

Gabriela de Oliveira Monteiro

**Análise de dados de luxação de quadril em
crianças com Zika Congênita**

Niterói - RJ, Brasil

15 de dezembro de 2022

Gabriela de Oliveira Monteiro

**Análise de dados de luxação de
quadril em crianças com Zika
Congênita**

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

Co-Orientador(a): Dra. Jakeline Oliveira da Fonseca

Niterói - RJ, Brasil

15 de dezembro de 2022

Gabriela de Oliveira Monteiro

**Análise de dados de luxação de quadril em
crianças com Zika Congênita**

Monografia de Projeto Final de Graduação sob o título “*Análise de dados de luxação de quadril em crianças com Zika Congênita*”, defendida por Gabriela de Oliveira Monteiro e aprovada em 15 de dezembro de 2022, 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. Mariana Albi de Oliveira Souza
Departamento de Estatística – UFF

Profa. Dra. Núbia Karla de Oliveira Almeida
Departamento de Estatística – UFF

Niterói, 15 de dezembro de 2022

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

M772a Monteiro, Gabriela de Oliveira
Análise de dados de luxação de quadril em crianças com
Zika Congênita / Gabriela de Oliveira Monteiro. - 2022.
42 f.

Orientador: Luis Guillermo Coca Velarde.
Coorientador: Jakeline Oliveira da Fonseca.
Trabalho de Conclusão de Curso (graduação)-Universidade
Federal Fluminense, Instituto de Matemática e Estatística,
Niterói, 2022.

1. Zika congênita. 2. Luxação de quadril. 3. Regressão
logística. 4. Inferência Bayesiana. 5. Produção
intelectual. I. Velarde, Luis Guillermo Coca, orientador. II.
Fonseca, Jakeline Oliveira da, coorientador. III. Universidade
Federal Fluminense. Instituto de Matemática e Estatística.
IV. Título.

CDD - XXX

Resumo

A Zika é uma doença transmitida pela picada do mosquito *Aedes Aegypti*, o mesmo transmissor da dengue. A doença foi identificada pela primeira vez no Brasil em abril de 2015, na cidade de Camaçari, próximo a Salvador (BA). Sua primeira descrição foi em um macaco em 1947, no país de Uganda, na floresta Zika, onde o vírus recebeu o nome do seu local de origem e o primeiro isolamento em humanos foi em 1952, também em Uganda. O vírus tem potencial de causar diversas formas de sintomas, de leves a graves, como exantema (manchas vermelhas na pele), dor de cabeça, febre baixa, conjuntivite e dores nas articulações. A Síndrome da Zika Congênita (SZC) se manifesta em recém-nascidos de mães que tenham sido infectadas pelo vírus durante a gravidez. Os principais sinais identificados nas crianças são alterações neurológicas (microcefalia), oftalmológicas e ortopédicas. O objetivo central do trabalho é descobrir possíveis fatores associados às luxações de quadril em crianças com Zika congênita. O estudo foi realizado com base nos dados de 31 crianças residentes no Estado do Rio de Janeiro e portadoras da Zika congênita. Foi usado um modelo de regressão logística sob um enfoque bayesiano, utilizando o Método de Monte Carlo via Cadeia de Markov (MCMC), para quantificar os efeitos dos diversos fatores sobre a condição de luxação das crianças do estudo. Inicialmente, foram considerados modelos com apenas uma variável regressora, e aquelas que se mostraram significativas foram consideradas no modelo múltiplo. Sendo assim, observou-se que os fatores que influenciam na ocorrência de luxação de quadril são a espasticidade, apresentando razão de chances de 19,79, e a alteração oftalmológica, com razão de chances de 7,35. Em outras palavras, as chances de uma criança ter luxação de quadril apresentando espasticidade são quase 20 vezes as chances de ter luxação quando a criança não apresenta espasticidade e a chance de apresentar luxação tendo observado a alteração oftalmológica é 7,35 vezes as chances em crianças que não possuem essa alteração.

Palavras-chave: Zika congênita. Luxação de quadril. Regressão logística. Inferência Bayesiana.

Dedicatória

Dedico esse trabalho aos meus familiares, por todo o apoio, em especial para minha avó, pelo amor que ela sempre me deu. Dedico também, para a minha companheira de curso e de vida; Carol, sem você nada disso seria possível, te amo.

Agradecimentos

Agradeço ao meu orientador, o Professor Dr. Luis Guillermo Coca Velarde por ter aceitado acompanhar-me neste projeto. A sua paciência, sua dedicação e suas motivações foram essenciais na minha caminhada até a conclusão desse trabalho. Obrigada por todo conhecimento compartilhado, te admiro como profissional e como pessoa.

Agradeço a minha co-orientadora, Dra. Jakeline Oliveira da Fonseca por todo o suporte na área médica, e por sua dedicação em me ajudar.

Expresso minha gratidão aos membros da banca, Profa. Dra. Mariana Albi de Oliveira Souza e a Profa. Dra. Núbia Karla de Oliveira Almeida, por fazerem parte do meu trabalho.

Por último, mas não menos importante, agradeço a todos os professores do departamento de Estatística da Universidade Federal Fluminense pelos ensinamentos ao longo da minha formação e aos meus colegas de curso.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Materiais e Métodos	p. 15
2.1	Dados	p. 15
2.1.1	Local e período de estudo	p. 15
2.1.2	Critérios de inclusão e exclusão	p. 15
2.1.3	Descrição dos dados	p. 16
2.2	Metodologia	p. 17
2.2.1	Regressão Logística	p. 18
2.2.1.1	Função de verossimilhança	p. 20
2.2.2	Enfoque bayesiano	p. 22
2.2.2.1	Teorema de Bayes	p. 22
2.2.2.2	Distribuições a priori e a posteriori	p. 23
2.2.3	Inferência usando enfoque Bayesiano	p. 24
2.2.3.1	Amostrador de Gibbs	p. 25
2.2.3.2	Algoritmo Metropolis-Hastings	p. 26
2.2.3.3	Comparação dos modelos	p. 27
3	Análise dos resultados	p. 28

3.1	Análise Descritiva	p. 28
3.2	Análise Inferencial	p. 32
4	Conclusão	p. 38
	Referências	p. 40
	Apêndice 1 – <i>Script</i> dos modelos implementados do <i>OpenBugs</i>	p. 42

Lista de Figuras

1	Raio-x do quadril (luxado e normal).	p. 12
2	Exemplos da função de resposta logística: (a) $E(Y) = 1/(1 + \exp^{-6-x})$; (b) $E(Y) = 1/(1 + \exp^{-6+x})$	p. 19
3	Frequência relativa das crianças que moram em comunidade.	p. 28
4	Distribuição da espasticidade nas crianças com luxação, sem luxação e na amostra geral.	p. 30
5	Distribuição da microcefalia nas crianças com luxação, sem luxação e na amostra geral.	p. 31
6	Distribuição da alteração oftalmológica nas crianças com luxação, sem luxação e na amostra geral.	p. 31
7	Distribuição do Trimestre Gestacional do Exantema Materno nas crianças com luxação, sem luxação e na amostra geral.	p. 32
8	Traço das cadeias dos parâmetros do modelo individual e da deviance da variável Sexo.	p. 34
9	Distribuição a posteriori dos parâmetros do modelo individual e da deviance da variável Sexo.	p. 35
10	Traço das cadeias dos parâmetros e da <i>deviance</i> do modelo múltiplo.	p. 36
11	Distribuição a posteriori dos parâmetros e da <i>deviance</i> do modelo múltiplo.	p. 37

Lista de Tabelas

1	Tabela de Frequências (absolutas e relativas) das Variáveis x Luxação de quadril.	p. 29
2	Razão de chances, intervalo de confiança Bayesiano e deviance dos modelos individuais.	p. 33
3	Razão de chances, intervalo de confiança Bayesiano e deviance do modelo múltiplo.	p. 35

1 Introdução

O vírus Zika, conhecido pela sigla ZIKV, é um arbovírus transmitido por artrópodes, nesse caso pela picada do mosquito *Aedes Aegypti* infectado. O vírus tem capacidade de se expressar de diferentes maneiras em adultos, podendo ter sintomas mais leves, como: exantema, conhecida pelas manchas vermelhas na pele, dor de cabeça, febre baixa, conjuntivite e dores nas articulações. Já em recém-nascidos, pode ser manifestada de forma mais severa, com sinais de alterações neurológicas, oftalmológicas e ortopédicas. Sua primeira descrição foi em um macaco em 1947, no país de Uganda, na floresta Zika, onde o vírus recebeu o nome do seu local de origem. O primeiro isolamento em humanos foi em 1952, em Uganda; inicialmente não havia indícios de que o vírus causasse doenças em humanos. Segundo Moore et al. (2017), o ZIKV foi identificado pela primeira vez no Brasil em 2015, num surto na região Nordeste e, meses após esse surto, os primeiros casos da síndrome da Zika congênita foram diagnosticados.

A Síndrome da Zika Congênita (SZC) é uma doença identificada em bebês cujas mães foram infectadas pelo vírus da Zika durante a gestação. No Brasil, os primeiros casos foram identificados mais precisamente na cidade de Camaçari - Bahia, em março de 2015 (CAMPOS; BANDEIRA; SARDI, 2015). Em setembro desse mesmo ano, pesquisadores notaram um aumento no número de recém-nascidos com microcefalia - uma malformação congênita em que a cabeça dos bebês é menor do que o esperado - nas mesmas áreas em que o vírus Zika foi relatado pela primeira vez (PETERSEN et al., 2016).

Apesar da microcefalia congênita ter sido o pontapé inicial para a análise da síndrome, outras anomalias também foram identificadas em recém-nascidos com a SZC, incluindo desproporção craniofacial, espasticidade, convulsões, irritabilidade, disfunção do tronco encefálico, como problemas de deglutição, anomalias auditivas e oftalmológicas e alterações ortopédicas, como a luxação de quadril, que será o foco deste trabalho (TEIXEIRA et al., 2020).

