

Trabalho de Conclusão de Curso

**Análise espacial de Cardiopatias Congênicas no  
Rio Grande do Sul**

Bruno Alano

5 de agosto de 2023

Bruno Alano

## Análise espacial de Cardiopatias Congênitas no Rio Grande do Sul

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientadora: Profa. Dr. Márcia Barbian

Porto Alegre  
7 de agosto de 2023

# Sumário

<b>1</b>	<b>Introdução</b>	<b>6</b>
<b>2</b>	<b>Metodologia</b>	<b>9</b>
2.1	Dados de Área . . . . .	10
2.2	Mapeamento de doenças . . . . .	11
2.3	Modelos Hierárquicos Bayesianos . . . . .	13
2.4	INLA . . . . .	14
2.5	PC prior . . . . .	15
2.6	Principais Modelos Hierárquicos Bayesianos para Mapeamento de Doenças . . . . .	16
2.6.1	O Modelo Besag . . . . .	16
2.6.2	O Modelo BYM . . . . .	17
2.6.3	O modelo de Leroux . . . . .	18
2.6.4	Modelo Dean . . . . .	18
2.6.5	Modelo BYM2 . . . . .	18
<b>3</b>	<b>Resultados</b>	<b>19</b>
<b>4</b>	<b>Conclusão</b>	<b>29</b>
	<b>Referências Bibliográficas</b>	<b>31</b>

## Lista de Figuras

Figura 2.1:	Contiguidade da torre e da rainha de primeira ordem. . . . .	10
Figura 2.2:	Grandes regiões do Brasil numeradas sendo 1 - Centro-Oeste, 2 - Nordeste, 3 - Norte, 4 - Sudeste e 5 - Sul . . . . .	11
Figura 3.1:	Mapa da estimativa do risco relativo $\lambda$ de cardiopatias congênitas através da TMP. . . . .	20
Figura 3.2:	Gráfico de correlação entre as variáveis agregados por município. no triângulo superior temos as correlações entre as variáveis, na diagonal principal a distribuição de densidade estimada via kernel e no triângulo inferior temos o gráfico de pontos entre as variáveis. as variáveis são respectivamente: logaritmo da taxa de mortalidade padrão; proporção de médicos geneticista, pediatras (intensivista e não intensivista) e cardiologista (intensivista e não intensivista); proporção de médicos; índice de escolaridade; proporção de extremamente pobres; renda per capita; proporção de mortalidade infantil até 5 anos presente no CENSO de 2010. . . . .	21
Figura 3.3:	Gráfico de correlação entre as variáveis com a transformação logarítmica agregados por município. no triângulo superior temos as correlações entre as variáveis, na diagonal principal a distribuição de densidade estimada via kernel e no triângulo inferior temos o gráfico de pontos entre as variáveis. as variáveis são respectivamente: logaritmo da taxa de mortalidade padrão; proporção de médicos geneticista, pediatras (intensivista e não intensivista) e cardiologista (intensivista e não intensivista); proporção de médicos; índice de escolaridade; proporção de extremamente pobres; renda per capita; proporção de mortalidade infantil até 5 anos presente no CENSO. . . . .	22
Figura 3.4:	Estimativa e Intervalos de Credibilidade das variáveis fixas . . . . .	23
Figura 3.5:	Risco Relativo BYM2, a direita temos a moda do risco relativo e na esquerda em vermelhos os RR que foram significativamente maiores que 1, em azul os RR que foram significativamente menores que 1 e em cinza os que não foram significativos. . . . .	24
Figura 3.6:	Risco relativo ordenado pela moda, cada barra representa 95% do intervalo de credibilidade. . . . .	28

## Lista de Tabelas

Tabela 3.1:	Tabela dos parâmetros fixos do modelo BYM2 com a moda, desvio padrão e intervalo de credibilidade. . . . .	24
Tabela 3.2:	Tabela dos hiperparâmetros do modelo BYM2 com a moda, desvio padrão e intervalo de credibilidade. . . . .	24
Tabela 3.3:	Tabela dos Municípios que tiveram um risco relativo significativamente maior que 1. . . . .	28

# 1 Introdução

Anomalias Congênitas (ACs) são alterações estruturais ou funcionais no embrião ou feto derivado de fatores anteriores ao nascimento, elas podem ser identificadas no acompanhamento pré-natal, ao nascimento ou ao longo da vida [World-Health-Organization](#). As ACs são observadas em 3% dos nascidos vivos e na maioria dos casos representam um problema grave de saúde, Sendo uma das maiores causas de mortalidade infantil [França \[2017\]](#).

Especificamente, as anomalias congênitas foram a principal causa de mortalidade infantil no Brasil em 2020 ([Ministério da Saúde \[2023\]](#)). Se forem considerados os diferentes períodos de óbito, segundo os autores [Pinto Júnior et al. \[2015\]](#) 20,3% dos óbitos ocorreram no período Neonatal precoce, 23% no período neonatal tardio e 23,7% no período pós-neonatal.

Devido à importância da temática, a Secretaria de Vigilância em Saúde e Ambiente trabalha na implantação de um modelo nacional de vigilância de anomalias congênitas na perspectiva da vigilância tripla, isto é, do monitoramento dos casos de anomalias, seus fatores de risco e desfechos em saúde associados [Ministério da Saúde \[2022\]](#).

Outro desafio está associado a identificação de ACs, pois muitas vezes, os sintomas de uma AC podem ser sutis ou facilmente confundidos com outras condições, o que pode levar a erros de diagnóstico ou à falta de identificação da anomalia. Além disso, em algumas regiões e comunidades, podem haver barreiras socioeconômicas, culturais e geográficas que dificultam o acesso aos serviços de saúde adequados. A falta de infraestrutura médica e recursos pode limitar a capacidade de diagnosticar e registrar anomalias congênitas corretamente. A falta de sistemas eficientes de notificação e registro de anomalias congênitas também contribui para a subnotificação, pois os casos podem passar despercebidos ou não serem registrados adequadamente. Sem dados precisos e abrangentes, é difícil avaliar a prevalência e o impacto das anomalias congênitas em uma população específica. Alguns trabalhos que envolvem a análise de subnotificação em ACs são: [Siefkes et al. \[2021\]](#) que abordam as implicações do preenchimento inadequado de formulários com informações de nascidos vivos pelos hospitais; [Zhang et al. \[2021\]](#) estudaram como o atraso do diagnóstico de cardiopatias congênitas influenciam na taxa de sobrevivência de recém nascidos em Beijing e os autores [Boyd et al. \[2005\]](#) avaliam quanto a interrupção de gravidez resultante do diagnóstico pré natal influenciam os registros nacionais de AC na Inglaterra.

Entre os trabalhos que buscam fatores de risco associados à AC, pode-se citar o artigo [Jenkins et al. \[2007\]](#) que faz uma revisão da literatura sobre as causas das

ACs. Neste trabalho é comentado que existe um grande avanço na literatura sobre as causas hereditárias de ACs, mas que as causas não herdadas e que podem ser evitadas ainda estão sendo descobertas. Entre as causas não herdadas, alguns fatores podem corroborar para o aumento da incidência de AC, como doenças maternas, exposições paternas, uso de drogas e exposições ambientais. Segundo os autores [Jenkins et al. \[2007\]](#) entre os fatores que podem reduzir o risco de doença cardíaca no feto está a ingestão de polivitamínico ou ácido fólico durante a gravidez.

No Brasil as informações sobre os nascimentos são registradas na Declaração de Nascidos Vivos (DNV) coletadas nos hospitais, estabelecimentos de saúde e cartórios (para partos domiciliares), os dados são enviados para as Secretarias Estaduais de Saúde que enviam as bases para o Ministério da Saúde que consolida e disponibiliza essas informações no Sistema de Informações sobre Nascidos Vivos (SINASC) [Ministério da Saúde \[c\]](#). O SINASC foi criado em 1990, tendo como sua base a a DNV, a qual é de preenchimento obrigatório [Ministério da Saúde \[b\]](#). Esse importante sistema é utilizado no Brasil para monitorar e identificar as características dos recém-nascidos, do pré-natal, gestação e parto. Esses dados são fundamentais para a construção de indicadores que acompanham a saúde materno-infantil e auxiliam no desenvolvimento de programas e políticas públicas voltadas para a melhoria dos serviços de saúde no país [Ministério da Saúde \[2022\]](#).

Nos últimos anos as ACs tiveram uma maior atenção do Ministério da Saúde, tendo lançado duas publicações sobre o tema, intituladas Saúde Brasil 2020/2021: anomalias congênitas prioritárias para a vigilância ao nascimento [Ministério da Saúde \[2022\]](#); e Saúde Brasil 2022: análise da situação de saúde e uma visão integrada sobre os fatores de risco para anomalias congênitas [Ministério da Saúde \[2023\]](#).

Entre os diferentes objetivos que demandam a análise de dados de AC podemos destacar aquelas que buscam investigar o processo envolvendo a dinâmica de um grupo específico de AC, as cardiopatias congênitas. Visto que no Brasil entre as ACs as malformações cardíacas congênitas são a principal causa de morte entre os portadores de malformações congênitas [Zimmerman et al. \[2020\]](#). Segundo os autores do artigo [Warren et al. \[2012\]](#), apesar da gravidade a prevenção dos defeitos cardiovasculares congênitos (DCCVs) tem sido um desafio devido à escassez de informações sobre fatores de risco modificáveis que afetam o desenvolvimento cardíaco.

Nos últimos anos, houve avanços significativos na compreensão das causas hereditárias dos DCCVs, incluindo a identificação de anormalidades genéticas específicas associadas a certos tipos de malformações. Embora haja menos conhecimento disponível sobre os fatores modificáveis não hereditários que podem ter um efeito adverso no coração fetal, uma crescente quantidade de pesquisas epidemiológicas aborda esse tema. Estudos indicam que até 30% dos casos de DCCVs podem ser potencialmente evitáveis por meio de alterações no ambiente fetal. No entanto, a carência de informações confiáveis sobre esses fatores de risco tem dificultado a implementação de estratégias populacionais para reduzir o impacto dessas doenças e tomar decisões de estilo de vida que diminuam a probabilidade de ter um filho com uma malformação cardíaca grave, como abordadas no artigo [Reece and Hulse \[2023\]](#). No trabalho [Botto et al. \[2001\]](#) foi utilizado o teste de Cochran-Armitage para analisar tendências temporais para cada tipo de cardiopatia congênita do período de 1968 até 1997 da região metropolitana de Atlanta.

Neste trabalho busca-se compreender a distribuição geográfica do registro das

cardiopatas congênitas no Estado do RS, através da análise espacial do risco relativo. Para isso, será considerado o número de nascidos vivos dado a presença ou não de cardiopatia congênita (CC), agrupados por município de residência materna. Logo, na análise será considerada a informação agregada, visto que as localizações das residências maternas não estão disponíveis devido à confidencialidade dos dados. Através da análise, busca-se identificar possíveis municípios com maior risco de notificação de CC o que pode corroborar com a identificação de fatores que associados ao registro desses casos. O modelo aplicado para estimar os Riscos Relativos (RR) é o BYM2 [Simpson et al. \[2017\]](#). O método é baseado no modelo BYM proposto por Besag-York-Mollié [Besag et al. \[1991\]](#) com uma modificação proposta por [Simpson et al. \[2017\]](#). Nesse modelo a distribuição *a priori* pertence à uma família de distribuições em que a complexidade é penalizada, a denominada PC prior proposta por [Simpson et al. \[2017\]](#). Uma das grandes vantagens desse modelo é que ele permite uma melhor interpretação dos parâmetros associados a estrutura de dependência espacial. Para o cálculo da distribuição *a posteriori* será utilizado o método integrated nested laplace approximation (INLA) [Rue et al. \[2009a\]](#) implementando no pacote R-INLA [Martins et al. \[2012\]](#).



