

**Anna Christina Teixeira da Silva Ferreira**

**Mortalidade Materna no Estado do Rio de Janeiro: Aplicação de Modelos Lineares Generalizados na Análise de Fatores de Risco**

Niterói - RJ, Brasil

11 de Dezembro de 2023

**Anna Christina Teixeira da Silva Ferreira**

**Mortalidade Materna no Estado do  
Rio de Janeiro: Aplicação de  
Modelos Lineares Generalizados na  
Análise de Fatores de Risco**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Mariana Albi de Oliveira Souza  
Co-Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Lusié Velozo da Costa


Niterói - RJ, Brasil

11 de Dezembro de 2023

**Anna Christina Teixeira da Silva Ferreira**


**Mortalidade Materna no Estado do Rio de Janeiro:  
Aplicação de Modelos Lineares Generalizados na  
Análise de Fatores de Risco.**

Monografia de Projeto Final de Graduação sob o título “*Mortalidade Materna no Estado do Rio de Janeiro: Aplicação de Modelos Lineares Generalizados na Análise de Fatores de Risco.*”, defendida por Anna Christina Teixeira da Silva Ferreira e aprovada em 11 de Dezembro de 2023, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Documento assinado digitalmente  
 **MARIANA ALBI DE OLIVEIRA SOUZA**  
Data: 14/12/2023 10:55:09-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Mariana Albi de Oliveira Souza**  
Departamento de Estatística – UFF

Documento assinado digitalmente  
 **PATRICIA LUSIE VELOZO DA COSTA**  
Data: 13/12/2023 11:41:51-0300  
Verifique em <https://validar.iti.gov.br>

---


**Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – UFF

Ana Beatriz Monteiro Fonseca  
abmfonseca@id.uff.br:88872564700

Assinado de forma digital por Ana Beatriz Monteiro  
Fonseca abmfonseca@id.uff.br:88872564700  
Dados: 2023.12.12 15:49:33 -03'00'

---

**Prof. Dr<sup>a</sup>. Ana Beatriz Monteiro Fonseca**  
Departamento de Estatística – UFF

Documento assinado digitalmente  
 **IURI DA COSTA LEITE**  
Data: 13/12/2023 11:16:44-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Iuri da Costa Leite**  
Escola Nacional de Saúde Coletiva Sérgio Arouca – ENSP

Niterói, 11 de Dezembro de 2023

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

F383m Ferreira, Anna Christina Teixeira da Silva  
Mortalidade Materna no Estado do Rio de Janeiro: Aplicação de Modelos Lineares Generalizados na Análise de Fatores de Risco / Anna Christina Teixeira da Silva Ferreira. - 2023. 72 f.

Orientador: Mariana Albi de Oliveira Souza.  
Coorientador: Patrícia Lusié Velozo da Costa.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2023.

1. Estatística. 2. Modelos Lineares Generalizados. 3. Mortalidade Materna. 4. Inferência Bayesiana. 5. Produção intelectual. I. Souza, Mariana Albi de Oliveira, orientadora. II. Costa, Patrícia Lusié Velozo da, coorientadora. III. Universidade Federal Fluminense. Instituto de Matemática e Estatística. IV. Título.

CDD - XXX

# Resumo

Mortalidade materna ocorre quando uma mulher morre durante a gravidez ou nos 42 dias seguintes após o parto, excluindo os casos de morte por causas acidentais. Segundo a Organização Pan-Americana de Saúde, a maioria destas mortes poderiam ser evitadas e ocorrem em países em desenvolvimento. Este trabalho analisou e modelou o número de mortes maternas no estado do Rio de Janeiro. Inicialmente, foram realizadas análises descritivas para examinar tendências temporais, padrões e características dos óbitos maternos no período de 2010 a 2019. Posteriormente usou-se um Modelo de Regressão Poisson para analisar a associação entre as variáveis explicativas e o número de óbitos maternos, investigando desta forma a influência de diferentes fatores, tais como idade, escolaridade, estado civil e raça/cor, nos óbitos maternos. Por fim, o uso da análise exploratória espacial teve por objetivo investigar a distribuição geográfica da mortalidade materna no estado do Rio de Janeiro, identificando agrupamentos espaciais de altas e baixas taxas de mortalidade. Os resultados revelaram padrões consistentes ao longo do período investigado, indicando uma redução significativa na mortalidade materna para mulheres abaixo de 40 anos em comparação com aquelas acima dessa idade. Por outro lado, a ausência de parceiro foi consistentemente associada a taxas de mortalidade materna mais elevadas. A análise da escolaridade apontou uma correlação inversa entre níveis educacionais e taxas de mortalidade materna, sugerindo uma tendência para taxas mais altas entre mulheres com menor grau de instrução. Além disso, a variável raça/cor mostrou uma associação entre mulheres não brancas e maiores taxas de mortalidade materna em comparação com mulheres brancas, destacando a necessidade de investigações mais profundas das disparidades socioeconômicas e de acesso aos serviços de saúde. Estes resultados enfatizam a urgência de políticas públicas e intervenções específicas para grupos vulneráveis, visando reduzir as taxas de mortalidade materna. Estratégias que promovam a educação, o acesso equitativo à saúde e o suporte social para mulheres nessas condições podem desempenhar um papel crucial na mitigação desses desafios. Recomenda-se a continuidade deste estudo, considerando análises mais detalhadas e a inclusão de outras variáveis socioeconômicas e de saúde para uma compreensão mais abrangente e eficaz. Futuros trabalhos podem explorar a dinâmica espacial da mortalidade materna, utilizando técnicas avançadas de modelagem espacial e incorporando variáveis contextuais para uma compreensão preditiva dos padrões e determinantes da mortalidade materna em diferentes regiões.

Palavras-chave: Estatística. Modelos Lineares Generalizados. Mortalidade Materna. Inferência Bayesiana.

# Sumário

**Lista de Figuras**

**Lista de Tabelas**

<b>Lista de Abreviações</b>	p. 11
<b>1 Introdução</b>	p. 12
1.1 Objetivos . . . . .	p. 14
1.2 Organização . . . . .	p. 14
<b>2 Materiais e Métodos</b>	p. 15
2.1 Mortalidade Materna . . . . .	p. 15
2.2 Modelos Lineares . . . . .	p. 19
2.2.1 Modelos Lineares Generalizados . . . . .	p. 21
2.2.1.1 Modelos para dados de contagem . . . . .	p. 22
2.3 Inferência Bayesiana . . . . .	p. 25
2.3.1 Estimação Bayesiana . . . . .	p. 26
2.3.1.1 Estimação Pontual . . . . .	p. 26
2.3.1.2 Estimação Intervalar . . . . .	p. 27
2.3.2 MLG Poisson . . . . .	p. 27
2.3.3 Monte Carlo via Cadeias de Markov . . . . .	p. 28
2.3.4 Amostrador de Gibbs . . . . .	p. 29
2.3.5 Algoritmo de Metropolis-Hastings . . . . .	p. 30

2.3.6	Monte Carlo Hamiltoniano . . . . .	p. 31
2.3.7	Teste de Convergência . . . . .	p. 32
2.4	Estatística Espacial . . . . .	p. 34
2.4.1	Dados de Área . . . . .	p. 35
2.4.1.1	Abordagens exploratórias para dados de área . . . . .	p. 36
2.4.1.2	Índice de Moran Global . . . . .	p. 37
<b>3</b>	<b>Análise dos Resultados</b>	p. 39
3.1	Análise Descritiva . . . . .	p. 39
3.2	Ajuste do Modelo . . . . .	p. 44
3.3	Análise Espacial . . . . .	p. 52
<b>4</b>	<b>Conclusões</b>	p. 57
	<b>Referências</b>	p. 59
	<b>Apêndice A – Cadeias via Monte Carlo Hamiltoniano dos anos de 2011 a 2018</b>	p. 61
	<b>Apêndice B – Função de Autocorrelação dos anos de 2011 a 2018</b>	p. 69

# Lista de Figuras

1	Mapa do estado do Rio de Janeiro por municípios. . . . .	p. 17
2	Exemplo de Critério de vizinhança por contiguidade Queen nos municípios dos Estado do Rio de Janeiro . . . . .	p. 37
3	Razão de Mortalidade Materna no Estado do Rio de Janeiro - Série Anual	p. 40
4	Razão de Mortalidade Materna no Estado do Rio de Janeiro - Variáveis: Faixa etária, Raça/Cor, Escolaridade e Estado Civil . . . . .	p. 43
5	Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2010. . . . .	p. 45
6	Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2019. . . . .	p. 46
7	Função de Autocorrelação dos valores gerados para os coeficientes - 2010.	p. 47
8	Função de Autocorrelação dos valores gerados para os coeficientes - 2019.	p. 48
9	Mapa Coroplético da Razão de Mortalidade Materna (RMM) do Estado do Rio de Janeiro de 2010. . . . .	p. 53
10	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2011. . . .	p. 53
11	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2012. . . .	p. 53
12	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2013. . . .	p. 54
13	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2014. . . .	p. 54
14	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2015. . . .	p. 54
15	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2016. . . .	p. 55
16	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2017. . . .	p. 55
17	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2018. . . .	p. 55
18	Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2019. . . .	p. 56



19	Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2011. . . . .	p. 61
20	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2012. . . . .	p. 62
21	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2013. . . . .	p. 63
22	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2014. . . . .	p. 64
23	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2015. . . . .	p. 65
24	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2016. . . . .	p. 66
25	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2017. . . . .	p. 67
26	Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2018. . . . .	p. 68
27	Função de Autocorrelação dos valores gerados para os coeficientes - 2011.	p. 69
28	Função de Autocorrelação dos valores gerados para os coeficientes - 2012.	p. 70
29	Função de Autocorrelação dos valores gerados para os coeficientes - 2013.	p. 70
30	Função de Autocorrelação dos valores gerados para os coeficientes - 2014.	p. 71
31	Função de Autocorrelação dos valores gerados para os coeficientes - 2015.	p. 71
32	Função de Autocorrelação dos valores gerados para os coeficientes - 2016.	p. 72
33	Função de Autocorrelação dos valores gerados para os coeficientes - 2017.	p. 72
34	Função de Autocorrelação dos valores gerados para os coeficientes - 2018.	p. 73

# Lista de Tabelas

1	Variáveis disponíveis pelo DATASUS e selecionadas para o estudo. . . .	p. 18
2	Exemplo de algumas linhas do banco de dados utilizados, referentes ao ano de 2019. . . . .	p. 19
3	Distribuição dos Óbitos Maternos, Nascidos Vivos e da Razão de Mortalidade Materna. . . . .	p. 40
4	Distribuição dos Óbitos Maternos, Nascidos Vivos e a Razão por variáveis, no período de 2010 a 2019. . . . .	p. 41
5	Distribuição dos Óbitos Maternos, Nascidos Vivos e a Razão por variáveis dicotômicas, no período de 2010 a 2019. . . . .	p. 42
6	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2010 . . . . .	p. 49
7	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2011 . . . . .	p. 49
8	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2012 . . . . .	p. 49
9	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2013 . . . . .	p. 50
10	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2014 . . . . .	p. 50
11	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2015 . . . . .	p. 50
12	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2016 . . . . .	p. 50
13	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95% Ano: 2017 . . . . .	p. 51

14	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95% Ano: 2018 . . . . .	p. 51
15	Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2019 . . . . .	p. 51
16	Autocorrelação espacial dos dados de 2010 a 2019 no estado do Rio de Janeiro. . . . .	p. 52

# Lista de Abreviações

**CID-10** 10<sup>a</sup> versão da Classificação Internacional de Doenças

**MCH** Monte Carlo Hamiltoniano

**MCMC** Monte Carlo via cadeias de Markov

**MLG** Modelo Linear Generalizado

**MH** Metropolis-Hastings

**OMS** Organização Mundial da Saúde

**ONU** Organização das Nações Unidas

**OPAS** Organização Pan-Americana da Saúde

**RMM** Razão de Mortalidade Materna

**SIM** Sistema de Informação sobre Mortalidade

# 1 Introdução

A avaliação do nível de vida da população tem sido um tema de grande interesse há muito tempo, não apenas para os Governos (Federais, Estaduais e Municipais), mas também para organizações não governamentais. Entre as diversas medidas utilizadas, destacam-se os indicadores de saúde, sejam em nível de países, estados ou municípios. Segundo Medronho et al. (2011), esses indicadores de saúde são métricas que fornecem informações sobre a saúde da população, permitindo monitorar e avaliar aspectos como morbidade, mortalidade, acesso aos serviços de saúde, qualidade de vida e outros determinantes da saúde. Um desses indicadores, que auxilia na identificação de problemas de saúde, na formulação de políticas públicas, no planejamento de ações de prevenção e no monitoramento da eficácia das intervenções implementadas, é chamado de Razão de Mortalidade Materna.

Reduzir o número de ocorrências de mortes maternas tem sido uma prioridade de nível global e esta meta foi inclusa nas metas dos Objetivos de Desenvolvimento Sustentável e na Agenda 2030 da Organização das Nações Unidas (ONU). Além disso, diversos países, ao estudarem as causas de morte materna, apresentaram resultados indicativos da negligência em direitos humanos das mulheres e da falta de atenção à saúde primária, pois a grande maioria das mortes poderiam ter sido evitadas se estivessem disponíveis tratamentos e cuidados essenciais.

Segundo Motta e Moreira (2021), a compreensão desta mortalidade exige uma abordagem que leve em consideração tanto as desigualdades regionais quanto questões como: qualidade da informação disponibilizada; políticas, programas e ações voltados para o enfrentamento da mortalidade materna; cuidados relacionados ao ciclo gravídico-puerperal; fatores de risco associados a essas mulheres e aborto ilegal.

Ruas et al. (2020) acreditam que é necessário ampliar o conhecimento acerca do perfil e da distribuição espacial da mortalidade materna para melhor entendimento sobre as causas que levam ao óbito das mulheres, com vistas a se produzir e divulgar informações

que possam subsidiar o planejamento de ações e políticas públicas que busquem a redução da mortalidade materna no cenário estudado. Essa abordagem metodológica permite interpretar os dados, levando em consideração a dinâmica espaço-temporal dos casos e identificar dependências espaciais.

Em estudos quantitativos de fenômenos relacionados à saúde, com o objetivo de compreender os resultados e os fatores que os explicam, os conceitos de epidemiologia e estatística espacial usualmente são abordados. Medronho et al. (2011) indicam que a epidemiologia espacial utiliza métodos estatísticos para analisar dados geográficos e espaciais, buscando identificar padrões e tendências que possam fornecer percepções importantes sobre a distribuição e os determinantes de doenças e outros eventos de saúde. Além disso, destaca-se que dados espaciais são definidos como sendo qualquer tipo de dado que descreva fenômenos aos quais esteja associada alguma dimensão espacial, enquanto os dados geográficos são dados cuja dimensão espacial está associada à sua localização na superfície da Terra num determinado instante ou período de tempo.

Percebe-se que, ao longo dos anos, a Mortalidade Materna continua a ser um desafio significativo de saúde pública em todo o mundo, representando não apenas um indicador crucial da qualidade dos cuidados de saúde, mas também um reflexo da complexidade das interações entre fatores biológicos, socioeconômicos e de acesso aos serviços médicos. A compreensão desses padrões complexos e a identificação de fatores associados à mortalidade materna são elementos essenciais na formulação de estratégias eficazes para reduzir esses trágicos eventos. É nesse contexto que os modelos estatísticos se destacam como ferramentas indispensáveis na análise e interpretação dos dados epidemiológicos.

Uma metodologia bastante empregada na tentativa de solucionar os problemas citados envolve o ajuste de Modelos Lineares Generalizados, como destacado por Conceição, Saldiva e Singer (2001). Esses modelos se apresentam como uma abordagem versátil e poderosa para analisar dados complexos, permitindo a incorporação de diferentes distribuições de probabilidade e funções de ligação adequadas à natureza dos dados em estudo. No entanto, a escolha e a aplicação adequada desses modelos tornam-se importantes para compreender as particularidades das associações entre variáveis explanatórias e a taxa de mortalidade materna.

Em Conceição, Saldiva e Singer (2001) ainda é mencionado que a escolha do modelo apropriado é de extrema importância, dada a difícil rede de relações presentes nos dados epidemiológicos. Modelos clássicos, como os de regressão linear, embora valiosos, podem não capturar a realidade de maneira abrangente, especialmente quando se trata

de fenômenos complexos, como a Mortalidade Materna, que frequentemente apresentam associações não-lineares e interações multifacetadas entre os fatores envolvidos.

## 1.1 Objetivos

O objetivo deste trabalho está na compreensão do comportamento dos dados de mortalidade materna no estado do Rio de Janeiro. Será usado um modelo de regressão generalizado sob a perspectiva Bayesiana para inferir sobre características da mortalidade materna no estado do RJ.

Os objetivos específicos do trabalho são:

- realizar uma análise descritiva e inferencial no conjunto de dados reais de mortalidade materna;
- estudar sobre modelos de regressão Poisson generalizados, que serão utilizados neste trabalho para compreender os fatores associados à mortalidade materna;
- utilizar Inferência Bayesiana para estimar os parâmetros desconhecidos do modelo de interesse;
- realizar uma análise espacial para dados de área, que possibilitará a identificação de padrões espaciais e a avaliação de autocorrelação espacial dos dados de mortalidade materna, permitindo propostas de modelagens futuras no conjunto de dados em estudo.

## 1.2 Organização

A estrutura deste trabalho está disposta em 4 capítulos. O segundo capítulo compõe-se da metodologia usada, onde são apresentadas 5 seções de conceitos e técnicas para descrever a Razão de Mortalidade Materna, Modelos Lineares Generalizados, Inferência Bayesiana, Métodos de Monte Carlo via Cadeias de Markov e Estatística Espacial, respectivamente. O terceiro capítulo apresenta as aplicações das técnicas expostas em um conjunto de dados reais de mortalidade materna, onde são apresentados os resultados obtidos. E por fim, o quarto capítulo finaliza o trabalho apresentando as conclusões obtidas por meio das análises e sugerindo extensões deste trabalho.

## 2 Materiais e Métodos

Neste Capítulo são apresentados os materiais usados e as metodologias implementadas para realização das análises dos dados, tendo como principais referências os livros de Banerjee, Carlin e Gelfand (2004) e Dobson (2018). Na Seção 2.1 há uma descrição dos dados obtidos sobre Mortalidade Materna. A Seção 2.2 contém uma revisão de modelos lineares e, em especial de Modelos Lineares Generalizados, que serão aplicados aos dados. Para inferir sobre os parâmetros desconhecidos no modelo, há um resumo sobre Inferência Bayesiana e sobre Métodos de Monte Carlo via Cadeias de Markov, na Seção 2.3. A Seção 2.4 introduz a área da Estatística Espacial e suas principais ferramentas exploratórias.

### 2.1 Mortalidade Materna

Conforme a Organização Pan-Americana da Saúde (OPAS) e Organização Mundial da Saúde (OMS), na 10<sup>a</sup> versão da Classificação Internacional de Doenças (CID-10), Morte Materna é definida como *...a morte de uma mulher durante a gestação ou dentro de um período de 42 dias após o término da gestação, independentemente da duração ou da localização da gravidez, devido a qualquer causa relacionada com ou agravada pela gravidez ou por medidas em relação a ela, porém não devido a causas acidentais ou incidentais*".

Os indicadores de saúde são ferramentas fundamentais utilizadas para medir, avaliar e monitorar diferentes aspectos do estado de saúde de uma população. Esses indicadores fornecem informações valiosas para profissionais de saúde, pesquisadores, formuladores de políticas e outros interessados, contribuindo para a compreensão da saúde de uma comunidade e aprimorando a tomada de decisões em saúde pública. São expressos na maioria das vezes por medidas de frequência como proporções, taxas ou coeficientes. Entretanto, também podem ser expressos em unidades de tempo, como nos casos dos anos de vida. As taxas, usualmente, são utilizadas para estimar o risco de ocorrência de um problema de saúde, adoecimento ou morte. Em Medronho et al. (2011) é apresentado um indicador de saúde considerado excelente não apenas para descrever a saúde da



mulher, mas também para a população como um todo, chamado de razão de mortalidade materna (RMM) e definido como o número de óbitos maternos por 100 mil nascidos vivos, possuindo então a seguinte forma:

$$RMM = \frac{\text{Óbitos Maternos}}{\text{Nascidos Vivos}} \cdot 100.000, \quad (2.1)$$

tendo no numerador o número de mortes maternas ocorridas até 42 dias após o término da gravidez, atribuídos a causas ligadas à gestação, ao parto e/ou ao puerpério, de mães residentes em determinado espaço geográfico, no ano considerado (com exclusão das mortes tardias e por sequelas obstétricas). No denominador, coloca-se a quantidade de nascidos vivos de mães também residentes no mesmo espaço geográfico, na mesma unidade de tempo.

Este é o indicador de saúde usado mundialmente para dimensionar a magnitude da mortalidade materna. Também permite a comparação da mortalidade materna entre diferentes regiões geográficas e ao longo do tempo, sendo fundamental para identificar disparidades e tendências, facilitando a implementação de medidas preventivas e de intervenção.

Existem dois grupos de causas obstétricas que estão dentro das mortes maternas, denominadas como morte obstétrica direta e indireta. Segundo SAÚDE (2009), os grupos possuem as seguintes definições:

- mortes obstétricas diretas: são resultantes de eventos que podem ocorrer durante a gravidez, no parto ou no puerpério, decorrentes de omissões, negligências, tratamentos incorretos, dentre outras cadeias de eventos resultantes dessas causas;
- mortes obstétricas indiretas: resultantes de doenças que existiam antes da gestação ou que tiveram desenvolvimento durante esse período, agravadas pelos efeitos fisiológicos da gravidez e que não foram provocadas por causas diretas como, por exemplo, a Diabetes Mellitus.

A OPAS e OMS, na CID-10, definiram que os óbitos do tipo direto incluem seis grupos: aborto, hipertensão, hemorragias, infecções, outras complicações obstétricas e complicações não antecipadas (em geral, relacionadas a procedimentos anestésicos), enquanto que as causas indiretas são compostas de apenas um grupo, que reúne todas as complicações não obstétricas e, por último, as causas desconhecidas representadas pelo código O95.

Além desses dois grupos, existem os casos de morte materna não obstétrica (também chamada de “morte não relacionada”) que são resultantes de causas acidentais ou incidentais que não possuem relação com a gravidez.

