

Marina Cardeal Sudré

**Modelo de regressão para dados
longitudinais: estudo do perímetro cefálico
de crianças expostas e não expostas ao vírus
Zika**

Niterói - RJ, Brasil

10 de maio de 2021

Marina Cardeal Sudré

**Modelo de regressão para dados
longitudinais: estudo do perímetro
cefálico de crianças expostas e não
expostas ao vírus Zika**

Trabalho de Conclusão de Curso

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

Orientador(a): Prof. Dr. Luis Guillermo Coca Velarde

Niterói - RJ, Brasil

10 de maio de 2021

Marina Cardeal Sudré

**Modelo de regressão para dados
longitudinais: estudo do perímetro cefálico
de crianças expostas e não expostas ao vírus
Zika**

Monografia de Projeto Final de Graduação sob o título “*Modelo de regressão para dados longitudinais: estudo do perímetro cefálico de crianças expostas e não expostas ao vírus Zika*”, defendida por Marina Cardeal Sudré e aprovada em 10 de maio de 2021, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Luis Guillermo Coca Velarde
Departamento de Estatística – UFF

Profa. Dra. Estelina Serrano de Marins Capistrano
Departamento de Estatística – UFF

Profa. Dra. Ludmilla da Silva Viana Jacobson
Departamento de Estatística – UFF

Niterói, 10 de maio de 2021

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

S943m Sudré, Marina Cardeal
Modelo de regressão para dados longitudinais: estudo do
perímetro cefálico de crianças expostas e não expostas ao
vírus Zika / Marina Cardeal Sudré ; Luis Guillermo Coca
Velarde, orientador. Niterói, 2021.
43 f.

Trabalho de Conclusão de Curso (Graduação em
Estatística)-Universidade Federal Fluminense, Instituto de
Matemática e Estatística, Niterói, 2021.

1. Microcefalia. 2. Dados longitudinais. 3. Perímetro
cefálico. 4. Produção intelectual. I. Coca Velarde, Luis
Guillermo, orientador. II. Universidade Federal Fluminense.
Instituto de Matemática e Estatística. III. Título.

CDD -

Resumo

A microcefalia é uma malformação congênita caracterizada pelo perímetro cefálico reduzido para a idade ou sexo, acompanhada de alterações no sistema nervoso central. No Brasil, no ano de 2015, foi observado um aumento inesperado no nascimento de crianças com microcefalia, coincidindo a epidemia de infecção pelo vírus Zika. Este fato foi alvo de muitas pesquisas no Brasil e no mundo. Diante disso, este estudo tem como objetivo modelar a evolução do perímetro cefálico de crianças acompanhadas no Hospital Antônio Pedro (UPC/HUAP-UFF) durante seu primeiro ano de vida e identificar quais são os fatores relacionados a essa evolução. O uso de modelos de regressão linear clássicos tem a limitação de que os dados precisam ser independentes, restrição que não permite explorar de forma adequada os dados deste estudo, dispostos de forma longitudinal, isto é, observações feitas em sequência temporal em uma mesma criança. Por isso, neste trabalho foram utilizados modelos de regressão para dados longitudinais sob enfoque Bayesiano. O modelo contou com 62 crianças com pelo menos 5 observações cada, totalizando 310 dados. A variável resposta, perímetro cefálico, foi analisada em função das variáveis explicativas comprimento, idade, peso, sexo e grupo de exposição (expostos ao Zika Vírus, não expostos ao Zika vírus, Zika suspeito e microcefalia sem exantema materno), em que comprimento, peso e idade são as variáveis que variam com o tempo. Após serem avaliadas separadamente, notou-se que as variáveis sexo e o grupo de expostos ao ZIKV não possuíam valor significativo em relação ao perímetro cefálico e, portanto, foram removidas. Já a variável idade, quando aplicada separadamente, mostrou exercer influência sobre o perímetro cefálico, mas quando testada em conjunto com as outras variáveis independentes não foi estatisticamente significativa. De acordo com os métodos de seleção de modelos Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiana (BIC), os resultados obtidos mostraram que o modelo completo apresentou melhor ajuste dos dados. Portanto, as variáveis comprimento, peso, grupo de Zika suspeito e grupo de microcefalia sem exantema materno se apresentaram associadas ao perímetro cefálico das crianças. Dessa forma, vimos que ajustar modelos para dados longitudinais sob enfoque Bayesiano permite analisar problemas altamente complexos, de forma eficiente e de fácil interpretação.

Palavras-chave: Microcefalia. Dados Longitudinais. Perímetro Cefálico.

Dedicatória

A minha filha, Malu, que é a força que me move e me faz querer ser uma pessoa melhor em todos os sentidos. A minha mãe, Valéria, minha maior apoiadora e incentivadora, que não mede esforços para me ver feliz. Ao meu esposo, Luciano, companheiro de todos os momentos. Ao meu pai, Marcos, que lá do céu está olhando por mim.

Agradecimentos

Em tempos tão difíceis de pandemia que estamos vivendo, esse capítulo de agradecimentos ganhou uma importância maior, pois acredito que sem uma rede de apoio concluir esta etapa importante seria praticamente impossível.

Diante disso, gostaria de começar esse capítulo agradecendo a Deus por ter me dado forças e saúde pra chegar até aqui.

Aos meus pais, por todo o apoio que sempre me deram, por me mostrarem que o estudo sempre será o melhor caminho e por me ensinarem a ser uma pessoa de caráter íntegro. Meu pai, mesmo não estando mais por perto, espero que aonde estiver esteja orgulhoso e com sensação de dever cumprido. À minha mãe, que desde que meu pai se foi, teve que ser pai e mãe, nunca deixou a peteca cair e sempre nos deu todo o amor e suporte necessário. Nenhuma palavra dita vai ser suficiente para te agradecer por tanto, mãe.

Ao meu esposo Luciano e à minha filha Malu, por compreenderem minha ausência todos esses meses, por tantos dias sem poder dar atenção a eles pra me dedicar a este trabalho. Vocês são o meu alicerce e tudo o que eu faço é por vocês.

Ao meu orientador Guillermo Velarde, que foi peça fundamental para a conclusão deste trabalho. Obrigada por ser tão solícito e estar sempre disponível para ajudar. Obrigada pela compreensão e paciência nos momentos de dificuldade. Obrigada por acreditar em mim e por toda a dedicação. Sei que não poderia ter escolhido alguém melhor para exercer esse papel tão importante.

Aos amigos que fiz ao longo desses anos na faculdade, em especial à Karolina, Rayssa, Bernardo e à minha companheira de orientação, Larissa, que me ajudou e apoiou em tantos momentos difíceis.

Por fim, gostaria de agradecer à Professora Claudete A. Araújo Cardoso e ao Luis Antonio Bataglin Dalcaste da Faculdade de Medicina da Universidade Federal Fluminense por ter cedido os dados do Projeto de Pesquisa “Monitoramento do perímetro cefálico no primeiro ano de vida em crianças expostas e não expostas ao vírus Zika: estudo prospectivo

de coorte” para que este trabalho pudesse ser realizado.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
1.0.1	Zika Vírus	p. 11
1.0.2	Microcefalia	p. 12
1.1	Motivação	p. 13
1.2	Objetivo	p. 13
2	Materiais e Métodos	p. 14
2.1	Dados	p. 14
2.1.1	População e período de estudo	p. 14
2.1.2	Local de estudo	p. 15
2.1.3	Crítérios de inclusão e exclusão	p. 15
2.1.4	Coleta dos dados	p. 16
2.2	Metodologia	p. 16
2.2.1	Estrutura de Variância e Covariância	p. 17
2.2.2	Enfoque Bayesiano	p. 19
2.2.3	Teorema de Bayes	p. 20
2.2.4	Distribuição a Priori	p. 20
2.2.5	Monte Carlo via Cadeias de Markov	p. 21
2.2.5.1	Cadeias de Markov	p. 21

2.2.5.2	Amostrador de Gibbs	p. 22
2.2.5.3	Algoritmo de Metropolis-Hastings	p. 22
2.2.6	Critério de seleção do modelo	p. 23
2.2.7	Variáveis incluídas no modelo	p. 24
3	Resultados e Discussão	p. 26
3.1	Análise do Modelo	p. 28
4	Conclusão	p. 34
	Referências	p. 36
	Anexo 1 – Esquema de escolha das consultas	p. 38
	Anexo 2 – Script OpenBugs	p. 39
	Anexo 3 – Modelos individuais apresentados na Tabela 6	p. 40
	Anexo 4 – Modelo completo	p. 42

Lista de Figuras

1	Medição do perímetro cefálico	p. 16
2	Dados longitudinais do perimetro cefalico das 62 crianças no primeiro ano de vida.	p. 30
3	Monitoração da Convergência dos Parâmetros do Modelo completo . .	p. 31
4	Densidades a Posteriori dos Parâmetros do Modelo completo	p. 31
5	Autocorrelações das cadeias geradas	p. 32

Lista de Tabelas

1	Estatísticas Descritivas das características das crianças na consulta 1 .	p. 29
2	Estatísticas Descritivas das características das crianças na consulta 2 .	p. 29
3	Estatísticas Descritivas das características das crianças na consulta 3 .	p. 29
4	Estatísticas Descritivas das características das crianças na consulta 4 .	p. 29
5	Estatísticas Descritivas das características das crianças na consulta 5 .	p. 29
6	Estimativas dos modelos individuais	p. 30
7	Estimativas dos parâmetros do Modelo Completo	p. 32
8	Avaliação da qualidade do modelo completo e dos modelos com apenas uma variável explicativa	p. 33

1 Introdução

1.0.1 Zika Vírus

No ano de 2015, foi observado no Brasil um surto de Zika vírus (ZIKV), um flavivírus pertencente à família Flaviviridae, cuja transmissão é feita, especialmente, através da picada do mosquito *Aedes Aegypti* (SCHULER-FACCINI et al., 2016). Seus principais sintomas incluem febre, cefaleia (dor de cabeça), dor muscular, manchas avermelhadas na pele e prurido (coceiras) (GARCIA, 2018). Em grande parte dos pacientes sintomáticos infectados pelo ZIKV, a doença é apresentada de forma branda com duração de aproximadamente uma semana (LUZ; SANTOS; VIEIRA, 2015).

