

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Programa de Pós-Graduação em Estatística

Maíra Soalheiro

**Modelagem Hierárquica Bayesiana de contatos: uma  
aplicação em modelos epidemiológicos compartimentais**

Belo Horizonte  
2022

Maíra Soalheiro

# **Modelagem Hierárquica Bayesiana de contatos: uma aplicação em modelos epidemiológicos compartimentais**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Adrian Pablo Hinojosa Luna

Belo Horizonte  
2022

Soalheiro, Maíra.

S676m      Modelagem hierárquica Bayesiana de contatos [manuscrito]:  
uma aplicação em modelos epidemiológicos compartimentais. /  
Maíra Soalheiro. – 2022.  
98 f. il.

Orientador: Adrian Pablo Hinojosa Luna.  
Dissertação (mestrado) - Universidade Federal de Minas  
Gerais, Instituto de Ciências Exatas, Departamento de Estatística  
Referências: f.69-73.

1. Estatística – Teses. 2. Modelos lineares hierárquicos –  
Teses. 3. COVID-19 Pandemia – Teses. 4. Análise de regressão  
– Teses. 5. Redes de relações sociais – Teses. I. Hinojosa Luna,  
Adrian Pablo. II. Universidade Federal de Minas Gerais, Instituto  
de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

## FOLHA DE APROVAÇÃO

**"Modelagem Hierárquica Bayesiana de contatos: uma aplicação em modelos epidemiológicos compartimentais"**

**MAÍRA SOALHEIRO**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 25 de fevereiro de 2022, pela banca constituída pelos membros:

Prof(a). Adrian Pablo Hinojosa Luna - Orientador  
DEST/UFMG

Prof(a). Luiz Henrique Duczmal  
DEST/UFMG

Prof(a). Silvio Segundo Salej Higgins  
FAFICH/UFMG

Prof(a). Gonzalo Panizo  
Instituto de Matemática e Ciencias Afines (IMCA-PERU)

Belo Horizonte, 25 de fevereiro de 2022.

# Agradecimentos

Os dois últimos anos foram repletos de emoções: alegrias, tristezas, “loucuras”, medo, tudo ainda intensificado pelo período pandêmico em que nos encontramos. Logo, ser grato é reconhecer que não passaria por tudo isso sozinha. Assim, gostaria de agradecer a todos que me ajudaram a chegar até aqui.

Primeiramente, gostaria de agradecer a Deus, pelas bênçãos e pela força ao longo desse período e em toda minha vida.

Agradeço aos meus pais Eliane e Gilberto, e aos meus irmãos Amanda e Igor, por sempre me apoiarem e me incentivarem a correr atrás dos meus sonhos e objetivos. E a Pucca, por ser a melhor “cãopanheira” que eu poderia ter. Eu seria nada sem vocês.

Agradeço a minha madrinha Cida e ao meu tio Helbert, por estarem sempre presentes. Às minhas avós Nhanhá e Odete e ao meu avô Levy, que mesmo no céu estão sempre comigo.

Muito obrigada aos meus amigos, em particular a Larissa, aos meus colegas de curso e a todos que me ajudaram durante esse curso de Mestrado e no desenvolvimento da minha dissertação. Em especial, gostaria de agradecer ao Otávio, Ricardo, Juliana, Reinaldo e Eduardo, sem vocês eu não teria conseguido.

Gostaria de agradecer a todos os professores, funcionários e colegas com quem tive o prazer de aprender ao longo destes anos na UFMG, tal como a todos os professores e funcionários, da educação infantil ao ensino médio, que participaram de minha formação. Assim como também a equipe de Gerência de Avaliação Educacional (Bia, Roberto, Val, Iáskara e Hamilton), pela aprendizagem pessoal e incentivo profissional.

Meu muito obrigada ao meu orientador, Professor Adrian Hinojosa, pelo auxílio, atenção, orientação e paciência nesses últimos três anos. Aos Professores Luiz Duczmal, Silvio Salej e Gonzalo Panizo por terem aceitado o convite para serem membros da banca.

Agradeço a CAPES pela bolsa que me foi conferida, assim como também a FAPEMIG e o CNPq. Sem a bolsa e os investimentos no Programa não seria possível concluir essa pós-graduação.

Enfim, meus sinceros agradecimentos a todos que acreditaram e torceram para que esse sonho se tornasse realidade.

# Resumo

Os Modelos Lineares Hierárquicos, também chamados de Modelos de Regressão Multi-níveis ou Modelos de Efeitos Mistos, é um método de modelagem para conjuntos de dados aninhados que apresentam uma estrutura hierárquica, sendo utilizado para estudos que buscam investigar os efeitos de variáveis a nível individual e a níveis de grupo (Snidjers, 2016), como também para estudos longitudinais, que contam com a presença de medidas repetidas. Esse tipo de ajuste explora a relação existente entre os indivíduos e o meio a ser estudado e, entende que por isso, todas as associações possíveis devem ser analisadas. O estudo em questão tem por objetivo propor um modelo multinível bayesiano para estimar as taxas de contatos entre os moradores do Aglomerado da Serra por grupos de idade e círculos sociais, embasado nos estudos do projeto POLYMOD (Mossong et al., 2008) e no artigo de Prem et al.(2017). As taxas estimadas serão projetadas para regiões da cidade de Belo Horizonte a fim de aplicá-las em um modelo SIR (Susceptible-Infected-Removed), como parte dos estudos para mitigar os impactos da COVID-19. Causada pelo novo coronavírus, SARS-CoV-2, a transmissão do vírus ocorre de uma pessoa contaminada para outra e com milhões de casos e mortes ao redor do mundo, buscar compreender os padrões das redes de contatos considerando as variações que podem ocorrer devido as faixas etárias e locais de interação é de suma importância, visto as quais podem levar às diferenças no efeito de medidas de distanciamento social.

**Palavras chave:** Modelos Multiníveis, Modelo Linear Hierárquico, Modelo SIR, modelos compartimentais, Coronavírus, Covid-19, redes de contatos sociais.

# Abstract

Hierarchical Linear Models, also called Multilevel Regression Models or Mixed-Effects Models, is a modeling method for nested data sets that present a hierarchical structure, being used for studies that pursue to investigate the effects of variables at the individual level and at group levels, as well as for longitudinal studies, which rely on the presence of repeated measures. This type of adjustment explores the relationship between individuals and the environment to be studied, and understands that for this reason, all possible associations must be analyzed. The study in question aims to propose a multilevel Bayesian model to estimate contact rates among residents of Agglomerado da Serra by age groups and social circles, based on the studies of the POLYMOD project (Mossong et al., 2008) and the article by Prem et al. (2017). The estimated rates will be projected for regions of the city of Belo Horizonte in order to apply them in a SIR (Susceptible-Infected-Removed) model, as part of the studies to mitigate the impacts of COVID-19. Caused by the new coronavirus, SARS-CoV-2, the transmission of the virus occurs from one infected person to another and with millions of cases and deaths around the world, seeking to understand the patterns of contact networks considering the variations that may occur due to age groups and places of interaction, is of paramount importance, as they can lead to differences in the effect of social distancing measures.

**Keywords:** Multilevel Models, Hierarchical Linear Models, SIR Model, Coronavirus, Covid-19, social contact networking.

## Lista de Figuras

1	Estrutura do modelo hierárquico. . . . .	13
2	Estados modelo SIR. . . . .	19
3	Frequência do número de entrevistados por bairro - Banco 1. . . . .	22
4	Frequência do número de moradores por domicílio - Banco 1. . . . .	23
5	Número de entrevistados por faixa etária e sexo - Banco 1. . . . .	24
6	Frequência do número de contatos - Banco 1. . . . .	25
7	Distribuição dos contatos nos círculos ‘Casa’ e ‘Outros’, respectivamente - Banco 1. . . . .	26
8	Número de entrevistados por faixa etária e sexo - Banco 2. . . . .	28
9	Frequência do número de domicílios por bairro - Banco 2. . . . .	29
10	Frequência do número de moradores por domicílio - Banco 2. . . . .	29
11	Número de indivíduos por faixa etária e sexo - Banco calibrado. . . . .	31
12	Frequência do número de domicílio por bairro - Banco calibrado. . . . .	31
13	Frequência do número de moradores por domicílio - Banco calibrado. . . . .	32
14	Frequência do número de contatos. . . . .	33
15	Diagnóstico PSIS - Modelo ‘casa7’. . . . .	37
16	Matriz de contatos por faixa etária: contexto Casa. Cores mais fortes indicam maiores médias de contatos. . . . .	40
17	Diagnóstico PSIS - Modelo ‘outros4’. . . . .	41
18	Taxa de contatos por faixa etária: contexto Outros. Cores mais fortes indicam maiores médias de contatos. . . . .	43
19	Diagnóstico PSIS - Modelo ‘geral3’. . . . .	45
20	Taxa de contatos por faixa etária: contexto Geral. Cores mais fortes indicam maiores médias de contatos. . . . .	47
21	Regiões do município de Belo Horizonte . . . . .	54
22	Cenário 1: Prevalência ( $D = 0.8$ ) . . . . .	56
23	Cenário 1: Estimação $R_t$ para $D = 0.8$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	57
24	Cenário 1: Prevalência ( $D = 0.5$ ) . . . . .	57
25	Cenário 1: Estimação $R_t$ para $D = 0.5$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	58
26	Cenário 1: Prevalência ( $D = 0.1$ ) . . . . .	58
27	Cenário 1: Estimação $R_t$ para $D = 0.1$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	59
28	Cenário 2: Prevalência ( $D = 0.8$ ) . . . . .	60
29	Cenário 2: Estimação $R_t$ para $D = 0.8$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	61
30	Cenário 2: Prevalência ( $D = 0.5$ ) . . . . .	61
31	Cenário 2: Estimação $R_t$ para $D = 0.5$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	62
32	Cenário 2: Prevalência ( $D = 0.1$ ) . . . . .	62



33	Cenário 2: Estimação $R_t$ para $D = 0.1$ . Regiões: <b>A)</b> Barreiro, <b>B)</b> Centro, <b>C)</b> Aglomerado da Serra, <b>D)</b> Centro-Sul, <b>E)</b> Leste, <b>F)</b> Nordeste, <b>G)</b> Noroeste, <b>H)</b> Norte, <b>I)</b> Oeste, <b>J)</b> Pampulha, <b>K)</b> Venda Nova. . . . .	63
34	Boxplot $R_t$ (dia 7 - dia 21): Cenário 1 ( $D = 0,8$ ). . . . .	64
35	Boxplot $R_t$ (dia 7 - dia 21): Cenário 2 ( $D = 0,8$ ). . . . .	64
36	Boxplot $R_t$ (dia 7 - dia 21): Cenário 1 ( $D = 0,5$ ). . . . .	65
37	Boxplot $R_t$ (dia 7 - dia 21): Cenário 2 ( $D = 0,5$ ). . . . .	65
38	Boxplot $R_t$ (dia 7 - dia 21): Cenário 1 ( $D = 0,1$ ). . . . .	66
39	Boxplot $R_t$ (dia 7 - dia 21): Cenário 2 ( $D = 0,1$ ). . . . .	66
40	Quantitativo de indivíduos por faixa etária e sexo: A) Banco 1, B) Banco 2 e C) Banco calibrado. . . . .	74
41	Frequência do número de moradores por domicílio: A) Banco 1, B) Banco 2 e C) Banco calibrado. . . . .	76
42	Frequência do número de moradores por bairro: A) Banco 1 e C) Banco calibrado. Frequência do número de domicílios por bairro: B) Banco 2. . . . .	77
43	Diagnóstico de convergência (1) - Modelo ‘casa7’. . . . .	87
44	Diagnóstico de convergência (2) - Modelo ‘casa7’. . . . .	87
45	Diagnóstico de convergência (3) - Modelo ‘casa7’. . . . .	88
49	Efeitos condicionais das marginais - Modelo ‘casa7’. . . . .	88
50	Efeitos marginais de interação e da variável tamanho do domicílio - Modelo ‘casa7’. . . . .	88
46	Diagnóstico de convergência (4) - Modelo ‘casa7’. . . . .	89
47	Diagnóstico de convergência (5) - Modelo ‘casa7’. . . . .	89
48	Diagnóstico de convergência (6) - Modelo ‘casa7’. . . . .	90
51	Verificação preditiva a posteriori - Modelo ‘casa7’. . . . .	90
52	Diagnóstico de convergência (1) - Modelo ‘outros4’. . . . .	91
53	Diagnóstico de convergência (2) - Modelo ‘outros4’. . . . .	91
54	Diagnóstico de convergência (3) - Modelo ‘outros4’. . . . .	92
58	Efeitos condicionais das marginais - Modelo ‘outros4’. . . . .	92
59	Efeitos marginais de interação e verificação preditiva a posteriori - Modelo ‘outros4’. . . . .	92
55	Diagnóstico de convergência (4) - Modelo ‘outros4’. . . . .	93
56	Diagnóstico de convergência (5) - Modelo ‘outros4’. . . . .	93
57	Diagnóstico de convergência (6) - Modelo ‘outros4’. . . . .	94
60	Diagnóstico de convergência (1) - Modelo ‘geral3’. . . . .	95
61	Diagnóstico de convergência (2) - Modelo ‘geral3’. . . . .	95
62	Diagnóstico de convergência (3) - Modelo ‘geral3’. . . . .	96
66	Efeitos condicionais das marginais - Modelo ‘geral3’. . . . .	96
67	Efeitos marginais de interação e verificação preditiva a posteriori- Modelo ‘geral3’. . . . .	96
63	Diagnóstico de convergência (4) - Modelo ‘geral3’. . . . .	97
64	Diagnóstico de convergência (5) - Modelo ‘geral3’. . . . .	97
65	Diagnóstico de convergência (6) - Modelo ‘geral3’. . . . .	98

## Lista de Tabelas

1	Total de entrevistados por faixa etária - Banco 1. . . . .	24
2	Total de contatos por círculo social - Banco 1. . . . .	25
3	Total de contatos por sexo - Banco 1. . . . .	26
4	Total de contatos por círculo social e faixas etárias - Banco 1. . . . .	27
5	Proporção da estrutura etária dos residentes - Banco 2 . . . . .	28
6	Proporção da estrutura etária - Banco calibrado. . . . .	30
7	Total de contatos por círculo social - Banco calibrado. . . . .	32
8	Total de contatos por sexo - Banco calibrado. . . . .	33
9	Total de contatos por círculo social e faixas etárias - Banco calibrado. . . . .	34
10	Modelos ajustados - Casa. . . . .	36
11	Comparativo LOO's - Casa. . . . .	37
12	Resumo do modelo 'casa7'. . . . .	39
13	Matriz de contatos por faixa etária: contexto Casa. . . . .	40
14	Modelos ajustados - Outros. . . . .	40
15	Comparativo LOO's - Outros. . . . .	41
16	Resumo do modelo 'outros4'. . . . .	42
17	Matriz de contatos por faixa etária: contexto Outros. . . . .	43
18	Modelos ajustados - Contexto Geral. . . . .	44
19	Comparativo LOO's - Contexto Geral. . . . .	44
20	Resumo do modelo 'geral3'. . . . .	46
21	Matriz de contatos por faixa etária: contexto Geral. . . . .	47
22	Matriz de contatos por faixa etária: Barreiro. . . . .	49
23	Matriz de contatos por faixa etária: Centro. . . . .	49
24	Matriz de contatos por faixa etária: Aglomerado da Serra. . . . .	49
25	Matriz de contatos por faixa etária: Centro-Sul. . . . .	50
26	Matriz de contatos por faixa etária: Leste. . . . .	50
27	Matriz de contatos por faixa etária: Nordeste. . . . .	50
28	Matriz de contatos por faixa etária: Noroeste. . . . .	50
29	Matriz de contatos por faixa etária: Norte. . . . .	50
30	Matriz de contatos por faixa etária: Oeste. . . . .	51
31	Matriz de contatos por faixa etária: Pampulha. . . . .	51
32	Matriz de contatos por faixa etária: Venda Nova. . . . .	51
33	População total, suscetíveis, infectados e recuperados por região. . . . .	53
34	Parâmetros do modelo SIR . . . . .	55
35	Proporção por sexo (%) - Comparação. . . . .	75
36	Proporção da estrutura etária (%) - Comparação. . . . .	75
37	Proporção por tamanho do domicílio (%) - Comparação. . . . .	76
38	Proporção por bairro (%) - Comparação. . . . .	77

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>11</b>
1.1	Motivação e Objetivos . . . . .	12
<b>2</b>	<b>Modelos Lineares Hierárquicos</b>	<b>14</b>
2.1	Modelo de Dois Níveis . . . . .	14
2.2	Variações do modelo . . . . .	16
2.2.1	Modelo de coeficientes aleatórios . . . . .	16
2.2.2	Modelo com intercepto aleatório . . . . .	16
2.2.3	Modelo com inclinação aleatória . . . . .	16
2.2.4	Modelo com intercepto e inclinações aleatórios . . . . .	16
2.3	Abordagem Bayesiana . . . . .	16
2.3.1	Modelos mistos . . . . .	17
<b>3</b>	<b>Modelagem Matemática em Epidemiologia</b>	<b>18</b>
3.1	Modelos Compartimentais: Modelo SIR . . . . .	18
3.1.1	Número de Reprodutibilidade Basal ( $R_0$ ) . . . . .	20
3.1.2	Número de Reprodução instantâneo ( $R_t$ ) . . . . .	20
<b>4</b>	<b>Descrição dos dados</b>	<b>22</b>
4.1	Banco 1 . . . . .	22
4.2	Banco 2 . . . . .	27
4.3	Banco calibrado . . . . .	30
<b>5</b>	<b>Aplicação 1: Modelos de contatos</b>	<b>35</b>
5.1	Modelo Hierárquico Bayesiano . . . . .	35
5.1.1	Círculo social: Casa . . . . .	35
5.1.2	Círculo social: Outros . . . . .	39
5.1.3	Contexto Geral . . . . .	43
5.2	Projeção . . . . .	48
<b>6</b>	<b>Aplicação 2: Modelos epidemiológicos</b>	<b>52</b>
6.1	Modelo compartimental: SIR acoplado em redes . . . . .	52
6.2	Cenário 1: probabilidade de transmissão = 0.09 . . . . .	56
6.2.1	Situação 1: $D = 0.8$ . . . . .	56
6.2.2	Situação 2: $D = 0.5$ . . . . .	57
6.2.3	Situação 3: $D = 0.1$ . . . . .	58
6.3	Cenário 2: probabilidade de transmissão = 0.05 . . . . .	60
6.3.1	Situação 1: $D = 0.8$ . . . . .	60
6.3.2	Situação 2: $D = 0.5$ . . . . .	61

6.3.3	Situação 3: $D = 0.1$ . . . . .	62
6.4	Comparação entre os $R'_t$ s dos Cenários 1 e 2 . . . . .	64
<b>7</b>	<b>Conclusão</b>	<b>67</b>
<b>8</b>	<b>Anexo 1</b>	<b>74</b>
<b>9</b>	<b>Anexo 2</b>	<b>78</b>
<b>10</b>	<b>Anexo 3</b>	<b>87</b>
<b>11</b>	<b>Anexo 4</b>	<b>91</b>
<b>12</b>	<b>Anexo 5</b>	<b>95</b>

# 1 Introdução

Ao longo da história, a humanidade sofreu em diversas epidemias, estas alavancadas pelas guerras, colonizações e revoluções (Ujvari, Stefan, 2003). Somado a isso, fatores como a migração e a globalização possibilitaram o aumento da disseminação de doenças, o que permitiu que estas viajassem de regiões a regiões e até entre continentes. As pandemias, como a peste bubônica e a gripe de 1918, já mataram milhões ao redor do mundo e, a mais recente delas, é a pandemia da COVID-19.

Causada pelo novo coronavírus, SARS-CoV-2, a COVID-19 foi identificada primeiramente em Wuhan, na China, em dezembro de 2019 e declarada pandemia pela Organização Mundial da Saúde (OMS) em março de 2020. Atualmente, se tem registrado mais de 26,5 milhões de casos e mais de 630 mil <sup>1</sup> mortes apenas no Brasil.

A transmissão do vírus ocorre de uma pessoa infectada para outra, seja por contato próximo por meio de toque do aperto de mão contaminada, gotículas de saliva, espirro ou tosse <sup>2</sup>. Dados exemplos da relevância do comportamento humano sobre o controle de infecções (Bauch, 2013), as redes de contatos sociais apresentam um papel essencial para a caracterização da propagação de doenças transmissíveis por contato, como o coronavírus. Outrossim, modelos que exploram interações comportamento-doença podem auxiliar na compreensão do desenrolar das pandemias, a partir de condições sociais e epidemiológicas (Pedro, Sansao A, 2020). E à vista disso, identificar os padrões de contatos considerando as variações que podem ocorrer devido as faixas etárias e locais de interação, é de suma importância, visto as quais podem levar às diferenças no efeito de medidas de distanciamento social.

A pretexto da pandemia do novo coronavírus muitos são os estudos que buscam modelar o efeito da propagação da doença, inclusive esta dissertação. Isto posto, algumas das principais referências para este ensaio são o modelo POLYMOD - “*Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases.*”, Mossong, J. et al. (2008); e os artigos “*Projecting social contact matrices in 152 countries using contact surveys and demographic data*”, Prem et al. (2017); “*Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era*”, Prem et al. (2020); “*Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK*”, Jarvis et al. (2020); “*Multi-city modeling of epidemics using spatial networks: Application to 2019-nCov (COVID-19) coronavirus in India*”, Pujari (2020). Assim como os artigos de Horby et al. (2011), Grijalva et al. (2015), Keplac et al. (2020), Chin et al. (2021).

O estudo de Mossong, J. et al. (2008) teve por propósito identificar os contatos dos indivíduos de 8 países da Europa (Bélgica, Alemanha, Finlândia, Grã-Bretanha, Itália,

<sup>1</sup>Disponível em: [https://qsprod.saude.gov.br/extensions/covid-19\\_html/covid-19\\_html.html](https://qsprod.saude.gov.br/extensions/covid-19_html/covid-19_html.html). Acesso em: 07 de fevereiro de 2022

<sup>2</sup>Disponível em: <https://coronavirus.saude.mg.gov.br/cidadao> Acesso em: 07 de fevereiro de 2022

Luxemburgo, Holanda e Polônia) diretamente. Este estudo contou com 7.290 participantes e foram coletados, a partir de um diário epidemiológico, 97.904 contatos entre maio de 2005 e setembro de 2006. Observou-se os padrões de mistura e as características do contato atentando-se as diferenças por estratos como idades, sexo e círculos sociais, devido a importância destes dados para estudos e modelagens de doenças infecciosas pelas vias respiratórias.

Em Prem et al. (2017) os dados obtidos pelo POLYMOD (Mossong et al., 2008) foram sintetizados com diversas fontes de dados: Pesquisas Domiciliares Demográficas (DHS) em grande escala, da divisão da população da ONU e de vários indicadores internacionais; com o intuito de, a partir das estruturas domiciliares, taxas de participação escolar e força do trabalho, projetar os contatos para países de diferentes perfis demográficos que não apresentam estes dados empíricos. Foram projetadas taxas de contatos específicas por idade para 152 países (95,9% da população mundial) considerando as faixas etárias e diferentes círculos sociais (Casa, Trabalho, Escola e Outros). Prem et al.(2020) atualiza as matrizes de contato de Prem et al.(2017), estendendo a projeção para 177 países (97,2% da população mundial), acrescentando a estratificação por áreas rurais e urbanas. Ademais, foram comparados os efeitos do uso das matrizes de contato empíricas e sintéticas ao modelar intervenções de distanciamento físico para a pandemia COVID-19.

Na Pesquisa CoMix (Jarvis et al., 2020) há um acompanhamento semelhante ao POLYMOD de Mossong et al.(2008). Com o intuito de avaliar o impacto das medidas de distanciamento social na disseminação do coronavírus no Reino Unido, através da estimação do número de reprodução básico ( $R_0$ ), foi realizada uma pesquisa de padrões de contato conforme as medidas de distância física. Foram aplicados questionários, a cada duas semanas, a uma amostra representativa da população do Reino Unido e, os respondentes foram abordados sobre seus contatos e suas atitudes no período pandêmico a fim de rastrear o  $R_0$  e os contatos médios ao longo do tempo, explorando as diferenças entre as regiões.

Por fim, Pujari (2020) propõe uma abordagem híbrida para o modelo SIR com populações bem misturadas dentro da cidade e acoplamento intermunicipal com base na rede de transporte para investigar a disseminação da COVID-19 na Índia. Nesta análise, um modelo SIR é utilizado para avaliar a dinâmica da doença dentre cidades, considerando uma mistura intra-cidade. Alternativamente, a migração intermunicipal é modelada por meio de redes espacialmente localizadas a partir do tráfego entre aeroportos e estações de trem.

## 1.1 Motivação e Objetivos

O presente estudo faz parte do Projeto: “COVID-19: proposta de um modelo epidemiológico que incorpora estruturas de contatos sociais”, este realizado em colaboração

entre o Departamento de Estatística-UFMG (Universidade Federal de Minas Gerais), GI-ARS (Grupo Interdisciplinar de Pesquisa em Análise de Redes Sociais-Departamento de Sociologia da FAFICH-UFMG), o Grupo de Modelagem Complexa da Faculdade de Engenharia e Ciências Físicas e Matemática da Universidade Central do Equador. O Projeto foi aprovado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) na chamada MCTIC/CNPq/ FNDCT/MS/ SCTIE/Decit No 07/2020 – Pesquisas para enfrentamento da COVID-19, suas consequências e outras síndromes respiratórias agudas graves.

O objetivo então, é construir, sob uma abordagem Bayesiana, modelos hierárquicos para estimar as taxas de contatos dos indivíduos do Aglomerado da Serra a fim de projetá-las para as regiões da cidade de Belo Horizonte e aplicá-las em modelos compartimentais como parte dos estudos para mitigar os impactos da COVID-19.

A estimação das taxas por modelos hierárquicos foi a metodologia adotada pelo embasamento do Projeto no artigo de Prem et al.(2017). Este método considera a estrutura de agrupamento no conjunto de dados e explora a relação existente entre os níveis individuais e de grupo, aqui retratada como os indivíduos que participaram da pesquisa e a faixa etária a qual pertencem.

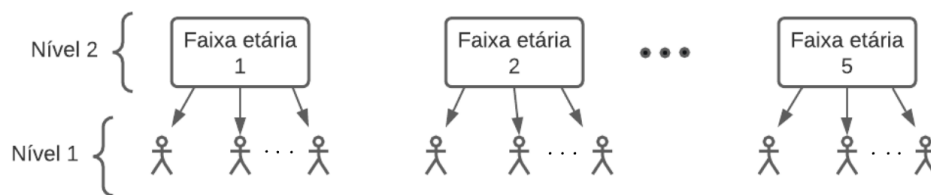


Figura 1: Estrutura do modelo hierárquico.

Esta dissertação está dividida em duas seções. A primeira seção abrange os capítulos 2 e 3, e aborda, de forma resumida, as principais questões das metodologias adotadas neste ensaio. A segunda seção é referente a aplicação destas metodologias e está organizada em 4 capítulos: o capítulo 4 que apresenta a descrição dos dados em estudo, os capítulos 5 e 6 as aplicações e análises realizadas; e o capítulo 7, que desenvolve as conclusões da pesquisa.

## 2 Modelos Lineares Hierárquicos

Os Modelos Lineares Hierárquicos (HLM), por Bryk e Raudenbush (1992) ou Modelos de Regressão Multiníveis, por Hox (1995) e Goldstein (1999); chamados também de modelos de efeitos mistos, são modelos utilizados em estudos dos conjuntos de dados aninhados em uma estrutura hierárquica. Este método é utilizado quando se busca analisar os efeitos de medidas repetidas (estudos longitudinais) ou os efeitos simultâneos de variáveis a níveis individuais e em grupos (Snidjers, 2016), sendo o *nível 1*, ou *nível mais baixo*, o nível mais detalhado, ou seja, o HLM explora as características individuais, assim como as referentes ao contexto em que o indivíduo observado se encontra. Um exemplo desta aplicação é em estudos educacionais, em que os estudantes (nível 1) estão aninhados em classes (nível 2), que estão aninhadas em escolas (nível 3).

Nos modelos de regressão usuais, assumimos a suposição da independência das observações, entretanto, em situações em que os dados são aninhados, não podemos assumir esse pressuposto visto que podem levar ao erro das estimativas (Hox, 2018). Logo, os modelos multiníveis, por considerarem a dependência entre as variáveis explicativas, são uma alternativa. Outrossim, podem ser descritos conceitualmente como “um sistema hierárquico de equações de regressão”, Hox (2018).

Os modelos hierárquicos podem apresentar vários níveis, sendo comumente mais utilizado os modelos de dois e três níveis. As subseções a seguir apresentam a estrutura do HLM de dois níveis para exemplificação, e algumas das características dessa metodologia sob a abordagem Bayesiana. Mais detalhes podem ser vistos em Hox(2018), Snidjers (2016), Migon (2008) e Gelman & Hill (2007).

### 2.1 Modelo de Dois Níveis

O modelo de dois níveis pode ser escrito como:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \quad (1)$$

em que

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad (3)$$

com  $i = 1, 2, \dots, n_j$  e  $j = 1, 2, \dots, m$ , onde:

- $Y_{ij}$ : variável resposta do  $i$ -ésimo indivíduo do nível 1, do  $j$ -ésimo grupo;
- $\beta_{0j}$ : intercepto para o  $j$ -ésimo grupo;



- $\beta_{1j}$ : inclinação da variável explicativa  $X_{ij}$ , do  $i$ -ésimo indivíduo do nível 1 para o  $j$ -ésimo grupo do nível 2;
- $X_{ij}$ : variável explicativa do  $i$ -ésimo indivíduo do nível 1, agrupadas ao  $j$ -ésimo grupo do nível 2;
- $\varepsilon_{ij}$ : erro aleatório associado ao  $i$ -ésimo indivíduo do nível 1, do  $j$ -ésimo grupo do nível 2. Com as suposições de distribuição  $N(0, \sigma_\varepsilon^2)$  e independência entre os erros.
- $\gamma_{00}$ : valor esperado dos interceptos no nível 2;
- $\gamma_{10}$ : valor esperado das inclinações no nível 2;
- $u_{0j}$ : efeito aleatório no intercepto  $\beta_{0j}$  do  $j$ -ésimo elemento do nível 2;
- $u_{1j}$ : efeito aleatório na inclinação  $\beta_{1j}$  do  $j$ -ésimo elemento do nível 2.