Por fim, define-se como mortes maternas tardias os óbitos ocorridos após os 42 dias, mas inferior a um ano após o término da gravidez, considerando tanto os casos de morte por causas obstétricas diretas ou indiretas. Estes tipos de óbitos não são inclusos no cálculo da razão de mortalidade materna (RMM). Além disso, se ocorrer um caso de óbito após mais de um ano do término da gravidez, pode-se ter uma morte por sequelas obstétricas.

A definição de mulher em idade fértil é internacionalmente entendida como aquela que tem idade entre 15 e 49 anos. No entanto, no Brasil, considera-se para essa definição a faixa etária dos 10 aos 49 anos, por constar historicamente idades inferiores a 15 anos em registros de gestações em bases de dados do Ministério da Saúde.

Conforme descrito no Capítulo 1, o principal objetivo deste trabalho está na compreensão do comportamento dos dados de mortalidade materna no estado do Rio de Janeiro; em particular, foram utilizados dados de casos de mortes maternas ocorridas nos municípios do estado do Rio de Janeiro, entre os anos de 2010 e 2019.

A área de estudo, que compreende o estado do Rio de Janeiro, situado na região Sudeste do Brasil, é o território de menor extensão entre os estados da região Sudeste, sendo subdividido em 92 municípios e com 16.054.524 habitantes. Na Figura 1 é possível observar a divisão territorial do estado do Rio de Janeiro.

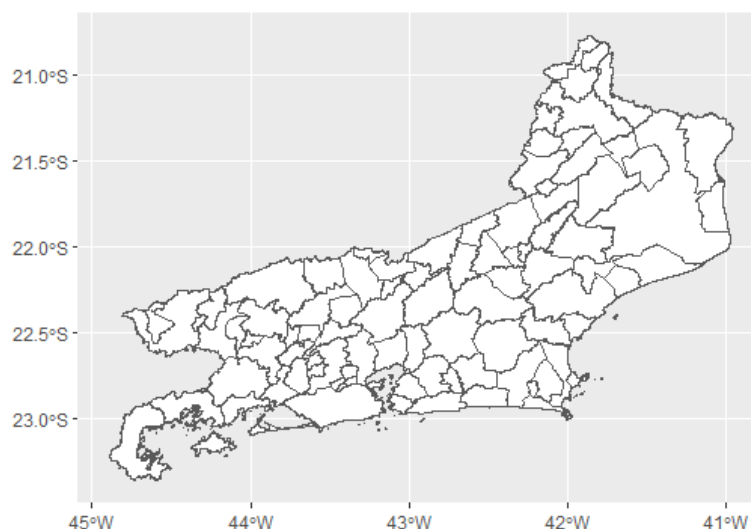


Figura 1: Mapa do estado do Rio de Janeiro por municípios.

As fontes de dados e informações utilizadas ao longo do trabalho estão disponíveis no portal DATASUS, do Ministério da Saúde. Foram utilizados dados das declarações de nascidos vivos através das bases do Sistema de Informações sobre Nascidos Vivos (SINASC) e o número de óbitos maternos declarados na base de Sistema de Informações de Mortalidade (SIM).

Inicialmente, foram consultadas duas bases de dados: a de nascidos vivos, obtida do SINASC, e a de mortes maternas, obtidas do SIM, filtrando os dados por município de residência do Estado do Rio de Janeiro. Dados com código de município<sup>1</sup> faltantes ou ignorados foram desconsiderados neste estudo. Foram considerados apenas os dados correspondentes aos anos de interesse, filtrando os dados com base nas variáveis DTOBITO, que indica a data do óbito no SIM, e DTNASC, que indica a data de nascimento no SINASC. Além disso, tendo em vista os conceitos e a construção do cálculo da RMM, foram considerados da base de mortalidade materna apenas os casos com código CID-10 XV, que indicam morte durante a gravidez, parto ou aborto e puerpério, e tipo de causa obstétrica “Morte materna obstétrica direta e indireta”.

Após esta filtragem, e com base na revisão de literatura feita acerca da Mortalidade Materna, foram selecionadas variáveis socioeconômicas das mães com potencial influência no risco de Mortalidade Materna; sendo o principal critério de inclusão para as variáveis neste estudo o fato da variável estar contida em ambos os sistemas (SINASC e SIM). Sendo assim, dentre as variáveis disponíveis, selecionou-se as descritas na Tabela 1.

Tabela 1: Variáveis disponíveis pelo DATASUS e selecionadas para o estudo.

Variável	Descrição	Valores
CODMUNRES	Código do município de residência	Numericos
IDADE	Idade da mãe ao falecer.	Numericos (em anos)
RACACOR	Cor informada pelo responsável pelas informações da falecida.	Branca Preta Amarela Parda Indígena
ESC	Escolaridade em anos.	Nenhuma de 1 a 3 anos de 4 a 7 anos de 8 a 11 anos 12 anos ou mais Ignorado
ESTCIV	Situação conjugal da falecida informada pelos familiares.	Solteira Casada Viúva Divorciada União estável Ignorado

<sup>1</sup>O Código do município de residência e de nascimento pelo IBGE e disponível em <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>.

Para fins de simplicidade na interpretação, a variável IDADE também foi categorizada, dando origem a variável FAIXAETARIA, com as seguintes categorias: “Até 18 anos”, “19 a 29 anos”, “30 a 39 anos” e “Acima de 40 anos”.

Com base nas variáveis descritas acima, duas novas variáveis também foram criadas: uma representando o total de óbitos maternos (denotada por NM) e outra o total de nascidos vivos (denotada por NNV). Estes totais foram obtidos contando o número de ocorrências (mortes ou nascimentos) por município, por ano, para cada diferente padrão socioeconômico definido pelo cruzamento dos valores das variáveis qualitativas apresentadas na Tabela 1, além da variável FAIXAETARIA. O banco de dados final utilizado neste trabalho foi definido pelo conjunto das 4 variáveis qualitativas apresentadas na Tabela 1, além da variável FAIXAETARIA e das duas novas variáveis criadas, daqui em diante identificadas por NM e NNV. Algumas linhas do banco de dados final podem ser vistas na Tabela 2. Na primeira linha desta Tabela há a seguinte informação: em 2019, houve 1.755 bebês nascidos vivos e 3 mortes maternas, no município do Rio de Janeiro, entre mulheres com idade de 30 a 39 anos, com raça/cor branca, com 8 a 11 anos de escolaridade e solteiras.

Tabela 2: Exemplo de algumas linhas do banco de dados utilizados, referentes ao ano de 2019.

CODMUNRES	FAIXAETARIA	RACACOR	ESC	ESTCIV	NM	NNV
330455	30 a 39 anos	Branca	8 a 11 anos	Solteira	3	1755
330100	19 a 29 anos	Preta	Nenhuma	Solteira	0	2
330300	19 a 29 anos	Parda	Nenhuma	Solteira	0	2
330010	19 a 29 anos	Indígena	Nenhuma	Solteira	0	2
330350	30 a 39 anos	Branca	8 a 11 anos	Divorciada	1	11
330250	Acima de 40 anos	Preta	8 a 11 anos	Casada	1	2

Para avaliar se a faixa de idade, a raça/cor, o estado civil e a escolaridade possuem relação com a mortalidade materna, modelos estatísticos podem ser ajustados. Neste trabalho, será utilizado modelos lineares generalizados e, portanto, na Seção há uma revisão sobre estes modelos.

## 2.2 Modelos Lineares

A compreensão dos determinantes por trás da mortalidade materna é essencial para desenvolver estratégias eficazes na área da saúde pública. Uma abordagem estatística fundamental para explorar essa relação é a utilização de modelos lineares. Na área da Estatística é comum descrever o comportamento de uma determinada variável resposta

(variável dependente) por uma ou mais variáveis explicativas (variável independente). Pode-se dizer que existe uma relação estatística entre duas variáveis  $X$  e  $Y$  se para cada valor de  $X$  existe uma distribuição de probabilidade para  $Y$ . Supondo que a variável resposta é contínua, é comum supor um modelo de regressão linear simples ou múltiplo, conforme descrito adiante.

**Modelo de regressão linear simples:** Analisa a relação entre uma variável explicativa e uma variável resposta, ambas contínuas. Este modelo é descrito da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (2.2)$$

onde

- $Y_i$  é o valor observado da variável resposta, referente ao  $i$ -ésimo elemento da amostra;
- $X_i$  é o valor observado da variável explicativa, referente ao  $i$ -ésimo elemento da amostra;
- $\beta_0$  e  $\beta_1$  são os parâmetros a serem estimados;
- $\varepsilon_i$ ,  $i = 1, \dots, n$ , são os erros aleatórios do modelo, possuindo as seguintes propriedades:  $E(\varepsilon_i) = 0$ ,  $Var(\varepsilon_i) = \sigma^2$  e  $Cov(\varepsilon_i, \varepsilon_j) = 0$ . É comum supor que  $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ .

**Modelo de regressão linear múltiplo:** É usado quando se quer determinar o valor da variável resposta com base nos valores de duas ou mais variáveis explicativas. Este modelo é descrito da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i. \quad (2.3)$$

Quando  $k = 1$ , tem-se o modelo de regressão linear simples como caso particular.

É comum escrever modelos de regressão linear em forma matricial:

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad (2.4)$$

onde

- $\mathbf{Y}$  é considerada a variável dependente, composta por  $n$  observações, representada pelo vetor de ordem  $n$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix};$$

- $\mathbf{X}$  é uma matriz de dimensão  $n \times (k+1)$ , composta por  $n$  observações das  $k$  variáveis explicativas

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix};$$

- $\boldsymbol{\beta}$  é um vetor de parâmetro desconhecidos de ordem  $k+1$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \text{ e}$$

- $\boldsymbol{\varepsilon}$  é um vetor aleatório com distribuição normal multivariada, cujas componentes são tais que, para  $i = 1, \dots, n$ ,  $E(\varepsilon_i) = 0$ ,  $Var(\varepsilon_i) = \sigma^2$ ,  $Cov(\varepsilon_i, \varepsilon_j) = 0$ , para  $j \neq i$ , e  $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ .

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Para mais informações, consulte Kubrusly (2014) e Dobson (2018).

### 2.2.1 Modelos Lineares Generalizados

Quando se está lidando com uma variável resposta que não segue uma distribuição normal, pode-se usar uma extensão do modelo de regressão linear que é denominado de Modelo Linear Generalizado (MLG). Nesse caso, também é possível modelar variáveis de interesse que assumam a forma de contagem, variáveis binárias ou categóricas, por exemplo. Alguns exemplos dos tipos de respostas e distribuições que podem ser assumidas em um MLG:

- Contagens: modelos de Poisson, Binomial ou Binomial Negativa;
- Binárias: modelo Bernoulli;
- Contínuas assimétricas: modelo Gama ou Normal Inversa;
- Categóricas: modelo Multinomial;

entre outros.

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes, quando condicionadas a um vetor paramétrico  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$ . Dizemos que a distribuição do vetor aleatório  $\mathbf{Y}$  pertencerá à família exponencial na forma canônica se puder ser escrita da forma

$$f(\mathbf{Y}) = \exp \left[ \sum_{i=1}^n Y_i b(\theta_i) + \sum_{j=1}^n c(\theta_j) + d(\mathbf{Y}) \right] \quad (2.5)$$

onde  $b(\cdot)$ ,  $c(\cdot)$  e  $d(\cdot)$  são funções reais.

Denote por  $E(Y_i) = \mu_i$  o valor esperado de  $Y_i$ . Em MLG, relaciona-se o valor esperado da variável de interesse  $Y_i$  com as covariáveis  $\mathbf{X}_i = (1 \ X_{i1} \ \dots \ X_{ik})^T$ , da seguinte forma

$$g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta},$$

onde  $g$  é função monótona e diferenciável, chamada função de ligação e  $\boldsymbol{\beta}$  é um vetor de parâmetros de ordem  $k + 1$ , sendo  $\beta_0$  chamado de intercepto.

A escolha da função de ligação é fundamental em um MLG, pois ela relaciona a média da distribuição à combinação linear dos preditores. Não existe uma única função de ligação possível. Baseado na família exponencial canônica, a função de ligação pode ser escolhida da seguinte forma

$$g(\mu_i) = b(\theta_i). \quad (2.6)$$

Para mais detalhes, veja Dobson (2018).

### 2.2.1.1 Modelos para dados de contagem

É comum que haja perguntas a responder na hora da construção de um MLG, sendo as 3 mais frequentes:

1. Qual a distribuição de probabilidade que modela o comportamento da variável resposta?

2. Quais são as variáveis explicativas?
3. Qual é a função do valor esperado que será modelada como função linear dessas variáveis?

Dobson (2018) define dados de contagem como “O número de vezes que um evento ocorre” e sua modelagem usualmente é feita através de distribuições de probabilidade como Binomial, Poisson ou Binomial Negativa.

Como neste trabalho a variável resposta de interesse é a quantidade de mortes maternas nos municípios do estado do Rio de Janeiro, contabilizadas a cada ano, para cada perfil de características socioeconômicas das mães, está-se lidando com dados de contagem e será admitido o uso da distribuição de probabilidade Poisson, na qual não há limite superior finito para os valores que podem ser observados na distribuição. Então, seja  $Y_i$  o número de ocorrências do evento de interesse em uma determinada unidade amostral  $i$ . Neste caso,  $Y_i \sim \text{Poisson}(\mu_i)$ ,  $i = 1, \dots, n$  e a função de probabilidade de  $Y_i$ , num dado ponto  $y_i$  é dada por

$$f(y_i | \mu_i) = \frac{\mu_i^{y_i} \exp\{-\mu_i\}}{y_i!}, \quad y_i = 0, 1, 2, \dots, \quad (2.7)$$

onde  $\mu_i$  é o número médio de ocorrências do evento de interesse para a unidade amostral  $i$ , sendo  $\mu_i > 0$ .

Note que a Equação dada em (2.7) pode ser escrita da seguinte forma

$$f(y_i | \mu_i) = \exp\{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\}, \quad (2.8)$$

e, portanto, esta distribuição pertence à família exponencial na forma canônica com  $b(\mu_i) = \ln(\mu_i)$ ,  $c(\mu_i) = -\mu_i$ ,  $d(y_i) = -\ln(y_i!)$ .

O valor esperado,  $E(Y_i) = \mu_i$ , pode ser relacionado às covariáveis usando a função de ligação natural, que no caso da Distribuição Poisson é a função logarítmica, dada por

$$\ln \mu_i = \mathbf{X}_i^T \boldsymbol{\beta} + O_i, \quad (2.9)$$

sendo  $O_i$  um termo chamado de *o set*, conhecido. Em modelos de contagem, é comum incluir o termo *o set* na função de ligação e serve para modificar a escala da variável resposta, tornando-a comparável em diferentes unidades amostrais. Para dar mais intuição sobre o uso de *o sets*, suponha a seguinte situação: supor 100 ocorrências de um certo evento em uma população de tamanho 100.000 não possui o mesmo impacto que as mesmas 100 ocorrências em uma população de tamanho 1.000; neste caso, o



tamanho de cada população  $i$  (conhecido e fixo) poderia ser considerado como um  $o$  set para corrigir a escala das respostas, tornando-as comparáveis. Neste trabalho, como as unidades amostrais levam em conta municípios com diferentes tamanhos de população, e, conseqüentemente, número de óbitos maternos relativamente desiguais ao longo do período analisado, optou-se por utilizar como  $o$  set o número de nascidos vivos de cada unidade amostral; ou seja, para cada resposta, associou-se o número de nascidos vivos no correspondente município, no referido ano, para cada perfil de características socioeconômicas das mães.

Dessa forma, tem-se que a dependência de  $\mu_i$  em relação às covariáveis é expressa por  $\mu_i = e^{\mathbf{X}_i^T \boldsymbol{\beta} + O_i}$ .

Por fim, um conceito fundamental para investigar a associação entre variáveis independentes e a taxa de ocorrência de eventos em um modelo de regressão Poisson é o chamado risco relativo (RR) ou razão de taxas. Em Dobson (2018), para um modelo de regressão Poisson, com função de ligação logarítmica, o risco relativo para variável indicadora de um certo fator  $X_j$ , que assume valores  $X_j = 0$  se o fator estiver ausente e  $X_j = 1$  se estiver presente, é dado por

$$\begin{aligned} RR &= \frac{E(Y_{ij}|X_{ij} = 1, X_{il} = x_{il}, l = 1, \dots, k, l \notin j)}{E(Y_{ij}|X_{ij} = 0, X_{il} = x_{il}, l = 1, \dots, k, l \notin j)} \\ &= \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \beta_{j+1} x_{i,j+1} + \dots + \beta_k x_{ik})}{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \beta_{j+1} x_{i,j+1} + \dots + \beta_k x_{ik})} \\ &= \exp(\beta_j), \end{aligned} \quad (2.10)$$

desde que todas as outras variáveis explicativas permaneçam as mesmas. Portanto, em geral, as estimativas dos parâmetros associados a variáveis binárias em modelos de regressão Poisson são frequentemente interpretadas na escala exponencial em termos das razões de taxas, conforme a seguir:

- se  $\exp(\beta_j)$  é igual a 1, não há efeito da presença da variável  $j$  na taxa de ocorrência do evento, quando comparado à ausência do fator;
- se  $\exp(\beta_j)$  é maior que 1, indica um aumento na taxa de ocorrência do evento na presença do fator  $j$  quando comparado à ausência deste fator; e
- se  $\exp(\beta_j)$  é menor que 1, há indícios de uma diminuição na taxa de ocorrência do evento na presença do fator  $j$ , quando comparado à ausência do fator.

De forma análoga, pode-se mostrar que para uma variável explicativa contínua  $X_j$ , um aumento de uma unidade no seu valor  $x_j$  resulta em um efeito multiplicativo de  $\exp(\beta_j)$  na taxa donde segue a interpretação das razões de taxas nestes casos.

Para mais detalhes, veja Dobson (2018).

## 2.3 Inferência Bayesiana

A compreensão dos fatores subjacentes à mortalidade materna requer uma análise abrangente e contextualizada. A Inferência Bayesiana surge como uma abordagem estatística poderosa para explorar esses complexos dados de saúde, diferindo da Inferência Clássica ao tratar o vetor de parâmetros populacionais, aqui denotado por  $\theta$ , como um vetor aleatório. Desta forma, a Inferência Bayesiana permite incorporar, na análise, evidências observadas em um experimento e crenças prévias de algum especialista.

O paradigma Bayesiano vê a incerteza como um grau de crença, calculada usando o conceito de probabilidade condicional. Vale ressaltar que a crença depende do conhecimento prévio sobre a situação em estudo, sendo representada por meio de distribuições de probabilidade.

- **Distribuição a priori:** É uma distribuição de probabilidade que expressa a crença inicial sobre os parâmetros de um modelo, antes da observação dos dados.
- **Distribuição a posteriori:** É uma distribuição de probabilidade dos parâmetros de um modelo, atualizada após a observação dos dados, por meio da aplicação do Teorema de Bayes. A distribuição a posteriori é calculada a partir da distribuição a priori e da função de verossimilhança dos dados.

Seja  $Y$  uma variável aleatória cuja distribuição é descrita por um vetor de parâmetros  $\theta$ . Para uma dada amostra observada  $\mathbf{y} = (y_1, \dots, y_n)$  de  $Y$ , a distribuição a posteriori de  $\theta$  pode ser obtida por

$$f(\theta | \mathbf{y}) = \frac{f(\theta; \mathbf{y})}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \theta) f(\theta)}{f(\mathbf{y})} \propto l(\theta; \mathbf{y}) f(\theta)$$

onde

- $f(\theta | \mathbf{y})$  é a distribuição a posteriori de  $\theta$ ;
- $l(\theta; \mathbf{y})$  é a função de verossimilhança de  $\theta$  considerando o vetor de observações  $\mathbf{y}$ ; e

- $f(\cdot)$  é a distribuição a priori de  $\theta$ .

A função de verossimilhança  $l(\theta; \mathbf{y})$  corresponde a função de (densidade de) probabilidade conjunta de  $\mathbf{Y}$ , condicionada em  $\theta$ , dada pela expressão  $f(\mathbf{y}|\theta)$ , após serem substituídos os valores da variável  $Y$  pelos observados na amostra.

### 2.3.1 Estimação Bayesiana

Segundo Ehlers (2003), a distribuição a posteriori de um vetor de parâmetros  $\theta$  contém toda a informação probabilística com relação a estes parâmetros. Além disso, a melhor descrição do processo de inferência é obtida através de um gráfico da sua função (de densidade) de probabilidade a posteriori. Porém, pode ocorrer de algumas vezes ser necessário resumir a informação contida na distribuição a posteriori através de estimadores pontuais e/ou intervalos de credibilidade.

#### 2.3.1.1 Estimação Pontual

A estimação pontual sob o ponto de vista Bayesiano é um processo de obtenção de resumos estatísticos sobre a distribuição a posteriori do(s) parâmetro(s) de interesse. Para a escolha da forma de fazer resumos, usa-se ferramentas da “Teoria da Decisão”. Com isso, é possível obter estimativas pontuais, a partir da distribuição a posteriori do parâmetro, tais como a média, a mediana ou a moda.

Ehlers (2003) define que, um problema de decisão fica especificado através de 3 tipos de espaços:

1. o espaço paramétrico ou estados da natureza;
2. o espaço amostral ou dos resultados possíveis; e
3. o espaço das decisões ou possíveis ações.

Uma Regra de Decisão  $\delta$  é uma função definida em um espaço amostral ( $\Omega$ ) que assume valores em possíveis ações ( $A$ ), isto é:

$$\delta : \Omega \rightarrow A$$

A cada decisão  $\delta$  e a cada possível valor do vetor de parâmetros  $\theta$ , pode-se associar uma perda  $L(\delta, \theta)$ . Ela pode ser usada para calcular a perda esperada associada a cada

decisão possível, levando em consideração a distribuição a posteriori dos parâmetros. A decisão final a ser tomada é, então, aquela que minimiza a perda esperada, levando em consideração as preferências pessoais.

O risco de uma regra de decisão é uma medida da perda esperada associada a esta regra. Denotado por  $R(\delta)$ , pode ser calculado como:

$$R(\delta) = E[L(\delta, \theta)] = \int L(\delta, \theta) f(\theta | \mathbf{y}) d\theta,$$

ou seja, é a média ponderada da perda esperada para todas as possíveis escolhas dos parâmetros, levando em consideração a distribuição a posteriori. Pode ser usado para comparar diferentes regras de decisão ou para avaliar a eficácia de uma regra de decisão em situações de incerteza.

