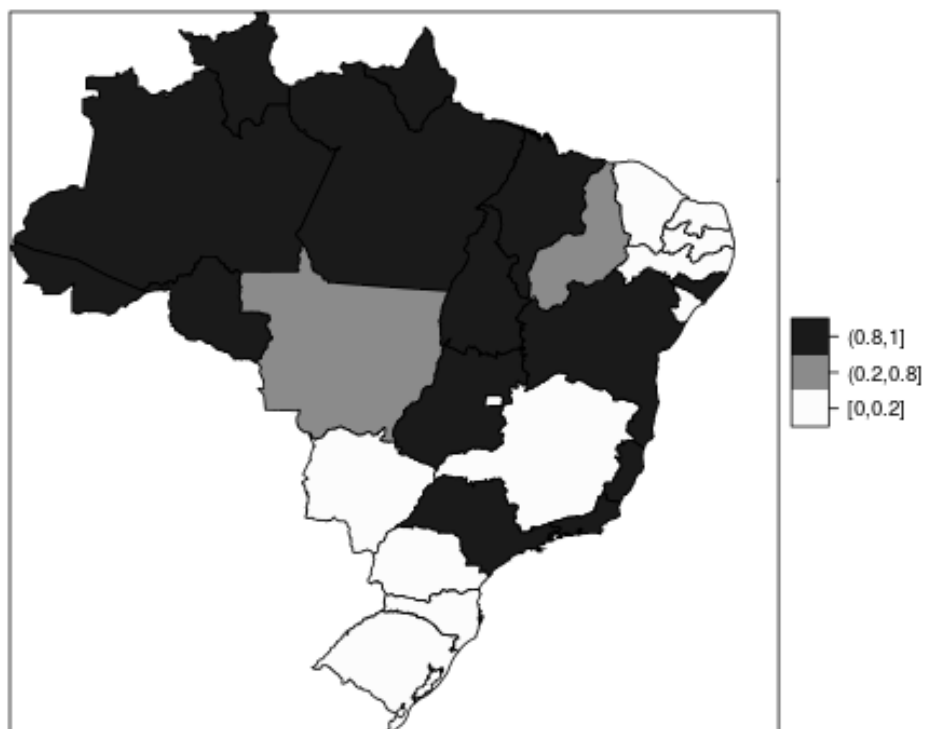


Figura 11 Probabilidade a posteriori  $p(\zeta_k > 1|\mathbf{y})$  para os efeitos espaciais para cada estado do Brasil referente aos anos de 2007 a 2014

Através da Figura 11, com alta probabilidade a posteriori as regiões sudeste, centro-oeste e norte, salvo exceções (cores brancas e cinzas dessas regiões), obtiveram um efeito espacial maior que 1 e, há uma probabilidade a posteriori muito baixa de que os efeitos espaciais dos estados das regiões sul e nordeste (com exceções de estados das cores pretas dessas regiões) sejam maiores que 1, isto é, provoquem um aumento no prêmio médio para suas áreas.

A Figura 12 mostra a média a posteriori da tendência temporal para os anos de 2007 a 2014. De uma maneira geral, houve um efeito crescente no valor médio do prêmio entre os anos de 2007 a 2014, com destaque para as quedas 2008, 2009 e 2011 e para as ascensões em 2013 e 2014.

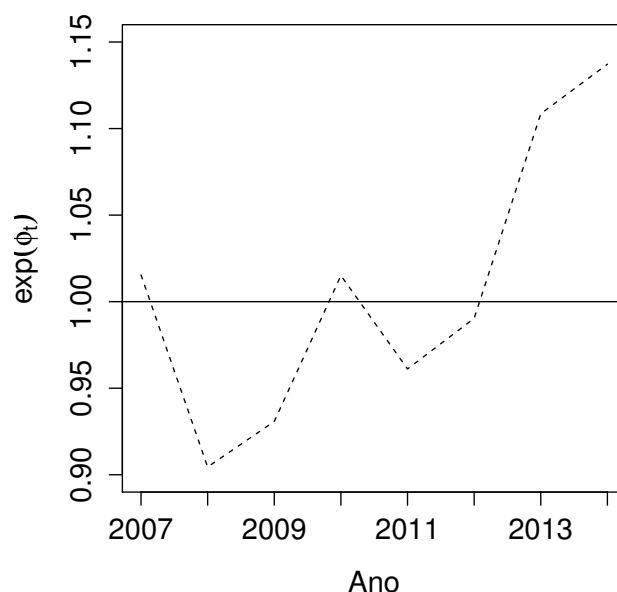


Figura 12 Tendência temporal da média a posteriori do valor médio do prêmio de seguro de automóveis no Brasil para os anos de 2007 a 2014