Com suposições de:  $u_{0j} \sim N(0, \tau_{00})$ ;  $u_{1j} \sim N(0, \tau_{11})$ ; todos os  $u_{0j}$  e  $u_{1j}$  são independentes entre si e de seus  $\varepsilon_{ij}$ . No qual:

- $\tau_{00}$ : variância populacional dos interceptos;
- $\tau_{11}$ : variância populacional das inclinações;
- $\tau_{01}$ : covariância entre  $\beta_{0j}$  e  $\beta_{1j}$ .

Em resumo:  $\beta_{0j} \sim N(\gamma_{00}, \tau_{00})$ ,  $\beta_{1j} \sim N(\gamma_{10}, \tau_{11})$  e  $Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}$ .

Nesse modelo, o intercepto ( $\beta_{0j}$ ) e o coeficiente de inclinação ( $\beta_{1j}$ ) são considerados coeficientes aleatórios. E pode-se dizer que, para o nível 2, teremos  $m$  modelos idênticos à equação (1).

Para calcular o quanto da variância total é explicada pela variabilidade de grupo, podemos calcular a correlação intra-classe (ICC). O ICC é a medida atribuída à proporção da variância total que é explicada pela variância em nível de grupo (Hox, 2017). É dada pela seguinte equação:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma_\varepsilon^2}, \quad (4)$$

onde  $\tau_{00}$  é a variância dos resíduos no nível 2 e  $\sigma_\varepsilon^2$  é a variância dos resíduos no nível 1. A soma das duas variações implica na variância total.

Se o valor do ICC é próximo de zero, isso significa que dentro do mesmo grupo, as observações a níveis individuais são heterogêneas; logo, se assumirmos o pressuposto de independência das observações, podemos utilizar um modelo de regressão simples para a análise.

## 2.2 Variações do modelo

Os HLM são usualmente utilizados em quatro variações: modelo de coeficiente aleatório, modelo com intercepto aleatório, modelos com inclinação aleatória e modelo com intercepto e inclinações aleatórios. Os três modelos são descritos abaixo, resumidamente.

### 2.2.1 Modelo de coeficientes aleatórios

No modelo de coeficientes aleatórios, ou modelo nulo, é composto apenas pelo intercepto e dois efeitos aleatórios, um para a diferença entre o nível dos indivíduos e outro para a diferença em nível de grupo (Rocha, 2017). É dado por:

$$Y_{ij} = \alpha + u_{0j} + \varepsilon_{ij}. \quad (5)$$

É utilizado para estimar os componentes de variância e, assim, calcular o coeficiente de correlação intra-classe.

### 2.2.2 Modelo com intercepto aleatório

Nos modelos com intercepto aleatório, temos um intercepto para cada grupo. Podem ser escritos da seguinte forma:

$$Y_{ij} = \alpha_{ij} + \beta X_{ij} + \varepsilon_{ij}. \quad (6)$$

### 2.2.3 Modelo com inclinação aleatória

Este modelo é composto por um intercepto global e uma inclinação para cada coeficiente a ser estimado por cada grupo. Sua equação é dada por:

$$Y_{ij} = \alpha + \beta_{ij} X_{ij} + \varepsilon_{ij}. \quad (7)$$

### 2.2.4 Modelo com intercepto e inclinações aleatórios

São os modelos de efeitos mistos, os que apresentam conjuntamente os efeitos fixos e aleatórios. Podendo ser escrito como:

$$Y_{ij} = \alpha_{ij} + \beta_{ij} X_{ij} + \varepsilon_{ij}. \quad (8)$$

## 2.3 Abordagem Bayesiana

A Inferência Bayesiana assume uma postura subjetivista (de Finetti, 1931), ou seja, assume que devemos estabelecer uma medida de probabilidade sobre a incerteza pessoal para tudo que seja desconhecido. Parte do pressuposto da permutabilidade (de Finetti,

1974) e segue o Princípio da Coerência (de Finetti, 1931; Loschi e Wechsler, 2002). Nessa abordagem, a melhor inferência sobre um parâmetro  $\theta$  é a distribuição a *posteriori*, e esta é proveniente da ponderação entre a informação dos dados amostrais (unicamente da amostra observada) e do conhecimento prévio ou referenciais teóricos que se tem do evento em estudo. Essas, resumidas pela função de verossimilhança ( $f(x|\theta)$ ) e distribuição a *priori* ( $\pi(\theta)$ ), respectivamente.

A distribuição a *posteriori* pode ser entendida como uma atualização da distribuição a *priori* e é oriunda da junção da *priori* com a função de verossimilhança, que tem como base, o Teorema de Bayes. Pode ser escrita como:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \quad (9)$$

Para maior aprofundamento nos conceitos de inferência com enfoque Bayesiano, verificar Migon et al.(2008), Migon et al.(2015) e Murteira et al. (2018).

### 2.3.1 Modelos mistos

Na abordagem Bayesiana, assumimos que os efeitos fixos se comportam de forma aleatória, e assim, devemos especificar uma distribuição a priori para cada parâmetro que se quer estimar. Desta forma, devido a estrutura hierárquica do conjunto de dados, as *prioris* dos hiperparâmetros (parâmetros dos níveis mais altos), chamadas de hiperprioris, são utilizados para inferir sobre a *posteriori* dos parâmetros de níveis mais baixos, o que auxilia na melhor estimação e ajuste do modelo. Para mais detalhes ver Gelman & Hill (2007) e Migon et al.(2008).

### 3 Modelagem Matemática em Epidemiologia

Modelos matemáticos são utilizados em epidemiologia com o propósito de descrever quantitativamente a propagação de doenças e fornecer informações epidemiológicas e dados estatísticos sobre os parâmetros estimados referentes a força da infecção, número de reprodutividade basal e taxas de contato (Sabeti, 2011), por exemplo. Existem diversos modelos empregados em epidemiologia matemática, como os modelos compartimentais, que podem apresentar estrutura determinística ou estocástica; modelos de rede; modelos estocásticos e modelos baseados em agentes (López-Flores et al., 2021). Nesta dissertação abordaremos apenas os modelos compartimentais, mais especificamente o modelo SIR.

#### 3.1 Modelos Compartimentais: Modelo SIR

Os modelos compartimentais são expressos por meio de sistemas de equações diferenciais, podendo ser empregados para modelar a propagação de doenças infecciosas como a gripe e a COVID-19 (López-Flores et al., 2021). Nestes modelos, a população é dividida em *compartimentos* (categorias), estas que dependem das características dos indivíduos relevantes à infecção sob estudo (Mercado Londoño, 2014) e são analisadas a evolução das frações de população nos compartimentos ao longo do tempo (Equações Diferenciais Ordinárias - EDO) ou tempo e espaço (Equações Diferenciais Parciais - EDP) (López-Flores et al., 2021).

Entre os modelos compartimentais, o modelo SIR (Susceptible-Infected-Removed), também chamado modelo epidêmico clássico (Hethcote, 2000), é um dos mais simples e mais utilizados para a análise de doenças infecciosas (Alvarenga, 2008). Foi proposto por Kendrick e McCormick em 1927 para o estudo da propagação de doenças transmissíveis e consiste na subdivisão da população em três compartimentos:

- Suscetíveis ( $S$ ): indivíduos que podem se infectar se forem expostos ao agente infeccioso;
- Infectados ( $I$ ): indivíduos que estão infectados e podem transmitir a doença;
- Recuperados ( $R$ ): indivíduos que contraíram a doença mas já estão recuperados e não podem mais ser infectados;

em que  $S(t)$ ,  $I(t)$  e  $R(t)$ , são frações das respectivas classes em função do tempo  $t$ . Em casos que a doença em estudo seja fatal apenas para uma pequena parcela dos infectados, as fatalidades são contabilizadas no compartimento  $R$ , que passa a ser nomeado como “Removidos” (López-Flores et al., 2021). Existe também o modelo SIR com dinâmica vital (considera os nascimentos e as mortes), o modelo endêmico clássico, para mais detalhes, verificar Hethcote (2000).



Figura 2: Estados modelo SIR.

A Figura 2 indica o esquema do sistema de compartimentos e as taxas de transição entre eles.

Neste modelo, a população se mantém constante, portanto não há nascimentos e mortes, somente as ocasionadas pela doença; e cada indivíduo infectado tem a mesma probabilidade de encontrar um suscetível (Bassanezi, 2002). Segundo Keeling e Rohani (2007), a progressão do indivíduo do compartimento S para o I é definido pela prevalência dos infectados, a estrutura de contato da população subjacente e a probabilidade de transmissão devido ao contato, ou seja, para a disseminação do patógeno, um indivíduo suscetível deve entrar em contato com um infectado, e a probabilidade disso ocorrer é dada pelos níveis de S e I e pela taxa de contágio. E os indivíduos passam do compartimento S para I, quando se recuperam da doença (Keeling e Rohani, 2007).

O modelo é homogêneo, independe do tempo e é descrito como:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \quad (10)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \quad (11)$$

$$\frac{dR}{dt} = \gamma I, \quad (12)$$

em que  $\beta$  é uma taxa de contágio diária e pode ser dada por:

$\beta = \text{taxa de contatos dos indivíduos} * \text{probabilidade de transmissão},$

$N$  é o tamanho da população e  $1/\gamma$  é o tempo médio em que os indivíduos estão infectados,  $\beta > 0, \gamma > 0$  e

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad (13)$$

$$\mathbf{S}(t) + \mathbf{I}(t) + \mathbf{R}(t) = N. \quad (14)$$

Para a análise da propagação da doença é necessário saber os valores iniciais das frações nos compartimentos, ou seja  $S(0)$ ,  $I(0)$  e  $R(0)$  e dos parâmetros  $\beta$ ,  $\gamma$  e  $R_0$  (Bassanezi, 2002).

### 3.1.1 Número de Reprodutibilidade Basal ( $R_0$ )

O número de reprodutibilidade basal (Sabeti, 2011) ou número de reprodução básico (López-Flores et al., 2021) é o número esperado de casos secundários produzidos por um indivíduo infectado (caso primário) durante todo o período infeccioso na situação de uma população totalmente suscetível (Diekmann et al., 1990; Diekmann and Heesterbeek 2000 apud Sabeti, 2011). O  $R_0$  é fundamental na área da epidemiologia, uma vez que mede a velocidade inicial do crescimento de uma epidemia e sendo usado, principalmente, para o cálculo da proporção da vacinação mínima essencial para a erradicação da doença (Sabeti, 2011). Tem-se,

$$R_0 = \beta * \frac{1}{\gamma} = \frac{\beta}{\gamma}. \quad (15)$$

Se o  $R_0 > 1$ , o número de indivíduos infectados cresce e propende a disseminação do agente infeccioso. Todavia, se  $R_0 < 1$ , o número de indivíduos infectados decresce e a doença tende a desaparecer. Se o  $R_0 = 1$ , diz-se que a doença encontra-se em equilíbrio epidemiológico, com um número constante de infectados por dia, ou seja, taxa de incidência constante. A estimacão do  $R_0$  é extremamente importante para o desenrolar de epidemias e criação de diretrizes de saúde pública, uma vez que é uma das medidas responsáveis na quantificação da transmissibilidade do patógeno (Cori et al., 2013).

### 3.1.2 Número de Reprodução instantâneo ( $R_t$ )

O número de reprodução instantâneo é o número esperado de infecções secundárias produzidas por um indivíduo infectado (caso primário) no momento  $t$  (Wallinga, Lipsitch, 2006). Pode ser calculado de diferentes modos, como por exemplo em função do  $R_0$  (Bettencout; Ribeiro, 2006):

$$R_t = \frac{S_t}{N} R_0, \quad (16)$$

em que  $S_t$  é o número de indivíduos suscetíveis na população no instante  $t$ .

Em função da relação entre os casos primários e secundários, utilizada nas áreas da Ecologia e na Demografia (Wallinga, Lipsitch, 2006):

$$R_t = \exp^{rT} \quad (17)$$

em que,

$$r = \frac{\ln \left[ \frac{\text{Infectados}}{\text{Infectantes}} \right]}{T}, \quad (18)$$

e  $T$  é o tempo médio de infecção.

E pelo método “Imperial College” (Cori et al., 2013; Thompson et al., 2019):

$$R_t = \frac{I_t}{\sum_1^a (I_{t-a} g_{t-a})}. \quad (19)$$

Em que a função  $g_{t-a}$  informa a probabilidade de geração de infecção dos casos em  $a$  dias (como exemplo de unidade de tempo) antes de  $t$  e  $I_t$  é o número de infectados no instante  $t$ . Considera-se que os indivíduos infectados no tempo  $t$ , em média, foram gerados pelos casos infectados em  $t - a$ .

## 4 Descrição dos dados

Esta dissertação tem por interesse analisar os níveis de propagação da COVID-19 nas regiões de Belo Horizonte através de um modelo SIR. Para isso, foram coletados os dados de padrão de mistura de uma amostra do Aglomerado da Serra, periferia na região centro-sul de Belo Horizonte, para que sejam estimadas as taxas de contato da região, por meio de um modelo hierárquico e, posteriormente, estas sejam projetadas para as demais regiões da cidade (Prem et al. 2017).

Para a análise, utilizou-se dois bancos de dados provenientes de questionários aplicados na região do Aglomerado. Estes questionários foram elaborados pelo Grupo Interdisciplinar de Pesquisa em Análise de Redes Sociais (GIARS), do Departamento de Sociologia da Faculdade de Filosofia e Ciências Humanas (FAFICH) da UFMG, fazem parte da Pesquisa “COVID-19: proposta de um modelo epidemiológico que incorpora estruturas de contatos sociais”, Parecer 4.649.110 e aprovado pelo comitê de ética em 14 de Abril de 2021.

### 4.1 Banco 1

O primeiro questionário foi aplicado a uma amostra de 1001 residentes do Aglomerado da Serra entre 04 e 10 de junho de 2021. A partir dele foram obtidos os dados socioeconômicos, a estrutura domiciliar e as situações de contatos dos entrevistados, entre outras variáveis para os demais estudos do Projeto.

Participaram da pesquisa 613 (61,24%) habitantes do sexo feminino e 388 (38,76%) do sexo masculino, distribuídos em 8 bairros: Fazendinha, Marçola, Nossa Senhora de Fátima, Nossa Senhora Aparecida, Nossa Senhora do Rosário, Nossa Senhora da Conceição, Novo São Lucas e Santana do Cafezal. A Figura 3 indica a frequência do número de entrevistados por bairro

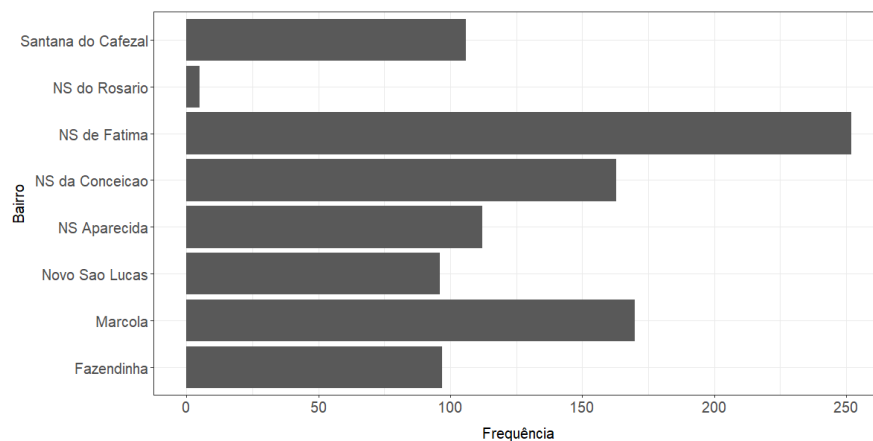


Figura 3: Frequência do número de entrevistados por bairro - Banco 1.



Nota-se que a maior parcela dos respondentes moram no bairro Nossa Senhora de Fátima, seguidos por Marçola e Nossa Senhora da Conceição. Destaca-se também, a partir da Figura 4, que os domicílios tendem a ter entre 2 e 4 moradores, média de 3,36 moradores por domicílio.

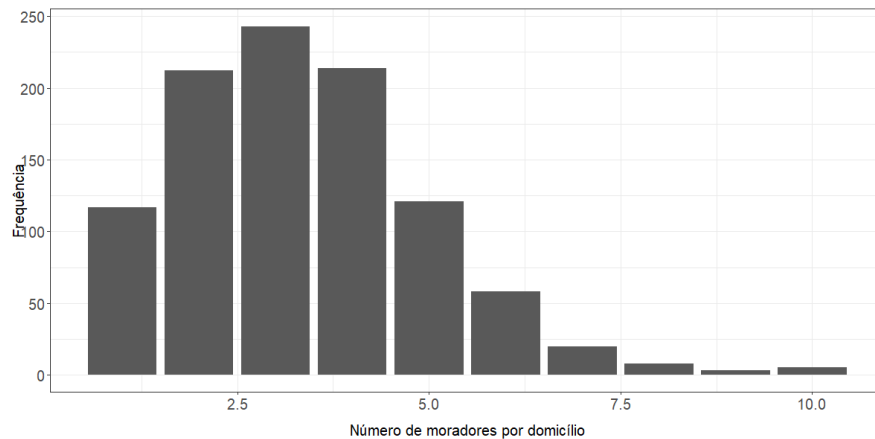


Figura 4: Frequência do número de moradores por domicílio - Banco 1.

Do total de entrevistados, as informações abaixo são referentes aos 995 respondentes que especificaram o sexo e a idade das pessoas com quem obtiveram contato, além do local em que este contato ocorreu. Desta amostra, 609 (61,21%) respondentes são do sexo feminino e 386 (38,79%) do sexo masculino e estão distribuídos em 5 faixas etárias:

- Faixa etária 1: 0 a 14 anos;
- Faixa etária 2: 15 a 19 anos;
- Faixa etária 3: 20 a 34 anos;
- Faixa etária 4: 35 a 59 anos;
- Faixa etária 5: +60 anos.

Obs.: Os grupos de idade foram definidos pelo GIARS, de acordo com suas metodologias de pesquisa.

Tabela 1: Total de entrevistados por faixa etária - Banco 1.

Faixa etária	N <sup>o</sup> de entrevistados	%
F1	147	14,7
F2	60	6,03
F3	222	22,31
F4	348	34,97
F5	218	21,91
Total	995	100

Percebe-se a partir da Tabela 1 que o maior percentual de entrevistados se encontra na faixa etária 4 (34,97%), seguida pela faixa etária 3 (22,31%), enquanto a faixa etária 2 apresenta o menor percentual, apenas 6,03%. A Figura 5 apresenta ainda as distribuições dos respondentes por faixa etária e sexo.

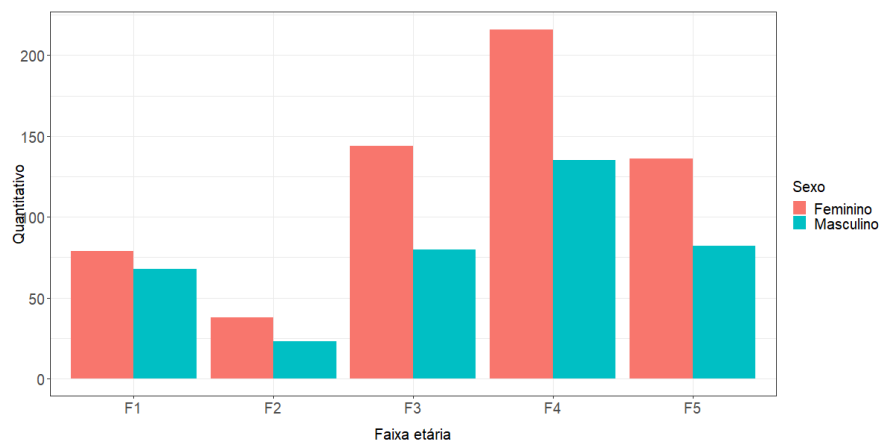


Figura 5: Número de entrevistados por faixa etária e sexo - Banco 1.

Observa-se-se que o percentual dos indivíduos que se declaram mulheres é maior em todas as faixas etárias, apresentando uma diferença mais acentuada na proporção dos sexos nas faixas etárias 3, 4 e 5.

- Variável resposta: Contatos

Os 1001 entrevistados relataram um total de 3539 contatos, entretanto para a análise descritiva a seguir, serão considerados os 3519 contatos pertencentes aos 995 respondentes selecionados. Contabilizou-se qualquer situação de contato realizado nas últimas 24 horas à entrevista, sejam conversas cara a cara ou contatos físicos (apertos de mão, abraços, beijos ou em práticas esportivas, por exemplo).

A Figura 6 apresenta a frequência dos contatos e nota-se que a grande parte dos entrevistados indicaram ter tido contatos com entre 1 e 4 pessoas (média de 3,54). Poucos

respondentes relataram não terem tido contato algum ou contatos com mais de 11 pessoas, não considerando os contatos pertinentes ao atendimento ao público.

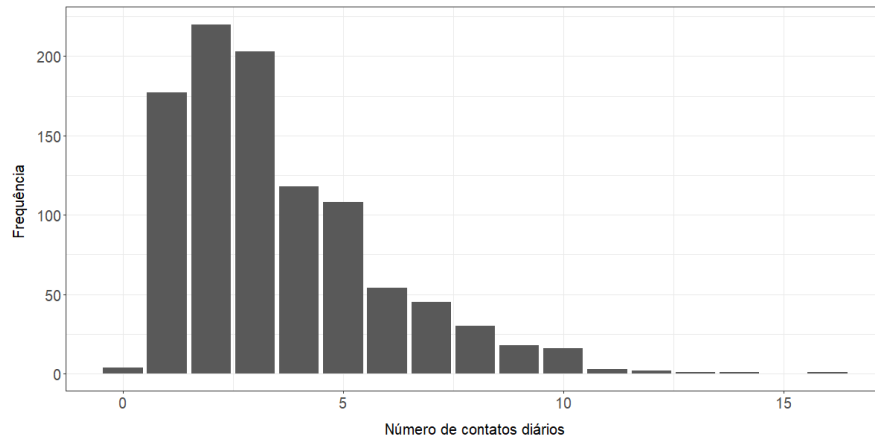


Figura 6: Frequência do número de contatos - Banco 1.

As situações de contato ocorreram em 8 diferentes locais: casa, escola, trabalho, igreja, velório, vizinhança, lazer e comércio/serviço. Entretanto, visto que o número de contatos realizados fora do contexto ‘Casa’ é pequeno para cada um dos demais locais, visando uma melhor análise, agrupamos as demais categorias em uma só: ‘Outros’, que será referente a todas as situações de contato que ocorreram fora do domicílio em que o respondente é residente.

A Tabela 2 apresenta o número de contatos por círculo social. Nota-se que o maior percentual dos contatos foram feitos no contexto ‘Casa’, 67,12%, isso provavelmente se deve ao período de coleta do banco de dados, período de distanciamento social, com o fechamento de escolas, trabalhos em *home office*, entre outros. Já os demais grupos, na categoria ‘Outros’, apresentam juntos os 32,88% contatos restantes. Podemos observar a distribuição dos contatos nos diferentes contextos na Figura 7.

Tabela 2: Total de contatos por círculo social - Banco 1.

Círculo social	N <sup>o</sup> de contatos	%
Casa	2362	67,12
Outros	1157	32,88
Total	3519	100

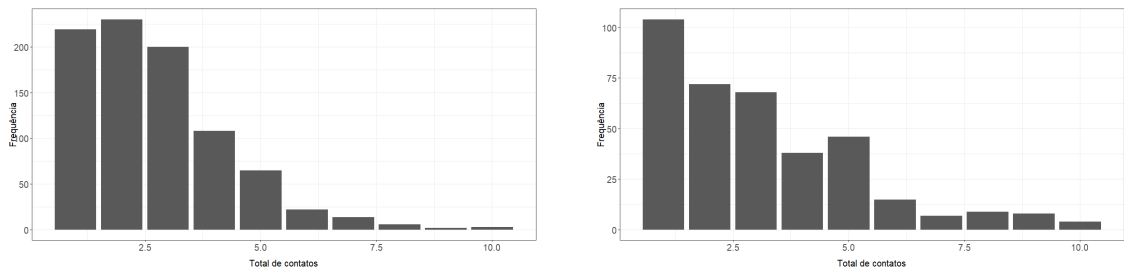


Figura 7: Distribuição dos contatos nos círculos ‘Casa’ e ‘Outros’, respectivamente - Banco 1.

A tabela 4 indica os contatos realizados por sexo. Aproximadamente 31% dos contatos foram realizados entre mulheres e 30% entre os homens, para mais, temos que 51,59% dos contatos foram realizados com alguém do sexo feminino e 48,51% com alguém do sexo masculino.

Tabela 3: Total de contatos por sexo - Banco 1.

Sexo (entrevistado)	Feminino*	Masculino*	Total
Feminino	1094	1054	2148
Masculino	718	653	1371
Total	1812	1707	3519

Obs.: “Feminino\*” e “Masculino\*” se referem ao sexo das pessoas com quem o entrevistado declarou ter o contato.

A tabela a seguir indica o quantitativo de contatos por faixas etárias e círculos sociais.

Tabela 4: Total de contatos por círculo social e faixas etárias - Banco 1.

Círculo social	Faixa etária do respondente	$f1^*$	$f2^*$	$f3^*$	$f4^*$	$f5^*$	Total
Casa	F1	136	42	166	119	24	487
	F2	35	20	26	70	9	160
	F3	183	35	151	132	33	534
	F4	160	103	239	203	86	791
	F5	54	30	93	120	93	390
Total(Casa)		568	230	675	644	245	2362
Outros	F1	44	11	31	37	11	134
	F2	18	22	14	18	2	74
	F3	41	32	193	88	12	366
	F4	34	18	132	173	48	405
	F5	15	6	48	72	37	178
Total(Outros)		152	89	418	388	110	1157
Total Geral		720	319	1093	1032	355	3519

Obs.: “ $f1^*$ ”, “ $f2^*$ ”, “ $f3^*$ ”, “ $f4^*$ ” e “ $f5^*$ ” se referem as faixas etárias das pessoas com quem o entrevistado declarou ter contato.

Observa-se que o maior número de entrevistados é da faixa etária 4, enquanto a faixa etária 2 apresenta o menor quantitativo. Temos também que a idade média dos respondentes foi de 39,88 anos (min = 1<sup>3</sup>, max = 90).

## 4.2 Banco 2

O segundo questionário contém um delineamento sociodemográfico e, foi aplicado de 26 a 28 de junho de 2021, em virtude da não realização do censo demográfico nos últimos anos, com o último sucedido no ano de 2010 e, por conseguinte, a desatualização do perfil dos habitantes da região. Este instrumento foi adotado com o intuito de viabilizar uma amostra final representativa da população alvo a partir da calibração (Chin et al., 2021, Ruiz e Silva, 2014) dos dados do Banco 1, uma vez que existe um vício das respostas para os dados sociodemográficos (Chin et al., 2021).

Foram coletas informações de 1366 pessoas distribuídas em 450 domicílios do Aglomerado da Serra. A composição da amostra é de 50,51% pessoas do sexo feminino e 49,49% do sexo masculino. A Tabela 5 apresenta a proporção dos moradores por faixa etária e a Figura 8 indica a distribuição dos indivíduos da amostra por faixa etária e sexo.

<sup>3</sup>Indivíduos menores de idade (< 18 anos), responderam ao questionário na presença de um responsável, ou com o auxílio do responsável presente.

Tabela 5: Proporção da estrutura etária dos residentes - Banco 2

Faixa etária	Número absoluto	%
F1	255	18,67
F2	104	7,61
F3	392	28,70
F4	449	32,87
F5	166	12,15
Total	1366	100

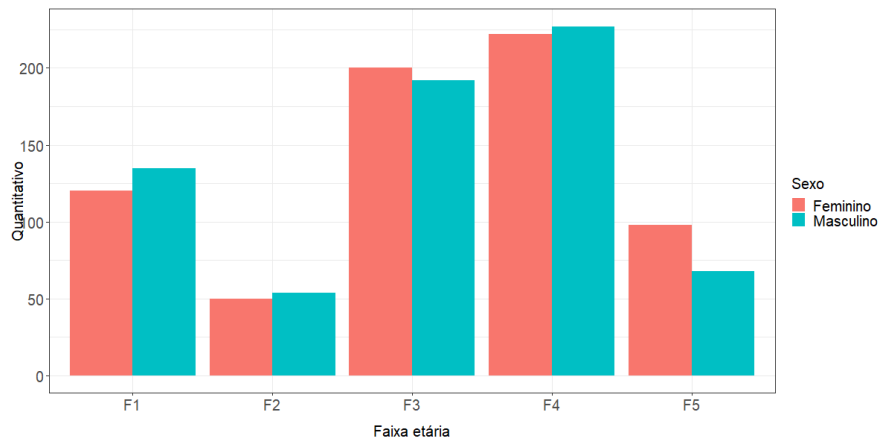


Figura 8: Número de entrevistados por faixa etária e sexo - Banco 2.

Com base nas tabela e figura anteriores, podemos citar que o maior percentual dos residentes se concentram nas faixas etárias 4 e 3, sendo que juntas representam 61,57% da amostra. Notamos também que não há diferenças discrepantes entre as proporções dos sexos nas 5 faixas etárias e que, enquanto as mulheres apresentam maiores proporções nas faixas 3 e 5, os declarados homens são mais representados nas faixas etárias 1, 2 e 4. Além disso, temos que a idade média dos habitantes é de 33,70 anos (min = 0, máx = 102).

A Figura 9 indica a frequência do número de domicílios por bairro.

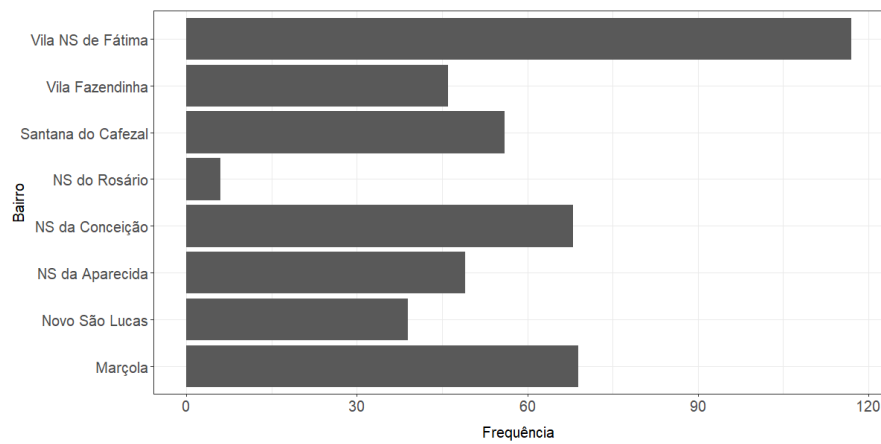


Figura 9: Frequência do número de domicílios por bairro - Banco 2.