Por fim, a regra de Bayes é a regra de decisão  $\delta$  ótima, ou seja, que possui risco mínimo frente às demais possíveis decisões, isto é,  $R(\delta) < R(\delta'), \forall \delta'$ .

### 2.3.1.2 Estimação Intervalar

A estimação pontual resulta em um número e uma dúvida natural é sobre a incerteza desta estimativa. Ao considerar a variabilidade dos estimadores pontuais é possível associar uma medida de incerteza, como, por exemplo, para a média a posteriori pode ser associado o desvio padrão; para a mediana, o intervalo interquartil, e para a moda, a curvatura neste ponto. Com a medida de incerteza é possível criar um estimador intervalar. Em Inferência Bayesiana, essa estimativa intervalar é chamada de intervalo de credibilidade (ou intervalo de confiança Bayesiano). Define-se  $C$  um intervalo de credibilidade de  $100(1-\alpha)\%$ , ou nível de credibilidade  $1 - \alpha$ , para  $\beta_j$  se

$$P(\beta_j \in C | \mathbf{y}) = 1 - \alpha \quad (2.11)$$

onde a definição é expressa de forma probabilística a pertinência ou não de  $\beta_j$  ao intervalo. Assim, quanto menor o tamanho do intervalo mais concentrada é a distribuição do parâmetro. Para mais detalhes, veja Ehlers (2003).

### 2.3.2 MLG Poisson

Para obter o estimador pontual e intervalar, é necessário obter inicialmente a distribuição posteriori. Sendo assim, para o MLG Poisson tem-se:

1. Função de probabilidade conjunta:

$$f(\mathbf{y}) = \prod_{i=1}^n \left[ \frac{\mu_i^{y_i} \exp\{-\mu_i\}}{y_i!} \right] \quad (2.12)$$

2. Função de Verossimilhança:

$$l(\boldsymbol{\mu}; \mathbf{y}) \propto \left[ \prod_{i=1}^n \mu_i^{y_i} \right] \exp \left\{ - \sum_{i=1}^n \mu_i \right\} \quad (2.13)$$

3. Conforme visto em Modelos Lineares Generalizado, para dados de contagem Poisson, temos que  $g(\mu_i) = \ln(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta} + O_i$ , logo,  $\mu_i = \exp\{\mathbf{X}_i^T \boldsymbol{\beta} + O_i\}$ . Substituindo:

$$l(\boldsymbol{\mu}; \mathbf{y}) \propto \exp \left\{ \sum_{i=1}^n (\mathbf{X}_i^T \boldsymbol{\beta} + O_i) y_i - \sum_{i=1}^n \exp\{\mathbf{X}_i^T \boldsymbol{\beta} + O_i\} \right\}. \quad (2.14)$$

4. Admitindo que a distribuição a priori conjunta foi construída a partir de distribuições univariadas normais independentes, denotadas por  $\beta_j \sim N(a_j, b_j)$ ,  $j = 1, \dots, k$ , a distribuição a posteriori conjunta para o vetor  $\boldsymbol{\beta}$  será dada por

$$f(\boldsymbol{\beta}; \mathbf{y}) = c \exp \left\{ \sum_{i=1}^n (\mathbf{X}_i^T \boldsymbol{\beta} + O_i) y_i - \sum_{i=1}^n \exp\{\mathbf{X}_i^T \boldsymbol{\beta} + O_i\} - \sum_{j=1}^k \frac{1}{2b_j} (\beta_j - a_j)^2 \right\}. \quad (2.15)$$

Por definição de função de probabilidade/densidade, tem-se que

$$\int_1 \dots \int_k f(\boldsymbol{\beta}; \mathbf{y}) d\beta_k \dots d\beta_1 = 1. \quad (2.16)$$

Logo, para encontrar a constante  $c$  é necessário resolver esta integral múltipla. Note que esta solução não é analiticamente conhecida. Neste trabalho, recorreu-se aos métodos de Monte Carlo via cadeias de Markov (MCMC), descritos abaixo.

### 2.3.3 Monte Carlo via Cadeias de Markov

O Monte Carlo via cadeias de Markov (MCMC) é uma técnica de simulação usada em áreas como Estatística, Engenharia, Ciência da Computação, dentre outras, com a ideia de obter amostras aleatórias de uma distribuição de probabilidade complexa. Sua diferença está no uso de técnicas de simulação iterativa com base em cadeias de Markov. Este método é amplamente utilizado em modelagem Bayesiana e a partir daqui será descrito considerando que o interesse está em obter uma amostra da distribuição a posteriori.

As amostras geradas pelo MCMC podem ser usadas para estimar os parâmetros, realizar testes de hipóteses, fazer previsões e outras tarefas de inferência estatística. Os métodos de MCMC mais utilizados em Inferência Bayesiana são: o amostrador de Gibbs, o algoritmo de Metropolis-Hastings e o Monte Carlo Hamiltoniano.

Em Inferência Bayesiana, os métodos de MCMC serão úteis quando a amostragem direta da distribuição a posteriori é difícil ou desconhecido, mas é possível avaliar a função (de densidade) de probabilidade até uma constante multiplicativa.

### 2.3.4 Amostrador de Gibbs

O Amostrador de Gibbs é um algoritmo usado para gerar amostras de uma distribuição multivariada. A ideia central é atualizar sequencialmente cada variável em relação às demais, condicionando na amostra obtida no passo anterior. O processo iterativo consiste em várias etapas, cada uma atualizando uma parte das variáveis, mantendo as outras fixas. Sendo assim, considere que o interesse está em obter uma amostra de um vetor  $\beta = (\beta_1, \dots, \beta_k)$ . Este vetor pode ser dividido em partes univariadas e/ou multivariadas. Por simplicidade de notação, suponha que este vetor foi dividido em  $k$  partes, sendo a parte  $j$  composta por  $\beta_j$ . As transições de um estado para outro são feitas usando distribuições chamadas de condicionais completas a posteriori (DCCP) e denotadas por  $f(\beta_j | \beta_{-j}, \mathbf{y})$ , onde  $\beta_{-j} = (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k)^T$ . Com isso, a distribuição condicional completa é a distribuição da  $j$ -ésima componente de  $\beta$  condicionada em todas as outras componentes. Ela é obtida da seguinte forma

$$f(\beta_j | \beta_{-j}, \mathbf{y}) = \frac{f(\beta, \mathbf{y})}{f(\beta_{-j}, \mathbf{y})} \propto f(\beta, \mathbf{y}) / f(\beta_{-j}). \quad (2.17)$$

Sendo assim, o amostrador de Gibbs é eficaz quando a amostragem direta da distribuição conjunta é difícil, mas a amostragem condicional é mais simples. O amostrador é definido pelos seguintes passos:

1. Iniciar com um contador de iterações da cadeia  $t = 0$  e especificar valores iniciais  $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_k^{(0)})^T$ .

2. Obter um novo valor  $\beta^{(t+1)}$  a partir de  $\beta^{(t)}$  através da geração sucessiva dos valores

$$\begin{aligned} \beta_1^{(t+1)} &= f(\beta_1 | \beta_2^{(t)}, \beta_3^{(t)}, \dots, \beta_k^{(t)}) \\ \beta_2^{(t+1)} &= f(\beta_2 | \beta_1^{(t+1)}, \beta_3^{(t)}, \dots, \beta_k^{(t)}) \\ &\vdots \end{aligned}$$

$$\beta_k^{(t+1)} = f(\beta_k, \beta_1^{(t+1)}, \beta_2^{(t+1)}, \dots, \beta_{k-1}^{(t+1)}).$$

3. Incrementar o contador de  $t$  para  $t + 1$  e retornar o passo 2 até obter convergência.

Sob o Amostrador de Gibbs, a cadeia sempre se move para um novo valor, pois não existe mecanismo de aceitação-rejeição. Além disso, após a convergência, os resultados formam uma amostra de  $f(\beta | \mathbf{y})$  e cada iteração se completa após  $k$  movimentos ao longo dos eixos das componentes de  $\beta$ . Espera-se que a convergência das cadeias de Markov seja alcançada após um número suficientemente grande de iterações e após um período inicial de aquecimento. O período de aquecimento refere-se às iterações iniciais necessárias até que a cadeia comece a convergir e costuma ser identificado visualmente ao analisar o traço da cadeia do parâmetro amostrado.

É importante observar que os parâmetros amostrados podem apresentar alta autocorrelação. Uma solução para lidar com esse problema é utilizar um espaçamento de ordem  $m$ , no qual uma amostra é selecionada a cada  $m$  iterações, corrigindo assim a autocorrelação da cadeia. O valor de  $m$  pode ser estimado por meio da análise do gráfico da função de autocorrelação do parâmetro amostrado.

No caso do MLG Poisson, as distribuições condicionais completas a posteriori de  $\beta_j$  possuem a seguinte forma

$$f(\beta_j | \beta_{-j}, \mathbf{y}) \propto \exp \left\{ \sum_{i=1}^n (\mathbf{X}_i^T \beta + O_i) y_i - \sum_{i=1}^n \exp(\mathbf{X}_i^T \beta + O_i) - \frac{1}{2b_j} (\beta_j - a_j)^2 \right\}. \quad (2.18)$$

Note que a DCCP de  $\beta_j$  não possui forma analítica conhecida. Para contornar este problema, é possível utilizar o amostrador de Gibbs com passos de Metropolis-Hastings.

### 2.3.5 Algoritmo de Metropolis-Hastings

Em Inferência Bayesiana, o Metropolis-Hastings é muito útil quando a distribuição a posteriori é univariada ou quando não é possível amostrar diretamente de pelo menos uma das DCCPs. Neste algoritmo, em cada iteração gera-se um valor, denotado por  $\beta^p$ , de uma distribuição auxiliar  $q(\beta^p | \beta^{(t)})$ , sendo  $\beta^{(t)}$  o valor obtido na iteração anterior. E aceita-se o ponto gerado com a seguinte probabilidade:

$$\alpha(\beta^{(t)}, \beta^p) = \min \left( 1, \frac{f(\beta^p | \mathbf{y}) q(\beta^{(t)} | \beta^p)}{f(\beta^{(t)} | \mathbf{y}) q(\beta^p | \beta^{(t)})} \right), \quad (2.19)$$

onde  $f$  é a distribuição de interesse. Percebe-se que só é preciso conhecer  $f$  parcialmente, a menos de uma constante. Além disso, a cadeia pode permanecer por muitas iterações no mesmo estado e na prática é usual monitorar calculando a porcentagem média de iterações para as quais novos valores são aceitos.

O Algoritmo de Metropolis-Hasting pode ser especificado através dos seguintes passos:

1. Inicie com um contador de iterações  $t = 0$  e especifique um valor inicial  $\theta^{(0)}$ .
2. Gere um novo valor  $\theta^*$  da distribuição  $q(j | \theta^{(t)})$ .
3. Calcule a probabilidade de aceitação  $\alpha(\theta^{(t)}, \theta^*)$  e gere  $u \sim U(0, 1)$ .
4. Se  $u \leq \alpha$  então aceite o novo valor e faça  $\theta^{(t+1)} = \theta^*$ , caso contrário, rejeite e faça  $\theta^{(t+1)} = \theta^{(t)}$ .
5. Incremente o contador de  $t$  para  $t + 1$  e volte ao passo 2 até que se obtenha convergência.

Se a variância da distribuição proposta for atribuído um valor pequeno, a convergência da Cadeia de Markov será lenta devido aos movimentos pequenos da cadeia. Por outro lado, se a variância for definida como um valor muito alto, a cadeia tende a ter pouca movimentação, pois a taxa de rejeição dos valores propostos será alta.

O algoritmo de Metropolis-Hastings segue os mesmos critérios de convergência mencionados na Subseção 2.3.4, sendo necessário um período de aquecimento e, possivelmente, o uso de espaçamento de ordem  $m$  a fim de evitar amostras altamente autocorrelacionadas. Mais detalhes sobre convergência serão discutidos na Subseção 2.3.7.

Escolher a distribuição auxiliar não é uma tarefa fácil pois dependendo da escolha a cadeia pode ter dificuldade em se mover. Quando os possíveis valores da variável a ser amostrada pertence ao conjunto dos números reais, é comum usar uma distribuição normal centrada em  $\theta^{(t)}$  e a variância da distribuição é escolhida de forma que a taxa de aceitação da cadeia esteja em torno de 44%.

### 2.3.6 Monte Carlo Hamiltoniano

O Monte Carlo Hamiltoniano (MCH) foi proposto para melhorar a eficiência do amostrador de Metropolis, especialmente em problemas de alta dimensionalidade. Proposto por Duane et al. (1987), este método fornece um poderoso algoritmo de



amostragem em MCMC que consiste em uma simulação dinâmica hamiltoniana paralela à regra de transição de Metropolis-Hastings. Este método simula valores aleatórios de uma distribuição alvo, percorrendo o suporte desta sem que haja necessidade de fazer uso de distribuições auxiliares.

O MCH aprimora a eficiência do Método de Metropolis-Hastings (MH) utilizando um esquema de geração de propostas direcionado. Especificamente, o MCH aproveita o gradiente do logaritmo da distribuição a posteriori para orientar a cadeia de Markov em direção a regiões de maior densidade a posteriori, onde a maioria das amostras é coletada. Como resultado, uma cadeia MCH bem calibrada aceitará propostas com uma taxa consideravelmente maior do que o algoritmo MH tradicional.

É um ótimo método para lidar com modelos estatísticos complexos, de alta dimensionalidade e em situações em que a amostragem de distribuições a posteriori é desafiadora. Ele pode ser implementado em várias bibliotecas e linguagens de programação, sendo a biblioteca Stan a mais conhecida. A implementação ou conceito da biblioteca Stan em softwares como o R, pode ser encontrada com mais detalhes em Duane et al. (1987). Neste trabalho utilizou-se o pacote *rstanarm*, que trata-se de um apêndice do pacote *rstan* que permite muitos dos modelos de regressão aplicados sejam estimados usando Monte Carlo via Cadeias de Markov. Uma dessas funções usada neste estudo é o *stan\_glm*, onde possui várias distribuições a priori possíveis para os coeficientes e, se aplicável, uma distribuição a priori para qualquer parâmetro auxiliar em um MLG.

### 2.3.7 Teste de Convergência

A análise de convergência em métodos de MCMC é crucial para garantir que as amostras geradas se aproximam da distribuição alvo e, para isso, é necessário que haja estacionariedade e ergodicidade. A estacionariedade implica na distribuição de probabilidade das amostras não mudar ao longo do tempo. E a ergodicidade significa que é possível alcançar qualquer estado a partir de qualquer outro em um número finito de passos.

Os métodos de MCMC consideram que após um certo período a cadeia converge para a distribuição alvo. Sendo assim, o período inicial que a cadeia leva para chegar na região de estacionariedade é chamado de *burn-in* ou de aquecimento da cadeia e é descartado da amostra final. Após eliminar o período de aquecimento, avalia-se a autocorrelação. Caso essa autocorrelação esteja alta, é necessário realizar um espaçamento, ou seja, a amostra final será composta pelos valores gerados a cada  $m$  iteração. Por exemplo, suponha que

gerou-se uma amostra de tamanho 1.050 e que descartou-se as 50 primeiras iterações. Ao notar a presença de uma autocorrelação alta, pode-se utilizar um espaçamento de 10 iterações e então a amostra final é composta pelo valor gerado na iteração 51, depois o valor da iteração 61, depois o da 71, e, assim por diante.

Existem várias ferramentas e diagnósticos que podem ser utilizados para avaliar essa convergência. A análise gráfica é a mais usada e fundamental. Essa análise é realizada através do traço da cadeia, da autocorrelação e do histograma de cada parâmetro amostrado. Recomenda-se gerar mais de uma amostra, sendo que cada uma deve inicializar de um ponto completamente diferente. Dessa forma, através do traço da cadeia pode-se verificar se as amostras convergiram para a mesma região. Além disso, tem-se que a suposição de convergência é atingida quando, após descartadas as iterações iniciais, percebe-se um comportamento aleatório por todo o espaço paramétrico no traço da cadeia. Pelo gráfico da autocorrelação, é possível definir o espaçamento necessário para a amostra final. E pelo histograma é possível verificar também a estacionariedade da cadeia.

A convergência do MCMC também pode ser verificada numericamente. Proposto por Gelman e Rubin (1992), o método indica se a convergência foi alcançada através da utilização de múltiplas cadeias. É um método formal que compara a variabilidade entre as cadeias de Markov simuladas com a variabilidade dentro de cada cadeia. Se as cadeias convergiram para a mesma distribuição, a variabilidade intra-cadeia deve ser semelhante à variabilidade inter-cadeia. O procedimento envolve os seguintes passos:

1. Executar várias cadeias de MCMC independentes.
2. Acompanhar a evolução de cada cadeia ao longo do tempo e certificar de que estão explorando a distribuição a posteriori de maneira semelhante.
3. Calcular a variância dentro de cada cadeia, fornecendo uma medida da variabilidade associada a cada cadeia individualmente.
4. Calcular a variância entre as médias das cadeias, fornecendo uma medida da variabilidade entre as cadeias.
5. Divida a variância entre cadeias pela variância dentro de cada cadeia e ajuste para o tamanho da amostra. O resultado obtido é conhecido como o fator de escala de Gelman-Rubin *R-hat*.

O critério de convergência de Gelman-Rubin sugere que, se as cadeias convergirem, o fator de escala  $R\text{-hat}$  será próximo de 1. Valores significativamente maiores que 1 indicam falta de convergência. Através da saída da função do *stan\_glm* que se encontra no pacote do *rstanarm* será fornecido o valor do  $R\text{-hat}$ . Para maiores detalhes sobre o assunto, veja Nogueira (2004) e Gelman e Rubin (1992).

## 2.4 Estatística Espacial

A Estatística Espacial avalia fenômenos ocorridos em um determinado espaço geográfico. É uma ferramenta que possui aplicações em diversas áreas como Agronomia, Epidemiologia, Climatologia, Demografia, entre outras. Um exemplo na área da saúde, que é o enfoque deste estudo, está na análise da distribuição espacial da ocorrência de uma determinada doença, onde deseja-se verificar a apresentação de algum padrão no espaço ou se existe associação com outra variável.

Cressie (1993) resume que a Estatística, desde a análise exploratória de dados até a teoria da distribuição assintótica de estimadores de parâmetros, depende de um modelo estocástico. Esse modelo, em dados espaciais, podem ser contínuos ou discretos; agregações espaciais ou observações em pontos no espaço; localizações espaciais regulares ou irregulares, e essas localizações podem ser de um conjunto espacial contínuo ou de um conjunto discreto. O modelo estocástico é usado para resumir dados existentes ou prever dados não observados.

Um conjunto de dados espaciais pode ser classificado em três tipos básicos: padrões pontuais (processos pontuais), superfícies contínuas (pertinentes à geoestatística) e dados de área.

- **Padrões Pontuais:** São fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. O principal interesse está em saber se os eventos observados exibem completa aleatoriedade ou algum padrão sistemático em determinada região. São exemplos: localização de crimes, ocorrências de doenças e localização de espécies vegetais.
- **Superfícies Contínuas:** Estimadas a partir de um conjunto de amostras de campo, que podem estar regularmente ou irregularmente distribuídas. Usualmente, este tipo de dado é resultante de levantamento de recursos naturais e que incluem mapas geológicos, topográficos, ecológicos, fitogeográficos e pedológicos (estudos

dos solos). Resumindo, são fenômenos de estudo variando de forma contínua em uma determinada região em que a sua localização é conhecida. Como exemplos de aplicação, pode-se citar: índices pluviométricos, níveis de pressão atmosférica, salinidade do solo, entre outros.

- **Dados de área:** São fenômenos de interesse ocorridos em uma determinada área (como bairros, municípios, estados, países, etc.). Neste caso é feito o agrupamento dos fenômenos ocorridos em diferentes pontos dessa área.

Para melhor compreensão dos dados de área temos como exemplo este trabalho, onde serão estudados os casos de mortalidade materna no Estado do Rio de Janeiro. Como não se tem a localização exata do evento em questão, mas sim o município de registro de cada óbito, será analisado o número de ocorrências/óbitos, de forma agregada, em cada município do Rio de Janeiro.

### 2.4.1 Dados de Área

Em Banerjee, Carlin e Gelfand (2004), tem-se a seguinte definição para dados de área: *“São regiões geográficas ou blocos que são denotados por  $B_i$ , e os dados são somas ou médias de variáveis sobre esses blocos”*. Resumindo, são eventos estudados em que não se conhece sua localização exata de ocorrência, mas, ao invés disso, se conhece um valor associado a uma área.

É comum surgirem dúvidas quanto à diferença entre os dados de área e os de padrões pontuais. Enquanto o primeiro inclui dados onde não se distingue a localização exata do evento estudado, ou seja, lida com uma quantidade que representa um padrão global com relação à variável de estudo, os dados do tipo padrão de pontos são compostos basicamente pelas coordenadas do evento. Um exemplo pode ser dado em relação aos casos de dengue no Estado do Rio de Janeiro. Enquanto os dados de área têm enfoque em analisar o somatório/percentual do número de ocorrências, os de padrões pontuais possuem interesse em prever a localização de casos naquela região.

Em Análise Exploratória, uma ferramenta muito usada na apresentação dos dados de área é o mapa coroplético. Através de cenários nos quais a variável de interesse foi observada de forma agregada, estes mapas diferenciam as regiões por cores, de acordo com a distribuição espacial do evento observado. Com isso, os mapas coropléticos permitem a visualização do comportamento espacial, identificando altas ou baixas incidências e regiões com agrupamento espacial (clusters).

### 2.4.1.1 Abordagens exploratórias para dados de área

Diversas ferramentas podem ser úteis para análise exploratória dos dados de área, principalmente para analisar casos em que se deseja verificar se um local está sendo influenciado não somente por outro local adjacente, mas talvez por todos os outros que o cercam. Com isso, o primeiro foco está na avaliação da matriz de proximidade (também conhecida como matriz de vizinhança), que é utilizada para estimar a variabilidade espacial dos dados, ou seja, descreve a relação espacial entre as unidades geográficas de interesse.

Em Banerjee, Carlin e Gelfand (2004) tem-se que a matriz de proximidade espacial é utilizada em cálculos de indicadores de análises exploratórias como, por exemplo, o índice de Moran. Geralmente é representada por uma matriz quadrada, na qual cada célula indica se duas unidades geográficas são vizinhas ou não. Para sua construção é necessário definir critérios de vizinhança para o qual é possível usar a distância entre os respectivos centros geográficos (centroides).

