

Modelo Poisson Zero-Modificado com efeito aleatório para dados longitudinais

Gabriela Cintra Raquel

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGES)

Gabriela Cintra Raquel

**Modelo Poisson Zero-Modificado com efeito aleatório para
dados longitudinais**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestra em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Marinho G. Andrade

Coorientadora: Profa. Dra. Katiane S. Conceição

**USP – São Carlos
Janeiro de 2020**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

R221m Raquel, Gabriela Cintra
Modelo Poisson Zero-Modificado com efeito
aleatório para dados longitudinais / Gabriela Cintra
Raquel; orientador Marinho Gomes Andrade;
coorientadora Katiane Silva Conceição. -- São
Carlos, 2020.
81 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2020.

1. Poisson Zero-Modificado. 2. Efeito aleatório.
3. Dados longitudinais. I. Andrade, Marinho Gomes,
orient. II. Conceição, Katiane Silva, coorient. III.
Título.

Gabriela Cintra Raquel

**Zero-Modified Poisson Model with Random Effect for
Longitudinal Data**

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC- USP and to the Department of Statistics – DEs- UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.
REVISED VERSION

Concentration Area: Statistics

Advisor: Prof. Dr. Marinho G. Andrade
Coadvisor: Profa. Dra. Katiane S. Conceição

**USP – São Carlos
January 2020**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Gabriela Cintra Raquel, realizada em 09/12/2019:



Prof. Dr. Marinho Gomes de Andrade Filho
USP



Profa. Dra. Roseli Aparecida Leandro
ESALQ/USP



Profa. Dra. Carolina Costa Mota Paraiba
UFBA

*Este trabalho é dedicado a todas as pessoas que trabalham na área de Estatística
e contribuem para o desenvolvimento dessa fascinante ciência.*

.

AGRADECIMENTOS

Agradeço imensamente ao meu orientador Marinho G. Andrade e à minha coorientadora Katiane S. Conceição pela dedicação, disponibilidade, carinho e esforço para me ajudar a realizar este trabalho.

Agradeço aos meus professores da graduação e da pós-graduação, Janete Crema, Mário de Castro, Fernando Manfio, Thais Jordão e Paulo Dattori, que sempre me incentivaram a estudar e foram grandes alicerces para mim durante a minha graduação, dividindo comigo todos os seus conhecimentos.

Aos meus amigos da graduação e da pós-graduação que sempre estiveram ao meu lado, me ajudando nas dificuldades, rindo comigo nas alegrias, compartilhando conhecimento, e me fazendo ser muito feliz longe da minha casa.

Agradeço em especial toda minha família. Aos meus pais, meus avós, minhas tias, meus tios, primos e primas. Eles sempre me apoiaram, me deram forças e acreditaram em mim. Minha mãe incessantemente me diz: "tudo vai dar certo", e meu avô João passou a vida toda dizendo para mim: "estuda, Gabriela".

Agradeço à CAPES por financiar meu projeto de pesquisa e à Superintendência de Tecnologia da Informação da Universidade de São Paulo por disponibilizar os recursos de HPC.

Agradeço também ao Instituto de Ciências Matemáticas e Computação (ICMC-USP), e à Universidade Federal de São Carlos (UFScar) por me cederem espaço físico e recursos para a realização do meu trabalho.

Por fim, agradeço a Deus. Por me capacitar e sempre ouvir minhas orações, me dar discernimento e fé para eu poder realizar meus sonhos e alcançar meus objetivos.

*“Foi o tempo que dedicaste à tua rosa
que a fez tão importante.”
(Antoine de Saint-Exupéry)*

RESUMO

RAQUEL, G. C.. **Modelo Poisson Zero-Modificado com efeito aleatório para dados longitudinais**. 2020. 82 f. Dissertação (Mestrado em em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Neste trabalho apresentamos os modelos Poisson Zero-Modificado com efeito aleatório Normal e Poisson Zero-Modificado com efeito aleatório Log-Gama Generalizado, os quais são extensões do modelo de Poisson Zero-Modificado. Sendo o efeito Log-Gama Generalizado uma generalização do efeito Normal que pode ser utilizado em situações atípicas em que o efeito Normal não é o mais adequado (ex.: dados assimétricos). O efeito aleatório induz uma correlação no modelo e acomoda a variabilidade intrínseca de cada indivíduo. Assim, os modelos nos permitem lidar com dados de contagem longitudinais, seja qual for sua quantidade de observações nulas (dados zero-inflacionados ou zero-deflacionados). Consideramos as abordagens clássica e bayesiana para estimar os parâmetros do modelo e desenvolvemos um estudo de simulação a fim de avaliar a performance dos estimadores. Com o intuito de ilustrar o procedimento proposto, analisamos um conjunto de dados reais referente a contagem de notificações de óbitos de crianças com idade entre 1 e 4 anos, nas cidades do Estado da Bahia, ao longo dos anos de 2014, 2015 e 2016. Os resultados mostraram que ambos os modelos são eficazes para modelar um conjunto de dados longitudinais sem o conhecimento preliminar da característica de inflação ou deflação de zeros existente.

Palavras-chave: Dados de Contagem, Dados Longitudinais, Dados Zero-Modificados, Dados Zero-Inflacionados, Dados Zero-Deflacionados, Efeito Aleatório, Log-Gama Generalizado..

ABSTRACT

RAQUEL, G. C.. **Modelo Poisson Zero-Modificado com efeito aleatório para dados longitudinais**. 2020. 82 f. Dissertação (Mestrado em em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação (ICM-C/USP), São Carlos – SP.

In this work we present the Zero-Modified Poisson model with Normal random effect, and the Zero-Modified Poisson model with Generalized Log-Gamma random effect, which are extensions of the Zero-Modified Poisson model. Since the Generalized Log-Gamma effect generalizes the Normal effect, it can be used in atypical situations where the Normal effect is not the most appropriate (e.g. asymmetric data). The random effect induces correlation in the model and accommodates the intrinsic variability of each individual. Thus, these models allow us to deal with longitudinal counting data, regardless of its number of null observations (zero-inflated or zero-deflated data). We consider the classical and Bayesian approaches to estimate the parameters of the model and we developed a simulation study to evaluate the performance of the estimators. In order to illustrate the proposed procedure, we analysed a set of real data regarding the count of reports of deaths of children aged 1 to 4 years, in the cities of the State of Bahia, Brazil, during the years 2014, 2015 and 2016. The results showed that both models are effective for modeling a longitudinal data set without the preliminary knowledge about the existing inflation or zero deflation characteristic.

Key-words: Counting Data, Longitudinal Data, Zero-Modified Data, Zero-Inflated Data, Zero-Deflated Data, Random Effect, Generalized Log-Gamma..

LISTA DE ILUSTRAÇÕES

Figura 1 – Perfil individual e médio das notificações de óbitos.	52
Figura 2 – Modelo ZMP com efeito aleatório Normal: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.	54
Figura 3 – Modelo ZMP com efeito aleatório Normal considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não-notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório Normal.	56
Figura 4 – Modelo ZMP com efeito aleatório LGG: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.	58
Figura 5 – Modelo ZMP com efeito aleatório LGG considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativa e limite do intervalo com 95% de confiança do parâmetro p . (b) Estimativa e limite do intervalo com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.	60
Figura 6 – Modelo ZMP com efeito aleatório Normal: (a) Estimativas e os intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson.	63

Figura 7 – Modelo ZMP com efeito aleatório Normal considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson do modelo ZMP com efeito aleatório Normal.	65
Figura 8 – Modelo ZMP com efeito aleatório LGG: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson.	66
Figura 9 – Modelo ZMP com efeito aleatório LGG considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson do modelo ZMP com efeito aleatório LGG.	68
Figura 10 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados completos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório Normal.	80
Figura 11 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados incompletos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório Normal.	81
Figura 12 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados completos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório LGG.	82
Figura 13 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados incompletos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório LGG.	82

LISTA DE TABELAS

Tabela 1	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	44
Tabela 2	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	44
Tabela 3	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	45
Tabela 4	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	45
Tabela 5	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	46
Tabela 6	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	46
Tabela 7	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	47
Tabela 8	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	47
Tabela 9	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	48
Tabela 10	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.	48
Tabela 11	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	49
Tabela 12	– Critério de avaliação do estimador de cada parâmetro do modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.	49
Tabela 13	– Distribuição de frequência das notificações de óbitos, em crianças de 1 a 4 anos de idade, das cidades da Bahia, durante os anos de 2014,2015 e 2016.	52
Tabela 14	– Número médio e desvio padrão (DP) de notificações de óbitos, em crianças de 1 a 4 anos de idade, das cidades da Bahia, em cada ano.	52
Tabela 15	– Estimativas dos parâmetros do modelo ZMP com efeito aleatório Normal , erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p considerando os dados completos de notificações de óbitos no Estado da Bahia.	53

Tabela 16 – Estimativas dos parâmetros do modelo ZMP com efeito aleatório Normal , erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana.	55
Tabela 17 – Estimativas dos parâmetros do modelo ZMP com efeito aleatório LGG , erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p considerando os dados completos de notificações de óbitos no Estado da Bahia.	57
Tabela 18 – Estimativas dos parâmetros do modelo ZMP com efeito aleatório LGG , erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana.	59
Tabela 19 – Valores do AIC e BIC calculados para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos infantis.	61
Tabela 20 – Sumário a posteriori para $\beta_0, \beta_1, \gamma_0, \gamma_1$ e σ^2 utilizando o modelo ZMP com efeito aleatório Normal e considerando os dados completos de notificações de óbitos no Estado da Bahia.	62
Tabela 21 – Sumário a posteriori para $\beta_0, \beta_1, \gamma_0, \gamma_1$ e σ^2 utilizando o modelo ZMP com efeito aleatório Normal e considerando os dados sem as notificações de óbitos referentes a Salvador e Feira de Santana.	64
Tabela 22 – Sumário a posteriori para $\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma^2$ e λ utilizando o modelo ZMP com efeito aleatório LGG e considerando os dados completos de notificações de óbitos no Estado da Bahia.	64
Tabela 23 – Sumário a posteriori para $\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma^2$ e λ utilizando o modelo ZMP com efeito aleatório LGG e considerando os dados sem as notificações de óbitos referentes a Salvador e Feira de Santana.	67
Tabela 24 – Valores do EAIC, BIC e DIC calculados para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos infantis.	69

LISTA DE ABREVIATURAS E SIGLAS

AIC	critério de informação de Akaike
BIC	critério de informação bayesiano de Schwarz
DIC	critério de informação de desvio
DP	desvio padrão
EAIC	critério de informação de Akaike extendido
EAM	erro absoluto médio
EBIC	critério de informação bayesiano de Schwarz extendido
EP	erros padrão
EQM	erro quadrático médio
GR	Gelman Rubin
LGG	Log-Gama Generalizada
MCMC	...	Monte Carlo em cadeia de Markov
PC	probabilidades de cobertura
PDF	função densidade de probabilidade
PDF	função densidade de probabilidade
PMF	função massa de probabilidade
PMF	função massa de probabilidade
Var	variância
ZDP	Poisson Zero-Deflacionada
ZIP	Poisson Zero-Inflacionada
ZMP	Poisson Zero-Modificada
ZMPS	Série de Potências Zero-Modificadas
ZTP	Poisson Zero-Truncada

SUMÁRIO

1	INTRODUÇÃO	21
2	DISTRIBUIÇÃO POISSON ZERO-MODIFICADA	25
2.1	Função Massa de Probabilidade	25
2.1.1	<i>Particularidades da Distribuição ZMP</i>	26
2.1.2	<i>Média e Variância</i>	26
2.1.3	<i>Índice de dispersão</i>	27
2.2	A Versão Hurdle da Distribuição ZMP	27
2.2.1	<i>Particularidades da Distribuição ZMP na Versão Hurdle</i>	28
2.2.2	<i>Média e Variância da Distribuição ZMP na Versão Hurdle</i>	29
3	MODELO ZMP COM EFEITO ALEATÓRIO PARA DADOS LONGITUDINAIS	31
3.1	Estrutura do Modelo ZMP com Efeito Aleatório	31
3.2	Distribuição Log-Gama Generalizada	32
3.3	Modelo ZMP com Intercepto Aleatório Normal e Log-Gama Generalizado	32
3.3.1	<i>Características do Modelo ZMP com Intercepto Aleatório Normal e Log-Gama Generalizado</i>	33
4	MÉTODOS DE ESTIMAÇÃO	35
4.1	Estimação por Máxima Verossimilhança Marginal	35
4.1.1	<i>Intervalos de Confiança</i>	37
4.1.2	<i>Teste Razão de Verossimilhança para o Efeito Aleatório</i>	38
4.2	Abordagem Bayesiana	38
4.3	Análise de Diagnóstico	39
4.3.1	<i>Resíduo de Pearson</i>	39
4.3.2	<i>Gráfico Envelope</i>	39
4.4	CrITÉRIOS para Seleção de Modelos	40
5	ESTUDO DE SIMULAÇÃO	41
5.1	Cenários da Simulação	41
5.2	Gerador de Variáveis Aleatórias ZMP Correlacionadas no Tempo	42
5.3	CrITÉRIOS de Avaliação dos Estimadores	42

5.4	Resultados com a Abordagem Clássica	43
5.5	Resultados com a Abordagem Bayesiana	46
6	APLICAÇÃO	51
6.1	Análise dos Dados de Notificações de Óbitos Infantis	51
6.2	Análise com a Abordagem Clássica	53
6.2.1	<i>Modelo ZMP com Efeito Aleatório Normal</i>	53
6.2.2	<i>Modelo ZMP com Efeito Aleatório LGG</i>	57
6.3	Análise com a Abordagem Bayesiana	61
6.3.1	<i>Modelo ZMP com Efeito Aleatório Normal</i>	61
6.3.2	<i>Modelo ZMP com Efeito Aleatório LGG</i>	64
7	CONCLUSÃO	71
	REFERÊNCIAS	73
	APÊNDICE A CÁLCULO DAS DERIVADAS	77
A.1	Parâmetro p	77
A.2	Parâmetro μ_{ZMP}	78
A.3	Parâmetro p_0	78
	APÊNDICE B AMOSTRAS DA DENSIDADE A POSTERIORI CON- JUNTA	79

INTRODUÇÃO

Dados de contagem aparecem em vários problemas práticos das áreas da saúde, economia, educação e ciências sociais, entre outras. Alguns exemplos podem ser encontrados em [Frome e Checkoway \(1985\)](#), [Joshua e Garber \(1990\)](#), [Madanat e Ibrahim \(1995\)](#) e [Gardner, Mulvey e Shaw \(1995\)](#). A distribuição de Poisson é uma das mais usadas para modelar dados de contagem, em que, usualmente, as observações são consideradas independentes ([FROME, 1983](#)).

Muitos dados de contagem apresentam características que são inerentes ao fenômeno que produzem os dados. Dentre estas, uma muito comum é o excesso de zeros (dados inflacionados de zero), e com menos ocorrência, mas igualmente importante, é o *deficit* de zeros (dados deflacionados de zero). Estes excessos ou *deficit* de zeros são sempre avaliados com referência a alguma distribuição de probabilidade, que supostamente seria adequada para modelagem dos dados de contagem.

Quando a frequência de zeros em um conjunto de dados é muito maior (ou menor) do que a frequência esperada segundo uma distribuição de contagem padrão, dizemos que tal conjunto de dados é zero-inflacionado (ou deflacionado). Neste caso, pressupõe-se haver mais de um processo gerador atuando na geração dos dados. Deste modo, a ocorrência de observações zeros pode ser caracterizada como zeros amostrais e zeros estruturais. Os zeros amostrais sucedem a partir de um processo gerador de contagens (ex.: Processo Poisson), enquanto os zeros estruturais dão-se pela ausência de determinada característica da população.

A solução encontrada por [Cohen \(1960\)](#) e [Umbach \(1981\)](#) para lidar com tal situação foi considerar uma mistura entre a distribuição de Poisson e uma distribuição degenerada em zero, introduzindo assim, a distribuição Poisson Zero-Inflacionada. Mais tarde, [Lambert \(1992\)](#) apresentou o modelo de regressão para dados com distribuição Poisson Zero-Inflacionada. O modelo de Poisson Zero-Modificado foi introduzido por [Dietz e Böhning \(2000\)](#), sendo capaz de modelar conjuntos de dados sem informação prévia de inflação ou deflação nos dados. Mais recentemente, [Conceição, Andrade e Louzada \(2013\)](#) propuseram uma abordagem

bayesiana para a inferência do modelo de Poisson Zero-Modificado. Um caso mais geral, o qual inclui distribuições da família Série de Potências (CORDEIRO; ANDRADE; CASTRO, 2009), foi amplamente estudado por Conceição (2013) no contexto de regressão. Tais modelos são denominados modelos Série de Potências Zero-Modificados.

Dados longitudinais são medidas repetidas feitas nas mesmas unidades experimentais (indivíduo ou região) ao longo do tempo (anos, meses, dias, horas). Potencialmente existe correlação entre essas medidas, uma vez que elas são tomadas em uma mesma unidade experimental. A partir desses dados podemos observar o comportamento de cada indivíduo ao longo do tempo.

Como exemplo de dados longitudinais, considere analisar o número de convulsões sofridas por pacientes com epilepsia durante cinco semanas (FABIO; PAULA; CASTRO, 2012, p. 1504). Cada paciente observado é uma unidade experimental, o tempo são as semanas, e a medida de interesse é o número de convulsões. Ou ainda, analisar o número de notificações de uma doença em determinado estado durante quatro anos (SEI, 2018). Cada município observado é uma unidade experimental, o tempo são os anos, e a medida de interesse é o número de notificações da doença estudada. Espera-se que as notificações obtidas na mesma região sejam correlacionadas.

A estrutura dos dados longitudinais pode ser caracterizada como regular, balanceada e completa (HAUSER, 2009). É regular, se durante todo o estudo, o período entre dois tempos subsequentes quaisquer é constante. É caracterizada como balanceada, se em cada unidade existe o mesmo número de observações e elas são feitas nos mesmos instantes de tempo. Por fim, é completa, se não apresentar observações perdidas. Na análise de dados longitudinais, é importante que essas características sejam levadas em consideração.

De acordo com Singer e Andrade (1986), uma das desvantagens de analisar dados longitudinais é o custo necessário para coletar os dados nos instantes pré-determinados, e além disso, o longo período de duração.

O modelo de regressão Poisson clássico não é adequado para modelar dados de contagem longitudinais, uma vez que os mesmos possuem dependência temporal. Uma forma de se levar em consideração a dependência entre esses dados é a inclusão de variáveis latentes no preditor linear, o que induz a correlação entre os indivíduos da mesma unidade experimental. Essa extensão é conhecida como modelos lineares mistos.

O modelo linear misto é composto por parâmetros de efeitos fixos, β , e parâmetros de efeitos aleatórios \mathbf{b} . Os efeitos fixos são compartilhados por todas as unidades experimentais, enquanto os efeitos aleatórios são específicos de cada uma (MCCULLOCH; NEUHAUS, 2001). Em geral assume-se que \mathbf{b} segue uma distribuição Normal com média zero e matriz de covariância \mathbf{G} , $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$, entretanto, distribuições alternativas podem ser usadas para definir o efeito aleatório. Hardin e Hilbe (2002), por exemplo, assumem a distribuição log-gama para o intercepto aleatório. Fabio, Paula e Castro (2012) consideram a distribuição Log-Gama Generalizada para a

variável latente.

Os modelos lineares mistos vêm sendo estudados há um longo tempo na literatura Estatística. [Searle *et al.* \(1974\)](#) propuseram o método de máxima verossimilhança para estimar os efeitos fixos e fazer previsões dos efeitos aleatórios de um modelo misto. Posteriormente, [Laird e Ware \(1982\)](#) apresentaram os modelos lineares mistos para dados longitudinais, os quais levam em consideração a dependência entre as observações na mesma unidade.

Modelos de regressão com efeito aleatório têm sido pesquisados por diversos autores: [Breslow e Clayton \(1993\)](#), [Costa \(2017\)](#), [Carvalho e Achcar \(2013\)](#) e [Trindade \(2014\)](#). Uma extensão da classe dos modelos lineares mistos são os modelos lineares generalizados mistos, abordado por [McCulloch e Neuhaus \(2001\)](#), sendo que se considera apenas as distribuições da família exponencial.

Considere a situação em que precisamos analisar um conjunto de dados longitudinais que, aparentemente, segue a distribuição de Poisson. Entretanto, sua frequência de zeros é maior, ou menor do que a frequência esperada segundo a distribuição de Poisson. Neste caso, precisamos de um modelo que seja adequado para lidar com a correlação que pode existir entre os dados que pertencem à mesma unidade experimental, e que, ao mesmo tempo, trate o problema de inflação ou deflação de zeros na amostra.