## 2 Metodologia

Em muitas aplicações, os pesquisadores analisam dados que possuem referências geográficas, esses dados são denominados de espaciais e são definidos como realizações de um processo estocástico indexado pelo espaço:

$$Y(\mathbf{s}) \equiv \{y(\mathbf{s}), \mathbf{s} \in \mathcal{D}\},$$

em que  $\mathcal{D}$  é um subconjunto (fixo) de  $\mathbb{R}^2$ .

Seguindo a distinção apresentada por [Cressie, 1993], Banerjee et al. [2004] e Gelfand et al. [2010], a estatística espacial é dividida em 3 grandes áreas dependendo da natureza dos dados observados: dados pontuais, geoestatística e dados de área.

1. Padrões de pontos (Diggle [2013]): referem-se a variáveis aleatórias  $y(s)$  em que a ocorrência ou não de um evento e as localizações desses eventos são aleatórias. O domínio espacial  $\mathcal{D}$  é bem definido e indica a área em que o conjunto de pontos pode acontecer no  $\mathbb{R}^2$ . Por exemplo, o estudo pode envolver localizações de árvores de uma espécie em uma floresta ou endereços de pessoas com uma doença específica. Nesse caso, embora as localizações  $\mathbf{s} \in \mathbb{R}^2$  sejam aleatórias, a variável  $y(s)$  assume valores 0 ou 1, ou seja, o evento ocorreu ou não ocorreu. Se informações adicionais como covariáveis estiverem disponíveis, o processo é denominado um padrão pontual marcado.
2. Geoestatística (Peter and Ribeiro [2007], Krainski et al. [2018]):  $y(s)$  indica uma variável aleatória observada em uma localização específica  $s$ , onde  $s$  pode variar continuamente no domínio  $\mathcal{D}$ . A localização  $s$  é tipicamente um vetor bidimensional com latitude e longitude, mas também pode incluir mais uma dimensão como o tempo ou a altura. Os dados são representados por uma coleção de observações  $\mathbf{y} = (y(s_1), \dots, y(s_n))$ , onde o conjunto  $(s_1, \dots, s_n)$  indica as localizações nas quais as medidas de um processo contínuo  $Y(\mathbf{s})$  são obtidas. Um interesse comum é através da amostra finita de pontos estimar a superfície de todo o plano  $\mathbf{Y}(\mathbf{s}) \in \mathcal{D}$ , o que é denominado de krigagem.
3. Dados de área (Blangiardo and Cameletti [2015], Banerjee et al. [2004], Richardson and Peter Oosterom [2002] e Lawson [2018]): referem-se a dados em que a variável aleatória  $y(s_i)$  representa um valor agregado sobre uma unidade area  $s_i$ , com fronteiras bem definidas em um domínio  $\mathcal{D}$ , em que  $\mathcal{D}$  é uma coleção enumerável finita de áreas disjuntas. Exemplos de análises que envolvam dados de área estão relacionados à fronteiras administrativas, como municípios, bairros ou estados.

Na modelagem de dados espaciais, o processo inferencial deve levar em consideração a dependência espacial das observações. Essa informação adicional pode fornecer alguma estrutura que auxilie na explicação da variabilidade presente nos dados, pois observações próximas no espaço tendem a apresentar valores semelhantes. Além disso, não considerar a dependência dessas observações pode levar a algum tipo de viés nas estimativas [Sherman \[2010\]](#).

## 2.1 Dados de Área

Segundo [Cressie \[1993\]](#) ao se trabalhar com dados de área a região  $\mathcal{D}$  está particionada em  $n$  pedaços mutuamente exclusivos  $(s_1, \dots, s_n)$ , em que em cada uma dessas áreas observa-se uma variável aleatória  $Y(s_i)$  de interesse. Essa variável aleatória pode indicar uma métrica da área  $i$  tais como pib, idh, taxa de analfabetismo ou um dado agregado como número de mortes por determinada doença, número de crimes, número de nascimentos, etc. O objetivo da análise é modelar o vetor de variáveis  $\mathbf{Y}(\mathbf{s}) = (Y(s_1), Y(s_2), \dots, Y(s_n)) = (Y_1, Y_2, \dots, Y_n)$  considerando a dependência espacial, para isso, é necessário especificar alguma estrutura de vizinhança entre as  $(s_1, \dots, s_n)$  áreas que compõem a região  $\mathcal{D}$ .

Existem várias maneiras de definir um vizinho, entre elas pode-se considerar que duas áreas são vizinhas se compartilham limites ou se estão próximas dado algum critério. Usualmente, áreas são consideradas vizinhas quando compartilham fronteira, isto é, são contíguas. Entre as diferentes formas de considerar contiguidade, pode-se mencionar o critério da Torre (Rook) e a contiguidade utilizando o critério da Dama (Queen) ([Blangiardo and Cameletti \[2015\]](#)).

Na esquerda da Figura 2.1 um exemplo de contiguidade considerando o critério da torre, em que a área representada pelo quadrado na cor roxa possui como áreas vizinhas de primeira ordem os quadrados representados na cor verde e de segunda ordem os de cor amarela. Analogamente, no lado direito da Figura 2.1 o que é conhecido como critério de Dama, onde os pontos verdes são vizinhos de primeira ordem do ponto roxo e os pontos amarelos são os vizinhos de segunda ordem.

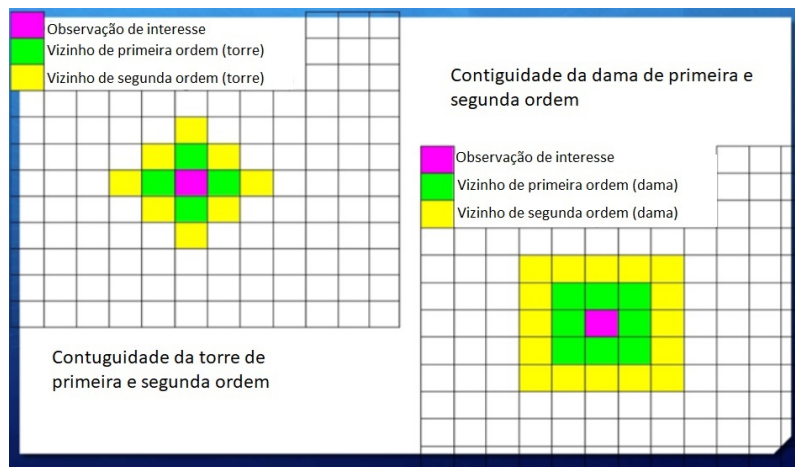


Figura 2.1: Contiguidade da torre e da rainha de primeira ordem.

A estruturada de vizinhança é representada através de uma matriz de vizinhança  $\mathbf{W}$  onde a entrada  $ij$  indica se a área  $i$  é vizinha da área  $j$ , ou seja, se elas possuem

alguma estrutura de dependência. Pela definição acima de contiguidade espacial, quando duas regiões são vizinhas recebem o valor 1 e 0 caso contrário.

$$W_{ij} = \begin{cases} n_{\delta_i}, & i = j \\ 1, & i \sim j \\ 0, & \text{Caso Contrário.} \end{cases}$$

Em que  $i \sim j$  denota que as áreas  $i$  e  $j$  são vizinhas,  $\delta_i$  representa o conjunto de vizinhos da região  $i$  e  $n_{\delta_i}$  a quantidade de vizinhos.

Para exemplificar uma estrutura de vizinha considere o mapa do Brasil dividido em 5 regiões numeradas de 1 a 5, a matriz abaixo indica qual a matriz  $\mathbf{W}$  desse mapa considerando o critério da Dama,

$$\mathbf{W} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 5 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 3 & 1 \\ 1 & 0 & 0 & 1 & 2 \end{pmatrix} \end{matrix} \quad (2.1)$$



Figura 2.2: Grandes regiões do Brasil numeradas sendo 1 - Centro-Oeste, 2 - Nordeste, 3 - Norte, 4 - Sudeste e 5 - Sul

## 2.2 Mapeamento de doenças

A área de mapeamento de doenças é um campo de estudo que se concentra na análise e visualização geográfica da distribuição e ocorrência de doenças em populações.

Entre os diferentes objetivos envolvendo o uso dessas metodologias estão aqueles que buscam entender os padrões de propagação de doenças, identificar áreas com alto ou baixo risco, buscar fatores que podem aumentar ou diminuir a chance de observar determinada moléstia, além de apoiar a tomada de decisões em saúde pública, entre os diversos trabalhos que abordam esse tipo de metodologia pode-se citar [Blangiardo and Cameletti \[2015\]](#), [Lawson \[2018\]](#) e [Waller and Gotway \[2004\]](#).

Considere a seguinte notação:

- $y_i$ : n° observado de casos da doença na área  $i$ ,  $i = 1, 2, \dots, n$ .
- $E_i$ : n° de casos esperados da doença na área  $i$ .
- $T_i$ : n° de pessoas em risco na área  $i$ .

Os casos  $Y_i$  são considerados variáveis aleatórias, enquanto  $E_i$  são fixos e calculados da seguinte forma:

$$E_i = T_i \left( \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n T_i} \right) = T_i \bar{r}. \quad (2.2)$$

O valor  $\bar{r}$  é a taxa de incidência global da doença em toda a região em estudo  $\mathcal{D}$ . O valor  $E_i$  é o número esperado de casos na área  $i$  caso a taxa de incidência da doença nessa área fosse igual à taxa de incidência global  $\bar{r}$ .

Comumente se utiliza o modelo de Poisson para representar a variável aleatória do número de casos da doença  $Y_i$  na área  $i$ :

$$Y_i | \lambda_i \sim \text{Poisson}(E_i \lambda_i), \quad (2.3)$$

em que  $\lambda_i$  é parâmetro que indica o risco relativo da doença na região  $i$ . Uma importante suposição desse modelo é que condicionado nos valores  $\lambda_1, \dots, \lambda_n$ , assume-se que as contagens  $Y_i$  são independentes.

Esse modelo pode ser facilmente modificado, permitindo a inclusão de  $p$  covariáveis  $x_{ik}$ , com coeficientes  $\beta_k$ , medidas em cada área  $i$ . A relação entre essas covariáveis e o vetor  $\mathbf{X}$  pode ser modelada através de um modelo linear generalizado [McCullagh and Nelder \[1989\]](#). Como  $Y_i$  possui distribuição de Poisson (Equação 2.3) uma função de ligação adequada é a logarítmica:

$$\log(\lambda_i) = \sum_{k=1}^p \beta_k x_{ki} + b_i. \quad (2.4)$$

Os efeitos aleatórios do modelo são denotados por  $b_i$ , que podem indicar a dependência espacial através da distribuição conjunta do vetor de efeitos aleatório  $\mathbf{b} = (b_1, \dots, b_n)$ .