Observa-se que a maior parcela dos domicílios se encontram na Vila Nossa Senhora de Fátima, seguidos por Marçola e Nossa Senhora da Conceição. Destaca-se também, a partir da Figura 10, que os domicílios tendem a ser formados por 2 ou 3 moradores (média de 3,04), com poucas residências apresentando mais de 7 pessoas.

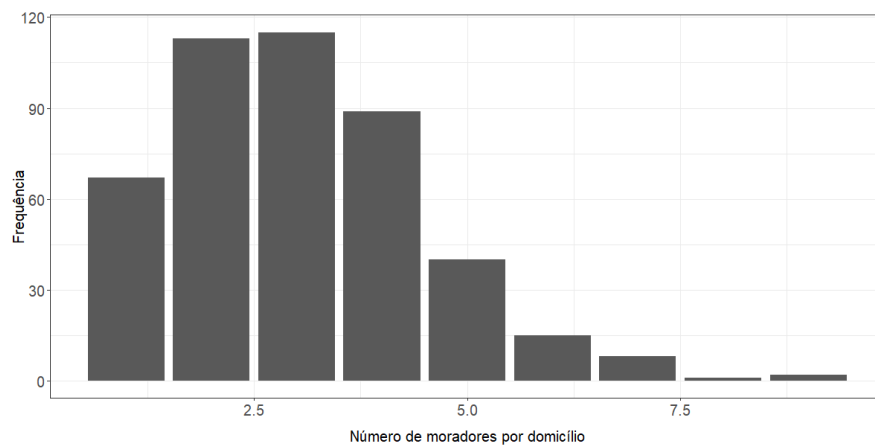


Figura 10: Frequência do número de moradores por domicílio - Banco 2.

### 4.3 Banco calibrado

Métodos de calibração e pós-estratificação são comumente utilizados para solucionar problemas de não representatividade dos dados e amostragem complexa (Chin et al, 2021; Lumley, T., 2004, 2010; Park et al, 2017; Gelman, A. e T. C. Little, 1997). O banco final para a análise é oriundo da calibração dos dados do Banco 1, por meio de pesos gerados a partir do banco 2, de forma similar à do artigo de Chin et al. (2021). Esses pesos foram calculados a nível do respondente, com o intuito de tornar a amostra coletada representativa utilizando *raking*, um método de ajuste proporcional iterativo (Deville and Särndal, 1992; Feehan and Mahmud, 2021). Para isso, utilizou-se o pacote `autumn`<sup>4</sup>, desenvolvido por Aaron Rudkin (2020).

A calibração foi feita a partir das seguintes covariáveis: sexo e faixa etária do indivíduo, bairro e tamanho do domicílio. Calculamos a proporção de cada uma delas no banco 2 (delineamento sociodemográfico) e utilizamos a função `harvest` para o cálculo dos pesos, em que terão média 1 e o valor máximo de 5. Uma coluna, com os pesos de cada indivíduo respondente foi adicionado ao conjunto de dados do primeiro levantamento (banco 1) e foi realizado uma reamostragem simples<sup>5</sup>, com reposição, de tamanho  $n = 1001$ , para a aquisição da nova amostra, sendo essa amostra calibrada, a utilizada nas análises posteriores. As figuras e tabelas a seguir apresentam a descrição do novo banco, e o Anexo 1 contém a comparação entre todos os conjuntos de dados citados.

Dos 1001 indivíduos da amostra, 50,45% são do sexo feminino e 49,55% do sexo masculino. As Tabela 6 e Figura 11 apresentam a proporção de pessoas por faixa etária, e sua distribuição por sexo, respectivamente.

Tabela 6: Proporção da estrutura etária - Banco calibrado.

Faixa etária	Nº Absoluto	%
F1	183	18,28
F2	86	8,59
F3	282	28,17
F4	306	30,57
F5	144	14,39
Total	1001	100

<sup>4</sup>Para maiores informações e procedimento de instalação do pacote, acessar: [https://github.com/aaronrudkin/autumn/?utm\\_source=pocket\\_mylist](https://github.com/aaronrudkin/autumn/?utm_source=pocket_mylist)

<sup>5</sup>Uma alternativa que permite o cálculo da variância desses estimadores consiste em repetir esse procedimento para a utilização do método Bootstrap, em abordagens Frequentistas e Bayesianas (ver Aitkin, Murray., 2008; Olsson et al., 2021; Rubin, Donald B, 1981; Goldstein et al., 2018).



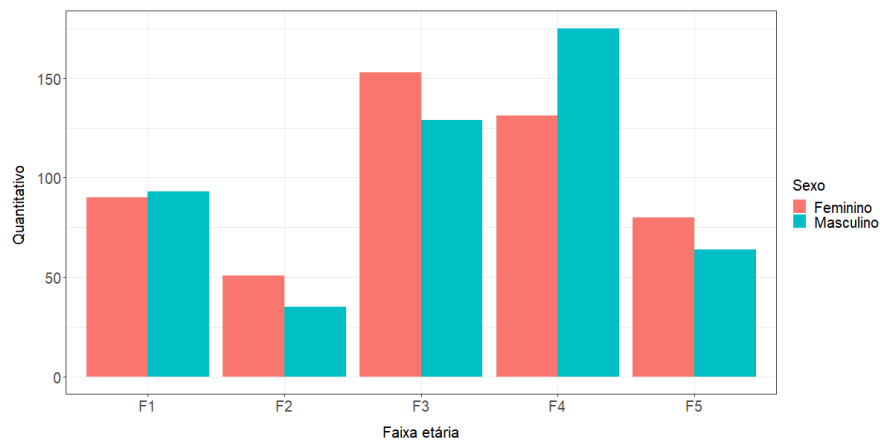


Figura 11: Número de indivíduos por faixa etária e sexo - Banco calibrado.

Nota-se que para ambos os sexos, a maior proporção de pessoas se encontram nas faixas etárias 3 e 4, em contrapartida a faixa 2 apresenta o menor percentual. Além disso, pode-se perceber também que os homens são maioria nas faixas etárias 1 e 4, e as mulheres nas demais. A idade média dos respondentes é de 35,08 anos (min = 1, máx = 90).

A Figura 12 indica a frequência do número de domicílios por bairro.

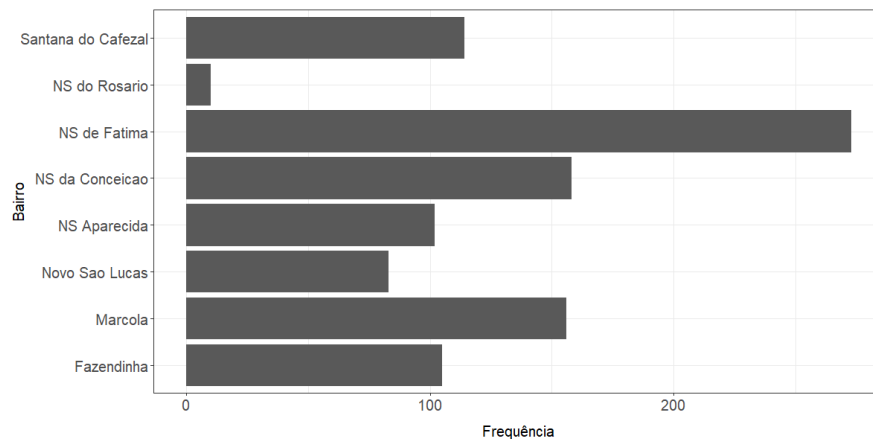


Figura 12: Frequência do número de domicílio por bairro - Banco calibrado.

Observa-se que a maior parcela dos domicílios se encontram na Vila Nossa Senhora de Fátima, seguidos por Nossa Senhora da Conceição e Marçola. Pode-se destacar também, a partir da Figura 13, que os domicílios tendem a ser formados por 2 ou 3 moradores (média de 3,04), com poucas residências apresentando mais de 7 pessoas.

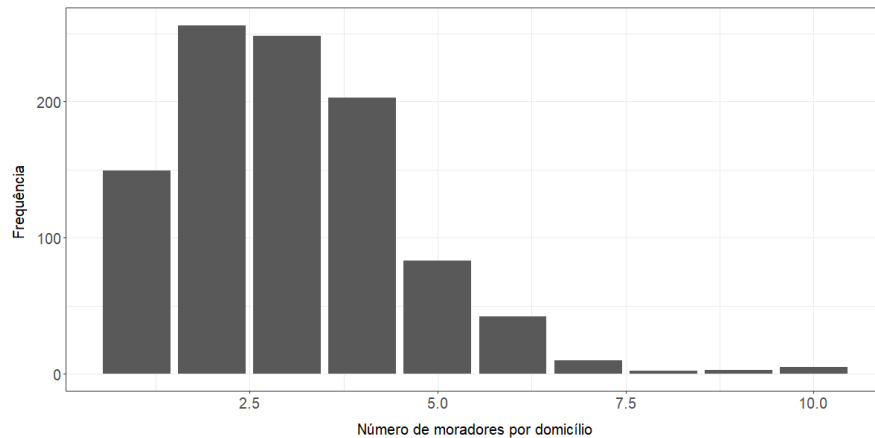


Figura 13: Frequência do número de moradores por domicílio - Banco calibrado.

- Variável resposta: Contatos

No banco calibrado, as 1001 pessoas da amostra relataram um total de 3450 contatos, entretanto para as análises foram considerados os 3435 dos 994 indivíduos que especificaram suas situações de contato por local e, sexo e idade do contato, adequadamente. Desses contatos, 60,79% foram realizados em ‘Casa’ e 39,21% fora do contexto domiciliar (‘Outros’).

Tabela 7: Total de contatos por círculo social - Banco calibrado.

Círculo social	Nº de contatos	%
Casa	2088	60,79
Outros	1347	39,21
Total	3435	100

A Figura 14 apresenta a frequência dos contatos e nota-se que a grande parte dos entrevistados indicaram ter tido contatos com entre 1 e 3 pessoas (média de 3,45, desvio padrão = 2,28). Poucos respondentes relataram não terem tido contato algum ou contatos com mais de 10 pessoas, não considerando os contatos pertinentes ao atendimento ao público.

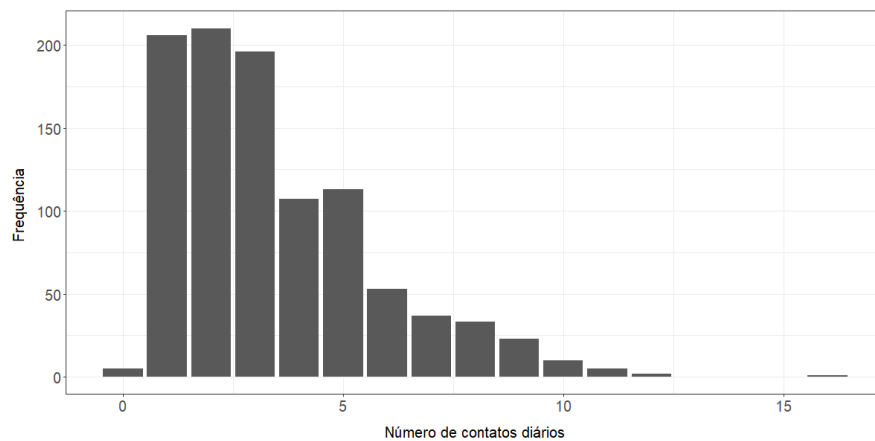


Figura 14: Frequência do número de contatos.

A Tabela 8 retrata os contatos por sexo. Destaca-se que 49,52% dos contatos foram feitos por pessoas que se declaram do sexo feminino, e 50,48% por pessoas do sexo masculino. Destaca-se também que 53,54% dos contatos foram feitos com pessoas do sexo feminino, 46,46% com pessoas do sexo masculino e 50,92% dos contatos foram realizados entre pessoas do mesmo sexo.

Tabela 8: Total de contatos por sexo - Banco calibrado.

Sexo (entrevistado)	Feminino*	Masculino*	Total
Feminino	927	774	1701
Masculino	912	822	1734
Total Geral	1839	1596	3435

Obs.: “Feminino\*” e “Masculino\*” se referem ao sexo das pessoas com quem o entrevistado declarou ter o contato.

A Tabela 9 descreve o total de contatos entre as faixas etárias e por círculos sociais, além do total geral.

Tabela 9: Total de contatos por círculo social e faixas etárias - Banco calibrado.

Círculo social	Faixa etária do respondente	$f1^*$	$f2^*$	$f3^*$	$f4^*$	$f5^*$	Total
Casa	$F1$	131	60	195	129	27	542
	$F2$	54	27	40	81	9	211
	$F3$	199	37	162	161	41	600
	$F4$	106	53	139	180	83	561
	$F5$	20	20	32	53	49	174
Total(Casa)		510	197	568	604	209	2088
Outros	$F1$	62	15	46	62	16	201
	$F2$	17	24	24	31	0	96
	$F3$	42	50	260	149	13	514
	$F4$	40	15	115	159	38	367
	$F5$	10	8	54	65	32	169
Total(Outros)		171	112	499	466	99	1347
Geral	$F1$	193	75	241	191	43	743
	$F2$	71	51	64	112	9	307
	$F3$	241	87	422	310	54	1114
	$F4$	146	68	254	339	121	928
	$F5$	30	28	86	118	81	343
Total (Geral)		681	309	1067	1070	308	3435

Obs.: “ $f1^*$ ”, “ $f2^*$ ”, “ $f3^*$ ”, “ $f4^*$ ” e “ $f5^*$ ” se referem as faixas etárias das pessoas com quem o entrevistado declarou ter contato.

Nota-se que para ambos os círculos e no total geral, as maiores parcelas dos contatos foram realizadas com as faixas etárias 3 e 4. Ademais, tem-se que os respondentes das faixas etárias 3 e 4, foram responsáveis pelos maiores números de contatos nas três situações; enquanto a faixa etária 2 apresentou o menor número nos contextos ‘Outros’ e ‘Geral’ e, a faixa etária 5 no círculo ‘Casa’.

## 5 Aplicação 1: Modelos de contatos

O presente estudo foi construído em três fases. A primeira consistiu na construção de um modelo hierárquico bayesiano para estimar as taxas de contato entre os grupos de idades dos indivíduos do Aglomerado da Serra em diferentes círculos sociais. A segunda fase se refere a projeção das taxas gerais para as regiões de Belo Horizonte (BH) e a terceira consiste na construção de um modelo epidemiológico SIR para as regiões da cidade de BH.

As subseções a seguir deste capítulo descrevem os modelos de contatos em estudo, para os contextos ‘Casa’, ‘Outros’, além do modelo sem distinção de local de contato (‘Geral’). Também descreve a projeção destas últimas taxas para as demais regiões de BH.

### 5.1 Modelo Hierárquico Bayesiano

O modelo hierárquico apresenta como variável resposta o total de contatos, uma variável quantitativa discreta, referente a dados de contagem. Por conseguinte, os modelos serão analisados nas famílias: Poisson, Binomial Negativa e Poisson Inflada de Zeros (ZIP), visto a grande quantidade de zeros no banco de dados devido à sua estrutura no formato longo.

Os modelos em análise são de dois níveis: indivíduos (*nível 1*), agrupados em faixas etárias (*nível 2*). Apresentam como variáveis explicativas a interação entre os grupos de idade do respondente e da pessoa com quem se teve o contato, além da variável referente ao tamanho do domicílio, no contexto do círculo social ‘Casa’. Os modelos foram ajustados utilizando o pacote `brms` (Bürkner, 2017), em 3 cadeias com 4000 iterações cada (2000 de warmup), e, os melhores ajustes foram selecionados pelo critério de Pareto smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO) (Vehtari et al, 2021; Vehtari et al., 2016).

#### 5.1.1 Círculo social: Casa

Para o contexto ‘Casa’ avaliou-se 12 dos modelos ajustados, assumindo diferentes famílias para a variável resposta (Poisson, Poisson Inflada de Zeros e Binomial Negativa) e diferentes prioris. As prioris dos modelos foram estabelecidas de acordo com o conhecimento prévio que temos sobre o conjunto de dados, assim como com prioris pouco informativas.

A tabela a seguir apresenta as famílias e prioris dos modelos analisados, além dos valores `p_loo` (número efetivo de parâmetros) e `elpd_loo` (“estimativa bayesiana LOO da densidade preditiva logarítmica esperada”<sup>6</sup>) desses, para a validação e comparação do ajuste. Os modelos de 1 a 6 foram ajustados com a interação entre a faixa etária do

<sup>6</sup><https://mc-stan.org/loo/reference/loo-glossary.html>, Vehtari, Gelman e Gabry (2017)

respondente e a faixa etária do contato como efeito fixo, e os de 7 a 12 com essa mesma interação mais a inclusão da variável ‘tamanho do domicílio’. Em ambos os grupos de modelos, temos também a adição do efeito aleatório a nível de indivíduo ( $\sigma_i$ ). A Tabela 10 expressa essa comparação.

Tabela 10: Modelos ajustados - Casa.

Modelo	Família	Priori ( $\beta$ 's)	Priori (sd $\sigma_i$ )	p_loo	elpd_loo
casa1	Poisson	$Normal(3, 2^2)$	$Gama(1, 1)$	27.2	-4862.3
casa2	Poisson	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	23.8	-4861.7
casa3	Bin Neg	$Normal(3, 2^2)$	$Gama(1, 1)$	27.5	-4863.6
casa4	Bin Neg	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	23.9	-4862.9
casa5	ZIP	$Normal(3, 2^2)$	$Gama(1, 1)$	27.3	-4863.2
casa6	ZIP	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	23.9	-4862.7
casa7	Poisson	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	24.0	-4798.3?
casa8	Poisson	$Normal(3, 2^2)$	$Gama(1, 1)$	26.5	-4789.5
casa9	Bin Neg	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	24.2	-4789.6
casa10	Bin Neg	$Normal(3, 2^2)$	$Gama(1, 1)$	26.0	-4790.2
casa11	ZIP	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	24.3	-4789.6
casa12	ZIP	$Normal(3, 2^2)$	$Gama(1, 1)$	26.0	-4790.0

Para todos os modelos listados, as estimativas de Pareto  $k$  são boas ( $k < 0.5$ ), o que assinala boa confiabilidade. Observamos também que os modelos ‘casa2’, ‘casa4’, ‘casa6’, ‘casa7’, ‘casa9’, ‘casa10’, ‘casa11’ e ‘casa12’ estão bem especificados, uma vez que o p\_loo, número efetivo estimado de parâmetros, é menor ou igual ao número total de parâmetros do modelo. Entretanto, ‘casa4’ e ‘casa6’ apresentaram transições divergentes, o que indica que as estimativas a posteriori dos parâmetros podem não ser confiáveis, logo eles foram desconsiderados na comparação entre os modelos, exposta na Tabela 11.

Tabela 11: Comparativo LOO's - Casa.

Modelo	elpd_diff	se_diff
<b>casa7</b>	0.0	0.0
casa11	-1.4	0.2
casa9	-1.4	0.3
casa12	-1.8	0.2
casa10	-1.9	0.3
casa2	-73.4	12.9

O modelo ‘casa7’ é o modelo com o melhor ajuste entre os testados, uma vez que na comparação, seus valores de `elpd_diff` e `se_diff` são a referência. A Figura 15 apresenta o diagnóstico PSIS, em que é possível observar a distribuição das estimativas de Pareto  $k$ , nota-se que todos os valores são menores que 0,5 e estão concentrados em torno do 0.

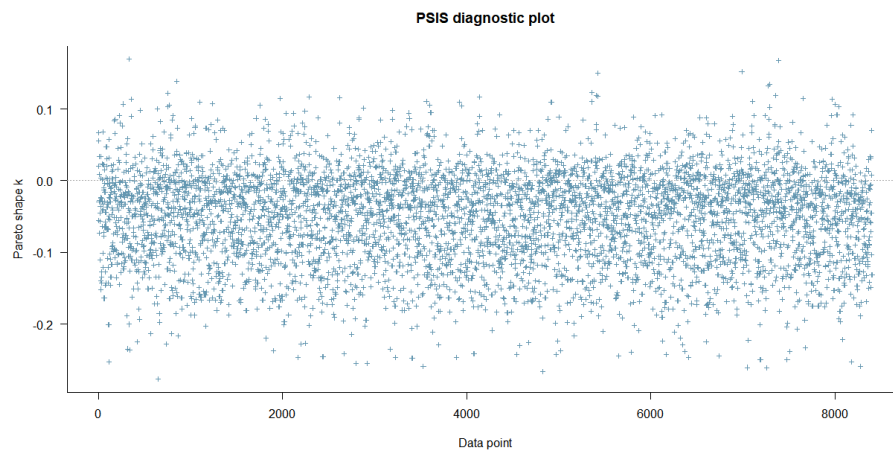


Figura 15: Diagnóstico PSIS - Modelo ‘casa7’.

O modelo ‘casa7’ pode ser escrito como:

$$X_{i,j}^C \sim \text{Poisson}(\mu_{i,\alpha}^C) \quad (20)$$

em que  $X_{i,j}^C$ , número de contatos do indivíduo  $i$  com alguém do grupo de idade  $j$ , no círculo social  $C$  (casa), assume uma distribuição  $\text{Poisson}(\mu_{i,\alpha}^C)$ . E:

$$\log(\mu_{i,\alpha}^C) = \beta_0^C + \sum_{a_i,\alpha} \beta_{a_i,\alpha}^C + v_i + \sigma_i, \quad a_i, \alpha = 0, 2, 3, 4, 5. \quad (21)$$

Logo,

$$\mu_{i,\alpha}^C = \exp^{\beta_0^C} \prod_{a_i,\alpha} \beta_{a_i,\alpha}^C v_i \sigma_i, \quad a_i, \alpha = 0, 2, 3, 4, 5 \quad (22)$$

onde os  $\beta'$ s representam o efeito das faixas etárias dos respondentes, dos contatos e a interação entre ambas,  $v_i$  é a covariável referente ao tamanho do domicílio e  $\sigma_i$  é o efeito aleatório a nível do indivíduo, o que faz com que os contatos não sejam fundamentados unicamente nas características dos contextos sociais, mas sim, considerando as características do indivíduo  $i$  em cada círculo.

Para este modelo, assumimos como prioris para os ( $\beta'$ s) uma distribuição *Normal*(3, 2<sup>2</sup>), uma vez que, a partir do conhecimento prévio oriundo da descrição do banco em estudo e de estudos anteriores há indícios que as pessoas tendem a ter, em média, 3 contatos por dia, com um desvio padrão de 2. E a priori para o desvio padrão dos efeitos aleatórios assume uma distribuição *Gama*(0.1, 0.1), uma priori pouco informativa devido ao desconhecimento e incerteza sobre este parâmetro.

A tabela abaixo detalha as estimativas a posteriori do modelo.

Nota-se que todos os parâmetros convergiram bem ( $\hat{R} = 1.00$ ) e apresentaram um tamanho efetivo da amostra (n\_eff) razoável. As estimativas relacionadas às faixas etárias 1's (dos contatos e do respondente) são capturadas pelo intercepto ( $\beta_0$ ). Os gráficos de diagnóstico de convergência e verificação preditiva a posteriori do modelo se encontram no Anexo 3.

As taxas de contato são dadas pela equação (21). Para exemplificação, as taxas entre as faixas etárias 1's ( $F_1 - f_1$ ) e entre um indivíduo da faixa etária 2 com uma alguma pessoa da faixa etária 3 ( $F_2 - f_3$ ), são dadas respectivamente por:

$$F_1 - f_1 : \exp [\log(\mu_{1,1}^C)] = \exp[\beta_0 + v_i] \quad (23)$$

$$F_2 - f_3 : \exp [\log(\mu_{2,3}^C)] = \exp[\beta_0 + \beta_{2,0} + \beta_{0,3} + \beta_{2,3} + v_i] \quad (24)$$

Assim, a matriz das taxas de contato entre o indivíduo  $i$ , do grupo de idade  $F_k$ , e alguém do grupo etário  $j$ , ( $f_j$ ), com  $k, j = 1, \dots, 5$ , é dada por:

A Figura 16 contém a representação gráfica dessas taxas.

Os grupos de idade que apresentaram as maiores médias de contatos são as faixas 3 com a 1, 1 com a 3 e 2 com a 4 (faixa etária do indivíduo respondente e de seu contato, respectivamente). Pode-se ressaltar pelas cores mais fortes da diagonal principal, a leve presença de homofilia, ou seja, as pessoas tendem a ter contatos com outras pessoas que apresentam características semelhantes as suas. Destaca-se ainda que as matrizes de contatos teriam que ser simétricas pela própria natureza do contato. Entretanto neste caso, isso não ocorre devido ao viés existente no respondente. Ademais, a simetria da matriz não foi realizada, pois a taxa de contatos específica por região a ser utilizada posteriormente é uma ponderação das taxas aqui calculadas.



Tabela 12: Resumo do modelo ‘casa7’.

Parâmetro	<i>Rhat</i>	n_eff	Média	sd	2.5%	97.5%
b_Intercept ( $\hat{\beta}_0$ )	1.00	2058	-1.69	0.10	-1.90	-1.49
b_fx.etaria_idF2 ( $\hat{\beta}_{2,0}$ )	1.00	1997	-0.10	0.16	-0.42	0.21
b_fx.etaria_idF3 ( $\hat{\beta}_{3,0}$ )	1.00	1836	0.25	0.11	0.04	0.47
b_fx.etaria_idF4 ( $\hat{\beta}_{4,0}$ )	1.00	2031	-0.47	0.13	-0.72	-0.23
b_fx.etaria_idF5 ( $\hat{\beta}_{5,0}$ )	1.00	2179	-1.05	0.24	-1.54	-0.61
b_faixa.etariafx2 ( $\hat{\beta}_{0,2}$ )	1.00	2348	-0.85	0.15	-1.15	-0.55
b_faixa.etariafx3 ( $\hat{\beta}_{0,3}$ )	1.00	1932	0.36	0.11	0.14	0.58
b_faixa.etariafx4 ( $\hat{\beta}_{0,4}$ )	1.00	1850	-0.06	0.12	-0.30	0.18
b_faixa.etariafx5 ( $\hat{\beta}_{0,5}$ )	1.00	1909	-1.65	0.21	-2.07	-1.26
b_tam_domicilio ( $\hat{v}_i$ )	1.00	6000	0.17	0.01	0.15	0.20
b_fx.etaria_idF2:faixa.etariafx2 ( $\hat{\beta}_{2,2}$ )	1.00	3155	0.21	0.28	-0.34	0.76
b_fx.etaria_idF3:faixa.etariafx2 ( $\hat{\beta}_{3,2}$ )	1.00	3301	-0.81	0.23	-1.26	-0.36
b_fx.etaria_idF4:faixa.etariafx2 ( $\hat{\beta}_{4,2}$ )	1.00	2802	0.18	0.23	-0.26	0.62
b_fx.etaria_idF5:faixa.etariafx2 ( $\hat{\beta}_{5,2}$ )	1.00	2953	0.90	0.35	0.21	1.58
b_fx.etaria_idF2:faixa.etariafx3 ( $\hat{\beta}_{2,3}$ )	1.00	2819	-0.60	0.23	-1.06	-0.15
b_fx.etaria_idF3:faixa.etariafx3 ( $\hat{\beta}_{3,3}$ )	1.00	2290	-0.54	0.15	-0.85	-0.24
b_fx.etaria_idF4:faixa.etariafx3 ( $\hat{\beta}_{4,3}$ )	1.00	2394	-0.06	0.17	-0.40	0.27
b_fx.etaria_idF5:faixa.etariafx3 ( $\hat{\beta}_{5,3}$ )	1.00	2567	0.17	0.30	-0.40	0.76
b_fx.etaria_idF2:faixa.etariafx4 ( $\hat{\beta}_{2,4}$ )	1.00	2267	0.51	0.21	0.10	0.92
b_fx.etaria_idF3:faixa.etariafx4 ( $\hat{\beta}_{3,4}$ )	1.00	2150	-0.14	0.16	-0.45	0.17
b_fx.etaria_idF4:faixa.etariafx4 ( $\hat{\beta}_{4,4}$ )	1.00	2166	0.61	0.17	0.29	0.94
b_fx.etaria_idF5:faixa.etariafx4 ( $\hat{\beta}_{5,4}$ )	1.00	2449	1.08	0.28	0.54	1.65
b_fx.etaria_idF2:faixa.etariafx5 ( $\hat{\beta}_{2,5}$ )	1.00	6000	-0.06	0.41	-0.88	0.73
b_fx.etaria_idF3:faixa.etariafx5 ( $\hat{\beta}_{3,5}$ )	1.00	2774	0.09	0.27	-0.43	0.63
b_fx.etaria_idF4:faixa.etariafx5 ( $\hat{\beta}_{4,5}$ )	1.00	2420	1.42	0.25	0.93	1.93
b_fx.etaria_idF5:faixa.etariafx5 ( $\hat{\beta}_{5,5}$ )	1.00	1979	2.59	0.33	1.97	3.25
sd_id__Intercept	1.00	3689	0.00	0.01	0.00	0.04

### 5.1.2 Círculo social: Outros

O círculo social ‘Outros’ aborda as situações de contato que ocorreram fora do núcleo domiciliar, ou seja, os contatos realizados “fora de casa”. Nesse contexto, foram contabilizados os contatos que advieram dos seguintes locais: vizinhança, escola, trabalho, comércio ou serviço, lazer, igreja e velório.

Assim como para o círculo social ‘Casa’, foram ajustados modelos assumindo diferentes famílias para a variável resposta: Poisson, Poisson Inflada de Zeros e Binomial Negativa; e diferentes prioris. Em todos os modelos a variável de interesse, os contatos, é explicada pela interação de duas covariáveis: a faixa etária do indivíduo respondente e a faixa etária do contato, que é disposta como um efeito fixo. A Tabela 14 descreve essas informações.