A luxação de quadril é a perda da ligação adequada da cabeça do fêmur com a pelve. O

quadril é uma das articulações do corpo que, ao sofrer alterações estruturais, pode levar a um desequilíbrio muscular que conduz a uma luxação ou sub-luxação. Tal anomalia pode ser identificada em crianças ao nascer e também há casos do desenvolvimento de luxação de quadril ao longo do tempo.

A identificação da luxação de quadril é feita através do exame de raio-X. É possível acompanhar a evolução da doença a partir dessas imagens radiológicas, onde é feito um cálculo para quantificar o percentual migratório, que indica o quanto o quadril está lateralizado em relação ao acetábulo (LINS; WATKINS; SHORE, 2019), conforme mostra a imagem abaixo.



Figura 1: Raio-x do quadril (luxado e normal).

O grande número de ocorrências desse tipo de anomalias em bebês despertou o interesse e a necessidade de estudar a doença. O reconhecimento desse fenótipo por médicos pode ajudar a estudar a causa da doença e, de forma apropriada, investigar clinicamente a abrangência de anomalias em um bebê afetado, bem como determinar o acompanhamento essencial e o cuidado adequado.

O objetivo principal desse trabalho é identificar os fatores associados à variável resposta, definida como ocorrência de luxação de quadril em crianças com SZC, e além disso, quantificar os efeitos de diversos fatores através da razão de chances¹ sobre a probabilidade de se observar a luxação de quadril. Para isso, será utilizado um modelo de regressão para modelar a variável resposta de interesse doravante denotada por Y . Como Y é binária, pode ser modelada através da distribuição Bernoulli, fazendo necessária a utilização dos Modelos Lineares Generalizados (MLG), mais especificamente a regressão logística, aqui tratada sob o ponto de vista bayesiano.

¹Tamanho do efeito que indica a chance de ocorrência de um desfecho em um grupo quando comparado a outro. Seus valores variam entre 0 e infinito.

Neste trabalho, é utilizada a regressão logística como principal ferramenta para modelagem de uma variável categórica dicotômica com base em uma ou mais variáveis contínuas e/ou binárias. Desta forma, pode-se determinar ou prever a probabilidade de que um evento ocorra. A regressão logística tem sido utilizada em estudos financeiros, ambientais e epidemiológicos para mostrar sua aplicabilidade nestas áreas, bem como sua capacidade de analisar dados de resposta categórica.

Um exemplo é o estudo de Leite et al. (2011) sobre a qualidade do ar de Uberlândia, MG. O objetivo é utilizar a regressão logística para investigar possíveis fatores que influenciam a qualidade do ar, tendo como resultado que os seguintes fatores contribuem positivamente: umidade relativa, o verão, a primavera, o sábado e o domingo. Em contrapartida, o inverno e o fluxo de veículos crescente conduzem a piora da qualidade do ar. Outros exemplos são os trabalhos de Beserra et al. (2022) e Araújo et al. (2011), que utilizaram a mesma técnica para analisar a concessão de crédito bancário e o perfil epidemiológico dos pacientes atendidos em um serviço de referência no MT, respectivamente.

Outros trabalhos utilizando diferentes ferramentas para dados de luxação de quadril foram encontrados recentemente na literatura. Murphy et al. (2022) apresentou uma ferramenta de triagem para pacientes com suspeita de luxação do quadril, com base na história clínica e nos resultados do exame de referência. Esse estudo apresentou uma amostra de tamanho 934 referente a crianças com suspeita de displasia do desenvolvimento do quadril, observadas por um período de três anos. Como ferramenta, foi utilizada a regressão logística multivariada para estabelecer preditores de luxação do quadril e, a partir disso, derivou-se um modelo preditivo que incorporou preditores significativos de luxação. O modelo preditivo final incluiu idade, sexo, histórico familiar, pélvico, problemas de marcha, diminuição da abdução, discrepância no comprimento das pernas e síndrome médica/neurológica. Foi visto que a predição clínica, com base no histórico prontamente disponível e nos parâmetros do exame, previu fortemente o risco de luxações no encaminhamento para displasia do quadril.

Já no trabalho de Bonsel et al. (2022) o objetivo foi usar a modelagem estatística em imagens de ultrassonografia para identificar as características da forma acetabular dos quadris de bebês, que permitem prever o desenvolvimento de quadris de diferentes tipos e identificar quais quadris se beneficiam do tratamento com o suspensório de Pavlik. A análise considerou o tratamento de 104 bebês de 3 a 4 meses com displasia de quadril, realizado entre 2009 e 2015. Foi utilizada uma regressão logística multivariada para sexo e tratamento (tratamento com suspensório de Pavlik versus observação ativa). Dentre

os 3 tipos de desenvolvimento de quadril, os tipos 2 e 3 da modelagem estatística foram associados à displasia persistente do quadril em imagens de ultrassonografia. O tipo 2 também foi associado com displasia residual do quadril em radiografias pélvicas. O termo de interação permaneceu significativo após a análise multivariada, indicando que o tratamento com o suspensório de Pavlik foi benéfico em pacientes com quadril do tipo 2.

No estudo de Fonseca (2022), que utiliza os mesmos dados deste trabalho, observou-se através da razão de chances, que as crianças com SZC e espasticidade (contratura muscular no quadril) apresentam cerca de 15,9 vezes mais chances de apresentarem luxação do quadril em relação àquelas sem essas alterações. De forma análoga, crianças com SZC e alteração oftalmológica apresentam cerca de 16,9 vezes mais chances de apresentarem a luxação do quadril. A título de comparação, na dissertação de Fonseca (2022) foram utilizados testes de hipóteses e regressão logística; portanto, ao utilizar inferência bayesiana neste trabalho, pode-se evidenciar o ganho com a inclusão dessa ferramenta.

No que segue, os demais capítulos deste trabalho estão organizados de forma a apresentar os dados e a metodologia no Capítulo 2, a análise e discussão dos resultados no Capítulo 3 e, por fim, a conclusão no Capítulo 4.

2 Materiais e Métodos

2.1 Dados

Os dados desse trabalho são do “Estudo clínico e epidemiológico de crianças expostas ao vírus Zika durante o período gestacional: estudo prospectivo de coorte”, coordenado pela pesquisadora Claudete Aparecida Araújo Cardoso na Faculdade de Medicina da Universidade Federal Fluminense. Esse estudo faz parte do “Projeto Zika”, composto por uma equipe interdisciplinar de pesquisadores, entre eles, pesquisadores da Fundação Oswaldo Cruz (Fiocruz) e da *La Jolla Institute for Allergy and Immunology*, da Califórnia. A pesquisa é proveniente da coleta de dados de mães que tiveram Zika na gestação e de seus filhos expostos ao vírus durante a gravidez e do acompanhamento dessas crianças, algumas delas até a atualidade.

2.1.1 Local e período de estudo

A coleta dos dados foi realizada na Unidade de Pesquisa Clínica do Hospital Universitário Antônio Pedro (HUAP/UFF), que atende os municípios de Niterói, Itaboraí, Maricá, Rio Bonito, São Gonçalo, Silva Jardim e Tanguá, no estado do Rio de Janeiro. O acompanhamento dessas crianças teve início em 2016 e estas são acompanhadas periodicamente. A cada 3 meses são realizadas avaliações clínicas, pediátricas, ortopédicas e oftalmológicas. Já o exame radiográfico da bacia é realizado a cada 6 meses.

2.1.2 Critérios de inclusão e exclusão

O critério de inclusão adotado na pesquisa selecionou crianças com diagnóstico de SZC acompanhadas no ambulatório de doenças exantemáticas.

Já os critérios de exclusão foram: crianças com diagnóstico perinatal de outras infecções congênicas (sífilis, toxoplasmose, rubéola, citomegalovírus e HIV), crianças cujas

mães foram RT-qPCR negativo para ZIKV, crianças com microcefalia de outra etiologia, crianças com asfixia perinatal e crianças cujas mães foram RT-qPCR positivo para ZIKV, mas assintomáticos para SZC.

2.1.3 Descrição dos dados

A população do estudo é composta por crianças com Síndrome da Zika Congênita, classificadas em três categorias:

1. crianças com SZC cujas mães apresentaram exantema e infecção pelo ZIKV confirmada por RT-qPCR durante a gestação;
2. crianças com SZC cujas mães apresentaram exantema, mas a infecção pelo ZIKV não foi confirmada durante a gravidez, ou seja, sem a coleta RT-qPCR;
3. crianças com diagnóstico clínico radiológico de SZC cujas mães não apresentaram exantema durante a gestação, e sem coleta RT-qPCR durante a gestação.

As variáveis disponíveis na base de dados foram descritas a seguir:

- Grupo: definida como a população de estudo pelas categorias 1, 2 e 3 conforme descrito acima.
- Sexo: identifica o sexo da criança, sendo 1 masculino e 0 feminino.
- Luxação de Quadril: variável binária, que identifica quando a articulação do quadril fica fora do lugar correto, sendo 1 quando a criança possui a luxação de quadril e 0 quando não possui.
- Espasticidade: variável binária que identifica a existência de contratura muscular, mais especificamente no quadril, sendo 1 quando possui espasticidade e 0 quando não possui.
- Alteração neurológica: variável binária, onde 1 identifica uma anormalidade cerebral e 0 quando não possui essa anormalidade.
- Microcefalia: variável binária, que identifica uma malformação congênita na cabeça dos recém-nascidos, onde é menor do que o esperado, se comparada com a de bebês do mesmo sexo e idade que não possuem a má formação; sendo 1 quando possui esta malformação e 0 quando não possui.