Seja um conjunto de  $n$  áreas definido por  $(A_1, A_2, \dots, A_n)$ . Uma matriz de vizinhança  $\mathbf{W}$ , de dimensão  $n \times n$ , é construída de forma que para cada elemento  $w_{ij}$  de  $\mathbf{W}$ ,  $i = 1, \dots, n$ ;  $j = 1, \dots, n$ , indica a medida de proximidade espacial entre as áreas  $A_i$  e  $A_j$ , onde  $w_{ii} = 0$ . Esta medida de proximidade pode ser definida a partir de um dos seguintes critérios:

1.  $w_{ij} = 1$  se o centroide de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário  $w_{ij} = 0$ . Em alguns casos, essa distância pode ser fixada como uma distância máxima. Por outro lado, em certos contextos, a distância pode não ser fixada como uma distância máxima. Em vez disso, pode depender de alguma medida de similaridade, densidade ou alguma lógica adaptativa baseada nos dados ou no problema em questão.
2.  $w_{ij} = 1$  se  $A_i$  compartilha a fronteira comum com  $A_j$ ; caso contrário  $w_{ij} = 0$ . Nesse caso, a contiguidade pode ser estabelecida de três maneiras: Bispo, Torre ou Rainha. Sendo apresentada na Figura 2 o tipo Rainha.
3.  $w_{ij} = \frac{l_{ij}}{l_i}$  onde  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$ , com  $l_i$  sendo o perímetro de  $A_i$ .

Segundo Camara et al. (2004), é comum padronizar as linhas da matriz de vizinhança  $W$ , efetuando a divisão de cada elemento  $w_{ij}$  pela soma da linha em que este elemento pertence, fazendo com que a soma de cada linha seja igual a 1.

A matriz de proximidade espacial pode ser generalizada para vizinhos de ordem superior a 1, ou seja, vizinhos dos vizinhos. Usando critério análogo ao adotado para a matriz de vizinhança de primeira ordem, pode-se construir as matrizes  $W^{(2)}; \dots; W^{(n)}$ . Para mais detalhes, consulte Camara et al. (2004).

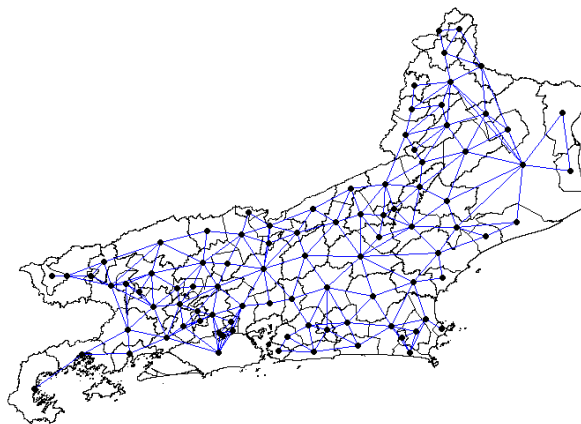


Figura 2: Exemplo de Critério de vizinhança por contiguidade Queen nos municípios dos Estado do Rio de Janeiro

A contiguidade Rainha entre as áreas é estabelecida de maneira que duas áreas são consideradas vizinhas se compartilham qualquer tipo de limite, não apenas um limite comum no sentido direto (horizontal ou vertical) como no caso do Torre. Em outras palavras, nesta abordagem, duas áreas são consideradas vizinhas se compartilham um limite direto ou diagonal, analogia com o movimento da rainha no jogo de xadrez, que pode se mover em todas as direções (horizontal, vertical e diagonal) pelo tabuleiro.

#### 2.4.1.2 Índice de Moran Global

Uma característica fundamental presente na análise exploratória espacial é a particularização da dependência espacial, indicando como os valores estão correlacionados no espaço. Duas estatísticas que medem a força da associação espacial entre unidades de área são os Índice de Moran e o Coeficiente de Geary. Entretanto, no presente estudo, será dada ênfase para o primeiro.

O Índice de Moran é uma medida estatística usada para avaliar a autocorrelação espacial, ou seja, a relação entre valores de uma variável em diferentes locais ou unidades geográficas. Assim como o coeficiente de correlação linear de Pearson mede a relação entre duas variáveis em um contexto não espacial, o Índice de Moran mede a relação entre os valores de uma variável em diferentes locais ou regiões adjacentes. Se o Índice de Moran for positivo e significativo, isso sugere que valores semelhantes estão próximos uns dos outros no mapa, indicando autocorrelação espacial positiva. Se for negativo e significativo, isso indica uma dispersão espacial dos valores, onde áreas com valores altos estão próximas de áreas com valores baixos. E caso o valor seja nulo, é indicativo de que os valores da variável estão distribuídos aleatoriamente no espaço, não mostrando nenhum padrão de similaridade espacial entre as unidades geográficas analisadas.

O cálculo do Índice de Moran é dado pela seguinte forma:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (2.20)$$

onde

- $n$  é o número total de observações;
- $w_{ij}$  é o elemento da matriz de vizinhança para  $i, j \in \{1, \dots, n\}$ ; e
- $Y_i$ ,  $Y_j$  e  $\bar{Y}$  indicam, respectivamente, o valor da variável na localização  $i$  e  $j$ , e a média da variável em toda a região.

É comum testar se o índice de Moran indica independência espacial ( $I = 0$ ) ou não ( $I \neq 0$ ), ou seja,  $H_0 : I = 0$  versus  $H_1 : I \neq 0$ . Para isso, calcula-se o índice de Moran observado. Depois realiza-se uma simulação de Monte Carlo ou uma permutação espacial para criar uma distribuição do índice de Moran sob a hipótese nula. Isso envolve a reordenação aleatória dos valores da variável entre as unidades espaciais para gerar uma nova configuração espacial. Calcula-se então o índice para cada permutação e compara-se o observado com a distribuição simulada. O p-valor do teste é a proporção de índices simulados que são iguais ou mais extremos do que o observado. Ele representa a probabilidade de observar um valor tão extremo quanto o observado, sob a hipótese nula de aleatoriedade espacial. Compara-se o p-valor com um limiar de significância pré-determinado (por exemplo, 0,05). Se for menor que o limiar, rejeita-se a hipótese nula em favor da hipótese alternativa de autocorrelação espacial significativa.

## 3 Análise dos Resultados

Este Capítulo contém os resultados das análises descritivas e inferenciais realizadas para os casos de mortalidade materna no estado do Rio de Janeiro durante o período de 2010 a 2019. Inicialmente, realizou-se uma análise exploratória dos dados de mortalidade, com o objetivo de compreender melhor o comportamento da variável resposta e das covariáveis. Em seguida, para cada ano, ajustou-se um modelo de regressão Poisson considerando como variável resposta o número de casos de mortalidade materna e como unidade amostral o perfil sociodemográfico combinado com o município. Posteriormente, fez-se uma análise descritiva espacial para o número de óbitos maternos encontrados nos municípios do RJ em cada ano. Nessa última análise, a unidade amostral era apenas os municípios.

A filtragem dos dados, conforme descrita na Seção 2.1, além das análises descritivas, espaciais e a modelagem dos dados foram realizadas usando o R Core Team (2023) e o Posit (2009).

### 3.1 Análise Descritiva

A análise descritiva dos dados foi feita a fim de compreender melhor a base do trabalho e sua natureza. A Tabela 3 apresenta o número de óbitos maternos, de nascidos vivos e a razão de mortalidade compreendidos no período de 2010 a 2019. Desde a realização do último Censo, no ano de 2010, é possível verificar pequenas oscilações na quantidade de casos ocorridos no estado do Rio de Janeiro. Os números de óbitos maternos variaram ao longo dos anos, com oscilações entre 107 e 143 casos por ano, sendo no ano de 2018 o menor número registrado conforme a Tabela 3. O número de nascidos vivos também variou, mas não parece ter uma tendência consistente em crescimento ou queda significativa ao longo dos anos. A RMM oscilou entre 50,20 e 66,54 ao longo do período de 10 anos.

Para melhor visualização da Razão de Mortalidade Materna anual, foi realizada uma análise de evolução temporal apresentada conformem a Figura 3. Percebe-se que no ano



Tabela 3: Distribuição dos Óbitos Maternos, Nascidos Vivos e da Razão de Mortalidade Materna.

Ano	Óbitos Maternos	Nascidos Vivos	RMM
2010	132	204.617	65,97
2011	114	213.875	56,67
2012	130	215.719	62,38
2013	127	215.726	59,82
2014	137	225.774	60,37
2015	127	226.665	59,09
2016	138	212.562	64,35
2017	143	214.905	65,86
2018	107	213.153	50,80
2019	122	199.993	61,07
Total	1.277	2.142.986	

de 2018 a razão de mortalidade materna registrou o seu menor valor, o que poderia indicar uma possível melhoria na qualidade dos cuidados a saúde materna, mas, no ano seguinte a razão já volta a crescer. Esses resultados indicam uma situação preocupante de mortalidade materna, apesar da ausência de uma tendência clara de aumento ou redução ao longo do tempo.

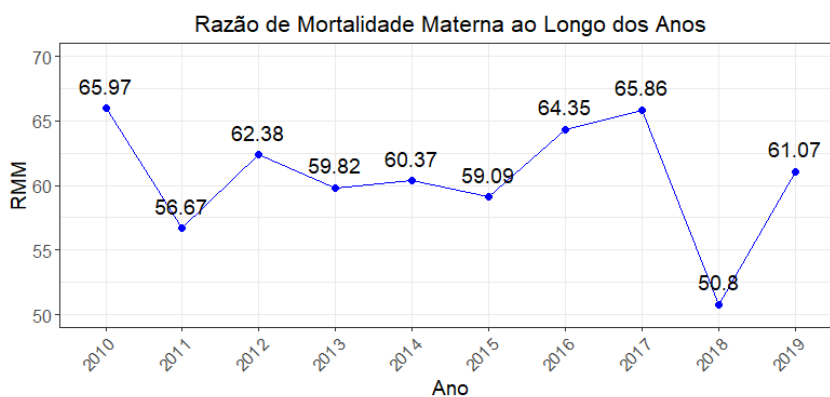


Figura 3: Razão de Mortalidade Materna no Estado do Rio de Janeiro - Série Anual

Entender as causas que influenciam no aumento e diminuição da razão de mortalidade é um trabalho minucioso que exige cautela do pesquisador na hora de interpretar. Conhecimentos prévios sobre a saúde primária da mulher auxiliam no estudo de identificação de possíveis variáveis que possam afetar esse indicador de saúde. Tendo em vista as variáveis disponíveis no banco de dados do Sistema de Informação sobre Mortalidade (SIM), a Tabela 4 reúne dados de óbitos maternos e nascidos vivos, além da razão entre esses números, segmentados por faixa etária, raça/cor, nível de escolaridade e estado civil. Essas variáveis sociodemográficas desempenham papéis significativos na

determinação dos riscos e desafios enfrentados pelas mulheres durante o período gravídico-puerperal.

Os resultados na Tabela 4 indicam que a razão de óbitos em relação aos nascidos vivos foi mais elevada na faixa etária acima de 40 anos (162,28), mulheres pretas (132,02), escolaridade de 1 a 3 anos completos (344,49) e viúvas (133,90). Porém, cabe ressaltar que as maiores frequências absolutas estão em mulheres na faixa etária de 19 a 29 anos, pardas, escolaridade de 8 a 11 anos completos e solteiras.

Percebe-se que houve casos de dados faltantes na variável escolaridade, dentre os 1.277 casos foram retirados 54 por faltas de informação. Isso ocorre pela dificuldade da equipe de saúde obter todas as informações da paciente. Mesmo uma região com excelente cobertura do registo, ainda ocorrem declarações de óbitos maternos que não são totalmente exatos.

Tabela 4: Distribuição dos Óbitos Maternos, Nascidos Vivos e a Razão por variáveis, no período de 2010 a 2019.

Variáveis	Óbitos Maternos	Nascidos Vivos	Razão
<b>Faixa Etária</b>			
Até 18 anos	141	262.161	53,78
19 a 29 anos	543	1.140.975	47,59
30 a 39 anos	490	676.383	72,44
Acima de 40 anos	103	63.470	162,28
<b>Raça/Cor</b>			
Amarela	2	5.440	36,76
Branca	412	820.427	50,22
Parda	566	1.090.839	51,89
Preta	297	224.972	132,02
<b>Escolaridade</b>			
Nenhuma	12	4.338	276,63
1 a 3 anos	167	48.477	344,49
4 a 7 anos	447	435.638	102,61
8 a 11 anos	465	1.228.238	37,86
12 e mais	132	426.298	30,96
<b>Estado Civil</b>			
Casada	324	672.608	48,17
Divorciada	27	28.964	93,22
Solteira	863	1.360.997	63,41
União estável	58	76.686	75,63
Viúva	5	3.734	133,90

Ao analisar os resultados para cada variável através da Tabela 4, é perceptível a identificação dos reais grupos que são fatores para o aumento da mortalidade materna. A

maioria dos casos estão na faixa etária a partir de 40 anos (162,28 por 100 mil nascidos vivos), pretas (132,02 por 100 mil nascidos vivos), com escolaridade até três anos completo (369,25 por 100 mil nascidos vivos) e sem parceiros (64,22 por 100 mil nascidos vivos).

Através dos resultados apresentados, buscando uma melhor precisão das análises, as variáveis que possuem alta razão de mortalidade materna foram transformadas em dicotômicas, visando separar os grupos que mais influenciam no aumento da razão da mortalidade materna. As variáveis transformadas, assim como a Razão de Mortalidade Materna associada a cada grupo (calculadas de acordo com a Equação dada em (2.1)), podem ser vistas na Tabela 5.

Tabela 5: Distribuição dos Óbitos Maternos, Nascidos Vivos e a Razão por variáveis dicotômicas, no período de 2010 a 2019.

Variáveis	Óbitos Maternos	Nascidos Vivos	Razão
<b>Faixa Etária</b>			
Abaixo de 40 anos	1.174	2.079.519	56,46
A partir de 40 anos	103	63.470	162,28
<b>Raça/Cor</b>			
Não preta	980	1.918.017	51,09
Preta	297	224.972	132,02
<b>Escolaridade</b>			
Até 3 anos	179	48.477	369,25
A partir de 4 anos	1.044	2.094.512	49,84
<b>Estado Civil</b>			
Com parceiro	382	749.294	50,98
Sem parceiro	895	1.393.695	64,22

Visando uma melhor compreensão visual da razão entre os grupos de cada variável, foram feitos os gráficos de barra que estão apresentados na Figura 4. Através deles, é possível compreender a dimensão de cada grupo de risco, dado por exemplo, a covariável escolaridade até 3 anos, que ultrapassa mais de 360 óbitos por 100 mil vivos.

Essa análise inicial sugere que a idade, raça/cor, escolaridade e estado civil podem estar associadas com a mortalidade materna no estado do Rio de Janeiro, sendo que em alguns grupos demográficos as taxas são consideravelmente mais altas. Isso pode ser crucial para direcionar políticas de saúde específicas para reduzir a mortalidade materna, considerando as disparidades identificadas entre esses grupos.

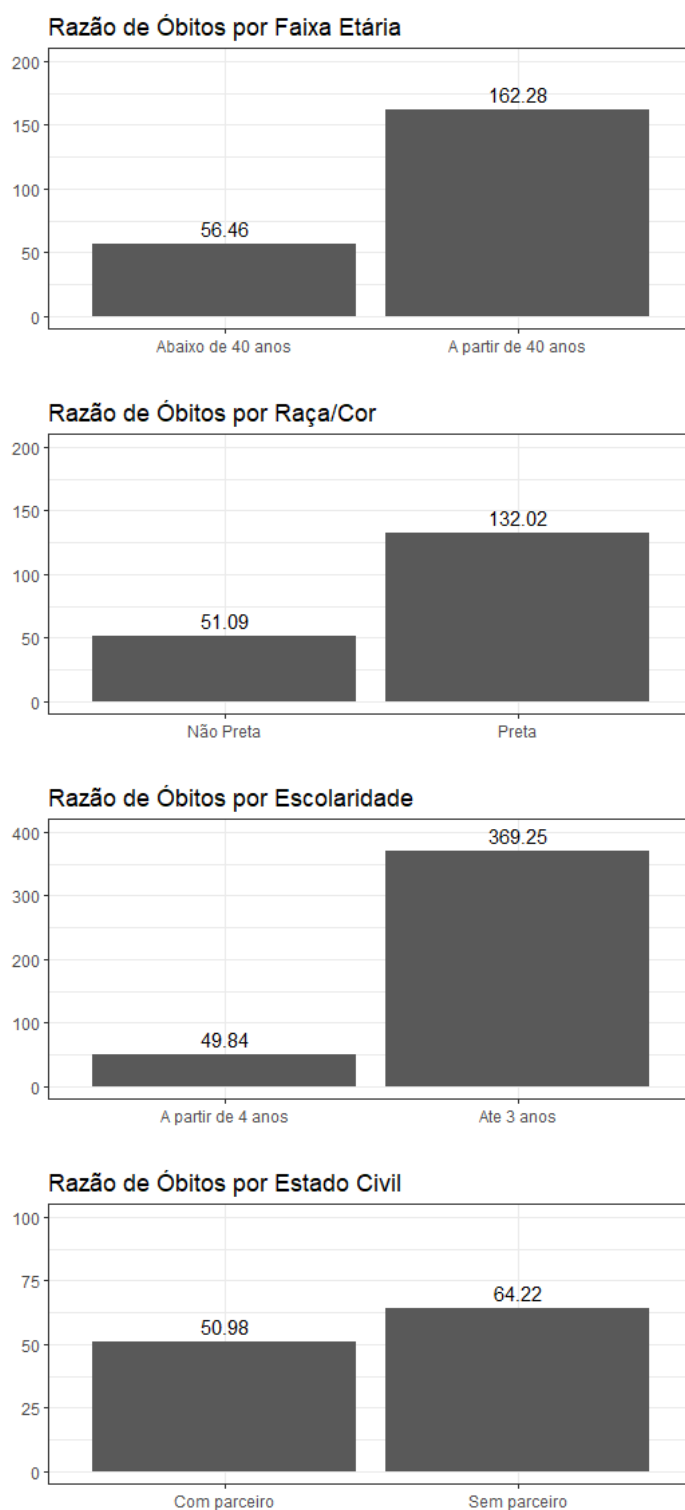


Figura 4: Razão de Mortalidade Materna no Estado do Rio de Janeiro - Variáveis: Faixa etária, Raça/Cor, Escolaridade e Estado Civil

## 3.2 Ajuste do Modelo

Visando relacionar e quantificar o efeito de variáveis como escolaridade e faixa etária com a mortalidade materna, ajustou-se um modelo linear generalizado sob a perspectiva Bayesiana. Considerou-se que, a cada ano de interesse,

$$\begin{aligned} Y_i & \sim \text{Pois}(\mu_i) \\ \ln(\mu_i) & = \mathbf{X}_i + O_i \end{aligned} \quad (3.1)$$

sendo  $Y_i$  o número de óbitos maternos da  $i$ -ésima combinação formada pelos municípios do estado do Rio de Janeiro e por cada configuração de covariáveis dos grupos de risco (variável NM, da Tabela 2) e  $O_i$  (termo conhecido como *o set*), como sendo o número de nascidos vivos da  $i$ -ésima unidade amostral (variável NNV, da Tabela 2), permitindo assim uma comparação adequada das taxas de óbitos maternos entre os diferentes municípios.

Para a matriz de covariáveis ( $\mathbf{X}_i$ ), assumiu-se uma coluna com todos os elementos iguais a 1, caracterizando o uso do intercepto; e, nas demais colunas, variáveis *dummies* indicando as variáveis Faixa Etária, Raça/Cor, Escolaridade e Estado Civil (conforme dicotomizações indicadas na Tabela 5). A escolha pelas variáveis dicotômicas foi feita para reduzir a complexidade dos dados, identificando padrões mais claros e permitindo a construção dos modelos mais precisos e eficazes na predição e identificação de fatores associados aos óbitos maternos.

Para a distribuição a priori conjunta para o vetor  $\beta$ , assumiu-se independência a priori para os efeitos  $\beta_j, j = 1, \dots, k$  (conforme descrito na Seção 2.3.3), com distribuição normal com média  $a_j = 0$  e variância  $b_j = 100^2$ . Ao definir a média como zero, não se está atribuindo nenhuma preferência prévia para valores positivos ou negativos dos parâmetros. Isso indica uma ausência de viés inicial em direção a qualquer valor específico. Quanto ao desvio padrão grande, significa que, a priori, permite-se uma ampla variabilidade de valores plausíveis para os parâmetros do modelo, caracterizando o uso de distribuições a priori não informativas, dando mais peso aos dados observados durante o processo de inferência. Conforme mostrado na Equação (2.15), a distribuição a posteriori de  $\beta$  não possui forma analítica conhecida e métodos de MCMC foram usados para amostrar desta distribuição. O método escolhido neste trabalho foi o Monte Carlo Hamiltoniano, utilizando-se a função *stan\_glm* do pacote *rstanarm* (Goodrich et al. (2023)). Utilizando a função gerou-se 2 cadeias para cada um dos parâmetros a serem estimados, sendo cada uma delas com 6.000 iterações.