Tabela 13: Matriz de contatos por faixa etária: contexto Casa.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,218711887	0,093480726	0,313486181	0,205975098	0,042003598
$F_2$	0,197898699	0,104350485	0,15567263	0,310366941	0,035793105
$F_3$	0,280831622	0,053397038	0,234570288	0,229925485	0,059012854
$F_4$	0,136695425	0,069948222	0,184519524	0,236927759	0,108609109
$F_5$	0,076535545	0,080459607	0,130028711	0,212247974	0,195929574

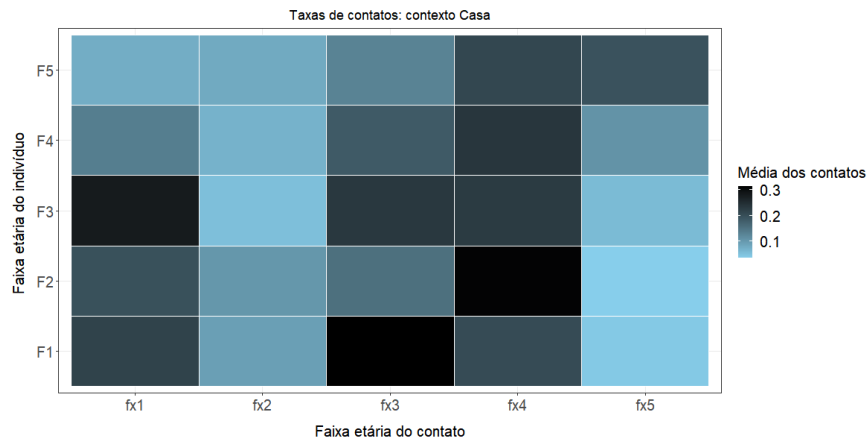


Figura 16: Matriz de contatos por faixa etária: contexto Casa. Cores mais fortes indicam maiores médias de contatos.

Tabela 14: Modelos ajustados - Outros.

Modelo	Família	Priori ( $\beta's$ )	Priori (sd $\sigma_i$ )	p_loo	elpd_loo
outros1	Poisson	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	139.0	-2841.4
outros2	Poisson	$Normal(3, 2^2)$	$Gama(1, 1)$	142.4	-2842.1
<b>outros3</b>	Bin Neg	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	26.9	-2753.6
outros4	Bin Neg	$Normal(3, 2^2)$	$Gama(1, 1)$	33.5	-2754.5
<b>outros5</b>	ZIP	$Normal(3, 2^2)$	$Gama(0.1, 0.1)$	30.3	-2757.0
outros6	ZIP	$Normal(3, 2^2)$	$Gama(1, 1)$	39.8	-2758.3

Os modelos ‘outros3’ e ‘outros5’ apresentaram transições divergentes, indicativo de que as posteriores podem não ser confiáveis. A Tabela 15 apresenta o comparativo dos modelos restantes.

Tabela 15: Comparativo LOO's - Outros.

Modelo	elpd_diff	se_diff
<b>outros4</b>	0.0	0.0
outros6	-3.8	5.0
outros1	-86.8	14.2
outros2	-87.6	14.2

O modelo ‘outros4’ apresentou o melhor ajuste entre os testados. Suas estimativas de Pareto  $k$  são boas ( $k < 0,5$ ), como apresentado na Figura 17, indicam boa confiabilidade. Entretanto, seu valor de  $p_{loo} = 33,5$ , sugere que o modelo não está bem especificado. Outros modelos foram ajustados, porém nenhum apresentou um valor de  $p_{loo}$  ideal. Diante disso, seguimos a análise com este modelo e, futuramente, avaliaremos novas prioridades.

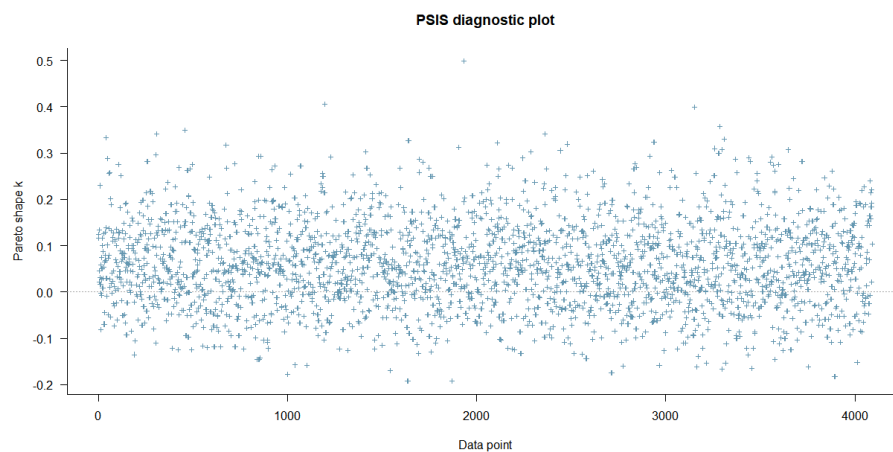


Figura 17: Diagnóstico PSIS - Modelo ‘outros4’.

No modelo ‘outros4’ a família da variável resposta é uma Binomial Negativa. Assumimos como prioris para o intercepto e parâmetros de efeitos no nível da população (efeitos fixos) uma distribuição  $Normal(3, 2^2)$ , para o desvio padrão dos efeitos em nível de grupo (desvio padrão do efeito aleatório -  $\sigma_i$ ) uma distribuição  $Gama(1, 1)$  e para o parâmetro específico da família,  $w_i$ , uma  $Gama(0.01, 0.01)$ . A tabela seguinte contém as estimativas a posteriori do modelo.

Tabela 16: Resumo do modelo ‘outros4’.

Parâmetro	<i>Rhat</i>	n_eff	Média	sd	2.5%	97.5%
b_Intercept	1.00	2810	-0.83	0.14	-1.11	-0.55
b_fx.etaria_idF2	1.00	4320	-0.10	0.31	-0.71	0.51
b_fx.etaria_idF3	1.00	3344	-1.06	0.21	-1.48	-0.65
b_fx.etaria_idF4	1.00	3432	-0.98	0.22	-1.42	-0.55
b_fx.etaria_idF5	1.00	3684	-1.84	0.33	-2.52	-1.22
b_faixa.etariafx2	1.00	4058	-1.46	0.30	-2.08	-0.88
b_faixa.etariafx3	1.00	3654	-0.30	0.22	-0.73	0.13
b_faixa.etariafx4	1.00	3467	-0.00	0.21	-0.40	0.42
b_faixa.etariafx5	1.00	4436	-1.44	0.29	-2.03	-0.87
b_fx.etaria_idF2:faixa.etariafx2	1.00	4610	1.98	0.48	1.06	2.92
b_fx.etaria_idF3:faixa.etariafx2	1.00	4237	1.67	0.37	0.94	2.41
b_fx.etaria_idF4:faixa.etariafx2	1.00	4827	0.53	0.43	-0.30	1.38
b_fx.etaria_idF5:faixa.etariafx2	1.00	4489	1.27	0.54	0.20	2.31
b_fx.etaria_idF2:faixa.etariafx3	1.00	4531	0.84	0.44	-0.02	1.71
b_fx.etaria_idF3:faixa.etariafx3	1.00	3717	2.15	0.29	1.59	2.70
b_fx.etaria_idF4:faixa.etariafx3	1.00	3706	1.40	0.30	0.82	1.97
b_fx.etaria_idF5:faixa.etariafx3	1.00	3567	2.01	0.40	1.26	2.81
b_fx.etaria_idF2:faixa.etariafx4	1.00	4539	0.80	0.41	-0.02	1.61
b_fx.etaria_idF3:faixa.etariafx4	1.00	3616	1.30	0.28	0.75	1.85
b_fx.etaria_idF4:faixa.etariafx4	1.00	3450	1.42	0.28	0.86	1.98
b_fx.etaria_idF5:faixa.etariafx4	1.00	3610	1.90	0.39	1.16	2.69
b_fx.etaria_idF2:faixa.etariafx5	1.00	6000	-1.44	1.01	-3.65	0.30
b_fx.etaria_idF3:faixa.etariafx5	1.00	6000	0.32	0.43	-0.54	1.14
b_fx.etaria_idF4:faixa.etariafx5	1.00	4176	1.43	0.38	0.70	2.19
b_fx.etaria_idF5:faixa.etariafx5	1.00	4276	2.62	0.45	1.75	3.52
sd_id__Intercept	1.00	2718	0.07	0.05	0.00	0.20
shape ( $\hat{w}_i$ )	1.00	6000	1.08	0.13	0.85	1.35

Percebe-se que todos os parâmetros convergiram bem ( $\hat{R} = 1.00$ ) e apresentaram um bom tamanho efetivo da amostra (n\_eff). Os gráficos de diagnóstico de convergência e verificação preditiva a posteriori do modelo se encontram no Anexo 4 e, a interpretação e o cálculo das taxas são feitas de maneira equivalente ao contexto anterior.

Através da Figura 18 observa-se a presença de homofilia mais forte do que se comparada ao contexto ‘Casa’, e que os indivíduos, em média, aparentam ter mais contatos com pessoas das faixas etárias 3 e 4. Essa matriz também não está simetrizada e devido ao agrupamento dos locais de contato fora do ambiente domiciliar, é difícil retratar de forma mais detalhada as características dos contatos.

Tabela 17: Matriz de contatos por faixa etária: contexto Outros.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,436049286	0,101266462	0,323033256	0,436049286	0,10331218
$F_2$	0,39455371	0,66365025	0,677056874	0,878095431	0,022148179
$F_3$	0,151071809	0,186373976	0,960789439	0,554327285	0,049291679
$F_4$	0,163654137	0,064570347	0,491644197	0,677056874	0,162025751
$F_5$	0,069252225	0,05726876	0,382892886	0,463013068	0,225372656

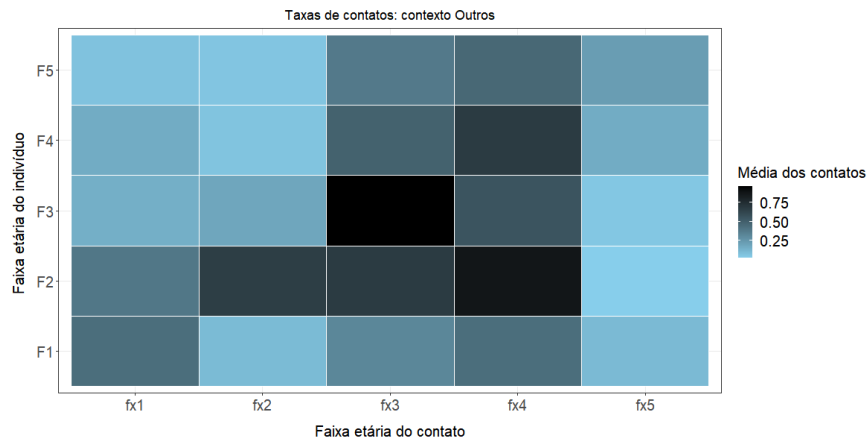


Figura 18: Taxa de contatos por faixa etária: contexto Outros. Cores mais fortes indicam maiores médias de contatos.

### 5.1.3 Contexto Geral

Com o intuito de explorar a disseminação do novo coronavírus nas regiões de Belo Horizonte, estimou-se as taxas de contato para a região do Aglomerado da Serra, desconsiderando os locais das situações de contato. Estas taxas foram projetadas para as demais regiões da cidade de acordo com a metodologia proposta por Prem et al. (2017), etapa que será explicada na próxima seção.

Os modelos foram ajustados de forma equivalente ao contextos sociais anteriores. Avaliou-se nas três famílias: Poisson, Poisson Inflada de Zeros e Binomial Negativa; e com diferentes prioris, tendo em vista o conhecimento prévio das descrições dos dados e pela pouca informação que temos. Os modelos foram dispostos da forma que os contatos são explicados pela interação entre a faixa etária do respondente  $i$  e a da pessoa com quem esse teve o contato. A Tabela 18 explana esses dados.

Tabela 18: Modelos ajustados - Contexto Geral.

Modelo	Família	Priori ( $\beta$ 's)	Priori (sd $\sigma_i$ )	p_loo	elpd_loo
geral1	Poisson	<i>Normal</i> (3, 2)	<i>Gama</i> (1, 1)	61.0	-7842.5
geral2	Poisson	<i>Normal</i> (3, 2)	<i>Gama</i> (0.1, 0.1)	34.0	-7840.5
geral3	Bin Neg	<i>Normal</i> (3, 2)	<i>Gama</i> (1, 1)	35.6	-7799.0
geral4	Bin Neg	<i>Normal</i> (3, 2)	<i>Gama</i> (0.1, 0.1)	27.6	-7798.0
geral5	ZIP	<i>Normal</i> (3, 2)	<i>Gama</i> (1, 1)	43.0	-7823.7
geral6	ZIP	<i>Normal</i> (3, 2)	<i>Gama</i> (0.1, 0.1)	29.8	-7823.0

Os modelos ‘geral2’, ‘geral4’ e ‘geral6’ apresentaram transições divergentes, o que sugere que as posteriores dos parâmetros estimados não são confiáveis. Seguiu-se então com a análise para os ajustes restantes: ‘geral1’, ‘geral3’ e ‘geral5’.

Para esses modelos, a estimativa de Pareto  $k$  são boas ( $k < 0, 5$ ), o que indica boa confiabilidade. Contudo, apresentam o valor de p\_loo maior que o número total de parâmetros, o que indica que provavelmente existe alguma especificação incorreta do modelo. Outros modelos além dos listados acima foram ajustados, todavia, os que apresentavam os valores mais próximos do p\_loo considerado ideal (menor ou igual ao número de parâmetros do modelo) possuíam transições divergentes. Acredita-se que avaliá-los com novas prioris, assim como com a inclusão de novas variáveis preditoras, possivelmente possa influenciar positivamente o ajuste e assim, talvez os modelos estariam melhores especificados. No entanto, dado que quer-se inicialmente as taxas de contato explicada apenas pela interação entre a faixa etária do indivíduo respondente e de seu contato, continuou-se o processo de análise com estes modelos e, futuramente pretende-se ajustá-los com novas covariáveis para diferentes aplicações. À vista disso, assumiu-se a qualidade de ajuste dos demais modelos, prosseguiu-se com a comparação destes, realizou-se as análises subsequentes e projeção para as demais regiões de BH.

A tabela a seguir exprime a comparação entre os modelos. A primeira linha da tabela apresenta o melhor modelo entre os comparados. E a Figura 19 caracteriza o PSIS diagnóstico do modelo, em que podemos observar as estimativa de Pareto  $k < 0, 5$ .

Tabela 19: Comparativo LOO's - Contexto Geral.

Modelo	elpd_diff	se_diff
<b>geral3</b>	0.0	0.0
geral5	-25.0	5.0
geral1	-43.5	11.0

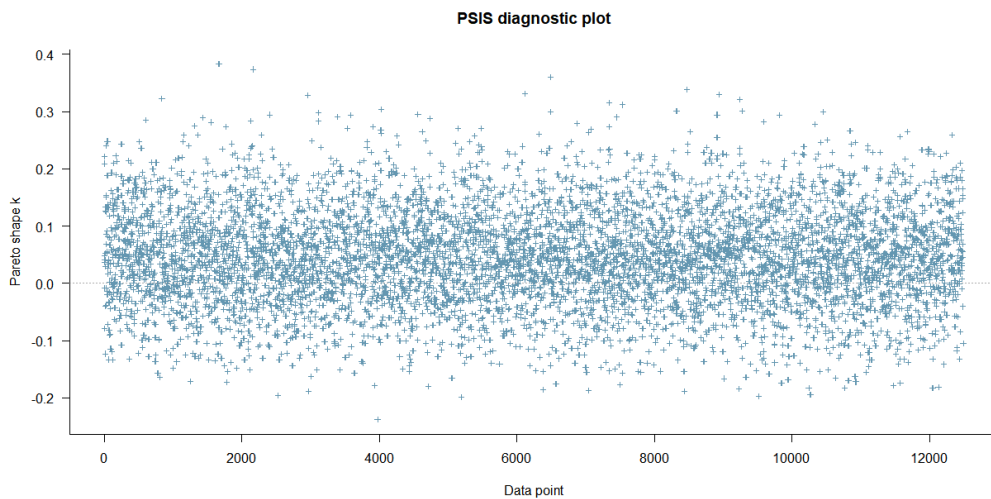


Figura 19: Diagnóstico PSIS - Modelo 'geral3'.

No modelo 'geral3' a família da variável resposta é uma Binomial Negativa. Assumimos como prioris para o intercepto e parâmetros de efeitos no nível da população (efeitos fixos) uma distribuição  $Normal(3, 2^2)$ , para o desvio padrão dos efeitos em nível de grupo (desvio padrão do efeito aleatório -  $\sigma_i$ ) uma distribuição  $Gama(1, 1)$  e para o parâmetro específico da família,  $w_i$ , uma  $Gama(0.01, 0.01)$ . A Tabela seguinte apresenta as estimativas a posteriori do modelo.

Tabela 20: Resumo do modelo ‘geral3’.

Parâmetro	<i>Rhat</i>	n_eff	Média	sd	2.5%	97.5%
b_Intercept	1.00	1910	-0.90	0.08	-1.05	-0.75
b_fx.etaria_idF2	1.00	3271	-0.19	0.15	-0.49	0.09
b_fx.etaria_idF3	1.00	2282	-0.22	0.10	-0.42	-0.02
b_fx.etaria_idF4	1.00	2582	-0.77	0.12	-0.99	-0.54
b_fx.etaria_idF5	1.00	2860	-1.53	0.20	-1.93	-1.15
b_faixa.etariafx2	1.00	2756	-0.98	0.14	-1.27	-0.72
b_faixa.etariafx3	1.00	2294	0.20	0.10	0.00	0.41
b_faixa.etariafx4	1.00	2404	-0.03	0.11	-0.25	0.18
b_faixa.etariafx5	1.00	3723	-1.55	0.17	-1.90	-1.22
b_fx.etaria_idF2:faixa.etariafx2	1.00	3945	0.70	0.24	0.24	1.17
b_fx.etaria_idF3:faixa.etariafx2	1.00	3560	-0.02	0.19	-0.39	0.35
b_fx.etaria_idF4:faixa.etariafx2	1.00	3693	0.24	0.21	-0.16	0.64
b_fx.etaria_idF5:faixa.etariafx2	1.00	3698	0.94	0.30	0.37	1.53
b_fx.etaria_idF2:faixa.etariafx3	1.00	3964	-0.25	0.21	-0.67	0.15
b_fx.etaria_idF3:faixa.etariafx3	1.00	2662	0.37	0.13	0.11	0.63
b_fx.etaria_idF4:faixa.etariafx3	1.00	2856	0.37	0.15	0.07	0.67
b_fx.etaria_idF5:faixa.etariafx3	1.00	3145	0.88	0.24	0.42	1.35
b_fx.etaria_idF2:faixa.etariafx4	1.00	3247	0.53	0.20	0.15	0.92
b_fx.etaria_idF3:faixa.etariafx4	1.00	2652	0.30	0.14	0.02	0.57
b_fx.etaria_idF4:faixa.etariafx4	1.00	2790	0.89	0.15	0.59	1.19
b_fx.etaria_idF5:faixa.etariafx4	1.00	3006	1.43	0.23	0.98	1.89
b_fx.etaria_idF2:faixa.etariafx5	1.00	6000	-0.43	0.38	-1.22	0.29
b_fx.etaria_idF3:faixa.etariafx5	1.00	3733	0.07	0.23	-0.38	0.52
b_fx.etaria_idF4:faixa.etariafx5	1.00	3553	1.37	0.22	0.96	1.80
b_fx.etaria_idF5:faixa.etariafx5	1.00	3464	2.56	0.27	2.04	3.11
sd_id__Intercept	1.00	2521	0.04	0.03	0.00	0.12
shape ( $\hat{w}_i$ )	1.00	6000	2.65	0.38	2.03	3.50

O  $\hat{R} = 1.00$  indica que houve convergência das estimativas, para grande parcela dessas, o 0 não está incluso no intervalo de credibilidade e o tamanho efetivo da amostra para elas são razoáveis. Maiores informações sobre o diagnóstico do modelo se encontram no Anexo 5.

A matriz de contatos está especificada na Tabela a seguir e sua representação gráfica na Figura 20.



Tabela 21: Matriz de contatos por faixa etária: contexto Geral.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,40656966	0,152590106	0,496585304	0,39455371	0,086293586
$F_2$	0,336216494	0,25410696	0,319819022	0,554327285	0,046421155
$F_3$	0,326279795	0,120031629	0,57694981	0,427414932	0,074273578
$F_4$	0,188247066	0,089815295	0,332871084	0,444858066	0,157237166
$F_5$	0,088036833	0,084584859	0,259240261	0,357006961	0,241714017

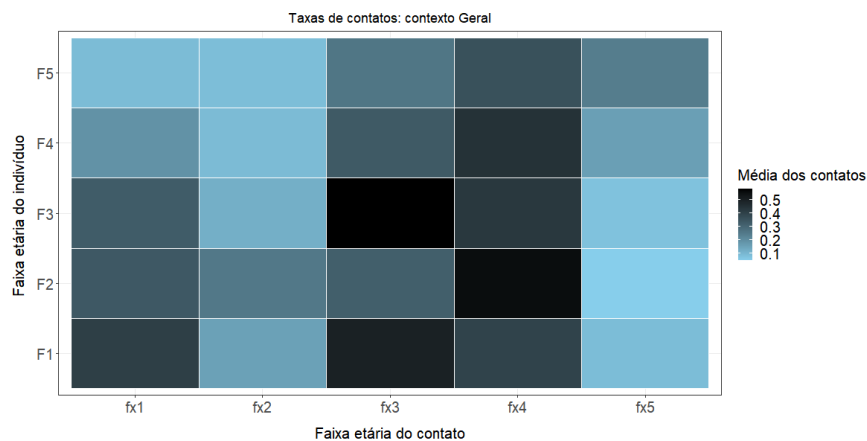


Figura 20: Taxa de contatos por faixa etária: contexto Geral. Cores mais fortes indicam maiores médias de contatos.

Observa-se pelas diagonais mais fortes, que há indícios de homofilia. Destaca-se também que, de forma geral, as pessoas tiveram em média, um maior número de contatos com indivíduos das faixas etárias 3 e 4, e, ressalta-se que a matriz não está simetrizada pela mesma razão relatada nos contextos anteriores.

## 5.2 Projeção

A projeção foi realizada seguindo a metodologia apresentada no artigo da Prem et al. (2017). A proposta consiste em reestimar as taxas de contato para diversos países que não ostentam um banco de dados que retrate as situações de contato, a partir dos dados coletados pelo Projeto POLYMOD (Mossong et al., 2008) e de outras fontes que contém as informações das estruturas sociodemográficas dos países para qual se procura estimar.

Para o presente estudo, adaptamos a metodologia proposta por Prem et al. (2017) e projetamos as taxas de contato do Aglomerado da Serra para 10 regiões de Belo Horizonte: Barreiro, Centro, Centro-Sul, Leste, Nordeste, Noroeste, Norte, Oeste, Pampulha e Venda Nova. Essa divisão se deu pela separação já existente do município de Belo Horizonte em 9 administrações regionais (Barreiro, Centro-Sul, Leste, Nordeste, Noroeste, Norte, Oeste, Pampulha e Venda Nova) e somado a isso, duas partições da regional Centro-Sul: o Aglomerado da Serra, região onde foram coletados os dados, e a região do Centro, delimitada pela Avenida do Contorno, região onde acreditamos que exista uma movimentação diferente da mobilização de transporte.

A projeção foi realizada com a colaboração de Reinaldo dos Santos, assessor da Secretaria de Planejamento, Orçamento e Gestão da Prefeitura de Belo Horizonte e Professor substituto do Departamento de Demografia da Faculdade de Ciências Econômicas-FACE da Universidade Federal de Minas Gerais.

A estimação das matrizes de contato se deu pela equação:

$$\mu_{a,\alpha}^{R_i} = \lambda_{a,\alpha} \times \frac{\pi_{\alpha}^{R_i}}{\pi_{\alpha}^{Ag}}, \quad i = 1, \dots, 10, \quad (25)$$

em que a taxa de contato  $\mu_{a,\alpha}^{R_i}$  do grupo etário  $a$  para o  $\alpha$  de uma região  $R_i$  é dado pelo produto da taxa de contato do Aglomerado para estes mesmos grupos ( $\lambda_{a,\alpha}$ ), com a diferença entre as estruturas etárias da região  $R_i$  e do Aglomerado da Serra ( $P_{\alpha}^{Ag}$ ).

Assim temos que

$$\begin{bmatrix} \pi_1(\lambda_{11} + \dots + \lambda_{15}) \\ \pi_2(\lambda_{21} + \dots + \lambda_{25}) \\ \pi_3(\lambda_{31} + \dots + \lambda_{35}) \\ \pi_4(\lambda_{41} + \dots + \lambda_{45}) \\ \pi_5(\lambda_{51} + \dots + \lambda_{55}) \end{bmatrix} \leftarrow \begin{bmatrix} \pi_1\lambda_{11} & \pi_1\lambda_{12} & \pi_1\lambda_{13} & \pi_1\lambda_{14} & \pi_1\lambda_{15} \\ \pi_2\lambda_{21} & \pi_2\lambda_{22} & \pi_2\lambda_{23} & \pi_2\lambda_{24} & \pi_2\lambda_{25} \\ \pi_3\lambda_{31} & \pi_3\lambda_{32} & \pi_3\lambda_{33} & \pi_3\lambda_{34} & \pi_3\lambda_{35} \\ \pi_4\lambda_{41} & \pi_4\lambda_{42} & \pi_4\lambda_{43} & \pi_4\lambda_{44} & \pi_4\lambda_{45} \\ \pi_5\lambda_{51} & \pi_5\lambda_{52} & \pi_5\lambda_{53} & \pi_5\lambda_{54} & \pi_5\lambda_{55} \end{bmatrix}$$

Logo, a taxa de contato específica por região será dada por

$$\mu^R = \pi_1(\lambda_{11} + \dots + \lambda_{15}) + \dots + \pi_5(\lambda_{51} + \dots + \lambda_{55}). \quad (26)$$

As tabelas abaixo reproduzem as taxas de contatos projetadas, específicas por grupos de idade, para cada uma das regiões. Para encontrarmos as taxas específicas por região, sem distinção por faixa etária, somamos todos os elementos da matriz. Essas se encontram retratadas na Tabela 34, no próximo capítulo.

Tabela 22: Matriz de contatos por faixa etária: Barreiro.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,329782083	0,122554145	0,484997581	0,484000166	0,108723054
$F_2$	0,272716306	0,204088339	0,312356107	0,679994867	0,058486962
$F_3$	0,264656321	0,096404505	0,563486798	0,524311121	0,0935788
$F_4$	0,152693414	0,072135979	0,325103601	0,545708663	0,19810632
$F_5$	0,071409583	0,067935107	0,253190939	0,437941461	0,304540432

Tabela 23: Matriz de contatos por faixa etária: Centro.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,135683678	0,093484956	0,511985283	0,492630363	0,25355765
$F_2$	0,112204857	0,155679673	0,329737169	0,692119841	0,136399927
$F_3$	0,108888702	0,073537871	0,594842033	0,533660101	0,218239092
$F_4$	0,062823316	0,055025709	0,343193998	0,555439183	0,462012161
$F_5$	0,029380356	0,051821262	0,267279755	0,445750386	0,710231671

Tabela 24: Matriz de contatos por faixa etária: Aglomerado da Serra.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,40656966	0,152590106	0,496585304	0,39455371	0,086293586
$F_2$	0,336216494	0,25410696	0,319819022	0,554327285	0,046421155
$F_3$	0,326279795	0,120031629	0,57694981	0,427414932	0,074273578
$F_4$	0,188247066	0,089815295	0,332871084	0,444858066	0,157237166
$F_5$	0,088036833	0,084584859	0,259240261	0,357006961	0,241714017

Tabela 25: Matriz de contatos por faixa etária: Centro-Sul.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,199905174	0,092554747	0,465027338	0,527870782	0,21046201
$F_2$	0,16531341	0,154130605	0,299494542	0,741630783	0,113216867
$F_3$	0,16042766	0,072806142	0,540284685	0,571835592	0,18114633
$F_4$	0,092558708	0,054478184	0,311717146	0,595172645	0,383486784
$F_5$	0,0432866	0,051305621	0,242765558	0,477637236	0,589517944

Tabela 26: Matriz de contatos por faixa etária: Leste.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,263105812	0,113370638	0,462496209	0,506711407	0,164104816
$F_2$	0,217577754	0,188795125	0,297864403	0,71190297	0,088279273
$F_3$	0,21114736	0,089180502	0,537343933	0,548913914	0,141246324
$F_4$	0,121821429	0,066730521	0,31002048	0,571315516	0,29901847
$F_5$	0,056971793	0,062844438	0,241444193	0,458491441	0,459668393

Tabela 27: Matriz de contatos por faixa etária: Nordeste.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,29212598	0,11577546	0,477875253	0,503783688	0,132406553
$F_2$	0,241576247	0,192799855	0,307769068	0,707789679	0,071227368
$F_3$	0,23443659	0,091072203	0,555211833	0,545742355	0,113963376
$F_4$	0,135258146	0,068146011	0,320329362	0,568014523	0,241260469
$F_5$	0,063255694	0,064177497	0,249472758	0,455842332	0,370879471

Tabela 28: Matriz de contatos por faixa etária: Noroeste.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,249173801	0,108208964	0,468504586	0,515032652	0,1684813
$F_2$	0,20605655	0,180199435	0,301734017	0,723593883	0,090633578
$F_3$	0,199966659	0,085120186	0,544324671	0,557928211	0,145013199
$F_4$	0,115370726	0,063692334	0,314048016	0,580697693	0,306992944
$F_5$	0,053955015	0,059983181	0,244580841	0,466020815	0,471927213

Tabela 29: Matriz de contatos por faixa etária: Norte.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,328557308	0,123098041	0,478869184	0,485076996	0,112457964
$F_2$	0,271703467	0,204994084	0,308409195	0,681507757	0,060496136
$F_3$	0,263673416	0,096832349	0,556366616	0,525477636	0,096793466
$F_4$	0,152126327	0,072456119	0,320995613	0,546922786	0,20491177
$F_5$	0,071144376	0,068236603	0,249991635	0,438915816	0,31500216

Tabela 30: Matriz de contatos por faixa etária: Oeste.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,279389767	0,109168147	0,485704307	0,510512501	0,136965458
$F_2$	0,23104392	0,181796754	0,312811263	0,717243308	0,073679807
$F_3$	0,22421554	0,085874706	0,564307895	0,55303159	0,117887263
$F_4$	0,129361113	0,064256914	0,325577333	0,575601238	0,249567337
$F_5$	0,06049785	0,060514882	0,253559882	0,461930813	0,383649266

Tabela 31: Matriz de contatos por faixa etária: Pampulha.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,27799741	0,111636395	0,486634323	0,515506135	0,13183153
$F_2$	0,229892497	0,185907106	0,313410227	0,724259103	0,070918038
$F_3$	0,223098147	0,087816299	0,56538842	0,55844113	0,113468449
$F_4$	0,128716434	0,065709737	0,326200741	0,581231544	0,240212709
$F_5$	0,060196355	0,0618831	0,254045393	0,46644924	0,369268794

Tabela 32: Matriz de contatos por faixa etária: Venda Nova.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$F_1$	0,323102931	0,120692632	0,490052644	0,48554237	0,110736367
$F_2$	0,26719292	0,200988377	0,315611751	0,682161583	0,059570013
$F_3$	0,259296176	0,094940187	0,569359943	0,525981771	0,095311674
$F_4$	0,149600879	0,071040283	0,328492111	0,547447494	0,201774816
$F_5$	0,069963309	0,066903219	0,255829913	0,439336905	0,310179854

## 6 Aplicação 2: Modelos epidemiológicos

Neste capítulo aplicou-se as taxas estimadas para a região do Aglomerado e as taxas projetadas para as 10 regiões da cidade de Belo Horizonte em modelos compartimentais SIR. Serão 11 modelos, um para cada região, em que se considerou um índice de mobilidade genérico entre as regiões.