Inicialmente, o ZIKV foi isolado em macacos rhesus em Uganda, em 1947 (DICK; KITCHEN; HADDOW, 1952) e, durante muito tempo, foi retratado como responsável por infecções raras em humanos em regiões na África e na Ásia, até 2007, quando ocorreu uma epidemia do vírus nas ilhas da Micronésia (DUFFY et al., 2009). Em 2013, foi registrada na Polinésia Francesa uma grande epidemia (ZANLUCA et al., 2015) e, posteriormente, o vírus se alastrou por regiões da Oceania, até chegar às Américas em 2014 (TOGNARELLI et al., 2016).

No Brasil, em localidades do Rio Grande do Norte, em outubro de 2014, foi reportada a ocorrência de uma doença cujos sintomas se pareciam com os da Chikungunya, todavia, os testes de detecção da doença resultaram negativos (ALBUQUERQUE et al., 2018). Em março do ano seguinte, o Centro de Informações Estratégicas de Vigilância em Saúde (Cievs) Nacional já havia recebido notificações parecidas de todos os Estados da Região Nordeste (FANTINATO et al., 2016). Esse Centro tem papel de detectar, monitorar e assegurar a administração de emergências de saúde pública, ampliando, assim, a capacidade de resposta a essas emergências, além de ser essencial para o enfrentamento de pandemias (CARMO; PENNA; OLIVEIRA, 2008).

Como o histórico do vírus nos poucos casos anteriormente reportados no mundo não era de casos graves ou fatais, a princípio não parecia ter motivos para grande apreensão.

Porém, um estado de alerta foi instalado no segundo semestre de 2015, quando foi reportado um aumento expressivo no número de casos de recém nascidos com microcefalia (SAÚDE, 2017).

1.0.2 Microcefalia

A microcefalia é uma condição caracterizada pelo perímetro cefálico reduzido em mais de 2 desvios padrões da referência para idade e sexo (DEVAKUMAR et al., 2018). A longo prazo, as consequências da microcefalia podem variar de atrasos leves no desenvolvimento a deficiências intelectuais e motoras severas, como paralisia cerebral (SCHULER-FACCINI et al., 2016).

No Brasil, no segundo semestre de 2015, foi observado um aumento repentino no número de casos de recém nascidos com microcefalia (ALBUQUERQUE et al., 2018), coincidentemente com o reporte do surto do vírus Zika no país. Em 2016 ficou comprovada a relação causal entre a infecção pelo ZIKV durante a gestação e o cenário de microcefalia em bebês (RASMUSSEN et al., 2016). Vale ressaltar que além do ZIKV, há vários fatores responsáveis pelo aparecimento dessa malformação, como anormalidades genéticas, exposição a álcool, drogas, toxinas, radiação e alguns vírus, como herpes, sífilis, HIV, rubéola, entre outros (NUNES et al., 2016).

As microcefalias podem ser classificadas como congênita, isto é, evidente ao nascimento, ou pós-natal, situação em que a criança apresenta circunferência craniana normal no nascimento e, em geral, dentro dos dois primeiros anos de vida, esta medida cai e fica com o tamanho abaixo do esperado. Ao passo que nas microcefalias congênicas os fatores agem no sistema nervoso central ainda durante a gestação e englobam as infecções maternas citadas anteriormente, na microcefalia pós-natal as razões mais habituais são traumatismo craniano, desnutrição grave, alterações vasculares, entre outras (NUNES et al., 2016).

Considerando que, mesmo após o nascimento, a microcefalia ainda pode ser desenvolvida, é essencial que o perímetro cefálico da criança seja monitorado no primeiro ano de vida. A medição do perímetro cefálico é um método de diagnóstico simples e confiável na detecção dessa malformação (NUNES et al., 2016).

1.1 Motivação

Mesmo decorridos cinco anos após a Declaração de Emergência em Saúde Pública pelo crescimento do número de recém-nascidos com microcefalia, dados do Boletim Epidemiológico da Síndrome Congênita associada à infecção pelo vírus zika (SCZ), lançado em Novembro de 2019, indicam que, apesar da redução significativa, novos casos de microcefalia ainda estão sendo registrados. Até outubro de 2019, 55 casos foram verificados em todo o Brasil, dos quais 29 foram de neonatos ou crianças que nasceram neste ano (MINISTÉRIO DA SAÚDE, SECRETARIA DE VIGILÂNCIA EM SAÚDE, 2019). Este fato ensejou este estudo, de forma a contribuir de alguma maneira para a evolução do enfrentamento a doença.

1.2 Objetivo

O objetivo do presente estudo é modelar a evolução do perímetro cefálico de crianças acompanhadas no Hospital Antônio Pedro (UPC/HUAP-UFF) durante seu primeiro ano de vida e identificar quais são os fatores relacionados a essa evolução, através de modelos de regressão para dados longitudinais.

2 Materiais e Métodos

2.1 Dados

A base de dados utilizada neste estudo é proveniente do Projeto de Pesquisa “Monitoramento do perímetro cefálico no primeiro ano de vida em crianças expostas e não expostas ao vírus Zika: estudo prospectivo de coorte” de Claudete A. Araújo Cardoso, pela Faculdade de Medicina da Universidade Federal Fluminense.

2.1.1 População e período de estudo

A população de estudo é formada por crianças nascidas no período de abril de 2016 a junho de 2017, cujas mães manifestaram exantema (manchas avermelhadas na pele, que podem ser acompanhadas por sintomas como febre e mal-estar) na gravidez, com ou sem microcefalia, além de crianças nascidas neste mesmo período, mas as mães não apresentaram exantema na atual gravidez. Este período de estudo foi escolhido com o objetivo de abranger participantes nascidos no período de maior circulação do vírus Zika, por se tratar de um período chuvoso. O acompanhamento dessas crianças participantes foi feito durante 12 meses, e, por esse motivo, o período de estudo foi até junho de 2018.

As crianças foram recrutadas via unidades de saúde (clínica da família, postos de saúde, maternidade, entre outros) ou por buscas espontâneas e a área de cobertura contemplou sete municípios: Niterói, São Gonçalo, Itaboraí, Maricá, Tanguá, Rio Bonito e Silva Jardim.

No total, 115 crianças foram acompanhadas e fizeram parte do estudo. Vale ressaltar que nem todas as crianças compareceram às 12 consultas estipuladas, isto é, cada criança compareceu a um número diferente de consultas que não estavam igualmente espaçadas no tempo.

2.1.2 Local de estudo

A equipe de pesquisadores foi responsável por todo o seguimento clínico-laboratorial dos pacientes e o acompanhamento destes foi feito periodicamente na Unidade de Pesquisa Clínica do Hospital Universitário Antônio Pedro da Universidade Federal Fluminense (UPC/HUAP-UFF).

2.1.3 Critérios de inclusão e exclusão

Os critérios de inclusão no estudo foram:

- *Grupo de expostos ao ZIKV (PCR+)*: formado por crianças cujas mães apresentaram exantema durante a gravidez, com diagnóstico positivo de infecção pelo ZIKV por RT-PCR (coleta feita dentro do período de exantema da gestante);
- *Grupo de não expostos ao ZIKV (PCR-)*: formado por crianças cujas mães apresentaram exantema na gravidez, porém o diagnóstico de infecção pelo ZIKV por RT-PCR foi negativo (coleta feita dentro do período de exantema da gestante);
- *Grupo Zika suspeito (sem PCR)*: composto por crianças cujas mães apresentaram exantema na gravidez, porém não há relato de coleta de RT-PCR ou a coleta foi realizada em período não adequado;
- *Grupo microcefalia sem exantema materno (SZC)*: composto por crianças com microcefalia cujas mães não apresentaram exantema na presente gravidez.