No mapeamento de doenças, o objetivo é estudar a variação espacial do risco relativo  $\lambda_i$ , assumindo que ele não é constante no mapa. Uma forma de estimar  $\lambda_i$  do modelo 2.3 é através do estimador de máxima verossimilhança, conhecido como taxa de mortalidade padronizada (TMP)

$$\text{TMP}_i = y_i / E_i. \quad (2.5)$$

Especificamente, no caso de doenças raras, os desafios envolvidos na análise são maiores, conforme indicado pelos trabalhos dos autores [Gelfand et al. \[2010\]](#), [Besag](#)

et al. [1991], Rue and Martino [2007], pois se  $E_i$  for próximo de zero, a estimativa  $TMP_i$  terá variância muito grande. Logo, os maiores valores estimados de  $\lambda_i$  serão associados à áreas em que não há um maior risco, mas à áreas que possuem uma pequena população. Além disso, a  $TMP_i$  considera que os  $Y_i$ 's são independentes, mas em alguns casos, uma parte desconhecida da variação do risco relativo pode ser causada por fatores não observados, dependentes geograficamente Gelfand et al. [2010]. Diante disso, uma das alternativas comumente empregadas é a utilização de modelos hierárquicos que consideram a estrutura espacial.

## 2.3 Modelos Hierárquicos Bayesianos

Considere o seguinte exemplo teórico em um contexto bayesiano, em que uma variável  $Z_i$  possui distribuição de poisson com média  $E_i\theta_i$ , sendo  $\theta_i$  o parâmetro de interesse e  $E_i$  uma constante conhecida,

$$Z_i|\theta_i \sim \text{Poisson}(E_i\theta_i). \quad (2.6)$$

Podemos estar interessados em explicar o parâmetro de interesse  $\theta_i$  por meio de uma variável  $x_i$ , para que isso seja possível o espaço do domínio do parâmetro tem que ser compatível com o domínio da variável  $x_i$ . Como a distribuição de Poisson pertence à família exponencial McCullagh and Nelder [1989], pode-se relacionar essas variáveis por uma função de ligação, a mais usual para esse caso é a transformação logarítmica. A relação linear entre o logaritmo de  $\theta_i$  e  $x_i$ , adicionando-se um termo de erro para permitir flutuações da relação linear produz a seguinte equação:

$$\theta_i = \exp(\beta_0 + \beta_1 x_i + \epsilon_i) \quad (2.7)$$

$$\log(\theta_i) = \beta_0 + \beta_1 x_i + \epsilon_i. \quad (2.8)$$

Pode-se atribuir distribuições *a priori* não informativas aos coeficientes  $\beta_0$  e  $\beta_1$ . A modelagem do termo de erro  $\epsilon_i$  comumente assume uma distribuição normal, com média zero e variância  $\sigma^2$ . No entanto, é comum que o parâmetro  $\sigma^2$  não disponha de informações *a priori* substanciais. Nesses casos, torna-se pertinente representar essa variância por meio de outra distribuição de probabilidade, que permita a incorporação de incerteza sobre o valor específico de  $\sigma^2$ . Introduzem-se, então, distribuições adicionais que não são do interesse do estudo denominadas de *hiperprioris* e seus parâmetros de hiperparâmetros.

$$\begin{aligned} \beta_0 &\sim \text{Normal}(0, 100000) \\ \beta_1 &\sim \text{Normal}(0, 100000) \\ \epsilon_i|\sigma^2 &\sim \text{Normal}(0, \sigma^2) \\ \sigma^2 &\sim h_{\sigma^2}(\rho), \end{aligned}$$

No contexto de mapeamento de doenças, o modelo hierárquico deve considerar mais uma informação, a estrutura de dependência espacial. Logo a modelagem não é realizada sobre as marginais dos parâmetros  $\theta_i$  e variáveis  $Z_i$ , mas sim sobre a conjunta  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$  e  $\mathbf{Z} = (Z_1, \dots, Z_n)$ , isto é, no vetor que representa as variáveis observadas em  $n$  áreas que compõem a região de estudo.

Uma forma de agregar a estrutura de dependência espacial é definir a distribuição *a priori* do vetor  $\boldsymbol{\theta}$  através de suas densidades condicionais, o denominado Campo Aleatório Markoviano [Rue and Held \[2005\]](#), o equivalente espacial de uma cadeia de Markov no tempo. A distribuição de  $\boldsymbol{\theta}$  é chamada um Campo Aleatório Markoviano se, para todo  $i$ , sua distribuição conjunta satisfaz a seguinte propriedade:

$$f(\theta_i|\boldsymbol{\theta}_{-i}) = f(\theta_i|\boldsymbol{\theta}_{\delta_i}), \quad (2.9)$$

em que  $\boldsymbol{\theta}_{-i}$  indica todos os elementos em  $\boldsymbol{\theta}$  exceto  $i$  e  $\delta_i$  indica o conjunto de áreas vizinhas da área  $i$ , compostas por todas as áreas  $j$  em que  $w_{ij} > 0$  na matriz de vizinhança  $\mathbf{W}$ . Logo, da equação 2.9, a distribuição de  $\theta_i$  condicionada à todas as áreas do mapa é a mesma distribuição de  $\theta_i$  condicionada somente às áreas vizinhas:

$$\theta_i \perp \boldsymbol{\theta}_{-i} | \boldsymbol{\theta}_{\delta_i}.$$

Dessa maneira áreas próximas terão riscos relativos semelhantes.

A distribuição conjunta mais utilizada como distribuição *a priori* de  $\boldsymbol{\theta}$  é a normal multivariada,

$$f(\boldsymbol{\theta}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{-n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^t \Sigma^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})\right), \quad (2.10)$$

em que  $\boldsymbol{\mu}$  é o vetor de médias e  $\Sigma$  é a matriz de variância e covariância. Nesses casos, os dados são modelados através de um Campo Aleatório Markovoviano Gaussiano, conhecidos como GMRF (*Gaussian Markov Random Fields*) [Rue and Held \[2005\]](#).

## 2.4 INLA

Em inferência bayesianana um dos principais objetivos é encontrar a distribuição *a posteriori* dos parâmetros desconhecidos do modelo. A distribuição *a posteriori* é obtida a partir da combinação da informação prévia (distribuição *a priori*) com a informação fornecida pelos dados observados (função de verossimilhança).

Através da distribuição *a posteriori* é possível fazer inferências sobre o parâmetro desconhecido, como calcular intervalos de credibilidade, estimar probabilidades do parâmetro ser maior ou menor que determinado valor, calcular testes de hipóteses e fazer previsões por meio da preditiva *a posteriori*. Apesar de sua grande flexibilidade, somente quando utilizamos o que é denominado distribuições *a priori* conjugadas é possível obter a distribuição *a posteriori* analiticamente. Para as demais situações, são necessários métodos numéricos para sua obtenção.

O método mais utilizado para obtenção da distribuição *a posteriori* é o Markov Chain Monte Carlo (MCMC) [Metropolis et al. \[1953\]](#). O MCMC é uma técnica poderosa e amplamente utilizada para amostragem de distribuições complexas, permitindo inferência em situações onde a amostragem direta ou soluções analíticas não estão disponíveis. Apesar de amplamente utilizada, alguns problemas podem surgir na sua utilização, como o alto custo computacional, definição da distribuição proposta e problema de convergência quando o modelo é complexo e as variáveis são altamente correlacionadas.

Em 2009 [Rue et al. \[2009b\]](#) propôs uma abordagem alternativa para o MCMC chamado de integrated nested Laplace approximations (INLAs), diferente do MCMC o algoritmo INLA é determinístico e fornece resultados mais rápidos e acurados na

classe de modelos gaussianos pois aproveita as propriedades dessa classe para produzir matrizes de precisão esparsas ganhando uma melhora considerável no desempenho computacional [Rue and Held \[2005\]](#), [Blangiardo and Cameletti \[2015\]](#).

Considerando o exemplo indicado nas Equações 2.6 e 2.7 um modelo gaussiano latente é definido como

$$\log(\theta_i) = \eta_i = \beta_0 + \sum_{k=1}^{n_\beta} \beta_k x_{ki} + \sum_{j=1}^{n_f} f_j(x'_{ij}) + \epsilon_i, \quad (2.11)$$

em que  $\eta_i$  é o preditor linear ligado ao parâmetro de interesse  $\theta_i$  por meio da função de ligação log,  $\eta_i$  é definido como uma função aditiva dos preditores. O vetor  $\boldsymbol{\beta} = \{\beta_0, \beta_1, \dots, \beta_{n_\beta}\}$  representa os coeficientes do modelo linear e  $\mathbf{x} = \{x_1, \dots, x_{n_\beta}\}$  são as covariáveis. O vetor  $\mathbf{f} = \{f_1(\cdot), \dots, f_{n_f}(\cdot)\}$  indica as funções não lineares das covariáveis  $\mathbf{x}' = \{x'_1, \dots, x'_{n_\beta}\}$ .

## 2.5 PC prior

As distribuições *a priori* desempenham um papel importante na análise bayesiana e sua definição deve ser considerada com muita cautela. Entre as várias possibilidades, pode-se citar as distribuições *a priori* não informativas como *a priori* proposta por [Jeffreys \[1946\]](#). Atualmente, uma família de distribuições *a priori* bastante flexível e que possibilita maior interpretabilidade em modelos complexos foi proposta pelos autores [Simpson et al. \[2017\]](#), esse novo conceito para construção de distribuições *a priori* é denominado *Penalised Complexity prior* a (PC prior).

O conceito mais importante da criação da PC prior é favorecer o modelo menos complexo, denominado modelo base, quando não existe evidência suficiente para sua rejeição. Para isso, quatro princípios são definidos:

- **Princípio 1 (menor complexidade):** O princípio da Navalha de Occam orienta que modelos mais simples devem ser preferidos aos modelos mais complexos. A PC prior mais simples é o modelo base, portanto, desejamos que a distribuição *a priori*  $\pi(\xi)$  para um parâmetro  $\xi$  penalize desvios do modelo base.
- **Princípio 2 (conexão com a priori de Jeffreys):** A PC prior utiliza a divergência de Kullback-Leibler (KDL) como medida de complexidade, a medida de Kullback-Leibler entre duas funções  $f$  e  $g$  é definida como:

$$KLD(f||g) = \int f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x}.$$

A função acima mensura a quantidade de "informação" perdida quando o modelo  $g(x)$  é usado para aproximar um modelo  $f(x)$ . No contexto da PC prior, o modelo base é o  $g(x)$  e o modelo mais complexo é o  $f(x)$ , logo a divergência de Kullback-Leibler medirá a distância do modelo base para o modelo proposto. Quanto maior essa distância maior é a evidência de que um modelo mais complexo deve ser preferível, por exemplo, se  $g(x)$  (modelo sem estrutura espacial) apresentar uma grande distância em relação à  $f(x)$  (modelo com estrutura espacial), então o modelo espacial deverá ser utilizado.