A Figura 13 mostra a média a posteriori da interação espaço-tempo para o valor do prêmio médio para os anos entre 2007 a 2014. Os estados com coloração mais escura em suas respectivas áreas, tiveram, para o ano específico, uma tendência temporal mais acentuada que a média da tendência temporal para o ano em questão. No ano de 2014, em que a média a posteriori da tendência temporal foi a mais alta dentre os anos considerados (ver Fig. 12) os estados como Rio de Janeiro, Espírito Santo, Amazonas, Bahia e todos os outros estados com coloração mais escura, tiveram uma média a posteriori da tendência temporal ainda mais

acentuada do que a média geral da tendência temporal para o ano de 2014. Por outro lado, em 2014, estados como o de Santa Catarina, Rio Grande do Sul, Maranhão e outros que possuem a área com coloração mais clara, tiveram uma média a posteriori da tendência temporal menos acentuada que a média a posteriori da tendência temporal para o ano de 2014.

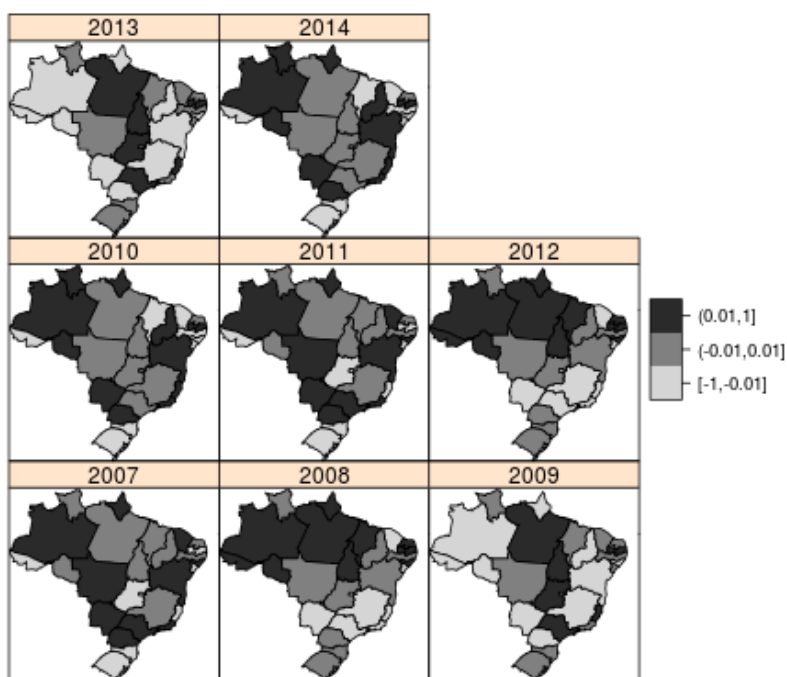


Figura 13 Média a posteriori da interação espaço-tempo  $\delta_{kt}$  para o valor médio do prêmio de seguro de automóveis no Brasil para os anos de 2007 a 2014

## 8.1 Discussões

Quanto ao modelo gama ajustado para o valor médio do prêmio de seguro de automóveis, cabe alguns destaques:

- O computador utilizado para os ajustes do modelo foi um Intel Core I7 CPU 3517U @ 1,9 GHz  $\times$  4 e o tempo de execução para o modelo selecionado foi de 14s utilizando a estratégia da aproximação gaussiana, 17s para a aproximação de Laplace simplificada e de 220s para a aproximação de Laplace (completa). Ademais, em termos inferenciais, não houve diferença entre as estimativas dos efeitos para o modelo ajustado com Aproximação de Laplace (completa) e Aproximação Gaussiana.
- Verifica-se que condutores do sexo feminino tem uma redução no prêmio em relação aos condutores do sexo masculino. A classe de condutores do sexo masculino entre 18 e 25 anos sofrem um aumento de 12,94% em relação ao prêmio médio pago por condutores do sexo feminino na mesma faixa etária que, inicialmente, pagam um prêmio médio equivalente a 4,05% da importância segurada média. Condutores com mais idade possuem uma redução do valor de seu prêmio em relação às outras classes de condutores, sendo ainda maior comparado com os que possuem entre 18 e 25 anos.
- Condutores dos estados das regiões Norte, Centro-oeste e Sudeste, exceto os estados de Minas Gerais, Mato Grosso do Sul e Brasília, tem um aumento na média do prêmio de seguro de automóveis, em especial os estados do Rio de Janeiro e Pará, áreas em que o aumento é ainda maior. Em contrapartida, os estados das regiões Sul e Nordeste, exceto os estados do Maranhão, Piauí, Bahia e Alagoas, tem uma diminuição no valor do prêmio médio, notadamente nos estados de Santa Catarina, Sergipe e Paraíba em que a redução é ainda maior.
- Existe uma tendência do valor médio do prêmio aumentar ao longo do anos, embora os nos de 2008, 2009, 2011 e 2012 se mostrem contrário a essa afirmação, há uma tendência de aumento nos valores médios dos prêmios

muito forte nos anos de 2013 e 2014 e esses valores devem manter-se ao menos no mesmo patamar durante os próximos anos.

- Na interação espaço-tempo, a tendência temporal da região Norte mostrou-se, na maioria dos anos, uma tendência temporal mais acentuada que a tendência temporal média dos respectivos anos. Os estados Rio Grande do Sul e Santa Catarina, pertencentes a região Sul, apresentaram, para os anos de 2007, 2010, 2011 e 2014, uma tendência temporal abaixo que a tendência temporal média para os seus respectivos anos. O estado do Paraná, por sua vez, apresentou uma tendência temporal acima da média, na maioria dos anos. Na região Sudeste, os estados de São Paulo e Rio de Janeiro apresentaram uma tendência temporal maior, enquanto que Minas Gerais, em sua maioria dos anos, apresentou uma tendência temporal menor que a média. O estado do Espírito Santo apresenta uma tendência temporal mais próxima da média. Já a região Nordeste, houve uma maior variação, com os seus estados ora apresentando uma tendência temporal maior, ora apresentando uma tendência temporal menor que a média.

## 9 CONCLUSÃO

Nessa dissertação foi apresentada uma análise bayesiana de modelos gaussianos latentes aplicados a dados de seguros de automóveis entre os anos de 2007 a 2014.

Com os resultados apresentados, pode-se observar que a metodologia INLA fornece uma nova, rápida e interessante maneira de aproximar as marginais a posteriori de uma ampla variedade de modelos bayesianos hierárquicos e pode ser utilizada por vários pesquisadores de diferentes áreas de aplicação. Dessa forma, a metodologia INLA representa um importante avanço no campo da estatística bayesiana e, por ser mais veloz, surge como uma boa alternativa às aproximações via MCMC, para o caso em que o modelo ajustado pertença à classe de modelos gaussianos latentes.

Observa-se que a modelagem que inclui efeitos agrupados (aleatórios) foi capaz de descrever melhor os dados de seguro de automóveis do que o modelo linear generalizado. Dentre esses modelos com efeitos agrupados, os mais precisos incluem efeitos espaciais.



## Referências

AGRESTI, A. **Categorical data analysis**. 3. ed. New Jersey: John Wiley & Sons, 2013. 792 p. (Wiley series in probability and statistics).

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, IEEE, v. 19, n. 6, p. 716–723, Dec. 1974.

ANTONIO, K.; BEIRLANT, J. Actuarial statistics with generalized linear mixed models. **Insurance: Mathematics and Economics**, Elsevier, v. 40, n. 1, p. 58–76, Jan. 2007.

BERNARDINELLI, L. et al. Bayesian analysis of space-time variation in disease risk. **Statistics in medicine**, Wiley Online Library, v. 14, n. 21-22, p. 2433–2443, Nov. 1995.