Já os critérios de exclusão utilizados foram:

- infecção confirmada por toxoplasmose, sífilis, herpes simples, citomegalovírus ou rubéola e/ou
- episódio de hipóxia (falta de oxigênio) no período perinatal e/ou
- microcefalia afastada durante o período de 12 meses de seguimento clínico da criança e/ou
- microcefalia proveniente de outra etiologia além do ZIKV.

2.1.4 Coleta dos dados

Durante o período de execução do projeto, todos os participantes que atenderam os critérios citados anteriormente e que foram atendidos no UPC/HUAP-UFF foram chamados para participar do estudo.

Os dados relativos ao perímetro cefálico foram levantados a partir dos registros da maternidade, do cartão da criança e das medidas coletadas pela equipe do estudo durante a visita clínica na UPC/HUAP-UFF.

A equipe de estudo realizou as medições do perímetro cefálico durante as visitas clínicas na UPC/HUAP-UFF, com o auxílio de fita métrica não extensível, no ponto mais proeminente da parte posterior do crânio e sobre as sobrancelhas, como mostra a Figura abaixo.



Ilustração: Filipe Alencar – NUCOM/GAB/SAS.

Figura 1: Medição do perímetro cefálico

2.2 Metodologia

Nesta seção serão apresentados os recursos utilizados para resolver o problema médico descrito até então. De maneira geral, tem-se um banco de dados formado por n observações ao longo de 12 meses de m indivíduos e o objetivo é utilizar uma ferramenta estatística que quantifique o efeito de fatores sobre a variável resposta. Por este motivo, em um primeiro momento, poderia se pensar em utilizar modelos de **regressão linear**

múltipla, ferramenta estatística que investiga a relação entre duas ou mais variáveis.

É chamado de modelo de regressão linear múltipla aquele que contém mais de uma variável independente e deseja-se determinar uma relação entre essas (p-1) variáveis independentes e a variável resposta Y, da forma:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (2.1)$$

onde $X_{i,k}$ equivale ao valor da k-ésima variável independente ($k = 1, 2, \dots, p - 1$) correspondente a i-ésima observação ($i = 1, \dots, m$), Y_i é o valor observado da variável resposta Y no i-ésimo elemento, $\beta_0, \beta_1, \dots, \beta_{p-1}$ são os parâmetros desconhecidos que deseja-se estimar e ε_i é o erro aleatório associado a i-ésima observação.

Os pressupostos do modelo de regressão linear são:

1. $E(\varepsilon_i) = 0 \quad \forall \quad i = 1, 2, \dots, m$
2. $\text{VAR}(\varepsilon_i) = \sigma^2 \quad \forall \quad i = 1, 2, \dots, m$
3. $\text{COV}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall \quad i \neq j = 1, 2, \dots, m$

Assim, pode-se assumir que $\varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, m$.

Visto que um dos pressupostos para a utilização dessa ferramenta é a independência dos dados, este tipo de modelo não pode ser utilizado neste estudo, cujos dados são de natureza longitudinal, observados ao longo do tempo e, portanto, naturalmente correlacionados para cada criança.

A seguir apresentamos a metodologia que será utilizada neste trabalho.

2.2.1 Estrutura de Variância e Covariância

Dados longitudinais são medidas feitas em sequência temporal em uma mesma unidade ou indivíduo (GUIMARÃES; HIRAKATA, 2012). Este é um estudo de coorte aberta cujos dados são desbalanceados, isto é, as medidas de cada indivíduo são observadas em momentos diferentes e não tem o mesmo tempo total de observação. Além disso, os dados longitudinais têm natureza hierárquica, em que as n medidas observadas de um mesmo indivíduo são dependentes com erro correlacionados e as observação entre os indivíduos são independentes. Assim, a matriz V_0 de variância-covariância para as observações de cada criança é definida por:

$$V_0 = \begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,n} \\ V_{2,1} & V_{2,2} & \cdots & V_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n,1} & V_{n,2} & \cdots & V_{n,n} \end{bmatrix}$$

Logo, a matriz de variância-covariância geral, supondo que esta possui a mesma estrutura para os m indivíduos, é definida por:

$$V = \begin{bmatrix} V_0 & 0 & \cdots & 0 \\ 0 & V_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_0 \end{bmatrix}_{mn \times mn}$$

Desta maneira, o modelo linear para dados longitudinais é dado por:

$$Y_{i,j} = \beta_1 X_{1,i,j} + \beta_2 X_{2,i,j} + \cdots + \beta_p X_{p,i,j} + \epsilon_{i,j} \quad (2.2)$$

$$\epsilon_{i,j} \sim N\left(0, \{V_0\}_{j,j}\right) \quad (2.3)$$

$$E[\epsilon_{i,j} \times \epsilon_{i,k}] = \{V_0\}_{j,k} \quad (2.4)$$

com $i = 1, \dots, m$ e $k = 1, \dots, n$.

E na forma matricial, temos que:

$$Y_{nm \times 1} \sim n(X_{nm \times p} \beta_{p \times 1}, V_{nm \times nm})$$

Já o estimador de mínimos quadrados ponderados é dado por:

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y \quad (2.5)$$

Como visto na equação 2.5, para estimar o parâmetro $\hat{\beta}$ por mínimos quadrados ponderados, é preciso que a matriz V seja conhecida. Para isso, pode-se assumir que, a medida que aumenta o tempo entre uma observação e outra de um mesmo indivíduo, a correlação entre elas diminui. Então, deriva-se um estimador indireto da matriz V de covariância, levando em consideração algumas condições necessárias.

A simetria é a primeira condição suficiente e necessária

$$V_{i,j} = V_{j,i}$$

Além disso, também precisamos gerar matrizes definidas não-negativas

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j V_{i,j} \geq 0$$

$$\forall n, \alpha_1, \dots, \alpha_n, \vec{s}_1, \dots, \vec{s}_n.$$

O modelo exponencial sugerido por (FERNANDES; SCHMIDT; MIGON, 2009) será utilizado neste trabalho, pois neste modelo considera-se que a medida que o tempo entre as observações de um mesmo indivíduo aumenta, a correlação entre elas diminui. Assim, o modelo exponencial é dado por:

$$\text{Cov}(h) = \sigma^2 e^{-\frac{h}{r}} \quad (2.6)$$

Onde,

σ^2 é a variância;

h é o tempo entre as observações;

$\frac{1}{r}$ é o tempo máximo de associação entre observações de um mesmo indivíduo.

Vale ressaltar que há outros modelos que podem ser utilizados, como o modelo de correlação uniforme, que, diferentemente do modelo de correlação exponencial, assume que a correlação entre duas medições de um mesmo indivíduo sempre será positiva. Para mais informações, veja o livro *Analysis of longitudinal data* (DIGGLE et al., 2002).

2.2.2 Enfoque Bayesiano

Por ser um método eficaz para lidar com modelos mais complexos de maneira mais simples, será utilizada uma abordagem Bayesiana para a estimação dos parâmetros desconhecidos. Diferentemente da abordagem clássica que assume que os parâmetros desconhecidos são quantidades fixas, neste tipo de enfoque distribuições de probabilidade são utilizadas para resumir a informação existente sobre a quantidade de interesse desconhecida e geralmente não observável θ .

2.2.3 Teorema de Bayes

É possível resumir toda a informação disponível sobre a quantidade de interesse θ desconhecida e geralmente não observável através da distribuição a priori $p(\theta)$. Utilizando a quantidade aleatória X , este conhecimento pode ser ampliado, considerando que $p(x|\theta)$ define a relação entre X e θ . A ideia é que, ao observar $X = x$, o montante de informação sobre o parâmetro θ aumenta e, para isso, utiliza-se o Teorema de Bayes.

$$p(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta} \quad (2.7)$$

Note que $\frac{1}{p(x)}$ não depende de θ , portanto atua como uma constante que tem finalidade de normalizar $p(\theta|x)$. A função de verossimilhança $L(\theta; x) = p(x|\theta)$, que aplicada a cada um dos valores de θ indica quais são mais plausíveis, e $p(\theta)$, chamada de distribuição a priori de θ , são combinadas, gerando a distribuição a *posteriori* de teta, $p(\theta | x)$. Por isso, a forma usual de representar a equação 2.7 é:

$$p(\theta | x) \propto L(\theta; x)p(\theta) \quad (2.8)$$

Isto é, a distribuição a *posteriori* é proporcional a verossimilhança vezes a distribuição a priori.

2.2.4 Distribuição a Priori

A distribuição a priori $p(\theta)$ deve retratar todo o conhecimento existente sobre o parâmetro θ . Caso haja conhecimento prévio de experimentos anteriores ou de algum especialista sobre o parâmetro desconhecido, definir prioris pode se tornar uma tarefa mais simples.

Porém, na maioria dos casos, há pouco ou nenhum conhecimento prévio sobre o parâmetro. Dessa forma, uma maneira de determinar a distribuição a priori de forma subjetiva é atribuindo a ela uma variância grande o suficiente de forma a torná-la vaga e pouco informativa.