- **Princípio 3:** (Penalização dada uma taxa constante) Na PC prior são atribuídas distâncias entre os modelos (princípio 2), em vez de atribuir distribuições aos parâmetros, considerando o princípio 2,

$$d(\xi) = \sqrt{2KLD(\xi)},$$

representa a distância entre dois modelos candidatos a distribuição *a priori* do parâmetro  $\xi$ . Na busca da PC prior adequada, atribui-se uma distribuição exponencial para essa distância:

$$\pi(d(\xi)) = \psi \exp^{-\psi d(\xi)}, \quad \psi > 0,$$

a função acima possui a moda localizada em  $d(\xi) = 0$ . Dado a propriedade da distribuição exponencial, a distribuição *a priori*  $\pi(d(\xi))$  satisfaz a propriedade de perda de memória, logo

$$\frac{\pi_\xi(d + \Delta)}{\pi_\xi(d)} = \psi^\Delta, \quad d, \Delta \geq 0,$$

em que a constante de decaimento é dada por  $\psi$ . Para mudar a forma da distribuição *a priori* de  $\xi$ , muda-se o parâmetro de decaimento  $\psi$ .

- **Princípio 4:** (Escala) Determinar o valor de  $\psi$  dado algum conhecimento sobre o parâmetro, com esse objetivo o último princípio busca definir a PC prior de forma que os usuários possam escolher a probabilidade que o parâmetro de interesse esteja em algum intervalo ou a chance de que ele seja maior que determinado valor, logo

$$P(h(\xi) > U) = \alpha,$$

em que  $h(\xi)$  é alguma função do parâmetro,  $U$  é o valor da distribuição definido pelo usuário, que determina a magnitude do efeito de um componente no modelo e quanto informativa a PC prior será e  $\alpha$  é a probabilidade na calda da distribuição.

## 2.6 Principais Modelos Hierárquicos Bayesianos para Mapeamento de Doenças

### 2.6.1 O Modelo Besag

Uma das mais populares abordagens para estruturar dependência espacial através de um Campo Aleatório Markoviano Gaussiano (GMRF) é o modelo Conditional Autoregressive (CAR) proposto por Besag [1974], nesse modelo o vetor de efeitos aleatórios  $\mathbf{b}$  da equação 2.4 possui distribuição normal multivariada de parâmetros  $\mathbf{b} \sim \mathcal{N}_n(\mathbf{0}, \frac{1}{\tau_b} \mathbf{Q}^{-1})$ . A matriz de precisão  $\mathbf{Q}$  controla a estrutura de autocorrelação espacial dos efeitos aleatórios e é baseada na vizinhança definida pela matriz de adjacência  $\mathbf{W}$ , as entradas da matriz  $\mathbf{Q}$  são:



$$Q_{ij} = \begin{cases} n_{\delta_i} & i = j \\ -1 & i \sim j \\ 0 & \text{Caso Contrário.} \end{cases}$$

As distribuição condicional do efeito aleatório da área  $i$ , o  $b_i$  da Equação 2.4 é definido como

$$b_i | \mathbf{b}_{-i}, \tau_b \sim \mathcal{N}\left(\frac{\rho \sum_{j \in \delta_i} b_j}{n_{\delta_i}}, \frac{1}{n_{\delta_i} \tau_b}\right) \quad (2.12)$$

$\rho$  é um parâmetro de autocorrelação espacial,  $\tau_b$  é o parâmetro de precisão e  $\mathbf{b}_{-i} = (b_1, \dots, b_{i-1}, b_{i+1}, \dots, b_n)^T$ . Neste modelo, a esperança condicional é a média dos efeitos aleatórios em áreas vizinhas, enquanto a precisão condicional é proporcional ao número de vizinhos. O último é apropriado porque, se os efeitos aleatórios forem espacialmente correlacionados, quanto mais vizinhos uma área tiver, mais informação haverá. Observe que apesar da distribuição condicional considerar todas as áreas do mapa, os parâmetros do modelo são dependentes somente das áreas vizinhas, conforme propriedade indicada na Equação 2.9. Quando  $\rho = 1$  no modelo 2.12 se obtém o impróprio ICAR. É chamado impróprio, porque a matriz de precisão que gera não possui inversa. Apesar dessa propriedade, o ICAR produz distribuições *a posteriori* próprias.

## 2.6.2 O Modelo BYM

O modelo CAR produz uma pequena correlação marginal que depende da estrutura de vizinhança  $\mathbf{W}$ , a variância não é constante e a correlação entre pares de áreas separadas pelo mesmo número de vizinhos não são necessariamente iguais [Rue and Martino \[2007\]](#), essas características fazem com o que modelo tenda a suavizar muito o risco relativo, tornando-o inadequado quando o objetivo é identificar áreas com valores extremos para o risco. No artigo [Besag et al. \[1991\]](#) os mesmos autores propuseram uma extensão para permitir a autocorrelação espacial fraca e forte, por meio da divisão do efeito aleatório:

$$\begin{aligned} \mathbf{b} &= \mathbf{v} + \mathbf{u} \\ \mathbf{v} &\sim \mathcal{N}(\mathbf{0}, \tau_v^{-1} \mathbf{I}) \\ \mathbf{u} &\sim \mathcal{N}_n\left(0, \frac{1}{\tau_u} \mathbf{Q}^{-1}\right). \end{aligned}$$

A matriz de covariância do vetor  $\mathbf{b}$  é

$$Var(\mathbf{b} | \tau_u, \tau_v) = \frac{1}{\tau_v} \mathbf{I} + \frac{1}{\tau_u} \mathbf{Q}^{-1}.$$

O modelo acima é conhecido como BYM (Besag-York-Mollié) [Besag et al. \[1991\]](#), ele surge para permitir a modelagem de um efeito aleatório específico de determinada área e não partilhada pelos vizinhos, para isso o antigo efeito espacial  $\mathbf{b}$  é decomposto como a soma de um efeito aleatório não estruturado ( $\mathbf{v}$ ) e um efeito aleatório espacial ( $\mathbf{u}$ ).

### 2.6.3 O modelo de Leroux

Os parâmetros do modelo BYM sofrem o problema de identificabilidade [Macnab \[2011\]](#), pois não é possível identificar os efeitos dos parâmetros  $u$  e  $v$  isoladamente, outro problema  $\tau_u$  e  $\tau_v$  não representarem a variabilidade na mesma escala, dificultando a escolha da distribuição *a priori* [Wakefield \[2006\]](#). Tendo isso em vista [Leroux et al. \[2000\]](#) propõe uma formula alternativa onde  $\mathbf{b}$  assume uma distribuição normal com esperança igual a zero e matriz de covariância

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \tau_b^{-1}((1 - \phi)\mathbf{I} + \phi\mathbf{Q})^{-1}, \quad \phi \in [0, 1].$$

Se  $\phi = 0$  o modelo se resume ao efeito aleatório puro, sem nenhuma dependência espacial, se  $\phi = 1$  os efeitos aleatórios são explicados somente pela estrutura de dependência espacial, como o modelo CAR [Macnab \[2011\]](#).

### 2.6.4 Modelo Dean

Dean [Dean et al. \[2001\]](#) propôs outra reparametrização do modelo BYM

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left( \sqrt{1 - \phi} \mathbf{v} + \sqrt{\phi} \mathbf{u} \right)$$

Possuindo matriz de covariância

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \tau_b^{-1}((1 - \phi)\mathbf{I} + \phi\mathbf{Q}^{-1}).$$

A decomposição de Dean é muito perspicaz pois permite uma maior flexibilidade na manipulação na matriz inversa generalizada  $\mathbf{Q}^{-1}$ .

### 2.6.5 Modelo BYM2

[Simpson et al. \[2017\]](#) Propôs uma modificação que resolve tanto o problema de identificabilidade como o de escalonamento do modelo BYM, denominado como bym2 no artigo de [Riebler et al. \[2016\]](#). O modelo usa um componente escalonado  $\mathbf{u}_*$  com variância generalizada calculada como a média geométrica das marginais igual a 1. Consequentemente o efeito aleatório é dado por:

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} (\sqrt{1 - \phi} \mathbf{v} + \sqrt{\phi} \mathbf{u}_*),$$

e matriz de covariância

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \frac{1}{\tau_b} ((1 - \phi)\mathbf{I} + \phi \mathbf{Q}_*^{-1}).$$

O valor de  $0 \leq \phi \leq 1$  mensura a proporção da variância marginal que é explicada pelo componente espacial, semelhante ao Modelo Leroux e ao modelo Dean, esse modelo se resume a sobredispersão caso  $\phi = 0$ , e ao modelo CAR caso  $\phi = 1$ .

Vale mencionar que o escalonamento de  $\mathbf{Q}_*^{-1}$  permite que  $\tau_b$  tenha a mesma interpretação independentemente do grafo  $\mathbf{W}$  que representa a estrutura de dependência espacial. Assim o escalonamento de  $\mathbf{Q}_*^{-1}$  facilita a interpretação de  $\tau_b$  e sua escolha *a priori*. Essa interpretação não é direta para os modelos anteriores, em que é necessário transformações dos parâmetros para mensurar a proporção da variabilidade espacial e não espacial explicada pelo modelo.

### 3 Resultados

O estudo tem caráter ecológico, de base populacional e retrospectivo, realizado a partir de dados secundários do Sistema de Informações sobre Nascidos Vivos (SINASC) - uma fonte de dados importante que possui destaque internacional como um dos maiores e mais completos sistemas de informação sobre nascimentos no mundo. A análise será referente ao período que abrange 2011 a 2021, períodos anteriores não foram considerados pois em 2011 houve um maior detalhamento na DNV [Ministério da Saúde](#) [c].

Serão considerados somente as DNV de nascidos vivos em que a residência materna está localizada no estado do Rio Grande do Sul, totalizando 1.655.702 nascimentos. A informação associadas às ACs estão disponíveis no SINASC de acordo com a Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde - 10<sup>a</sup> revisão (CID 10) [Ministério da Saúde](#) [a]. Na análise serão considerados os CIDs que indicam malformação cardíaca, representados pelos seguintes códigos: Q20, Q21, Q22, Q23, Q24, Q25, Q26, Q27 e Q28, totalizando um registro de 1996 casos. Importante destacar que uma criança pode apresentar mais de um CID em sua DNV. Além disso, as análises consideraram algumas variáveis socio-econômicas, obtidas no CENSO de 2010.

O estado do Rio Grande do Sul é composto por 497 municípios. No entanto, cabe ressaltar que o município de Pinto Bandeira emancipou-se de Bento Gonçalves em outubro de 2012, tendo a posse de prefeito em 2013. Por essa razão, para os propósitos deste estudo, consideramos Pinto Bandeira como parte de Bento Gonçalves, resultando em 496 municípios a serem considerados em nossas análises. A maioria desses municípios possui baixa densidade populacional, o que se reflete em um baixo número de nascimentos em diversas regiões. Essa característica pode resultar em estimativas com grande variabilidade ao analisarmos a relação entre o número de nascimentos e a presença de cardiopatias congênitas. Tal variabilidade se dá pela presença de municípios pequenos sem casos de cardiopatias, enquanto outros podem apresentar valores muito altos.

Diante dessa situação, uma das alternativas para lidar com a variabilidade é a aplicação de métodos que levem em consideração a estrutura de dependência espacial entre os municípios. Dessa forma, o valor estimado para uma determinada localidade (município)  $i$  será influenciado não apenas por suas características locais, mas também pela informação proveniente de seus municípios vizinhos. Essa abordagem permitirá que a variabilidade dos resultados seja suavizada, uma vez que a presença de valores extremos será mitigada pela estrutura de dependência espacial.