BESAG, J.; KOOPERBERG, C. On conditional and intrinsic autoregressions. **Biometrika**, Biometrika Trust, v. 82, n. 4, p. 733–746, Apr. 1995.

BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration, with two applications in spatial statistics. **Annals of the institute of statistical mathematics**, Springer, v. 43, n. 1, p. 1–20, Mar. 1991.

BIVAND, R.; LEWIN-KOH, N. **maptools: Tools for Reading and Handling Spatial Objects**. [S.l.], 2016. R package version 0.8-39. Disponível em: <<https://CRAN.R-project.org/package=maptools>>.

BIVAND, R.; PIRAS, G. Comparing implementations of estimation methods for spatial econometrics. **Journal of Statistical Software**, v. 63, n. 18, p. 1–36, 2015. Disponível em: <<http://www.jstatsoft.org/v63/i18/>>.

BLANGIARDO, M.; CAMELETTI, M. **Spatial and Spatio-temporal Bayesian Models with R-INLA**. 1. ed. Reino Unido: John Wiley & Sons, 2015. 308 p.

BOWERS, N. et al. **Actuarial Mathematics**. 2. ed. Schaumburg, Illinois: Society of Actuaries, 1997. 753 p.

BRASIL. Lei n. 10.406, de 10 de janeiro de 2002. **Código Civil**, Revista dos Tribunais, São Paulo, 2002.

CARLIN, B. P.; LOUIS, T. A. **Bayesian methods for data analysis**. 3. ed. New York, NY: CRC Press, 2008. 552 p.

CASELLA, G.; BERGER, R. L. **Statistical inference**. 2. ed. Austrália, Pacific Grove, CA: Thomson Learning, 2002. 660 p.

CAVALIERI FILHO, S. **Programa de responsabilidade civil**. 14. ed. São Paulo: Atlas, 2014. 648 p.

CHRISTENSEN, R. et al. **Bayesian ideas and data analysis: an introduction for scientists and statisticians**. New York, NY: CRC Press, 2011. 498 p.

CONGDON, P. D. **Applied Bayesian hierarchical methods**. New York, NY: CRC Press, 2010. 604 p.

CORDEIRO, G. M.; DEMÉTRIO, C. G. **Modelos lineares generalizados e Extensões**. Recife: UFRPE, 2010. 249 p.

CZADO, C.; GNEITING, T.; HELD, L. Predictive model assessment for count data. **Biometrics**, Wiley, International Biometric Society, v. 65, n. 4, p. 1254–1261, Dec. 2009.

DEMIDENKO, E. **Mixed models: theory and applications with R**. 2. ed. New York, NY: John Wiley & Sons, 2013. 754 p.

DENUIT, M. et al. **Actuarial theory for dependent risks: measures, orders and models**. New York, NY: John Wiley & Sons, 2006. 458 p.

DICKSON, D. C. **Insurance risk and ruin**. United Kingdom: Cambridge University Press, 2010. 242 p.

DOBSON, A. J.; BARNETT, A. **An introduction to generalized linear models**. 3. ed. New York, NY: CRC press, 2008. 320 p.

EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. **Modelling extremal events: for insurance and finance**. Berlin, Germany: Springer Science & Business Media, 1997. 644 p.

FARAWAY, J. J. **Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models**. New York, NY: CRC press, 2005. 312 p.

FONG, Y.; RUE, H.; WAKEFIELD, J. Bayesian inference for generalized linear mixed models. **Biostatistics**, Biometrika Trust, v. 11, n. 3, p. 397–412, Dec. 2009.

FREES, E. W.; DERRIG, R. A.; MEYERS, G. **Predictive modeling applications in actuarial science**. United Kingdom: Cambridge University Press, 2014. 565 p.

GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 74, n. 365, p. 153–160, Mar. 1979.

GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 85, n. 410, p. 398–409, Nov. 1990.

GELMAN, A. et al. **Bayesian data analysis**. New York, NY: Taylor & Francis, 2013. 675 p.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, n. 6, p. 721–741, Nov. 1984.

GNEITING, T.; BALABDAOUI, F.; RAFTERY, A. E. Probabilistic forecasts, calibration and sharpness. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 69, n. 2, p. 243–268, Mar. 2007.