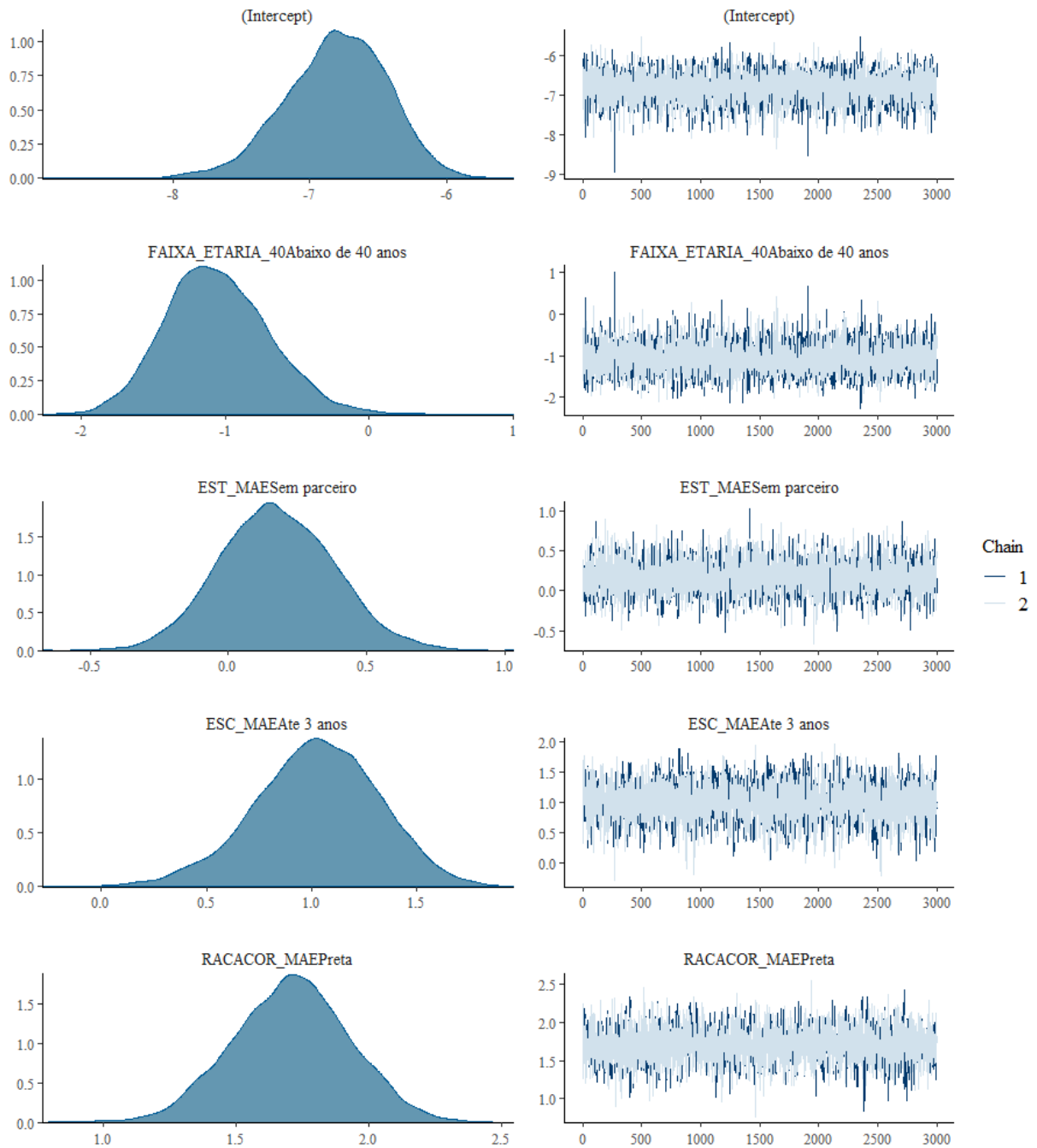


Figura 5: Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2010.

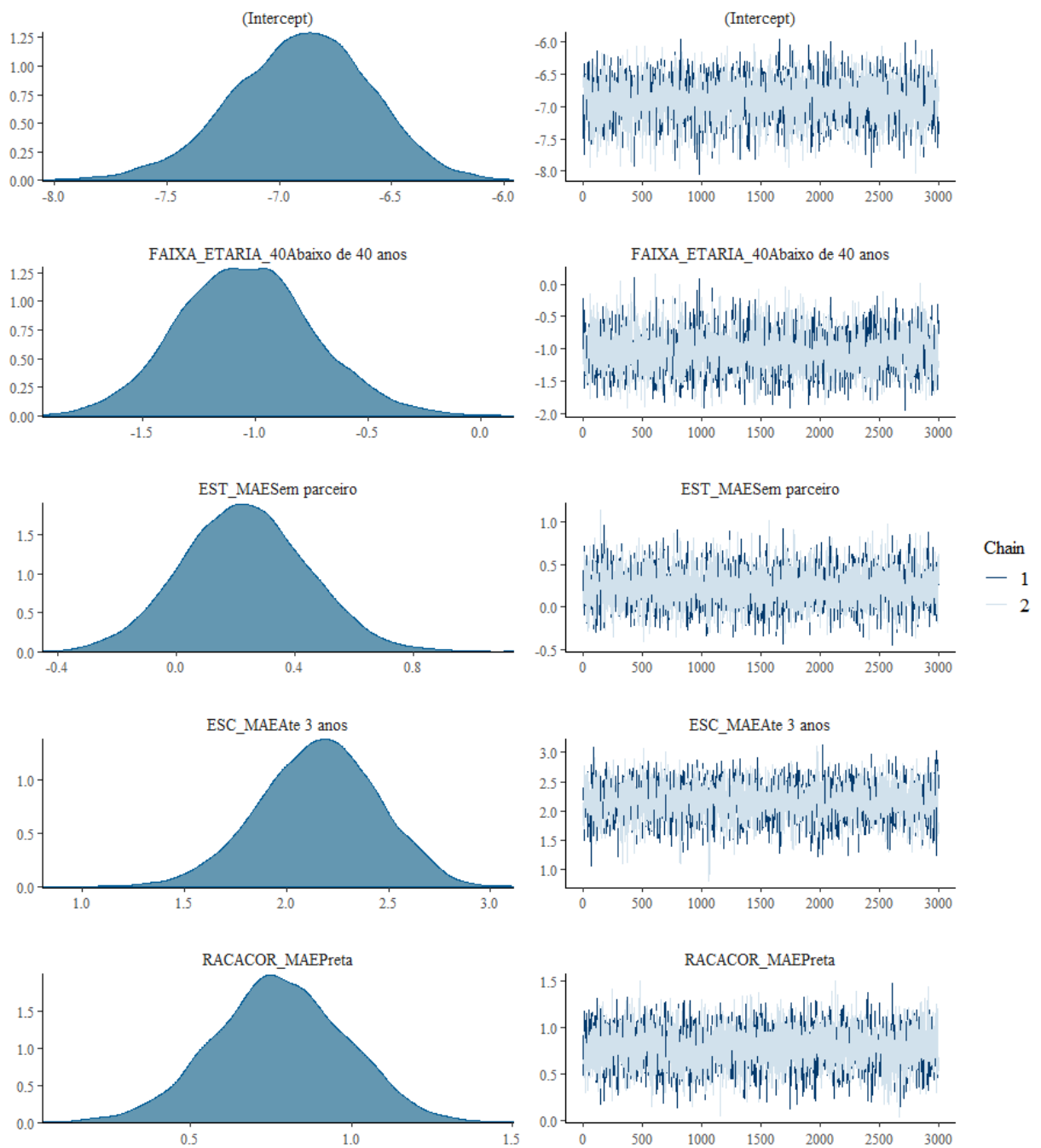


Figura 6: Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2019.

Para avaliar a convergência para a distribuição de interesse, foi verificado de forma visual por meio das observações dos traços das cadeias dos parâmetros. Foram usadas as Figuras 5 e 6 para interpretação dos resultados dos anos de 2010 e 2019, escolhidos para representar as análises realizadas. Observando as densidades aproximadas, percebe-se que, são aproximadamente simétrica para as covariáveis, além disso, a posição central da distribuição indica o valor provável do parâmetro a posteriori. É possível notar a semelhança e estabilidade nas trajetórias das cadeias dos parâmetros, dando indícios de convergência para a distribuição a posteriori. A convergência indica que as Cadeias de Markov atingiram a estacionariedade, permitindo a obtenção de estimativas confiáveis e precisas para os parâmetros do modelo. O comportamento das cadeias para os parâmetros dos demais anos podem ser consultados no Apêndice A junto com as densidades aproximadas.

Nas Figuras 7 e 8 é possível observar os gráficos de autocorrelação das cadeias geradas para os parâmetros para os anos de 2010 e 2019. Em ambos os casos, percebe-se uma rápida diminuição na autocorrelação, indicando correlações não significativas ao longo das iterações, não havendo necessidade de tomar espaçamentos (*lags*) ao longo das cadeias. O Apêndice B apresenta os resultados para anos de 2011 a 2018, com comportamentos semelhantes aos discutidos para os anos de 2010 e 2019.

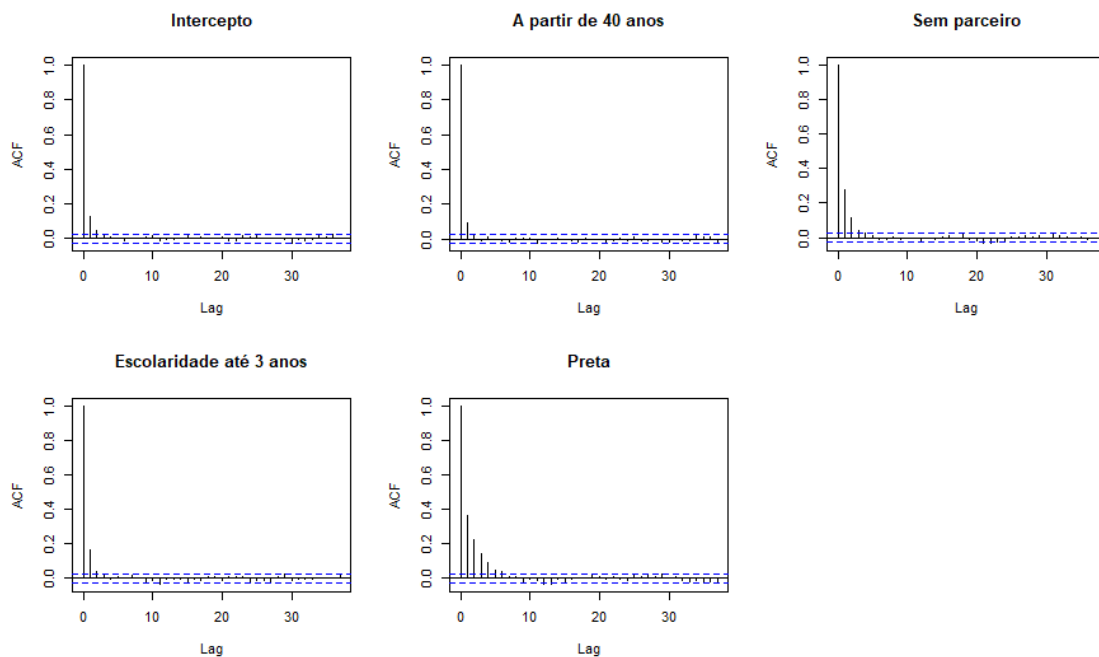


Figura 7: Função de Autocorrelação dos valores gerados para os coeficientes - 2010.



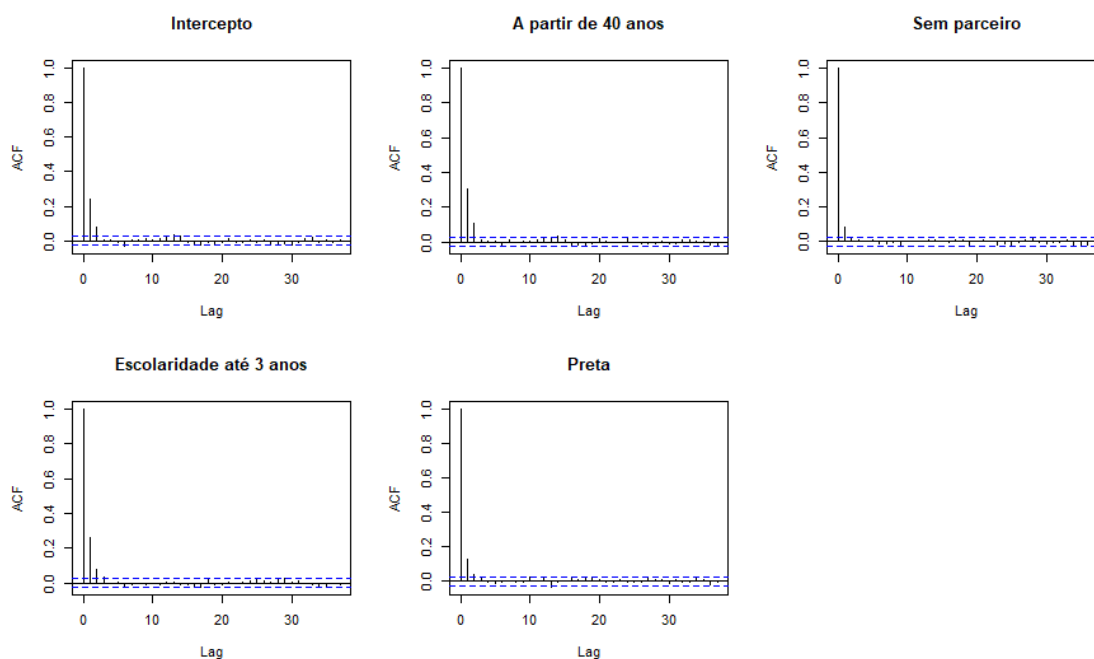


Figura 8: Função de Autocorrelação dos valores gerados para os coeficientes - 2019.

Para uma avaliação numérica da convergência, analisou-se a estatística  $R$ -hat de Gelman-Rubin, obtida por meio da função *stan\_glm*. Os valores do  $R$ -hat encontrados para todas as covariáveis e anos foram iguais a 1, sugerindo boa convergência e consistência entre as múltiplas cadeias de Markov utilizadas no modelo.

Nas Tabelas 6 a 15 encontram-se as estimativas pontuais (médias a posteriori) e os intervalos de credibilidade de 95% para os parâmetros dos modelos ajustados para os anos de 2010 a 2019, respectivamente; além das estimativas pontuais para a razão de taxas (calculadas conforme descrito na Seção 2.2.1.1). Interpretando as Tabelas, pode-se observar que, no ano de 2010, a estimativa para o coeficiente associado à variável indicadora de mulheres abaixo de 40 anos foi de -1,1, com razão de taxas de 0,341; o que indica que mulheres com menos de 40 anos em comparação com mulheres acima de 40 anos, tem uma redução de aproximadamente 66% na taxa de óbitos maternos. Além disso, essa interpretação é semelhante para todos os demais anos pois a razão de taxas permaneceu abaixo de 1,0. Ainda sobre o ano de 2010, percebemos que mulheres sem parceiro têm taxas de mortalidade cerca de 18,6% maiores em comparação com aquelas que possuem um parceiro, mantendo as outras covariáveis constantes. Com exceção dos anos de 2016 e 2018, onde houveram uma redução de taxa de mortalidade para mulheres sem parceiros quando comparadas com as que possuem, os demais anos tiveram aumento dessa taxa.

Na Tabela 14, a covariável escolaridade até 3 anos teve seu coeficiente aumentado para 2,3, indicando uma razão de taxa de 9,972. Isso sugere que, em 2018, mulheres com escolaridade até 3 anos tiveram uma taxa de óbitos maternos aproximadamente 9 vezes maior do que aquelas com maior escolaridade. E para mulheres pretas, o coeficiente diminuiu para 0,9, indicando uma razão de taxa de 2,587; ou seja, mulheres pretas tiveram uma taxa de óbitos maternos aproximadamente 59% maior do que mulheres não pretas. Para os demais anos, a interpretação das covariáveis é semelhante pois os valores são próximos, com exceção do ano de 2010, onde as mulheres pretas tiveram a taxa de mortalidade 5 vezes maior do que aquelas que não são.

Tabela 6: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2010

Covariáveis	<b>Coeficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,8	0,001	[-7,60 ; -6,12]
Abaixo de 40 anos	-1,1	0,341	[-1,71 ; -0,30]
Sem Parceiro	0,2	1,186	[-0,23 ; 0,61]
Escolaridade até 3 anos	1,0	2,795	[0,40 ; 1,57]
Preta	1,7	5,569	[1,27 ; 2,12]

Tabela 7: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2011

Covariáveis	<b>Coeficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-7,4	0,001	[-8,26 ; -6,68]
Abaixo de 40 anos	-0,8	0,428	[-1,52 ; -0,03]
Sem Parceiro	0,4	1,492	[-0,02 ; 0,87]
Escolaridade até 3 anos	1,8	6,060	[1,31 ; 2,24]
Preta	1,1	2,983	[0,63 ; 1,53]

Tabela 8: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2012

Covariáveis	<b>Coeficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,9	0,001	[-7,61 ; -6,23]
Abaixo de 40 anos	-0,9	0,396	[-1,55 ; -0,20]
Sem Parceiro	0,1	1,116	[-0,27 ; 0,50]
Escolaridade até 3 anos	1,7	5,252	[1,15 ; 2,12]
Preta	0,4	1,558	[-0,02 ; 0,87]

Tabela 9: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2013

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-7,6	0,001	[-8,60 ; -6,77]
Abaixo de 40 anos	-0,3	0,726	[-1,09 ; 0,69]
Sem Parceiro	0,3	1,342	[-0,09 ; 0,70]
Escolaridade até 3 anos	1,7	5,301	[1,12 ; 2,14]
Preta	0,6	1,827	[0,14 ; 1,00]

Tabela 10: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2014

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-7,5	0,001	[-8,38 ; -6,70]
Abaixo de 40 anos	-0,4	0,650	[-1,19 ; 0,47]
Sem Parceiro	0,2	1,262	[-0,15 ; 0,63]
Escolaridade até 3 anos	2,0	7,387	[1,47 ; 2,46]
Preta	0,6	1,818	[0,15 ; 1,03]

Tabela 11: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2015

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,9	0,001	[-7,56 ; -6,25]
Abaixo de 40 anos	-1,0	0,355	[-1,64 ; -0,32]
Sem Parceiro	0,1	1,075	[-0,32 ; 0,48]
Escolaridade até 3 anos	1,6	5,058	[1,00 ; 2,18]
Preta	0,9	2,512	[0,49 ; 1,32]

Tabela 12: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2016

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,5	0,002	[-7,30 ; -5,91]
Abaixo de 40 anos	-0,9	0,385	[-1,51 ; -0,17]
Sem Parceiro	-0,4	0,642	[-0,80 ; -0,05]
Escolaridade até 3 anos	2,3	9,782	[1,74 ; 2,76]
Preta	0,5	1,673	[0 ; 0,93]

Tabela 13: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95% Ano: 2017

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-7,6	0,001	[-8,48 ; -6,90]
Abaixo de 40 anos	-0,4	0,694	[-1,05 ; 0,48]
Sem Parceiro	0,5	1,669	[0,12 ; 0,92]
Escolaridade até 3 anos	2,0	7,270	[1,42 ; 2,46]
Preta	0,8	2,337	[0,46 ; 1,20]

Tabela 14: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95% Ano: 2018

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,8	0,001	[-7,51 ; -6,24]
Abaixo de 40 anos	-1,1	0,331	[-1,66 ; -0,39]
Sem Parceiro	-0,2	0,842	[-0,59 ; 0,25]
Escolaridade até 3 anos	2,3	9,972	[1,70 ; 2,84]
Preta	0,9	2,587	[0,48 ; 1,37]

Tabela 15: Estimativas pontuais para os parâmetros e para as Razões de taxas e Intervalo de Credibilidade de 95%. Ano: 2019

Covariáveis	<b>Coefficiente</b>	<b>Razão de Taxas</b>	Intervalo de Credibilidade
	Est. Pontual	Est. Pontual	
Intercepto	-6,9	0,001	[-7,43 ; 2,01]
Abaixo de 40 anos	-1,0	0,350	[-2,40 ; -0,52]
Sem Parceiro	0,2	1,247	[-0,55 ; 0,58]
Escolaridade até 3 anos	2,2	8,766	[1,67 ; 3,22]
Preta	0,8	2,179	[0,43 ; 1,52]

### 3.3 Análise Espacial

Visando identificar algum padrão espacial nos dados de mortalidade materna, segue-se para a Análise Espacial dos dados, conforme proposto na Seção 2.4.1.1.

Nesta Seção a unidade amostral dos anos foi alterada, considerando agora que a unidade amostral é apenas os municípios do estado do RJ.

Foi calculada a matriz de proximidade espacial seguindo o critério de contiguidade. Como já dito anteriormente, este critério define se duas unidades espaciais são consideradas vizinhas por compartilharem uma fronteira ou limite comum. Além disso, é usado também em análises de autocorrelação espacial, como o Índice de Moran.

Os valores apresentados na Tabela 16 são resultados da análise de autocorrelação espacial dos dados dos anos de 2010 a 2019, utilizando o Índice de Moran. Para todos os anos listados, os valores do Índice de Moran estão próximos de zero, variando entre -0,021 e 0,116. Isso sugere uma ausência de forte padrão de autocorrelação espacial nos dados ao longo desses anos. Quanto aos p-valores, a maioria está acima de 0,05, indicando que os resultados não são estatisticamente significativos para afirmar a presença de autocorrelação espacial dado um nível de significância de 5%.

Em resumo, os resultados da análise de autocorrelação espacial sugerem que não há um padrão espacial significativo nos dados de óbitos maternos ao longo dos anos de 2010 a 2019, indicando uma distribuição mais aleatória desses eventos ao longo das unidades geográficas consideradas.

Tabela 16: Autocorrelação espacial dos dados de 2010 a 2019 no estado do Rio de Janeiro.

Anos	Índice de Moran	P-Valor
2010	-0,021	0,867
2011	-0,030	0,768
2012	-0,013	0,958
2013	0,018	0,654
2014	-0,046	0,535
2015	0,116	0,054
2016	-0,013	0,968
2017	0,044	0,372
2018	0,084	0,064
2019	-0,048	0,552

Nas Figuras de 9 a 18, são dispostos os mapas de 2010 a 2019 a nível estadual do Rio de Janeiro, com dados da razão de mortalidade materna. Foi usado o mapa coroplético por ser uma técnica cartográfica muito útil para a representação de informações quantitativas

sobre variáveis geográficas e para a identificação de padrões espaciais em diferentes regiões geográficas. Além disso, ele é frequentemente utilizado em análises espaciais e em estudos que envolvem a compreensão da distribuição geográfica de diferentes fenômenos conforme foi descrito na Seção 2.4.

2010

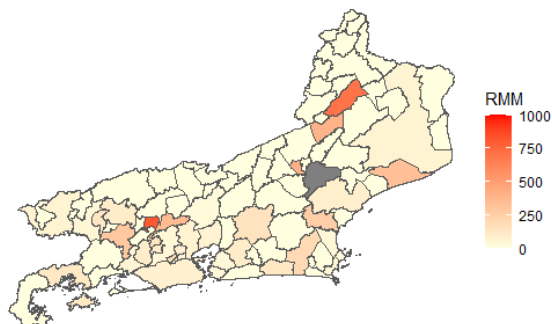


Figura 9: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2010.