O primeiro modelo a ser considerado na análise da distribuição de cardiopatias é

o Estimador de Máxima Verossimilhança do risco relativo  $\lambda$ , a TMP. Este estimador, considera que os riscos das áreas  $i = (1, 2, \dots, n)$  do modelo 2.3 são independentes. A TMP estimada é mostrada no mapa da Figura 3, através dessa imagem é possível observar a alta variância do estimador, os riscos estimados possuem uma grande amplitude, por exemplo, o município de Coqueiro Baixo a  $TMP=7.97$  (1 caso de cardiopatia em 104 nascidos vivos) e o município de Relvado com  $TMP=0$  (0 casos de cardiopatia congênita em 166 nascidos vivos). Tais valores não devem representar a realidade local, provavelmente não há um surto de cardiopatias congênitas na cidade de Coqueiro Baixo, nem o risco de cardiopatia é nulo no município de Relvado.

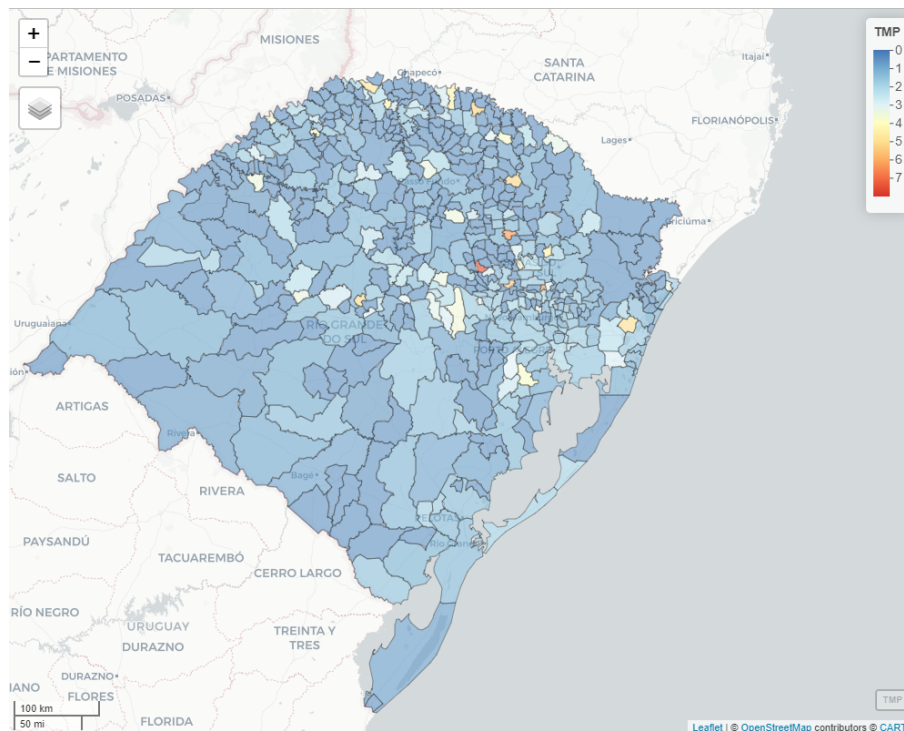


Figura 3.1: Mapa da estimativa do risco relativo  $\lambda$  de cardiopatias congênitas através da TMP.

Como comentado no capítulo 2 uma das alternativas é a utilização de modelos hierárquicos que consideram a estrutura espacial, o segundo modelo a ser considerado é o BYM-2 apresentado na seção 2.6 do capítulo 2. As covariáveis consideradas no modelo são: proporção de médicos geneticista, pediatras (intensivista e não intensivista) e cardiologista (intensivista e não intensivista); proporção de médicos; índice de escolaridade; proporção de extremamente pobres; renda per capita; proporção de mortalidade infantil até 5 anos presente no CENSO de 2010. As covariáveis que mensuram a proporção consideram como denominados o número de nascidos vivos por município.

Nas Figuras 3.2 e foi explorado a relação entre as variáveis socioeconômicas e a TMP entre os diferentes municípios, é possível observar que algumas variáveis apresentam algum tipo de associação linear (com e com o logaritmo da taxa de mortalidade padrão, aplicamos o logaritmo devido a função de ligação do modelo 2.3, as variáveis que tiveram maiores correlações com o logaritmo da TMP foram índice de escolaridade, logaritmo da renda per capita e logaritmo dos médicos selecionados. Além disso, é possível observar uma associação linear muito forte entre algumas covariáveis, como proporção de extremamente pobres e renda per capita.

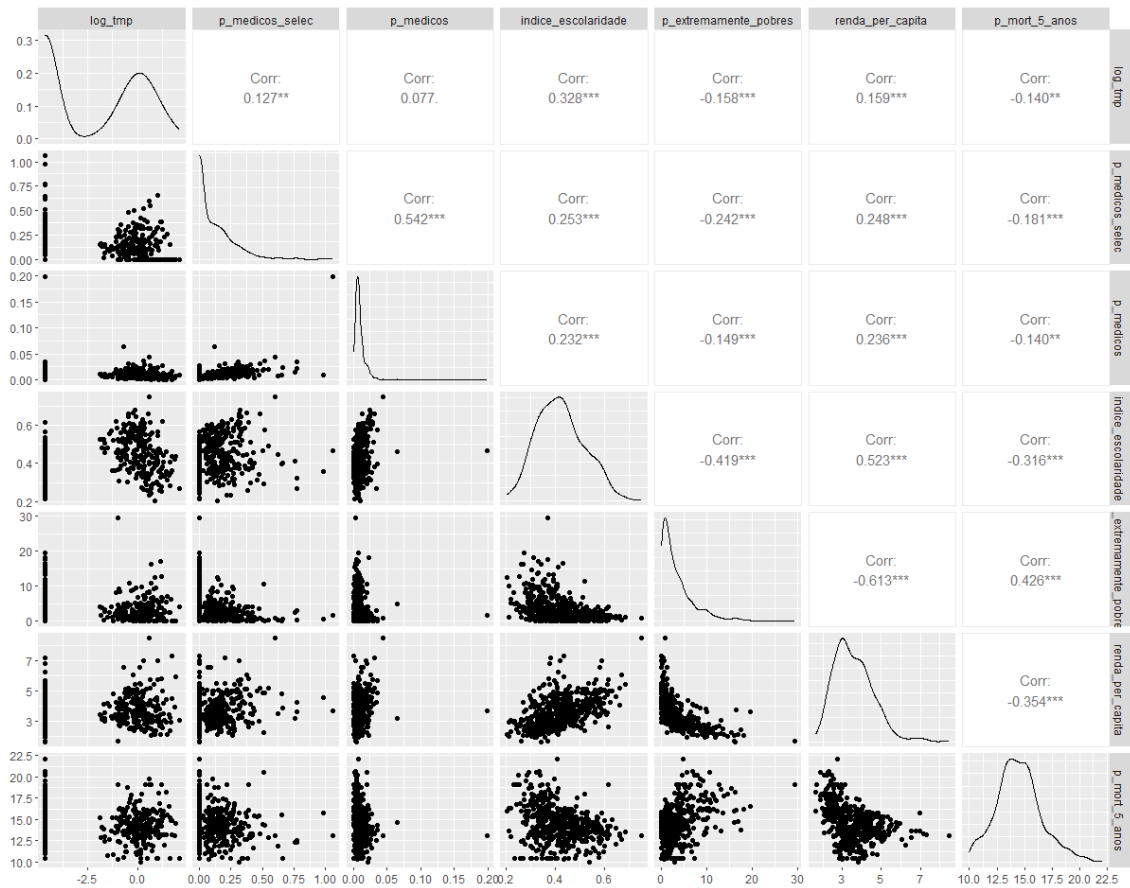


Figura 3.2: Gráfico de correlação entre as variáveis agregadas por município. no triângulo superior temos as correlações entre as variáveis, na diagonal principal a distribuição de densidade estimada via kernel e no triângulo inferior temos o gráfico de pontos entre as variáveis. as variáveis são respectivamente: logaritmo da taxa de mortalidade padrão; proporção de médicos geneticista, pediatras (intensivista e não intensivista) e cardiologista (intensivista e não intensivista); proporção de médicos; índice de escolaridade; proporção de extremamente pobres; renda per capita; proporção de mortalidade infantil até 5 anos presente no CENSO de 2010.

Para avaliar a presença de associação não linear entre as preditoras e os casos de cardiopatias congênitas, foi aplicada a função logaritmo nas covariáveis, os resultados estão indicados na Figura 3.3, observa-se que não houve uma grande mudança entre associação dos dados transformados.

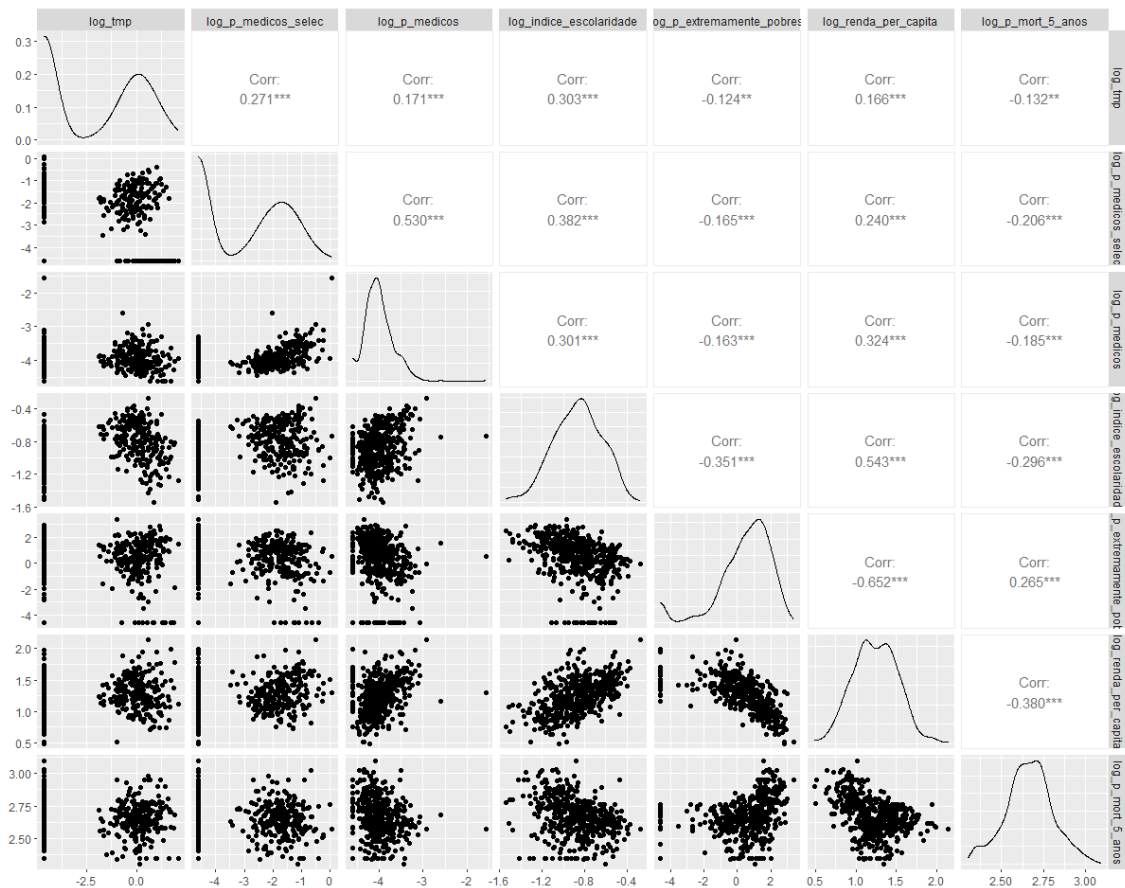


Figura 3.3: Gráfico de correlação entre as variáveis com a transformação logarítmica agregados por município. no triângulo superior temos as correlações entre as variáveis, na diagonal principal a distribuição de densidade estimada via kernel e no triângulo inferior temos o gráfico de pontos entre as variáveis. as variáveis são respectivamente: logaritmo da taxa de mortalidade padrão; proporção de médicos geneticista, pediatras (intensivista e não intensivista) e cardiologista (intensivista e não intensivista); proporção de médicos; índice de escolaridade; proporção de extremamente pobres; renda per capita; proporção de mortalidade infantil até 5 anos presente no CENSO.