HARRINGTON, S. E.; NIEHAUS, G. R.; HARRINGTON, N. Risk management and insurance. **The Journal of Finance**, JSTOR, v. 54, n. 3, p. 1187–1189, Jun. 1999.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, Biometrika Trust, v. 57, n. 1, p. 97–109, Apr. 1970.

JEFFREYS, H. **Theory of probability**. 3. ed. Oxford: Clarendon Press, 1961. 459 p.

JENNRICH, R. I.; SAMPSON, P. Newton-raphson and related algorithms for maximum likelihood variance component estimation. **Technometrics**, Taylor & Francis Group, v. 18, n. 1, p. 11–17, Feb. 1976.

JONG, P. D.; HELLER, G. Z. et al. **Generalized linear models for insurance data**. United Kingdom: Cambridge University Press Cambridge, 2008.

KAAS, R. et al. **Modern actuarial risk theory: using R**. 2. ed. New York, NY: Springer Science & Business Media, 2008. 382 p.

KELLY, D.; SMITH, C. **Bayesian inference for probabilistic risk assessment: a practitioner's guidebook**. London: Springer Science & Business Media, 2011. 228 p.

KNORR-HELD, L. Bayesian modelling of inseparable space-time variation in disease risk. **Statistics in medicine**, Wiley Online Library, v. 19, n. 1718, p. 2555–2567, Aug. 2000.

MARTINS, J. M. B. **O contrato de seguro: comentado conforme as disposições**

**do novo código civil, lei no. 10.406, de 10 de janeiro de 2002.** São Paulo: Forense Universitária, 2003. 212 p.

MARTINS, J. M. B.; MARTINS, L. d. S. **Direito de seguro: responsabilidade civil das seguradoras, doutrina, legislação, jurisprudência.** São Paulo: Forense Universitária, 2008. 323 p.

MARTINS, T. G. et al. Bayesian computing with inla: new features. **Computational Statistics & Data Analysis**, Elsevier, v. 67, p. 68–83, Nov. 2013.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models.** 2. ed. New York, NY: CRC press, 1989. 532 p.

MCCULLOCH, C. E.; SEARLE, S. R.; NEUHAUS, J. M. **Generalized, Linear, and Mixed Models.** 2. ed. New York, NY: Wiley, 2008. 424 p.

METROPOLIS, N. et al. Equation of state calculations by fast computing machines. **The journal of chemical physics**, AIP Publishing, v. 21, n. 6, p. 1087–1092, Jun. 1953.

NATARIO, I. Métodos computacionais: Inla, integrated nested laplace approximation. **Boletim Sociedade Portuguesa de Estatística**, n. 2, p. 52–57, 2013.

OHLSSON, E.; JOHANSSON, B. **Non-life insurance pricing with generalized linear models.** Berlin: Springer, 2010. 174 p.

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

PIMENTA, M. C. **Seguro de responsabilidade civil.** São Paulo: Atlas, 2010. 184 p.

PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-PLUS**. New York, NY: Springer Science & Business Media, 2006. 528 p.

PITA, R.; DOMINGUEZ, A. **Seguro de automóvel**. Rio de Janeiro: Escola Nacional de Seguros, 2011. 224 p.

PLUMMER, M. Penalized loss functions for bayesian model comparison. **Biostatistics**, Biometrika Trust, v. 9, n. 3, p. 523–539, Jul. 2008.

PRESS, S. J. **Subjective and objective Bayesian statistics: principles, models, and applications**. 2. ed. New York, NY: John Wiley & Sons, 2002. 600 p.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org/>>.

RENCHEER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. 2. ed. New York, NY: John Wiley & Sons, 2008. 688 p.

ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. 2. ed. New York, NY: Springer Science & Business Media, 2004. 649 p.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. **Journal of the royal statistical society: Series B (statistical methodology)**, Wiley Online Library, v. 71, n. 2, p. 319–392, Apr. 2009.

SARKAR, D. **Lattice: Multivariate Data Visualization with R**. New York: Springer, 2008. ISBN 978-0-387-75968-5. Disponível em: <<http://lmdvr.r-forge.r-project.org>>.

SCHRÖDLE, B. et al. Using integrated nested laplace approximations for the evaluation of veterinary surveillance data from switzerland: a case-study. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 60, n. 2, p. 261–279, Mar. 2011.