- Alteração oftalmológica: variável binária, sendo 1 quando a criança apresenta alguma alteração oftalmológica e 0 quando não possui.
- Exantema materno: identifica a ocorrência de exantema (manchas vermelhas na pele) na mãe durante a gestação, sendo 1 quando houve a ocorrência do exantema e 0 quando não houve.
- Estimulação: variável binária que identifica se a criança teve uma estimulação com um fisioterapeuta depois de ter sido diagnosticada com luxação de quadril, sendo 1 quando a criança teve a estimulação e 0 quando não teve.
- Trimestre Gestacional do exantema materno: variável categórica, criada para identificar em qual trimestre da gestação a mãe teve o exantema, categorizada como “primeiro trimestre”, “segundo trimestre” e “terceiro trimestre”.

A amostra foi construída com 31 crianças incluídas no projeto e que passaram pelos critérios de inclusão. A ferramenta utilizada neste trabalho foi o *software OpenBugs* (OPENBUGS, 2014) e o nível de significância Bayesiano adotado foi de 5%.

2.2 Metodologia

Nesta Seção serão descritos os procedimentos estatísticos necessários para o desenvolvimento deste trabalho. Mais detalhes estão disponíveis em Dobson e Barnett (2018) e Ehlers (2003).

Com o objetivo de modelar dados em que possivelmente existe uma relação de causa e efeito entre as variáveis dependente e regressoras, geralmente se utiliza um modelo de regressão. Entretanto, os modelos lineares clássicos assumem que a variável dependente (Y) seja proveniente de uma distribuição Normal. Visto que a variável resposta deste trabalho não satisfaz a condição de normalidade, pode-se recorrer aos Modelos Lineares Generalizados (MLG), os quais admitem que Y possa ser proveniente de outras distribuições que pertençam à família exponencial.

Definição 2.1 (*Família Exponencial com um parâmetro*) *A distribuição de uma variável aleatória Y pertence à família exponencial quando sua função (densidade) de probabilidade puder ser reescrita da seguinte forma:*

$$f(y|\theta) = a(y) \cdot b(\theta) \cdot \exp\{c(y) \cdot d(\theta)\} \quad (2.1)$$

onde $a(y)$, $b(\theta)$, $c(y)$ e $d(\theta)$ são funções conhecidas.

No caso $Y \sim \text{Bernoulli}(\theta)$:

$$\begin{aligned} f(y|\theta) &= \theta^y(1-\theta)^{1-y}I_{\{0,1\}}(y) \\ &= I_{\{0,1\}}(y)\exp\{y(\ln(\theta) - \ln(1-\theta))\}(1-\theta) \end{aligned} \quad (2.2)$$

onde $a(y) = I_{\{0,1\}}(y)$, $b(\theta) = 1 - \theta$, $c(y) = y$ e $d(\theta) = \ln(\theta) - \ln(1 - \theta)$. Logo, a distribuição de Y pertence a família exponencial.

2.2.1 Regressão Logística

A teoria apresentada nesta subseção foi obtida de Myers et al. (2010).

Neste trabalho a variável resposta é a ocorrência de luxação de quadril em crianças que tiveram Zika congênita e ela pode ser modelada usando a distribuição de Bernoulli.

O modelo de regressão logística é um caso especial dos MLG's. Esta técnica é utilizada para estabelecer a relação entre uma variável dependente que é binária ($Y_i \sim \text{Bernoulli}(\theta_i)$), isto é, indicadora de ocorrência ou não de um evento, e um conjunto de variáveis explicativas.

Vamos supor que o modelo tem a seguinte forma:

$$Y_i = \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i, \quad (2.3)$$

onde:

- Y_i é o valor da variável resposta para a i -ésima observação,
- $\alpha, \beta_1, \dots, \beta_k$ são parâmetros desconhecidos,
- X_{i1}, \dots, X_{ik} é o valor das k variáveis explicativas para a i -ésima observação, e
- ϵ_i é o termo de erro aleatório independente, sendo $\epsilon_i \sim N(0, \sigma^2)$.

Uma vez que $E(\epsilon_i) = 0$, temos que:

$$E(Y_i) = \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} = \theta_i, \quad (2.4)$$

porém há uma restrição na função de resposta: já que estamos falando sobre probabilidades, os valores tem que estar entre 0 e 1, ou seja,

$$0 \leq E(Y_i) = \theta_i \leq 1. \quad (2.5)$$

Essa restrição pode causar problemas na escolha de uma função linear, pois deseja-se que os valores da função resposta fiquem entre 0 e 1. Geralmente, quando a variável resposta é binária, há evidências que a forma da função resposta deva ser não-linear, ultrapassando os valores de 0 e 1. Por esse motivo, busca-se a curva, no formato de “S”, conhecida pela função logística. A função logística tem a seguinte forma:

$$E(Y_i) = \frac{\exp^{\alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + \exp^{\alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} = \frac{1}{1 + \exp^{-(\alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik})}} \quad (2.6)$$

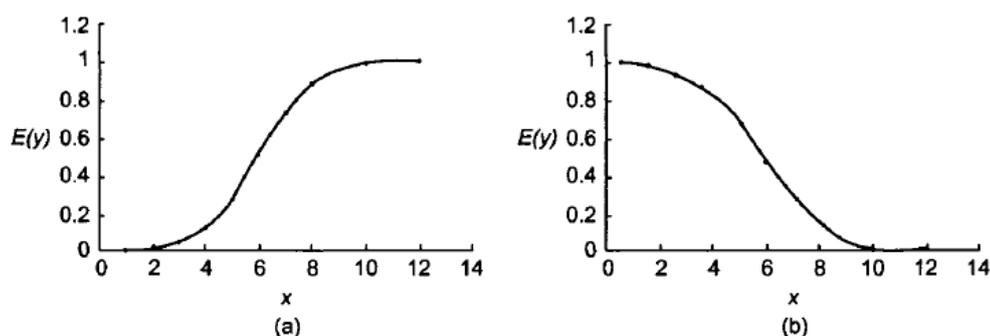


Figura 2: Exemplos da função de resposta logística: (a) $E(Y) = 1/(1 + \exp^{-6-x})$; (b) $E(Y) = 1/(1 + \exp^{-6+x})$.

(Fonte: (MYERS et al., 2010))

A função logística pode ser linearizada. Considere, então, o preditor linear:

$$\eta_i = \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} \quad (2.7)$$

onde η_i é definido pela transformação

$$\eta_i = \ln \left(\frac{\theta_i}{1 - \theta_i} \right), i = 1, \dots, n. \quad (2.8)$$

Esta transformação é chamada de *transformação logit* e $\frac{\theta_i}{1 - \theta_i}$ é chamada de chances. Tal transformação é conhecida como função de ligação, que fornece o relacionamento entre o preditor linear e a esperança da variável resposta. As funções de ligação podem variar

de uma família de distribuições para outra. Portanto, ao utilizar a função de ligação *logit*, também chamada de logito, a regressão passa a ser chamada de regressão logística.

A título de interpretação, considere uma única variável regressora numérica. Para identificar o que acontece ao acréscimo de uma unidade na variável regressora, defina $X = k$ e $X = k + 1$, portanto, tem-se as seguintes regressões:

$$\text{logit}(\theta)_{X=k+1} = a + b(k + 1) = a + bk + b \quad (2.9)$$

$$\text{logit}(\theta)_{X=k} = a + bk \quad (2.10)$$

Fazendo a diferença,

$$\begin{aligned} \text{logit}(\theta)_{X=k+1} - \text{logit}(\theta)_{X=k} &= a + bk + b - a - bk \\ \log\left(\frac{\theta}{1-\theta}\right)_{X=k+1} - \log\left(\frac{\theta}{1-\theta}\right)_{X=k} &= b \\ \log\frac{\left(\frac{\theta}{1-\theta}\right)_{X=k+1}}{\left(\frac{\theta}{1-\theta}\right)_{X=k}} &= b \\ \frac{\left(\frac{\theta}{1-\theta}\right)_{X=k+1}}{\left(\frac{\theta}{1-\theta}\right)_{X=k}} &= e^b \end{aligned} \quad (2.11)$$

Logo, a Equação (2.11) expressa a razão de chances quando existe um aumento unitário na variável X .

Estes resultados vistos acima podem ser estendidos para o caso em que a variável regressora é binária. Ou seja, no ponto de vista prático, os resultados deste trabalho são apresentados em relação a razão de chances, que é calculada a partir da exponencial dos coeficientes de regressão para as variáveis regressoras.

2.2.1.1 Função de verossimilhança

A função de verossimilhança contém toda a informação dos dados relevante para fazer inferências sobre os parâmetros. Portanto, o princípio da verossimilhança aponta a importância do que de fato foi observado, tomando o resultado que o experimento gerou e não como o experimento foi realizado.

Como a variável resposta desse trabalho segue uma distribuição de Bernoulli ($Y_i \sim$

$Bernoulli(\theta_i)$), sua função de probabilidade é definida da seguinte forma:

$$P(Y_i = y_i | \theta_i) = \theta_i^{y_i} (1 - \theta_i)^{1-y_i} I_{\{0,1\}}(y_i), i = 1, 2, \dots, n, \quad (2.12)$$

onde θ_i representa a probabilidade da criança i ter luxação de quadril.