2011

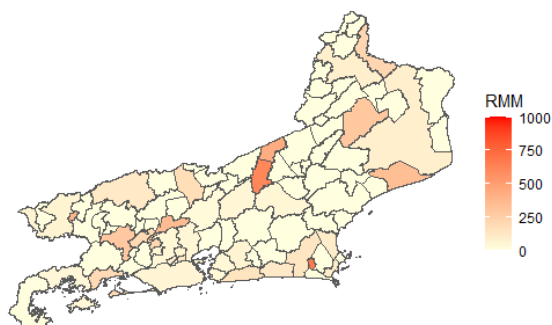


Figura 10: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2011.

2012

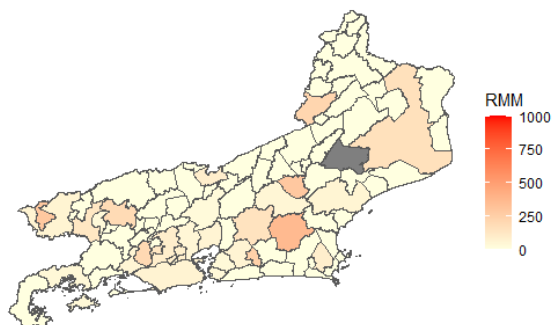


Figura 11: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2012.

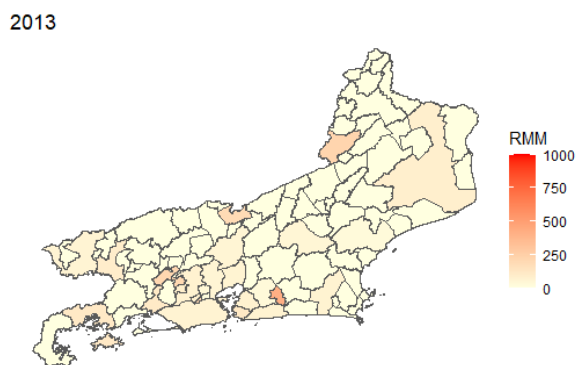


Figura 12: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2013.

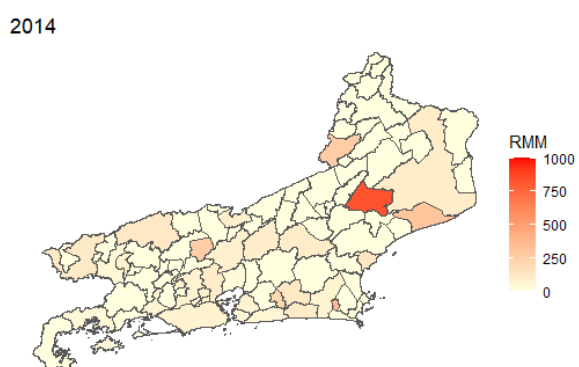


Figura 13: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2014.

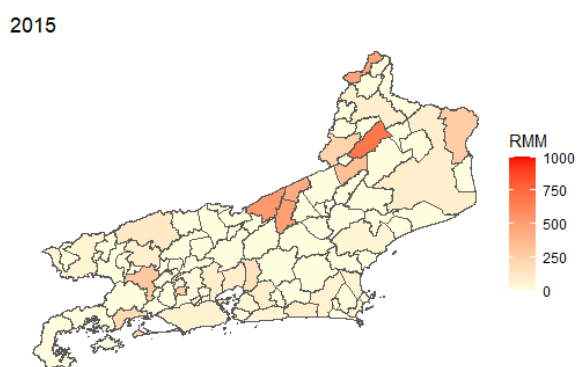


Figura 14: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2015.

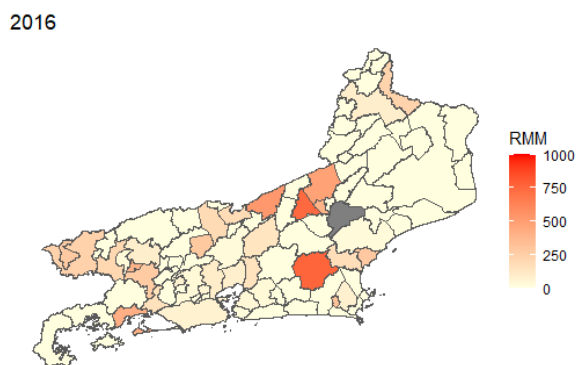


Figura 15: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2016.

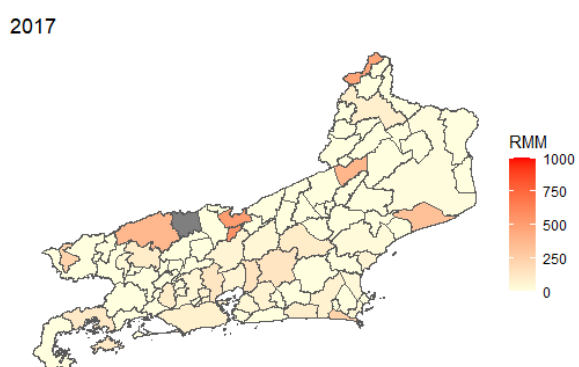


Figura 16: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2017.

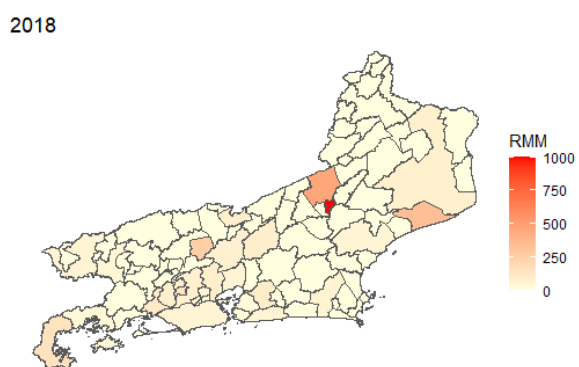


Figura 17: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2018.



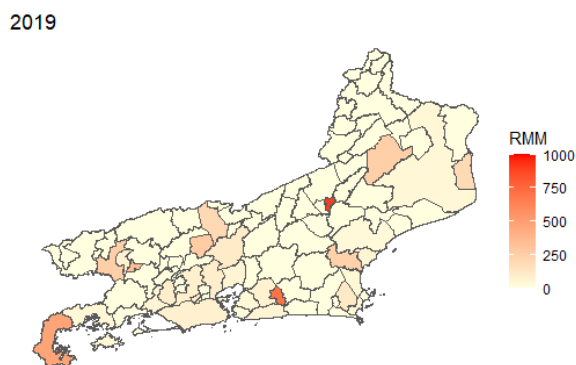


Figura 18: Mapa Coroplético da RMM do Estado do Rio de Janeiro de 2019.

Usando o mesmo limite de classes para cada ano, percebe-se que a razão de mortalidade esteve em alta em municípios localizados nas regiões Centro Fluminense, Noroeste Fluminense e Norte Fluminense, para os demais anos, houve dispersão na razão. Além disso, alguns municípios possuem a coloração cinza na sua área em determinados anos, podendo indicar não haver casos de óbitos maternos e conseqüentemente não foi possível calcular sua razão.

Embora esta análise tenha revelado insights valiosos sobre a distribuição espacial da razão de mortalidade materna no Estado do Rio de Janeiro ao longo dos anos, a modelagem preditiva não foi explorada devido a limitações de tempo para desenvolvimento deste trabalho. Esta lacuna representa uma oportunidade significativa para pesquisas futuras. A aplicação de modelos estatísticos avançados ou técnicas pode oferecer uma compreensão mais profunda dos fatores subjacentes e das tendências futuras associadas à mortalidade materna.

## 4 Conclusões

O estudo realizado buscou investigar os determinantes da mortalidade materna ao longo de um período de dez anos, empregando uma abordagem estatística Bayesiana. As análises foram fundamentadas através dos dados de nascidos vivos e óbitos maternos, utilizando o modelo Poisson para compreender a influência de variáveis como faixa etária, estado civil, escolaridade e raça/cor nas taxas de mortalidade. A investigação dos resultados obtidos revelou padrões consistentes ao longo dos anos, notando-se que mulheres abaixo de 40 anos apresentaram, em média, uma redução significativa na mortalidade materna em comparação com aquelas acima dessa faixa etária. No entanto, a ausência de parceiro, demonstrou consistentemente uma associação positiva com taxas mais elevadas.

A análise da escolaridade evidenciou uma correlação inversa entre os níveis educacionais e as de mortalidade materna, indicando que mulheres com menor nível de instrução têm uma tendência maior de apresentar taxas mais altas. Além disso, a variável étnica revelou uma associação positiva entre mulheres não brancas e maiores taxas de mortalidade materna em comparação com mulheres brancas, aspecto que demanda uma análise mais aprofundada das disparidades socioeconômicas e de acesso aos serviços de saúde entre diferentes grupos.

Estes resultados sugerem a necessidade de políticas públicas e intervenções direcionadas para grupos específicos, visando a redução das taxas de mortalidade materna. Estratégias que promovam a educação, o acesso equitativo à saúde e o apoio social para mulheres em situação de vulnerabilidade podem desempenhar um papel crucial na redução desses desafios. Ademais, ressalta-se a importância da continuidade deste estudo, incorporando análises mais detalhadas e considerando outras variáveis socioeconômicas e de saúde que possam impactar as taxas de mortalidade materna, visando um entendimento mais abrangente e efetivo para o desenvolvimento de políticas mais assertivas e inclusivas. Sendo interessante para trabalhos futuros, explorar a dinâmica espacial da mortalidade materna, utilizando técnicas avançadas de modelagem

espacial e incorporando variáveis contextuais para uma compreensão mais abrangente e preditiva dos padrões e determinantes da mortalidade materna em diferentes regiões.

## Referências

- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. *Hierarchical Modeling and Analysis for Spatial Data*. [S.l.]: Chapman & Hall/CRC, 2004.
- CAMARA, G. et al. *Análise Espacial de Dados Geográficos*. 2004. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/i>.
- CONCEIÇÃO, G. M. de S.; SALDIVA, P. H. N.; SINGER, J. da M. *Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo*. 2001. Disponível em: <https://www.scielo.org/pdf/rbepid/2001.v4n3/206-219/pt.i>.
- CRESSIE, N. A. *Statistics For Spatial Data*. Wiley series in probability and mathematical statistics. [S.l.]: John Wiley, 1993.
- DOBSON, A. J. *An Introduction to Generalized Linear Models*. 4. ed. [S.l.]: Chapman and Hall, 2018.
- DUANE, S. et al. *Hybrid Monte Carlo*. [S.l.]: Physics letters, 1987.
- EHLERS, R. S. *Introdução à Inferência Bayesiana*. 2003. Disponível em: <http://www.leg.ufpr.br/paulojus/CE227/ce227.pdf>.
- GELMAN, A.; RUBIN, D. B. *Inference from Iterative Simulation Using Multiple Sequences*. Vol 7, no 4. [S.l.]: Statistical Science, Institute of Mathematical Statistics, 1992.
- GOODRICH, B. et al. *rstanarm: Bayesian applied regression modeling via Stan*. 2023. R package version 2.26.1. Disponível em: <https://mc-stan.org/rstanarm/i>.
- KUBRUSLY, J. *Apostila de Modelos Lineares*. 2014. Disponível em: [https://www.professores.uff.br/jessica/wp-content/uploads/sites/137/2017/09/notas\\_de\\_aula\\_modelos\\_lineares.pdf](https://www.professores.uff.br/jessica/wp-content/uploads/sites/137/2017/09/notas_de_aula_modelos_lineares.pdf).
- MEDRONHO, R. A. et al. *Epidemiologia*. 2. ed. [S.l.]: Atheneu, 2011.
- MOTTA, C. T.; MOREIRA, M. R. *Uma análise sobre a mortalidade materna, de 1996 a 2018*. 2021. Disponível em: <https://www.scielo.br/j/csc/a/4pPdjk3DDSH6B8c5X3TNsKy/?format=pdf&lang=pt.i>.
- NOGUEIRA, D. A. *Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov: casos uni e multivariados*. Universidade federal de lavras. [S.l.: s.n.], 2004.
- POSIT. *The Open-Source Data Science Company*. 2009. Disponível em: <https://posit.co/download/rstudio-desktop/i>.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2023. Disponível em: <https://www.R-project.org/>.

RUAS, C. A. M. et al. *Per 1 e distribuicao espacial da mortalidade materna*. 2020. Disponível em: <https://www.scielo.br/j/rbsmi/a/sMMfc47dLMPRZgBDcjy5LP/?format=pdf&lang=pt>.

SAÚDE, M. D. *UMA ANALISE DA SITUAÇÃO DE SAÚDE E DA AGENDA NACIONAL E INTERNACIONAL DE PRIORIDADES EM SAÚDE*. 2009. Disponível em: [https://bvsmms.saude.gov.br/bvs/publicacoes/saude\\_brasil\\_2009.pdf](https://bvsmms.saude.gov.br/bvs/publicacoes/saude_brasil_2009.pdf).

# APÊNDICE A – Cadeias via Monte Carlo Hamiltoniano dos anos de 2011 a 2018

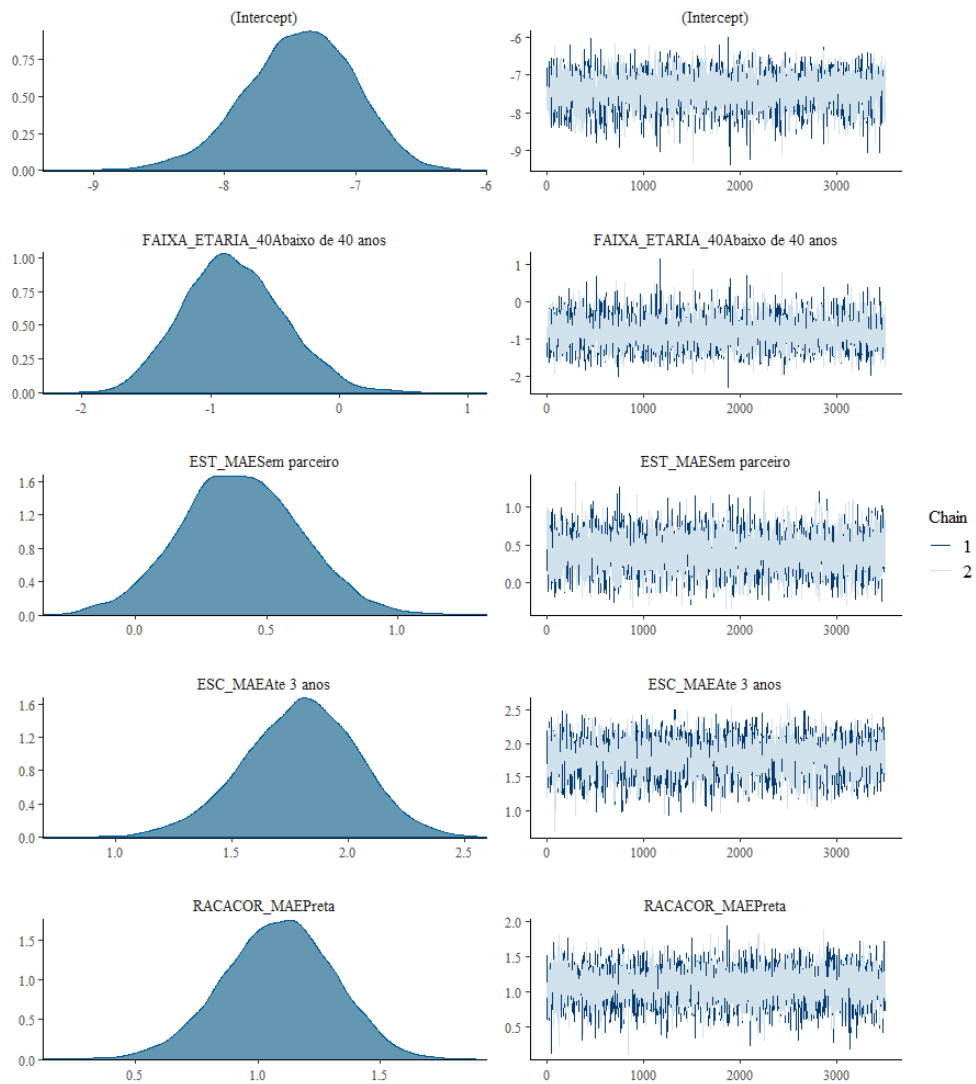


Figura 19: Densidades aproximadas e Cadeias obtidas via HCM dos valores gerados para os coeficientes - 2011.

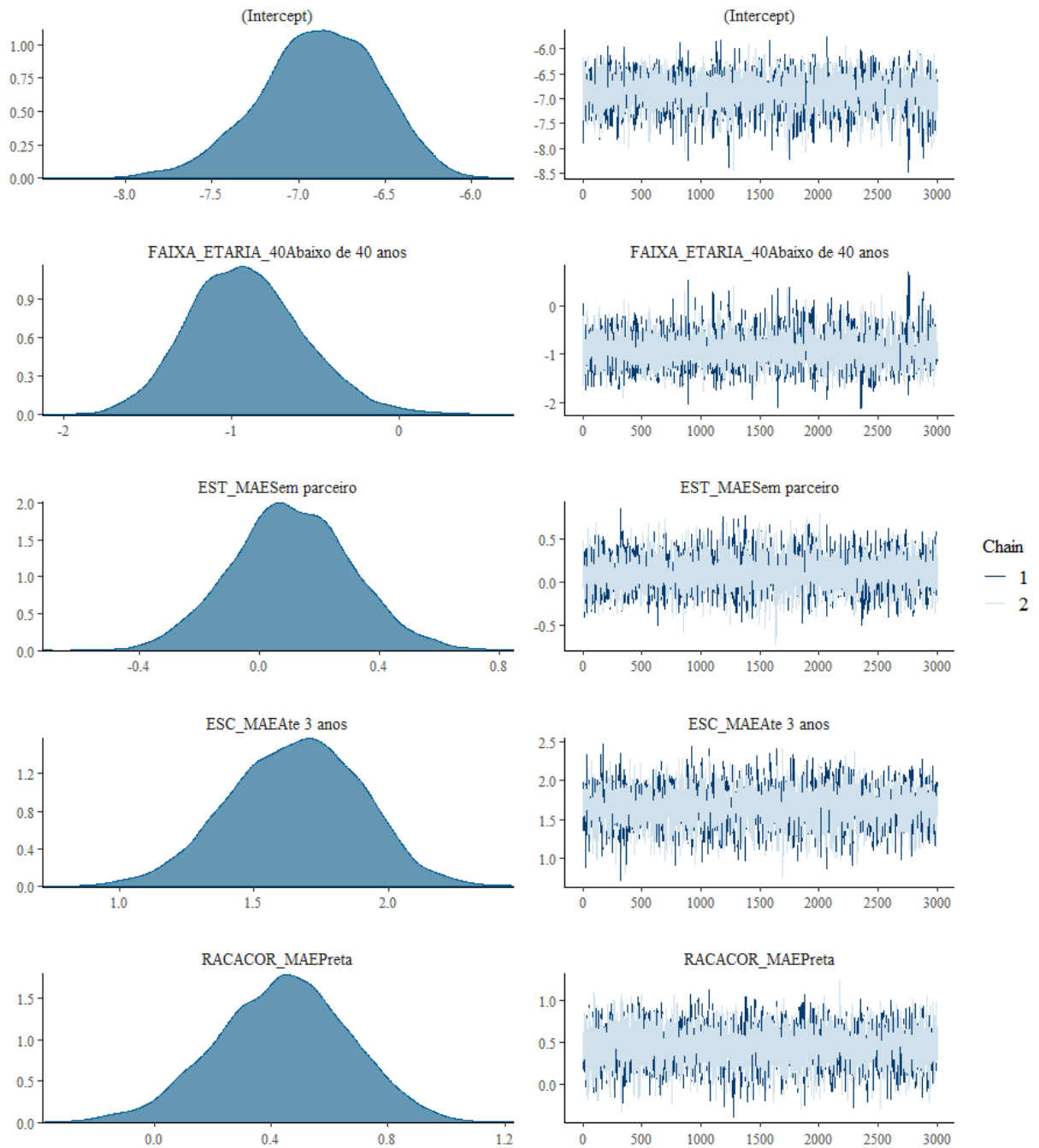


Figura 20: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2012.

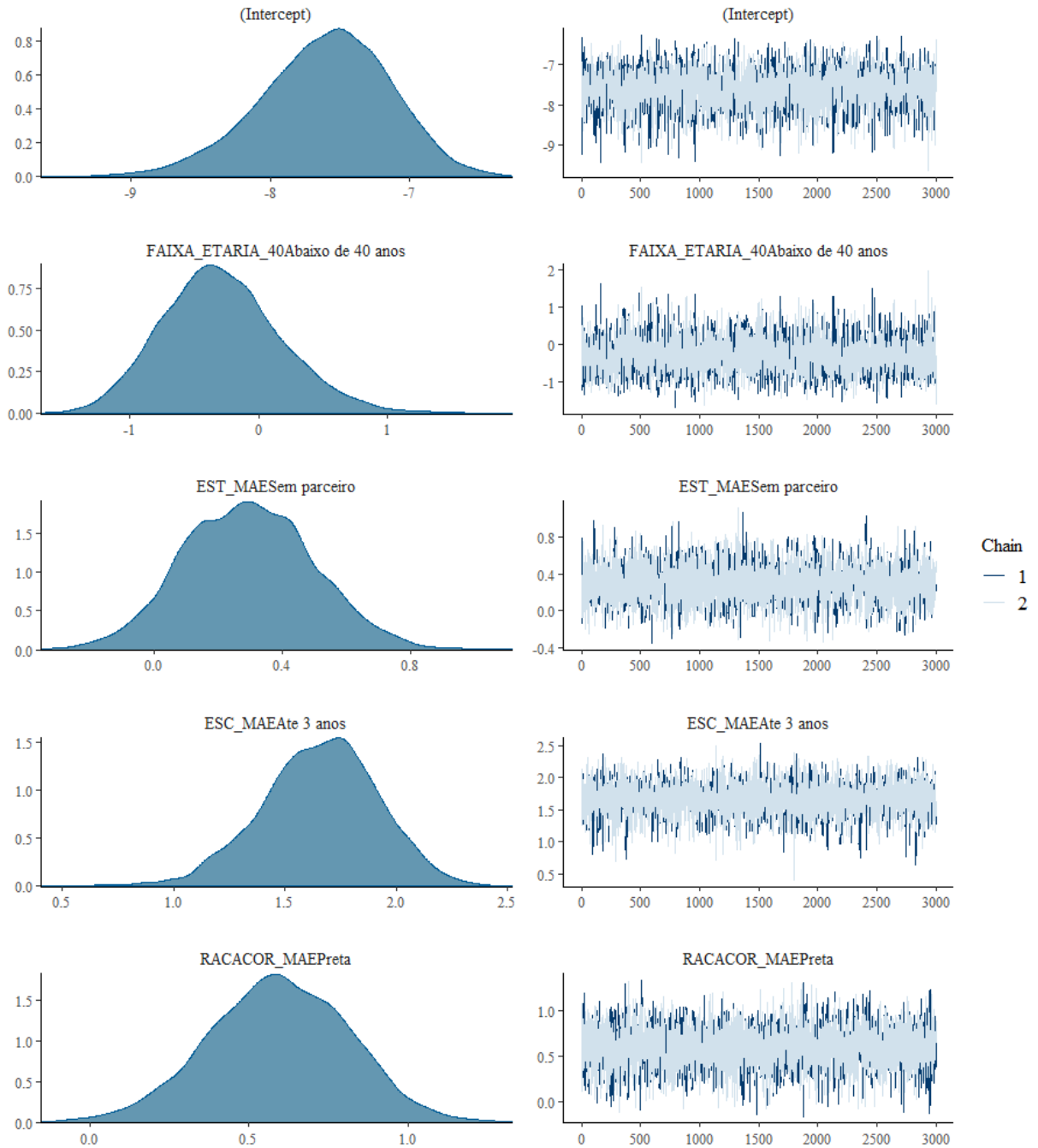


Figura 21: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2013.