Assim, o principal objetivo desta dissertação é estender o modelo de regressão Poisson Zero-Modificado para um modelo a fim de tratar o excesso e o *deficit* de zeros em dados dependentes ao longo do tempo. Neste trabalho, similarmente ao texto de [Fabio, Paula e Castro \(2012\)](#), os modelos propostos para modelar dados com tais características são o Poisson Zero-Modificado com efeito aleatório Normal e o Poisson Zero-Modificado com efeito aleatório Log-Gama Generalizado. A distribuição Log-Gama Generalizada ([PRENTICE, 1974](#)) tem como caso particular a distribuição Normal e pode apresentar assimetrias à esquerda ou à direita. Sendo assim, o efeito aleatório com distribuição Log-Gama Generalizada pode ser utilizado em situações atípicas (ex.: dados assimétricos) em que o efeito Normal não é o mais adequado. De fato, eles modelam os dados sem o conhecimento prévio da característica de inflação ou deflação de zeros, e além disso, modelam a variabilidade intrínseca de cada indivíduo.

Este trabalho está organizado da seguinte maneira. No [Capítulo 2](#) é apresentada a distribuição Poisson Zero-Modificada e suas particularidades; no [Capítulo 3](#) definimos os modelos Poisson Zero-Modificado com efeito aleatório Normal e Poisson Zero-Modificado com efeito aleatório Log-Gama Generalizado para dados longitudinais e algumas de suas características; no [Capítulo 4](#) apresentamos uma abordagem clássica e bayesiana para estimar os parâmetros do modelo, exibimos o procedimento para o cálculo dos intervalos de confiança assintóticos e informações sobre a análise de diagnóstico para os modelos; no [Capítulo 5](#) realizamos um estudo de simulação a fim de avaliar o processo de estimação dos parâmetros; no [Capítulo 6](#) apresentamos uma aplicação utilizando um conjunto de dados reais; e por fim, no [Capítulo 7](#) apresentamos as conclusões do presente trabalho.

DISTRIBUIÇÃO POISSON ZERO-MODIFICADA

Neste capítulo apresentamos a distribuição de Poisson Zero-Modificada (ZMP), a qual pertence à família de distribuições Série de Potências Zero-Modificadas (ZMPS), abordadas em [Conceição *et al.* \(2017\)](#). Além disso, exibimos sua versão Hurdle, em que a distribuição ZMP é escrita como uma distribuição de mistura, e ressaltamos algumas de suas características.

2.1 Função Massa de Probabilidade

Considere a distribuição de Poisson com média desconhecida $\mu > 0$, cuja função massa de probabilidade (PMF) é dada por

$$\pi_P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, \dots$$

A distribuição de Poisson Zero-Truncada (ZTP), adequada em situações em que nenhum zero é observado, é dada por

$$\pi_{ZTP}(y; \mu) = \frac{\pi_P(y; \mu)}{1 - \pi_P(0; \mu)} = \frac{e^{-\mu} \mu^y}{y!(1 - e^{-\mu})}, \quad y = 1, 2, \dots \quad (2.1)$$

Uma variável aleatória Y segue a distribuição ZMP ([CONCEIÇÃO *et al.*, 2017](#)) se sua PMF é dada por

$$\pi_{ZMP}(y; \mu, p) = (1 - p)I(y) + p\pi_P(y; \mu), \quad y = 0, 1, \dots \quad (2.2)$$

em que $I(y)$ é a função indicadora na qual

$$I(y) = \begin{cases} 1, & \text{se } y = 0 \\ 0, & \text{se } y > 0, \end{cases}$$

$\mu > 0$ é a média da distribuição de Poisson tradicional, e p é o parâmetro de modificação da probabilidade de zeros com referência a distribuição de Poisson tradicional, tal que

$$0 \leq p \leq \frac{1}{1 - \pi_P(0; \mu)}.$$

A função dada em (2.2) é chamada distribuição ZMP e é denotada por $ZMP(\mu, p)$.

2.1.1 Particularidades da Distribuição ZMP

A partir da avaliação da proporção de zeros adicionais ou faltantes, podemos observar que a frequência de zeros da distribuição $ZMP(\mu, p)$ depende dos valores atribuídos a p :

$$\begin{aligned} \pi_{ZMP}(0; \mu, p) - \pi_P(0; \mu) &= 1 - p + p\pi_P(0; \mu) - \pi_P(0; \mu) \\ &= (1 - p)[1 - \pi_P(0; \mu)]. \end{aligned} \quad (2.3)$$

De acordo com (2.3) temos:

- Se $p = 0$ em (2.3), então $\pi_{ZMP}(0; \mu, p) = 1$. Assim, a expressão em (2.2) é a distribuição Poisson degenerada com toda massa em zero.
- Quando $0 < p < 1$ em (2.3), então $(1 - p)[1 - \pi_P(0; \mu)] > 0$. Portanto $\pi_{ZMP}(0; \mu, p) > \pi_P(0; \mu, \phi)$, e a expressão em (2.2) é definida como a distribuição Poisson Zero-Inflacionada (ZIP), que tem uma proporção extra de zeros.
- Se $p = 1$ em (2.3), então $\pi_{ZMP}(0; \mu, p) - \pi_P(0; \mu) = 0$. Assim, $\pi_{ZMP}(0; \mu, p) = \pi_P(0; \mu)$, e a expressão (2.2) é definida como a distribuição de Poisson tradicional, sem modificação.
- Quando $1 < p < \frac{1}{1 - \pi_P(0; \mu)}$ em (2.3), então $(1 - p)[1 - \pi_P(0; \mu)] < 0$. Logo, $\pi_{ZMP}(0; \mu, p) < \pi_P(0; \mu)$, e a expressão (2.2) é definida como a distribuição Poisson Zero-Deflacionada (ZDP), que tem uma proporção de zeros faltantes.
- Se $p = \frac{1}{1 - \pi_P(0; \mu)}$ em (2.3), então $\pi_{ZMP}(0; \mu, p) = 0$. Assim, a equação (2.2) é definida como a distribuição ZTP.

2.1.2 Média e Variância

A média e a variância de uma variável aleatória Y com distribuição $ZMP(\mu, p)$ são dadas respectivamente por

$$\mu_{ZMP} = p\mu \text{ e } \sigma_{ZMP}^2 = p[\mu + (1 - p)\mu^2], \quad (2.4)$$

em que μ é a média da distribuição Poisson associada a Y .

2.1.3 Índice de dispersão

O índice de dispersão de uma variável aleatória Y é dado por $I(Y) = V(Y)/E(Y)$. Quando esse índice é igual a 1 dizemos que os dados são equidispersos, quando o índice é maior ou menor do que 1 dizemos que os dados são, respectivamente sobredispersos e subdispersos. A distribuição de Poisson admite índice de dispersão maior, igual ou menor do que 1. Logo, tal distribuição é adequada para ajustar dados zero-inflacionados ou deflacionados, os quais são sobredispersos ou subdispersos, respectivamente. O índice de dispersão (CONCEIÇÃO *et al.*, 2017) para dados de uma distribuição ZMP é dado por

$$I_{ZMP} = \frac{\sigma_{ZMP}^2}{\mu_{ZMP}} = \frac{\sigma^2}{\mu} + \mu(1-p) = I_p + \mu(1-p), \quad (2.5)$$

em que I_p é o índice de dispersão referente a distribuição de Poisson padrão, e o termo $\mu(1-p)$ representa a sobredispersão ou a subdispersão causada pela modificação na frequência dos zeros. Conforme é exibido a seguir, existe uma relação entre o índice de dispersão e as características de inflação ou deflação de zeros nos dados:

- Distribuição de Poisson zero-inflacionada: $0 < p < 1, I_{ZMP} > I_p$;
- Distribuição de Poisson padrão: $p = 1, I_{ZMP} = I_p$;
- Distribuição de Poisson zero-deflacionada: $1 < p < (1 - \pi_p(0; \mu))^{-1}, I_{ZMP} < I_p$.

Portanto, a sobredispersão e a subdispersão são fortemente causadas pela inflação e deflação de zeros.

2.2 A Versão Hurdle da Distribuição ZMP

A PMF da distribuição ZMP, dada em (2.2), pode ser escrita como uma distribuição Hurdle, ou seja, uma distribuição de mistura, dada por

$$\pi_{ZMP}(y; \mu, \omega) = (1 - \omega)I(y) + \omega\pi_{ZTP}(y; \mu), \quad y = 0, 1, \dots, \quad (2.6)$$

em que $0 \leq \omega \leq 1$ e $\pi_{ZTP}(y; \mu)$ é a distribuição ZTP definida em (2.1).

A partir de (2.2) podemos deduzir (2.6):

(i) Para $y = 0$, $\pi_{ZMP}(0; \mu, p) = 1 - p + p\pi_p(0; \mu)$;

(ii) Para todo $y > 0$, $\pi_{ZMP}(y; \mu, p) = p\pi_p(y; \mu)$.

Assim, com base em (i) e (ii), a expressão de $\pi_{ZMP}(y; \mu, p)$ pode ser construída por meio da função indicadora $I(y)$:

$$\begin{aligned}\pi_{ZMP}(y; \mu, p) &= \pi_{ZMP}(0; \mu, p)I(y) + \pi_{ZMP}(y; \mu, p)[1 - I(y)] \\ &= [1 - p + p\pi_P(0; \mu)]I(y) + [p\pi_P(y; \mu)][1 - I(y)] \\ &= \{1 - p[1 - \pi_P(0; \mu)]\}I(y) \\ &\quad + p[1 - \pi_P(0; \mu)] \left\{ \frac{\pi_P(y; \mu)}{1 - \pi_P(0; \mu)} [1 - I(y)] \right\}.\end{aligned}\tag{2.7}$$

Deste modo, a função massa de probabilidade em (2.2) pode ser reescrita usando a distribuição ZTP:

$$\pi_{ZMP}(y; \mu, p) = \{1 - p(1 - \pi_P(0; \mu))\}I(y) + p(1 - \pi_P(0; \mu))\pi_{ZTP}(y; \mu), \quad y = 0, 1, \dots$$

Dado que $0 \leq p \leq \frac{1}{1 - \pi_P(0; \mu)}$, então $0 \leq p(1 - \pi_P(0; \mu)) \leq 1$. Denotando $\omega = p(1 - \pi_P(0; \mu))$, temos

$$\pi_{ZMP}(y; \mu, \omega) = (1 - \omega)I(y) + \omega\pi_{ZTP}(y; \mu).$$

Observação 1. É importante ressaltar que os parâmetros ω e p estão relacionados pela seguinte equação:

$$p = \frac{\omega}{1 - e^{-\mu}}.\tag{2.8}$$

A versão Hurdle da distribuição ZMP é denotada por $ZMP(\mu, \omega)$.

A distribuição ZMP apresentada em (2.6) pode ser interpretada como uma distribuição de mistura. Podemos observar que na versão Hurdle, a probabilidade de $y = 0$ é $1 - \omega$, enquanto a probabilidade de $y > 0$ é $\omega\pi_{ZTP}(y; \mu)$. Ou seja, uma parte produz apenas observações positivas vindas de uma distribuição ZTP, enquanto a outra parte produz apenas observações de valor zero.

2.2.1 Particularidades da Distribuição ZMP na Versão Hurdle

Para diferentes valores de ω obtemos distribuições $ZMP(\mu, \omega)$ com diferentes frequências de zero. Uma vez que $\omega = p(1 - \pi_P(0; \mu))$, podemos reescrever a equação (2.3) como:

$$\pi_{ZMP}(0; \mu, \omega) - \pi_P(0; \mu) = 1 - \omega - \pi_P(0; \mu).\tag{2.9}$$

O parâmetro ω controla a frequência de zeros da seguinte forma:

- Se $\omega = 0$ em (2.9), então $\pi_{ZMP}(0; \mu, \omega) = 1$. Assim, a expressão (2.6) é definida como a distribuição de Poisson degenerada com toda massa em zero.
- Quando $0 < \omega < 1 - \pi_P(0; \mu)$ em (2.9), então $1 - \omega - \pi_P(0; \mu) > 0$. Logo, $\pi_{ZMP}(0; \mu, \omega) > \pi_P(0; \mu)$, e a expressão (2.6) é definida como a distribuição ZIP.

- Se $\omega = 1 - \pi_P(0; \mu)$ em (2.9), então $\pi_{ZMP}(0; \mu, \omega) - \pi_P(0; \mu) = 0$. Assim, $\pi_{ZMP}(0; \mu, \omega) = \pi_P(0; \mu)$, e a expressão (2.6) é definida como a distribuição Poisson usual.
- Quando $1 - \pi_P(0; \mu) < \omega < 1$ em (2.9), então $1 - \omega - \pi_P(0; \mu) < 0$. Então, $\pi_{ZMP}(0; \mu, \omega) < \pi_P(0; \mu)$, e a expressão (2.6) é definida como a distribuição ZDP.
- Se $\omega = 1$ em (2.9), então $\pi_{ZMP}(0; \mu, \omega) = 0$. Assim, a equação (2.6) é definida como a distribuição ZTP.

2.2.2 Média e Variância da Distribuição ZMP na Versão Hurdle

A média e a variância de uma variável aleatória Y , com distribuição $ZMP(\mu, \omega)$, são dadas por (CONCEIÇÃO *et al.*, 2017)

$$\mu_{ZMP} = \frac{\mu \omega}{1 - \pi_P(0; \mu)} \quad \text{e} \quad \sigma_{ZMP}^2 = \frac{\omega(\sigma^2 + \mu^2)}{1 - \pi_P(0; \mu)} - \left[\frac{\mu \omega}{1 - \pi_P(0; \mu)} \right]^2,$$

em que μ e σ^2 são, respectivamente, a média e a variância da distribuição Poisson associada a Y .

MODELO ZMP COM EFEITO ALEATÓRIO PARA DADOS LONGITUDINAIS

Dados longitudinais são medidas efetuadas nas mesmas unidades amostrais ao longo do tempo. A possível correlação entre as medidas tomadas na mesma unidade amostral deve ser levada em consideração na análise de dados longitudinais. Uma forma de se modelar dados com tais características, é a inclusão de um efeito aleatório no preditor linear. Tal efeito é responsável por induzir uma correlação entre os dados e modelar a variabilidade intrínseca de cada indivíduo.

Neste capítulo, apresentamos dois modelos: o modelo ZMP com intercepto aleatório Normal e o modelo ZMP com intercepto aleatório Log-Gama Generalizado. Além disso, exibimos algumas de suas características.

3.1 Estrutura do Modelo ZMP com Efeito Aleatório

A estrutura em (3.1) descreve um modelo de regressão ZMP com efeitos aleatórios para dados longitudinais. Considerando a versão Hurdle da distribuição ZMP dada em (2.6), seja Y_{ij} uma variável aleatória assumindo independência condicional no vetor de efeitos aleatórios \mathbf{b} , em que $i = 1, \dots, N$ são unidades experimentais no tempo $j = 1, \dots, n_i$ (medidas por unidade). Suponha $Y_{ij}|\mathbf{b}$ variáveis aleatórias independentes, com PMF pertencente à família ZMP, tal que

$$Y_{ij}|\mathbf{b} \sim ZMP(y_{ij}|\mathbf{b}; \mu_{ij}, \omega_{ij}),$$

e

$$\log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}, \quad \log\left(\frac{\omega_{ij}}{1 - \omega_{ij}}\right) = \mathbf{w}'_{ij}\boldsymbol{\gamma}, \quad (3.1)$$

em que $\mathbf{x}_{ij} = (1, x_{ij1}, \dots, x_{ijp})'$ e $\mathbf{w}_{ij} = (1, w_{ij1}, \dots, w_{ijr})'$ são vetores de covariáveis relacionados aos vetores de parâmetros dos efeitos fixos $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ e $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_r)'$, respectivamente, e $\mathbf{z}_{ij} = (1, z_{ij1}, \dots, z_{ijq})'$ é o vetor de covariáveis relacionado ao vetor de

efeitos aleatórios $\mathbf{b} = (b_0, b_1, \dots, b_q)'$. Além do mais, consideramos as funções de ligação $h_1(\mu_{ij}) = \log(\mu_{ij})$ e $h_2(\omega_{ij}) = \log\left(\frac{\omega_{ij}}{1-\omega_{ij}}\right)$, que são diferenciáveis e relacionam, respectivamente, μ_{ij} e ω_{ij} aos seus preditores lineares com parâmetros β , \mathbf{b} e γ , os quais devem ser estimados. Para finalizar a descrição do modelo, associamos ao vetor de efeitos aleatórios \mathbf{b} uma função densidade de probabilidade π_B .

3.2 Distribuição Log-Gama Generalizada

Seja S uma variável aleatória com distribuição Log-Gama Generalizada (LGG), $S \sim LGG(\mu, \sigma, \lambda)$. Assim, sua função densidade de probabilidade (PDF) (LAWLESS, 2011) é dada por

$$\pi_{LGG}(s; \mu, \sigma, \lambda) = \begin{cases} \frac{c(\lambda)}{\sigma} \exp\left[\frac{(s-\mu)}{\lambda\sigma} - \frac{1}{\lambda^2} \exp\left\{\frac{\lambda(s-\mu)}{\sigma}\right\}\right], & \text{se } \lambda \neq 0 \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\}, & \text{se } \lambda = 0, \end{cases}$$

em que $s \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$, e $\lambda \in \mathbb{R}$, correspondem aos parâmetros de localização, escala e forma, respectivamente. Com $c(\lambda) = \frac{|\lambda|}{\Gamma(\lambda^{-1})}(\lambda^{-2})^{\lambda^{-2}}$, em que $\Gamma(\cdot)$ é a função gama. Para $\lambda < 0$, a PDF possui uma assimetria à direita, para $\lambda > 0$, a PDF possui uma assimetria à esquerda, e para $\lambda = 0$ a PDF é de uma distribuição Normal. Para $\lambda \neq 0$, a esperança e a variância de S são dadas por

$$E(S) = \mu + \sigma \left\{ \frac{\psi(\lambda^{-2}) + \log \lambda^{-2}}{|\lambda|} \right\} \text{ e } Var(S) = \frac{\sigma^2 \Psi'(\lambda^{-2})}{\lambda^2},$$

em que $\Psi(\cdot)$ e $\Psi'(\cdot)$ são as funções digama e trigama, respectivamente. Para $\lambda = 0$, $E(S) = \mu$ e $Var(S) = \sigma^2$.

3.3 Modelo ZMP com Intercepto Aleatório Normal e Log-Gama Generalizado

Neste trabalho apresentamos o modelo ZMP com Intercepto Aleatório Normal e o modelo ZMP com Intercepto Aleatório Log-Gama Generalizado, aos quais atribuímos aos efeitos aleatórios as distribuições $N(0, \sigma^2)$ e $LGG(0, \sigma, \lambda = -\sigma)$, respectivamente. Visto que a distribuição de dados zero-inflacionados e zero-deflacionados possui uma assimetria maior à direita, e tomando um caso particular apresentado em Fabio, Paula e Castro (2012), consideramos a parametrização $\lambda = -\sigma$ na distribuição LGG.

Considere Y_{ij} uma variável aleatória condicionalmente independente com $i = 1, \dots, N$ unidades no tempo $j = 1, \dots, n_i$ (medidas por unidade), e suponha que $Y_{ij}|b_i$ siga a distribuição ZMP. Assim, assumimos os seguintes modelos ZMP com intercepto aleatório:

$$Y_{ij}|b_i \sim ZMP(y_{ij}|b_i; \mu_{ij}, \omega_{ij}),$$

e

$$\log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_i, \quad \log\left(\frac{\omega_{ij}}{1 - \omega_{ij}}\right) = \mathbf{z}'_{ij}\boldsymbol{\gamma}, \quad (3.2)$$

em que $\mathbf{x}_{ij} = (1, x_{ij1}, \dots, x_{ijp})'$ e $\mathbf{z}_{ij} = (1, z_{ij1}, \dots, z_{ijr})'$ são vetores de covariáveis relacionados aos vetores de parâmetros com efeitos fixos $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ e $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_r)'$ respectivamente, e b_i um efeito aleatório para cada unidade i . Para completar a especificação dos modelos, consideramos que o efeito pode assumir as distribuições Normal ou Log-Gama Generalizada. Ou seja, $b_i \sim \pi_b(b_i)$ tal que

$$\pi_b(b_i; \mu, \sigma, \lambda) = \begin{cases} \frac{c(-\sigma)}{\sigma} \exp\left[\frac{-b_i}{\sigma^2} - \frac{e^{-b_i}}{\sigma^2}\right], & \text{se } \lambda \neq 0, \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{b_i^2}{2\sigma^2}\right\}, & \text{se } \lambda = 0. \end{cases} \quad (3.3)$$

Observação 2. É importante ressaltar que o tamanho de uma amostra no modelo (3.2) é dado por $n = N \times n_i$, ou seja, a quantidade do número de observações realizadas em cada unidade por tempo. Além disso, vale frisar que no modelo (3.1), \mathbf{z}_{ij} é um vetor de covariáveis e \mathbf{b} é um vetor de efeitos aleatórios. Já no modelo (3.2) $\mathbf{z}_{ij} = 1$. Logo, b_i exerce o papel de intercepto aleatório.