2.2.5 Monte Carlo via Cadeias de Markov

Em problemas complexos, fazer contas analíticas para encontrar *posteriors* pode ser inviável. Nesses casos, é possível recorrer aos métodos de Monte Carlo via Cadeias de Markov (MCMC) no lugar de métodos não iterativos. Através deste método pode-se obter amostras da distribuição a *posteriori* de θ com o objetivo de calcular as estimativas amostrais da distribuição de interesse, possibilitando, assim, fazer inferências sobre o vetor de parâmetros desconhecidos. Para isso, basta que o núcleo da expressão da função de densidade ou de probabilidade dessa distribuição seja conhecido. Mais detalhes podem ser encontrados em (ROBERT; CASELLA, 2013). Nesta seção, serão abordados os métodos de MCMC mais utilizados e que também serão aplicados neste trabalho, o amostrador de Gibbs e o algoritmo de Metropolis-Hastings. Uma discussão mais ampla pode ser vista em (GAMERMAN; LOPES, 2006).

2.2.5.1 Cadeias de Markov

Uma cadeia de Markov é um tipo de processo estocástico, de forma que as distribuições condicionais de estados futuros, X_{n+j} para $j \geq 1$ dado X_1, \dots, X_n , dependa somente de X_n e não de estados passados X_1, \dots, X_{n-1} , isto é:

$$P(X_{n+1} | X_0, X_1, \dots, X_n) = P(X_{n+1} | X_n) \quad (2.9)$$

De acordo com (GAMERMAN; LOPES, 2006), para que os métodos de MCMC possam ser utilizados, as cadeias devem ser:

- Homogêneas, isto é, as probabilidades de transição são estacionárias, portanto são constantes em relação ao tempo;
- Irredutível, isto é, cada estado pode ser alcançado a partir de qualquer outro estado;
- Aperiódica, não há estados absorventes.

Na prática, valores iniciais são gerados e, conforme o número de iterações aumenta, esses valores iniciais são esquecidos (período conhecido como aquecimento) até que ocorra a convergência para a distribuição de equilíbrio da cadeia de Markov.

2.2.5.2 Amostrador de Gibbs

O Amostrador de Gibbs é uma cadeia de Markov que não possui um mecanismo de aceitação-rejeição, isto é, a cadeia sempre irá se movimentar para um novo valor. Esse algoritmo irá gerar, a partir das distribuições condicionais completas, uma amostra da distribuição conjunta de interesse. Considere que $\pi(\theta)$ seja a distribuição que se deseja amostrar, onde θ é um vetor ou matriz que está particionado em d componentes $\theta = (\theta_1, \dots, \theta_d)$. As transições de um estado para outro ocorrem através das distribuições condicionais completas a *posteriori* $\pi(\theta_i | \theta_{-i})$, para $i = 1, \dots, d$ e $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$. O amostrador de gibbs opera da seguinte forma:

1. iniciar o contador de iterações de cadeia em $t = 0$;
2. definir valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$;
3. a partir de $\theta^{(t-1)}$, obter um novo valor de $\theta^{(t)}$, através da geração sucessiva de valores;

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

4. incrementar em uma unidade o contador t e voltar para o passo 2 até a obtenção da convergência.

A convergência da cadeia ocorre após o período de aquecimento e é facilmente identificada analisando o gráfico da cadeia. Os valores resultantes formam a amostra de $\pi(\theta)$.

2.2.5.3 Algoritmo de Metropolis-Hastings

Este mecanismo utiliza a ideia dos métodos de rejeição em que, a partir de uma distribuição auxiliar, um valor é gerado e aceito com certa probabilidade. Os algoritmos de Metropolis-Hastings asseguram a convergência da cadeia para uma determinada distribuição de equilíbrio, que, sob enfoque bayesiano, pode ser a distribuição a *posteriori*.

Considere uma cadeia no estado θ e um novo valor θ' gerado por uma distribuição proposta $q(\cdot | \theta)$. Repare que esta distribuição proposta pode depender do atual estado

da cadeia. A probabilidade de aceitação do valor gerado θ' é:

$$\alpha(\theta, \theta') = \min \left(1, \frac{\pi(\theta') q(\theta, \theta')}{\pi(\theta) q(\theta', \theta)} \right) \quad (2.10)$$

em que π é a distribuição de interesse.

Um aspecto importante para aplicações bayesianas, onde não conhecemos totalmente a distribuição a *posteriori*, é que só é necessário conhecer π parcialmente, isto é, a menos de uma constante, já que a probabilidade 2.10 é inalterável.

O algoritmo de Metropolis-Hastings, na prática, opera da seguinte forma:

1. iniciar o contador de iterações em $t = 0$ e determinar um valor inicial $\theta^{(0)}$;
2. gerar um novo valor θ' da distribuição proposta $q(\cdot | \theta)$;
3. calcular a probabilidade de aceitação 2.10 e gerar $u \sim U(0, 1)$;
4. aceitar o novo valor se $u \leq \alpha$ e fazer $\theta^{(t+1)} = \theta'$. Caso contrário, rejeitar e fazer $\theta^{(t+1)} = \theta$;
5. incrementar o contador t em uma unidade e voltar para o passo 2, até que a convergência seja obtida.

A escolha da distribuição proposta $q(\cdot | \theta)$, ainda que possa ser feita de forma arbitrária, é fundamental para o funcionamento do algoritmo. Em aplicações bayesianas, geralmente as *prioris* são utilizadas como distribuição proposta e a *posteriori* é a distribuição de interesse, isto é, $\pi = p(\theta | x)$ e a probabilidade de aceitação assume a seguinte forma:

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{p(x | \theta') p(\theta') q(\theta | \theta')}{p(x | \theta) p(\theta) q(\theta' | \theta)} \right\} \quad (2.11)$$

2.2.6 Critério de seleção do modelo

Há várias maneiras de avaliar a qualidade do ajuste de um modelo de regressão. Um deles é o Critério de Informação de Akaike (AIC), definido por:

$$AIC(p') = n \ln(SQRes_p) + 2p' - n \ln(n) \quad (2.12)$$

onde:

- p' é o número de parâmetros do modelo;
- n é o número de observações;
- $SQRes_p = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ é a soma dos quadrados dos resíduos.

Conforme o $SQRes$ aumenta, o AIC também aumenta. Portanto, quanto menor o AIC, melhor é o ajuste do modelo, pois espera-se que a soma dos resíduos tenha baixo valor. Este critério é o mais utilizado, embora que priorize modelos com mais variáveis explicativas.

Um critério alternativo ao AIC é Critério de Informação Bayesiana (BIC), definido por:

$$BIC(p') = n \ln(SQRes_p) + [\ln(n)]p' - n \ln(n) \quad (2.13)$$

Pode ser notado que, na segunda parcela da equação 2.13 o 2 que multiplicava o p' foi substituído pelo $\ln(n)$, de modo que os modelos que possuam maior número de variáveis explicativas sejam penalizados de forma mais severa que no AIC. No entanto, assim como no AIC, quanto menor o BIC, melhor a qualidade do modelo.

2.2.7 Variáveis incluídas no modelo

Serão consideradas na modelagem estatística deste estudo as seguinte variáveis:

- Perímetro cefálico;
- Comprimento
- Peso
- Idade na consulta
- Sexo (masculino e feminino)
- Grupo (expostos ao ZIKV, não expostos ao ZIKV, Zika suspeito, microcefalia sem exantema materno)

em que perímetro cefálico é a variável de interesse e o restante das variáveis são explicativas.

As variáveis explicativas peso, comprimento, e idade na consulta são longitudinais e foram medidas ao longo dos doze meses de acompanhamento da criança, em até doze consultas. Em cada um desses exames pós natais realizados, todas as medidas citadas anteriormente foram aferidas e registradas. Entretanto, vale ressaltar que nem todas as crianças participantes compareceram às doze consultas de avaliação.

Já as variáveis sexo e grupo, independem do tempo, isto é, são iguais em todas as medições. Além disso, vale ressaltar que como essas variáveis são qualitativas, para adicioná-las ao modelo, utilizam-se as variáveis indicadoras, também conhecidas com variáveis dummy.

Por exemplo, para sabermos o efeito que a variável qualitativa com k categorias tem sobre a variável resposta, temos que definir $k - 1$ variáveis indicadoras. No caso das variáveis dicotômicas, com apenas dois níveis, usualmente define-se o número 1 quando a característica em questão está presente e 0 caso contrário.

No exemplo abaixo, a variável sexo, quando esta representa o sexo biológico, apresenta apenas duas categorias, masculino e feminino. Dessa forma, foi criada apenas uma variável indicadora, definida por:

$$X_{i1} = \begin{cases} 1, & \text{se o } i\text{-ésimo indivíduo é do sexo masculino} \\ 0, & \text{se o indivíduo é do sexo feminino} \end{cases}$$

Já para a variável grupo, que apresenta quatro categorias, foram criadas 3 variáveis dummy: expostos ao Zika (PCR+), Zika suspeito (sPCR) e microcefalia sem exantema materno (SZC). Caso o indivíduo fizesse parte de um desses grupos, definiu-se o número 1 para o grupo em questão e 0 para o restante dos grupos. Caso o indivíduo não pertencesse a nenhum dos 3 grupos citados, definiu-se 0 para todos e automaticamente, designa-se que o indivíduo pertence ao grupo de não expostos ao Zika (PCR-).