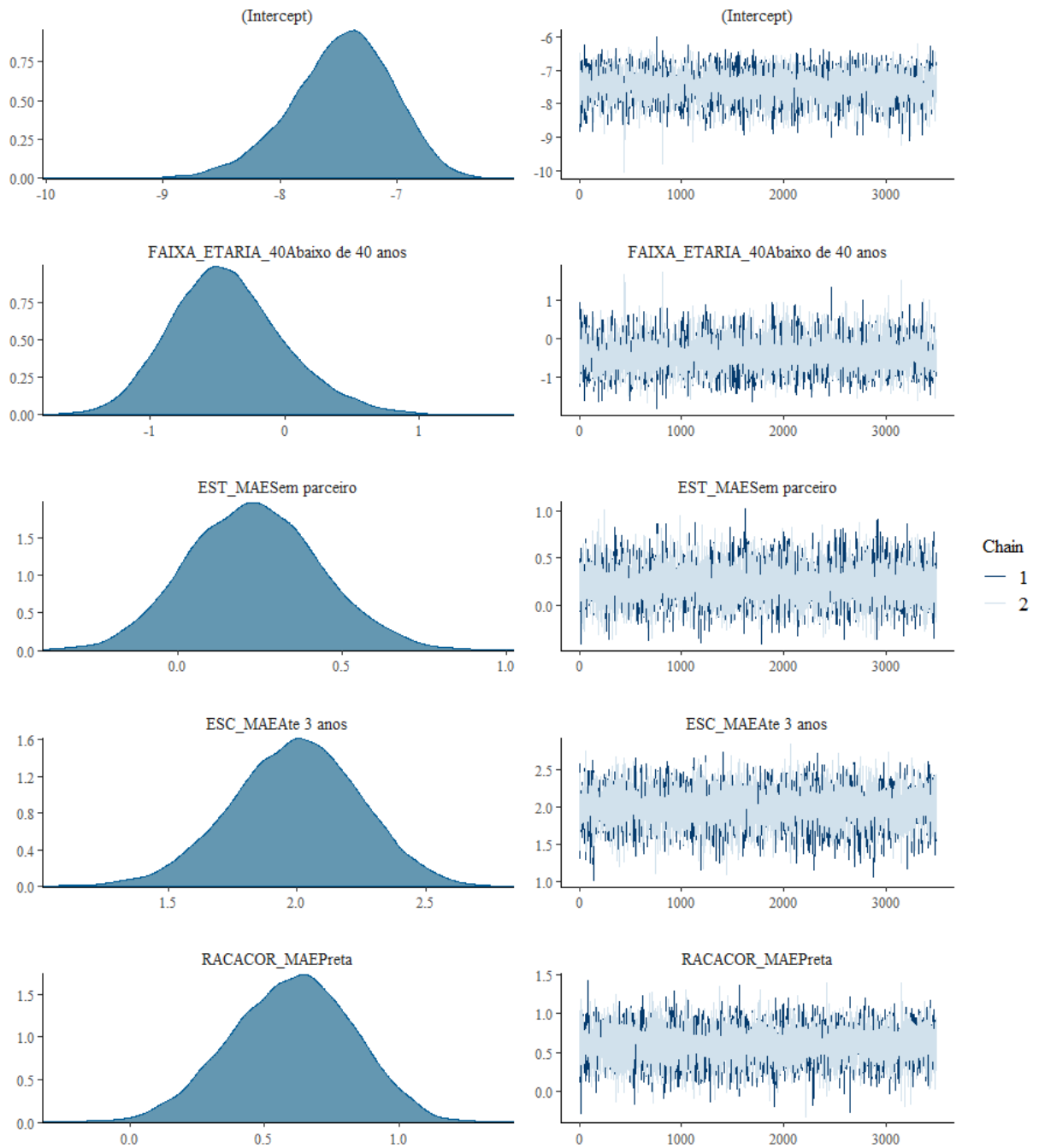


Figura 22: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2014.

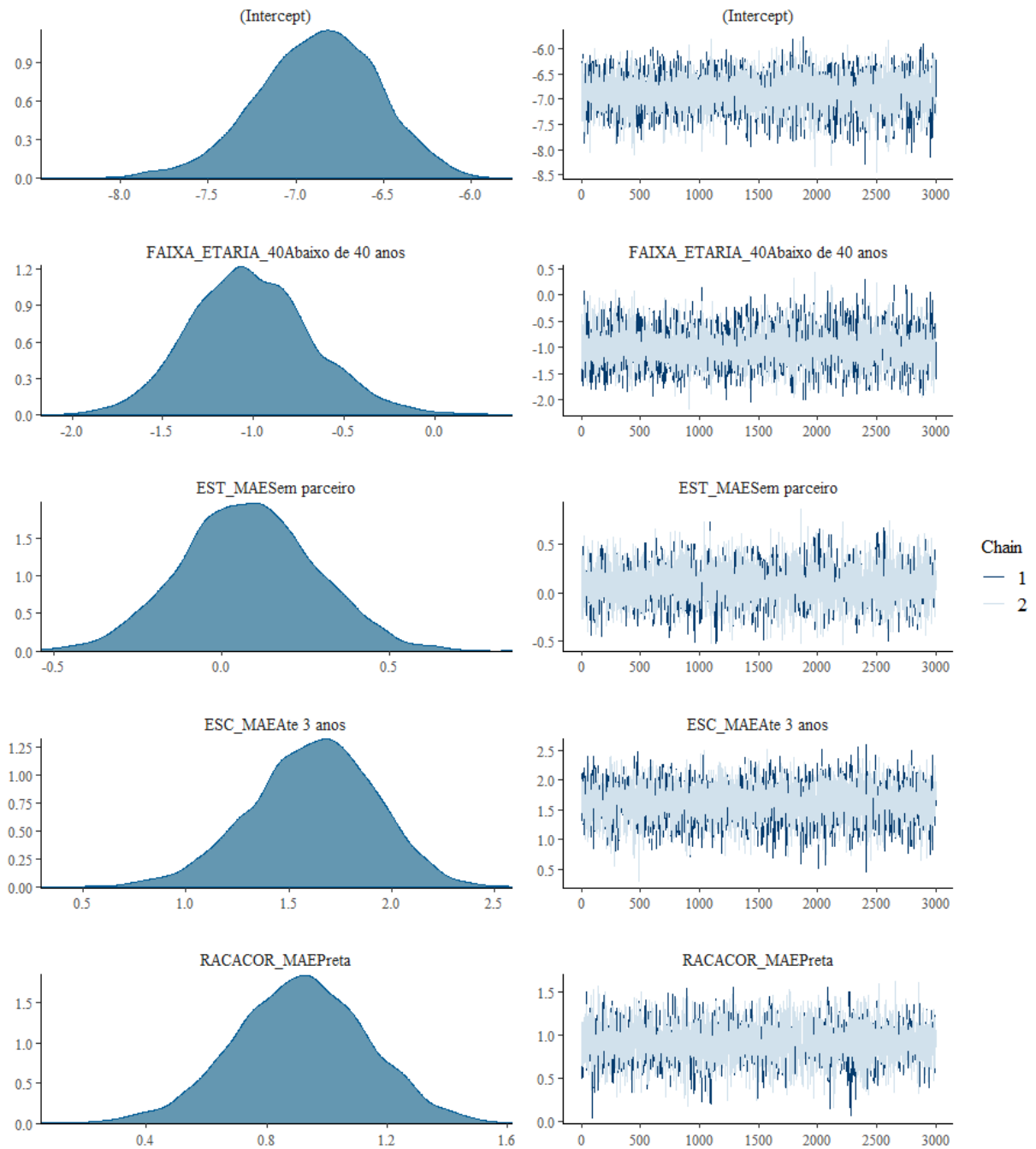


Figura 23: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2015.

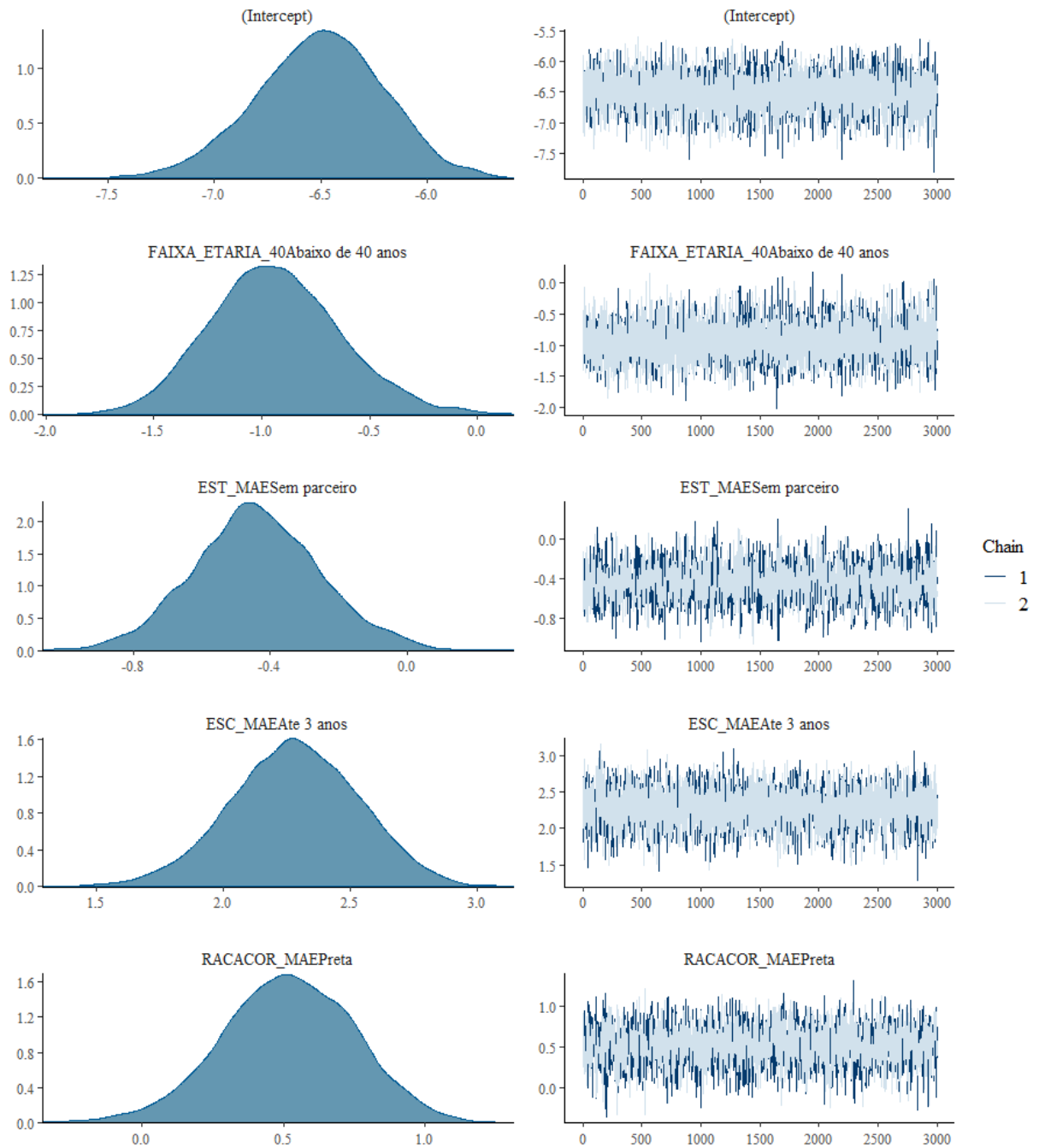


Figura 24: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2016.

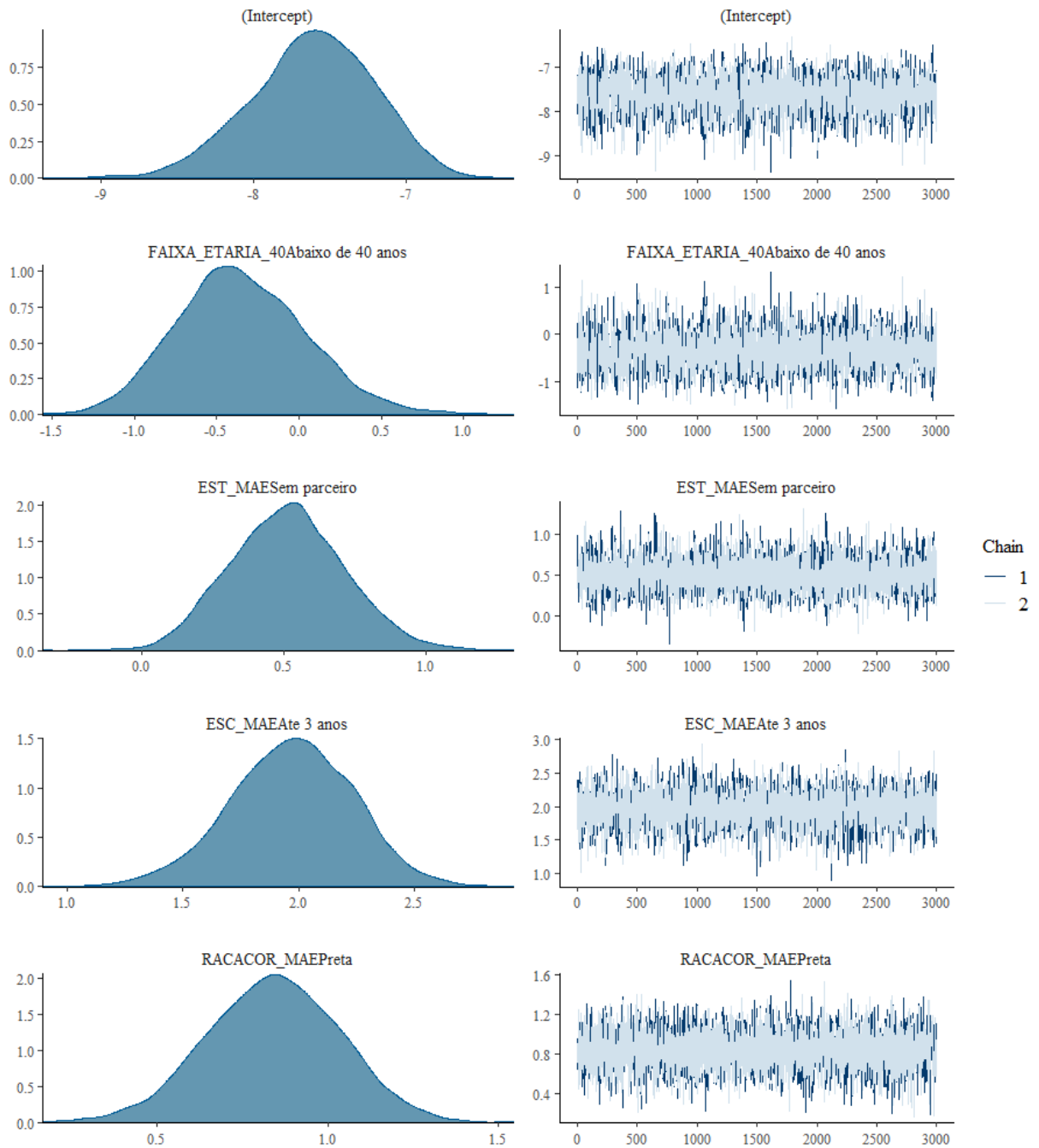


Figura 25: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2017.

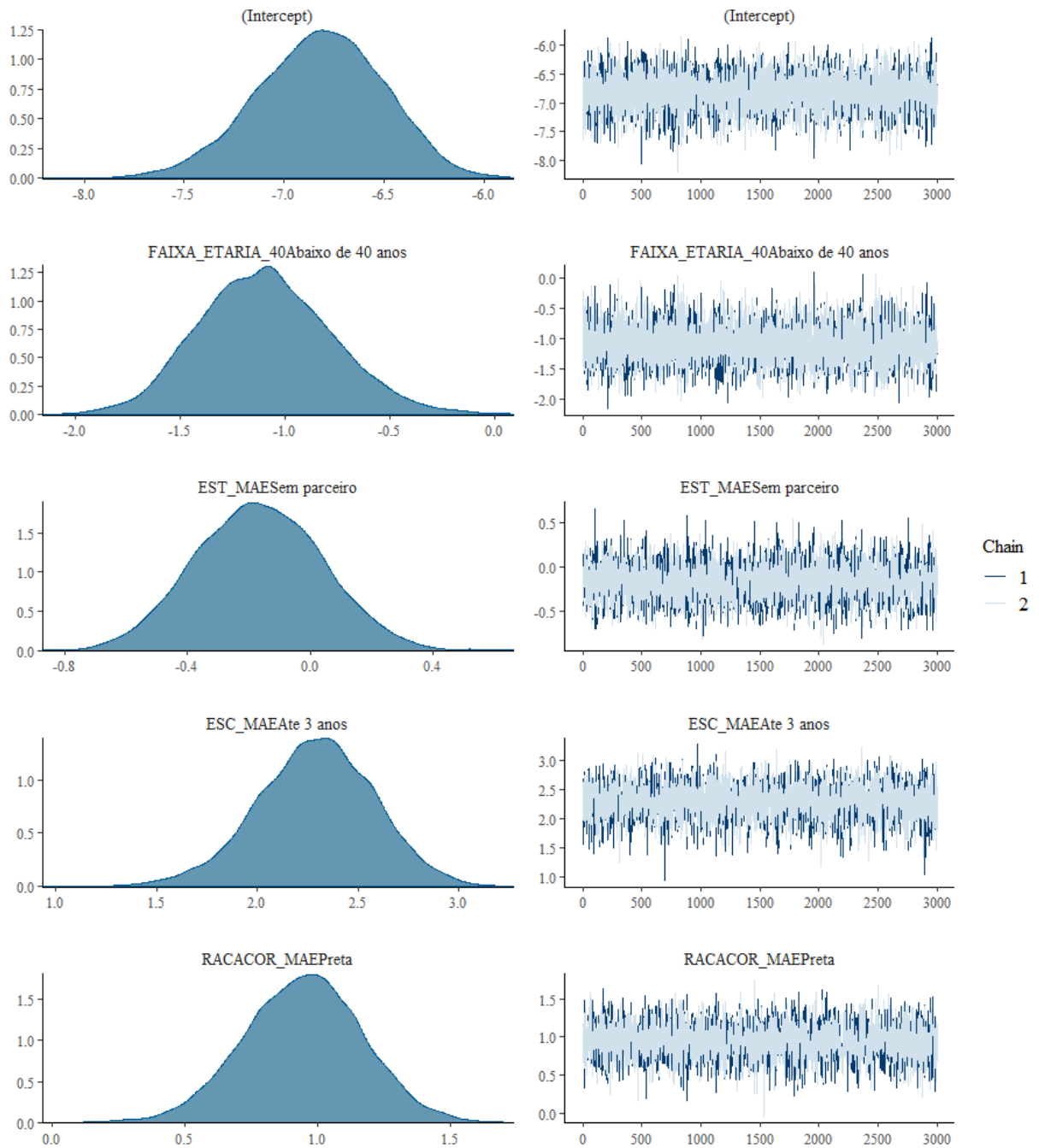


Figura 26: Densidades aproximadas e Cadeia obtida via HCM dos valores gerados para os coeficientes - 2018.