3.3.1 Características do Modelo ZMP com Intercepto Aleatório Normal e Log-Gama Generalizado

O valor esperado e a variância da variável aleatória Y_{ij} no modelo (3.2), e a covariância entre Y_{ij_1} e Y_{ij_2} , com $j_1 \neq j_2$, são calculadas em seguida.

A princípio, vamos considerar que $Y_{ij}|b_i$ segue a distribuição ZMP(μ_{ij}, p_{ij}). Conforme (2.4), observa-se que $E(Y_{ij}|b_i) = p_{ij}\mu_{ij}$, logo, por (2.8), (3.2) e (3.3) tem-se

$$\begin{aligned} E(Y_{ij}) &= E(E(Y_{ij}|b_i)) = E(p_{ij}\mu_{ij}) \\ &= E\left(\frac{\omega_{ij}}{1 - e^{-e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i}}} \times e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i}\right) = \omega_{ij} \int_{-\infty}^{\infty} \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i} \times \pi_b(b_i; \mu, \sigma, \lambda)}{1 - e^{-e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i}}} db_i. \end{aligned} \quad (3.4)$$

Da mesma forma, considerando inicialmente que $Y_{ij}|b_i$ segue a distribuição ZMP(μ_{ij}, p_{ij}), segundo (2.4), temos que $\text{Var}(Y_{ij}|b_i) = p_{ij}[\mu_{ij} + \mu_{ij}^2(1 - p_{ij})]$. Como $\text{Var}(Y_{ij}) = \text{Var}(E(Y_{ij}|b_i)) +$

$E(\text{Var}(Y_{ij}|b_i))$, por (2.8), (3.2) e (3.3) tem-se

$$\begin{aligned}
\text{Var}(Y_{ij}) &= \text{Var}(p_{ij}\mu_{ij}) + E(p_{ij}\{\mu_{ij} + \mu_{ij}^2(1 - p_{ij})\}) \\
&= \text{Var}(p_{ij}\mu_{ij}) + E(p_{ij}\mu_{ij} + p_{ij}\mu_{ij}^2 + p_{ij}^2\mu_{ij}^2) \\
&= E(p_{ij}^2\mu_{ij}^2) - E(p_{ij}\mu_{ij})^2 + E(p_{ij}\mu_{ij}) + E(p_{ij}\mu_{ij}^2) - E(p_{ij}^2\mu_{ij}^2) \\
&= -E(p_{ij}\mu_{ij})^2 + E(p_{ij}\mu_{ij}) + E(p_{ij}\mu_{ij}^2) \\
&= -E\left(\frac{\omega_{ij}}{1 - e^{-e^{x'_{ij}\beta + b_i}}} \cdot e^{x'_{ij}\beta + b_i}\right)^2 + E\left(\frac{\omega_{ij}}{1 - e^{-e^{x'_{ij}\beta + b_i}}} \cdot e^{x'_{ij}\beta + b_i}\right) \\
&\quad + E\left(\frac{\omega_{ij}}{1 - e^{-e^{x'_{ij}\beta + b_i}}} \cdot e^{2x'_{ij}\beta + 2b_i}\right) \\
&= -\left(\omega_{ij} \int_{-\infty}^{\infty} \frac{e^{x'_{ij}\beta + b_i} \times \pi_b(b_i; \mu, \sigma, \lambda)}{1 - e^{-e^{x'_{ij}\beta + b_i}}} db_i\right)^2 + \omega_{ij} \int_{-\infty}^{\infty} \frac{e^{x'_{ij}\beta + b_i} \times \pi_b(b_i; \mu, \sigma, \lambda)}{1 - e^{-e^{x'_{ij}\beta + b_i}}} db_i \\
&\quad + \omega_{ij} \int_{-\infty}^{\infty} \frac{e^{2x'_{ij}\beta + 2b_i} \times \pi_b(b_i; \mu, \sigma, \lambda)}{1 - e^{-e^{x'_{ij}\beta + b_i}}} db_i.
\end{aligned} \tag{3.5}$$

A covariância entre Y_{ij_1} e Y_{ij_2} é dada por

$$\text{Cov}(Y_{ij_1}, Y_{ij_2}) = \text{Cov}(E(Y_{ij_1}|b_i), E(Y_{ij_2}|b_i)) + E(\text{Cov}(Y_{ij_1}, Y_{ij_2}|b_i)). \tag{3.6}$$

Assumindo independência condicional entre Y_{ij_1} e Y_{ij_2} , temos que

$$\text{Cov}(Y_{ij_1}, Y_{ij_2}|b_i) = E(Y_{ij_1}|b_i Y_{ij_2}|b_i) - E(Y_{ij_1}|b_i)E(Y_{ij_2}|b_i) = 0.$$

Logo, a equação em (3.6) é reescrita como

$$\text{Cov}(Y_{ij_1}, Y_{ij_2}) = \text{Cov}(E(Y_{ij_1}|b_i), E(Y_{ij_2}|b_i)).$$

Considerando inicialmente que $Y_{ij}|b_i$ segue a distribuição ZMP(μ_{ij}, p_{ij}), por (2.4), (2.8) e (3.2) tem-se

$$\begin{aligned}
\text{Cov}(Y_{ij_1}, Y_{ij_2}) &= \text{Cov}(p_{ij_1}\mu_{ij_1}, p_{ij_2}\mu_{ij_2}) \\
&= \text{Cov}\left(\frac{\omega_{ij_1}}{1 - e^{-e^{x'_{ij_1}\beta + b_i}}} \times e^{x'_{ij_1}\beta + b_i}, \frac{\omega_{ij_2}}{1 - e^{-e^{x'_{ij_2}\beta + b_i}}} \times e^{x'_{ij_2}\beta + b_i}\right).
\end{aligned}$$

Além disto, também podemos calcular o parâmetro p_{ij} .

Como $E(p_{ij}|b_i) = \omega_{ij}/1 - e^{-e^{x'_{ij}\beta + b_i}}$, então

$$p_{ij} = E[E(p_{ij}|b_i)] = E\left[\frac{\omega_{ij}}{1 - e^{-e^{x'_{ij}\beta + b_i}}}\right] = \omega_{ij} \int_{-\infty}^{\infty} \frac{\pi_b(b_i; \mu, \sigma, \lambda)}{1 - e^{-e^{x'_{ij}\beta + b_i}}} db_i. \tag{3.7}$$

As integrais calculadas sobre os efeitos aleatórios em (3.4), (3.5) e (3.7) não podem ser resolvidas analiticamente. Entre os diversos métodos de integração numérica possíveis para realizar esse cálculo, usamos o método de integração Quadratura de Gauss-Hermite (para mais detalhes, vide Bonat *et al.* (2012, p. 137)).

MÉTODOS DE ESTIMAÇÃO

Neste capítulo abordamos o processo para a estimação dos parâmetros do modelo ZMP com efeito aleatório utilizando a abordagem clássica e bayesiana. Apresentamos o procedimento para o cálculo dos intervalos de confiança assintóticos, o teste de razão de verossimilhança para o efeito aleatório, e informações sobre a análise de diagnóstico para o modelo.

4.1 Estimação por Máxima Verossimilhança Marginal

Na abordagem clássica, a estimação dos parâmetros do modelo (3.2) pode ser feita maximizando a função de verossimilhança marginal de $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)'$, em que $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$. Considere $\pi_{ZMP}(y_{ij}|b_i; \beta, \gamma)$ e $\pi_b(b_i; \mu, \sigma, \lambda)$, sendo, respectivamente, a função massa de probabilidade (PMF) de $y_{ij}|b_i$, e a função densidade de probabilidade (PDF) de b_i . Assim, a PMF marginal de \mathbf{y} tem a seguinte forma:

$$\pi_{ZMP}(\mathbf{y}; \beta, \gamma, \sigma, \lambda) = \prod_{i=1}^N \pi_{ZMP}(\mathbf{y}_i; \beta, \gamma, \sigma, \lambda) \quad (4.1)$$

em que a contribuição de cada unidade amostral para a verossimilhança é

$$\pi_{ZMP}(\mathbf{y}_i; \beta, \gamma, \sigma, \lambda) = \int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{n_i} \pi_{ZMP}(y_{ij}|b_i; \beta, \gamma) \right\} \pi_b(b_i; \mu, \sigma, \lambda) db_i.$$

Logo, a função log-verossimilhança para o modelo proposto, obtida por (4.1), é dada por

$$\ell(\beta, \gamma, \sigma, \lambda; \mathbf{y}) = \sum_{i=1}^N \log \int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{n_i} \pi_{ZMP}(y_{ij}|b_i; \beta, \gamma) \right\} \pi_b(b_i; \mu, \sigma, \lambda) db_i. \quad (4.2)$$

Dado que

$$\begin{aligned}
\pi_{ZMP}(y_{ij}|b_i; \beta, \gamma) &= (1 - \omega_{ij})I(y_{ij}) + \omega_{ij}\pi_{ZTP}(y_{ij}|b_i; \beta) \\
&= (1 - \omega_{ij})^{I(y_{ij})} \omega_{ij}^{(1-I(y_{ij}))} [\pi_{ZTP}(y_{ij}|b_i; \beta)]^{(1-I(y_{ij}))} \\
&= (1 - \omega_{ij})^{I(y_{ij})} \omega_{ij}^{(1-I(y_{ij}))} \left[\frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{(1 - e^{-\mu_{ij}}) y_{ij}!} \right]^{(1-I(y_{ij}))},
\end{aligned} \tag{4.3}$$

e substituindo (4.3) em (4.2), obtemos

$$\begin{aligned}
\ell(\beta, \gamma, \sigma, \lambda; \mathbf{y}) &= \sum_{i=1}^N \log \int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{n_i} (1 - \omega_{ij})^{I(y_{ij})} \omega_{ij}^{(1-I(y_{ij}))} \left[\frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{(1 - e^{-\mu_{ij}}) y_{ij}!} \right]^{1-I(y_{ij})} \right\} \\
&\quad \times \pi_b(b_i; \mu, \sigma, \lambda) db_i \\
&= \left\{ \sum_{i=1}^N \sum_{j=1}^{n_i} I(y_{ij}) \log(1 - \omega_{ij}) + (1 - I(y_{ij})) (\log(\omega_{ij}) - \log(y_{ij}!)) \right\} \\
&\quad + \sum_{i=1}^N \log \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} \left[\frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{(1 - e^{-\mu_{ij}})} \right]^{1-I(y_{ij})} \times \pi_b(b_i; \mu, \sigma, \lambda) db_i \\
&= \left\{ \sum_{i,j} I(y_{ij}) \log \left(1 - \frac{e^{z'_{ij}\gamma}}{1 + e^{z'_{ij}\gamma}} \right) + (1 - I(y_{ij})) \left(\log \left(\frac{e^{z'_{ij}\gamma}}{1 + e^{z'_{ij}\gamma}} \right) - \log(y_{ij}!) \right) \right\} \\
&\quad + \sum_{i=1}^N \log \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} \left[\frac{e^{-e^{x'_{ij}\beta+b_i}} e^{(x'_{ij}\beta+b_i)y_{ij}}}{(1 - e^{-e^{x'_{ij}\beta+b_i}})} \right]^{1-I(y_{ij})} \times \pi_b(b_i; \mu, \sigma, \lambda) db_i \\
&= \left\{ \sum_{i,j} I(y_{ij}) \log \left(1 - \frac{e^{z'_{ij}\gamma}}{1 + e^{z'_{ij}\gamma}} \right) + (1 - I(y_{ij})) \left(\log \left(\frac{e^{z'_{ij}\gamma}}{1 + e^{z'_{ij}\gamma}} \right) - \log(y_{ij}!) \right) \right\} \\
&\quad + \sum_{i,j} (1 - I(y_{ij})) x'_{ij} \beta y_{ij} \\
&\quad + \sum_{i=1}^N \log \int_{-\infty}^{+\infty} \frac{\exp \left\{ \sum_{j=1}^{n_i} (1 - I(y_{ij})) [y_{ij} b_i - e^{x'_{ij}\beta+b_i}] \right\}}{\prod_{j=1}^{n_i} (1 - e^{-e^{x'_{ij}\beta+b_i}})^{1-I(y_{ij})}} \times \pi_b(b_i; \mu, \sigma, \lambda) db_i.
\end{aligned} \tag{4.4}$$

O que torna mais difícil a maximização da expressão (4.4) é o cálculo das N integrais sobre os efeitos aleatórios, já que as mesmas não possuem solução analítica. Existem algumas exceções, assim como podemos ver no trabalho de [Fabio, Paula e Castro \(2012\)](#), em que as integrais presentes na função de verossimilhança possuem expressões em forma fechada. Entre os diversos métodos possíveis para calcular as integrais em (4.4), adotamos o método da aproximação de Laplace (para mais detalhes, vide [Bonat et al. \(2012, p. 140\)](#)). Como sugerido em [Bonat et al. \(2012\)](#), para estimar os parâmetros do modelo (3.2), usamos a função `mle2` ([ROSENBAUM; RALL, 2018](#)) disponível na linguagem `R`, versão 3.5.0 (os códigos utilizados neste trabalho podem ser solicitados à autora).

4.1.1 Intervalos de Confiana

Os intervalos de confiana assintticos para os parmetros do modelo (3.2), β , γ , σ e λ , so calculados a partir da matriz de covarincias assinttica dos parmetros. Alm do mais, utilizando o mtodo Delta (MIGON; GAMERMAN; LOUZADA, 1999), podemos encontrar os intervalos de confiana para os parmetros p_{ij} , $\mu_{ZMP_{ij}}$ e $p_{0_{ij}}$, em que $\mu_{ZMP_{ij}} = E(Y_{ij})$ e $p_{0_{ij}} = \pi_{ZMP}(0, \mu_{ij}, \omega_{ij})$.

A matriz de covarincia dos estimadores  baseada na matriz de informao observada, a qual tem a seguinte forma: a partir da funo log-verossimilhana dada em (4.4), definimos $J_{\theta_1 \theta_2} = \frac{\partial^2 l(\beta, \gamma, \sigma, \lambda)}{\partial \theta_1 \partial \theta_2}$, em que $\theta_1, \theta_2 = \beta_0, \beta_1, \beta_2, \dots, \beta_p, \gamma_0, \gamma_1, \gamma_2, \dots, \gamma_r, \sigma, \lambda$. Logo, a matriz de informao observada  dada por

$$J = - \begin{bmatrix} J_{\beta_0 \beta_0} & \dots & J_{\beta_0 \beta_p} & J_{\beta_0 \gamma_0} & \dots & J_{\beta_0 \gamma_r} & J_{\beta_0 \sigma} & J_{\beta_0 \lambda} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ J_{\beta_p \beta_0} & \dots & J_{\beta_p \beta_p} & J_{\beta_p \gamma_0} & \dots & J_{\beta_p \gamma_r} & J_{\beta_p \sigma} & J_{\beta_p \lambda} \\ J_{\gamma_0 \beta_0} & \dots & J_{\gamma_0 \beta_p} & J_{\gamma_0 \gamma_0} & \dots & J_{\gamma_0 \gamma_r} & J_{\gamma_0 \sigma} & J_{\gamma_0 \lambda} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ J_{\gamma_r \beta_0} & \dots & J_{\gamma_r \beta_p} & J_{\gamma_r \gamma_0} & \dots & J_{\gamma_r \gamma_r} & J_{\gamma_r \sigma} & J_{\gamma_r \lambda} \\ J_{\sigma \beta_0} & \dots & J_{\sigma \beta_p} & J_{\sigma \gamma_0} & \dots & J_{\sigma \gamma_r} & J_{\sigma \sigma} & J_{\sigma \lambda} \\ J_{\lambda \beta_0} & \dots & J_{\lambda \beta_p} & J_{\lambda \gamma_0} & \dots & J_{\lambda \gamma_r} & J_{\lambda \sigma} & J_{\lambda \lambda} \end{bmatrix},$$

cujos elementos podem ser encontrados apenas numericamente. Assim sendo, a matriz de covarincias dos estimadores  escrita como $COV(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda}) = J^{-1}$. Os intervalos de confiana com um coeficiente de confiana de $100(1 - \alpha)\%$, para os parmetros β , γ , σ e λ , podem ser calculados avaliando a matriz de covarincia com as estimativas de mxima verossimilhana de β , γ , σ e λ ,

$$(\hat{\beta}_0, \dots, \hat{\beta}_p, \hat{\gamma}_0, \dots, \hat{\gamma}_r, \hat{\sigma}, \hat{\lambda})' = \pm z_{\alpha/2} \times \sqrt{\text{diag}(\widehat{COV}(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda}))},$$

em que $z_{\alpha/2}$  o quantil da distribuo Normal padro acumulada, tal que $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Conforme (3.4), (2.8) e (3.2), os parmetros p_{ij} , $\mu_{ZMP_{ij}}$ e $p_{0_{ij}}$ so funoes de β , γ , σ e λ . Logo, vamos utilizar o mtodo Delta para encontrar os intervalos de confiana assintticos para p_{ij} , $\mu_{ZMP_{ij}}$ e $p_{0_{ij}}$. Deste modo, considere $\eta = h(\beta, \gamma, \sigma, \lambda)$, tal que h seja uma funo diferencivel $h: \mathbb{R} \rightarrow \mathbb{R}$. A varincia de $\hat{\eta}$  dada por

$$\text{Var}(\hat{\eta}) = D' \times COV(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda}) \times D \quad (4.5)$$

em que $D = \left(\frac{\partial h}{\partial \beta_0}, \dots, \frac{\partial h}{\partial \beta_p}, \frac{\partial h}{\partial \gamma_0}, \dots, \frac{\partial h}{\partial \gamma_r}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \lambda} \right)'$. Assim, podemos calcular um intervalo de confiana para η , com um coeficiente de confiana de $100(1 - \alpha)\%$, da seguinte forma

$$\hat{\eta} \pm z_{\alpha/2} \times \sqrt{\text{Var} \hat{\eta}}.$$

4.1.2 Teste Razão de Verossimilhança para o Efeito Aleatório

Consideremos os modelos ZMP com efeito aleatório apresentados em (3.2). Para verificar se a variância do efeito aleatório é significativa vamos aplicar o teste razão de verossimilhança (STRAM; LEE, 1994) (SELF; LIANG, 1987).

Seja $\ell_0 = \ell(\hat{\beta}, \hat{\gamma})$ a função log-verossimilhança do modelo sem o efeito aleatório, avaliada nas estimativas dos parâmetros, e $\ell_1 = \ell(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda})$ a função log-verossimilhança do modelo com o efeito aleatório avaliada nas estimativas dos parâmetros. Estamos testando as hipóteses $H_0 : \sigma = 0$ versus $H_1 : \sigma > 0$.

Com um nível de significância α , rejeitamos H_0 se $T > \chi_{2\alpha,1}^2$, em que $T = -2(\ell_0 - \ell_1)$ é a estatística do teste razão de verossimilhança e $\chi_{2\alpha,1}^2$ é o $(1 - 2\alpha)$ -ésimo quantil da distribuição χ^2 com 1 grau de liberdade. O valor-p é dado por $P[\chi_1^2 \geq T]/2$.

4.2 Abordagem Bayesiana

Considere novamente y_{ij} , observações condicionalmente independentes da variável aleatória Y_{ij} , com $i = 1, \dots, N$ unidades no tempo $j = 1, \dots, n_i$ (medidas por unidade), em que $Y_{ij}|b_i \sim ZMP(y_{ij}|b_i; \mu_{ij}, \omega_{ij})$. Além disso, seja b_i um efeito aleatório para cada unidade i , seguindo a distribuição $\pi_b(b_i; \mu, \sigma, \lambda)$.