A seleção das variáveis a serem consideradas no modelo é uma etapa crucial desta pesquisa, pois as covariáveis socioeconômicas consideradas possuem alta multicolinearidade, tanto entre si quanto em termos espaciais. Essa interdependência complexa pode dificultar a identificação das contribuições individuais de cada variável para a ocorrência das cardiopatias congênitas. Diante desses desafios, optamos por restringir o conjunto de variáveis a serem incluídas no modelo.

Considerando o Modelo BYM2 com dependência espacial baseada na contiguidade da dama, adotou-se o seguinte conjunto de variáveis:

$$\log(\lambda_i) = \beta_0 + \beta_1 (\log p \text{ medicos selec})_i + \beta_2 (\log \text{ renda per capita})_i \quad (3.1)$$

$$+ \beta_3 (p \text{ mort } 5 \text{ anos})_i + b_i \quad (3.2)$$

Optamos por não incluir o índice de escolaridade por ele não apresentar um bom desempenho no modelo e possuir alta correlação com a renda per capita, Para os coeficientes  $\beta$  foi escolhido distribuições *a priori* não informativas, para os hiperparâmetros utilizamos a ideia desenvolvidas pela PC prior, será aplicado uma distribuição *a priori* tipo-2 Gumbel para  $\tau_\beta$  com  $U_\tau$  e  $\alpha_\tau$  tal que  $P(\frac{1}{\sqrt{\tau}} > U_\tau) = \alpha_\tau$ , lembrando que  $\tau$  é o parâmetro de precisão, isso significa que  $\frac{1}{\sqrt{\tau}}$  representa o inverso do desvio padrão, escolhendo  $U_\tau = 1.6$  e  $\alpha_\tau = 0.01$  temos que a probabilidade do desvio ser maior que 1.6 é igual a 0.01. A distribuição *a priori* para  $\phi$  não possui formula fechada, mas seus hiperparâmetros são obtidos da seguinte relação  $P(\phi < U_\phi) = \alpha_\phi$ , escolhendo  $U_\phi = 0.5$  e  $\alpha_\phi = 0.5$  considera-se mesma probabilidade para  $P(\phi > 0.5) = P(\phi < 0.5)$ .

Na Figura 3.4 comparamos o modelo BYM2 com o modelo que considera os efeitos aleatórios independentes, nota-se que apesar das pequenas diferenças nas estimativas e intervalos, não houve mudanças de sinal ou diferenças de significância entre os modelos, assim podemos concluir que houve pouco confundimento espacial.

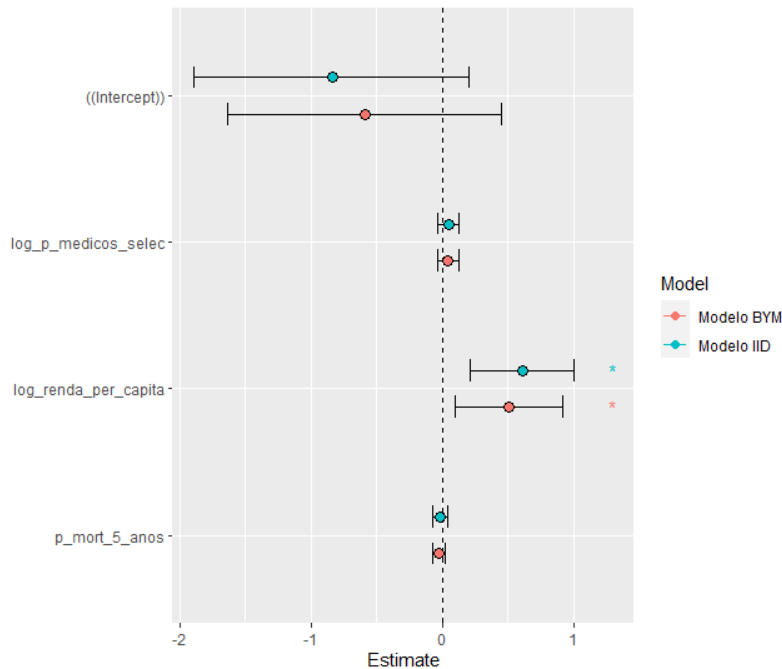


Figura 3.4: Estimativa e Intervalos de Credibilidade das variáveis fixas

Na Figura 3.5 observamos que boa parte dos municípios da região sudoeste e alguns da região noroeste, sudeste e nordeste tiveram um risco relativo significativamente menor. Os municípios que tiveram um risco significativamente maior são da região metropolitana (Porto Alegre, Cachoeirinha, Gravataí e Alvorada), Centro oriental (Vera Cruz e Santa Cruz) e o município de Bento Gonçalves. Nas tabelas 3.3 e 3 foram listados todos os municípios que apresentaram risco relativo significativo.

Na tabela 3.1, apenas log da renda per capita teve uma associação positiva significativa, logo se espera que municípios com maiores rendas per capita tenham



maiores registros de cardiopatias congênitas, o intercepto desse modelo foi significativo mas não possui interpretação. Na tabela 3.2 notamos que a moda do parâmetro de precisão  $\tau_b$  foi de 3.83, isso dá um desvio padrão próximo de 0.5 para o efeito aleatório  $b$ , para  $\phi$  temos a moda igual 0.92, mostrando evidências de haver uma alta dependência espacial.

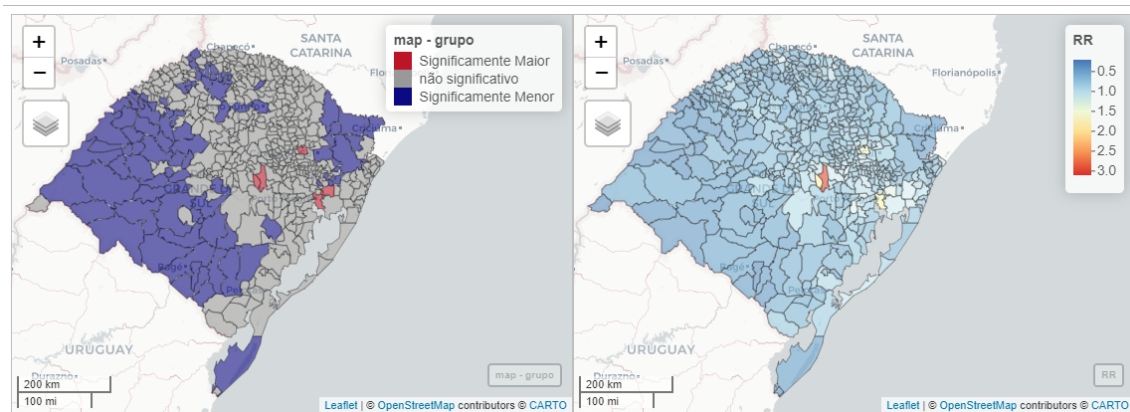


Figura 3.5: Risco Relativo BYM2, a direita temos a moda do risco relativo e na esquerda em vermelhos os RR que foram significativamente maiores que 1, em azul os RR que foram significativamente menores que 1 e em cinza os que não foram significativos.

	$\hat{\beta}_i$	Desvio Padrão	2,5%	97,5%
Intercepto	-5.07	1.43	-7.88	-2.26
log_p_medicos_selec	0.03	0.04	-0.04	0.11
log_renda_per_capita	0.86	0.22	0.43	1.29
p_mort_5_anos	-0.02	0.03	-0.07	0.04

Tabela 3.1: Tabela dos parâmetros fixos do modelo BYM2 com a moda, desvio padrão e intervalo de credibilidade.

	Moda	Desvio Padrão	2,5%	97,5%
$\tau_b$	3.83	1.04	2.98	7.06
$\phi$	0.92	0.09	0.66	1.00

Tabela 3.2: Tabela dos hiperparâmetros do modelo BYM2 com a moda, desvio padrão e intervalo de credibilidade.

Município	número de nascidos vivos	n_cardio	$\hat{\lambda}_i$	2,5%	97,5%
Redentora	2272	1	0.15	0.06	0.39
Hulha Negra	889	0	0.21	0.08	0.58
Dezesseis de Novembro	271	0	0.22	0.09	0.57
Jaquirana	608	0	0.23	0.09	0.63
São Nicolau	678	0	0.23	0.10	0.59

Continua na próxima página



Tabela 3.4 – Continuação

Município	número de nascidos vivos	n_cardio	$\hat{\lambda}_i$	2,5%	97,5%
Garruchos	326	0	0.24	0.09	0.64
Unistalda	302	0	0.25	0.11	0.57
Manoel Viana	855	0	0.26	0.11	0.66
Itacurubi	360	0	0.27	0.12	0.61
Pedras Altas	235	0	0.27	0.11	0.67
São José dos Ausentes	621	0	0.27	0.10	0.80
Quaraí	3376	0	0.27	0.13	0.57
Lajeado do Bugre	373	0	0.27	0.12	0.66
Pirapó	211	0	0.28	0.11	0.82
Itaqui	6551	0	0.28	0.15	0.55
São Valério do Sul	473	0	0.28	0.13	0.67
Santana da Boa Vista	752	0	0.29	0.14	0.65
Monte Alegre dos Campos	420	1	0.30	0.14	0.68
Benjamin Constant do Sul	405	0	0.30	0.13	0.78
Jari	416	1	0.30	0.15	0.68
Capão Bonito do Sul	197	0	0.30	0.12	0.92
Derrubadas	383	0	0.30	0.13	0.79
Bossoroca	717	0	0.31	0.15	0.65
Sant'Ana do Livramento	13109	4	0.31	0.17	0.59
Dois Irmãos das Missões	349	0	0.31	0.14	0.71
Capão do Cipó	385	0	0.32	0.14	0.77
Vitória das Missões	367	0	0.32	0.16	0.68
Santo Antônio das Missões	1320	0	0.32	0.16	0.69
Aceguá	716	0	0.32	0.14	0.80
Santa Vitória do Palmar	4550	1	0.33	0.13	0.92
Pinhal da Serra	262	1	0.33	0.14	0.94
Sagrada Família	333	0	0.33	0.15	0.78
Cacequi	1436	0	0.33	0.17	0.68
Jaboticaba	430	0	0.33	0.16	0.77
Esperança do Sul	423	0	0.33	0.13	0.98
São José das Missões	320	0	0.33	0.16	0.76
Bom Jesus	1903	0	0.33	0.16	0.75
São Pedro das Missões	266	0	0.34	0.16	0.78
Jacuizinho	286	0	0.34	0.17	0.76
Dilermando de Aguiar	328	0	0.34	0.16	0.80
Inhacorá	343	0	0.34	0.16	0.79
Roque Gonzales	788	0	0.34	0.17	0.76
Candiota	1561	0	0.34	0.16	0.80
Rosário do Sul	5352	1	0.34	0.19	0.64
Braga	490	0	0.34	0.16	0.80
São Francisco de Assis	2079	1	0.34	0.19	0.68
Maçambará	589	1	0.35	0.17	0.79
Mata	530	0	0.35	0.17	0.82
Jóia	1028	0	0.36	0.18	0.79
Cacique Doble	660	0	0.36	0.17	0.88