3 Resultados e Discussão

A base de dados inicial era composta por 115 crianças que foram acompanhadas no estudo, como foi explicado em 2.1.1. Entretanto, diante do fato que as crianças possuem números diferentes de consultas ao longo dos 12 meses, neste trabalho foi utilizado o modelo composto por 62 crianças que compareceram a pelo menos 5 consultas. Nos casos em que as crianças possuíam mais de 5 observações, as consultas foram selecionadas da maneira mais equidistante possível, como mostra o esquema no Anexo 1.

Com o objetivo de identificar quais variáveis estão relacionadas a evolução do perímetro cefálico das crianças acompanhadas, foram utilizados modelos de efeitos mistos. Este tipo de modelagem foi escolhida pois algumas variáveis independem do tempo, e outras, de natureza longitudinal, variam com o tempo, caracterizando, assim, uma estrutura hierárquica.

Devido a complexidade do modelo, será utilizado enfoque Bayesiano para estimar os parâmetros desconhecidos. Assim, o modelo Bayesiano proposto foi:

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,\dots}) \\
 \mu_{i,j} &= \mu_i + \beta_{1,1}X_{1,i,j} + \beta_{1,2}X_{2,i,j} + \dots + \beta_{1,p_1-1}X_{p_1-1,i,j} \\
 \mu_i &= \beta_0 + \beta_{2,1}X_{1,i} + \beta_{2,2}X_{2,i} + \dots + \beta_{2,p_2-1}X_{p_2-1,i} \\
 \{V_0\}_{j,k} &= \frac{1}{\phi} \exp\{-\tau h_{j,k}\}
 \end{aligned}$$

onde:

- $Y_{i,j}$ é o perímetro cefálico do indivíduo i no tempo j ; i varia de 1 a 62 pacientes, j varia de 1 a 5 consultas
- $\beta_{1,1}X_{1,i,j} + \dots + \beta_{1,p_1-1}X_{p_1-1,i,j}$ é o efeito longitudinal do modelo do indivíduo i no tempo j ;
- $\beta_{2,1}X_{1,i} + \dots + \beta_{2,p_2-1}X_{p_2-1,i}$ é a soma dos efeitos das variáveis que não têm influência longitudinal do indivíduo i ;

- $V_{0,\dots}$ é a matriz de variância-covariância de cada criança i ;
- τ é o parâmetro que indica a distância máxima de associação entre os dados definida pelos meses entre as consultas, e é igual ao inverso da amplitude r ;
- ϕ é a precisão dos dados e é equivalente ao inverso da variância σ^2 ;
- h é a distância calculada através das diferenças das idades das crianças em cada consulta

Vale ressaltar que a covariância entre os $Y_{i,j}$ e $Y_{i,k}$ são definidos pela distância h entre as consultas. Portanto, como foi discutido em 2.2.1, quanto maior for a distância entre as consultas, menor é a covariância entre elas.

- Definição das prioris dos parâmetros

As distribuições a priori dos parâmetros β , τ e ϕ foram definidas da seguinte forma:

$$\begin{aligned}\beta_{1,k_1} &\sim N(0, 1000), k_1 = 1, \dots, p_1 - 1 \\ \beta_{2,k_2} &\sim N(0, 1000), k_2 = 1, \dots, p_2 - 1 \\ \phi &\sim \text{Gama}(0.44, 0.67) \\ \tau &\sim \text{Gama}(2.2, 10)\end{aligned}$$

A priori, foram definidas distribuições não informativas para os parâmetros β , com média zero e precisão $\phi = 0,001$. Como ϕ é o inverso da variância, pode-se dizer que, a priori, existe uma grande incerteza sobre os parâmetros.

Já para definir as prioris de ϕ e τ , pela dificuldade em modelar esses parâmetros da matriz de variância-covariância, foi utilizada a metodologia aplicada em (FERNANDES; SCHMIDT; MIGON, 2009), em que os parâmetros das distribuições Gama são pouco informativos, escolhidos de forma ótima. Dessa forma, o parâmetro ϕ tem média 0,66 e variância ampla, e o parâmetro τ tem média 0,22 e a variabilidade também é ampla.

Para estimar os parâmetros desconhecidos, é necessário definir primeiramente a verossimilhança:

$$L(\mu; \phi; \tau | \mathbf{y}) = \prod_{i=1}^m \prod_{j=1}^n \frac{1}{\sqrt{2\pi}} \left(\frac{\phi}{e^{-\tau h}} \right)^{\frac{1}{2}} \exp \left[\frac{1}{2} \frac{\phi}{e^{-\tau h}} \sum_{i=1}^m \sum_{j=1}^n - (y_{i,j} - \mu_{i,j})^2 \right] \quad (3.1)$$

$$= (2\pi)^{\frac{-mn}{2}} \left(\frac{\phi}{e^{-\tau h}} \right)^{\frac{mn}{2}} \exp \left\{ \frac{1}{2} \frac{\phi}{e^{-\tau h}} \sum_{i=1}^m \sum_{j=1}^n - [y_{i,j} - (\mu_i + \beta_{1,1}X_{1,i,j} + \dots + \beta_{1,p_1-1}X_{p_1-1,i,j})]^2 \right\} \quad (3.2)$$

Fazendo as substituições em (3.2) tem-se que $L(\mu; \phi; \tau | \mathbf{y})$ é igual a:

$$= (2\pi)^{\frac{-mn}{2}} \left(\frac{\phi}{e^{-\tau h}} \right)^{\frac{mn}{2}} \exp \left\{ \frac{1}{2} \frac{\phi}{e^{-\tau h}} \sum_{i=1}^m \sum_{j=1}^n - [y_{i,j} - (\beta_0 + \beta_{2,1}X_{1,i} + \dots + \beta_{2,1}X_{1,i} + \beta_{2,p_2-1}X_{p_2-1,i} + \beta_{1,1}X_{1,i,j} + \dots + \beta_{1,p_1-1}X_{p_1-1,i,j})]^2 \right\} \quad (3.3)$$

Combinando a verossimilhança acima com a priori, tem-se a seguinte *posteriori*:

$$p(\mu; \phi; \tau | \mathbf{y}) \propto L(\mu; \phi; \tau | \mathbf{y}) p(\beta_{1,1}) \dots p(\beta_{1,p_1-1}) p(\beta_{2,1}) \dots p(\beta_{2,p_2-1}) p(\phi) p(\tau)$$

Como pode ser notado acima, fazer o uso contas analíticas para encontrar a *posteriori* pode não ser eficiente. Por isso, para a obtenção de amostras das *posteriors* dos parâmetros, foi utilizado o software OpenBugs, que permite obter essas amostras através do algoritmo de Amostrador de Gibbs de forma automática.

No software OpenBugs, foram simuladas 10.000 rodadas de aquecimento, conhecido como *burn-in*. Posteriormente, mais 10.000 amostras foram simuladas para a obtenção das distribuições a *posteriori* dos parâmetros de interesse. O script do software OpenBugs se encontra no Anexo 2.

3.1 Análise do Modelo

O modelo utilizado contou com 62 indivíduos com pelo menos 5 consultas, dos quais 40 (64,52%) são do sexo masculino e 22 (35,48%) do sexo feminino. Além disso, com relação a variável grupo, definida em 2.1.3, 18 crianças (29,03%) eram do grupo de expostos ao ZIKV, 10 (16,13%) pertenciam ao grupo de não expostos ao ZIKV, 26 (41,93%) do grupo de zika suspeito e 8 (12,90%) do grupo microcefalia sem exantema materno. As Tabelas 1 à 5 mostram algumas estatísticas descritivas sobre as características das crianças em

cada consulta.