## APÊNDICE B – Função de Autocorrelação dos anos de 2011 a 2018

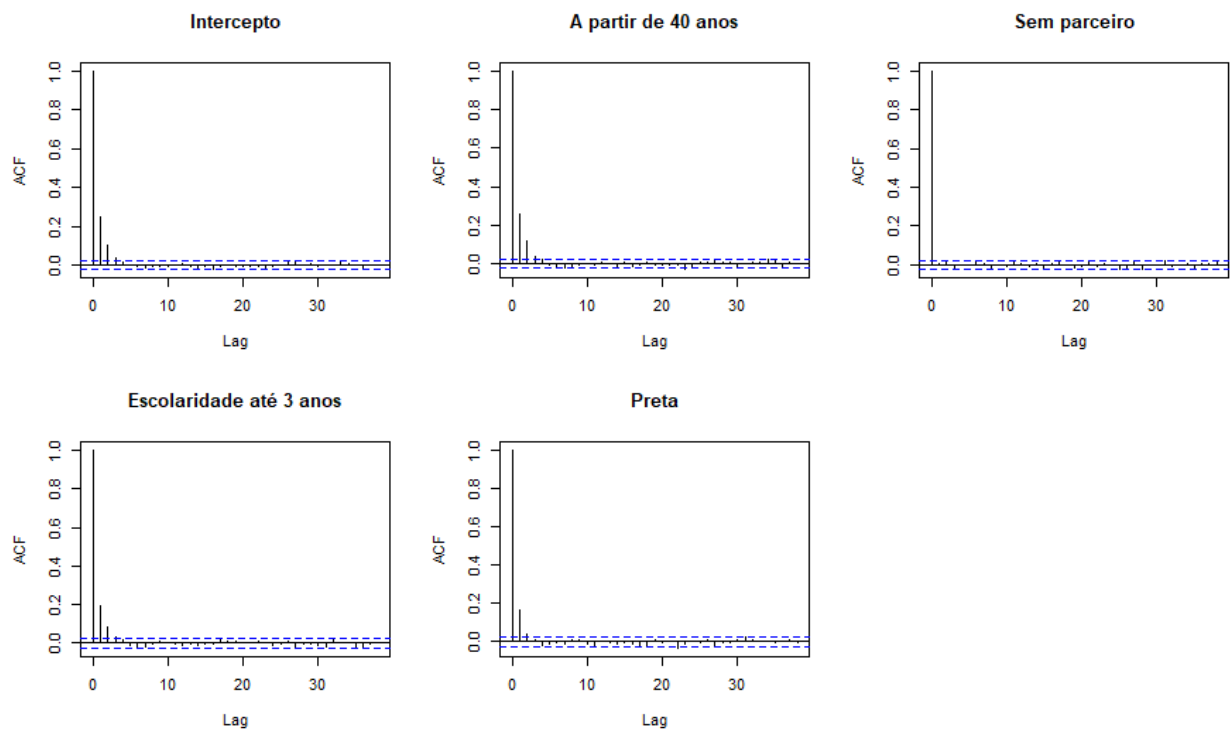


Figura 27: Função de Autocorrelação dos valores gerados para os coeficientes - 2011.

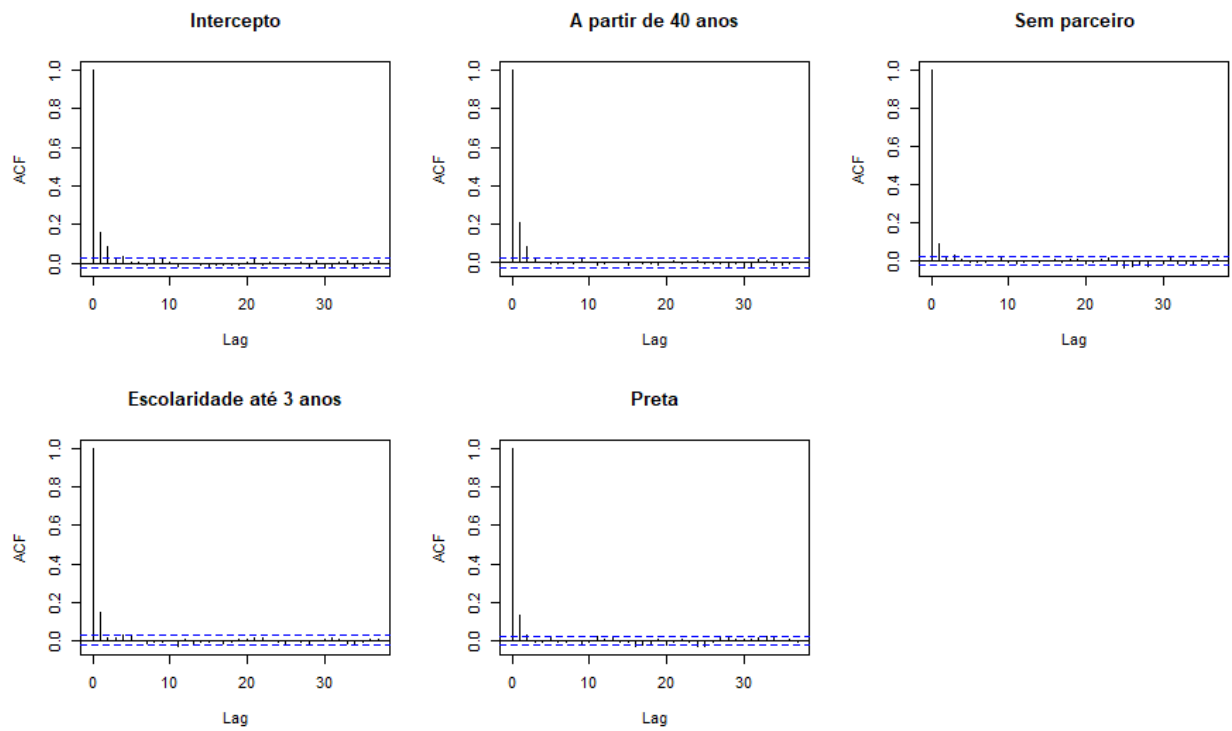


Figura 28: Função de Autocorrelação dos valores gerados para os coeficientes - 2012.

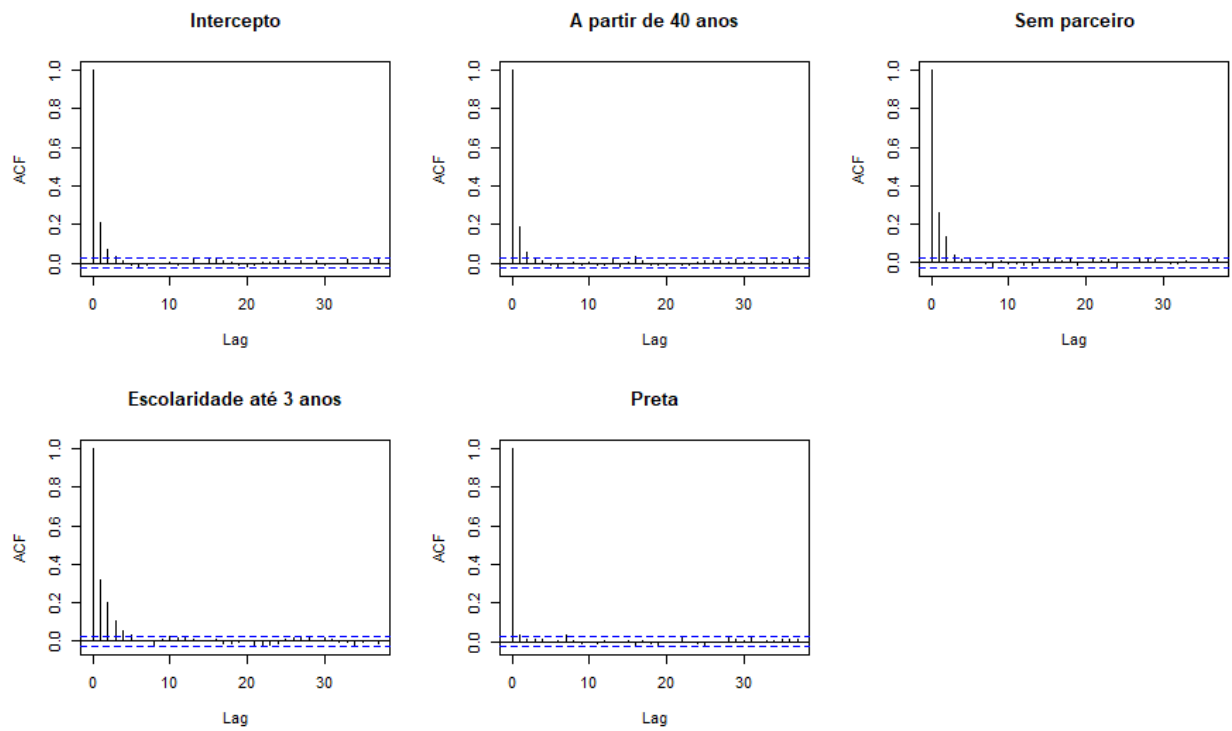


Figura 29: Função de Autocorrelação dos valores gerados para os coeficientes - 2013.

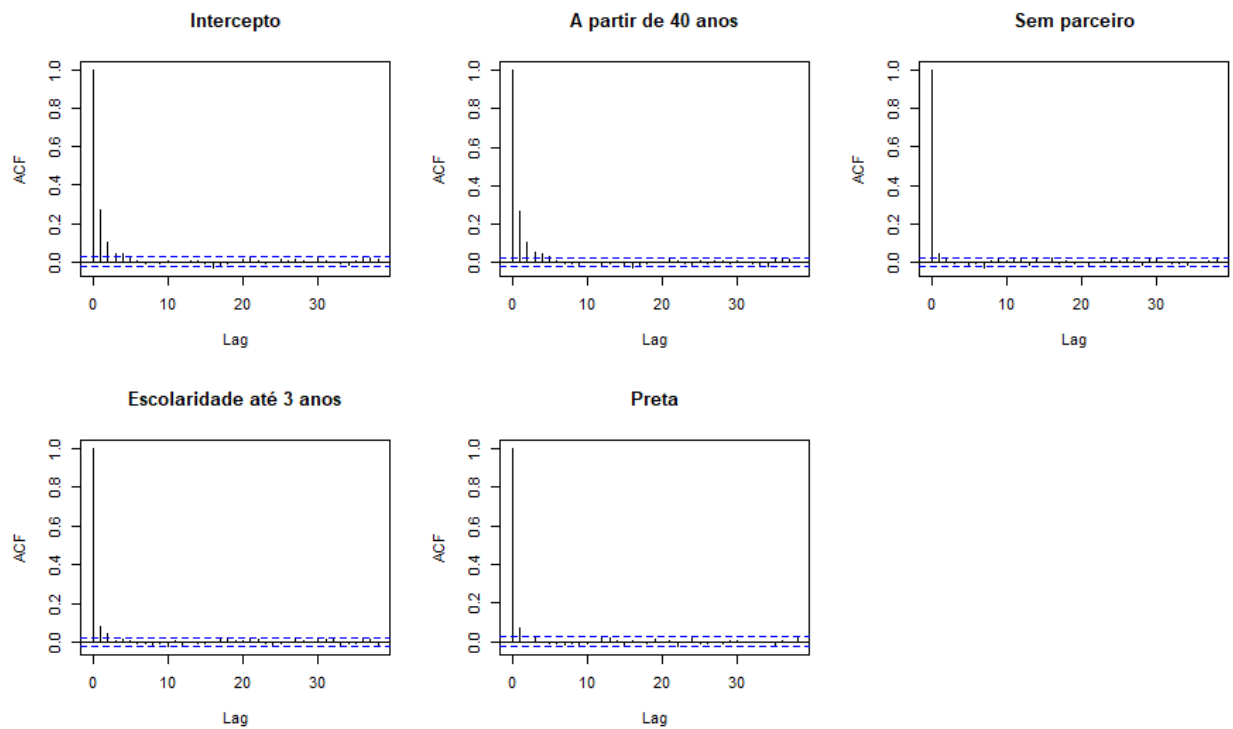


Figura 30: Função de Autocorrelação dos valores gerados para os coeficientes - 2014.

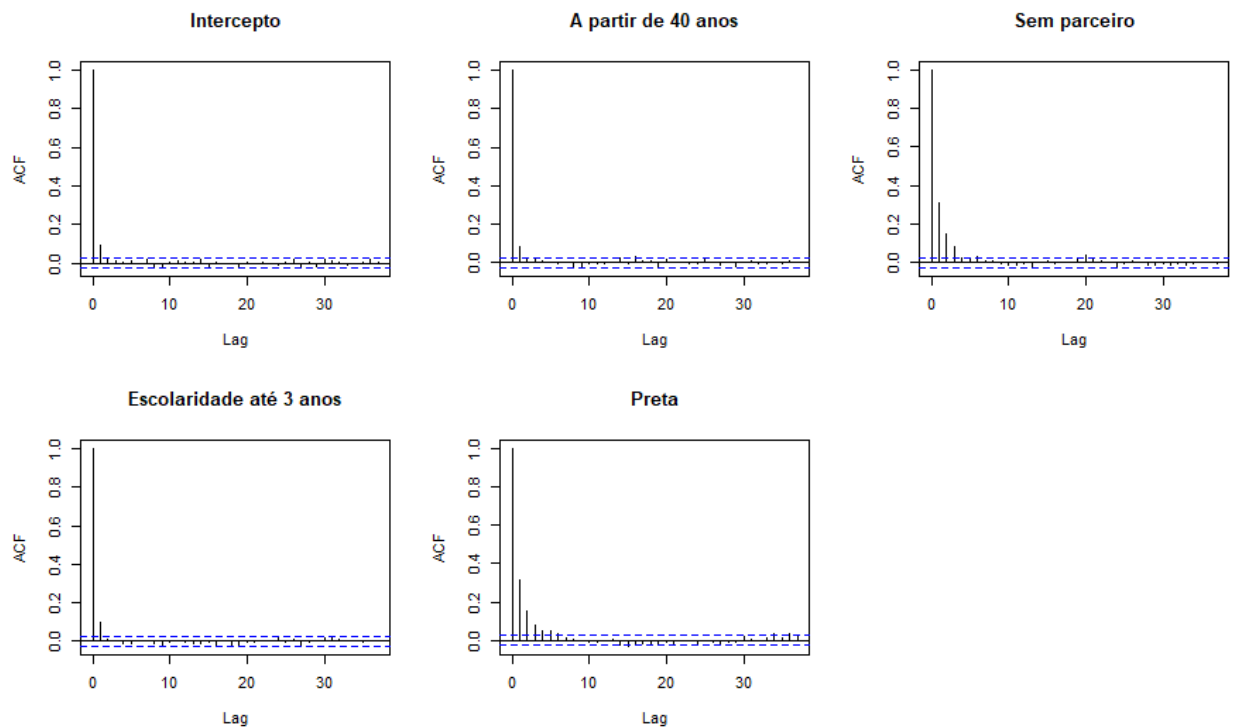


Figura 31: Função de Autocorrelação dos valores gerados para os coeficientes - 2015.



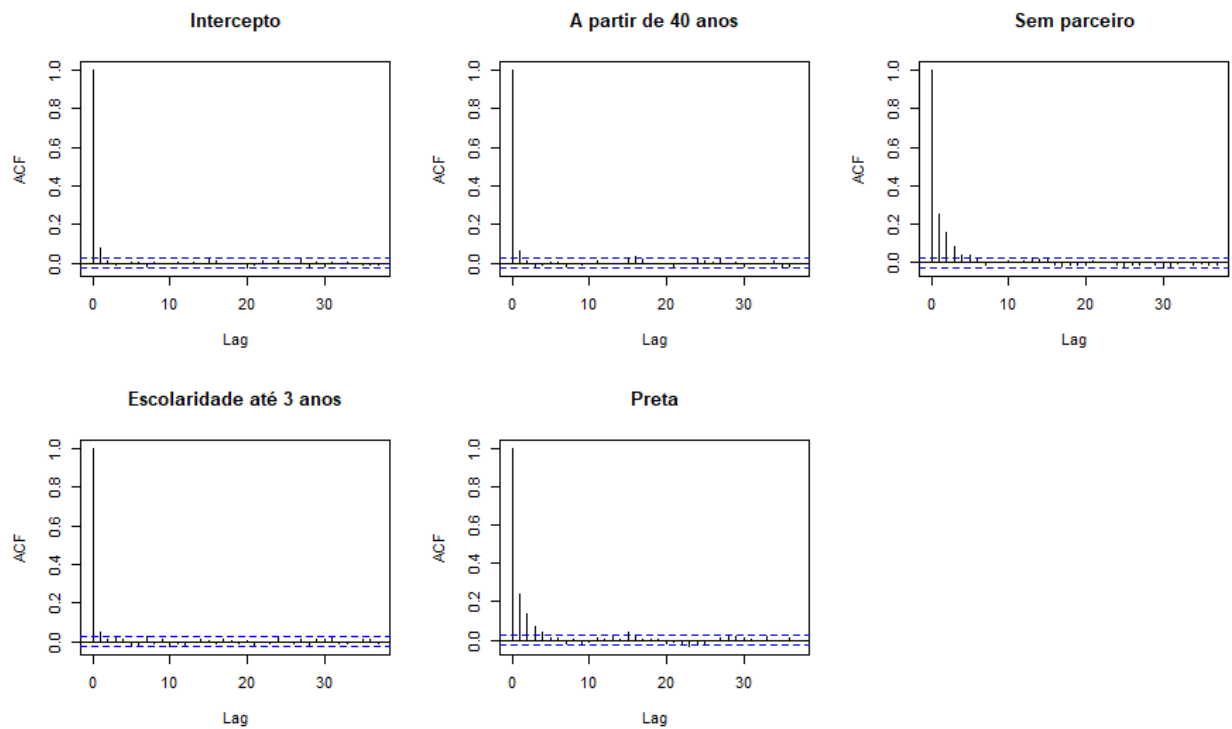


Figura 32: Função de Autocorrelação dos valores gerados para os coeficientes - 2016.

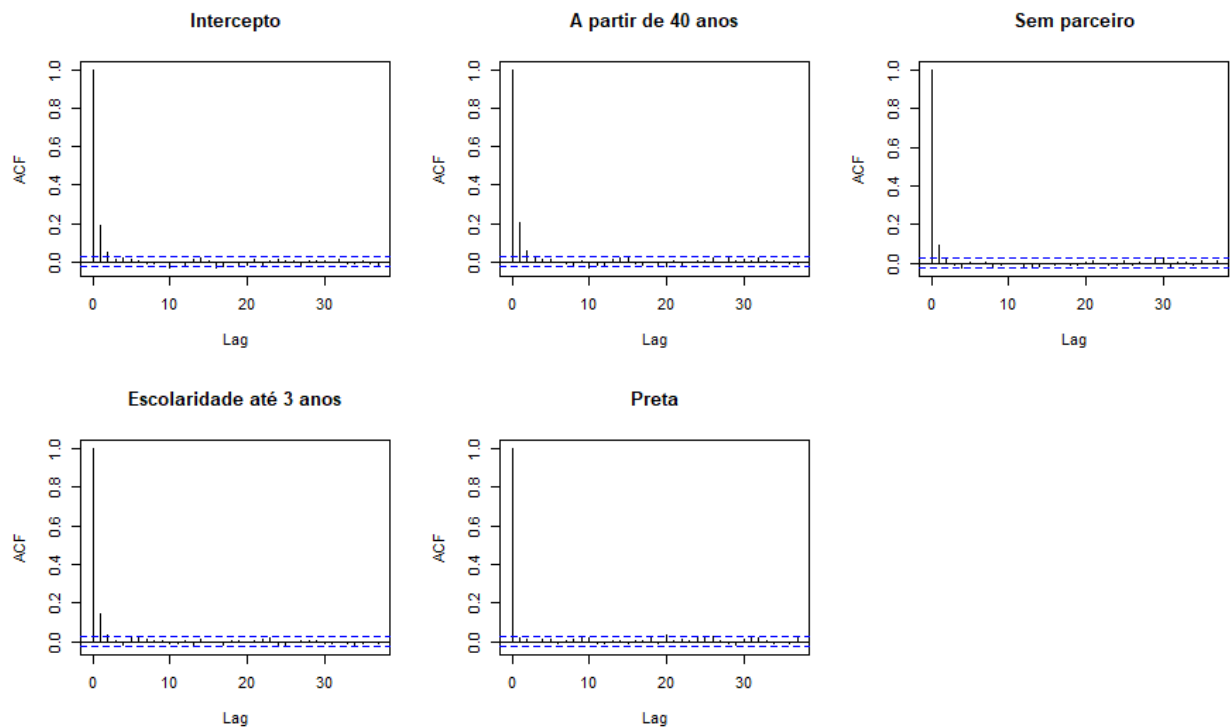


Figura 33: Função de Autocorrelação dos valores gerados para os coeficientes - 2017.

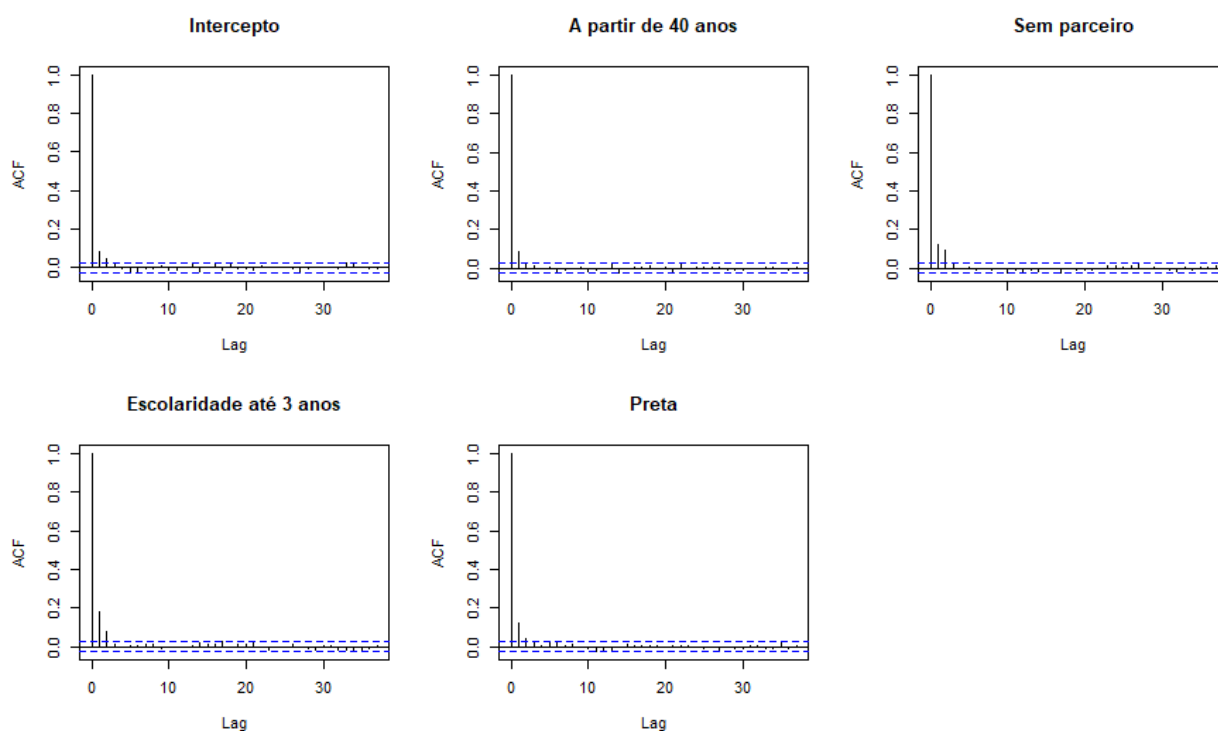


Figura 34: Função de Autocorrelação dos valores gerados para os coeficientes - 2018.