Continua na próxima página

Tabela 3.4 – Continuação

Município	número de nascidos vivos	n_cardio	$\hat{\lambda}_i$	2,5%	97,5%
Charrua	404	0	0.36	0.17	0.87
Herval	810	0	0.37	0.18	0.79
Miraguaí	713	0	0.37	0.18	0.84
Coronel Bicaco	1143	1	0.37	0.19	0.81
São Borja	9286	4	0.37	0.21	0.69
Quevedos	260	0	0.38	0.19	0.82
Vila Nova do Sul	370	0	0.38	0.17	0.95
São Miguel das Missões	1009	2	0.39	0.21	0.78
Dom Feliciano	1859	0	0.39	0.21	0.77
Ervál Seco	1008	0	0.39	0.20	0.81
Rolador	249	1	0.39	0.19	0.87
Esmeralda	406	0	0.39	0.18	0.93
São Vicente do Sul	1024	0	0.39	0.20	0.81
Bagé	18982	8	0.40	0.25	0.66
Dom Pedrito	5251	3	0.40	0.22	0.77
Bom Progresso	345	0	0.40	0.20	0.90
São Luiz Gonzaga	5432	1	0.40	0.23	0.74
Fontoura Xavier	1273	1	0.41	0.21	0.87
Novo Barreiro	528	0	0.42	0.20	0.95
Pinheiro Machado	1307	0	0.42	0.22	0.83
Lavras do Sul	862	1	0.42	0.23	0.85
Erebango	487	0	0.42	0.20	0.97
Alegria	406	0	0.42	0.21	0.94
Pontão	495	1	0.42	0.23	0.86
Coxilha	472	0	0.43	0.21	0.95
Uruguaiana	21726	13	0.43	0.27	0.71
Cerro Grande	350	0	0.43	0.21	0.97
Campo Novo	864	1	0.43	0.22	0.93
São Gabriel	9279	6	0.43	0.27	0.74
Independência	801	0	0.44	0.22	0.98
São Francisco de Paula	3232	0	0.44	0.27	0.77
Tenente Portela	2560	1	0.45	0.24	0.91
Entre-Ijuís	1242	0	0.45	0.24	0.92
Alegrete	10668	7	0.45	0.28	0.76
Caçapava do Sul	4605	1	0.45	0.25	0.84
Chiapetta	562	0	0.46	0.24	0.98
Palmeira das Missões	5853	1	0.47	0.28	0.80
Boa Vista das Missões	292	0	0.47	0.25	0.99
Parobé	9212	4	0.47	0.27	0.87
São Pedro do Sul	1906	2	0.47	0.27	0.93
Tupanciretã	3239	2	0.47	0.28	0.84
Quatro Irmãos	221	0	0.47	0.26	0.96
São Sepé	2974	0	0.48	0.27	0.87
Ronda Alta	1366	0	0.48	0.26	0.97
Canela	7757	4	0.49	0.28	0.91

Continua na próxima página

**Tabela 3.4 – Continuação**

<b>Município</b>	número de nascidos vivos	n_cardio	$\hat{\lambda}_i$	2,5%	97,5%
Santa Bárbara do Sul	1132	0	0.49	0.28	0.91
Giruá	2305	1	0.49	0.27	0.96
Piratini	2141	2	0.49	0.28	0.96
Santo Augusto	2214	1	0.51	0.29	0.97
Restinga Sêca	1946	1	0.51	0.29	0.98
Carazinho	9929	4	0.52	0.33	0.85
Canguçu	6646	3	0.52	0.32	0.91
Vacaria	11519	8	0.53	0.33	0.93
Soledade	4487	1	0.54	0.33	0.93
Sapiranga	13809	7	0.55	0.36	0.87
Santiago	6758	4	0.55	0.34	0.98
Lagoa Vermelha	4554	3	0.56	0.34	0.99
Taquara	7966	2	0.59	0.37	0.94
São Leopoldo	35022	23	0.61	0.44	0.87
Passo Fundo	33981	26	0.64	0.47	0.91
Caxias do Sul	72054	67	0.79	0.64	0.99

Tabela com o número de nascidos vivos, número de nascidos vivos com cardiopatia congênita, moda estimada do risco relativo e limites de 2,5% e 95% para os intervalos de credibilidade do risco relativo  $\lambda_i$  para os municípios que tiveram um risco relativo significativamente menor que 1 .

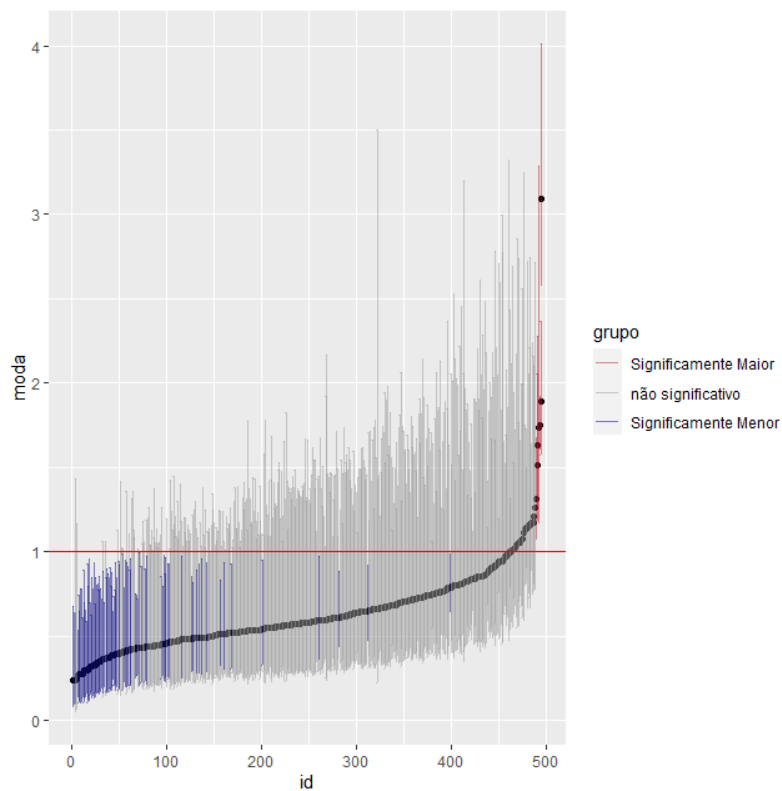


Figura 3.6: Risco relativo ordenado pela moda, cada barra representa 95% do intervalo de credibilidade.

<b>Município</b>	n nascidos vivos	n cardio	$\hat{\lambda}_i$	2,5%	97,5%
Santa Cruz do Sul	19081	80	3.10	2.58	4.01
Alvorada	37429	91	1.88	1.57	2.35
Vera Cruz	3326	11	1.83	1.26	3.38
Porto Alegre	216437	456	1.74	1.59	1.91
Bento Gonçalves	18100	41	1.64	1.27	2.28
Cachoeirinha	22049	44	1.54	1.21	2.08
Gravataí	41303	69	1.32	1.08	1.68

Tabela 3.3: Tabela dos Municípios que tiveram um risco relativo significativamente maior que 1.

## 4 Conclusão

O objetivo deste estudo foi compreender a distribuição das cardiopatias congênitas no Rio Grande do Sul durante o período de 2010 até 2021, agregadas pelo município de residência materna. Para alcançar esse propósito, utilizamos uma abordagem espacial bayesiana empregando o modelo BYM2 proposto por [Simpson et al. \[2017\]](#). Com o intuito de obter as estimativas da distribuição *a posteriori*, o modelo foi implementado utilizando o método INLA proposto por [Rue et al. \[2009b\]](#).

No modelo podemos perceber uma alta dependência espacial, a estimativa do do hiperparâmetro  $\hat{\phi} = 0,921$  ficou muito próximo de 1, para evitar multicolinearidade e confundimento espacial consideramos apenas 3 covariáveis: log proporção de médicos selecionados (geneticista, pediatras e cardiologista), proporção de mortalidade até 5 anos e log renda per capita, sendo só a renda significativa no modelo, Dessa forma, observou-se que municípios com renda per capita mais elevada tendem a apresentar um maior registro de casos de cardiopatias congênitas. Apesar disso, não foi possível identificar associações fortes entre variáveis socioeconômicas e o risco relativo de ocorrência das cardiopatias congênitas.

É importante ressaltar que, por se tratar de um estudo ecológico, as associações encontradas não podem ser interpretadas como efeitos causais. Outra limitação enfrentada foi a presença de subnotificação nos registros de cardiopatias congênitas, o que pode impactar os resultados obtidos, para trabalhos futuros espera-se trabalhar conjuntamente com o Sistema de Informações de Mortalidade (SIM) e com o Sistema de Informação de Nascidos Vivos (SINASC) e a possibilidade de abranger os outros grupos de anomalias congênitas.

A contribuição relevante deste estudo está em ressaltar a importância de explorar a dependência espacial em modelos que tentam explicar cardiopatias congênitas. Além disso, o trabalho identificou regiões específicas no Rio Grande do Sul com prevalências maiores ou menores em relação ao resto do Estado, o que pode auxiliar os gestores da área da saúde na formulação de planos de ação mais efetivos e direcionados.

ovos estudos serão necessários para avaliar o impacto da covid-19 nos indicadores aqui avaliados, tendo em vista que o Brasil foi o país latino com maior quantidade de casos de covid-19 registrados até 2020, sendo 0,260 a 18 anos.<sup>32, 33</sup> Apesar da doença acometer indivíduos de qualquer idade, foi observado uma incidência considerável em crianças e adolescentes.<sup>33, 34</sup> Embora as manifestações clínicas sejam mais leves, foi identificada uma complicação rara, chamada de síndrome inflamatória multissistêmica pediátrica temporalmente associada à covid-19 (SIM-P), com potencial de agravamento, podendo levar ao óbito.<sup>34, 35</sup> 66 Secretaria de Vigilância em Saúde e Ambiente | MS Saúde Brasil: análise da situação de saúde e uma visão integrada sobre os fatores de risco para anomalias congênitas Tendo em vista os aspectos observados, vale destacar também a necessidade de ampliação da investigação dos óbitos fetais.<sup>19,21</sup> No mais, é importante garantir o progresso já alcançado e investir na qualificação do atendimento ao parto e do pré-natal considerando no planejamento, as especificidades de cada região brasileira. A produção deste trabalho permitiu observar uma série de melhorias relacionadas aos indicadores de mortalidade infantil no Brasil. Porém, desafios como as diferenças regionais, mortes por causas evitáveis e aumento de óbitos provindos de fatores maternos evidenciam cada vez mais a importância do tema com prioridade para que a redução da mortalidade infantil continue acontecendo. A mobilização não somente de todas as esferas de governo, mas de toda a sociedade e de cada cidadão, é importante para consolidar essa redução, num movimento em defesa da vida.<sup>8</sup> Desse modo, para a redução das desigualdades regionais, faz-se necessário investimento em políticas de saúde que assegurem cuidados à gestação, ao parto e ao nascimento, assim como uma melhoria no acesso, práticas e medidas direcionadas à qualificação da rede assistencial à gestante e ao recém-nascido durante o período pré-natal e pós-natal. Tais melhorias aumentariam a probabilidade de redução das mortes por causas evitáveis e, consequentemente, redução significativa das taxas de mortalidade infantil e na infância. Esse cuidado apropriado ressalta a importância do trabalho em promoção da saúde que direciona ao aprimoramento de políticas públicas por meio da priorização de intervenções de acordo com o perfil epidemiológico identificado, o que pode gerar um resultado mais positivo para a qualidade de vida dos brasileiros.

e com o diagnóstico precoce e o acompanhamento médico adequado muitas dessas mortes podem ser evitadas e seus efeitos diminuídos.