A função de verossimilhança é dada por:

$$L(\underline{\theta}; \underline{y}) = \prod_{i=1}^n \theta_i^{y_i} (1 - \theta_i)^{1-y_i} \quad (2.13)$$

Lembrando que $E(Y_i) = \theta_i$, de acordo com (2.6). Como visto na Seção 2.2.1, o modelo logito resulta na seguinte função de ligação:

$$\begin{aligned} \ln \left(\frac{\theta_i}{1 - \theta_i} \right) &= \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} \\ \left(\frac{\theta_i}{1 - \theta_i} \right) &= e^{\alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}. \end{aligned} \quad (2.14)$$

Nas contas abaixo iremos usar apenas X_i , uma única variável regressora, apenas para simplificar. Logo temos:

$$L(\underline{\theta}; \underline{y}) = \prod_{i=1}^n \left(\frac{e^{\alpha + \beta_1 x_{i1}}}{1 + e^{\alpha + \beta_1 x_{i1}}} \right)^{y_i} \left(1 - \frac{e^{\alpha + \beta_1 x_{i1}}}{1 + e^{\alpha + \beta_1 x_{i1}}} \right)^{1-y_i}; \quad (2.15)$$

portanto,

$$\begin{aligned}
\ln L(\theta; \underline{y}) &= \sum_{i=1}^n y_i \ln \left(\frac{e^{\alpha + \beta_1 x_{i1}}}{1 + e^{\alpha + \beta_1 x_{i1}}} \right) + \sum_{i=1}^n (1 - y_i) \ln \left(\frac{1}{1 + e^{\alpha + \beta_1 x_{i1}}} \right) \\
&= \sum_{i=1}^n y_i \ln(e^{\alpha + \beta_1 x_{i1}}) - \sum_{i=1}^n y_i \ln(1 + e^{\alpha + \beta_1 x_{i1}}) - \sum_{i=1}^n (1 - y_i) \ln(1 + e^{\alpha + \beta_1 x_{i1}}) \\
&= \sum_{i=1}^n y_i (\alpha + \beta_1 x_{i1}) - \sum_{i=1}^n y_i \ln(1 + e^{\alpha + \beta_1 x_{i1}}) - \sum_{i=1}^n \ln(1 + e^{\alpha + \beta_1 x_{i1}}) \\
&\quad + \sum_{i=1}^n y_i \ln(1 + e^{\alpha + \beta_1 x_{i1}}) \\
&= \alpha \sum_{i=1}^n y_i + \beta_1 \sum_{i=1}^n y_i x_{i1} - \sum_{i=1}^n \ln(1 + e^{\alpha + \beta_1 x_{i1}}) \tag{2.16}
\end{aligned}$$

2.2.2 Enfoque bayesiano

Ao invés de usar o método clássico, onde os parâmetros desconhecidos são considerados valores fixos, a abordagem Bayesiana será usada neste estudo. A abordagem Bayesiana assume que os parâmetros desconhecidos não são valores fixos, mas sim variáveis aleatórias que apresentam distribuições de probabilidade que resumem o conhecimento atual sobre a quantidade não observável e geralmente desconhecida. A inferência Bayesiana leva em conta o conhecimento prévio sobre o vetor de parâmetros, a fim de fazer inferências sobre este. Nesta metodologia, os modelos de probabilidade representam a incerteza do parâmetro, portanto todos os parâmetros são considerados como variáveis aleatórias.

2.2.2.1 Teorema de Bayes

Suponha que há interesse em alguma quantidade desconhecida θ (que geralmente não é observável). Ao reunir dados sobre θ e sobre alguma variável aleatória Y observável e que esteja relacionada a θ , pode-se quantificar esta relação usando uma distribuição amostral $p(y|\theta)$. Neste caso, após observar $Y = y$, é possível aumentar a informação sobre θ e o teorema de Bayes é utilizado para quantificar este aumento, de tal forma que

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{2.17}$$

onde $p(y)$ é dada por:

$$p(y) = \int p(\theta, y) d\theta. \tag{2.18}$$

Sabe-se que $1/p(y)$, que não depende de θ , funciona como uma constante normalizadora de $p(\theta|y)$, que é chamada de distribuição a posteriori. Para um valor fixo de y , a função $l(\theta; y) = p(y|\theta)$ fornece a verossimilhança de cada um dos possíveis valores de θ , enquanto $p(\theta)$ é chamada distribuição a priori de θ .

2.2.2.2 Distribuições a priori e a posteriori

A distribuição a priori de θ , $p(\theta)$, é definida em inferência Bayesiana como sendo a distribuição do parâmetro de interesse θ antes da realização do experimento aleatório, ou seja, ela deve conter todo o conhecimento prévio sobre o parâmetro.

Combinando as duas fontes de informação, a distribuição a priori e a função de verossimilhança, a distribuição a posteriori de θ é definida por $p(\theta|y)$. Assim, pode-se reescrever o teorema de Bayes, de forma usual, como:

$$p(\theta|y) \propto l(\theta; y)p(\theta). \quad (2.19)$$

Em outras palavras, a distribuição a posteriori é proporcional ao produto da verossimilhança com a distribuição a priori.

Portanto, a distribuição a posteriori contém toda a informação já existente na distribuição a priori e no experimento realizado, presente na função de verossimilhança. Desta forma, a distribuição a posteriori pode ser vista como uma distribuição com informação atualizada do parâmetro de interesse.

Voltando ao exemplo simplificado na Seção 2.2.1.1, utilizando apenas uma variável regressora, precisa-se definir as distribuições a priori para os parâmetros. Nesse caso, serão utilizadas distribuições a priori não-informativas, que expressam ignorância relativa à informação sobre qual é o verdadeiro valor dos parâmetros, que pode ser fornecida por um dado experimento. Quando o experimento é alterado, então espera-se que a expressão de relativa ignorância possa ser alterada correspondentemente (BOX; TIAO, 1973). Isso será realizado para α e β_1 , que são os parâmetros do modelo, onde α é o intercepto e o β_1 é o parâmetro associado à variável regressora.

Então, as distribuições a priori não-informativas para os parâmetros α e β_1 serão definidos com distribuição Normal com média 0 e variância σ_α^2 e $\sigma_{\beta_1}^2$, respectivamente. Vale destacar que tais variâncias são, a priori, consideradas grandes. Ou seja,

$$\alpha \sim N(0, \sigma_\alpha^2) \text{ e } \beta_1 \sim N(0, \sigma_{\beta_1}^2)$$

Supondo que tais parâmetros são, a priori, independentes, a distribuição a priori conjunta é definida como o produto das distribuições a priori marginais:

$$\begin{aligned} p(\alpha, \beta_1) &= \left(\frac{1}{\sigma_\alpha \sqrt{2\pi}} \exp \left\{ -\frac{\alpha^2}{2\sigma_\alpha^2} \right\} \right) \left(\frac{1}{\sigma_{\beta_1} \sqrt{2\pi}} \exp \left\{ -\frac{\beta_1^2}{2\sigma_{\beta_1}^2} \right\} \right) \\ &= \frac{1}{\sigma_\alpha \sigma_{\beta_1} 2\pi} \exp \left\{ -\frac{1}{2} \left(\frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta_1^2}{\sigma_{\beta_1}^2} \right) \right\}. \end{aligned} \quad (2.20)$$

Definida a distribuição a priori conjunta e dada a função de verossimilhança na Equação (2.16), utilizando o Teorema de Bayes adaptado, em escala logaritmica, a log posteriori será:

$$\ln(p(\underline{\theta}|y)) = \ln(l(\underline{\theta}; y)) + \ln(p(\underline{\theta})) - \ln(p(y)) \quad (2.21)$$

onde $\underline{\theta} = (\alpha, \beta_1)$.

Portanto, a log posteriori pode ser descrita como:

$$\begin{aligned} \ln(p(\alpha, \beta_1|y, \underline{x})) &= \alpha \sum_{i=1}^n y_i + \beta_1 \sum_{i=1}^n y_i x_{i1} - \sum_{i=1}^n \ln(1 + e^{\alpha + \beta_1 x_{i1}}) \\ &\quad - \ln(\sigma_\alpha \sigma_{\beta_1} 2\pi) - \frac{1}{2} \left(\frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta_1^2}{\sigma_{\beta_1}^2} \right). \end{aligned} \quad (2.22)$$

2.2.3 Inferência usando enfoque Bayesiano

Para avaliar se um fator tem associação significativa foram feitas estimações utilizando o Método de Monte Carlo via Cadeia de Markov (MCMC), que permite obter amostras da distribuição a posteriori. Optou-se especificamente pelos Algoritmos de Metropolis-Hastings e Amostrador de Gibbs; além disso, foram considerados intervalos de confiança Bayesiano de 95%. Tais teorias foram consultadas em Gamerman e Lopes (2006) e Ehlers (2003).

O Método MCMC surge em Inferência Bayesiana no contexto em que as integrais utilizadas para o cálculo da distribuição a posteriori são frequentemente difíceis de calcular

por métodos numéricos, impossibilitando a definição das distribuições a posteriori, intervalos de confiança Bayesiano, entre outras medidas estatísticas. O método consiste em gerar valores de uma distribuição condicional a posteriori para cada parâmetro, através dos métodos computacionalmente intensivos, como o Amostrador de Gibbs e o Algoritmo de Metropolis-Hastings.

A Cadeia de Markov é um processo estocástico, na qual as distribuições condicionais dos estados futuros, X_{n+j} para $j \geq 1$ dado X_1, \dots, X_n , são independentes dos estados passados X_1, \dots, X_{n-1} , ou seja:

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

Para que o método MCMC possa ser utilizado, a cadeia deve ser homogênea, irredutível ou aperiódica (GAMERMAN; LOPES, 2006). Uma cadeia de Markov é homogênea quando as probabilidades de transição são constantes ao longo do tempo. Em uma cadeia de Markov irredutível, cada estado pode ser alcançado de qualquer outro estado; e uma cadeia aperiódica não apresenta estados absorventes. Empiricamente são gerados valores iniciais e, à medida que o número de iterações aumenta, os valores iniciais são esquecidos, correspondendo ao período de aquecimento, até que ocorra a convergência para a distribuição de equilíbrio da Cadeia de Markov (GABRY et al., 2019).

2.2.3.1 Amostrador de Gibbs

O Amostrador de Gibbs não possui um mecanismo de rejeição ou aceitação. Assim, é possível obter todas as distribuições condicionais realizando uma amostra de Gibbs, até a última etapa. Considere que $\pi(\theta)$ é a distribuição que se deseja amostrar, onde θ é um vetor ou matriz dividida em d elementos $\underline{\theta} = (\theta_1, \dots, \theta_d)$. O Amostrador funciona em três etapas:

1. fixar um contador de iterações da cadeia, $t = 1$ e estabelecer valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$;
2. obter um novo valor $\theta^{(t)}$, por meio de geração sucessiva de valores, a partir de $\theta^{(t-1)}$,

conforme

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

3. somar uma unidade no contador ($t = t + 1$) e voltar à etapa 2 até atingir a convergência.