Tabela 1: Estatísticas Descritivas das características das crianças na consulta 1

Variáveis	Consulta 1		
	Média (desvio padrão)	Min.	Max
Idade	1,45 (1,11)	0,26	6,03
Perímetro cefálico	35,24 (3,76)	28,5	45,0
Peso	4,29 (1,13)	2,66	8,50
Comprimento	53,36 (4,70)	46,0	70,0

Tabela 2: Estatísticas Descritivas das características das crianças na consulta 2

Variáveis	Consulta 2		
	Média (desvio padrão)	Min.	Max
Idade	3,86 (1,41)	1,61	7,61
Perímetro cefálico	38,61 (3,79)	30,2	46,0
Peso	6,29 (1,25)	4,2	9,40
Comprimento	60,99 (4,20)	52,0	74,0

Tabela 3: Estatísticas Descritivas das características das crianças na consulta 3

Variáveis	Consulta 3		
	Média (desvio padrão)	Min.	Max
Idade	6,34 (1,36)	3,06	9,48
Perímetro cefálico	40,35 (3,97)	30,5	46,0
Peso	7,56 (1,43)	5,03	12,21
Comprimento	66,15 (3,96)	57,0	75,0

Tabela 4: Estatísticas Descritivas das características das crianças na consulta 4

Variáveis	Consulta 4		
	Média (desvio padrão)	Min.	Max
Idade	8,90 (1,35)	4,42	11,06
Perímetro cefálico	41,83 (4,36)	30,8	47,5
Peso	8,46 (1,47)	5,63	12,85
Comprimento	69,75 (3,88)	62,0	77,0

Tabela 5: Estatísticas Descritivas das características das crianças na consulta 5

Variáveis	Consulta 5		
	Média (desvio padrão)	Min.	Max
Idade	11,39 (1,07)	6,0	12,48
Perímetro cefálico	42,83 (4,43)	31,2	48,5
Peso	9,08 (1,59)	6,17	13,10
Comprimento	73,17 (3,83)	64,0	84,0

Para ilustrar como estão dispostos os dados longitudinais da variável resposta, perímetro cefálico, das 62 crianças incluídas no modelo, ao longo das 5 consultas, durante os 12 meses de acompanhamento, foi gerado o gráfico abaixo.

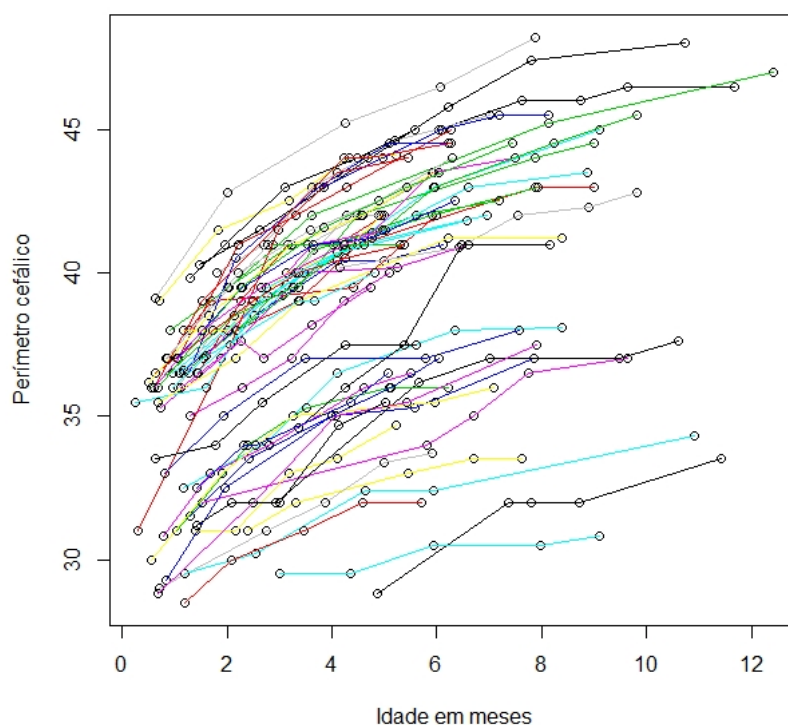


Figura 2: Dados longitudinais do perímetro cefálico das 62 crianças no primeiro ano de vida.

Para testar quais variáveis explicativas seriam significativas, primeiramente elas foram testadas uma a uma separadamente, formando cinco modelos individuais (apresentados no Anexo 3). Na Tabela 6 se encontram as estimativas geradas.

Tabela 6: Estimativas dos modelos individuais

Modelo	Variável regressora	Média	Desvio Padrão	Intervalo de Credibilidade
Modelo I_1	Comprimento	0,3747	0,35	[0,35;0,3992]
Modelo I_2	Idade	0,7603	0,03483	[0,6917;0,8281]
Modelo I_3	Sexo	0,541	1,284	[-1,915;3,114]
Modelo I_4	Peso	1,443	0,0486	[1,35;1,539]
Modelo I_5 *	PCR+	-2,548	1,743	[-5,927;0,9227]
	sem PCR	-3,867	1,666	[-7,161;-0,5615]
	SZC	-8,358	2,086	[-12,41;-4,192]

*O Modelo I_5 é composto pela variável qualitativa grupo que possui 4 categorias e, conseqüentemente, 3 variáveis dummy (PCR+, sem PCR, SZC) que foram testadas juntas.

De acordo com a Tabela 6, pode-se observar que os intervalos de credibilidade dos parâmetros β_{sexo} e β_{PCR+} contém o valor 0. Portanto, conclui-se que esses parâmetros não possuem valor significativo e podem ser removidos do modelo completo.

Antes de verificar o ajuste, estimar os parâmetros do modelo com base no método de MCMC exige que haja convergência nas cadeias geradas. Por esse motivo, a fim de mostrar o cumprimento dessa exigência, foram gerados gráficos das iterações das cadeias e das distribuições dos parâmetros do modelo completo (já tendo sido feita a exclusão das variáveis sexo e PCR+) como mostram as Figuras 3 e 4.

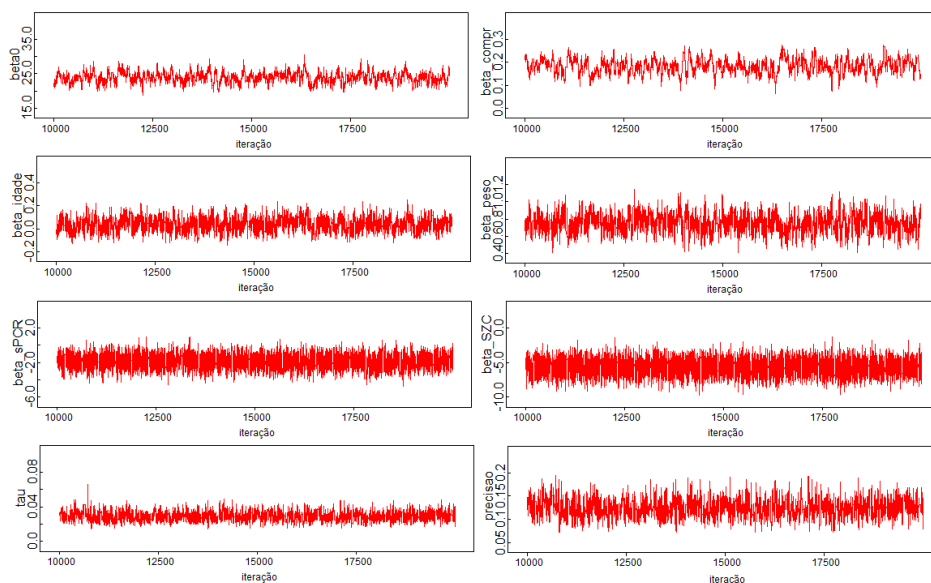


Figura 3: Monitoração da Convergência dos Parâmetros do Modelo completo

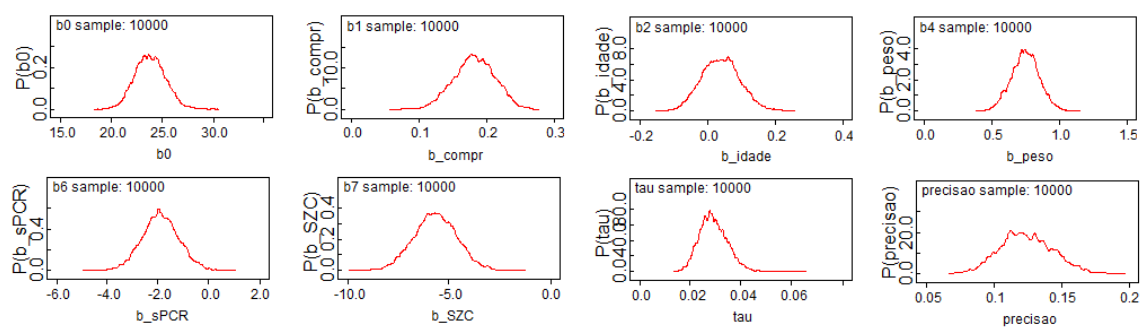


Figura 4: Densidades a Posteriori dos Parâmetros do Modelo completo

Através da Figura 3, é possível notar indícios de que os parâmetros atingiram convergência, visto que, no geral, as cadeias estão bem comportadas, contidas dentro de um intervalo. Também é possível observar por meio da Figura 4 que as *posteriori* têm formato semelhante aos das *prioris*.