Na abordagem bayesiana para o modelo proposto (3.2) consideremos a seguinte função de verossimilhança conjunta:

$$L(\beta, \gamma, b_i, \mu, \sigma, \lambda | \mathbf{y}) = \prod_{i=1}^N \left\{ \prod_{j=1}^{n_i} \pi_{ZMP}(y_{ij}|b_i; \beta, \gamma) \right\} \pi_b(b_i; \mu, \sigma, \lambda),$$

a qual é associada ao vetor de observações $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)'$, em que $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$, e que foi obtida a partir de (4.1). Além disso, tomemos densidades a priori vagas para cada elemento dos vetores $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ e $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_r)'$, e para os parâmetros σ , μ e λ . Assumindo independência dos parâmetros, a priori conjunta é dada por $\pi(\beta, \gamma, \mu, \sigma, \lambda) = \pi(\beta)\pi(\gamma)\pi(\mu)\pi(\sigma)\pi(\lambda)$. Portanto, escrevemos a densidade a posteriori conjunta como:

$$\pi(\beta, \gamma, b_i, \sigma, \lambda | \mathbf{y}) \propto L(\beta, \gamma, b_i, \mu, \sigma, \lambda | \mathbf{y})\pi(\beta, \gamma, \mu, \sigma, \lambda). \quad (4.6)$$

Do ponto de vista bayesiano, para fazer inferências sobre os parâmetros usamos suas densidades a posteriori marginais, que podemos obter quando integramos a densidade a posteriori conjunta (CONCEIÇÃO, 2013) (LESAFFRE; LAWSON, 2012). Em tal situação, como não podemos obter soluções analíticas para as integrais, utilizamos o algoritmo *Metropolis-Hastings* (CHIB S.; GREENBERG, 1995), da classe dos métodos de simulação de Monte Carlo em cadeia de Markov (MCMC). Neste trabalho, todas as implementações computacionais envolvendo inferência bayesiana foram feitas utilizando o programa Jags (PLUMMER, 2003), através da

linguagem R. Para trabalhos futuros, o programa Stan (KRUSCHKE, 2014) também é uma opção a ser utilizado. Os estimadores de Monte Carlo foram obtidos considerando as funções de perda quadrática e absoluta, e o critério de Gelman-Rubin (GELMAN A.; RUBIN, 1992) foi utilizado para diagnosticar a convergência das cadeias.

4.3 Análise de Diagnóstico

A análise de diagnóstico é fundamental para avaliar a bondade do ajuste de um modelo de regressão e, além disso, identificar a presença de observações discrepantes (outliers) ou influentes. Neste trabalho, para a análise de diagnóstico, vamos adotar o gráfico normal de probabilidade com envelope gerado considerando o resíduo de Pearson. A seguir definimos o resíduo de Pearson e apresentamos mais informações sobre o Gráfico Envelope.

4.3.1 Resíduo de Pearson

Baseado na diferença entre os valores observados e os valores preditos pelo modelo, o resíduo de Pearson é um dos mais utilizados em análises de diagnóstico. Considerando o modelo (3.2), sua expressão para a ij -ésima observação é dada por

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ZMP_{ij}}}{\sqrt{\widehat{Var}(y_{ij})}}, \quad (4.7)$$

em que $\hat{\mu}_{ZMP_{ij}}$ e $\widehat{Var}(y_{ij})$ são, respectivamente, a média e a variância ajustada, $i = 1, \dots, N$ e $j = 1, \dots, n_i$.

4.3.2 Gráfico Envelope

O gráfico Envelope foi proposto por Atkinson (1981) e resume-se em obter um tipo de banda de confiança para os resíduos, através de simulações, em um gráfico normal de probabilidade. A sugestão de Atkinson (1981) é realizar os seguintes passos, no mínimo, $m = 19$ vezes: gerar n observações a partir do modelo ajustado, ajustar o modelo com essas novas observações e obter o vetor de resíduos. Ordenando cada vetor de resíduos de maneira crescente, os limites superiores e inferiores da banda de confiança serão, respectivamente, o máximo e o mínimo comparando todos os vetores de resíduos. Neste trabalho, alternativamente, consideramos $m = 100$. Além disso, usamos como limites superiores e inferiores valores que produzam uma banda com 95% de confiança. Ou seja, o limite superior é definido pela média dos vetores $r_{[ij],[97]}$ e $r_{[ij],[98]}$, enquanto o limite inferior é definido pela média dos vetores $r_{[ij],[2]}$ e $r_{[ij],[3]}$, em que $i = 1, \dots, N$ e $j = 1, \dots, n_i$ (PAULA, 2004). O objetivo é verificar, por meio dos resíduos, se as observações podem ser consideradas provenientes de uma amostra aleatória da distribuição que

estamos supondo aos dados. Consequentemente, se muitos pontos não pertencerem ao envelope, o ajuste não é adequado.

4.4 Critérios para Seleção de Modelos

Na estimação clássica, para selecionar o modelo que melhor se ajusta a um conjunto de dados vamos considerar o critério de informação bayesiano de Schwarz (BIC) (SCHWARZ, 1978), e o critério de informação de Akaike (AIC) (AKAIKE, 1974). O BIC e o AIC penalizam a função log-verossimilhança a partir da complexidade do modelo e são calculados por

$$BIC = -2\ell(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda}) + s \log(n) \quad , \quad AIC = -2\ell(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda}) + 2s,$$

em que $\ell(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, \hat{\lambda})$ é o valor da função log-verossimilhança avaliada nas estimativas dos parâmetros, s é o número de parâmetros do modelo, e n é o número de observações. O modelo com menor valor de BIC e AIC é considerado o mais adequado.

Na abordagem bayesiana considera-se o desvio bayesiano (DEMPSTER, 1997) $D(\beta, \gamma, \sigma, \lambda) = -2\ell(\beta, \gamma, \sigma, \lambda)$, logo, o desvio bayesiano esperado pode ser aproximado por

$$\bar{D} = E(D(\beta, \gamma, \sigma, \lambda)) \approx \sum_{k=1}^M D(\beta^k, \gamma^k, \sigma^k, \lambda^k) / M \quad (4.8)$$

em que M é o tamanho da cadeia gerada pelo método MCMC. Vamos utilizar o critério de informação de Akaike extendido (EAIC), o critério de informação bayesiano de Schwarz extendido (EBIC) (BROOKS, 2002), e o critério de informação de desvio (DIC) (SPIEGELHALTER *et al.*, 2002). O EAIC e BAIC são penalizados como o AIC e BIC, respectivamente

$$EBIC = E(D(\beta, \gamma, \sigma, \lambda)) \approx \bar{D} + s \log(n) \quad , \quad EAIC = E(D(\beta, \gamma, \sigma, \lambda)) \approx \bar{D} + 2s.$$

Finalmente, o DIC é dado por

$$DIC = 2E(D(\beta, \gamma, \sigma, \lambda)) - D(E(\beta), E(\gamma), E(\sigma), E(\lambda))$$

em que $D(E(\beta), E(\gamma), E(\sigma), E(\lambda))$ é o desvio bayesiano nas médias a posteriori e pode ser aproximado por

$$D(E(\beta), E(\gamma), E(\sigma), E(\lambda)) \approx D \left(\sum_k^M \beta^{(k)} / M, \sum_k^M \gamma^{(k)} / k, \sum_k^M \sigma^{(k)} / M, \sum_k^M \lambda^{(k)} / M \right).$$

ESTUDO DE SIMULAÇÃO

Neste capítulo é apresentado um estudo de simulação para os modelos ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG. O objetivo é avaliar o processo de estimação dos parâmetros a partir das abordagens clássica e bayesiana, e suas propriedades assintóticas quando a unidade $N \rightarrow \infty$ e o tempo n_i é fixado.

5.1 Cenários da Simulação

Para este estudo de simulação foram considerados diferentes cenários e diferentes caracterizações de conjuntos de dados: conjuntos de dados caracterizados como zero-inflacionados e zero-deflacionados. Para cada situação foram geradas $L = 1000$ réplicas, com diferentes unidades ($N = 30, 100$) e tempos $n_i = 5$, totalizando em amostras de tamanhos 150 e 500 ($n = N \times n_i$). Além disso, as observações seguem a distribuição $ZMP(\mu_{ij}, p_{ij})$, com $i = 1, \dots, N$, $j = 1, \dots, n_i$, e são correlacionadas no tempo. Considerando a matriz modelo $X_{N \times n_N}$, dada por

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n_1} \\ 1 & x_{21} & x_{22} & \dots & x_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Nn_N} \end{pmatrix},$$

em que os vetores de covariáveis \mathbf{x} são constituído dos elementos das colunas da matriz $X_{N \times n_N}$ e são obtidos da seguinte forma: gere x_{ij} considerando uma $U(0, 1)$, para $i = 1, \dots, N$.

Os conjuntos de dados zero-inflacionados correlacionados no tempo foram gerados de um modelo ZMP com efeito aleatório Normal e de um modelo ZMP com efeito aleatório LGG com $\beta' = (1, 3)$, $\gamma' = (-1, 2)$, e diferentes valores de σ^2 . $\sigma^2 = 0.25$ em que $\lambda = -0.5$ e $\sigma^2 = 1$ no qual $\lambda = -1$, tal que $p_{ij} < 1$. Já os conjuntos de dados zero-deflacionados correlacionados no tempo, foram gerados de um modelo ZMP com efeito aleatório Normal e de um modelo ZMP com efeito aleatório LGG com $\beta' = (-0.5, -1)$, $\gamma' = (2, -2)$, e da mesma forma, com

diferentes valores de σ^2 . $\sigma^2 = 0.2$ em que $\lambda = -0.45$ e $\sigma^2 = 0.7$ no qual $\lambda = -0.84$, tal que $p_{ij} > 1$.

5.2 Gerador de Variáveis Aleatórias ZMP Correlacionadas no Tempo

O esquema para gerar variáveis aleatórias correlacionadas no tempo para o modelo ZMP com efeito aleatório é descrito a seguir:

1. Gere $b_i \sim \pi_b$, para $i = 1, \dots, N$.
2. Faça $\mathbf{b} = (b_1^{(1)}, b_1^{(2)}, \dots, b_1^{(n_1)}, b_2^{(1)}, b_2^{(2)}, \dots, b_2^{(n_2)}, \dots, b_N^{(1)}, b_N^{(2)}, \dots, b_N^{(n_N)})$ a quantidade de tempos observados.
3. Faça $i = 1$ e $j = 1, \dots, n_1$.
4. Calcule $\mu_{ij} = e^{x'_{ij}\beta + b_i}$ e a probabilidade de zero, p_0 , a qual é dada por

$$p_0 = 1 - \frac{e^{x'_{ij}\gamma}}{1 + e^{x'_{ij}\gamma}}.$$

5. Gere $u \sim U(0, 1)$ e faça $y_{ij} = 0$. Se $u \leq p_0$, faça $y_{ij} = 0$.

Caso contrário, faça $\alpha_2 = p_0$,

Enquanto $y_{ij} = 0$ {

Faça $\alpha_1 = \alpha_2$

Faça $\alpha_1 = \alpha_2 + (1 - p_0)\pi_{ZTP}(q, \mu_{ij})$.

Se $(\alpha_1 < 1 < \alpha_2)$, faça $y_{ij} = q$.

Caso contrário, faça $q = q + 1$.

}

6. Faça $i = i + 1$ e vá para o passo 4.

5.3 Critérios de Avaliação dos Estimadores

Para avaliar o estimador de cada parâmetro θ do modelo, foram calculados o vício ($\mathcal{B}(\hat{\theta})$) do estimador, a raiz quadrada da razão entre o erro quadrático médio (EQM) e a variância (Var) do estimador $\left(\sqrt{EQM(\hat{\theta})/Var(\hat{\theta})}\right)$, o erro absoluto médio (EAM) e a razão entre a média dos erros padrão (EP) assintóticos e a raiz quadrada da variância dos estimadores

$\left(EP(\hat{\theta})/\sqrt{Var(\hat{\theta})}\right)$. Além disso, na abordagem clássica obtemos os intervalos assintóticos com 95% de confiança, e na abordagem bayesiana os intervalos com 95% de credibilidade para cada parâmetro, a fim de verificar suas probabilidades de cobertura (PC).

O vício, a variância, o EQM e o EAM dos estimadores são calculados por:

$$\mathcal{B}(\hat{\theta}) = \bar{\hat{\theta}} - \theta, \quad Var(\hat{\theta}) = \frac{1}{L} \sum_{i=1}^L (\hat{\theta}_i - \theta)^2,$$

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + \mathcal{B}(\hat{\theta})^2, \quad EAM(\hat{\theta}) = \frac{1}{L} \sum_{i=1}^L |\hat{\theta}_i - \theta|,$$

em que

$$\bar{\hat{\theta}} = \frac{1}{L} \sum_{i=1}^L \hat{\theta}_i.$$

Na abordagem clássica calculamos os estimadores a partir do método de máxima verossimilhança, enquanto na abordagem bayesiana os estimadores de Monte Carlo foram obtidos considerando a função de perda quadrática (média a *posteriori*). Usando o software Jags, foram geradas duas cadeias de tamanho 50.000 para cada parâmetro, com um período *burn in* de 10.000 iterações em cada cadeia e saltos de tamanho 10. Resultando em cadeias de tamanho 8.000 para cada parâmetro. A convergência das cadeias foi verificada através do critério de Gelman Rubin, e as densidades a *priori* utilizadas foram: $\beta_0 \sim N(0, 1000)$, $\beta_1 \sim N(0, 1000)$, $\gamma_0 \sim N(0, 1000)$, $\gamma_1 \sim N(0, 1000)$, $\tau \sim Gama(0.01, 0.001)$ em que $\tau = 1/\sigma^2$ e $\lambda = -\sqrt{1/\tau}$.

Para a boa eficácia dos estimadores, é esperado que os valores de \mathcal{B} e EAM sejam próximos de zero, e que as razões $\sqrt{EQM/Var}$ e EP/\sqrt{Var} sejam próximas de um. Também espera-se que as probabilidades de cobertura para os parâmetros estejam próximas de 95%.

5.4 Resultados com a Abordagem Clássica

Inicialmente vamos analisar os resultados considerando o modelo ZMP com efeito aleatório Normal. Observando a [Tabela 1](#) e [Tabela 2](#), correspondentes a análise dos dados zero-inflacionados correlacionados no tempo, podemos perceber que de forma geral, os valores do \mathcal{B} e do EAM se aproximam de zero conforme aumentamos o tamanho da amostra (ou seja, o número de unidades N). Da mesma forma, observa-se que o valor do $\sqrt{EQM/Var}$ e do EP/\sqrt{Var} se aproxima de 1 quando aumentamos o tamanho da amostra, o que nos mostra que os estimadores são assintoticamente não viciados e que a média dos erros padrão assintóticos estão próximas da \sqrt{Var} . Tais resultados comprovam que as estimativas estão próximas dos verdadeiros valores do parâmetro. Observando as PC do conjunto de dados zero-inflacionados, nota-se que para a amostra de tamanho $n = 150$ e com σ^2 igual a 0.25 e 1, apenas o parâmetro σ^2 possui probabilidade menor que 90%. Entretanto, quando aumentamos o tamanho da amostra, a probabilidade de cobertura apresenta valor a cerca de 0.95.

Tabela 1 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal**, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
N = 30 $n_i = 5$ n = 150	β_0	1	-0.007	1.000	0.195	0.966	0.935
	β_1	3	0.015	1.000	0.315	0.958	0.937
	γ_0	-1	-0.032	1.003	0.298	0.952	0.945
	γ_1	2	0.056	1.003	0.523	0.968	0.953
	σ^2	0.25	-0.021	1.036	0.064	1.001	0.855
N = 100 $n_i = 5$ n = 500	β_0	1	-0.003	1.000	0.104	0.984	0.941
	β_1	3	0.002	1.000	0.166	0.990	0.943
	γ_0	-1	-0.007	1.000	0.155	0.991	0.952
	γ_1	2	0.013	1.000	0.270	0.986	0.951
	σ^2	0.25	-0.009	1.021	0.036	0.989	0.918

Tabela 2 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal**, considerando um conjunto de dados zero-inflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
N = 30 $n_i = 5$ n = 150	β_0	1	-0.007	1.000	0.360	0.935	0.926
	β_1	3	0.022	1.000	0.583	0.942	0.932
	γ_0	-1	-0.025	1.002	0.289	0.987	0.952
	γ_1	2	0.046	1.002	0.501	0.996	0.956
	σ^2	1	-0.070	1.026	0.245	0.923	0.839
N = 100 $n_i = 5$ n = 500	β_0	1	-0.002	1.000	0.193	0.962	0.936
	β_1	3	0.011	1.000	0.312	0.970	0.943
	γ_0	-1	-0.012	1.001	0.156	0.977	0.946
	γ_1	2	0.024	1.002	0.275	0.972	0.944
	σ^2	1	-0.026	1.013	0.126	1.024	0.927

Analisando a [Tabela 3](#) e [Tabela 4](#), referentes aos dados zero-deflacionados correlacionados no tempo, podemos notar que com ambos valores de σ^2 , 0.2 e 0.7, o \mathcal{B} e o EAM se aproximam de zero conforme aumentamos o tamanho da amostra. Além disso, também observa-se que os estimadores são assintoticamente não viciados e que a média dos erros padrão assintóticos estão próximas da raiz quadrada da variância, pois os valores do $\sqrt{EQM/Var}$ e do EP/\sqrt{Var} se aproximam de 1 quando aumentamos o tamanho da amostra. Analisando as probabilidades de cobertura do conjunto de dados zero-deflacionados, podemos observar o impacto da variância no modelo. O aumento da variância e do tamanho da amostra causa uma melhora na probabilidade de cobertura do parâmetro σ^2 . De um modo geral, as probabilidades de cobertura estão próximas de 0.90.

Agora vamos analisar os resultados considerando o modelo ZMP com efeito aleatório LGG. A partir da [Tabela 5](#) e [Tabela 6](#), correspondentes a análise dos dados zero-inflacionados correlacionados no tempo, podemos notar que de forma geral, conforme aumentamos o tamanho

Tabela 3 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	-0.5	-0.021	1.001	0.333	0.952	0.923
	β_1	-1	-0.075	1.003	0.683	0.947	0.945
	γ_0	2	0.079	1.014	0.367	0.967	0.957
	γ_1	-2	-0.111	1.011	0.591	0.957	0.956
	σ^2	0.2	-0.012	1.001	0.202	0.903	0.659
$N = 100$ $n_i = 5$ $n = 500$	β_0	-0.5	-0.028	1.008	0.177	1.023	0.965
	β_1	-1	-0.006	1.000	0.354	0.987	0.957
	γ_0	2	0.014	1.001	0.194	0.976	0.948
	γ_1	-2	-0.017	1.000	0.306	0.977	0.959
	σ^2	0.2	0.031	1.016	0.136	1.007	0.879

Tabela 4 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-deflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	-0.5	-0.002	1.000	0.405	0.935	0.914
	β_1	-1	-0.061	1.001	0.779	0.910	0.934
	γ_0	2	0.058	1.007	0.367	0.950	0.954
	γ_1	-2	-0.087	1.006	0.586	0.956	0.946
	σ^2	0.7	-0.078	1.012	0.387	0.890	0.841
$N = 100$ $n_i = 5$ $n = 500$	β_0	-0.5	-0.013	1.001	0.217	1.111	0.923
	β_1	-1	-0.005	1.000	0.390	1.188	0.937
	γ_0	2	0.008	1.000	0.187	1.007	0.952
	γ_1	-2	-0.000	0.999	0.306	0.983	0.946
	σ^2	0.7	-0.007	1.000	0.197	0.945	0.904

da amostra, os valores do \mathcal{B} e do EAM se aproximam de zero. Da mesma forma, observa-se que os estimadores são assintoticamente não viciados e que a média dos erros padrão assintóticos estão próximas da \sqrt{Var} . Observando as PC do conjunto de dados zero-inflacionados, nota-se que para a amostra de tamanho $n = 150$ e com σ^2 verdadeiro igual a 0.25 e 1, apenas os parâmetros σ^2 e λ possuem probabilidades menores que 90%. Entretanto, quando aumentamos o tamanho da amostra, a probabilidade de cobertura apresenta valores próximos de 0.95.