A taxa de mortalidade infantil representa um indicador importante do nível econômico e de desenvolvimento de uma sociedade. No livro *saúde Brasil ...* é mencionado que houve uma redução significativa da mortalidade infantil e na infância a partir da década de 1990, com uma queda de 60% na mortalidade na infância. No entanto, ainda existe uma alta carga global de morbidade em crianças e jovens, com 7,4 milhões de mortes de crianças em 2019, a maioria por causas evitáveis ou tratáveis.

Os fatores principais que contribuem para as altas taxas de mortalidade infantil e na infância são os problemas sociais decorrentes de condições socioeconômicas precárias, idade materna, doenças infecciosas e dificuldades de acesso a cuidados pré-natais de qualidade. A Organização Mundial da Saúde (OMS) estabeleceu metas nos Objetivos de Desenvolvimento do Milênio (ODM), concluídos em 2015, e nos Objetivos de Desenvolvimento Sustentável (ODS), que incluem a redução da mortalidade materna, de recém-nascidos e de menores de 5 anos, além de garantir acesso à saúde de qualidade e promover o bem-estar para todas as idades até 2030.

Essas metas visam chamar a atenção para o tema e contribuir com políticas públicas para a redução da mortalidade infantil e a eliminação de mortes evitáveis.

## Referências Bibliográficas

- Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2004.
- Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):pp. 192–236, 1974. ISSN 00359246. URL <http://www.jstor.org/stable/2984812>.
- Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991. URL <https://EconPapers.repec.org/RePEc:spr:aistmt:v:43:y:1991:i:1:p:1-20>.
- Marta Blangiardo and Michela Cameletti. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley, 2015.
- Lorenzo D. Botto, Adolfo Correa, and J. David Erickson. Racial and Temporal Variations in the Prevalence of Heart Defects. *Pediatrics*, 107(3):e32–e32, 03 2001. ISSN 0031-4005. doi: 10.1542/peds.107.3.e32. URL <https://doi.org/10.1542/peds.107.3.e32>.
- P. A. Boyd, B. Armstrong, H. Dolk, B. Botting, S. Pattenden, L. Abramsky, J. Rankin, M. Vrijheid, and D. Wellesley. Congenital anomaly surveillance in england: Ascertainment deficiencies in the national system. *BMJ: British Medical Journal*, 330(7481):27–29, 2005. ISSN 09598138, 17561833. URL <http://www.jstor.org/stable/25458567>.
- Noel A. C. Cressie. *Statistics for spatial data*. Wiley, revised edition edition, January 1993. ISBN 0471002550. URL <http://www.worldcat.org/isbn/0471002550>.
- C Dean, M. Ugarte, and Ana Militino. Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57:197–202, 04 2001. doi: 10.1111/j.0006-341X.2001.00197.x.
- Peter J. Diggle. *Analysis of Spatial Point Patterns (3rd edition)*. Chapman and Hall/CRC, 2013.
- Elisabeth Barboza et al. França. Principais causas da mortalidade na infância no brasil, em 1990 e 2015: estimativas do estudo de carga global de doença. *Revista Brasileira de Epidemiologia [online]*, 2017.



- A.E. Gelfand, M. Fuentes, P. Guttorp, and P. Diggle. *Handbook of Spatial Statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2010. ISBN 9781420072877. URL <http://books.google.com/books?id=EFbbcMFZ2mMC>.
- Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461, 1946. ISSN 00804630. URL <http://www.jstor.org/stable/97883>.
- Kathy J Jenkins, Adolfo Correa, Jeffrey A Feinstein, Lorenzo Botto, Amy E Britt, Stephen R Daniels, Marsha Elixson, Carole A Warnes, and Catherine L Webb. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the american heart association council on cardiovascular disease in the young: endorsed by the american academy of pediatrics. *Circulation*, 115(23):2995–3014, 2007.
- Elias Krainski, Virgilio Gómez-Rubio, Haakon Bakka, Castro-Camilo Daniela Lenzi, Amanda, Daniel Simpson, Finn Lindgren, and Håvard Rue. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and Inla*. CRC Press, 2018.
- Andrew B Lawson. *Bayesian Disease Mapping - Hierarchical Modeling in Spatial Epidemiology, Third Edition*. CRC Press (2018), 2018.
- Brian G. Leroux, Xingye Lei, and Norman E. Breslow. Estimation of disease rates in small areas: A new mixed model for spatial dependence. 2000.
- Ying Macnab. On gaussian markov random fields and bayesian disease mapping. *Statistical methods in medical research*, 20:49–68, 02 2011. doi: 10.1177/0962280210371561.
- Thiago G. Martins, Daniel Simpson, Finn Lindgren, and Håvard Rue. Bayesian computing with inla: new features. *Computational Statistics and Data Analysis*, 2012.
- Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman and Hall/CRC, 1989.
- Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 12 1953. ISSN 0021-9606. doi: 10.1063/1.1699114. URL <https://doi.org/10.1063/1.1699114>.
- Ministério da Saúde. *Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde 10 (CID-10)*, a. [acesso em: 2021 Fev 24]. Disponível em: <http://tabnet.datasus.gov.br/cgi/sih/mxqid10lm.htm>.
- Ministério da Saúde. *Declaração de Nascido Vivo, Manual de instruções para preenchimento*, b. [acesso em: 2022 Jul 02]. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/vigilancia/declaracao-de-nascido-vivo-manual-de-instrucoes-para-preenchimento>.

- Ministério da Saúde. *Sistema de Informações sobre Nascidos Vivos (SINASC)*, c. [acesso em: 2021 Fev 12]. Disponível em: <http://svs.aids.gov.br/dantps/cgiae/sinasc/apresentacao/#wrapper>.
- Ministério da Saúde. *Saúde Brasil 2020/2021: anomalias congênitas prioritárias para a vigilância ao nascimento*. Ministério da Saúde, 2022.
- Ministério da Saúde. *Saúde Brasil 2022: análise da situação de saúde e uma visão integrada sobre os fatores de risco para anomalias congênitas*. Ministério da Saúde, 2023.
- J. Diggle Peter and Paulo J. Ribeiro. *Model-based Geostatistics*. Springer Series in Statistics, 2007.
- Valdester Cavalcante Pinto Júnior, Klébia Magalhães P. Castello Branco, Rodrigo Cardoso Cavalcante, Waldemiro Carvalho Junior, José Rubens Costa Lima, Sílvia Maria de Freitas, Maria Nazaré de Oliveira Fraga, and Nayana Maria Gomes de Souza. Epidemiology of congenital heart disease in brazil. *Brazilian Journal of Cardiovascular Surgery*, 30(2):219–224, Mar 2015. ISSN 0102-7638. doi: 10.5935/1678-9741.20150018. URL <https://doi.org/10.5935/1678-9741.20150018>.
- Albert Stuart Reece and Gary Kenneth Hulse. Patterns of cannabis- and substance-related congenital general anomalies in europe: A geospatiotemporal and causal inferential study. *Pediatric Reports*, 15(1):69–118, 2023. ISSN 2036-7503. doi: 10.3390/pediatric15010009. URL <https://www.mdpi.com/2036-7503/15/1/9>.
- Dianne E. Richardson and 10th International Symposium on Spatial Data Handling Peter Oosterom. *Advances in Spatial Data Handling*. Springer Berlin, Heidelberg, 2002.
- Andrea Riebler, Sigrunn H Sørbye, Daniel Simpson, and Håvard Rue. An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165, 2016. doi: 10.1177/0962280216660421. URL <https://doi.org/10.1177/0962280216660421>. PMID: 27566770.
- Havard Rue and Leonhard Held. *Gaussian Markov Random Fields Theory and Applications*. 2005.
- Håvard Rue and Sara Martino. Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192, 2007. ISSN 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2006.07.016>. URL <https://www.sciencedirect.com/science/article/pii/S0378375807000845>. Special Issue: Bayesian Inference for Stochastic Processes.
- Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319392, 2009. , 2009a.

- Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2): 319–392, 04 2009b. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2008.00700.x. URL <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Michael Sherman. *Spatial Statistics and Spatio[**U+2010**]Temporal Data: Covariance Functions and Directional Properties*. Wiley, 2010.
- Heather Siefkes, Laura R. Kair, Annamarie Saarinen, and Satyan Lakshminrusimha. Inadequacies of hospital-level critical congenital heart disease screening data reports: implications for research and quality efforts. *Journal of Perinatology*, 41(7):1611–1620, Jul 2021. ISSN 1476-5543. doi: 10.1038/s41372-020-00783-z. URL <https://doi.org/10.1038/s41372-020-00783-z>.
- Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28, 2017. ISSN 08834237, 21688745. URL <http://www.jstor.org/stable/26408114>.
- Jon Wakefield. Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–183, 06 2006. ISSN 1465-4644. doi: 10.1093/biostatistics/kxl008. URL <https://doi.org/10.1093/biostatistics/kxl008>.
- A. Waller, Lance and Carol A. Gotway. *Applied Spatial Statistics for Public Health Data*. Wiley, 2004.
- Joshua Warren, Montserrat Fuentes, Amy Herring, and Peter Langlois. Bayesian spatial–temporal model for cardiac congenital anomalies and ambient air pollution risk assessment. *Environmetrics*, 23(8):673–684, 2012. doi: <https://doi.org/10.1002/env.2174>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2174>.
- World-Health-Organization. *Congenital anomalies*. [acesso em: 2021 Feb 12]. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/congenital-anomalies>.
- Wen Zhang, Hong-Yan Xu, Yan-Chun Zhang, and Kai-Bo Liu. Delayed diagnosis of critical congenital heart defects predicting risk factors and survival rate in newborns in beijing: a retrospective study. *Journal of International Medical Research*, 49(7):03000605211028028, 2021. doi: 10.1177/03000605211028028. URL <https://doi.org/10.1177/03000605211028028>. PMID: 34264137.
- Meghan Zimmerman, Alison Smith, Craig Sable, Michelle Echko, Lauren Wilner, Helen Olsen, Hagos Tasew, Ashish Awasthi, Zulfiqar Bhutta, Jackie Boucher, Franz Castro, Paolo Cortesi, Manisha Dubey, Florian Fischer, Samer Hamidi, Simon Hay, Chi Hoang, Christopher Hugo-Hamman, Kathy Jenkins, and Nicholas Kassebaum. Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the global burden of disease study 2017. 4: 185–200, 01 2020. doi: 10.1016/S2352-4642(19)30402-X.