A Cadeia de Markov converge após a conclusão da fase de aquecimento e pode ser identificada pelo gráfico da Cadeia de Markov.

2.2.3.2 Algoritmo Metropolis-Hastings

O Algoritmo Metropolis-Hastings emprega técnicas de rejeição onde um valor é gerado e aceito com uma determinada probabilidade a partir de uma distribuição auxiliar. Este método garante a convergência da cadeia para uma certa distribuição de interesse, que no caso da Inferência Bayesiana é a distribuição a posteriori. A cadeia tem o estado θ e um novo valor θ' que foi gerado por uma nova distribuição $q(\cdot|\theta)$. A aceitação de θ' é dada com a seguinte probabilidade, onde π é a distribuição de interesse:

$$\alpha(\theta, \theta') = \min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right). \quad (2.23)$$

A abordagem Bayesiana exige que se conheça π de forma parcial, a menos de uma constante, porque a probabilidade $\alpha(\theta, \theta')$ não se altera com a constante. Na prática o algoritmo de Metropolis-Hastings opera da seguinte forma:

1. iniciar o contador de iterações $t = 0$ e indicar o valor inicial $\theta^{(0)}$;
2. gerar um novo valor θ da distribuição $q(\cdot|\theta)$;
3. calcular a probabilidade de aceitação $\alpha(\theta, \theta')$ e gerar $u \sim U(0, 1)$;
4. se $u \leq \alpha$ então aceitar o novo valor e fazer $\theta^{(t+1)} = \theta'$, caso contrário, rejeitar e fazer $\theta^{(t+1)} = \theta$;
5. somar uma unidade ao contador t e voltar ao passo 2.

Geralmente, na abordagem Bayesiana, a distribuição a priori é utilizada como distribuição $q(\cdot|\theta)$ e a distribuição a posteriori como distribuição de interesse, ou seja, $\pi(\theta) = p(\theta|x)$.

2.2.3.3 Comparação dos modelos

Existem diversas métricas para se fazer uma comparação de qualidade dos modelos. Um método muito utilizado para esta abordagem é a função desvio ou *Deviance*, que mede a qualidade do ajuste e ajuda na decisão do modelo mais adequado. Isso será feito de acordo com o estabelecido em Gelfand e Ghosh (1998).

Para utilizar a função desvio da distribuição Bernoulli, foi usada a função da Binomial a seguir, considerando $m = 1$, onde m é o número de realizações da Binomial:

$$D(y; \hat{\mu}) = 2 \sum \{(y + 0, 1) \ln((y + 0, 1)/\hat{\mu}) + (m - (y + 0, 1)) \ln[(m - (y + 0, 1))/(m - \hat{\mu})]\} \quad (2.24)$$

onde y representa o valor observado e $\hat{\mu}$ o valor previsto pelo modelo.

A análise para decisão do melhor modelo é que, quanto menor o valor da *Deviance*, melhor é o ajuste. Isso acontece porque a *Deviance* expressa a diferença entre o modelo proposto e o modelo mais completo possível, ou seja, quanto mais próximos estes modelos estiverem, menor será o valor da função de desvio.

3 Análise dos resultados

3.1 Análise Descritiva

Depois de aplicados os critérios de inclusão e exclusão foram observadas na análise 31 crianças com idade média de 4 meses, onde 61% dessas crianças possuem luxação de quadril e aproximadamente 42% reside em comunidades de Niterói e de cidades circunvizinhas (Figura 3). A Tabela 1 mostra os resultados comparativos entre os grupos com e sem luxação de quadril.



Figura 3: Frequência relativa das crianças que moram em comunidade.

Tabela 1: Tabela de Frequências (absolutas e relativas) das Variáveis x Luxação de quadril.

Variável	Categorias	Total	Luxação	
			Sim	Não
Sexo	Feminino	10 (32,26%)	6 (31,58%)	4 (33,33%)
	Masculino	21 (67,74%)	13 (68,42%)	8 (66,67%)
Grupo	1	11 (35,48%)	6 (31,58%)	5 (41,67%)
	2	10 (32,26%)	7 (36,84%)	3 (25%)
	3	10 (32,26%)	6 (31,58%)	4 (33,33%)
Espasticidade	Sim	21 (67,74%)	17 (89,47%)	4 (33,33%)
	Não	10 (32,26%)	2 (10,53%)	8 (66,67%)
Microcefalia	Sim	27 (87,10%)	19 (100%)	8 (66,67%)
	Não	4 (12,90%)	0 (0%)	4 (33,33%)
Alteração Oftalmológica	Sim	23 (74,19%)	17 (89,47%)	6 (50%)
	Não	8 (25,81%)	2 (10,53%)	6 (50%)
Exantema materno	Sim	21 (67,74%)	13 (68,42%)	8 (66,67%)
	Não	10 (32,26%)	6 (31,58%)	4 (33,33%)
Estimulação	Sim	28 (93,33%)	17 (89,47%)	11 (100%)
	Não	2 (6,67%)	2 (10,53%)	0 (0%)
Trimestre gestacional do exantema materno	1º trimestre	16 (76,19%)	12 (92,31%)	4 (50%)
	2º trimestre	4 (19,05%)	0 (0%)	4 (50%)
	3º trimestre	1 (5%)	1 (8%)	0 (0%)
Mora comunidade	Sim	13 (41,94%)	6 (46,15%)	7 (38,89%)
	Não	18 (58,06%)	7 (53,85%)	11 (61,11%)

Observa-se, através dos resultados da Tabela 1, que a maioria dessas crianças são meninos (67,74%). Além disso, 35,48% são crianças com SZC cujas mães apresentaram exantema e infecção pelo ZIKV confirmada por RT-qPCR durante a gestação, 67,74% apresentam espasticidade e 87,1% possuem microcefalia. A alteração oftalmológica está presente em 74,19% dessas crianças e 93,33% foram crianças que tiveram uma estimulação com um fisioterapeuta depois de ter sido diagnosticada com luxação de quadril. Por fim, foi identificada a ocorrência de exantema (manchas vermelhas na pele) na mãe durante a gestação em 67,74% das crianças e, em 76,19% dessas crianças, as mães apresentaram o exantema no 1º trimestre gestacional.

É importante ressaltar que 100% das crianças observadas apresentaram alteração neu-

rológica, independente de ter apresentado luxação de quadril ou não.

Pode ser observado que 89,47% das crianças que possuem luxação de quadril, também possuem espasticidade. Já as crianças que não possuem luxação, apenas 33,33% apresenta espasticidade (Figura 4). Quanto à ocorrência de microcefalia, observa-se que, das crianças que possuem luxação, 100% também apresenta microcefalia. Das crianças que não possuem luxação, 66,67% apresenta microcefalia (Figura 5).

Além disso, nota-se que 89,47% das crianças com luxação também apresentam alteração oftalmológica e, das crianças que não possuem luxação, metade possui esta alteração (Figura 6). Por fim, das crianças que possuem luxação, 92,31% foram de mães que apresentaram exantema no primeiro trimestre gestacional e 8% apresentaram no terceiro trimestre; enquanto das crianças que não possuem luxação, 50% das mães apresentaram exantema no primeiro trimestre gestacional e 50% no segundo trimestre.

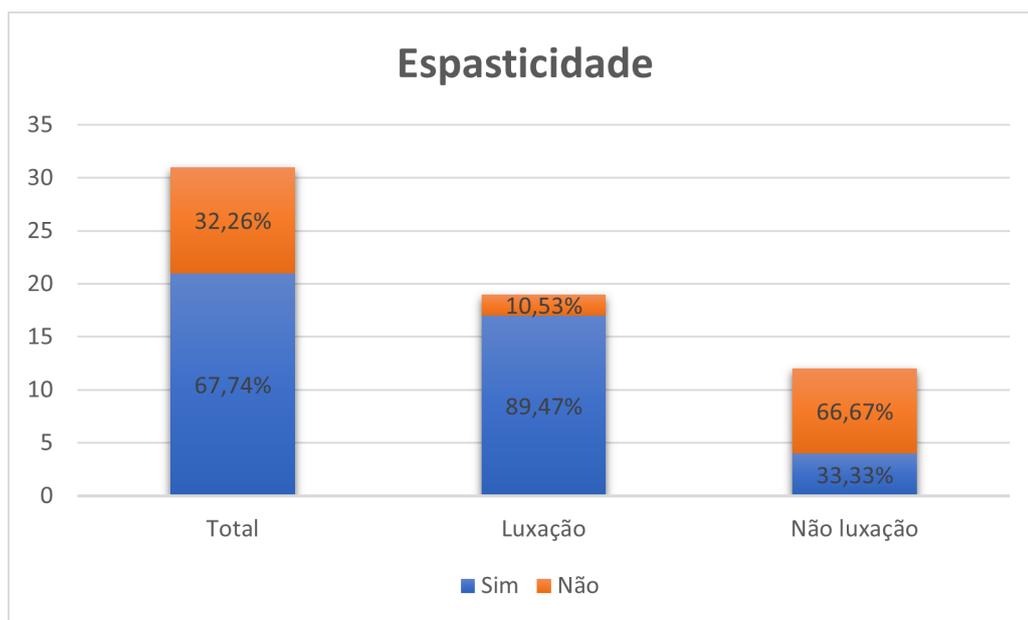


Figura 4: Distribuição da espasticidade nas crianças com luxação, sem luxação e na amostra geral.