### 6.1 Modelo compartimental: SIR acoplado em redes

Os modelos em análise foram construídos possuindo como referência o artigo de Pujari (2020). Deste modo, explora-se a disseminação do novo coronavírus dentro de cada região a partir de um modelo SIR (Susceptible-Infected-Removed) acopladas por meio de redes de transporte, abordando os índices de circulação. Assim, consideramos então, as misturas intra e inter regiões.

Nesta abordagem, o modelo SIR é descrito como:

$$\frac{dS_i}{dt} = -\beta \frac{S_i(t)I_i(t)}{\phi_i} - \eta_i S_i(t) + \sum_j \frac{\mathbf{A}_{ij}\eta_j}{k_j} S_j, \quad (27)$$

$$\frac{dI_i}{dt} = \beta \frac{S_i(t)I_i(t)}{\phi_i} - \frac{1}{\gamma} I_i(t) - \eta_i I_i(t) + \sum_j \frac{\mathbf{A}_{ij}\eta_j}{k_j} I_j, \quad (28)$$

$$\frac{dR_i}{dt} = \frac{1}{\gamma} I_i(t) - \eta_i R_i(t) + \sum_j \frac{\mathbf{A}_{ij}\eta_j}{k_j} R_j, \quad (29)$$

em que temos:

- $S_i$ ,  $I_i$  e  $R_i$ , as frações de suscetíveis, infectados e recuperados na metapopulação  $i$ , respectivamente;
- $\phi_i$  o tamanho da metapopulação  $i$ , inde  $\phi_i = S_i + I_i + R_i$ ;
- $\beta$  e  $\frac{1}{\gamma}$  são as taxas de infecção e recuperação, respectivamente;
- $\mathbf{A}$  é a matriz de adjacência da rede, com  $\mathbf{A}_{ij} = 1$  se  $i$  e  $j$  estão conectados, ou seja, são regiões vizinhas e, 0 caso contrário;
- A taxa de emigração da região  $i$  é proporcional ao seu tamanho  $\phi_i$ , e neste estudo será uma valor genérico.

Ademais, tem-se também que:

$$\eta_i \phi_i = \sum_j \frac{\mathbf{A}_{ij}\eta_j}{k_j} \phi_j \quad (30)$$

uma equação autoconsistente (Pujari, 2020) onde considera-se que a população de uma região tende a permanecer constante durante o período da pandemia, mesmo com as migrações. E por isso, a migração líquida de uma região deve ser zero (Pujari, 2020). Os  $\eta$ 's obtidos são estipulados até uma constante de proporcionalidade  $D$ , chamada índice de mobilidade ou índice de circulação.

A Tabela 33 apresenta o número absoluto das pessoas presentes em cada compartimento do modelo SIR no estado inicial, suscetíveis ( $S_i(0)$ ), infectados ( $I_i(0)$ ) e recuperados ( $R_i(0)$ ); e da população total de cada uma das  $i$  regiões, com  $i = 1, \dots, 11$ . Sendo essas: Barreiro, Centro, Aglomerado da Serra, Centro-Sul, Leste, Nordeste, Noroeste, Norte, Oeste, Pampulha e Venda Nova.

Tabela 33: População total, suscetíveis, infectados e recuperados por região.

Região	População total	Suscetíveis	Infectados	Recuperados
Barreiro (BA)	282.552	256.770	219	25.563
Centro (C1)	84.941	73.739	100	11.102
Aglomerado da Serra (C2)	38.405	36.272	10	2.123
Centro-Sul (CS)	160.430	144.548	138	15.744
Leste (LE)	238.539	217.987	169	20.383
Nordeste (NE)	290.353	264.653	234	25.466
Noroeste (NO)	268.038	244.891	202	22.945
Norte (NT)	212.055	194.794	138	17.123
Oeste (OE)	308.549	277.217	268	31.064
Pampulha (PA)	226.110	200.753	225	25.132
Venda Nova (VN)	265.179	241.254	199	23.726
Total	2.375.151	2.152.878	1.902	220.371

Esses dados são referentes a população do período de 04 a 10 de junho de 2021, intervalo em que os dados das situações de contatos foram coletados. Enquanto o número de recuperados é específico do último dia da coleta, o número de infectados é a média dos valores dos casos ativos registrados durante esse tempo.

As regiões do município de BH estão dispostas na seguinte forma:

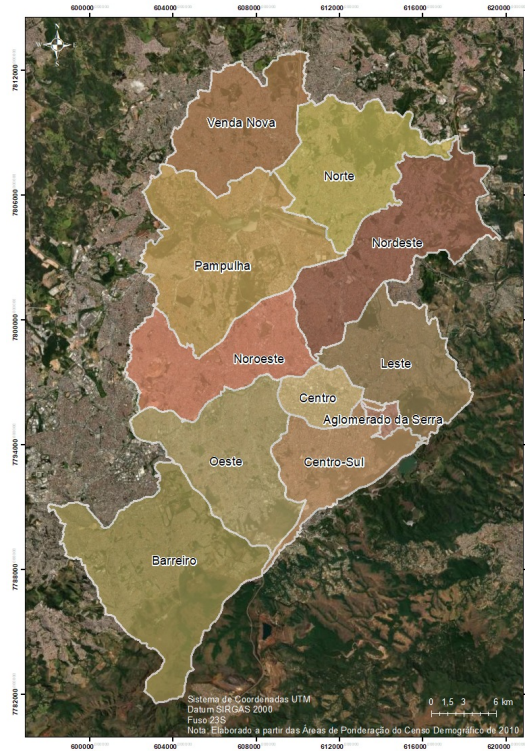


Figura 21: Regiões do município de Belo Horizonte

logo, a matriz de adjacência é dada por:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

em que as entradas da matriz seguem a ordem das regiões na Tabela 33.



Analisou-se a disseminação do coronavírus em 2 diferentes cenários para o nível de contágio. No primeiro, a probabilidade de transmissão do patógeno é de 0,09 e no segundo, a probabilidade é de 0,05. A Tabela 34 contém as taxas de contato, contágio ( $\beta(1)$  para o primeiro cenário e  $\beta(2)$  para o segundo) e o período, em dias, em que o indivíduo permanece infeccioso ( $\gamma$ ) (Li, R. et al, 2020). Utilizou-se valores genéricos para o índice de mobilidade  $D$ , sendo  $D = 0,8$ ,  $D = 0,5$  e  $D = 0,1$ , para um índice alto, intermediário e baixo de circulação, respectivamente. O valor de 0,8 foi determinado a partir da proporção estimada de habitantes do Aglomerado que utilizam o transporte público, já os demais valores de  $D$ , foram elegidos a fim de comparação, uma vez que 0,5 é o valor central e 0,1, o menor valor possível diferente de zero.

Tabela 34: Parâmetros do modelo SIR

Região	Taxas de contato	$\beta(1)$	$\beta(2)$	$\gamma$
Barreiro (BA)	7,0289	0,6326	0,3514	3,47
Centro (C1)	7,4256	0,6683	0,3713	3,47
Aglomerado da Serra (C2)	6,8160	0,6134	0,3408	3,47
Centro-Sul (CS)	7,3380	0,6604	0,3669	3,47
Leste (LE)	7,1904	0,6471	0,3595	3,47
Nordeste (NE)	7,1202	0,6408	0,3560	3,47
Noroeste (NO)	7,2212	0,6499	0,3611	3,47
Norte (NT)	7,0350	0,6332	0,3518	3,47
Oeste (OE)	7,1481	0,6433	0,3574	3,47
Pampulha (PA)	7,1501	0,6435	0,3575	3,47
Venda Nova (VN)	7,0411	0,6337	0,3521	3,47

Os modelos SIR analisados foram contruídos utilizando o pacote `EpiModel` (Jenness SM, 2018). A estimação do  $R_t$  foi realizada pelo método “Imperial College” (Cori et al., 2013; Thompson et al., 2019), utilizou-se o pacote `EpiEstim` (Cori et al., 2013) e, assumiu-se que a estimativa média de  $a$  é de 4,7 dias, com desvio padrão de 2,9 (Nishiura, Linton e Akhmetzhanov, 2020).

## 6.2 Cenário 1: probabilidade de transmissão = 0.09

Neste cenário, a taxa de contágio  $\beta(1)$  é dada pelo produto da taxa de contatos com a probabilidade de transmissão igual a 0,09. Entre os cenários analisados, é neste que o patógeno é mais contagioso.

### 6.2.1 Situação 1: $D = 0.8$

A primeira situação considera um alto índice de mobilidade entre as regiões. A Figura 22 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte, enquanto a Figura 23 exprime a estimação do  $R_t$ .

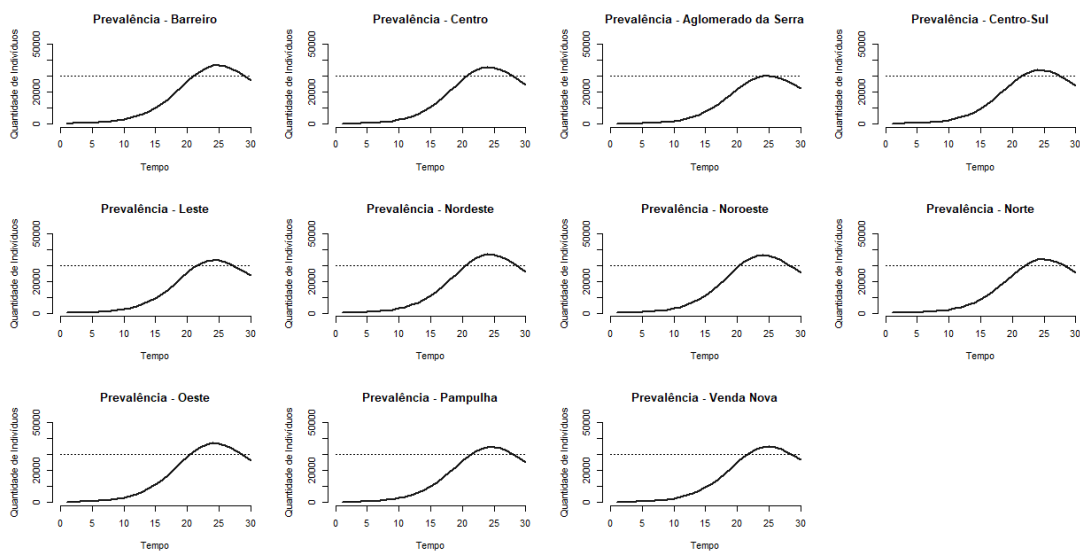


Figura 22: Cenário 1: Prevalência ( $D = 0.8$ )

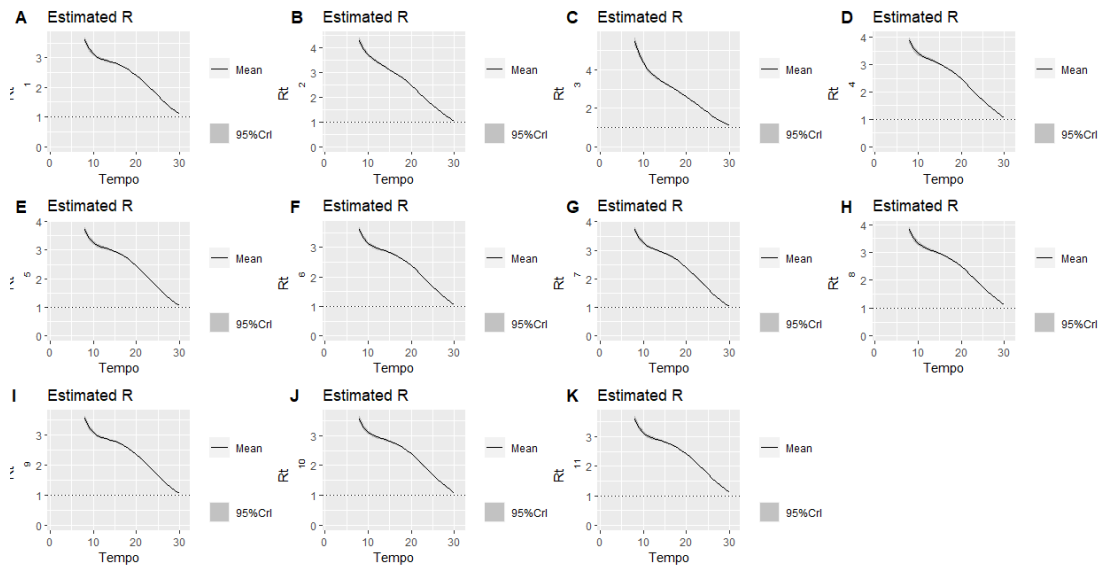


Figura 23: Cenário 1: Estimação  $R_t$  para  $D = 0.8$ . Regiões: **A)**Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

### 6.2.2 Situação 2: $D = 0.5$

A segunda situação considera um índice intermediário de circulação entre as regiões. A Figura 24 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte enquanto a Figura 25 exprime a estimação do  $R_t$ .

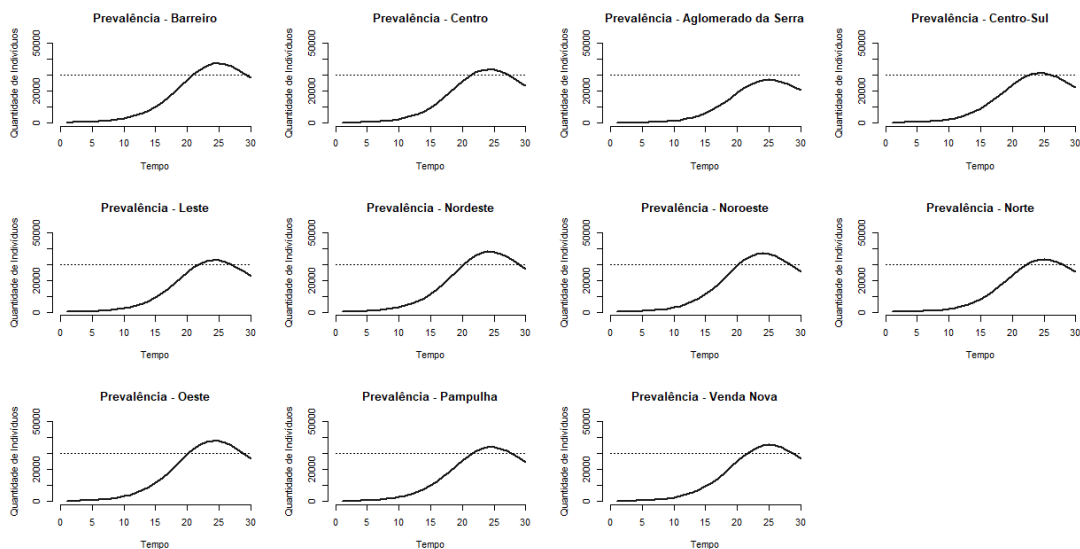


Figura 24: Cenário 1: Prevalência ( $D = 0.5$ )

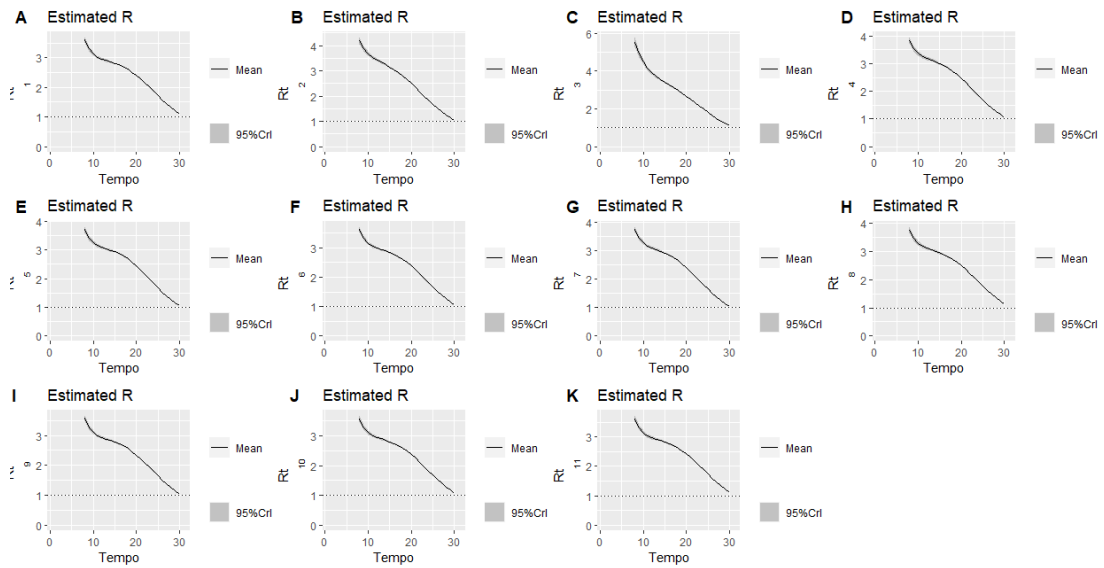


Figura 25: Cenário 1: Estimação  $R_t$  para  $D = 0.5$ . Regiões: **A)**Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

### 6.2.3 Situação 3: $D = 0.1$

A terceira situação considera um baixo índice de mobilidade entre as regiões. A Figura 26 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte enquanto a Figura 27 exprime a estimação do  $R_t$ .

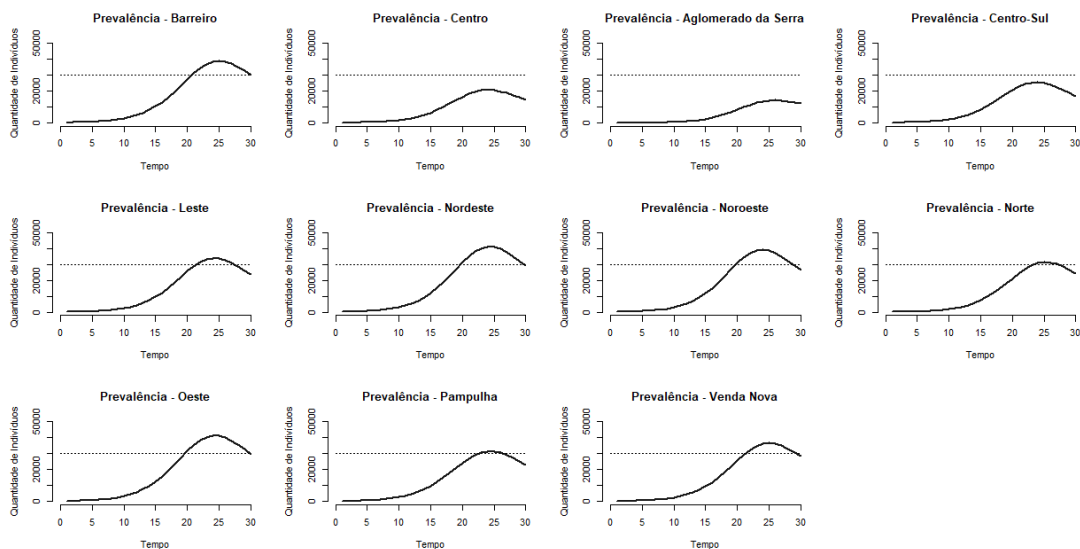


Figura 26: Cenário 1: Prevalência ( $D = 0.1$ )

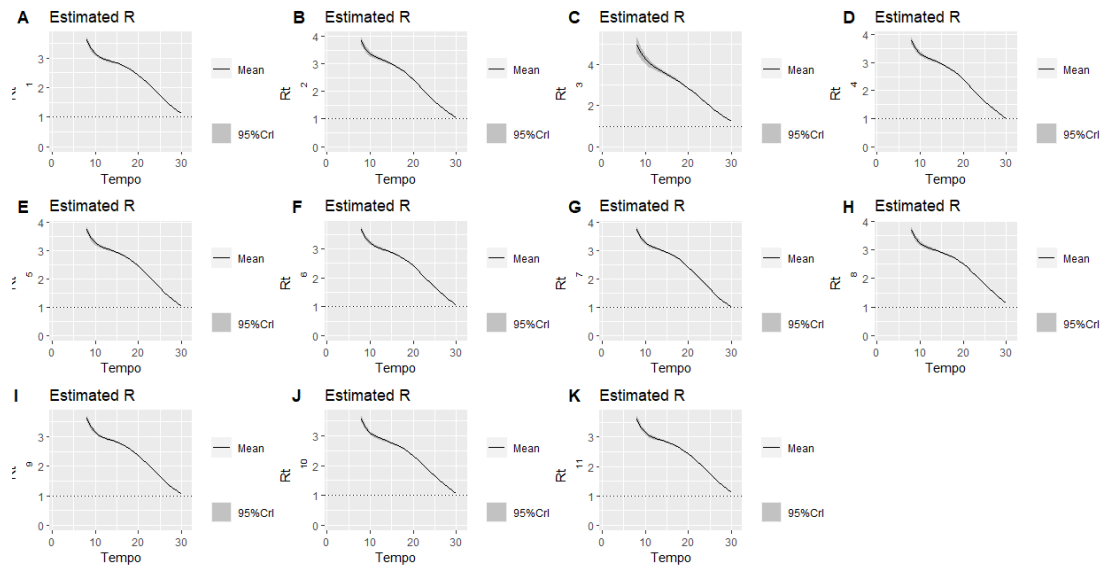


Figura 27: Cenário 1: Estimação  $R_t$  para  $D = 0.1$ . Regiões: **A)** Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

Observa-se que, de forma geral, os gráficos apresentam comportamentos semelhantes, isso se deve a similaridade das estruturas etárias entre algumas regiões, somado ao fato de se estar utilizando uma constante  $D$  genérica para elas, mesmo que apresentem diferentes índices de circulação. Apenas o Aglomerado da Serra apresenta uma diferença mais nítida na curvatura com a diminuição do índice de circulação, possivelmente é resultado do fato de ser a região que apresentou a menor taxa de contatos. O total de indivíduos infectados para as regiões Barreiro, Centro-Sul, Leste, Nordeste, Noroeste, Norte, Oeste, Pampulha e Venda Nova é maior que 30.000 (linha pontilhada) para as três situações.

Destaca-se ainda, as flutuações que ocorrem na prevalência do patógeno e no valor do  $R_t$  estimado para cada uma das 11 regiões com a diminuição do valor de  $D$ . Como exemplo, para o Centro, um maior índice de mobilidade resulta em um maior número de infectados que vai diminuindo com o decréscimo da constante  $D$ ; para a região Nordeste essa situação se inverte, a medida que diminui o o valor de  $D$ , o número de infectados aumenta. Mais estudos são necessários para justificar essa mudança e novas análises com índices de mobilidade específicos por regiões devem auxiliar em uma melhora na caracterização e entendimento.

### 6.3 Cenário 2: probabilidade de transmissão = 0.05

Neste cenário, a taxa de contágio  $\beta(2)$  é dada pelo produto da taxa de contatos com a probabilidade de transmissão igual a 0,05. Entre os cenários analisados, é neste que o patógeno é menos contagioso.

#### 6.3.1 Situação 1: $D = 0.8$

A primeira situação considera um alto índice de mobilidade entre as regiões. A Figura 22 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte, enquanto a Figura 23 exprime a estimação do  $R_t$ .

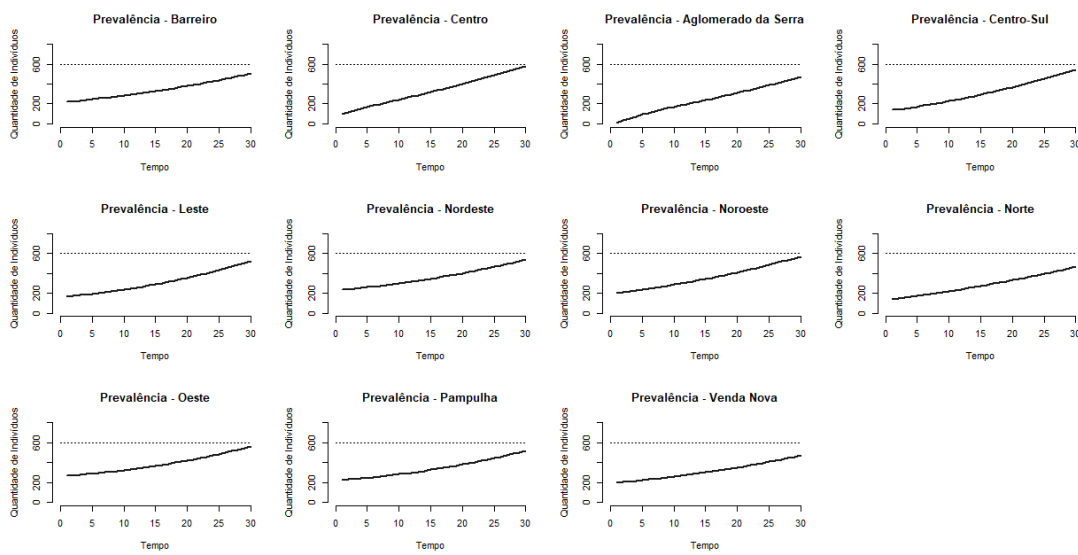


Figura 28: Cenário 2: Prevalência ( $D = 0.8$ )

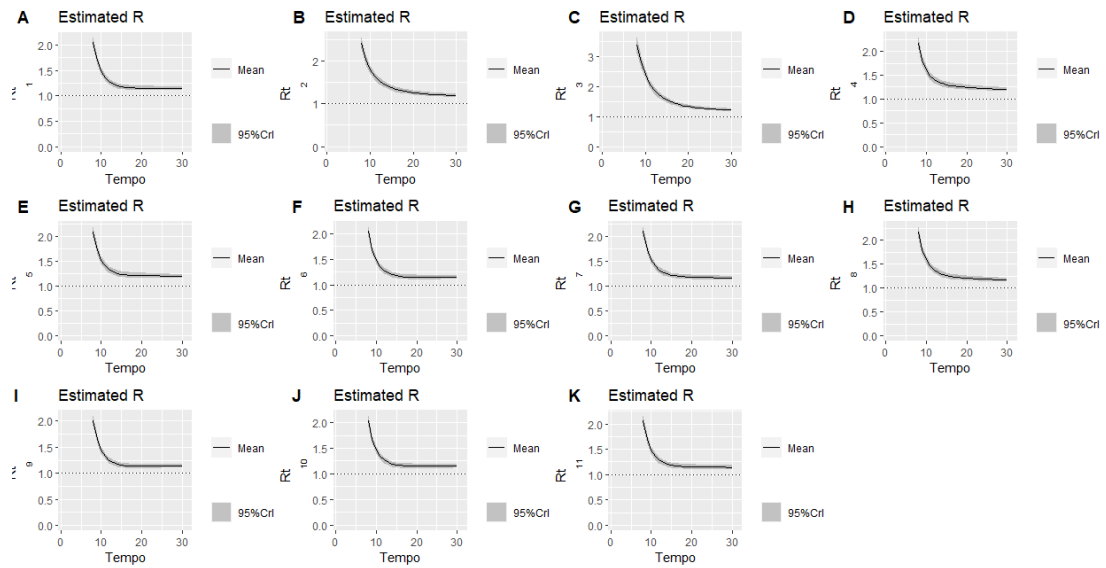


Figura 29: Cenário 2: Estimação  $R_t$  para  $D = 0.8$ . Regiões: **A)** Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

### 6.3.2 Situação 2: $D = 0.5$

A segunda situação considera um índice intermediário de circulação entre as regiões. A Figura 24 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte enquanto a Figura 25 exprime a estimação do  $R_t$ .

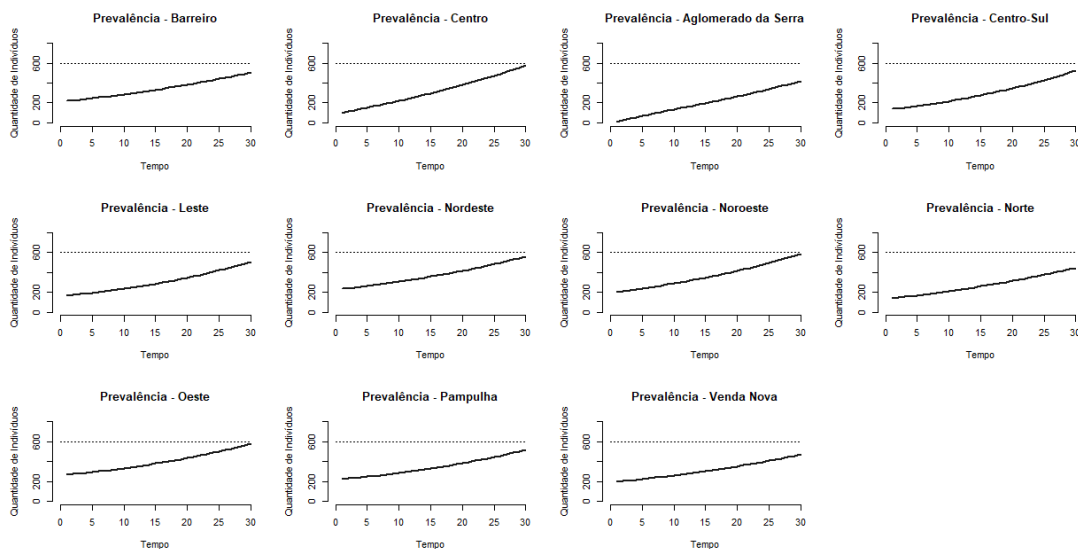


Figura 30: Cenário 2: Prevalência ( $D = 0.5$ )

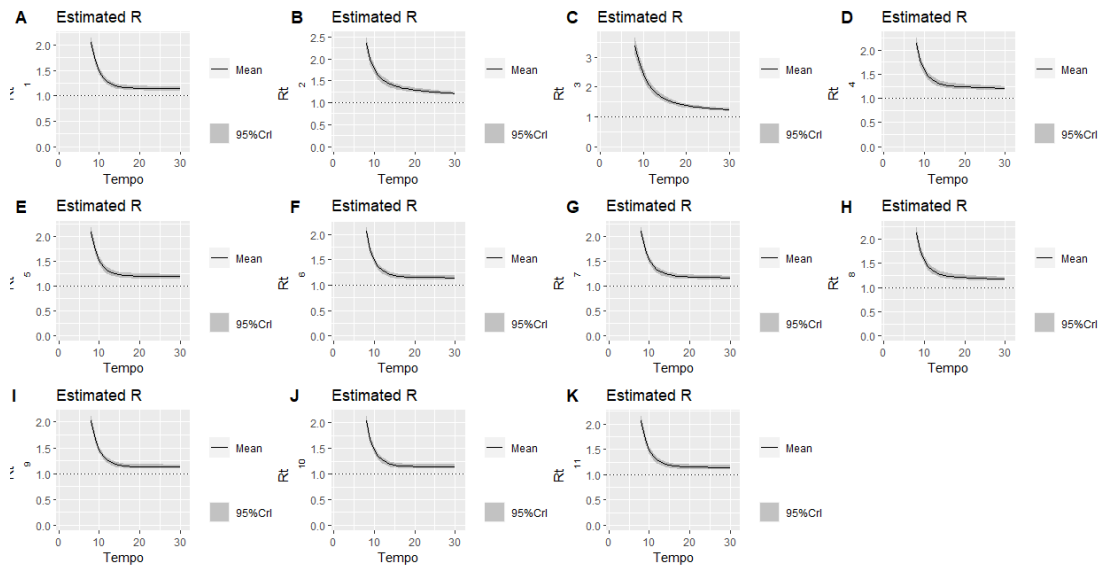


Figura 31: Cenário 2: Estimação  $R_t$  para  $D = 0.5$ . Regiões: **A)** Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

### 6.3.3 Situação 3: $D = 0.1$

A terceira situação considera um baixo índice de mobilidade entre as regiões. A Figura 26 retrata a prevalência do novo coronavírus nas regiões de Belo Horizonte enquanto a Figura 27 exprime a estimação do  $R_t$ .