O resultado dos gráficos de autocorrelação, que indica como o valor de cada iteração da cadeia influencia as estimativas vizinhas, será apresentado na Figura 5:

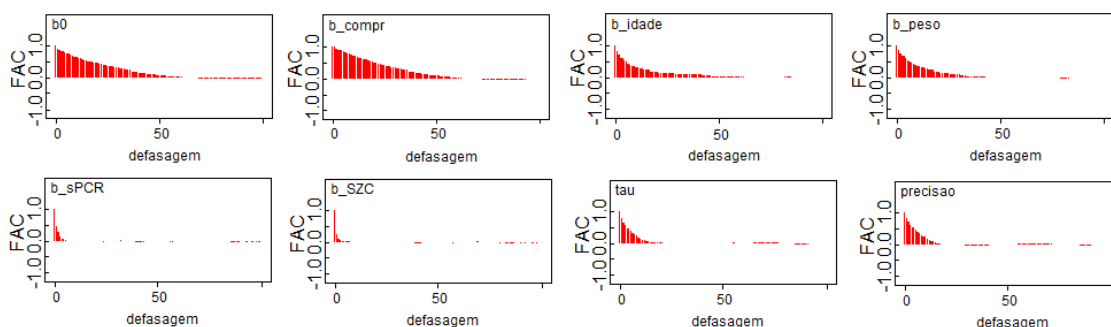


Figura 5: Autocorrelações das cadeias geradas

É possível notar que apenas as primeiras defasagens das cadeias das estimativas dos parâmetros se apresentaram de forma autocorrelacionada e, a partir dessas defasagens, os gráfico de autocorrelação se apresentaram de forma desejável. Por este motivo, usamos uma defasagem de não menos de 50 para coletar as amostras da *posteriori*.

Finalmente, depois de conferir as suposições, os resultados das estatísticas ligadas a cada parâmetro do modelo completo estão dipostos na Tabela 7.

Tabela 7: Estimativas dos parâmetros do Modelo Completo

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade
β_0	23,86	1,5	[21,06;26,83]
$\beta_{comprimento}$	0,1814	0,03104	[0,1196;0,2395]
β_{idade}	0,03907	0,05672	[-0,06858;0,1529]
β_{peso}	0,7484	0,1043	[0,5443;0,9546]
β_{sPCR}	-1,908	0,7391	[-3,346;-0,4169]
β_{SZC}	-5,679	1,084	[-7,812;-3,56]
τ	0,02912	0,00553	[0,01944;0,0408]
ϕ	0,1245	0,02002	[0,08776;0,1654]

O parâmetro β_{idade} , quando testado separadamente, possuía valor significativo. No entanto, pelos resultados mostrados na Tabela 7, é possível visualizar que, quando testado em conjunto no novo modelo proposto, contém o valor 0 em seu intervalo, indício de que a variável idade não é significativa para o modelo completo.

Já a variável explicativa comprimento não contém zero em seu intervalo de credibilidade, o que indica associação significativa. De acordo com o que foi mostrado na Tabela 7, tem-se que ao aumentar em 1 centímetro o comprimento da criança, o perímetro cefálico

aumenta em 0,1814.

Em relação ao peso, nota-se que essa variável explicativa também se apresentou de maneira significativa e, para cada um quilo acrescido no peso, o perímetro cefálico aumentaria em 0,7484.

Para as variáveis indicadoras grupo zika suspeito e microcefalia sem exantema materno, que também se mostraram ser estatisticamente significantes, tem-se que o perímetro cefálico das crianças ZIKV suspeito é menor que das crianças que são do grupo de não expostos ao ZIKV, e essa diferença foi de 1,908 centímetros. Já as crianças do grupo microcefalia sem exantema materno também possuem perímetro cefálico menor em relação as crianças não expostas ao ZIKV, e essa diferença foi de 5,679 centímetros.

Por fim, analisando as estimativas dos parâmetros τ e ϕ , nota-se que ambos não possuem o valor zero contido em seu intervalo de credibilidade, logo, são significativos para o Modelo completo. Além disso, a precisão ϕ apresentou um valor de 0,1245, que corresponde a uma variância de $\frac{1}{\phi} = \sigma^2 = 8,0321$.

Com o objetivo de avaliar a qualidade do ajuste dos modelos, foram calculados o AIC e o BIC para cada um dos modelos com as variáveis incluídas no modelo completo e para o modelo completo (já feita a exclusão da variável idade), como está apresentado na Tabela 8:

Tabela 8: Avaliação da qualidade do modelo completo e dos modelos com apenas uma variável explicativa

Modelo	Variáveis Regressoras	AIC	BIC
Modelo 1 (completo)	Comprimento, peso, sPCR e SZC	868,2	885,9
Modelo 2	Comprimento	971,3	975,7
Modelo 3	Peso	989,3	993,8
Modelo 4	sPCR	1226,0	1230,0
Modelo 5	SZC	1182,0	1186,0

*sPCR e SZC são os grupos Zika suspeito e microcefalia sem exantema materno, respectivamente

É possível observar na Tabela 8 que, os valores de AIC e BIC foram menores para o Modelo 1 e, portanto, pode-se concluir que o modelo completo é o mais adequado, pois possui o melhor ajuste (o modelo completo se encontra no Anexo 4). Vale ressaltar que, como visto na Tabela 8, no Modelo 1 (completo), as variáveis comprimento, peso, zika suspeito e microcefalia sem exantema materno apresentam influência sobre perímetro cefálico das crianças.

4 Conclusão

De acordo com os resultados encontrados e apresentados neste estudo, pode-se concluir que o modelo bayesiano se mostrou consistente para a modelagem do perímetro cefálico de crianças expostas e não expostas ao vírus Zika. Através dos métodos de MCMC, temos que as cadeias estão bem comportadas e as amostras geradas são provenientes das distribuições a *posteriori*. Ademais, o modelo bayesiano proporcionou uma interpretação fácil para um modelo altamente complexo.

Inicialmente, quando foram testadas separadamente, as variáveis sexo e PCR positivo não se mostraram significativas para o modelo e, por isso, foram removidas. Posteriormente, utilizou-se o software OpenBugs para gerar as estimativas dos parâmetros do modelo, obtidas a partir de amostras das posteriores através do algoritmo do Amostrador de Gibbs. Deste ponto, foi possível perceber que a variável idade também não se apresentou de maneira significativa. Isso significa dizer que as variáveis sexo, idade e a indicadora grupo de expostos ao ZIKV não têm influência sobre o perímetro cefálico das crianças.

Além disso, viu-se que as variáveis comprimento, peso, zika suspeito e microcefalia sem exantema materno mostraram associações significantes com o perímetro cefálico. Vale ressaltar que, com relação a variável grupo, faz todo o sentido que as crianças cujas mães apresentaram exantema na gravidez, porém não fizeram coleta de PCR e as crianças do grupo com microcefalia embora que a mãe não tenha apresentado exantema tenham um perímetro cefálico menor em relação às crianças cujas mães apresentaram exantema mas tiveram resultado negativo para ZIKV, fato apontado pela análise dos resultados das estimativas dos parâmetros do modelo.

Por fim, foi possível testar o ajuste dos modelos (completo e com apenas uma variável explicativa), através dos critérios AIC e BIC de seleção do modelo. Concluiu-se que o modelo completo, com as variáveis explicativas peso, comprimento, zika suspeito e microcefalia sem exantema materno, foi considerado com maior poder de prever um novo conjunto de dados com a mesma estrutura observada, pois apresentou o menor AIC

e o menor BIC em comparação com os modelos simples.

Feitas essas considerações, pode-se dizer que vale a pena investir em modelos para dados longitudinais em estudos como este na área da saúde e em outras áreas, pois são capazes de refletir de forma eficiente e fidedigna uma situação altamente complexa através de uma interpretação mais simples.

Referências

- ALBUQUERQUE, M. d. F. P. M. d. et al. Epidemia de microcefalia e vírus zika: a construção do conhecimento em epidemiologia. *Cadernos de Saúde Pública*, SciELO Public Health, v. 34, p. e00069018, 2018.
- CARMO, E. H.; PENNA, G.; OLIVEIRA, W. K. d. Emergências de saúde pública: conceito, caracterização, preparação e resposta. Instituto de Estudos Avançados da Universidade de São Paulo, 2008.
- DEVAKUMAR, D. et al. Infectious causes of microcephaly: epidemiology, pathogenesis, diagnosis, and management. *The Lancet Infectious Diseases*, Elsevier, v. 18, n. 1, p. e1–e13, 2018.
- DICK, G.; KITCHEN, S.; HADDOW, A. Zika virus (i). isolations and serological specificity. *Transactions of the royal society of tropical medicine and hygiene*, Oxford University Press, v. 46, n. 5, p. 509–520, 1952.
- DIGGLE, P. et al. *Analysis of longitudinal data*. [S.l.]: Oxford University Press, 2002.
- DUFFY, M. R. et al. Zika virus outbreak on yap island, federated states of micronesia. *New England Journal of Medicine*, Mass Medical Soc, v. 360, n. 24, p. 2536–2543, 2009.
- FANTINATO, F. F. S. T. et al. Descrição dos primeiros casos de febre pelo vírus zika investigados em municípios da região nordeste do brasil, 2015. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 25, p. 683–690, 2016.
- FERNANDES, M. V.; SCHMIDT, A. M.; MIGON, H. S. Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, SAGE Publications Sage India: New Delhi, India, v. 9, n. 1, p. 3–25, 2009.
- GAMERMAN, D.; LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. [S.l.]: CRC Press, 2006.
- GARCIA, L. P. *Epidemia do vírus Zika e microcefalia no Brasil: emergência, evolução e enfrentamento*. [S.l.], 2018.
- GUIMARÃES, L. S. P.; HIRAKATA, V. N. Uso do modelo de equações de estimativas generalizadas na análise de dados longitudinais. *Revista HCPA. Porto Alegre. Vol. 32, n. 4 (2012), p. 503-511*, 2012.
- LUZ, K. G.; SANTOS, G. I. V. d.; VIEIRA, R. d. M. Febre pelo vírus zika. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 24, p. 785–788, 2015.
- MINISTÉRIO DA SAÚDE, SECRETARIA DE VIGILÂNCIA EM SAÚDE. *Síndrome Congênita associada a infecção pelo vírus Zika*. [S.l.], 2019.