Analisando a [Tabela 7](#) e [Tabela 8](#), referentes aos dados zero-deflacionados correlacionados no tempo, podemos notar que com ambos valores de σ^2 , 0.2 e 0.7, o \mathcal{B} e o EAM se aproximam de zero, e o $\sqrt{EQM/Var}$ se aproxima de 1 conforme aumentamos o tamanho da amostra. Além disso, é possível comparar o impacto da variância no modelo. Com $\sigma^2 = 0.2$ e $\lambda = -0.45$ obtemos valores bem abaixo de 1 para o EP/\sqrt{Var} e probabilidades de cobertura bem abaixo de 0.95, para os parâmetros σ^2 e λ . Entretanto, quando aumentamos a variância para $\sigma^2 = 0.7$ tal que $\lambda = -0.84$, obtemos valores satisfatórios do EP/\sqrt{Var} e da PC. Entretanto,

Tabela 5 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-inflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	1	0.010	1.000	0.203	0.972	0.942
	β_1	3	-0.004	1.000	0.318	0.965	0.934
	γ_0	-1	-0.011	1.000	0.294	0.966	0.952
	γ_1	2	0.024	1.000	0.515	0.972	0.945
	σ^2	0.25	-0.019	1.035	0.061	0.955	0.865
	λ	-0.5	0.026	1.056	0.064	0.947	0.890
	$N = 100$ $n_i = 5$ $n = 500$	β_0	1	0.007	1.001	0.109	1.001
β_1		3	-0.003	1.000	0.172	0.995	0.950
γ_0		-1	-0.002	1.000	0.152	1.013	0.959
γ_1		2	0.008	1.000	0.270	1.004	0.952
σ^2		0.25	-0.004	1.005	0.033	0.997	0.937
λ		-0.5	0.006	1.010	0.033	0.994	0.950

Tabela 6 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório LGG, considerando um conjunto de dados zero-inflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	1	0.018	1.000	0.387	0.959	0.933
	β_1	3	0.012	1.000	0.612	0.967	0.932
	γ_0	-1	-0.023	1.002	0.288	0.986	0.947
	γ_1	2	0.031	1.001	0.499	0.999	0.957
	σ^2	1	-0.065	1.036	0.200	0.969	0.875
	λ	-1	0.041	1.053	0.104	0.962	0.90
	$N = 100$ $n_i = 5$ $n = 500$	β_0	1	0.001	1.000	0.213	0.958
β_1		3	0.014	1.000	0.342	0.959	0.942
γ_0		-1	-0.003	1.000	0.153	1.007	0.958
γ_1		2	0.021	1.001	0.275	0.980	0.951
σ^2		1	-0.020	1.011	0.104	1.006	0.931
λ		-1	0.012	1.016	0.053	1.002	0.940

podemos observar que quando aumentamos o tamanho da variância a probabilidade de cobertura dos parâmetros β_0 e β_1 tem uma queda considerável.

5.5 Resultados com a Abordagem Bayesiana

Considerando a abordagem bayesiana vamos analisar os resultados do modelo ZMP com efeito aleatório Normal. Observando a [Tabela 9](#) e [Tabela 10](#), correspondentes a análise dos dados zero-inflacionados correlacionados no tempo, podemos perceber que de forma geral, os valores do \mathcal{B} e do EAM se aproximam de zero conforme aumentamos o tamanho da amostra (ou seja, o número de unidades N). Da mesma forma, observa-se que o valor do $\sqrt{EQM/Var}$ se aproxima

Tabela 7 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório LGG**, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	-0.5	-0.008	1.000	0.337	0.992	0.923
	β_1	-1	-0.072	1.003	0.638	1.001	0.936
	γ_0	2	0.023	1.001	0.359	0.982	0.951
	γ_1	-2	-0.015	1.000	0.567	0.991	0.950
	σ^2	0.2	-0.010	1.001	0.157	0.841	0.633
	λ	-0.45	0.107	1.073	0.239	0.895	0.79
	$N = 100$ $n_i = 5$ $n = 500$	β_0	-0.5	-0.070	1.038	0.202	0.976
β_1		-1	-0.003	1.000	0.328	1.053	0.960
γ_0		2	0.032	1.008	0.196	0.957	0.936
γ_1		-2	-0.028	1.002	0.311	0.956	0.942
σ^2		0.2	0.057	1.111	0.107	0.981	0.871
λ		-0.45	-0.031	1.017	0.125	0.756	0.860

Tabela 8 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório LGG**, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	EP/\sqrt{Var}	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	-0.5	0.006	1.000	0.429	0.936	0.884
	β_1	-1	-0.014	1.000	0.813	0.895	0.910
	γ_0	2	0.041	1.004	0.351	0.993	0.942
	γ_1	-2	-0.046	1.001	0.579	0.973	0.946
	σ^2	0.7	-0.040	1.011	0.210	0.983	0.880
	λ	-0.84	0.045	1.029	0.137	0.916	0.930
	$N = 100$ $n_i = 5$ $n = 500$	β_0	-0.5	-0.053	1.016	0.228	0.929
β_1		-1	0.053	1.004	0.430	0.914	0.890
γ_0		2	0.019	1.003	0.187	1.003	0.948
γ_1		-2	-0.026	1.002	0.291	1.020	0.952
σ^2		0.7	0.025	1.014	0.118	0.923	0.906
λ		-0.84	-0.010	1.006	0.070	0.928	0.90

de 1 quando aumentamos o tamanho da amostra, o seja, o vício diminui conforme aumentamos o tamanho da amostra. Tais resultados comprovam que as estimativas estão próximas dos verdadeiros valores do parâmetro. Além disso, todos os parâmetros possuem probabilidades de cobertura próximas de 0.95. É importante observar o impacto da variância no modelo. Podemos notar, de um modo geral, que os valores do \mathcal{B} e do EAM estão mais próximos de zero e os $\sqrt{EQM/Var}$ estão mais próximos de 1 quando $\sigma^2 = 0.25$, do que quando $\sigma^2 = 1$.

Analisando a [Tabela 11](#) e [Tabela 12](#), correspondentes a análise dos dados zero-deflacionados correlacionados no tempo, podemos perceber que de forma geral, os valores do \mathcal{B} e do EAM se aproximam de zero conforme aumentamos o tamanho da amostra. Também observa-se que o valor do $\sqrt{EQM/Var}$ se aproxima de 1 quando aumentamos o tamanho da amostra, o que nos

Tabela 9 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	1	-0.040	1.011	0.208	0.938
	β_1	3	0.053	1.007	0.331	0.941
	γ_0	-1	0.032	1.003	0.302	0.944
	γ_1	2	-0.037	1.001	0.516	0.949
	σ^2	0.25	0.032	1.051	0.079	0.945
$N = 100$ $n_i = 5$ $n = 500$	β_0	1	-0.021	1.011	0.110	0.945
	β_1	3	0.025	1.006	0.174	0.937
	γ_0	-1	0.065	1.052	0.166	0.928
	γ_1	2	-0.091	1.034	0.287	0.945
	σ^2	0.25	0.005	1.006	0.037	0.955

Tabela 10 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal, considerando um conjunto de dados zero-inflacionado** correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	1	-0.085	1.014	0.393	0.925
	β_1	3	0.119	1.011	0.629	0.927
	γ_0	-1	0.167	1.083	0.343	0.924
	γ_1	2	-0.188	1.036	0.564	0.934
	σ	1	0.125	1.054	0.294	0.945
$N = 100$ $n_i = 5$ $n = 500$	β_0	1	-0.072	1.043	0.200	0.943
	β_1	3	0.091	1.028	0.312	0.949
	γ_0	-1	0.164	1.285	0.213	0.875
	γ_1	2	-0.202	1.157	0.320	0.922
	σ^2	1	0.037	1.022	0.137	0.948

mostra que o vício diminui. Podemos notar que todos os parâmetros possuem probabilidades de cobertura próximas de 0.95.

Podemos observar que as probabilidades de cobertura dos parâmetros do modelo ZMP com efeito aleatório Normal com a abordagem bayesiana estão mais próximas de 0.95 do que as probabilidades de cobertura encontradas com a abordagem clássica. Tal comportamento mostra que a abordagem bayesiana é mais eficaz.

Tabela 11 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal**, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	PC
$N = 30$ $n_i = 5$ $n = 150$	β_0	-0.5	-0.051	1.006	0.337	0.935
	β_1	-1	-0.157	1.014	0.710	0.938
	γ_0	2	0.092	1.018	0.391	0.940
	γ_1	-2	-0.110	1.010	0.599	0.935
	σ^2	0.20	0.093	1.025	0.204	0.984
$N = 100$ $n_i = 5$ $n = 150$	β_0	-0.5	0.005	1.000	0.171	0.935
	β_1	-1	-0.046	1.005	0.360	0.935
	γ_0	2	0.018	1.002	0.191	0.945
	γ_1	-2	-0.018	1.001	0.311	0.942
	σ^2	0.20	-0.023	1.010	0.129	0.939

Tabela 12 – Critério de avaliação do estimador de cada parâmetro do **modelo ZMP com efeito aleatório Normal**, considerando um conjunto de dados zero-deflacionado correlacionado no tempo.

n	Parâmetros	Valor verdadeiro	\mathcal{B}	$\sqrt{EQM/Var}$	EAM	PC
$N = 30$ $n_i = 5$ $n = 500$	β_0	-0.5	-0.017	1.000	0.416	0.936
	β_1	-1	-0.178	1.014	0.807	0.945
	γ_0	2	0.053	1.006	0.365	0.939
	γ_1	-2	-0.052	1.002	0.572	0.954
	σ^2	0.7	0.174	1.024	0.553	0.894
$N = 100$ $n_i = 5$ $n = 500$	β_0	-0.5	0.005	1.000	0.212	0.953
	β_1	-1	-0.046	1.004	0.397	0.943
	γ_0	2	0.024	1.005	0.187	0.949
	γ_1	-2	-0.024	1.001	0.297	0.948
	σ^2	0.7	0.027	1.004	0.224	0.933

APLICAÇÃO

A seguinte aplicação foi realizada com o intuito de mostrar a flexibilidade dos modelos ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG em se ajustar a um conjunto de dados longitudinais em que as variáveis podem ser correlacionadas no tempo, e cuja frequência de zeros difere da frequência esperada sob a suposição da distribuição de Poisson tradicional.

6.1 Análise dos Dados de Notificações de Óbitos Infantis

Nesta aplicação consideramos um conjunto de dados referente a notificações de óbitos em crianças com faixa etária de 1 a 4 anos de idade registradas nas cidades do Estado da Bahia, durante os anos de 2014, 2015 e 2016 (SEI, 2018). Foram consideradas para o estudo todas as 417 cidades do Estado da Bahia. Para ilustrar o potencial do efeito aleatório LGG modificamos as contagens referentes a cidade de Feira de Santana.

A [Tabela 13](#) apresenta a distribuição de frequências do conjunto de dados em cada ano. É possível observar uma grande quantidade de notificações zero nos três anos (2014, 2015 e 2016), porém, vale ressaltar, que esses valores podem ser altos devido a falta de notificações de óbitos nas cidades, e não necessariamente pela ausência de óbitos no local. A [Tabela 14](#) exhibe o número médio e desvio padrão (DP) das notificações de óbitos nas cidades da Bahia em cada ano, separadamente. Pode-se verificar que os valores da média e desvio padrão estão muito próximas nos três anos observados. Entretanto, o ano de 2016 contabilizou a maior média e maior DP das notificações, comparado com os anos de 2014 e 2015. A [Figura 1](#) ilustra o perfil individual e médio das notificações de óbitos em cada cidade durante os anos de 2014, 2015 e 2016. Podemos visualizar a trajetória das contagens em todas as cidades ao longo dos anos. Nota-se que 99% das cidades notificaram um número de 0 a 10 óbitos. Também é possível visualizar que o número de notificações de óbitos de uma mesma cidade não tem uma grande variação.

As notificações de óbitos relatadas na mesma cidade podem ser correlacionadas ao longo

Tabela 13 – Distribuição de frequência das notificações de óbitos, em crianças de 1 a 4 anos de idade, das cidades da Bahia, durante os anos de 2014, 2015 e 2016.

2014	y	0	1	2	3	4	6	7	8	9	13	46	65
	f	228	109	46	15	6	3	1	2	3	2	1	1
2015	y	0	1	2	3	4	5	6	9	13	47	61	
	f	239	95	41	24	6	5	3	1	1	1	1	
2016	y	0	1	2	3	4	5	6	7	8	9	45	70
	f	219	111	48	19	5	5	2	2	2	2	1	1

Tabela 14 – Número médio e desvio padrão (DP) de notificações de óbitos, em crianças de 1 a 4 anos de idade, das cidades da Bahia, em cada ano.

Anos	Média	DP
2014	1.14	4.16
2015	1.07	3.95
2016	1.16	4.25

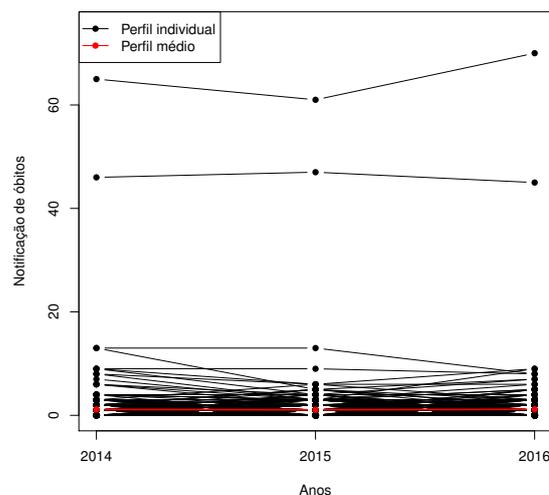


Figura 1 – Perfil individual e médio das notificações de óbitos.

dos anos de 2014, 2015 e 2016. Além disso, apresentam uma alta frequência de zeros. Assim sendo, os modelos propostos para analisar tal conjunto de dados são: modelo ZMP com intercepto aleatório Normal e modelo ZMP com intercepto aleatório LGG, ambos abordados na [seção 3.3](#). A variável explicativa utilizada nos modelos é o índice de desenvolvimento humano (IDH) ([UNDP, 2018](#)). O IDH é um índice de comparação usado para medir o grau de desenvolvimento de uma população a partir de sua educação, riqueza e expectativa média de vida ao nascer. Valores de IDH entre 0,8 e 1 são considerados muito altos, entre 0,7 e 0,79 são considerados altos, entre 0,6 e 0,69 são considerados médios, entre 0,5 e 0,59 são considerados baixos e entre 0,4 e 0,49 são considerados muito baixos.

Deste modo, considere a variável aleatória Y_{ij} , que representa as notificações de óbitos de crianças de 1 a 4 anos de idade registradas na i -ésima cidade da Bahia, no j -ésimo ano ($i = 1, \dots, 417$ e $j = 1, 2, 3$). Logo,

$$Y_{ij}|b_i \sim ZMP(y_{ij}|b_i; \mu_{ij}, \omega_{ij}),$$

$$\log(\mu_{ij}) = \beta_0 + X_{1ij}\beta_1 + b_i \quad (6.1)$$

e

$$\log\left(\frac{\omega_{ij}}{1 - \omega_{ij}}\right) = \gamma_0 + X_{1ij}\gamma_1,$$

em que X_1 é a variável explicativa que corresponde ao IDH das cidades. Além disso, b_i é um efeito aleatório para cada cidade i seguindo as distribuições $N(0, \sigma^2)$ ou $LGG(0, \sigma, \lambda = -\sigma)$, com σ desconhecido.

6.2 Análise com a Abordagem Clássica

Na abordagem clássica vamos utilizar a estimação por máxima verossimilhança marginal para estimar os parâmetros dos modelos, assim como está retratado na [seção 4.1](#).

6.2.1 Modelo ZMP com Efeito Aleatório Normal

Consideremos o modelo apresentado em (6.1) com $b_i \sim N(0, \sigma^2)$ para modelar o conjunto de dados referente as notificações de óbitos infantis descrito na [seção 6.1](#).

Na [Tabela 15](#) encontram-se as estimativas, os erros padrão, os intervalos de confiança assintóticos com 95% de confiança e os valores-p. Verificou-se, a um nível de significância de 5%, que a covariável X_1 é significativa para p e para μ , ou seja, o valor do IDH tem uma relação com o número de notificações de óbitos nas cidades.

Tabela 15 – Estimativas dos parâmetros do **modelo ZMP com efeito aleatório Normal**, erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p **considerando os dados completos de notificações de óbitos no Estado da Bahia**.

Parâmetro	Estimativas	Erro padrão	IC(%)	Valor p
β_0	-7.867	0.833	(-9.499 ; -6.235)	< 0.001
β_1	12.819	1.340	(10.193 ; 15.446)	< 0.001
γ_0	-6.238	0.879	(-7.962 ; -4.515)	< 0.001
γ_1	10.171	1.477	(7.277 ; 13.066)	< 0.001
σ^2	0.534	0.096	(0.346 ; 0.721)	< 0.001

A fim de mostrar a significância do efeito aleatório no ajuste do modelo vamos utilizar o teste razão de verossimilhança apresentado na [subseção 4.1.2](#). Sendo a estatística do teste $T = 571.413$ e correspondente valor-p < 0.001, com um nível de confiança de 95%, rejeitamos

a hipótese nula. Tal resultado comprova a significância do efeito aleatório no ajuste do modelo e indica uma correlação entre o número de notificações de óbitos da mesma cidade.

Os gráficos exibidos na [Figura 2](#) apresentam: (a) estimativas e limites dos intervalos com 95% de confiança do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório Normal. Mais detalhes sobre o cálculo dos intervalos de confiança são dados no [Apêndice A](#).

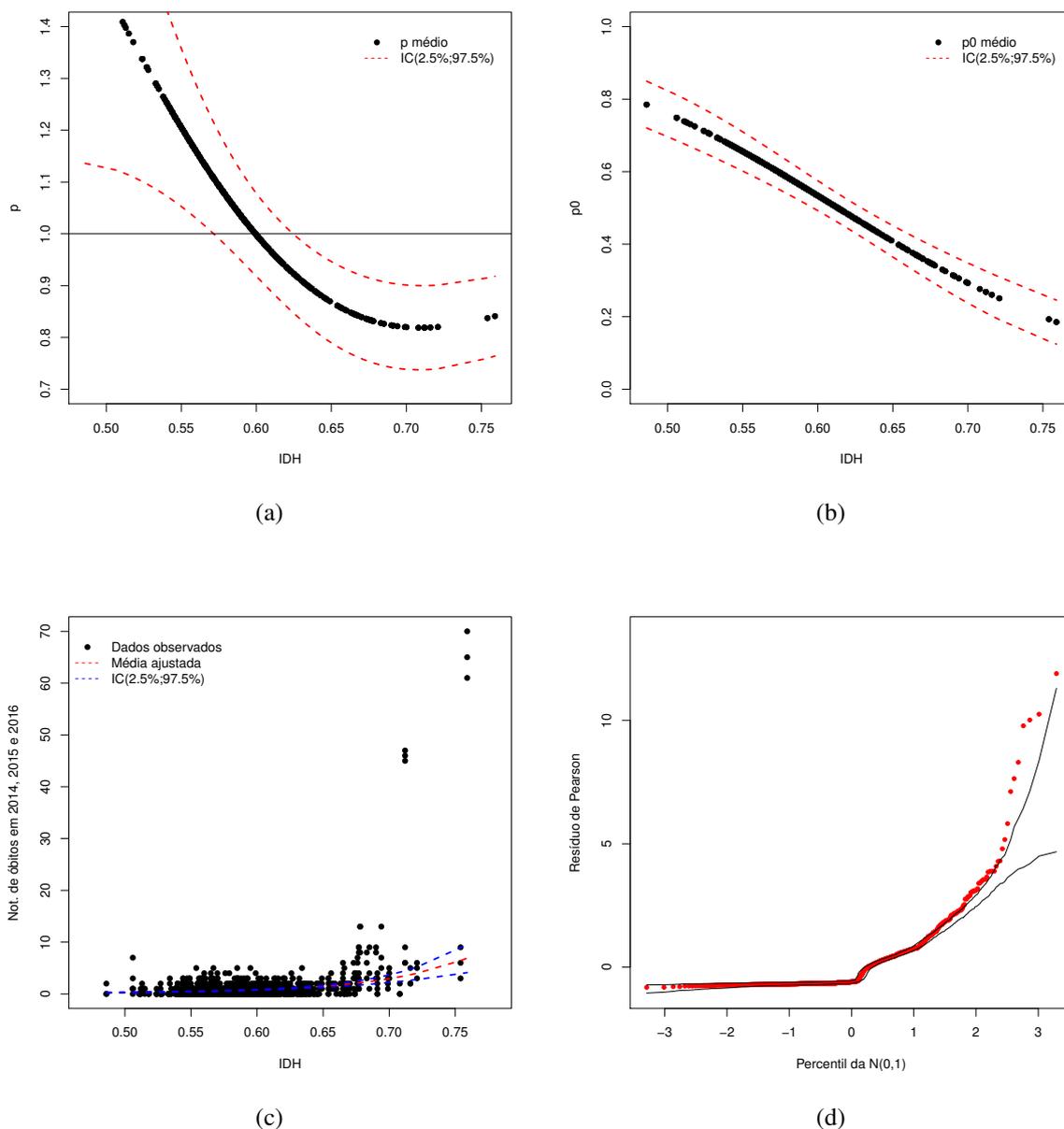


Figura 2 – Modelo ZMP com efeito aleatório Normal: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.