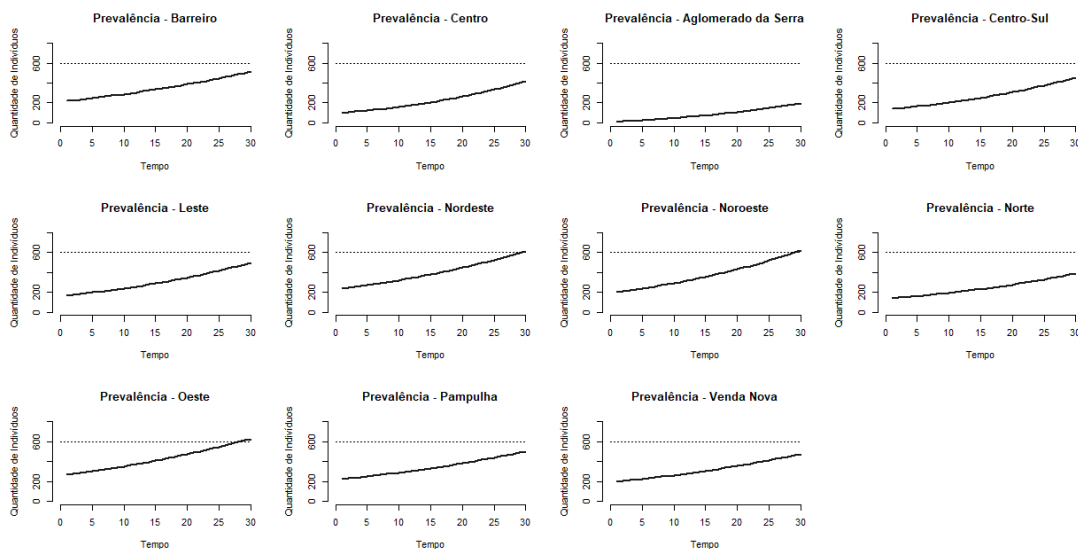


Figura 32: Cenário 2: Prevalência ( $D = 0.1$ )



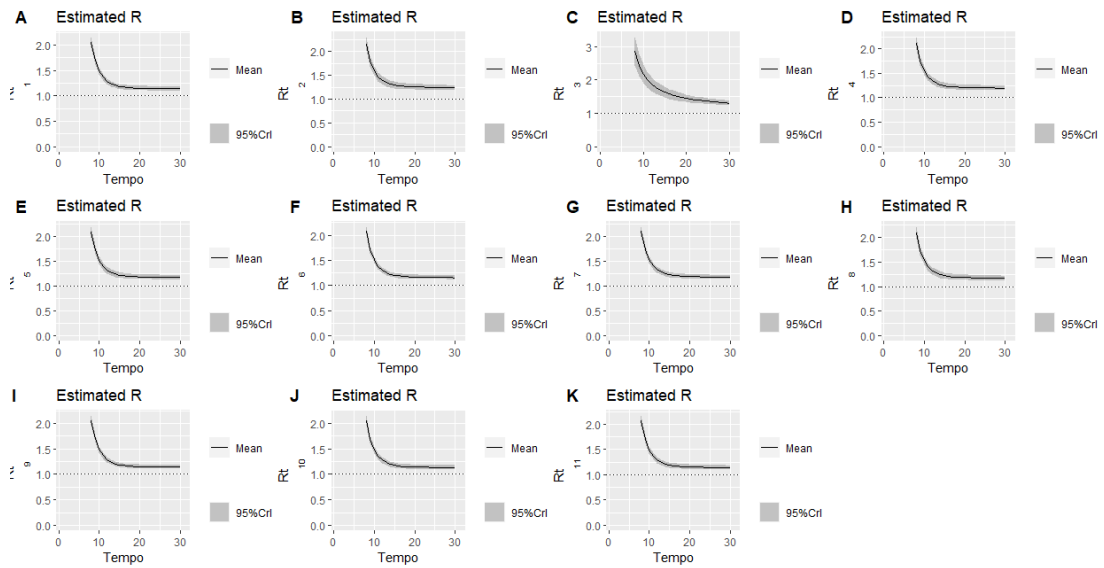


Figura 33: Cenário 2: Estimação  $R_t$  para  $D = 0.1$ . Regiões: **A)** Barreiro, **B)** Centro, **C)** Aglomerado da Serra, **D)** Centro-Sul, **E)** Leste, **F)** Nordeste, **G)** Noroeste, **H)** Norte, **I)** Oeste, **J)** Pampulha, **K)** Venda Nova.

Nota-se que de forma geral, assim como no cenário anterior, os gráficos apresentam comportamentos semelhantes entre as regiões para uma mesma probabilidade de transmissão. Diferentemente do Cenário 1, em que o patógeno é mais contagioso, neste cenário a disseminação do vírus ainda está em expansão, logo, para o período de 30 dias, as curvas de prevalência não chegaram ao seu “pico”, ou seja, o número de infectados ainda está em crescimento.

As regiões Nordeste, Noroeste e Norte, entre todas as regiões e situações neste cenário, foram as únicas a ultrapassarem a marca de 600 infectados (linha pontilhada), na situação 3. E assim como no cenário anterior, o Aglomerado da Serra está entre as regiões que possuíram o menor número de infectados, independentemente do valor do índice de circulação.

É relevante observar também, que assim como no Cenário 1, as flutuações que ocorrem na prevalência do patógeno e no valor do  $R_t$  estimado para cada uma das 11 regiões com a diminuição do valor de  $D$ . Como exemplo, para a região do Centro, um maior índice de mobilidade resulta em um maior número de infectados, esse que vai diminuindo com o decréscimo do índice de circulação. Já para a região Nordeste, essa situação se inverte, a medida que diminui o valor de  $D$ , o número de infectados aumenta. Ademais, no Cenário 2 a curva da estimação do  $R_t$  decai de forma exponencial, o que representa que, em comparação ao cenário anterior, será necessário mais tempo para que a doença se encontre em equilíbrio epidemiológico.

## 6.4 Comparação entre os $R_t$ 's dos Cenários 1 e 2

Os box-plots abaixo evidenciam a estimação do número de reprodução instantâneo ( $R_t$ ) para cada uma das regiões, entre os dias 7 e 21 do período analisado, nos diferentes cenários e situações listados anteriormente. O valor máximo do boxplot é o  $R_t$  estimado para o dia 7 e, o valor mínimo é o  $R_t$  estimado para o dia 21.

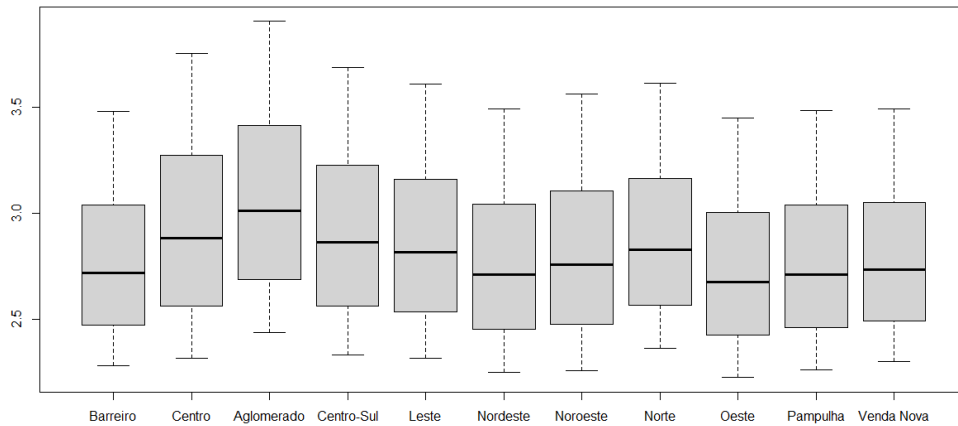


Figura 34: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 1 ( $D = 0,8$ ).

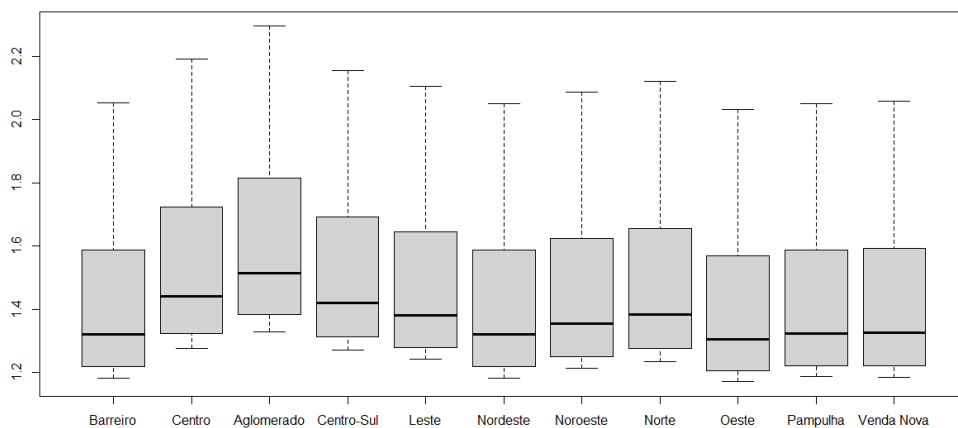


Figura 35: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 2 ( $D = 0,8$ ).

Observa-se que para  $D = 0,8$ , no Cenário 1, em que há a maior taxa de contágio, os valores estimados dos  $R_t$ 's são mais altos se comparados ao Cenário 2. E nota-se ainda que, em ambos cenários, existem leves diferenças entre os  $R_t$ 's estimados por região, com o Aglomerado da Serra apresentando as maiores estimativas.

As Figuras 36 e 37 apresentam os box-plots para os Cenários 1 e 2, respectivamente. Para esses, o índice de mobilidade é igual a 0,5.

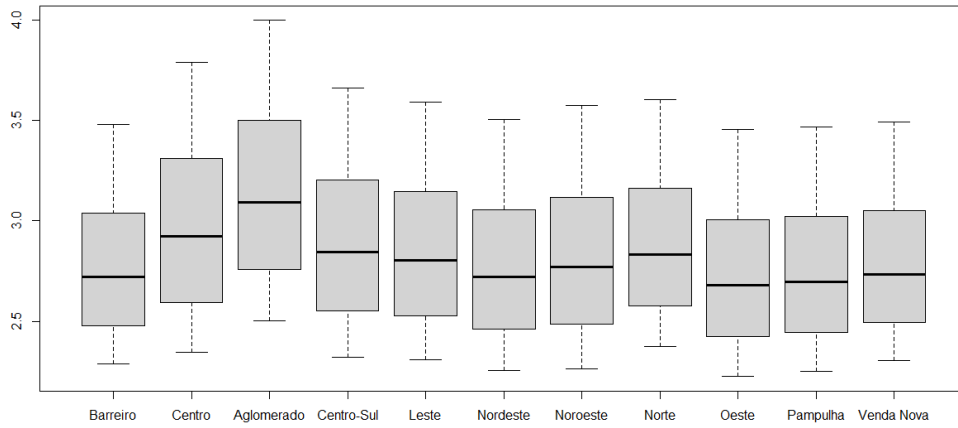


Figura 36: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 1 ( $D = 0,5$ ).

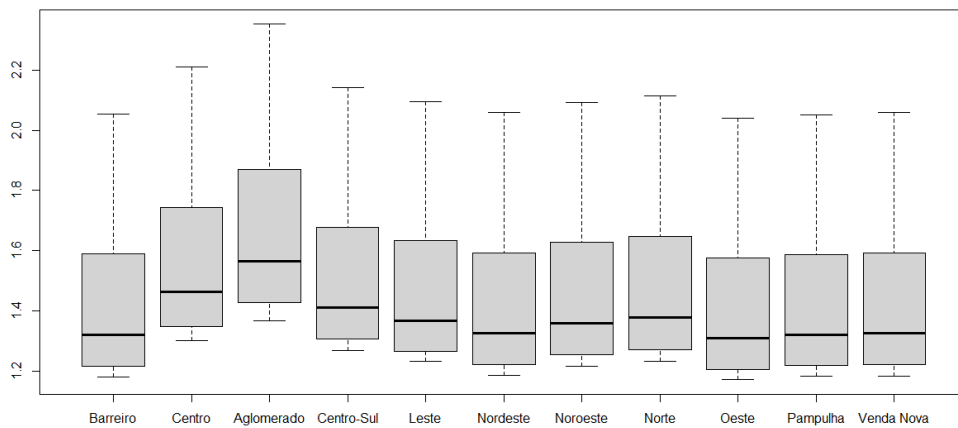


Figura 37: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 2 ( $D = 0,5$ ).

Assim como na situação anterior, existem leves diferenças entre as regiões, sendo mais perceptíveis no primeiro cenário. Ademais, as estimativas do  $R_t$  são mais altas no Cenário 1, em que há maior taxa de contágio.

As Figuras 38 e 39 apresentam os box-plots para os Cenários 1 e 2, respectivamente. Para esses, o índice de mobilidade é igual a 0,1.

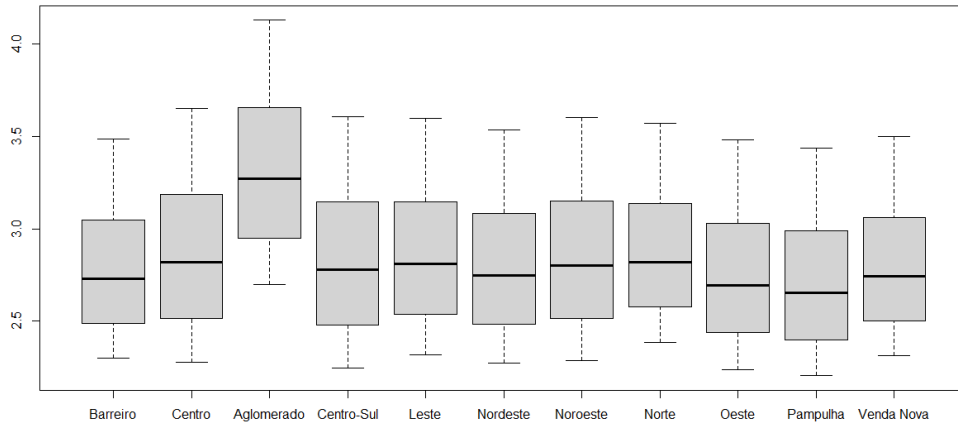


Figura 38: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 1 ( $D = 0,1$ ).

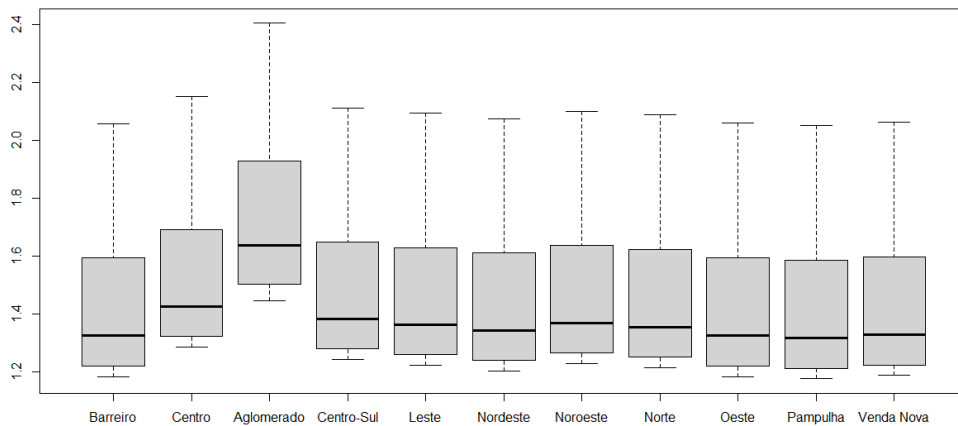


Figura 39: Boxplot  $R_t$  (dia 7 - dia 21): Cenário 2 ( $D = 0,1$ ).

Nota-se que para ambos os cenários, as estimativas do  $R_t$  entre as regiões são mais semelhantes do que nas situações anteriores, com exceção do Aglomerado da Serra, região que aparenta apresentar uma diferença significativa em relação as demais.

Destaca-se que há indícios das mudanças de comportamento do número de reprodução instantâneo por regiões e para diferentes graus de circulação. Estudos mais aprofundados são indicados para que se desenvolva a caracterização e entendimento dos fatores que levam a essas diferenças, e assim, efetivar as ações de políticas públicas para a contenção da disseminação da doença.

## 7 Conclusão

O presente estudo teve por objetivo estimar as taxas de contato dos moradores da região do Aglomerado da Serra, por meio de modelos hierárquicos e, projetá-las para as regiões de Belo Horizonte com o propósito de aplicá-las em um modelo SIR e assim analisar a propagação do novo coronavírus na cidade. Esse ensaio foi embasado no Projeto POLYMOD (Mossong, J. et al., 2008) e no artigo de Prem et al. (2017) e, a metodologia de modelos hierárquicos foi adotada devido ao agrupamento das observações do banco de dados por faixa etária, sendo realizada sob a abordagem Bayesiana em virtude da flexibilidade dos modelos e pela aleatoriedade dos coeficientes.

Para a estimação das taxas, primeiramente foi necessário a calibração da amostra coletada, resultando em um banco de dados representativo à população. A calibração foi realizada utilizando *raking*, de maneira similar a apresentada por Chin et al. (2021). As taxas foram estimadas para diferentes círculos sociais: casa, locais externos ao ambiente domiciliar (nomeado como ‘Outros’) e no contexto geral, em que não há a distinção por círculo. Para os três contextos sociais notou-se a presença de homofilia, ou seja, as médias de contatos tendem a ser maiores entre pessoas com características semelhantes. Observou-se também que, em média, os contatos foram mais realizados com indivíduos das faixas etárias 3 e 4, e, além disso, deve-se atentar ao viés de memória e a não simetria da matriz das taxas.

As taxas do contexto ‘Geral’ foram ainda projetadas para as regionais administrativas de BH, com a divisão da regional Centro-Sul, entre Centro, Aglomerado da Serra e o restante da região. A projeção se deu pelo produto das taxas específicas por grupo de idade do Aglomerado, com a proporção da estrutura etária da região para qual se queria projetar, com a estrutura etária do Aglomerado (Prem et al., 2017). As taxas encontradas são específicas pelos grupos de idade, assim, somou-se as então para uma taxa geral por região. E essa última foi utilizada na aplicação no modelo compartimental SIR, que considera o índice de mobilidade.

Notou-se que a prevalência do patógeno era similar entre algumas regiões, em função da semelhança entre suas estruturas etárias e pela utilização de um índice de mobilidade genérico e comum a todas. Destaca-se também a flutuação que ocorreu no número de infectados por região, com o decréscimo da circulação. Em que, enquanto para algumas um maior índice de mobilização resultou em um maior número de infectados, uma situação que se é esperada, para outras regiões, essa suspeita não se concretizou e, um baixo índice de mobilidade sucedeu em um aumento do número de casos.

Dado que alguns dos modelos selecionados apresentavam algum grau de incerteza em suas especificações, e, alguns do que continham, apresentavam evidências de que a posterioris dos parâmetros estimados não eram confiáveis, pretende-se para trabalhos futuros, realizar novos ajustes com diferentes famílias e prioris, assim como também,

com a inclusão de novas covariáveis a fim de obter uma melhora na estimativa dessas, e procurar compreender quais outros fatores contribuem na explicação dos padrões dos contatos. Concomitantemente, busca-se realizar novas análises da propagação da doença, admitindo as diferenças do comportamento dessa, pelas diferenças das taxas de contatos por faixas etárias e pelos índices de mobilidade específicos por região.

Por fim, ressalta-se a importância de estudos referentes a padrões de mistura para compreender o desenrolar das epidemias e, assim, auxiliar nas diretrizes e tomadas de decisões de enfrentamento, para a melhora da saúde pública.

## Referências

- [1] Aitkin, Murray. (2008). *Applications of the Bayesian Bootstrap in Finite Population Inference*. Journal of official statistics. 24. 21-51.
- [2] Allen, L. J. S. *An Introduction to stochastic models*. Em: *Modeling the interplay between human behavior and the spread of infectious diseases*. Ed. por P. Manfredi e A. d'Onofrio. Springer, pp. 1-19.
- [3] Allen, L. J. S. *An Introduction to Stochastic Processes with Applications to Biology*. 2da.ed. Prentice Hall, 2011.
- [4] Alvarenga, Lucymara de Resende. *Modelagem de epidemias através de modelos baseados em indivíduos*. Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais. 2008.
- [5] Anderson, R. M; May, R. M. *Infectious Disease of Humans*. Nova Iorque: Oxford University Press, 1991.
- [6] Bassanezi, Rodney. *Ensino - aprendizagem com Modelagem matemática*. 2002. isbn = 85-7244-207-3.
- [7] Bettencourt, Luis Ma; Ribeiro, Ruy M. “*Real time bayesian estimation of the epidemic potential of emerging infectious diseases*”. PLoS One, v. 3, n. 5, 2008.
- [8] Bürkner P (2017). “*brms: An R Package for Bayesian Multilevel Models Using Stan*”. Journal of Statistical Software, 80(1), 1–28. doi: 10.18637/jss.v080.i01.
- [9] Bürkner P (2018). “*Advanced Bayesian Multilevel Modeling with the R Package brms*”. The R Journal, 10(1), 395–411. doi: 10.32614/RJ-2018-017.
- [10] Chin, Taylor and Feehan, Dennis M. and Buckee, Caroline O. and Mahmud, Ayesha S. *Contact surveys reveal heterogeneities in age-group contributions to SARS-CoV-2 dynamics in the United States*. 2021. doi: 10.1101/2021.09.25.21264082.
- [11] Cori, Anne; Ferguson, Neil M., Fraser, Christophe; Cauchemez, Simon. *A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics*. American Journal of Epidemiology. Volume 178, Issue 9, 1 November 2013, Pages 1505–1512. <https://doi.org/10.1093/aje/kwt133>
- [12] Cruz, Cláudia Catarina Mendes Silva da. *Modelos multi-nível : fundamentos e aplicações*. Dissertação apresentada na Universidade Aberta para obtenção do grau de Mestre em Matemática, Estatística e Computação (especialização em Estatística Computacional). Lisboa, 2010. Lisboa : [s.n.], 2010. 182 p.. Identificador: <http://hdl.handle.net/10400.2/1729>.

- [13] de Finetti, Bruno. "*Sul significato soggettivo della probabilità.*" *Fundamenta Mathematicae* 17.1 (1931): 298-329. <<http://eudml.org/doc/212523>>.
- [14] de Finetti, B. (1974). *Theory of Probability (Volume 1)*. New York: John Wiley & Sons.
- [15] Deville, Jean-Claude and Särndal, Carl-Erik. *Calibration estimators in survey sampling*. *Journal of the American statistical Association*, 87(418):376–382, 1992.
- [16] Diekmann O. and Heesterbeek J.A.P. "*Mathematical Epidemiology of infectious Diseases: Model Building, Analysis, and Interpretation*". John Wiley and Sons Ltd, England, 2000.
- [17] Downes M, Gurrin LC, English DR, Pirkis J, Currier D, Spittal MJ, Carlin JB. "*Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities From Highly Selected Survey Samples*". *Am J Epidemiol*. 2018 Aug 1;187(8):1780-1790. doi: 10.1093/aje/kwy070. PMID: 29635276.
- [18] Ehlers, R. S. (2011). "*Inferência Bayesiana*". USP, São Paulo.
- [19] Faraway, Julian J. "*Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*". Second edition. ©2016 by Taylor & Francis Group, LLC. Version Date: 20160301.
- [20] Feehan, Dennis M. and Mahmud, Ayesha S. *Quantifying population contact patterns in the united states during the covid-19 pandemic*. *Nature communications*, 12(1):1–9, 2021.
- [21] Gelman, A., & Hill, J. *Data analysis using regression and multilevel/ hierarchical models*. New York: Cambridge University Press. 2007.
- [22] Gelman, A., and T. C. Little. 1997. "*Postratification into Many Categories Using Hierarchical Logistic Regression.*" *Survey Methodology* 23:127–135.
- [23] Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, et al. (2015). *A Household-Based Study of Contact Networks Relevant for the Spread of Infectious Diseases in the Highlands of Peru*. *PLoS ONE* 10(3): e0118457. <https://doi.org/10.1371/journal.pone.0118457>
- [24] Goldstein, Harvey & Carpenter, James Kenward, Michael. (2018). *Bayesian models for weighted data with missing values: a bootstrap approach*. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 67. 10.1111/rssc.12259.
- [25] Hethcote H. "*The mathematics of infectious diseases*". *SIAM Rev.*, 42 (2000), 599–653. doi: 10.1137/S0036144500371907



- [26] Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD, et al. (2011) *Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases*. PLoS ONE 6(2): e16965. <https://doi.org/10.1371/journal.pone.0016965>
- [27] Hox, J.J., Moerbeek, M., Schoot, R.V.de. *Multilevel Analysis: Techniques and Applications*. Quantitative Methodology. 3rd Edition, Routledge. (isbn = 1138121363,9781138121362). 2017.
- [28] Jarvis, C.I., Van Zandvoort, K., Gimma, A. et al. *Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK*. BMC Med 18, 124 (2020). <https://doi.org/10.1186/s12916-020-01597-8>
- [29] Jenness SM, Goodreau SM, Morris M (2018). “*EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks*.”. Journal of Statistical Software, 84(8), 1–47. doi: 10.18637/jss.v084.i08.
- [30] Kiss, J.C. Miller e P.L. Simon (2017). *Mathematics of epidemics on networks: from exact to approximate models*. Vol. 46. Interdisciplinary Applied Mathematics. Springer, Cham, pp. viii+413. MR:3644065.
- [31] Klepac, Petra and Kucharski, Adam J and Conlan, Andrew JK and Kissler, Stephen and Tang, Maria L and Fry, Hannah and Gog, Julia R. *Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project*. 2020. Cold Spring Harbor Laboratory Press. medRxiv. elocation-id: 2020.02.16.20023754. doi: 10.1101/2020.02.16.20023754. <https://www.medrxiv.org/content/early/2020/03/05/2020.02.16.20023754.full.pdf>.
- [32] Keeling, Matt J., Rohani, Pejman. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press. 1<sup>a</sup> Ed. 2007. ISBN: 978-0-691-11617-4.
- [33] Kermack, W.O., McKendrick, A.G. *Contribuições para a teoria matemática das epidemias*. Proc. Royal Soc. A, 115, 700-721 (1927).
- [34] Li R., Pei S., Chen B., Song Y., Zhang T., Yang W., Shaman J. “*Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)*”. Science. 2020 May 1;368(6490):489-493. doi: 10.1126/science.abb3221. Epub 2020 Mar 16. PMID: 32179701; PMCID: PMC7164387.
- [35] Loschi, Rosangela H., and Sergio Wechsler. “*COHERENCE, BAYES’S THEOREM AND POSTERIOR DISTRIBUTIONS*”. Brazilian Journal of Probability and Statistics, vol. 16, no. 2, [Brazilian Statistical Association, Institute of Mathematical Statistics], 2002, pp. 169–85, <http://www.jstor.org/stable/43601015>.

- [36] Lumley, T. (2004). *Analysis of Complex Survey Samples*. Journal of Statistical Software, 9(8), 1–19. <https://doi.org/10.18637/jss.v009.i08>.
- [37] Lumley, T. *Complex surveys: A guide to analysis using R*. 2010. Wiley Series in Survey Methodology. ISBN = 9780470284308; 0470284307.
- [38] Migon, Helio S., Souza, Aparecida D.P., Schmidt, Alexandra M. *Modelos Hierárquicos e Aplicações*. 2008. 18<sup>o</sup> SINAPE. ABE - Associação Brasileira de Estatística.
- [39] Mohebbi, M., Wolfe, R. Jolley, D. *A poisson regression approach for modelling spatial autocorrelation between geographically referenced observations*. BMC Med Res Methodol 11, 133 (2011). <https://doi.org/10.1186/1471-2288-11-133>
- [40] Marchesin, Dan; López-Flores, Marlon M.; Matos, Vítor; Schecter, Stephen. *"Equações diferenciais e modelos epidemiológicos"*. Editora do IMPA. 2021. ISBN: 978-65-89124-42-9.
- [41] Mercado Londoño, Sergio Luis. *Estimação do número de reprodução basal em modelos compartimentais*. Campinas, SP : [s.n.], 2014. Orientador: Luiz Koodi Hotta. Dissertação (Mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.
- [42] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. *Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases*. PLoS Med. 2008; 5: e74. <https://doi.org/10.1371/journal.pmed.0050074> PMID: 18366252
- [43] Nishiura, Hiroshi; Linton, Natalie M.; Akhmetzhanov, Andrei R. *Serial interval of novel coronavirus (COVID-19) infections*. International journal of infectious diseases, 2020.
- [44] Olsson, H., de Bruin, W. B., Galesic, M., Prelec, D. *Combining survey questions with a Bayesian bootstrap method yields accurate election forecasts*. 2021. <https://doi.org/10.31219/osf.io/nqcg8>
- [45] Park, D., Gelman, A., Bafumi, J. (2017). *Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls*. Political Analysis, 12(4), 375-385. doi:10.1093/pan/mpw024
- [46] Pujari, Bhalchandra S. and Shekatkar, Sneha. *"Multi-city modeling of epidemics using spatial networks: Application to 2019-nCov (COVID-19) coronavirus in India"*. medRxiv 2020.03.13.20035386. <https://doi.org/10.1101/2020.03.13.20035386>.