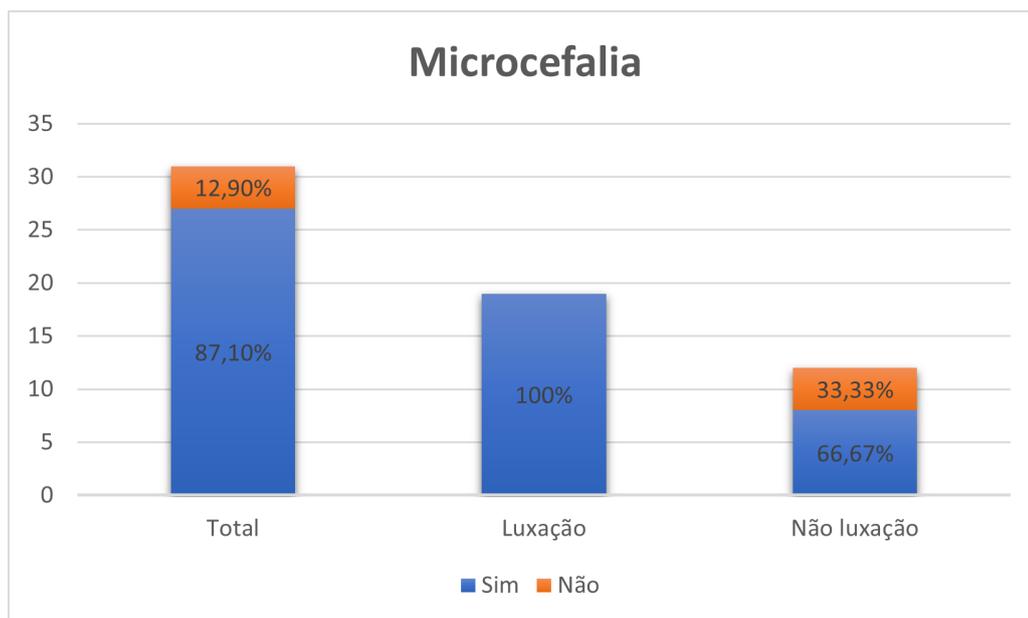


Figura 5: Distribuição da microcefalia nas crianças com luxação, sem luxação e na amostra geral.

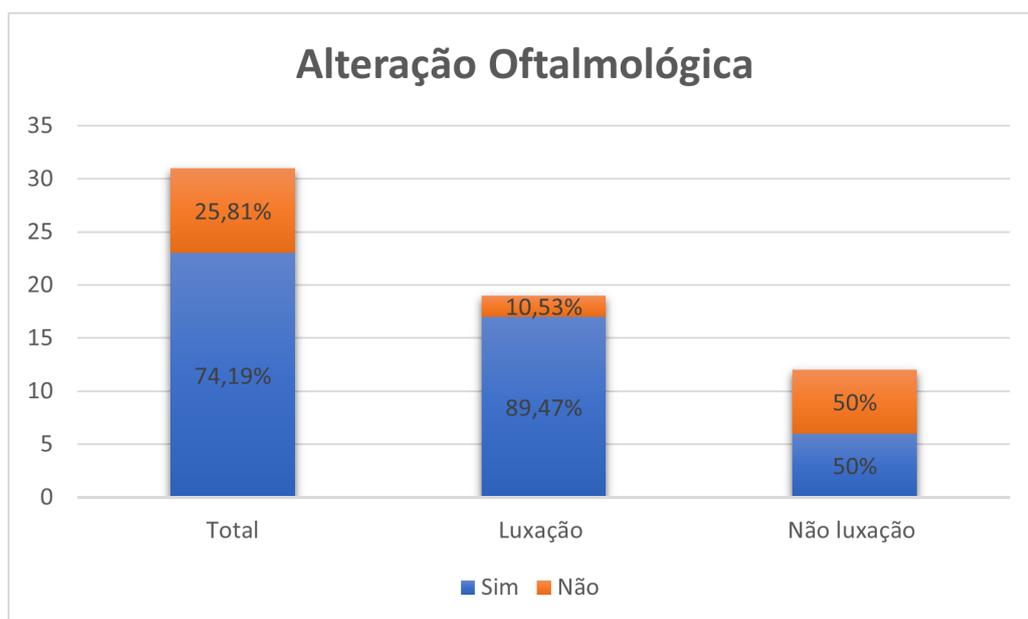


Figura 6: Distribuição da alteração oftalmológica nas crianças com luxação, sem luxação e na amostra geral.

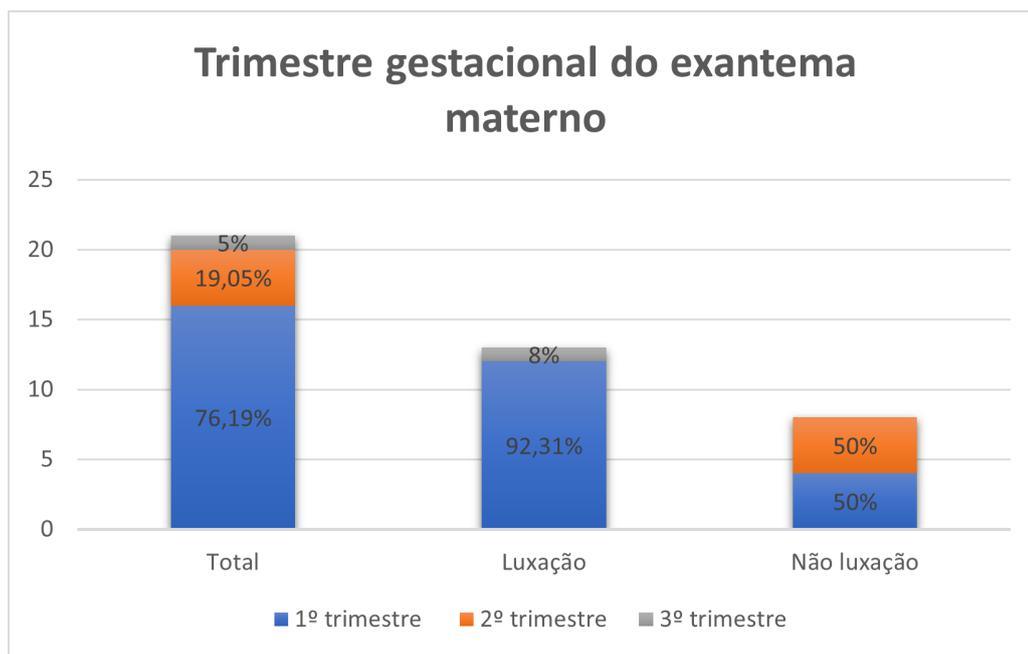


Figura 7: Distribuição do Trimestre Gestacional do Exantema Materno nas crianças com luxação, sem luxação e na amostra geral.

3.2 Análise Inferencial

Primeiro foi rodado o modelo logístico com enfoque Bayesiano de forma individual, com uma única variável regressora, a fim de analisar quais variáveis possuem associação significativa. Após a identificação dessas variáveis foi rodado o modelo múltiplo, incluindo apenas tais variáveis que em conjunto são significativas. Estes resultados estão apresentados nas Tabelas 2 e 3.

A título de exemplificação para análise dos modelos individuais, será feita uma descrição dos resultados para a variável sexo. Foram simuladas 1.000 rodadas de aquecimento e, após isso, mais 1.000 amostras foram geradas com a finalidade de se obter as distribuições a posteriori para os parâmetros de interesse. A partir da Figura 8 observa-se a convergência dos parâmetros, já que as cadeias apresentam valores dentro de um intervalo. Além disso, pode-se observar o formato da distribuição a posteriori dos parâmetros utilizando as amostras obtidas após convergência (Figura 9). Tais procedimentos foram feitos de forma análoga para todas as outras variáveis, e os resultados foram resumidos na Tabela 2.

Tabela 2: Razão de chances, intervalo de confiança Bayesiano e deviance dos modelos individuais.

Variável	Categorias	Razão de chances	IC	Deviance
Grupo	1	1	-	45,82
	2	2,17	[0,38; 15,64]	
	3	1,28	[0,23; 7,77]	
Sexo	Feminino	1	-	45,21
	Masculino	1,08	[0,18; 5,98]	
Espasticidade	Não	1	-	37,04
	Sim	22,7	[3,31; 176,09]	
Alteração Oftalmológica	Não	1	-	40,43
	Sim	9,68	[1,68; 70,18]	
Exantema materno	Não	1	-	45,21
	Sim	1,08	[0,18; 5,98]	
Estimulação	Não	1	-	44,77
	Sim	0,76	[0,08; 8,79]	
Mora comunidade	Não	1	-	43,43
	Sim	0,29	[0,06; 1,24]	

As variáveis Alteração Neurológica, Microcefalia e Trimestre Gestacional do Exantema Materno foram desconsideradas do modelo individual. No caso da variável Alteração Neurológica, foi observado que todas as crianças da base (com e sem luxação) possuem essa condição. Já no caso da Microcefalia, 100% das crianças que possuem luxação apresentam também a microcefalia, ou seja, não há observações de crianças que apresentam luxação e não apresentam microcefalia. Por fim, na variável Trimestre Gestacional 100% das crianças com luxação a mãe apresentou exantema no primeiro e terceiro trimestre, não possuindo observações de crianças com luxação cuja mãe teve exantema no segundo trimestre. Tais informações impedem que o modelo seja rodado.

Ainda observando os resultados da Tabela 2, vale ressaltar que a variável sexo não apresentou associação significativa sobre a chance de luxação de quadril, porque estatisticamente as chances são iguais para ambos os sexos (uma vez que o intervalo de confiança Bayesiano contém o valor 1). Para interpretar a razão de chance, sabendo que esta foi de 1,08, isto significa que a chance de luxação de quadril em crianças do sexo masculino é 1,08 vezes das chances de luxação em crianças do sexo feminino.

Esta mesma análise foi realizada para todos os modelos com uma variável regressora, tendo sido observada a convergência após 1.000 iterações.