Analisando a [Figura 2 \(a\)](#) observamos que para altos valores de IDH o conjunto de dados é caracterizado como zero-inflacionado, pois o parâmetro p assume valores menores que 1. Entretanto, para observações com valores baixos de IDH, as estimativas de p possuem intervalos de confiança que contém o valor 1, o que caracteriza uma parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Deste modo, outra forma de analisar tal conjunto de dados é dividi-lo em dois subconjuntos. Na [Figura 2 \(b\)](#), podemos observar que a probabilidade de não notificação de óbitos diminuiu a medida que o IDH cresce. Ou seja, a probabilidade de não notificar óbitos de crianças de 1 a 4 anos, em cidades com altos valores de IDH, é baixa. Tal resultado também nos dá indícios da insuficiência do sistema de saúde em detectar óbitos infantis nas cidades com baixo IDH. A [Figura 2 \(c\)](#) ilustra a média ajustada aos dados. Podemos notar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam, e que existem alguns pontos discrepantes. Assim como abordado na [seção 4.3](#), foi realizada uma análise de diagnóstico para verificar a bondade de ajuste do modelo ZMP com efeito aleatório Normal. Analisando a [Figura 2 \(d\)](#), a qual exibe o envelope simulado do modelo com o resíduo de Pearson, podemos observar que existem alguns pontos fora da banda de confiança do envelope. O que nos traz dúvidas sobre o bom ajuste do modelo.

Na [Figura 2 \(c\)](#) é notável a presença de pontos discrepantes. Tais observações correspondem as notificações de óbitos nas cidades de Salvador e em Feira de Santana nos três anos. A fim de testar o impacto de tais observações no ajuste do modelo vamos considerar o conjunto de dados apresentado na [seção 6.1](#) sem as observações referentes as cidades de Salvador (contagens: 61, 65 e 70) e Feira de Santana (contagens: 45, 46 e 47). Removendo tais observações o tamanho da amostra é $n = 415 \times 3 = 1245$.

A [Tabela 16](#) exibe as estimativas, os erros padrão, os intervalos assintóticos com 95% de confiança e os valores-p do ajuste. Verificou-se, a um nível de significância de 5%, que o IDH é significativo para p e para μ . Também é possível perceber que a estimativa de σ^2 apresentada na [Tabela 16](#) é menor do que a estimativa calculada para os dados completos, $\hat{\sigma}^2 = 0.534$.

Tabela 16 – Estimativas dos parâmetros do **modelo ZMP com efeito aleatório Normal**, erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p **considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana**.

Parâmetro	Estimativas	Erro padrão	IC(%)	Valor p
β_0	-6.392	0.771	(-7.904 ; -4.860)	< 0.001
β_1	10.483	1.249	(8.034 ; 12.932)	< 0.001
γ_0	-5.991	0.894	(-7.743 ; -4.238)	< 0.001
γ_1	9.746	1.502	(6.801 ; 12.692)	< 0.001
σ^2	0.353	0.078	(0.200 ; 0.506)	< 0.001

Realizando o teste razão de verossimilhança obtemos a estatística do teste $T = 92.131$ com correspondente valor-p < 0.001. Ou seja, com um nível de confiança de 95%, rejeitamos a hipótese nula de que $\sigma = 0$.

Os gráficos na [Figura 3](#) apresentam: (a) estimativas e limites dos intervalos com 95% de confiança do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório Normal.

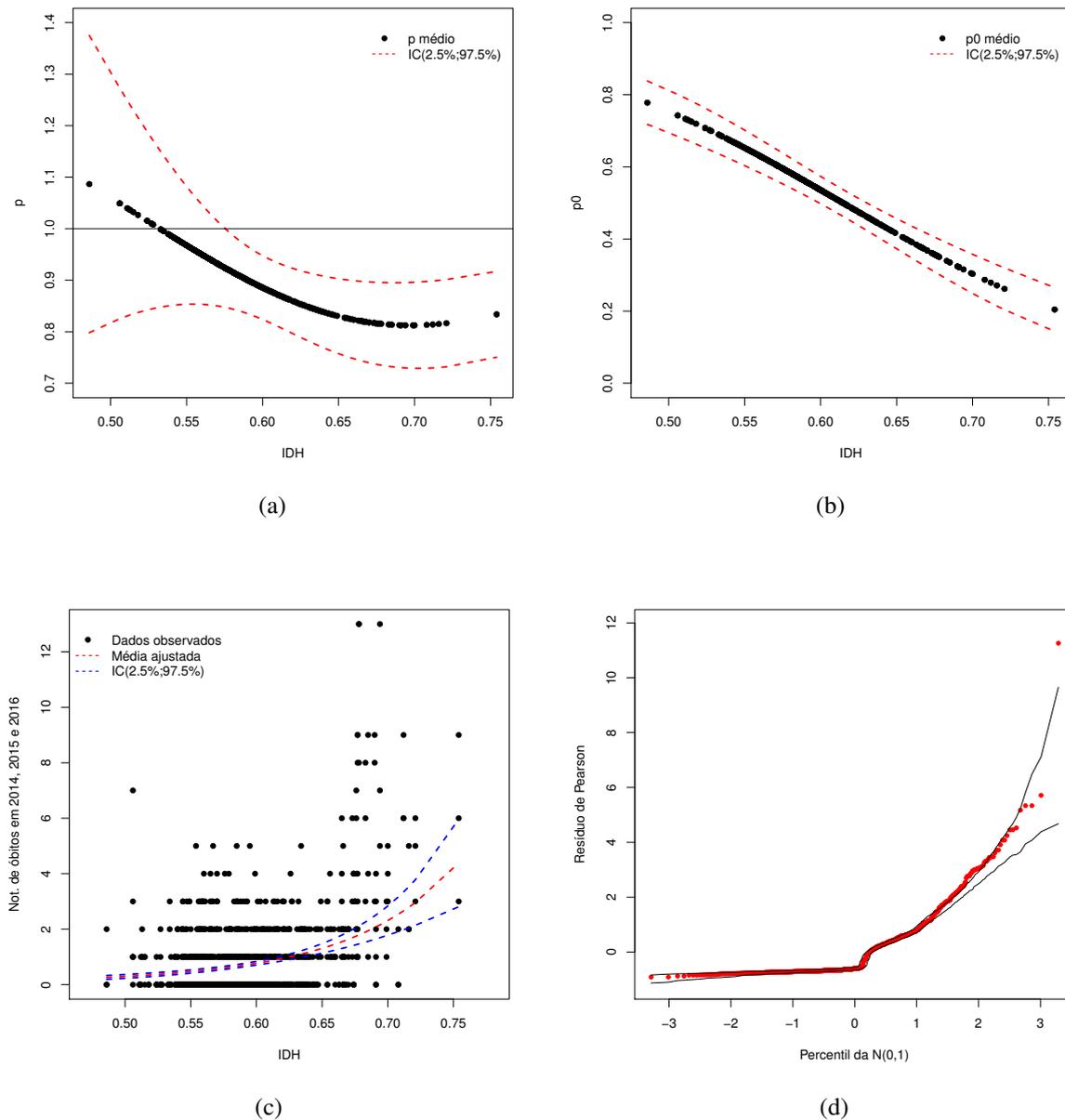


Figura 3 – Modelo ZMP com efeito aleatório Normal considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não-notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório Normal.

Observando a [Figura 3 \(a\)](#), verificamos que para valores de IDH maiores que 0.58, o conjunto de dados é caracterizado como zero-inflacionado, pois o parâmetro p assume valores entre 0 e 1. Porém, para observações com valores de IDH menores que 0.58, os intervalos

de confiança contém o valor 1, o que caracteriza parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na [Figura 3 \(b\)](#), nota-se que a probabilidade de não-notificação de óbitos diminui a medida que o IDH cresce. A [Figura 3 \(c\)](#) exibe a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam. E se comparada a média obtida com os dados completos, na [Figura 2 \(c\)](#), a média obtida sem as notificações referentes as cidades de Salvador e Feira de Santana teve uma leve queda. Observando a [Figura 3 \(d\)](#), a qual exibe o envelope simulado do modelo, nota-se que ainda existem algumas observações atípicas no conjunto de dados. Porém, o resíduo que assume o maior valor se encontra na banda de confiança do envelope. Podemos notar que houve uma melhora considerável quando o comparamos com o envelope simulado com os dados completos na [Figura 2 \(d\)](#). Em geral, verifica-se que o modelo ZMP com efeito aleatório Normal fez um bom ajuste com os dados sem as observações referentes as cidades de Salvador e Feira de Santana.

6.2.2 Modelo ZMP com Efeito Aleatório LGG

Nesta Seção vamos considerar o modelo apresentado em (6.1) com $b_i \sim LGG(0, \sigma, \lambda = -\sigma)$ para modelar o conjunto de dados referente as notificações de óbitos infantis descrito na [seção 6.1](#).

A [Tabela 17](#) exibe as estimativas, os erros padrão, os intervalos de confiança assintóticos com 95% de confiança e os valores-p. Observou-se, a um nível de significância de 5%, que a covariável X_1 é significativa para p e para μ , ou seja, existe uma relação entre o número de notificações de óbitos nas cidades e seus respectivos valores de IDH.

Tabela 17 – Estimativas dos parâmetros do **modelo ZMP com efeito aleatório LGG**, erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p **considerando os dados completos de notificações de óbitos no Estado da Bahia**.

Parâmetro	Estimativas	Erro padrão	IC(%)	Valor p
β_0	-7.597	0.870	(-9.301 ; -5.892)	< 0.001
β_1	12.176	1.400	(9.433 ; 14.919)	< 0.001
γ_0	-6.237	0.883	(-7.969 ; -4.506)	< 0.001
γ_1	10.169	1.483	(7.262 ; 13.077)	< 0.001
σ^2	0.338	0.053	(0.234 ; 0.442)	< 0.001
λ	-0.582	0.046	(-0.671; -0.492)	-

Com o objetivo de mostrar a significância do parâmetro σ no ajuste do modelo vamos utilizar o teste razão de verossimilhança. Sendo a estatística do teste $T = 578.116$ com correspondente valor-p < 0.001, a um nível de confiança de 95%, rejeitamos a hipótese nula. Ou seja, o efeito aleatório é significativo.

Os gráficos exibidos na [Figura 4](#) apresentam: (a) estimativas e limites dos intervalos com 95% de confiança do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de

confiança da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório LGG.

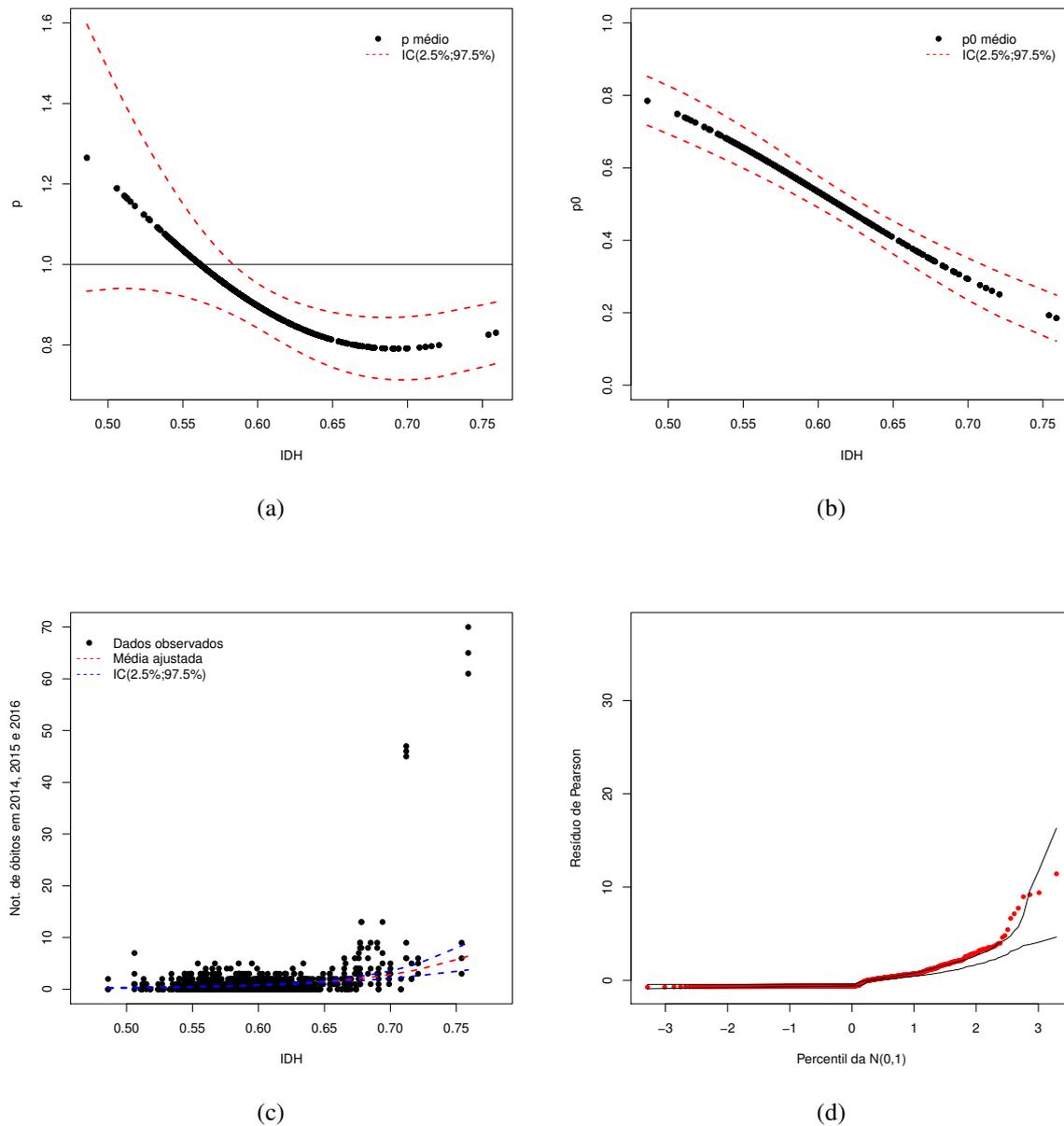


Figura 4 – Modelo ZMP com efeito aleatório LGG: (a) Estimativas e limites dos intervalos com 95% de confiança do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.

Analisando a Figura 4 (a) podemos notar que para altos valores de IDH o conjunto de dados é caracterizado como zero-inflacionado pois as estimativas do parâmetro p assumem valores entre 0 e 1. Entretanto, para observações com valores de IDH menores que 0.6, as estimativas de p assumem valores maiores que 1, o que caracteriza uma parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na Figura 4 (b), podemos observar que a probabilidade de não notificação de óbitos diminui a medida que o IDH cresce. Ou seja, a

probabilidade de não notificar óbitos de crianças de 1 a 4 anos, em cidades com altos valores de IDH, é baixa. A [Figura 4 \(c\)](#) ilustra a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam, e que existem alguns pontos discrepantes. Para a análise de diagnóstico utilizamos o envelope simulado com os resíduos de Pearson. Analisando a [Figura 4 \(d\)](#), podemos observar que existem alguns pontos fora da banda de confiança do envelope. Entretanto, se o compararmos com o envelope obtido do modelo ZMP com efeito aleatório Normal na [Figura 2 \(d\)](#), podemos verificar que, aparentemente, o modelo com efeito aleatório LGG se ajustou melhor aos dados.

Retirando novamente as observações discrepantes correspondentes as notificações de óbitos nas cidades de Salvador e em Feira de Santana para realizar o ajuste do modelo ZMP com efeito aleatório LGG, obtemos os seguintes resultados.

Na [Tabela 18](#) é apresentado as estimativas, os erros padrão, os intervalos assintóticos com 95% de confiança e os valores-p do ajuste. Podemos observar que, a um nível de significância de 5%, o IDH é significativo para p e para μ . Além disso, utilizando o teste razão de verossimilhança obtemos a estatística do teste $T = 92.636$ com correspondente valor-p < 0.001 . Ou seja, com um nível de confiança de 95%, o efeito aleatório é significativo no modelo.

Tabela 18 – Estimativas dos parâmetros do **modelo ZMP com efeito aleatório LGG**, erros padrão, intervalos de confiança assintóticos com 95% de confiança e valores-p **considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana.**

Parâmetro	Estimativas	Erro padrão	IC(%)	Valor p
β_0	-6.582	0.818	(-8.185 ; -4.979)	< 0.001
β_1	10.620	1.316	(8.040 ; 13.198)	< 0.001
γ_0	-5.991	0.894	(-7.743 ; -4.238)	< 0.001
γ_1	9.746	1.503	(6.801 ; 12.692)	< 0.001
σ^2	0.258	0.050	(0.161 ; 0.355)	< 0.001
λ	-0.508	0.048	(-0.603 ; -0.413)	-

Os gráficos na [Figura 5](#) apresentam: (a) estimativas e limites dos intervalos com 95% de confiança do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de confiança da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) envelope simulado com o resíduo de Pearson do modelo ZMP com efeito aleatório LGG.

Observando a [Figura 5 \(a\)](#), verificamos que para valores de IDH maiores que 0.57, o conjunto de dados é caracterizado como zero-inflacionado, pois o parâmetro p assume valores entre 0 e 1. Porém, para observações com valores de IDH menores que 0.57, os intervalos de confiança contêm o valor 1, o que caracteriza parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na [Figura 5 \(b\)](#), nota-se que a probabilidade de não notificação de óbitos diminui a medida que o IDH cresce. A [Figura 5 \(c\)](#) exibe a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam. Observando a [Figura 5 \(d\)](#), a qual exibe o envelope simulado do modelo,

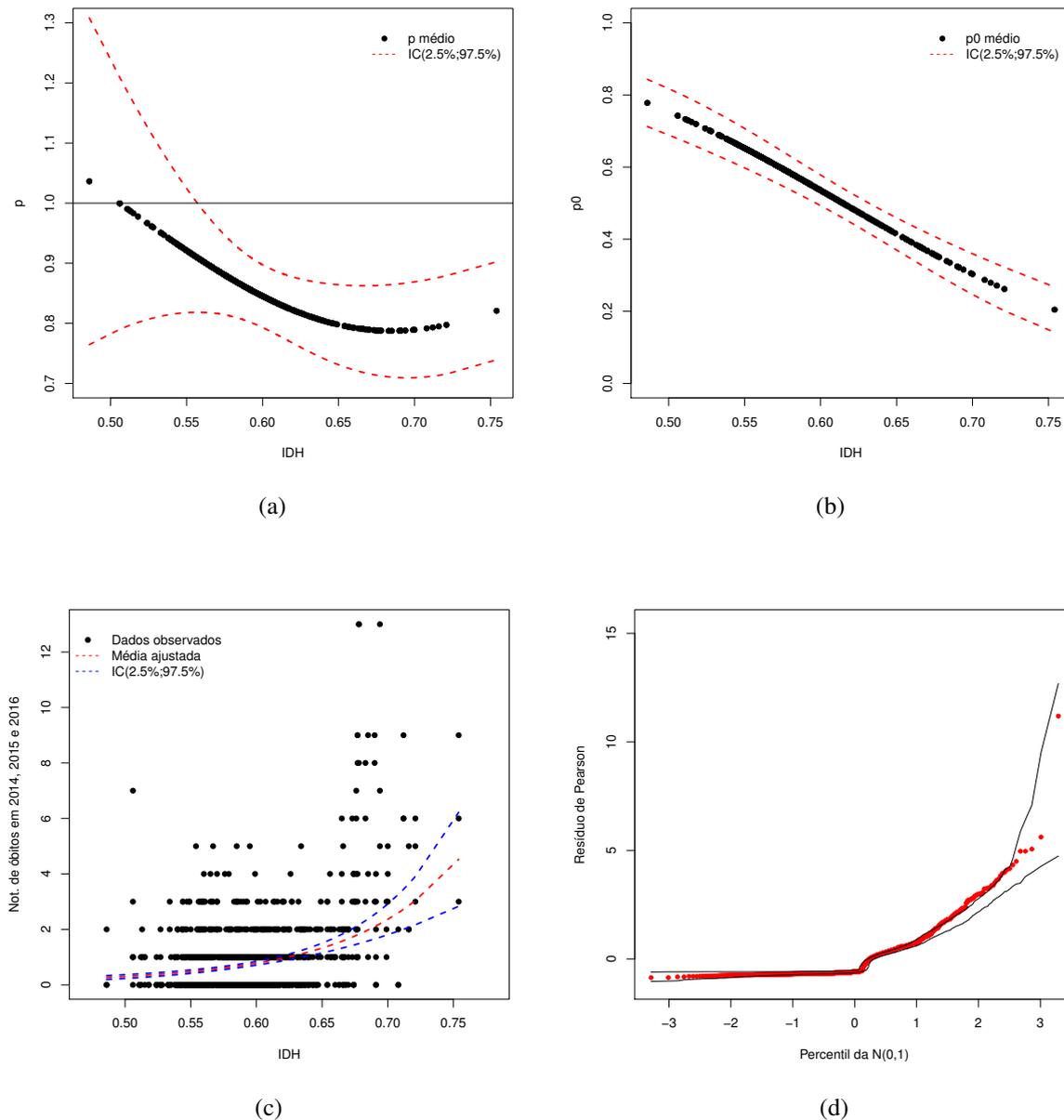


Figura 5 – Modelo ZMP com efeito aleatório LGG considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativa e limite do intervalo com 95% de confiança do parâmetro p . (b) Estimativa e limite do intervalo com 95% de confiança da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Envelope simulado com o resíduo de Pearson.

nota-se que existem poucas observações fora da banda do envelope. Houve uma melhora notável quando o comparamos com o envelope simulado com os dados completos na Figura 4 (d), pois existem menos pontos fora da banda do envelope. No geral, verifica-se que o modelo ZMP com efeito aleatório LGG fez um bom ajuste com os dados sem as observações referentes às cidades de Salvador e Feira de Santana.