- [47] Prem K, Cook AR, Jit M (2017) . *Projecting social contact matrices in 152 countries using contact surveys and demographic data*. PLOS Computational Biology 13(9): e1005697. <https://doi.org/10.1371/journal.pcbi.1005697>
- [48] Prem K, Zandvoort, Kevin van; Klepac P.; Eggo, Rosalind M; Davies N.; CMMID COVID-19 working group, Cook, Alex R & Jit M. "*Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era*". 2020. Disponível em: <https://cmmid.github.io/topics/covid19/synthetic-contact-matrices.html>.
- [49] Rocha, Alex Luiz Martins Matheus da. *Modelos lineares mistos: uma abordagem bayesiana*. 2017. viii, 49 f., il. Dissertação (Mestrado em Estatística)—Universidade de Brasília, Brasília, 2017. <https://repositorio.unb.br/handle/10482/31708>
- [50] Rubin, Donald B. *The Bayesian Bootstrap*. The Annals of Statistics, vol. 9, no. 1. Institute of Mathematical Statistics, 1981, pp. 130–34. <http://www.jstor.org/stable/2240875>.
- [51] Ruiz, Charles Martín Miguel; Silva, Pedro Luis do Nascimento. "*Explorando alternativas para a calibração dos pesos amostrais da Pesquisa Nacional por Amostra de Domicílios*". Trabalho apresentado no VI Congreso de la Asociación Latinoamericana de Población, realizado em Lima-Peru, de 12 a 15 de agosto de 2014.
- [52] Storopoli (2021, Aug. 1). *Estatística Bayesiana com R e Stan: Modelos Multiníveis Bayesianos*. Obtido em [https://storopoli.io/Estatistica-Bayesiana/10-Regressao\\_Multinivel.html](https://storopoli.io/Estatistica-Bayesiana/10-Regressao_Multinivel.html).
- [53] Thompson, R. N. et al. *Improved inference of time-varying reproduction numbers during infectious disease outbreaks*. Epidemics, v. 29, p. 100356, 2019.
- [54] Vehtari, Aki and Gelman, Andrew and Gabry, Jonah. *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*. 2016. Statistics and Computing. Springer Science and Business Media LLC. Vol 27. pages 1413–1432. ISSN: 1573-1375. DOI: 10.1007/s11222-016-9696-4. <http://dx.doi.org/10.1007/s11222-016-9696-4>.
- [55] Vehtari, Aki; Simpson, Daniel; Gelman, Andrew; Yao, Yuling; Gabry, Jonah. *Pareto Smoothed Importance Sampling*. 2021. <https://arxiv.org/abs/1507.02646>.
- [56] Wallinga, Jacco; Lipsitch, Marc. *How generation intervals shape the relationship between growth rates and reproductive numbers*. Proceedings of the Royal Society B: Biological Sciences, v. 274, n. 1609, p. 599-604, 2007.

## 8 Anexo 1

### Comparações entre os bancos

Os gráficos e tabelas a seguir retratam os quantitativos e proporções, respectivamente, das covariáveis utilizadas para a calibração do banco de dados coletado.

A Figura 40 apresenta o quantitativo de indivíduos por faixa etária e sexo para os três bancos de dados. O gráfico **A** se refere ao levantamento das situações de contato na região do Aglomerado (Banco 1), o gráfico **B** à amostra coletada do perfil sociodemográfico que foi utilizada para o cálculo dos pesos (Banco 2) e, o gráfico **C** à amostra calibrada (Banco calibrado).

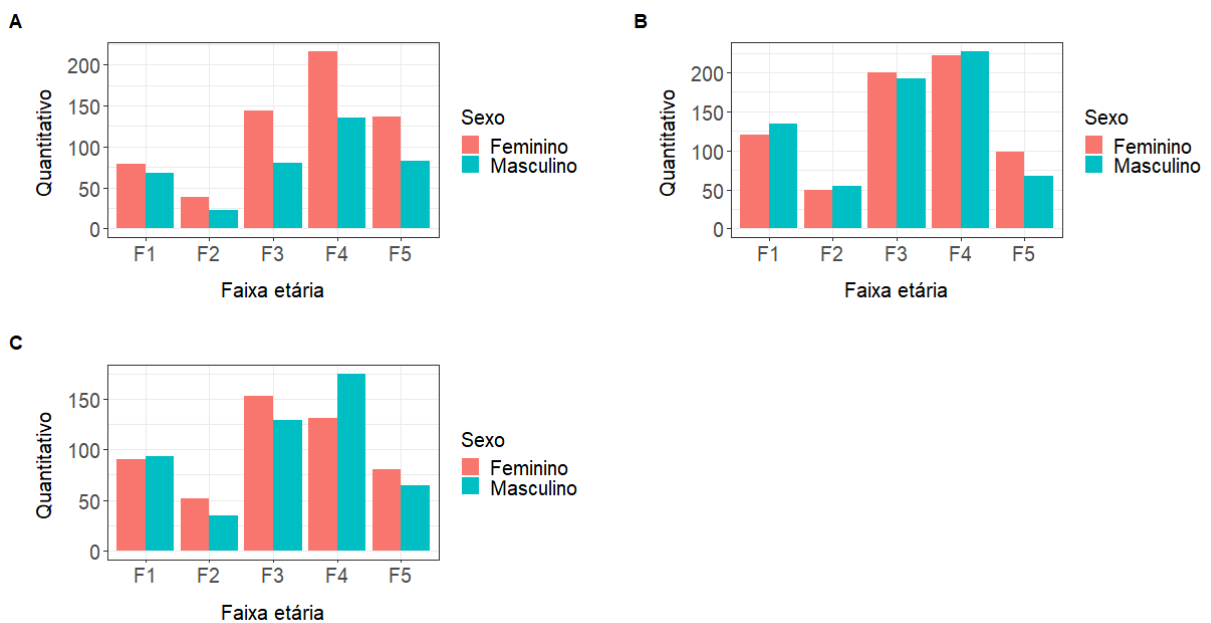


Figura 40: Quantitativo de indivíduos por faixa etária e sexo: A) Banco 1, B) Banco 2 e C) Banco calibrado.

Nota-se que a o número de mulheres parece ser consideravelmente maior no Banco 1 se compararmos aos bancos 2 e calibrado. Percebe-se também que a maioria das pessoas se concentram nas faixas etárias 3 e 4 para os três bancos. Além disso, é possível observar que os gráficos **B** e **C** são bem semelhantes, o que sugere uma calibração adequada do conjunto de dados.

As Tabelas 35 e 36 descrevem a proporção de indivíduos por sexo e faixa etária.

Tabela 35: Proporção por sexo (%) - Comparação.

Sexo	Banco 1	Banco 2	Banco calibrado
Feminino	61,24	50,51	50,45
Masculino	38,76	49,49	49,55
Total	100	100	100

Tabela 36: Proporção da estrutura etária (%) - Comparação.

Faixa etária	Banco 1	Banco 2	Banco calibrado
<i>F1</i>	14,77	18,67	18,28
<i>F2</i>	6,03	7,61	8,59
<i>F3</i>	22,31	28,70	28,17
<i>F4</i>	34,97	32,87	30,57
<i>F5</i>	21,91	12,15	14,39
Total	100	100	100

A Tabela 35 confirma a suspeita de que no Banco 1, o maior percentual de pessoas declarou ser do sexo feminino (61,24%); e a Tabela 36 que a maior parte das pessoas estão centralizadas nas faixas etárias 3 e 4. Para mais, ambas tabelas corroboram a boa calibração dos dados para essas covariáveis.

As Tabela 37 e Figura 41 apresentam a proporção e o quantitativo do número de pessoas por domicílio.

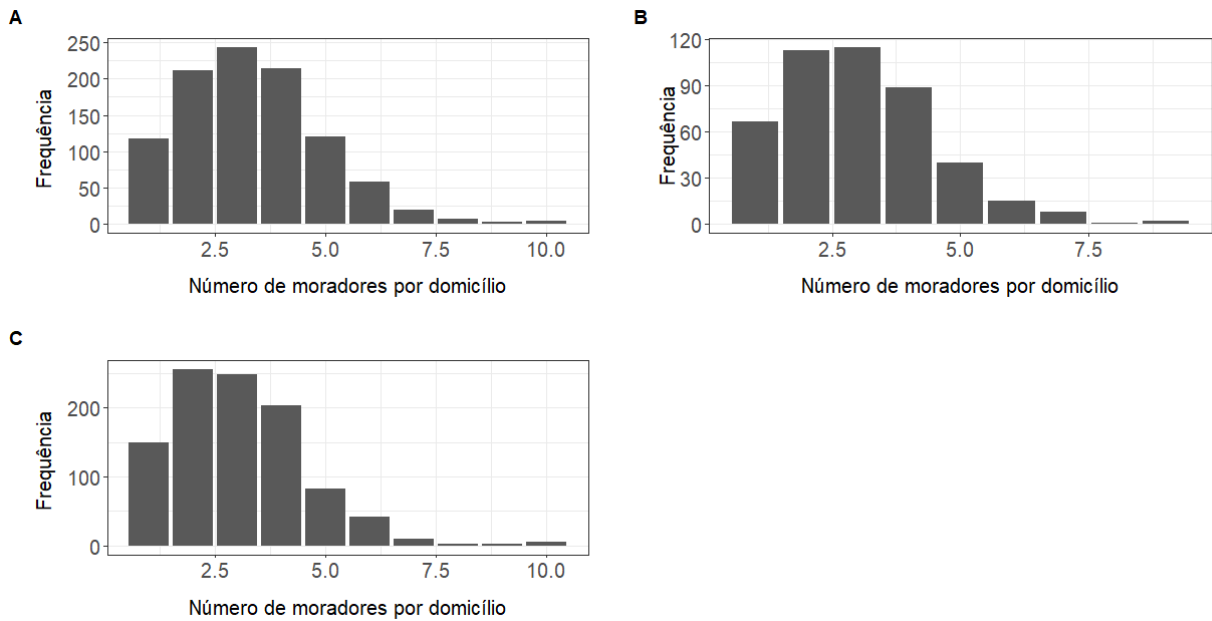


Figura 41: Frequência do número de moradores por domicílio: A) Banco 1, B) Banco 2 e C) Banco calibrado.

Tabela 37: Proporção por tamanho do domicílio (%) - Comparação.

Tamanho do domicílio	Banco 1	Banco 2	Banco calibrado
1	11,69	14,89	14,89
2	21,18	25,11	25,57
3	24,28	25,56	24,78
4	21,38	19,78	20,28
5	12,09	8,89	8,29
6	5,79	3,33	4,20
7 ou mais	3,60	2,44	2,00
Total <sup>7</sup>	100*	100	100*

Para os três bancos, os domicílios tendem a ter entre 2 e 4 moradores, não existindo diferenças expressivas entre eles. Para esta covariável também há indícios de uma boa calibração.

As Figura 42 e Tabela expressam a frequência do número de moradores por bairro para os bancos 1 (A) e calibrado (C), e a frequência do número de domicílios por bairro para o Banco 2 (B).

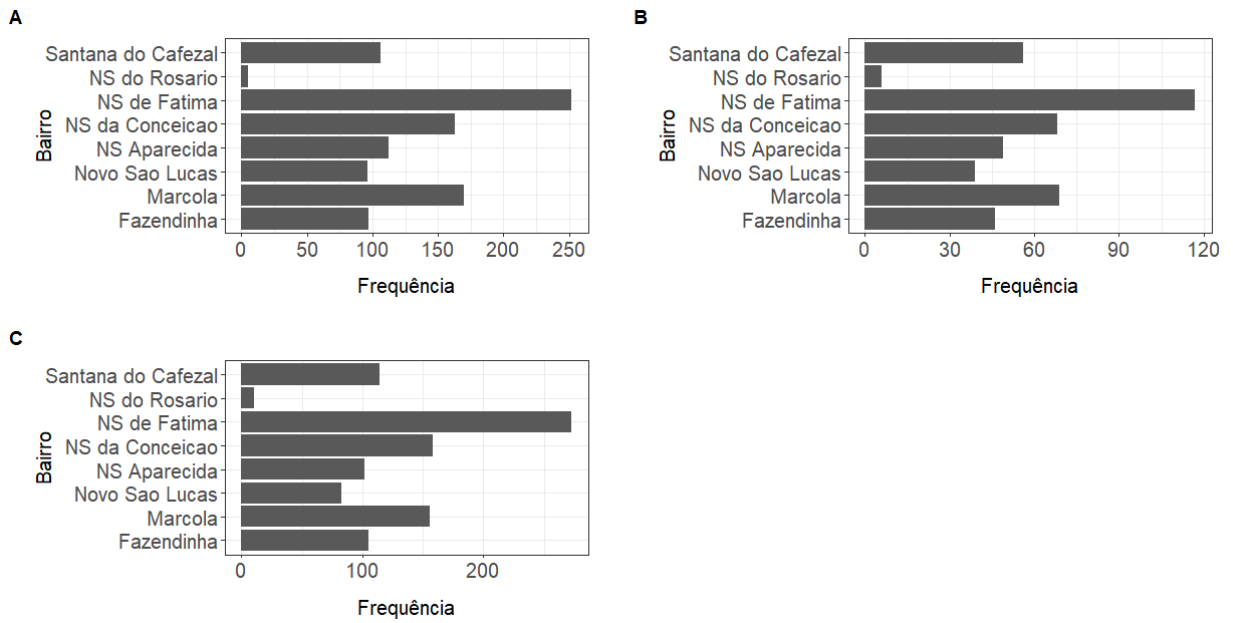


Figura 42: Frequência do número de moradores por bairro: A) Banco 1 e C) Banco calibrado. Frequência do número de domicílios por bairro: B) Banco 2.

Tabela 38: Proporção por bairro (%) - Comparação.

Sexo	Banco 1	Banco 2	Banco calibrado
Vila Fazendinha	9,69	10,22	10,49
Marçola	16,98	15,33	15,58
Novo São Lucas	9,59	8,67	8,29
NS Aparecida	11,19	10,89	10,19
NS da Conceição	16,28	15,11	15,78
NS de Fátima	25,17	26,0	27,27
NS do Rosário	0,50	1,33	1,00
Santana do Cafezal	10,59	12,44	11,39
Total <sup>8</sup>	100*	100*	100*

Podemos perceber que para os três bancos a proporção dos indivíduos/ domicílios por bairro são bem semelhantes, sendo também, essa covariável um indicativo da boa calibração.

De forma geral, podemos assumir que a partir dos pesos gerados pela proporção do Banco 2, obtivemos uma calibração satisfatória dos dados do Banco 1, gerando um banco representativo para as análises.

## 9 Anexo 2

### Modelos ajustados

```

# library(brms)
#set.seed(13)

# modelos para o círculo social: casa -----

# familia Poisson
casa1 <- brm(data = casa,
             family = poisson(),
             formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
             prior = c(prior(normal(3,2), class = Intercept),
                       prior(normal(3,2), class = b),
                       prior(gamma(1,1), class = sd)),
             iter = 4000, warmup = 2000, chains = 3, cores = 3,
             control = list(adapt_delta = .975, max_treedepth = 20))

casa2 <- brm(data = casa,
             family = poisson(),
             formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
             prior = c(prior(normal(3,2), class = Intercept),
                       prior(normal(3,2), class = b),
                       prior(gamma(0.1,0.1), class = sd)),
             iter = 4000, warmup = 2000, chains = 3, cores = 3,
             control = list(adapt_delta = .975, max_treedepth = 20))

#4divergencias

# família Binomial Negativa

casa3 <- brm(data = casa,
             family = negbinomial(),
             formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
             prior = c(prior(normal(3,2), class = Intercept),
                       prior(normal(3,2), class = b),
                       prior(gamma(1,1), class = sd)),
             iter = 4000, warmup = 2000, chains = 3, cores = 3,

```



```

control = list(adapt_delta = .975, max_treedepth = 20))

casa4 <- brm(data = casa,
  family = negbinomial(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(0.1,0.1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .975, max_treedepth = 20))
# 1 transicao divergente

# familia Poisson Inflada de Zeros
casa5 <- brm(data = casa,
  family = zero_inflated_poisson(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(1,1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .975, max_treedepth = 20))

casa6 <- brm(data = casa,
  family = zero_inflated_poisson(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(0.1,0.1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .975, max_treedepth = 20))
# 2 transicoes divergentes

##### adicionando a covariavel referente ao tamanho do domicilio
# familia Poisson

casa7 <- brm(data = casa,
  family = poisson(),

```

```

formula = cont ~ fx.etaria_id*faixa.etaria +
tam_domicilio + (1 | id) ,
prior = c(prior(normal(3,2), class = Intercept),
          prior(normal(3,2), class = b),
          prior(gamma(0.1,0.1), class = sd)),
iter = 4000, warmup = 2000, chains = 3, cores = 3,
control = list(adapt_delta = .975, max_treedepth = 20))

casa8 <- brm(data = casa,
             family = poisson(),
             formula = cont ~ fx.etaria_id*faixa.etaria +
tam_domicilio + (1 | id) ,
             prior = c(prior(normal(3,2), class = Intercept),
                       prior(normal(3,2), class = b),
                       prior(gamma(1,1), class = sd)),
             iter = 4000, warmup = 2000, chains = 3, cores = 3,
             control = list(adapt_delta = .975, max_treedepth = 20))

# família Binomial Negativa
casa9 <- brm(data = casa,
             family = negbinomial(),
             formula = cont ~ fx.etaria_id*faixa.etaria +
tam_domicilio + (1 | id),
             prior = c(prior(normal(3,2), class = Intercept),
                       prior(normal(3,2), class = b),
                       prior(gamma(0.1,0.1), class = sd)),
             iter = 4000, warmup = 2000, chains = 3, cores = 3,
             control = list(adapt_delta = .975, max_treedepth = 20))

casa10 <- brm(data = casa,
              family = negbinomial(),
              formula = cont ~ fx.etaria_id*faixa.etaria +
tam_domicilio + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(1,1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

```

```

# familia Poisson Inflada de Zeros
casa11 <- brm(data = casa,
              family = zero_inflated_poisson(),
              formula = cont ~ fx.etaria_id*faixa.etaria +
              tam_domicilio + (1 | id) ,
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(0.1,0.1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

casa12 <- brm(data = casa,
              family = zero_inflated_poisson(),
              formula = cont ~ fx.etaria_id*faixa.etaria +
              tam_domicilio + (1 | id) ,
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(1,1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

#-----

# validacao e comparacao de modelos

loo1 <- loo(casa1, save_psis = TRUE)
loo2 <- loo(casa2, save_psis = TRUE)
loo3 <- loo(casa3, save_psis = TRUE)
loo4 <- loo(casa4, save_psis = TRUE)
loo5 <- loo(casa5, save_psis = TRUE)
loo6 <- loo(casa6, save_psis = TRUE)
loo7 <- loo(casa7, save_psis = TRUE)
loo8 <- loo(casa8, save_psis = TRUE)
loo9 <- loo(casa9, save_psis = TRUE)
loo10 <- loo(casa10, save_psis = TRUE)
loo11 <- loo(casa11, save_psis = TRUE)
loo12 <- loo(casa12, save_psis = TRUE)

```

```

loo_compare(loo2, loo7, loo9, loo10, loo11, loo12)

#####

# modelos para o círculo social: outros -----
# familia Poisson
outros1 <- brm(data = outros,
              family = poisson() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(0.1,0.1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

outros2 <- brm(data = outros,
              family = poisson() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(1,1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

# binomial negativa -----

outros3 <- brm(data = outros,
              family = negbinomial() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(0.1,0.1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

# transicoes divergentes

```

```

outros4 <- brm(data = outros,
              family = negbinomial() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(1,1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

```

```
# zero inflada -----
```

```

outros5 <- brm(data = outros,
              family = zero_inflated_poisson() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(0.1,0.1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

```

```
# transicao divergente
```

```

outros6 <- brm(data = outros,
              family = zero_inflated_poisson() ,
              formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id),
              prior = c(prior(normal(3,2), class = Intercept),
                        prior(normal(3,2), class = b),
                        prior(gamma(1,1), class = sd)),
              iter = 4000, warmup = 2000, chains = 3, cores = 3,
              control = list(adapt_delta = .975, max_treedepth = 20))

```

```
#-----
```

```
# validacao e comparacao de modelos
```

```

loo1 <- loo(outros1, save_psis = TRUE)
loo2 <- loo(outros2, save_psis = TRUE)
loo3 <- loo(outros3, save_psis = TRUE)
loo31 <- loo(outros31, save_psis = TRUE)
loo4 <- loo(outros4, save_psis = TRUE)
loo5 <- loo(outros5, save_psis = TRUE)

```



```

iter = 4000, warmup = 2000, chains = 3, cores = 3,
control = list(adapt_delta = .975, max_treedepth = 20))

geral4 <- brm(data = banco_final,
  family = negbinomial(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(0.1,0.1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .99, max_treedepth = 20))
#transicao diverg

# familia Poisson Inflada de Zeros
20))
geral5 <- brm(data = banco_final,
  family = zero_inflated_poisson(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(1,1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .975, max_treedepth = 20))

geral6 <- brm(data = banco_final,
  family = zero_inflated_poisson(),
  formula = cont ~ fx.etaria_id*faixa.etaria + (1 | id) ,
  prior = c(prior(normal(3,2), class = Intercept),
    prior(normal(3,2), class = b),
    prior(gamma(0.1,0.1), class = sd)),
  iter = 4000, warmup = 2000, chains = 3, cores = 3,
  control = list(adapt_delta = .975, max_treedepth = 20))
#18 transicoes divergentes

#-----

# validacao e comparacao de modelos

```

```
loo1 <- loo(geral1, save_psis = TRUE)
loo2 <- loo(geral2, save_psis = TRUE)
loo3 <- loo(geral3, save_psis = TRUE)
loo4 <- loo(geral4, save_psis = TRUE)
loo5 <- loo(geral5, save_psis = TRUE)
loo6 <- loo(geral6, save_psis = TRUE)
```

```
loo_compare(loo1, loo3, loo5)
```



## 10 Anexo 3

### Modelo ‘casa7’

As figuras a seguir apresentam os diagnósticos de convergência do modelo ‘casa7’, feitos das marginais e verificação preditiva do modelo.

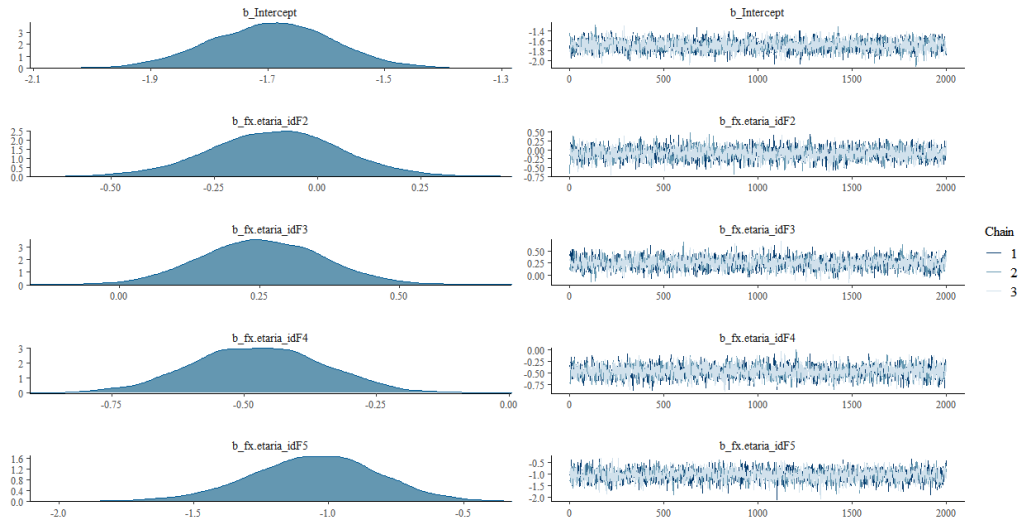


Figura 43: Diagnóstico de convergência (1) - Modelo ‘casa7’.

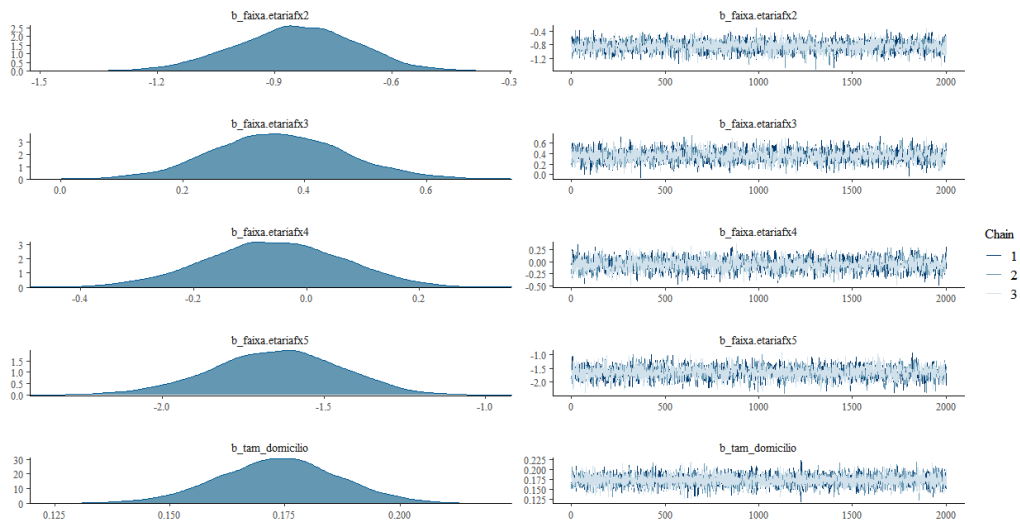


Figura 44: Diagnóstico de convergência (2) - Modelo ‘casa7’.

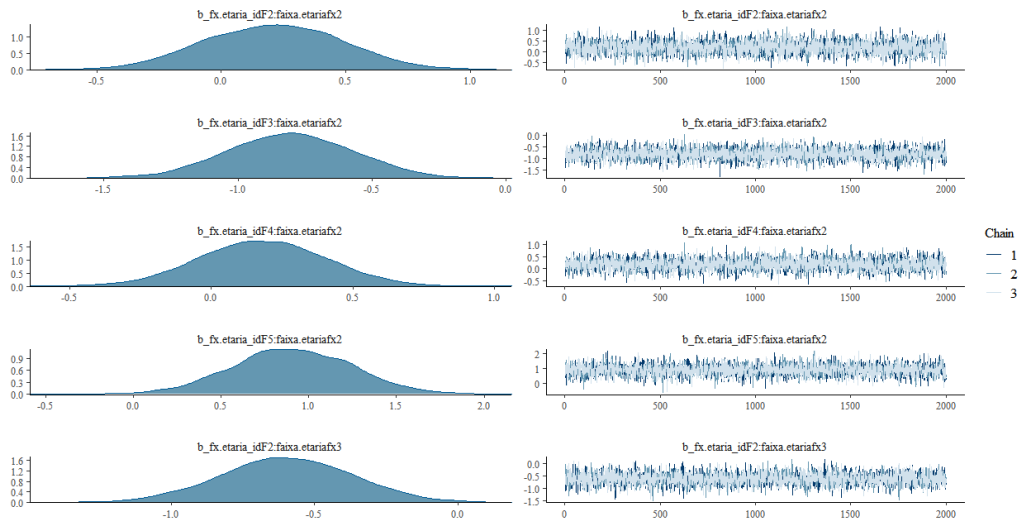


Figura 45: Diagnóstico de convergência (3) - Modelo ‘casa7’.

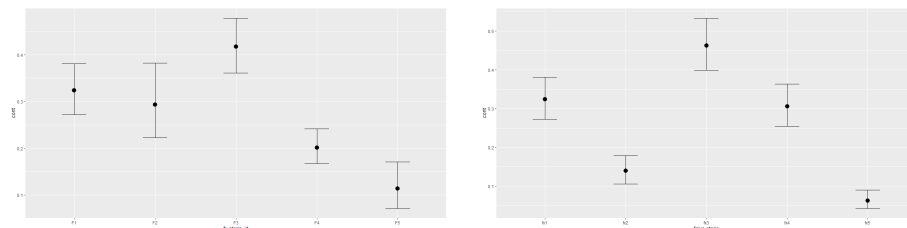


Figura 49: Efeitos condicionais das marginais - Modelo ‘casa7’.

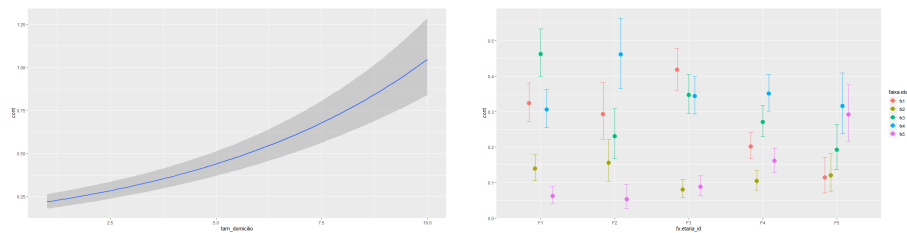


Figura 50: Efeitos marginais de interação e da variável tamanho do domicílio - Modelo ‘casa7’.