- NUNES, M. L. et al. Microcephaly and zika virus: a clinical and epidemiological analysis of the current outbreak in brazil. *Jornal de Pediatria (Versão em Português)*, Elsevier, v. 92, n. 3, p. 230–240, 2016.
- RASMUSSEN, S. A. et al. Zika virus and birth defects—reviewing the evidence for causality. *New England Journal of Medicine*, Mass Medical Soc, v. 374, n. 20, p. 1981–1987, 2016.
- ROBERT, C.; CASELLA, G. *Monte Carlo statistical methods*. [S.l.]: Springer Science & Business Media, 2013.
- SAÚDE, S. d. V. e. S. Ministério da. *Vírus Zika no Brasil: a resposta do SUS*. [S.l.], 2017.
- SCHULER-FACCINI, L. et al. Possible association between zika virus infection and microcephaly—brazil, 2015. *Morbidity and Mortality Weekly Report*, JSTOR, v. 65, n. 3, p. 59–62, 2016.
- TOGNARELLI, J. et al. A report on the outbreak of zika virus on easter island, south pacific, 2014. *Archives of virology*, Springer, v. 161, n. 3, p. 665–668, 2016.
- ZANLUCA, C. et al. First report of autochthonous transmission of zika virus in brazil. *Memórias do Instituto Oswaldo Cruz*, SciELO Brasil, v. 110, n. 4, p. 569–572, 2015.

ANEXO 1 – Esquema de escolha das consultas

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Pront	Compr_1	Compr_2	Compr_3	Compr_4	Compr_5	Compr_6	Compr_7	Compr_8	Compr_9	Compr_10	Compr_11	Compr_12		
2	822919	70,0	73,0	75,0	77,0	78,0									
4	820919	57,0	62,0	64,0	71,0	71,0	71,5								
6	819886	51,5	57,5	62,0	68,0	75,0									
7	828619	51,0	56,0	60,5	62,5	65,0	68,0	69,0	73,5						
10	821211	52,0	57,5	61,5	64,0	66,5	72,0								
11	819894	52,0	58,0	62,5	65,0	69,0	71,0	76,0	76,0						
12	822038	54,0	59,0	60,5	64,0	66,0	69,0	70,5	76,5						
13	818424	58,5	65,0	67,0	69,0	74,0	76,5								
14	826415	56,0	62,0	69,0	71,0	75,0									
15	820512	52,0	55,0	58,5	62,0	64,0									
17	826388	52,5	57,0	60,0	67,0	69,0	71,0	73,0							
19	823124	53,0	57,0	62,0	64,0	66,5	68,5	73,0							
23	818146	51,5	53,5	57,0	60,5	64,5	66,0	66,0	68,0						
24	824679	54,0	54,0	61,0	62,5	65,0	67,5		70,5	71,0	74,5				
25	821655	52,0	55,0	59,0	61,5	65,0	65,0	66,0	69,0						
26	820716	48,0	55,0	57,0	59,0	59,5	63,0	64,5	65,5	68,0					
32	818145	55,5	61,0	66,0	67,0	72,0	73,5	76,0	77,0						
33	823591	54,0	59,0	60,0	64,0	66,0	66,0	71,0	72,0	74,0	77,0	78,0			
38	819397	58,0	63,0	63,0	69,0	72,0	72,0	75,0	76,0						
40	824691	62,0	65,0	66,0	68,0	70,0	70,0	74,0							
41	830557	51,5	58,0	63,0	64,0	70,0									
44	840177	47,0	54,0	59,0	61,0	63,0	67,0								
45	820694	51,0	53,0	56,5	59,0	64,0	65,0	67,0	67,5	67,5	68,0	69,0	70,0		
46	824274	52,0	56,0	59,0	59,0	63,0	64,0	65,5	66,0	67,0	69,5				
47	820704	53,0	66,0	66,0	70,5	71,0	72,5								
48	829952	55,5	58,5	61,0	63,5	67,5	68,0	71,5	73,5	74,5					
49	820918	54,0	59,5	61,5	68,0	71,0	71,0								
52	820300	60,5	67,0	67,0	68,0	69,0	71,0	72,5	75,5	77,5					

ANEXO 2 – Script OpenBugs

```

model
{
  b0 ~ dnorm(0,0.001)
  b1 ~ dnorm(0,0.001)
  b2 ~ dnorm(0,0.001)
  b3 ~ dnorm(0,0.001)
  b4 ~ dnorm(0, 0.001)
  b5 ~ dnorm(0, 0.001)
  b6 ~ dnorm(0, 0.001)
  b7 ~ dnorm(0, 0.001)
  precisao ~ dgamma(2.2,10)
  phi ~ dgamma(0.44,0.67)
  for(k in 1:62){
  for(i in 1:N){
  for(j in 1:N){
  Omega[k,i,j] <- exp(-phi*Distancia[k,i,j])/precisao
  }
  }
  }
  for(i in 1:M){
  varcovar[i, 1:N, 1:N]<-inverse(Omega[i,1:N, 1:N])
  Perimetro_cefalico[i, 1:N] ~ dnorm(mu[i,1:N], varcovar[i,,])
  Z[i,1:N] ~ dnorm(mu[i,1:N],varcovar[i,,])
  for(j in 1:N){
  mu[i,j]<- beta[i]
  + b1 * Comprimento[i,j]
  + b2 * Idade[i,j]
  + b4 * Peso[i,j]
  difquad[i,j]<-pow(Perimetro_cefalico[i,j]-Z[i,j],2)
  }
  beta[i]<- b0
  + b3 * Sexo[i]
  + b5 * PCRpositivo[i]
  + b6 * SemPCR[i]
  + b7 * SZCsExan[i]
  }
  quadmed<-mean(difquad[,])
  aic<-M*N*log(quadmed)+2*6
  bic<-M*N*log(quadmed)+6*log(310*2)
  }

```

ANEXO 3 – Modelos individuais apresentados na Tabela 6

- Modelo 1

– Variável: Comprimento

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i + \beta_{1,1} \text{Comprimento}_{i,j} \\
 \mu_i &= \beta_0 \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{1,1} &\sim N(0, 1000) \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$

- Modelo 2

– Variável: Idade

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i + \beta_{1,1} \text{Idade}_{i,j} \\
 \mu_i &= \beta_0 \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{1,1} &\sim N(0, 1000) \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$

- Modelo 3

– Variável: Sexo

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i \\
 \mu_i &= \beta_0 + \beta_{2,1}Sexo_{1,i} \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{2,1} &\sim N(0, 1000) \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$

• Modelo 4

– Variável: Peso

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i + \beta_{1,1}Peso_{i,j} \\
 \mu_i &= \beta_0 \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{1,1} &\sim N(0, 1000) \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$

• Modelo 5

– Variáveis: Grupos de expostos ao ZIKV, Zika suspeito e microcefalia sem exantema materno

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i \\
 \mu_i &= \beta_0 + \beta_{2,1}PCR_{1,i}^+ + \beta_{2,2}sPCR_{2,i} + \beta_{2,3}SZC_{3,i} \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{2,k_2} &\sim N(0, 1000), k_2 = 1, 2, 3 \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$

ANEXO 4 – Modelo completo

- Variáveis: Comprimento, peso, Zika suspeito e microcefalia sem exantema materno

$$\begin{aligned}
 Y_{i,j} &\sim N(\mu_{i,j}, V_{0,..}) \\
 \mu_{i,j} &= \mu_i + \beta_{1,1} \text{Comprimento}_{i,j} + \beta_{1,2} \text{Peso}_{i,j} \\
 \mu_i &= \beta_0 + \beta_{2,1} sPCR_{1,i} + \beta_{2,2} SZC_{2,i} \\
 \{V_0\}_{j,k} &= \phi \exp\{-\tau h_{j,k}\} \\
 \beta_{1,k_1} &\sim N(0, 1000), k_1 = 1, 2 \\
 \beta_{2,k_2} &\sim N(0, 1000), k_2 = 1, 2 \\
 \phi &\sim \text{Gama}(0.44, 0.67) \\
 \tau &\sim \text{Gama}(2.2, 10)
 \end{aligned}$$