A Tabela 19 mostra os critérios de seleção de modelo AIC e BIC para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos

infantis apresentado na [seção 6.1](#). Vamos considerar os dados com as notificações completas (dados completos) e os dados sem as notificações referentes às cidades de Salvador e Feira de Santana (dados incompletos). Os valores do AIC e BIC reforçam os resultados obtidos com os gráficos dos envelopes simulados na [Figura 2 \(d\)](#), [Figura 3 \(d\)](#), [Figura 4 \(d\)](#) e [Figura 5 \(d\)](#). Os modelos ZMP com efeito Normal e com efeito LGG se ajustam melhor ao conjunto de dados incompleto. Entretanto, quando consideramos o conjunto de dados completo, o modelo ZMP com efeito LGG se ajusta melhor do que o modelo ZMP com efeito Normal. Tal resultado nos mostra que o efeito aleatório LGG lida melhor com dados assimétricos.

Tabela 19 – Valores do AIC e BIC calculados para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos infantis.

	ZMP com efeito Normal		ZMP com efeito LGG	
	AIC	BIC	AIC	BIC
Dados completos	3081.139	3106.797	3074.436	3100.094
Dados incompletos	3004.539	3030.173	3004.034	3029.668

6.3 Análise com a Abordagem Bayesiana

A partir do procedimento iterativo de amostragem Metropolis-Hastings, utilizando o programa Jags, foram geradas duas cadeias de tamanho 50.000 para cada parâmetro, com um período burn in de 10.000 iterações em cada cadeia e saltos de tamanho 10. Resultando em cadeias de tamanho 8.000 para cada parâmetro. A convergência das cadeias foi verificada através do critério de Gelman Rubin, e as densidades a priori utilizadas foram: $\beta_0 \sim N(0, 1000)$, $\beta_1 \sim N(0, 1000)$, $\gamma_0 \sim N(0, 1000)$, $\gamma_1 \sim N(0, 1000)$ e $\tau \sim \text{Gama}(0.01, 0.001)$ em que $\tau = 1/\sigma^2$.

6.3.1 Modelo ZMP com Efeito Aleatório Normal

Consideremos o modelo apresentado em (6.1) com $b_i \sim N(0, \sigma^2)$ para modelar o conjunto de dados referente as notificações de óbitos infantis descrito na [seção 6.1](#).

A [Tabela 20](#) apresenta: a média, a mediana e o desvio padrão da densidade a posteriori, as quais foram obtidas pelo método de Monte Carlo. O intervalo de 95% de credibilidade, estimado a partir dos quantis da amostra gerada. E o critério de Gelman Rubin (GR), o qual indica a convergência da cadeia. Podemos verificar que o valor zero não pertence a nenhum dos intervalos de credibilidade, o que indica que todos os parâmetros são significativos no modelo. Ou seja, o valor do IDH tem uma relação com o número de notificações de óbitos nas cidades. Os histogramas das cadeias são exibidos na [Figura 11](#) do [Apêndice B](#).

Os gráficos exibidos na [Figura 2](#) apresentam: (a) estimativas e limites dos intervalos com 95% de credibilidade para o parâmetro p ; (b) estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos; (c) média ajustada em função do

Tabela 20 – Sumário a posteriori para β_0 , β_1 , γ_0 , γ_1 e σ^2 utilizando o **modelo ZMP com efeito aleatório Normal e considerando os dados completos de notificações de óbitos no Estado da Bahia.**

Parâmetros	Média	Mediana	Desvio-Padrão	IC(95%)	GR
β_0	-8.099	-8.117	0.773	(-9.656; -6.586)	1.027
β_1	13.199	13.236	1.247	(10.752; 15.674)	1.022
γ_0	-6.233	-6.262	0.947	(-8.091; -4.252)	1.005
γ_1	10.164	10.207	1.591	(6.820; 13.298)	1.003
σ^2	0.530	0.520	0.097	(0.366; 0.744)	1.001

IDH; (d) resíduo de Pearson do modelo ZMP com efeito aleatório Normal. Mais detalhes sobre o cálculo dos intervalos de confiança são dados no [Apêndice A](#).

Analisando a [Figura 6 \(a\)](#) observamos que para altos valores de IDH o conjunto de dados é caracterizado como zero-inflacionado, pois as estimativas de p assumem valores menores que 1. Entretanto, para observações com valores de IDH baixos, as estimativas de p possuem intervalos de credibilidade que contém o valor 1, o que caracteriza uma parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na [Figura 6 \(b\)](#), podemos observar que a probabilidade de não notificação de óbitos diminuiu a medida que o IDH cresce. Ou seja, a probabilidade de não notificar óbitos de crianças de 1 a 4 anos, em cidades com altos valores de IDH, é baixa. Tal resultado também nos dá indícios da insuficiência do sistema de saúde em detectar óbitos infantis nos municípios com baixo IDH. A [Figura 6 \(c\)](#) ilustra a média ajustada aos dados. Podemos notar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam, e que existem alguns pontos discrepantes. Analisando a [Figura 6 \(d\)](#) com os resíduos de Pearson, podemos observar que existem alguns resíduos com valores altos, alcançando quase 20.

Assim como no caso da aplicação clássica, na [Figura 6 \(c\)](#) é notável a presença de pontos discrepantes. A fim de testar o impacto de tais observações no ajuste do modelo vamos considerar o conjunto de dados apresentado na [seção 6.1](#) sem as observações referentes aos municípios de Salvador e Feira de Santana.

A [Tabela 21](#) exhibe a média, a mediana e o desvio padrão da densidade a posteriori juntamente com seus intervalos de 95% de credibilidade, e o critério de Gelman Rubin (GR). Podemos verificar que todos os parâmetros do modelo são significativos pois o valor zero não pertence a nenhum dos intervalos de credibilidade. Além disso, os valores do critério de GR próximos de 1 indicam que as cadeias são convergentes.

Os gráficos na [Figura 7](#) apresentam: (a) estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) resíduos de Pearson do modelo ZMP com efeito aleatório Normal.

Observando a [Figura 7 \(a\)](#), verificamos que para valores de IDH maiores que 0.56, o conjunto de dados é caracterizado como zero-inflacionado, pois o parâmetro p assume valores

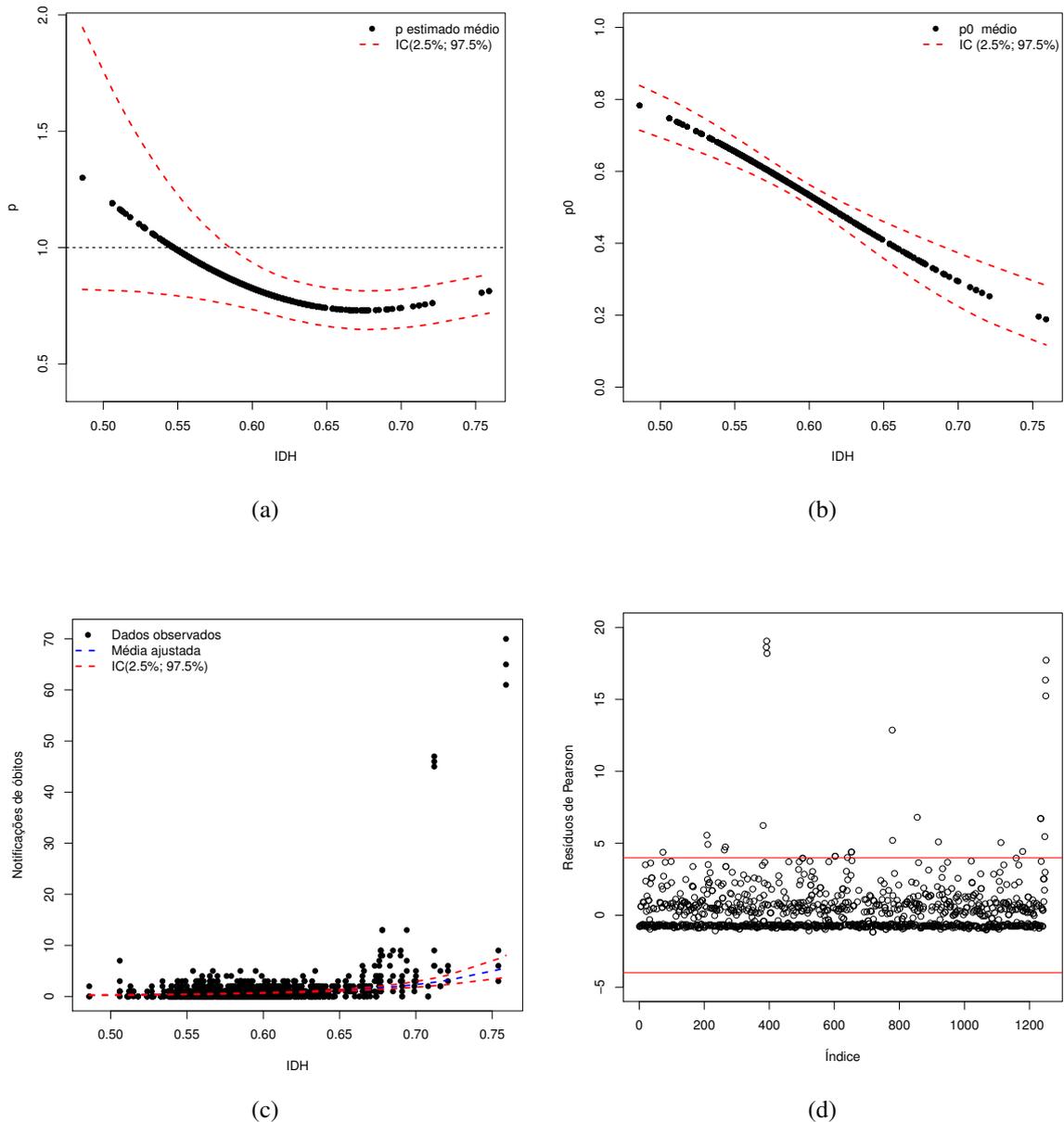


Figura 6 – Modelo ZMP com efeito aleatório Normal: (a) Estimativas e os intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson.

entre 0 e 1. Porém, para observações com valores de IDH menores que 0.56, os intervalos de credibilidade das estimativas de p contêm o valor 1, o que caracteriza parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na Figura 7 (b), nota-se que a probabilidade de não-notificação de óbitos diminui a medida que o IDH cresce. A Figura 7 (c) exibe a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam. E se comparada a média obtida com os dados completos, na Figura 6 (c), a média obtida sem as notificações referentes as cidades de Salvador e Feira de Santana teve uma leve queda. Observando a Figura 7 (d), a qual exibe os resíduos do modelo, nota-se que a

Tabela 21 – Sumário a posteriori para β_0 , β_1 , γ_0 , γ_1 e σ^2 utilizando o **modelo ZMP com efeito aleatório Normal e considerando os dados sem as notificações de óbitos referentes a Salvador e Feira de Santana.**

Parâmetros	Média	Mediana	Desvio-Padrão	IC(95%)	GR
β_0	-6.360	-6.372	0.787	(-7.836; -4.773)	1.054
β_1	10.442	10.469	1.276	(7.864; 12.840)	1.045
γ_0	-6.074	-6.080	0.936	(-7.921; -4.201)	1.021
γ_1	9.886	9.896	1.570	(6.726; 12.986)	1.025
σ^2	0.342	0.335	0.076	(0.215; 0.508)	1.003

grande maioria dos pontos assumem valores de resíduo em torno de zero.

6.3.2 Modelo ZMP com Efeito Aleatório LGG

Nesta Seção vamos considerar o modelo apresentado em (6.1) com $b_i \sim LGG(0, \sigma, \lambda = -\sigma)$ para modelar o conjunto de dados referente as notificações de óbitos infantis descrito na seção 6.1.

A Tabela 22 exibe a média, a mediana, o desvio padrão a posteriori, os intervalos com 95% de credibilidade e o critério GR. Observou-se que a covariável X_1 é significativa para p e para μ pois o valor zero não pertence a nenhum dos intervalos de credibilidade. Ou seja, existe uma relação entre o número de notificações de óbitos nas cidades e seus respectivos valores de IDH. Os histogramas das cadeias são exibidos na Figura 12 do Apêndice B.

Tabela 22 – Sumário a posteriori para β_0 , β_1 , γ_0 , γ_1 , σ^2 e λ utilizando o **modelo ZMP com efeito aleatório LGG e considerando os dados completos de notificações de óbitos no Estado da Bahia.**

Parâmetros	Média	Mediana	Desvio-Padrão	IC(95%)	GR
β_0	-8.583	-8.554	0.721	(-9.986; -7.243)	1.011
β_1	14.456	14.405	1.163	(12.278; 16.716)	1.012
γ_0	-6.164	-6.178	0.835	(-7.783; -4.549)	1.006
γ_1	10.046	10.070	1.403	(7.321; 12.779)	1.004
σ^2	0.646	0.639	0.108	(0.456; 0.876)	1.001
λ	-0.801	-0.799	0.067	(-0.936; -0.675)	1.001

Os gráficos exibidos na Figura 8 apresentam: (a) estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) resíduo de Pearson do modelo ZMP com efeito aleatório LGG.

Analisando a Figura 8 (a) podemos notar que para altos valores de IDH o conjunto de dados é caracterizado como zero-inflacionado pois as estimativas do parâmetro p assumem valores entre 0 e 1. Entretanto, para observações com valores de IDH menores que 0.55, os intervalos das estimativas de p contém o valor 1, o que caracteriza que uma parte do conjunto de dados é proveniente da distribuição de Poisson tradicional. Na Figura 8 (b), podemos observar

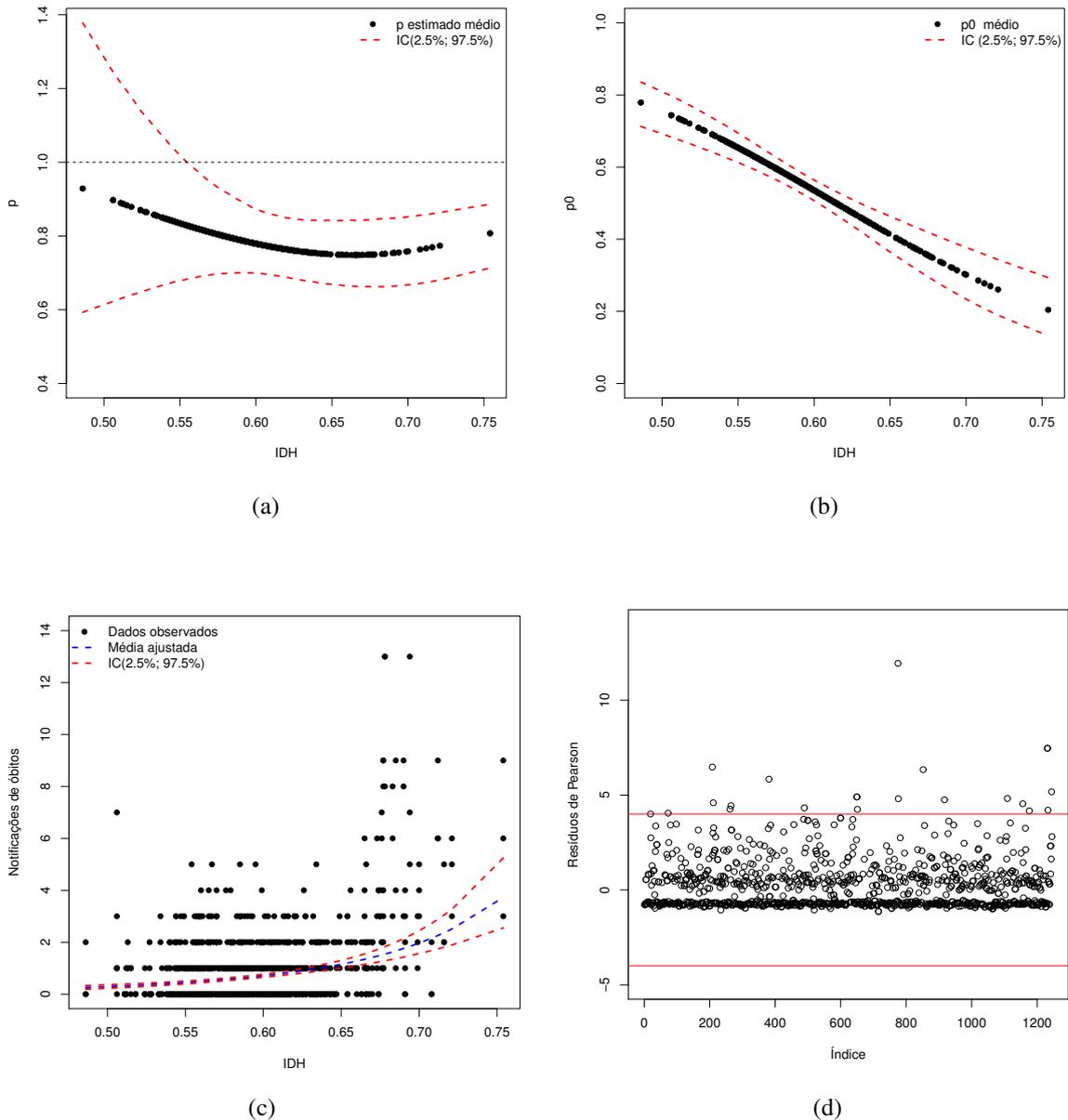


Figura 7 – Modelo ZMP com efeito aleatório Normal considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de c da probabilidade de não-notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson do modelo ZMP com efeito aleatório Normal.

que a probabilidade de não notificação de óbitos diminui a medida que o IDH cresce. Ou seja, a probabilidade de não notificar óbitos de crianças de 1 a 4 anos, em cidades com altos valores de IDH, é baixa. A Figura 8 (c) ilustra a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam, e que existem alguns pontos discrepantes. Para a análise de diagnóstico utilizamos os resíduos de Pearson. Analisando a Figura 8 (d), podemos observar que existem alguns resíduos com valores próximos de 12. Entretanto, se o compararmos com os resíduos obtidos do modelo ZMP com efeito aleatório

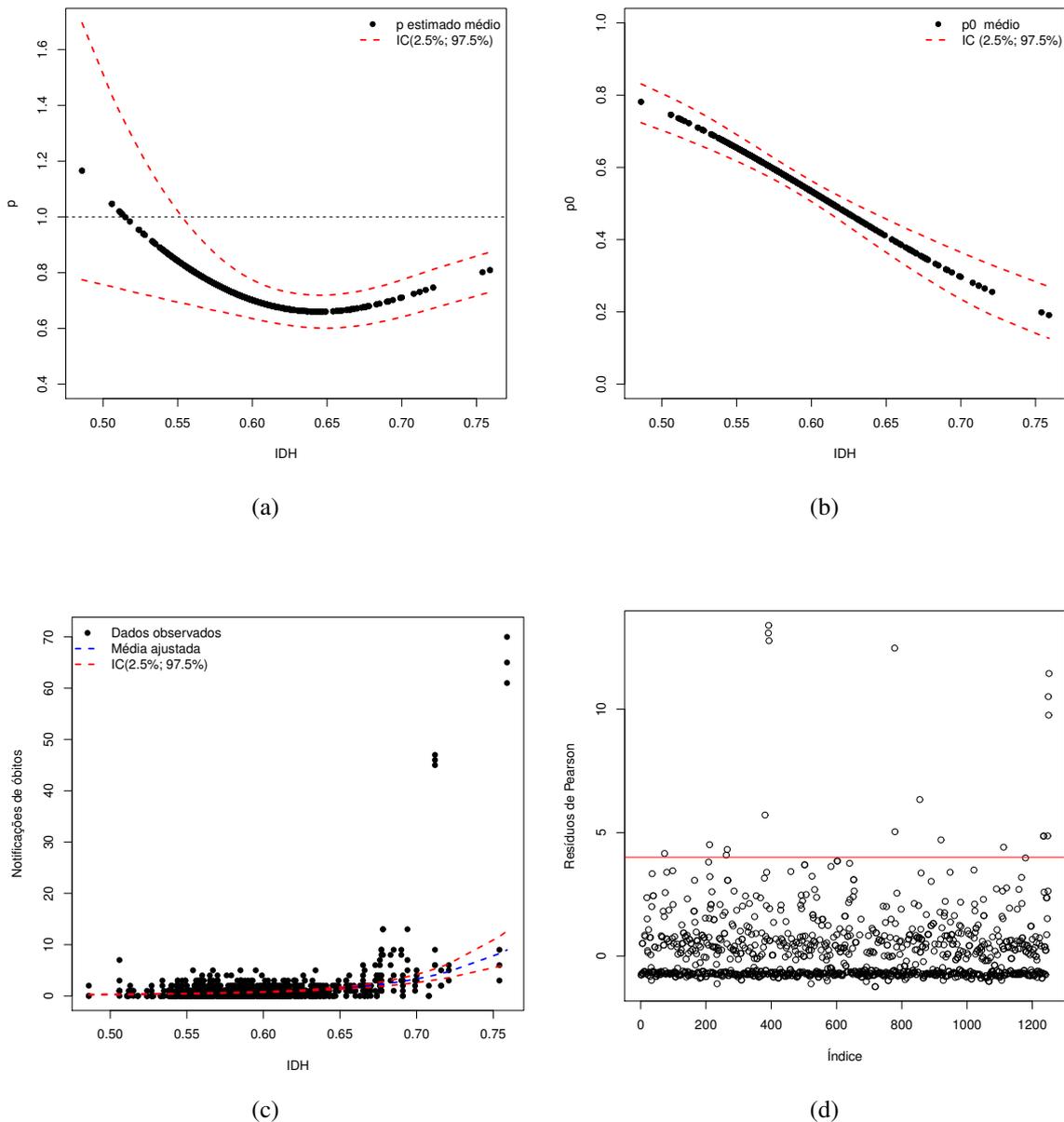


Figura 8 – Modelo ZMP com efeito aleatório LGG: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson.