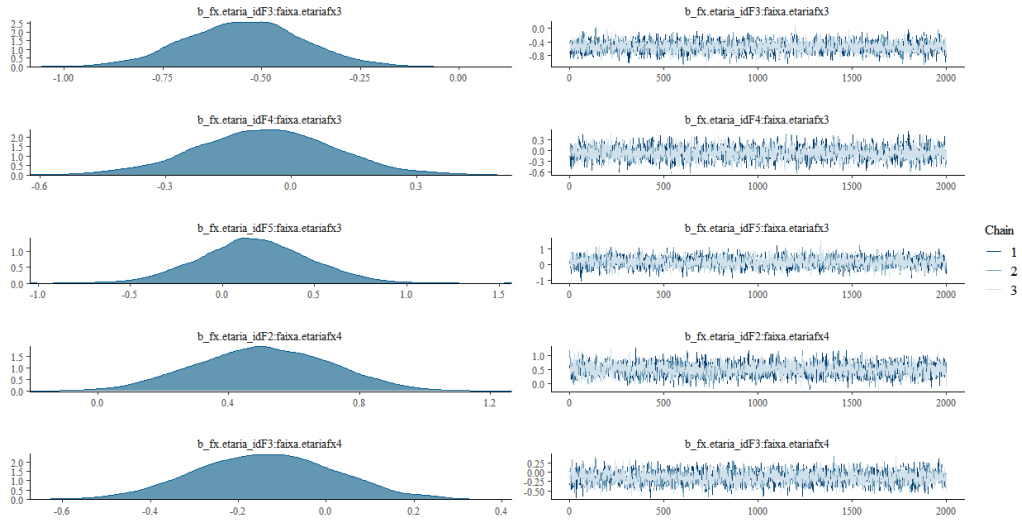


Figura 46: Diagnóstico de convergência (4) - Modelo ‘casa7’.

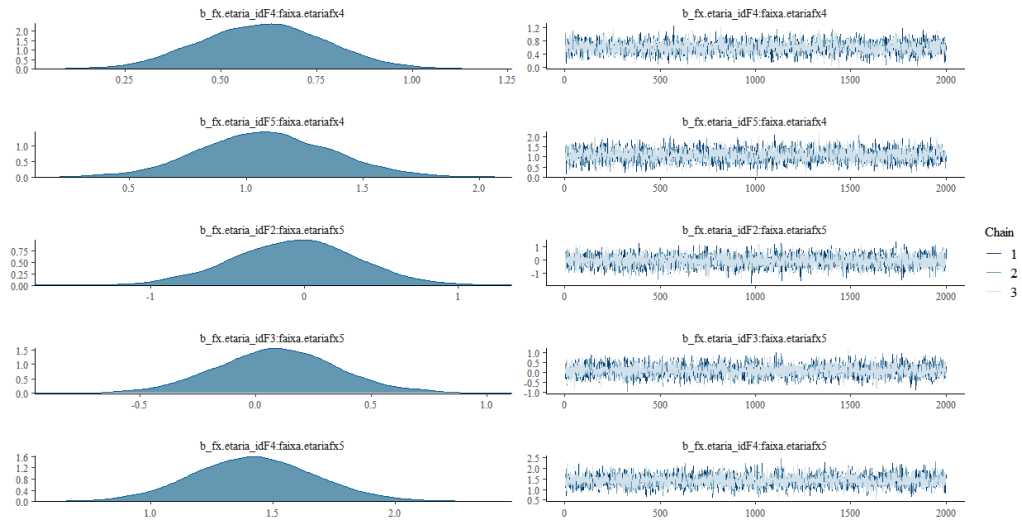


Figura 47: Diagnóstico de convergência (5) - Modelo ‘casa7’.

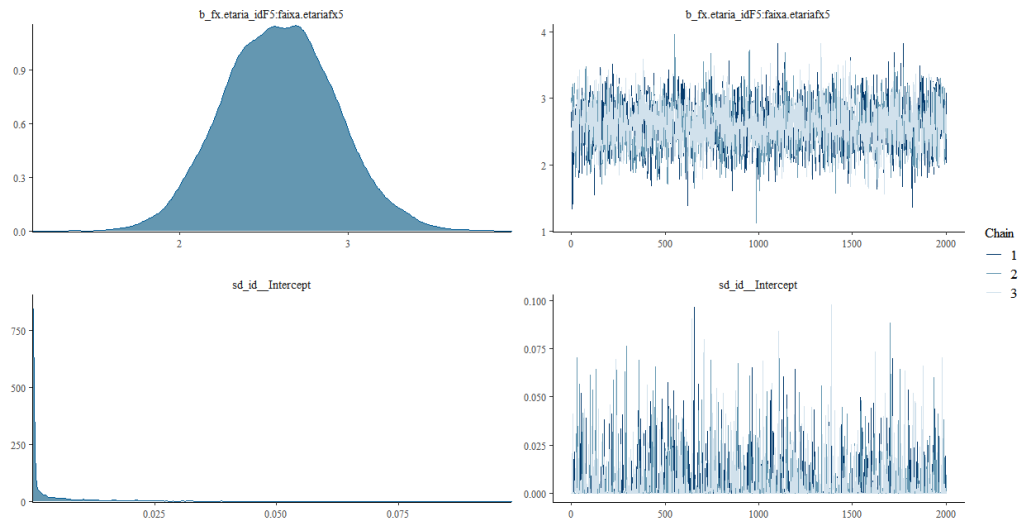


Figura 48: Diagnóstico de convergência (6) - Modelo 'casa7'.

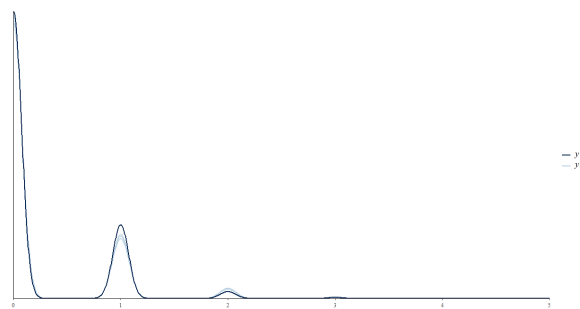


Figura 51: Verificação preditiva a posteriori - Modelo 'casa7'.

## 11 Anexo 4

### Modelo ‘outros4’

As figuras a seguir apresentam os diagnósticos de convergência do modelo ‘outros4’, feitos das marginais e verificação preditiva do modelo.

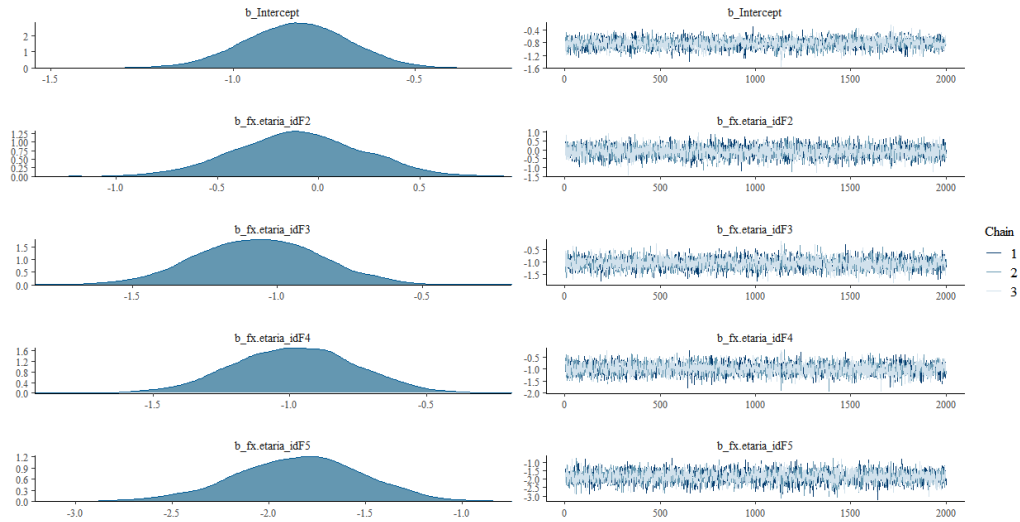


Figura 52: Diagnóstico de convergência (1) - Modelo ‘outros4’.

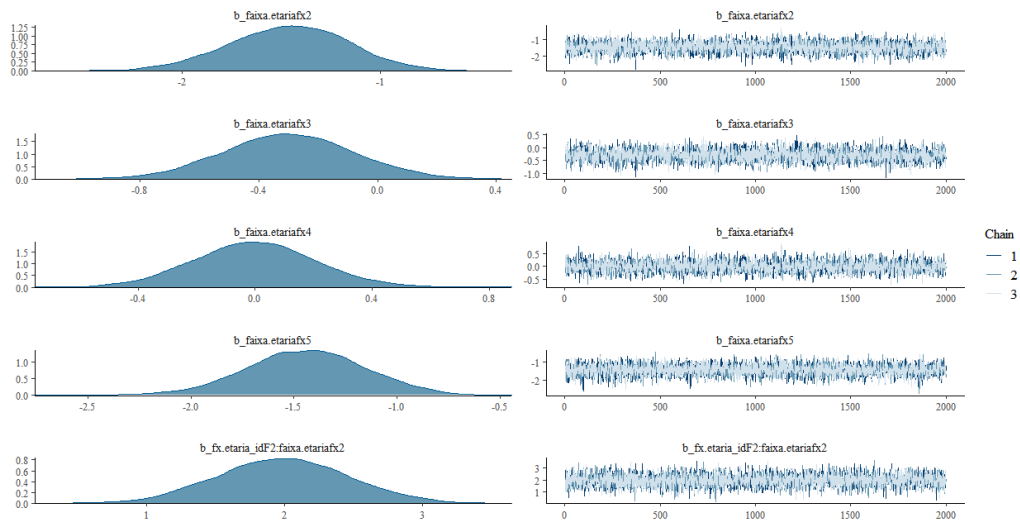


Figura 53: Diagnóstico de convergência (2) - Modelo ‘outros4’.

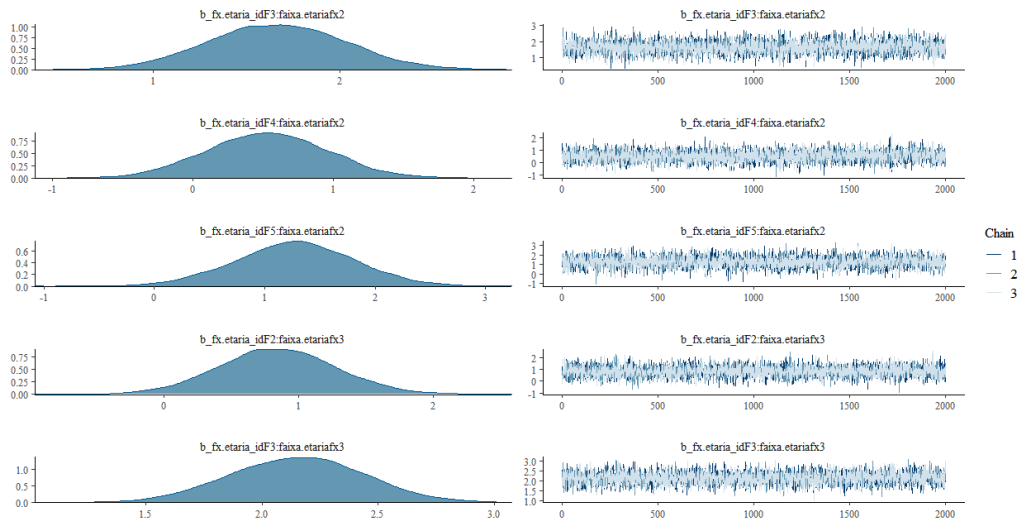


Figura 54: Diagnóstico de convergência (3) - Modelo ‘outros4’.

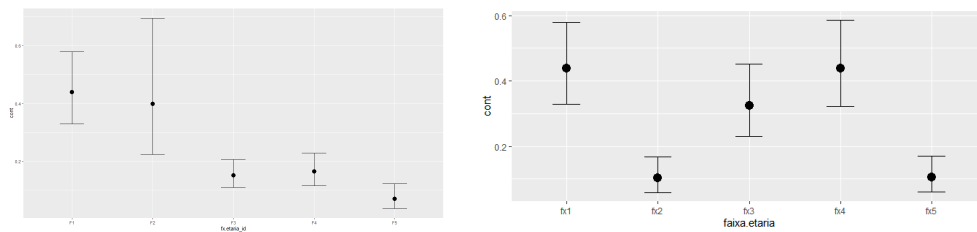


Figura 58: Efeitos condicionais das marginais - Modelo ‘outros4’.

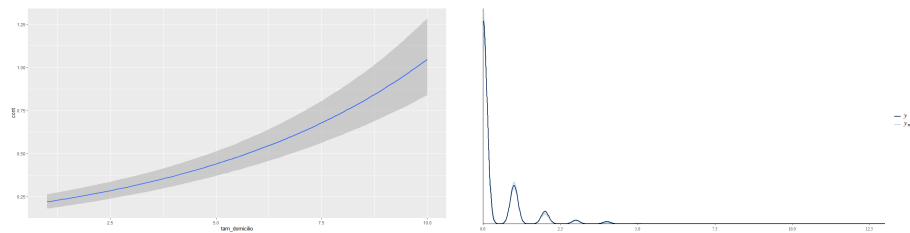


Figura 59: Efeitos marginais de interação e verificação preditiva a posteriori - Modelo ‘outros4’.

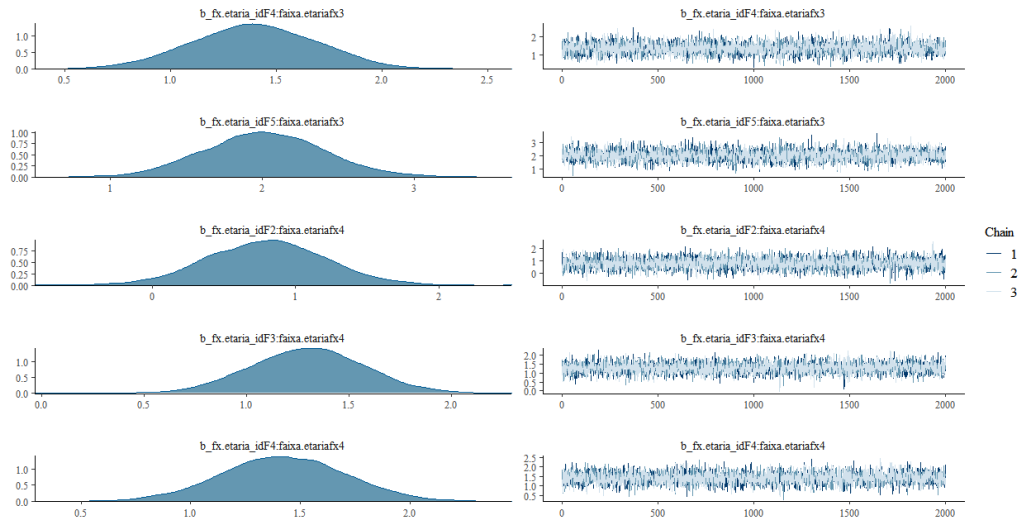


Figura 55: Diagnóstico de convergência (4) - Modelo 'outros4'.

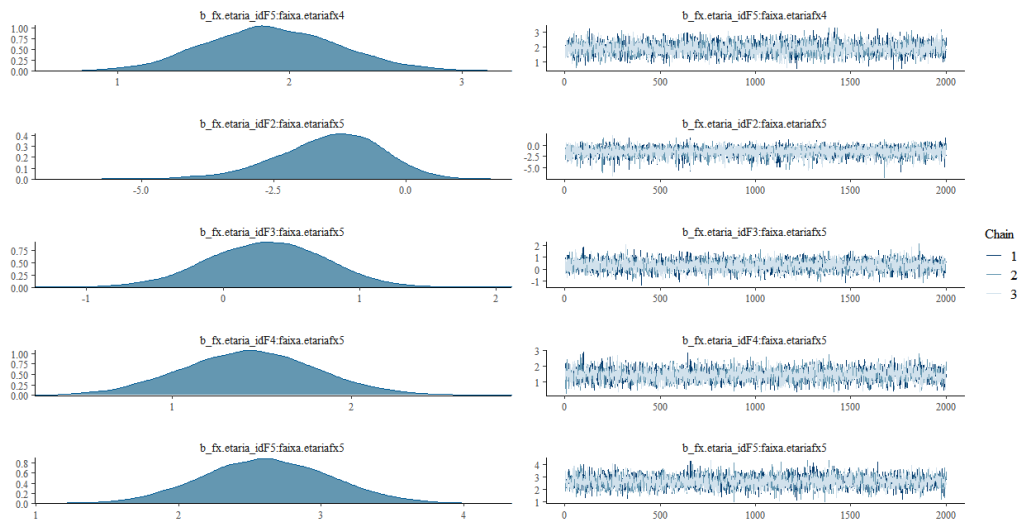


Figura 56: Diagnóstico de convergência (5) - Modelo 'outros4'.

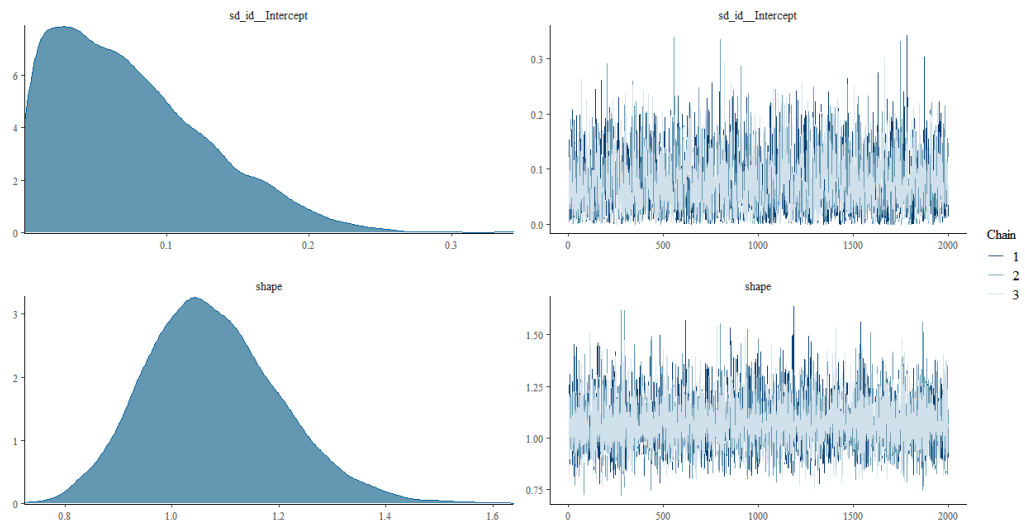


Figura 57: Diagnóstico de convergência (6) - Modelo 'outros4'.



## 12 Anexo 5

### Modelo 'geral3'

As figuras a seguir apresentam os diagnósticos de convergência do modelo 'geral3', efeitos das marginais e verificação preditiva do modelo.

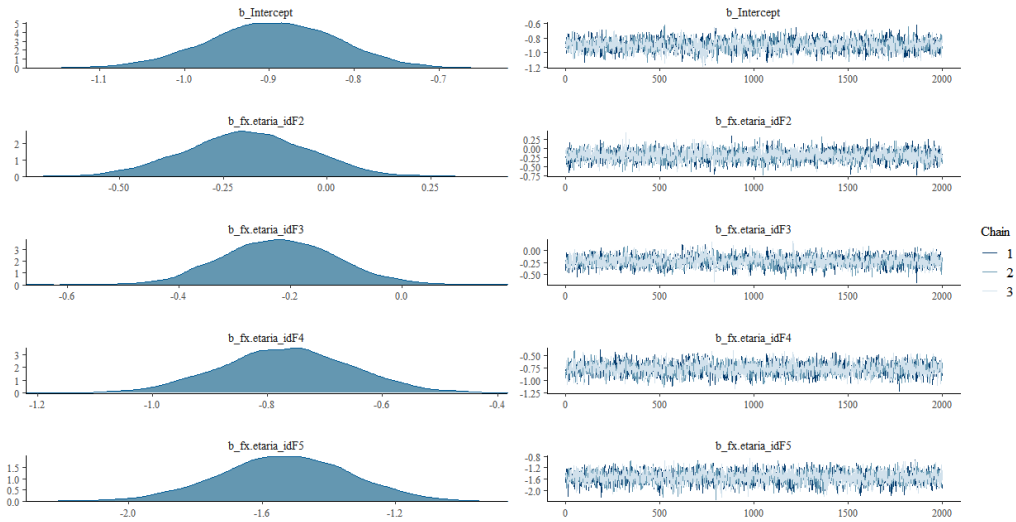


Figura 60: Diagnóstico de convergência (1) - Modelo 'geral3'.

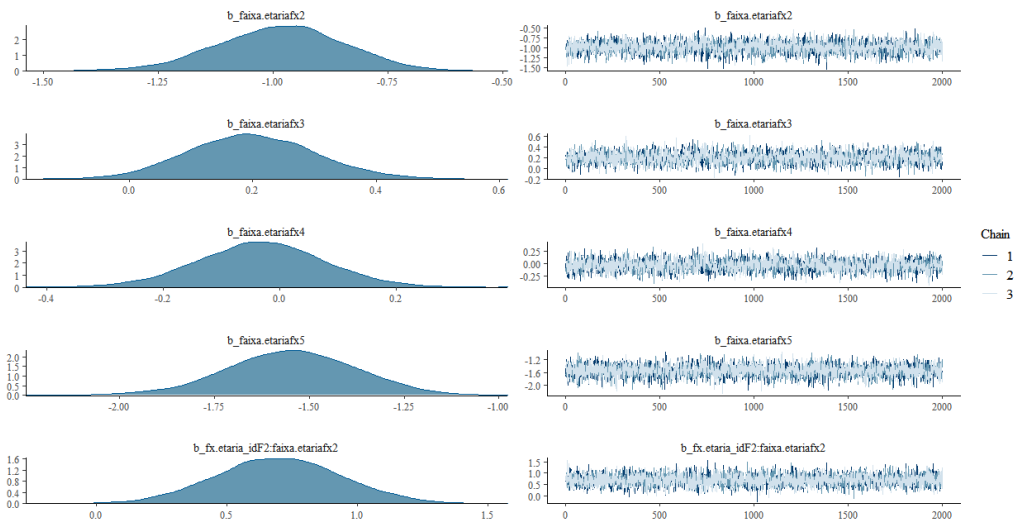


Figura 61: Diagnóstico de convergência (2) - Modelo 'geral3'.

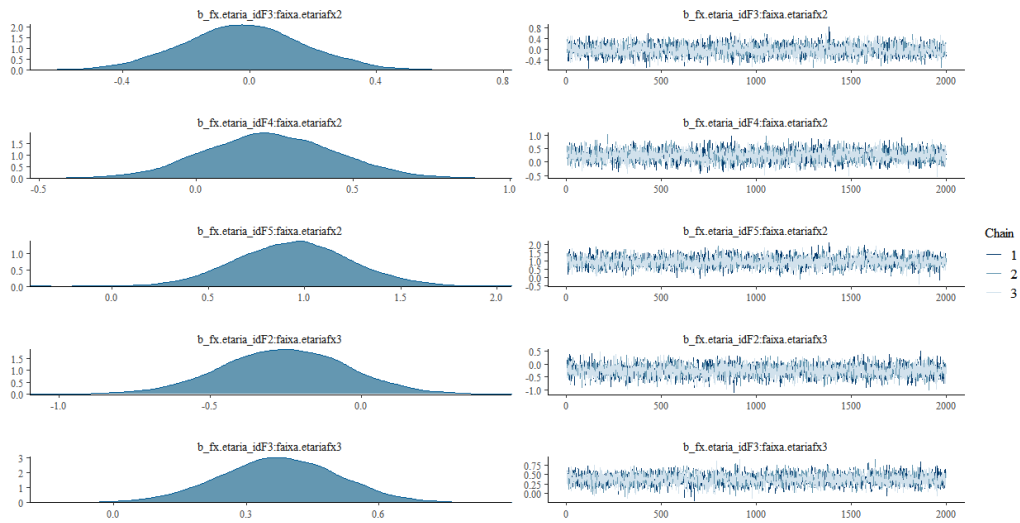


Figura 62: Diagnóstico de convergência (3) - Modelo ‘geral3’.

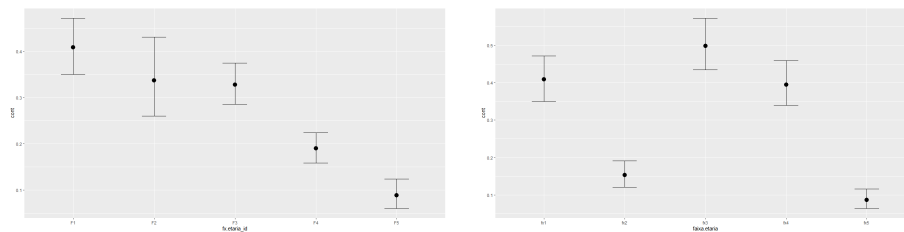


Figura 66: Efeitos condicionais das marginais - Modelo ‘geral3’.

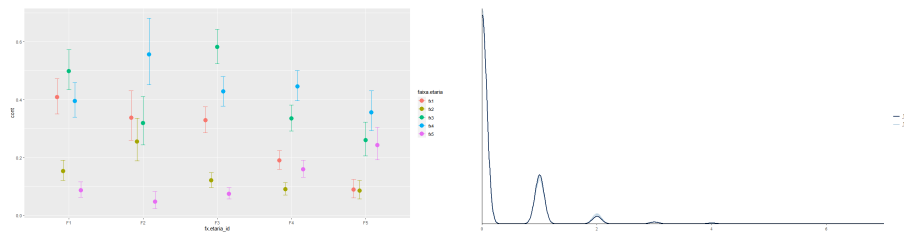


Figura 67: Efeitos marginais de interação e verificação preditiva a posteriori- Modelo ‘geral3’.

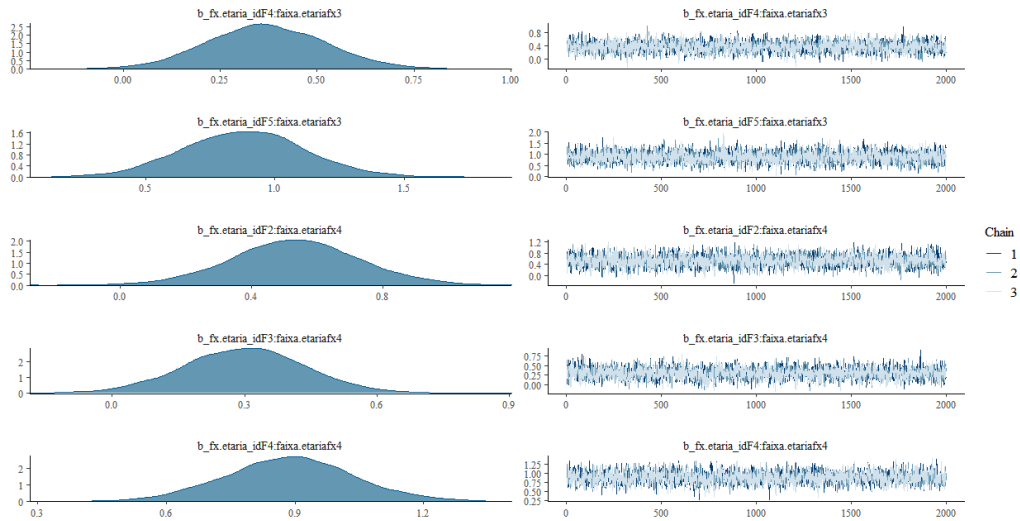


Figura 63: Diagnóstico de convergência (4) - Modelo 'geral3'.

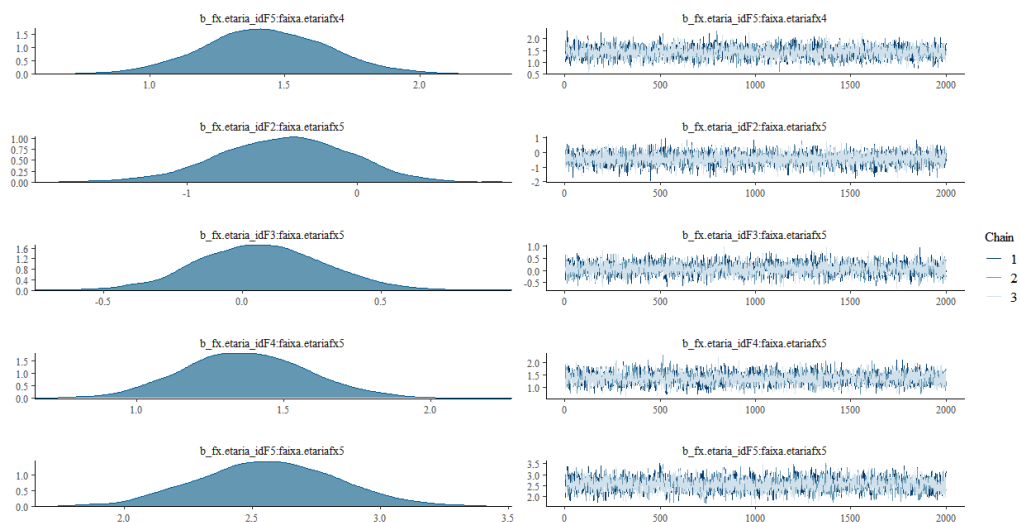


Figura 64: Diagnóstico de convergência (5) - Modelo 'geral3'.

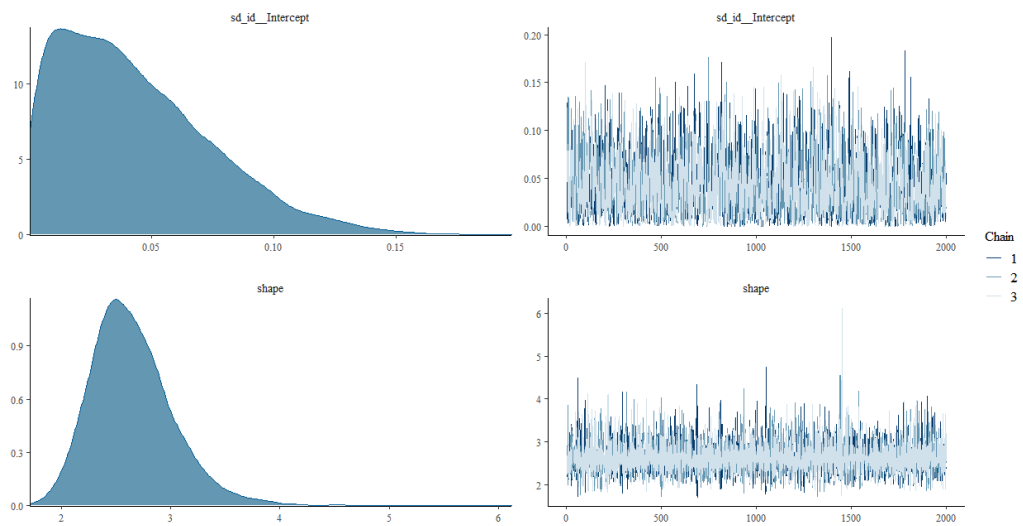


Figura 65: Diagnóstico de convergência (6) - Modelo 'geral3'.