SCHWARZ, G. et al. Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, Mar. 1978.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York, NY: John Wiley & Sons, 2009. 536 p.

SHI, P. Insurance ratemaking using a copula-based multivariate tweedie model. **Scandinavian Actuarial Journal**, Taylor & Francis, v. 2016, n. 3, p. 198–215, 2016.

SILVA, I. O. **Curso de direito do seguro**. 2. ed. São Paulo: Saraiva, 2012. 311 p.

SOUZA, S. D. **Seguros: contabilidade, atuária e auditoria**. 2. ed. São Paulo: Saraiva, 2007. 225 p.

SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, Oct. 2002.

TSE, Y.-K. **Nonlife actuarial models: theory, methods and evaluation**. United Kingdom: Cambridge University Press, 2009. 542 p.

## APÊNDICE - Rotina utilizada para a metodologia INLA

Após baixar o mapa do Brasil no Ministério do Meio Ambiente (são três arquivos: .dbf, .shx e .shp) e após carregar o conjunto de dados para o ajuste do modelo. A análise ocorreu do seguinte modo:

```
# carrega o mapa na variável brasil.
brasil <- readShapePoly("estadosl_2007")

# data.frame com os dados a serem analisados
data <- data.frame(NAME = UF, y = y, x1 = x1,
x2 = x2, IS_MEDIA = IS_MEDIA , u1 = u1, u2 = u2,
nincendio = nincendio, ncolisao = ncolisao,
noutras = noutras, indcol = indcol,
indinc = indinc, indout = indout)

# número de estados (27)
Nareas <- length(brasil$SIGLAUF3)

# Nesta parte criamos o respectivo gráfico
temp <- poly2nb(brasil)
nb2INLA("LDN.graph", temp)

#Assim criamos um arquivo chamado ``LDN-INLA.adj``
# próprio para o INLA

LDN.adj <- paste(getwd(), "/LDN.graph", sep="")

# A ordem dos dados necessita ser a mesma entre
# os dados e o objeto polígono espacial, que
# informa a estrutura de viinhança assim,
# necessitamos reordenar os dados de modo que
# fique de acordo

boroughs <- brasil
data.boroughs <- attr(boroughs, "data")
order <- match(data.boroughs$SIGLAUF3,
```



```

data$NAME)
order      <- rep(order, each = 80)
ID         <- rep(1:27, each = 80)
data       <- cbind(order, ID, data)

#### interacao espaco tempo
ID.area.year <- rep(1:(8*27),10)

formula.int <- PREMIO_MEDIO ~ SEXO + FAIXA_ETARIA +
nincendio + indcol +
# modela a componente espacial
f(u1,model="bym",graph="LDN.graph") +
# modela a componente temporal
f(u2, model = "iid") +
# modela a interação espaço-tempo
f(ID.area.year, model="iid")

inla.gamma5 <- inla(formula.int, data = data,
# offset importancia segurada
offset = log(IS_MEDIA) ,
# Reposta Gama
family = "gamma",
# Estratégia de aprox. Laplace Completa
control.inla = list(strategy = "laplace"),
control.predictor = list(compute = TRUE),
# Computar o DIC e a CPO
control.compute = list(dic=TRUE, cpo = TRUE))

# obtém o resumo dos efeitos do modelo
summary(inla.gamma5)

# DIC Total
inla.gamma5$dic$dic

# DIC Segregado
round(inla.gamma5$dic$p.eff, 2)

round(inla.gamma5$dic$mean.deviance, 2)

```

```

# Obtém a LPML
sum(log(inla.gamma5$cpo$cpo))

# Mostra se houve falha na estimação
# Não houve falha caso o valor igual 0
# Houve falha caso o valor seja 1
# Retorna a soma das falhas

sum(inla.gamma5$cpo$failure)

#Plota o histograma da
#transformação integral de probabilidade
hist(inla.gamma5$cpo$pit, 10, freq = FALSE)

# Obtém os y* preditos
yhat <- inla.gamma5$summary.fitted.values$mean

# plota o gráfico de dispersão entre o
# predito e o observado
plot(inla.gamma5$summary.fitted.values$mean,
PREMIO_MEDIO)

#obter o mapa de interacao espaco-tempo
delta.intI <- data.frame(
delta=inla.gamma5$summary.random$ID.area.year[,2],
year=data$u2, ID.area=data$u1)

delta.intI.matrix <- matrix(delta.intI[,1],
270,8,byrow=FALSE)
rownames(delta.intI.matrix)<- delta.intI[1:270,3]

# define os intervalos das interações
cutoff.interaction <- c(-1,-0.01,0.01,1)

delta.intI.factor <- data.frame(NAME=data$NAME)
for(i in 1:8){
delta.factor.temp <- cut(delta.intI.matrix[,i],
breaks=cutoff.interaction,include.lowest=TRUE)