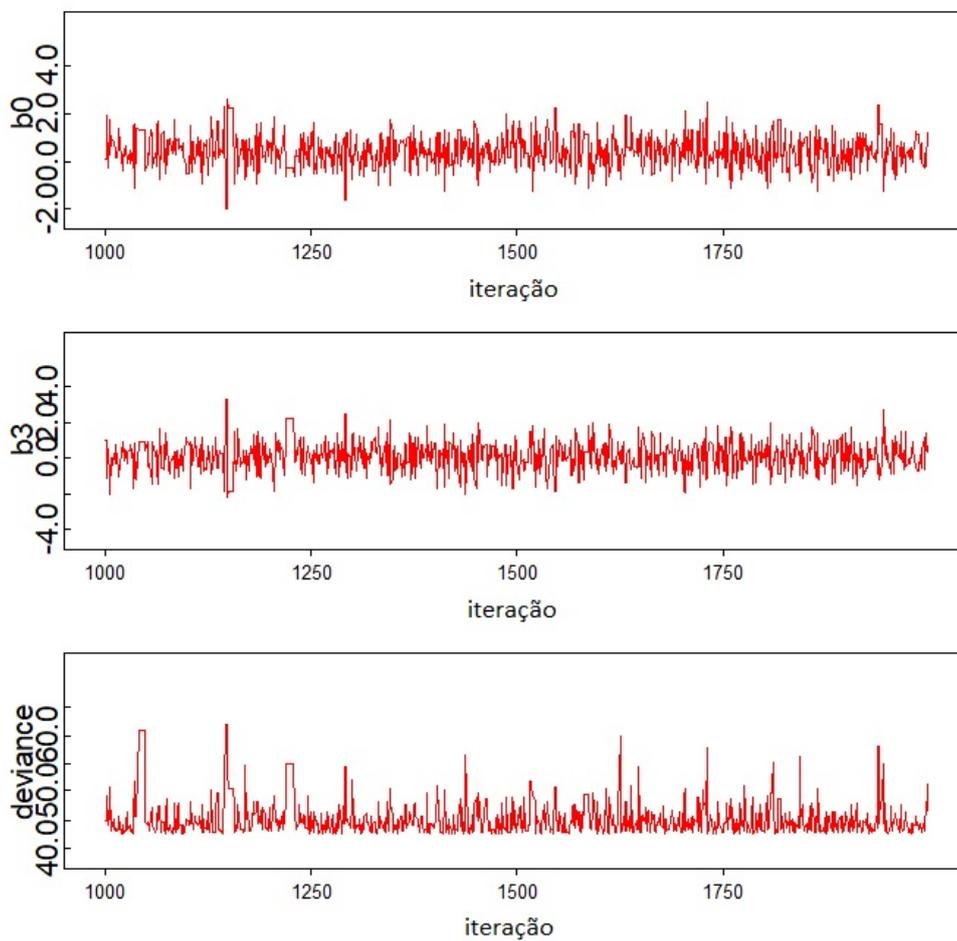


Figura 8: Traço das cadeias dos parâmetros do modelo individual e da deviance da variável Sexo.

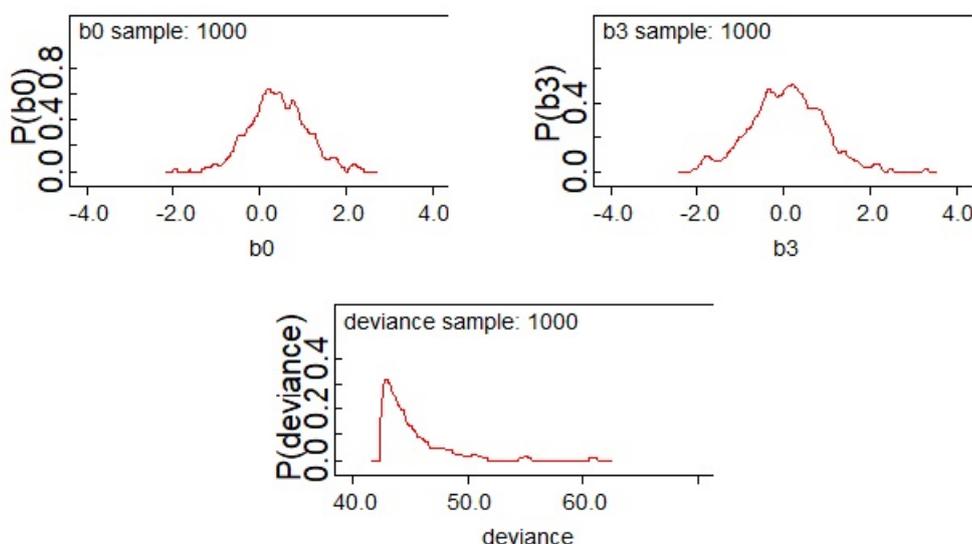


Figura 9: Distribuição a posteriori dos parâmetros do modelo individual e da deviance da variável Sexo.

Para verificar a significância da variável analisa-se o intervalo de confiança Bayesiano (IC) de 95%. Se o IC contiver o número um, significa que a variável não é significativa, já que as chances são semelhantes para as diferentes categorias da variável regressora. Sendo assim, nota-se que apenas as variáveis Espasticidade e Alteração Oftalmológica apresentaram ser significativas e, conseqüentemente, foram consideradas no modelo múltiplo que está apresentado na Tabela 3.

Tabela 3: Razão de chances, intervalo de confiança Bayesiano e deviance do modelo múltiplo.

Variável	Categorias	Razão de chances	IC	Deviance
Espasticidade	Não	1	-	38,13
	Sim	19,79	[2,36; 227,01]	
Alteração Oftalmológica	Não	1	-	
	Sim	7,35	[1,06; 127,87]	

Logo, observa-se que a chance de luxação de quadril em crianças que apresentaram espasticidade foi de quase 20 vezes das chances de luxação em crianças que não apresentaram espasticidade, considerando alteração oftalmológica constante. O intervalo de confiança Bayesiano de 95% para essa razão de chance foi de [2, 36; 227, 01]. Analogamente, tem-se que a chance de luxação de quadril em crianças que apresentaram alteração oftalmológica foi 7,35 vezes das chances de luxação em crianças que não possuem esta condição. O intervalo de confiança Bayesiano de 95% para essa razão de chance foi de [1, 06; 127, 87].

Na Figura 10 é possível observar a convergência dos parâmetros, onde foram simuladas 2.000 rodadas de aquecimento e, após isso, mais 6.000 amostras foram geradas para se obter as distribuições a posteriori para os parâmetros do modelo múltiplo. Além disso, nota-se o formato da distribuição a posteriori dos parâmetros utilizando as amostras obtidas após convergência (Figura 11).

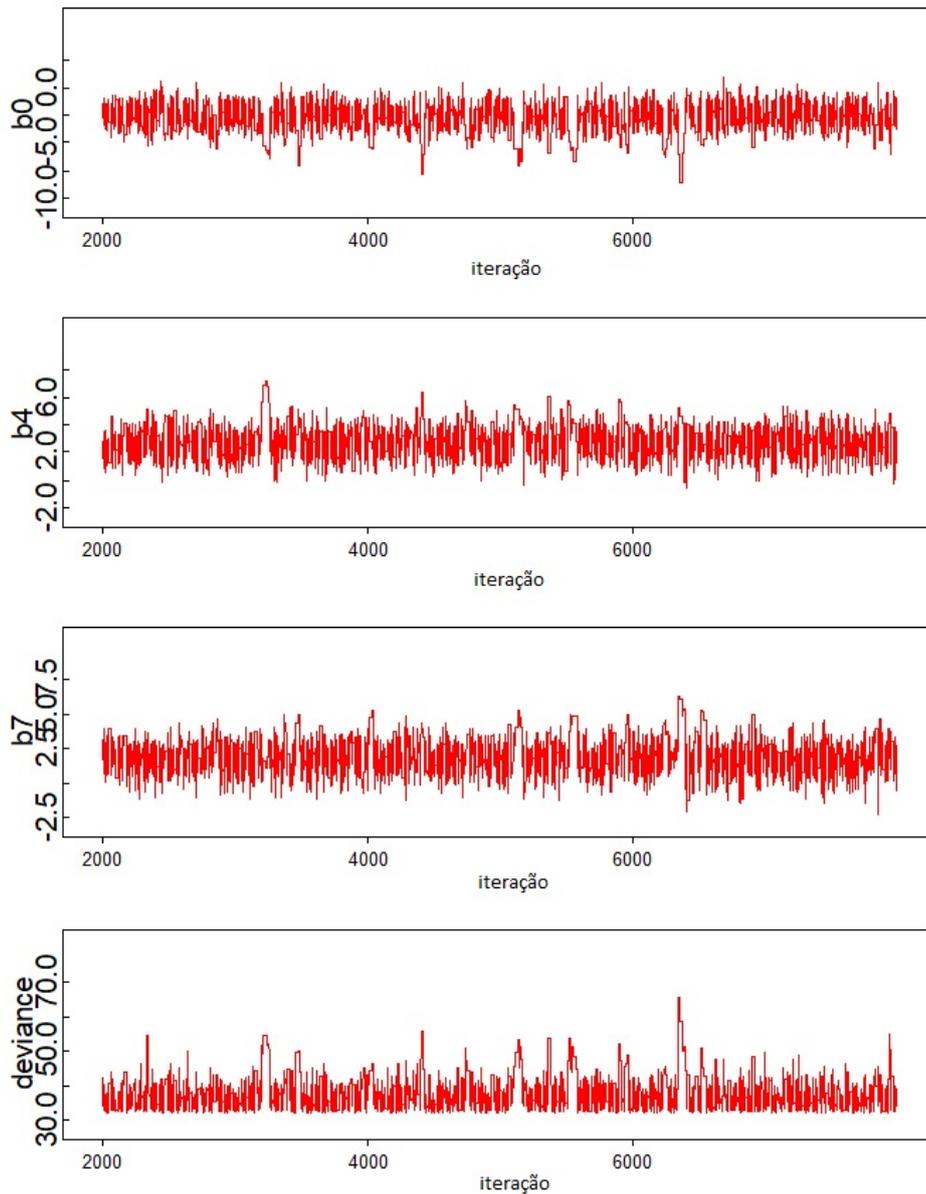


Figura 10: Traço das cadeias dos parâmetros e da *deviance* do modelo múltiplo.