```

```

delta.intI.factor <- cbind(delta.intI.factor,
  delta.factor.temp)
}

colnames(delta.intI.factor)<- c("NAME",
  seq(2007,2014))

attr(london.gen, "data") <- cbind(attr(london.gen,
  "data"),
  intI=delta.intI.factor)
trellis.par.set(axis.line=list(col=NA))
sppplot(obj=london.gen,
zcol=c("intI.2007","intI.2008",
"intI.2009","intI.2010",
"intI.2011","intI.2012",
"intI.2013","intI.2014"),
col.regions=gray(2.5:0.5/3),
names.attr=seq(2007,2014),main="")

# obter mapa espacial dos efeitos e
# da prob a posteriori de tais efeitos
# sejam maior que 1

csi <- inla.gamma5$marginals.random$u1[1:Nareas]
zeta <- lapply(csi,
function(x) inla.emarginal(exp,x))
#obter o mapa
#intervalos
zeta.cutoff <- c(0.8,0.90,1.0,1.1,1.25)

# transformar zeta em variavel categorica

cat.zeta <- cut(unlist(zeta),breaks=zeta.cutoff,
include.lowest=TRUE)
table(cat.zeta)
# Cria um data.frame com todas informações
# necessarias para o mapa

maps.cat.zeta <- data.frame(ID=1:27,

```

```

cat.zeta=cat.zeta)

# Adiciona a variavel categorcia zeta
# ao poligono espacial

data.boroughs <- attr(london.gen, "data")
attr(london.gen, "data") <- cbind(attr(london.gen,
"data"),
  maps.cat.zeta)

spplot(obj=london.gen, zcol= "cat.zeta",
col.regions=gray(seq(0.99,0.1,length=4)),
  asp=1)
####
a <- 0
csi <- inla.gamma5$marginals.random$u1[1:Nareas]
prob.csi <- lapply(csi, function(x)
{1 - inla.pmarginal(a, x)})
zeta.cutoff <- c(0,0.2,0.8,1)
cat.zeta <- cut(unlist(prob.csi),
breaks=zeta.cutoff,
include.lowest=TRUE)
table(cat.zeta)
# Cria um data.frame com todas
# informações necessarias para o mapa
maps.cat.zeta <- data.frame(ID=1:27,
  cat.zeta=cat.zeta)

#Adiciona a variavel categorcia zeta
#ao poligono espacial
data.boroughs <- attr(london.gen, "data")
attr(london.gen, "data") <- cbind(
attr(london.gen,
  "data"),
maps.cat.zeta)

spplot(obj=london.gen, zcol= "cat.zeta",
col.regions=gray(seq(0.99,0.1,length=3)), asp=1)

```

```
temporal.CAR <-lapply(  
  inla.gamma5$marginals.random$ID.year1,  
  function(X){  
    marg <- inla.tmarginal(  
      function(x) exp(x), X)  
      inla.emarginal(mean, marg)  
    })  
temporal.IID <- lapply(  
  inla.gamma5$marginals.random$u2,  
  function(X){  
    marg <- inla.tmarginal(  
      function(x) exp(x), X)  
      inla.emarginal(mean, marg)  
    })  
  
# Obter o gráfico temporal  
ano <- 2006 + 1:8  
plot(1:8, seq(0.9, 1.15, length=8),  
     type="n", xlab="t",  
     ylab=expression(exp(phi[t])))  
lines(unlist(temporal.IID))  
lines(vetor, lty=2)  
lines(unlist(temporal.CAR), lty=2)  
abline(h=1, lty=1)
```