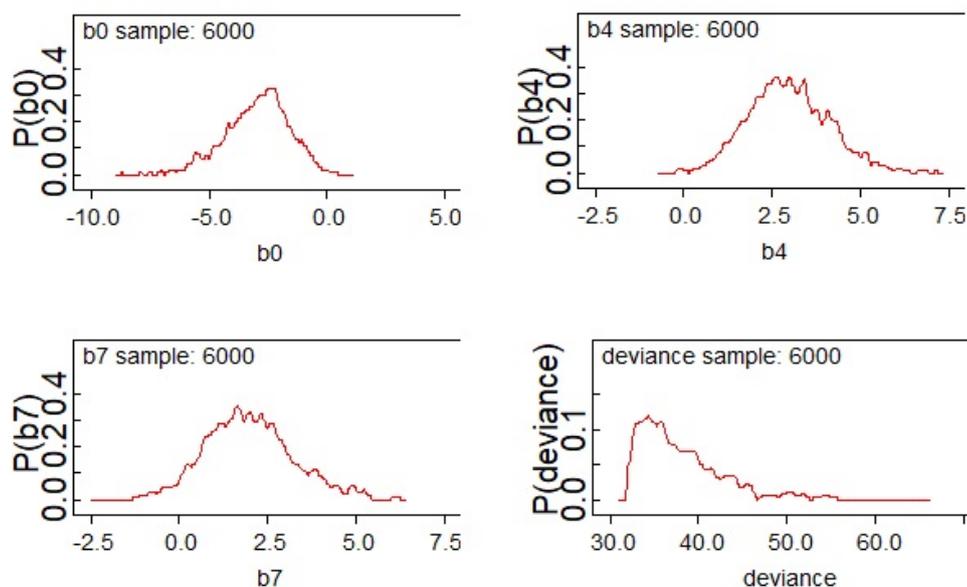


Figura 11: Distribuição a posteriori dos parâmetros e da *deviance* do modelo múltiplo.

Analisando os resultados da *Deviance* de todos os modelos propostos, os que apresentaram a menor *Deviance* foram o modelo que considera apenas a variável Espasticidade (37,04) e o modelo múltiplo, que considera a Espasticidade e Alteração Oftalmológica (38,13). A rigor, o modelo que possui o melhor ajuste é o modelo que apresenta menor *Deviance*, portanto, o modelo individual para Espasticidade. Entretanto, cabe aos médicos especialistas analisarem a importância de considerar a variável Alteração Oftalmológica, visto que o valor da *Deviance* do modelo que a considera é próximo da menor *Deviance* observada.

4 Conclusão

O objetivo do presente trabalho foi identificar os fatores associados a luxação de quadril em crianças com a síndrome da zika congênita. Nesse sentido, foram utilizados modelos de regressão logística sob enfoque Bayesiano, para que fosse possível quantificar os efeitos desses fatores sobre a condição de luxação de quadril.

Primeiramente, considerou-se cada modelo com apenas uma variável regressora. Posteriormente, foi rodado o modelo final que considerou todas as variáveis que foram significativas em conjunto. Como resultado dos modelos individuais, observou-se que as variáveis Microcefalia, Alteração Neurológica e Trimeste Gestacional do Exantema Materno retornaram erro no *software OpenBugs* e foram desconsideradas da análise. As variáveis que apresentaram significância estatística foram Espasticidade e Alteração Oftalmológica.

Sendo assim, o modelo múltiplo proposto considerou tais variáveis como sendo influentes sobre a condição de luxação de quadril, apresentando razão de chances de 19,79 para a Espasticidade e 7,35 para a Alteração Neurológica. Em outras palavras, as chances de uma criança ter luxação de quadril apresentando espasticidade são 19,79 vezes das chances de ter luxação quando a criança não apresenta espasticidade e a chance de apresentar luxação tendo observado a alteração oftalmológica é 7,35 vezes das chances em crianças que não possuem essa alteração.

Os resultados aqui obtidos podem ser considerados semelhantes aos resultados de Fonseca (2022), porém, neste trabalho, seria possível fornecer algumas informações que não poderiam ser calculadas utilizando o enfoque clássico, pelo fato de contar com uma amostra da distribuição a posteriori dos parâmetros; como, por exemplo, calcular a probabilidade do parâmetro ser positivo, ou seja, da razão de chances ser maior a 1. De qualquer forma, os intervalos de confiança apresentados foram bem largos, no caso específico da Espasticidade vale notar que, mesmo com a amplitude do intervalo sendo muito grande, o limite inferior está acima do valor 1, portanto, espera-se que a probabilidade da razão de chances ser maior a 1 seja maior que 0,95.

Como forma de comparação dos modelos propostos, realizou-se uma análise da qualidade de ajuste dos dados, através da função desvio, conhecida como *Deviance*. O provável melhor modelo é o que possui menor valor de *Deviance*, que no caso foi o modelo considerando apenas a variável regressora Espasticidade, porém o valor da *Deviance* do modelo múltiplo, que considera também a variável Alteração Oftalmológica, não apresentou discrepância em relação ao modelo individual. Neste caso, cabe análise de um especialista para julgar a importância de usar tais variáveis em conjunto.

Referências

- ARAÚJO, C. de et al. Perfil clínico e epidemiológico de pacientes com câncer atendidos em um serviço de referência na região sul de Mato Grosso. *Eng Sanit Ambient*, v. 16, n. 1, p. 175–180, 2011.
- BESERRA, R. S. et al. Modelagem com regressão logística para análise de concessão de crédito. *Research, Society and Development*, v. 11, n. 7, 2022.
- BONSEL, J. M. et al. Statistical shape modeling of us images to predict hip dysplasia development in infants. *Radiology*, v. 303, n. 2, p. 425–432, 2022.
- BOX, G. E. P.; TIAO, G. C. *Bayesian Inference in Statistical Analysis*. [S.l.]: Addison-Wesley Publishing Company, 1973.
- CAMPOS, G. S.; BANDEIRA, A. C.; SARDI, S. I. Zika Virus Outbreak, Bahia, Brazil. *Emerging Infectious Diseases*, v. 21, n. 10, p. 1885–1886, 2015. Disponível em: <https://doi.org/10.3201/eid2110.150847>.
- DOBSON, A. J.; BARNETT, A. G. *An Introduction to Generalized Linear Models*. [S.l.]: CRC Press, 2018.
- EHLERS, R. S. *Introdução a inferência bayesiana*. [S.l.], 2003. Disponível em: <http://www.leg.ufpr.br/~paulojus/CE227/ce227.pdf>.
- FONSECA, J. O. da. *Avaliação do quadril de crianças com Síndrome da Zika Congênita: Estudo prospectivo observacional*. Dissertação (Mestrado) — Ciências Médicas da Universidade Federal Fluminense, 2022.
- GABRY, J. et al. Visualization in bayesian workflow. *Journal of the Royal Statistical Society*, v. 182, n. 2, p. 389–402, 2019.
- GAMERMAN, D.; LOPES, H. *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. [S.l.]: Chapman & Hall/CRC, 2006.
- GELFAND, A. E.; GHOSH, S. K. Model choice: A minimum posterior predictive loss approach. *Oxford University Press*, v. 85, n. 1, p. 1–11, 1998.
- LEITE, R. C. M. et al. Utilização de regressão logística simples na verificação da qualidade do ar atmosférico de Uberlândia. *Eng Sanit Ambient*, v. 16, n. 1, p. 175–180, 2011.
- LINS, L. A. B.; WATKINS, C. J.; SHORE, B. J. Natural history of spastic hip disease. *J Pediatr Orthop*, v. 39, p. 33–37, 2019.
- MOORE, C. A. et al. Characterizing the Pattern of Anomalies in Congenital Zika Syndrome for Pediatric Clinicians. *JAMA Pediatrics*, v. 171, n. 3, p. 288–295, 03 2017. Disponível em: <https://doi.org/10.1001/jamapediatrics.2016.3982>.

- MURPHY, E. P. et al. Children's hip predictive (chip) score: A triage tool for hip dislocation in children referred with suspected hip dysplasia. *J Pediatr Orthop*, v. 42, n. 10, p. 552–557, 2022.
- MYERS, R. H. et al. *Generalized Linear Models: With Applications in Engineering and the Sciences*. [S.l.]: Wiley, 2010.
- OPENBUGS. Openbugs examples. 2014. Disponível em: https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol1.pdf.
- PETERSEN, L. R. et al. Zika virus. *New England Journal of Medicine*, v. 374, n. 16, p. 1552–1563, 2016. Disponível em: <https://doi.org/10.1056/NEJMra1602113>.
- TEIXEIRA, G. A. et al. Análise do conceito síndrome congênita pelo Zika vírus. *Ciência & Saúde Coletiva*, v. 25, n. 2, p. 567–574, 2020.

APÊNDICE 1 – *Script* dos modelos implementados do *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)
  b8 ~ dnorm(0,0.001)
  b9 ~ dnorm(0,0.001)
  b10 ~ dnorm(0,0.001)
  b11 ~ dnorm(0,0.001)
  b12 ~ dnorm(0,0.001)
  for(i in 1:31){
    Y[i] ~ dbern(mu[i])
    #mu[i]<- exp(b0 + b1 * grupo[i])/(1+exp(b0 + b1 * grupo[i]))
```

```
logit(mu[i]) <- b0
+ b1 * grupo2[i]
+ b2 * grupo3[i]
+ b3 * sexo[i]
+ b4 * espast[i]
+ b5 * alt.neuro[i]
+ b6 * microcef[i]
+ b7 * alt.oft[i]
+ b8 * exant.mat[i]
+ b9 * estimulacao[i]
+ b10 * mora.comum[i]
+ b11 * ig.exant.mat2[i]
+ b12 * ig.exant.mat3[i]
desvio[i] <- 2*((Y[i]+0.1)*log((Y[i]+0.1)/mu[i])+(1-Y[i]+0.1)*log((1-Y[i]+0.1)/(1-mu[i])))
}
deviance <- sum(desvio[])
}
```