Normal na Figura 6 (d), podemos verificar que o modelo com efeito aleatório LGG possui resíduos com valores menores.

Retirando novamente as observações discrepantes correspondentes as notificações de óbitos nas cidades de Salvador e em Feira de Santana para realizar o ajuste do modelo ZMP com efeito aleatório LGG, obtemos os seguintes resultados.

Na Tabela 23 é apresentado a média, a mediana e o desvio padrão da densidade a posteriori juntamente com seus intervalos de 95% de credibilidade, e o critério GR. Podemos observar que todos os parâmetros do modelo são significativos, inclusive o efeito aleatório, pois

o valor zero não pertence a nenhum dos intervalos de credibilidade. Além disso, os valores do critério de GR próximos de 1 indicam que as cadeias são convergentes.

Tabela 23 – Sumário a posteriori para $\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma^2$ e λ utilizando o **modelo ZMP com efeito aleatório LGG e considerando os dados sem as notificações de óbitos referentes a Salvador e Feira de Santana.**

Parâmetros	Média	Mediana	Desvio-Padrão	IC(95%)	GR
β_0	-6.222	-6.247	0.682	(-7.555; -4.914)	1.004
β_1	10.499	10.541	1.106	(8.411; 12.675)	1.003
γ_0	-6.119	-6.127	0.893	(-7.791; -4.334)	1.019
γ_1	9.959	9.980	1.498	(6.976; 12.766)	1.024
σ^2	0.376	0.369	0.086	(0.230; 0.565)	1.002
λ	-0.609	-0.607	0.070	(-0.751; -0.480)	1.001

Os gráficos na [Figura 9](#) apresentam: (a) estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p ; (b) estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos; (c) média ajustada em função do IDH; (d) resíduos de Pearson do modelo ZMP com efeito aleatório LGG.

Observando a [Figura 9 \(a\)](#), verificamos que para valores de IDH maiores que 0.52, o conjunto de dados é caracterizado como zero-inflacionado, pois o parâmetro p assume valores entre 0 e 1. Porém, para observações com valores de IDH menores que 0.52, os intervalos de confiança contêm o valor 1, o que caracteriza parte do conjunto de dados proveniente da distribuição de Poisson tradicional. Na [Figura 9 \(b\)](#), nota-se que a probabilidade de não notificação de óbitos diminui a medida que o IDH cresce. A [Figura 9 \(c\)](#) exibe a média ajustada aos dados. Podemos observar que a média de notificação de óbitos cresce conforme os valores de IDH aumentam. Observando a [Figura 9 \(d\)](#), com os resíduos de Pearson, nota-se que existe apenas uma observação com resíduo alto. Houve uma melhora quando o comparamos com o gráfico dos resíduos com os dados completos na [Figura 8 \(d\)](#). No geral, verifica-se que o modelo ZMP com efeito aleatório LGG fez um bom ajuste com os dados sem as observações referentes às cidades de Salvador e Feira de Santana.

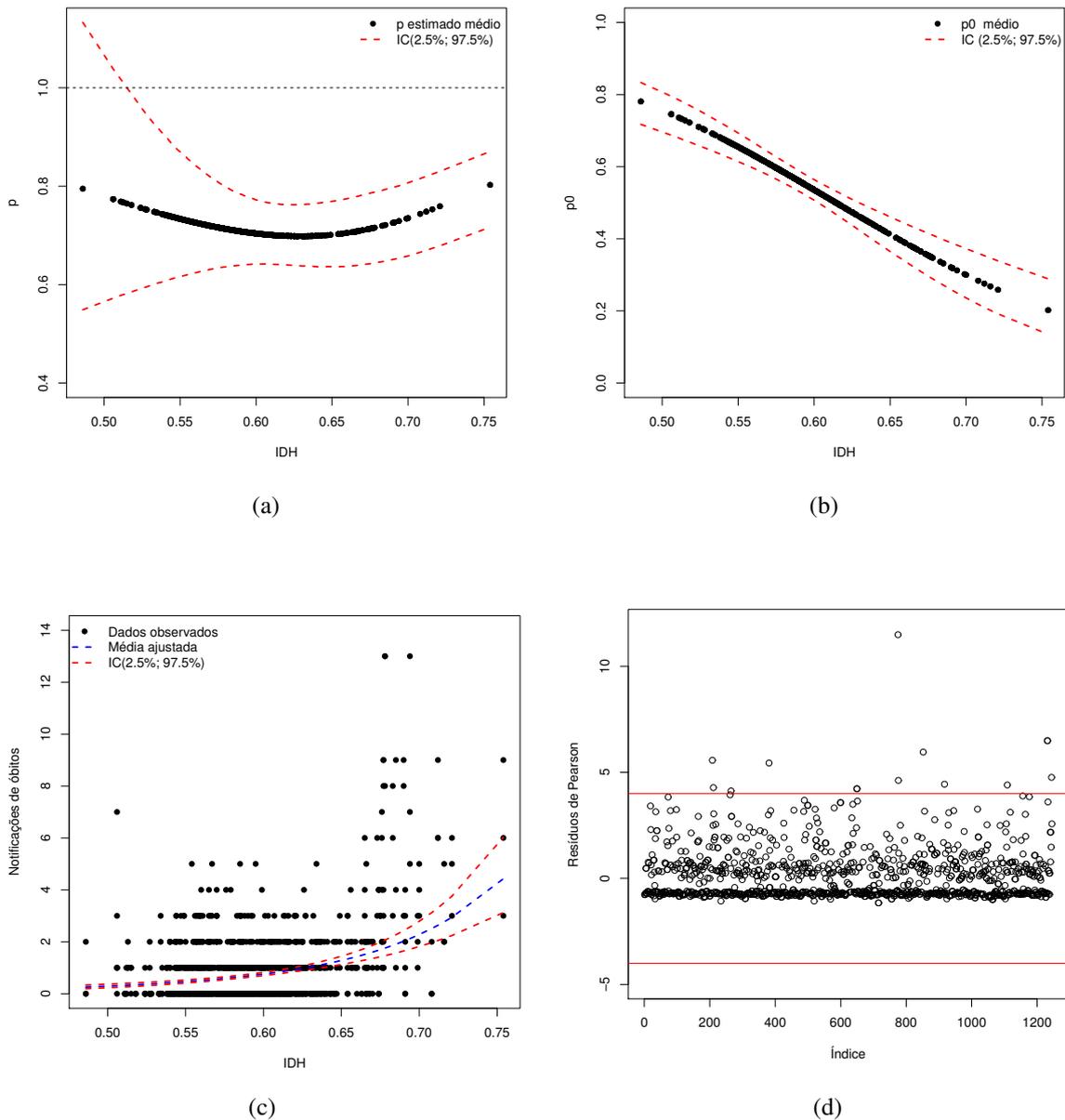


Figura 9 – Modelo ZMP com efeito aleatório LGG considerando os dados de notificação de óbitos no Estado da Bahia sem as observações referentes às cidades de Salvador e Feira de Santana: (a) Estimativas e limites dos intervalos com 95% de credibilidade do parâmetro p . (b) Estimativas e limites dos intervalos com 95% de credibilidade da probabilidade de não notificação de óbitos. (c) Média ajustada em função do IDH. (d) Resíduos de Pearson do modelo ZMP com efeito aleatório LGG.

A [Tabela 24](#) mostra os critérios de seleção de modelo EAIC, EBIC e DIC para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos infantis apresentado na [seção 6.1](#). Considerando os dados com as notificações completas (dados completos) e os dados sem as notificações referentes às cidades de Salvador e Feira de Santana (dados incompletos) obtemos:

Tabela 24 – Valores do EAIC, BIC e DIC calculados para os modelos ZMP com efeito aleatório Normal e efeito aleatório LGG ajustados aos dados de notificação de óbitos infantis.

	ZMP com efeito Normal			ZMP com efeito LGG		
	EAIC	EBIC	DIC	EAIC	EBIC	DIC
Dados completos	4166.437	4192.096	4164.781	3876.454	3907.245	3874.017
Dados incompletos	3155.799	3181.434	3152.466	3117.629	3148.391	3111.686

Assim como na abordagem clássica, comparando os valores do EAIC, EBIC e DIC na [Tabela 24](#) e também considerando os gráficos dos resíduos na [Figura 6 \(d\)](#), [Figura 7 \(d\)](#), [Figura 9 \(d\)](#) e [Figura 8 \(d\)](#), podemos concluir que os modelos ZMP com efeito Normal e com efeito LGG se ajustam melhor ao conjunto de dados incompleto. Entretanto, quando comparamos o modelo ZMP com efeito Normal e o modelo ZMP com efeito LGG, podemos notar que o modelo ZMP com efeito LGG se ajusta melhor aos dados completos e incompletos.

CONCLUSÃO

O modelo Poisson Zero-Modificado com efeito aleatório nos permite trabalhar com dados de contagem que possuem excesso (ou *deficit*) de zeros, e que sejam correlacionados no tempo. Sua maior vantagem é que ele modela dados longitudinais sem o conhecimento prévio da sua característica de inflação ou deflação de zeros, e o efeito aleatório é responsável por induzir uma correlação entre os dados e modelar a variabilidade intrínseca de cada indivíduo.

Neste trabalho propomos os modelos ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG. O uso do efeito aleatório Normal e LGG no modelo ZMP são viáveis para modelar dados longitudinais. Sendo o efeito LGG uma generalização do efeito Normal, este pode ser utilizado em situações atípicas em que o efeito Normal não é o mais adequado. Por exemplo: para dados que possuem assimetria á direita, o efeito LGG pode ser mais apropriado para ser usado no modelo.

Ao longo dessa dissertação exibimos algumas características dos modelo ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG e apresentamos uma abordagem clássica e bayesiana para a estimação dos parâmetros dos modelos. Na abordagem clássica, assim como mencionado ao longo do texto, devido a complexidade para resolver algumas integrais e o custo computacional, encontramos certas dificuldades para calcular os estimadores de máxima verossimilhança e os intervalos de confiança assintóticos. Foi realizado um estudo de simulação a fim de investigar a performance dos estimadores e tais modelos se mostraram eficazes.

Ajustamos os modelos ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG considerando um conjunto de dados da contagem de notificações de óbitos infantis no Estado da Bahia, ao longo de 2014, 2015 e 2016, através das abordagens clássica e bayesiana. A partir dos resultados podemos concluir que ambos os modelos se ajustam bem ao conjunto de dados, entretanto, o modelo ZMP com efeito LGG é mais eficaz para lidar com dados assimétricos. Além disso, verificamos a importância do IDH ao explicar as notificações de óbitos nas cidades.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974. Citado na página 40.
- ATKINSON, A. C. Two graphical displays for outlying and influential observations in regression. **Biometrika**, v. 68, n. 1, p. 13–20, 1981. Citado na página 39.
- BONAT, W. H.; JR, P. J. R.; KRAINSKI, E. T.; ZEVIANI, W. M. **Métodos Computacionais em Inferência Estatística**. [S.l.], 2012. Disponível em: <<http://www.leg.ufpr.br/mcie>>. Citado 2 vezes nas páginas 34 e 36.
- BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, v. 88, n. 421, p. 9–25, 1993. Citado na página 23.
- BROOKS, S. Discussion of the paper bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society**, v. 64, p. 616–618, 2002. Citado na página 40.
- CARVALHO, H. G. de; ACHCAR, J. A. Uma análise bayesiana para dados de contagem longitudinais na presença de uma covariável com efeito espaço-temporal. **Revista Brasileira de Biometria**, v. 31, n. 4, p. 617–630, 2013. Citado na página 23.
- CHIB S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The American Statistician**, v. 49, n. 4, p. 327–335, 1995. Citado na página 38.
- COHEN, A. C. An extension of a truncated poisson distribution. **Biometrics**, v. 16, n. 3, p. 446–450, 1960. Citado na página 21.
- CONCEIÇÃO, K. S. **Modelos Série de Potência Zero-Modificados**. Tese (Doutorado em Estatística) — Universidade Federal de São Carlos, São Carlos – SP, 2013. Citado 2 vezes nas páginas 22 e 38.
- CONCEIÇÃO, K. S.; ANDRADE, M. G.; LOUZADA, F. Zero-modified poisson model: Bayesian approach, influence diagnostics, and an application to a brazilian leptospirosis notification data. **Biometrical Journal**, v. 55, n. 5, p. 661–678, 2013. Citado na página 21.
- CONCEIÇÃO, K. S.; LOUZADA, F.; ANDRADE, M. G.; HELOU, E. S. Zero-modified power series distribution and its hurdle distribution version. **Journal of Statistical Computation and Simulation**, v. 87, n. 9, p. 1842–1862, 2017. Citado 3 vezes nas páginas 25, 27 e 29.
- CORDEIRO, G. M.; ANDRADE, M. G.; CASTRO, M. D. Power series generalized nonlinear models. **Computational Statistics & Data Analysis**, v. 53, n. 4, p. 1155–1166, 2009. Citado na página 22.
- COSTA, S. C. D. **Modelos lineares generalizados mistos para dados longitudinais**. Tese (Doutorado em Estatística) — Universidade de São Paulo, ESALQ, Piracicaba – SP, 2017. Citado na página 23.

DEMPSTER, A. P. The direct use of likelihood for significance testing. **Statistics and Computing**, v. 7, n. 4, p. 247–252, 1997. Citado na página 40.

DIETZ, E.; BÖHNING, D. On estimation of the poisson parameter in zero-modified poisson models. **Computational Statistics & Data Analysis**, v. 34, n. 4, p. 441–459, 2000. Citado na página 21.

FABIO, L. C.; PAULA, G. A.; CASTRO, M. D. A poisson mixed model with nonnormal random effect distribution. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1499–1510, 2012. Citado 4 vezes nas páginas 22, 23, 32 e 36.

FROME, E. L. The analysis of rates using poisson regression models. **Biometrics**, p. 665–674, 1983. Citado na página 21.

FROME, E. L.; CHECKOWAY, H. Use of poisson regression models in estimating incidence rates and ratios. **American Journal of Epidemiology**, v. 121, n. 2, p. 309–323, 1985. Citado na página 21.

GARDNER, W.; MULVEY, E. P.; SHAW, E. C. Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. **Psychological Bulletin**, v. 118, n. 3, p. 392, 1995. Citado na página 21.

GELMAN A.; RUBIN, D. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, n. 4, p. 457–472, 1992. Citado na página 39.

HARDIN, J. W.; HILBE, J. M. **Generalized estimating equations**. [S.l.]: Chapman Hall/CRC:London, 2002. Citado na página 22.

HAUSER, L. Curva de crescimento usando o modelo de coeficientes aleatórios: uma aplicação na progressão da doença de machado-joseph. **Revista do Hospital das Clínicas de Porto Alegre**, v. 29, n. 1, p. 5–7, 2009. Citado na página 22.

JOSHUA, S. C.; GARBER, N. J. Estimating truck accident rate and involvements using linear and poisson regression models. **Transportation Planning and Technology**, v. 15, n. 1, p. 41–58, 1990. Citado na página 21.

KRUSCHKE, J. **Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan**. [S.l.]: Academic Press, 2014. Citado na página 39.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, v. 33, n. 4, p. 963–974, 1982. Citado na página 23.

LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, v. 34, n. 1, p. 1–14, 1992. Citado na página 21.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. v. 362. Citado na página 32.

LESAFFRE, E.; LAWSON, J. W. A. B. **Bayesian Biostatistics**. [S.l.]: Taylor & Francis, 2012. Citado na página 38.

MADANAT, S.; IBRAHIM, W. H. W. Poisson regression models of infrastructure transition probabilities. **Journal of Transportation Engineering**, v. 121, n. 3, p. 267–272, 1995. Citado na página 21.

- MCCULLOCH, C. E.; NEUHAUS, J. M. **Generalized linear mixed models**. [S.l.]: Chapman and Hall: London, 2001. Citado 2 vezes nas páginas 22 e 23.
- MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. **Statistical inference: an Integrated Approach**. [S.l.]: Arnold: London, 1999. Citado na página 37.
- PAULA, G. A. **Modelos de regressão com apoio computacional**. [S.l.]: São Paulo: IME-USP, 2004. Citado na página 39.
- PLUMMER, M. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: VIENNA, AUSTRIA. **Proceedings of the 3rd international workshop on distributed statistical computing**. [S.l.], 2003. v. 124, n. 125, p. 10. Citado na página 38.
- PRENTICE, R. L. A log gamma model and its maximum likelihood estimation. **Biometrika**, v. 61, n. 3, p. 539–544, 1974. Citado na página 23.
- ROSENBAUM, B.; RALL, B. C. Manual: Fitting functional responses—direct parameter estimation by simulating differential equations. 2018. Citado na página 36.
- SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978. Citado na página 40.
- SEARLE, S. R. *et al.* Prediction, mixed models, and variance components. **Reliability and Biometry**, p. 229–266, 1974. Citado na página 23.
- SEI. **Superintendência de Estudos Econômicos e Sociais da Bahia (banco de dados)**. 2018. Disponível em: <<http://www.sei.ba.gov.br>>. Acesso em: 05/06/2018. Citado 2 vezes nas páginas 22 e 51.
- SELF, S. G.; LIANG, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. **Journal of the American Statistical Association**, v. 82, n. 398, p. 605–610, 1987. Citado na página 38.
- SINGER, J. M.; ANDRADE, D. d. Análise de dados longitudinais. **SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA**, v. 7, 1986. Citado na página 22.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 64, n. 4, p. 583–639, 2002. Citado na página 40.
- STRAM, D. O.; LEE, J. W. Variance components testing in the longitudinal mixed effects model. **Biometrics**, p. 1171–1177, 1994. Citado na página 38.
- TRINDADE, D. d. B. **Modelagem para Dados Longitudinais de Contagem**. Dissertação (Mestrado em Estatística) — Universidade Federal de Pernambuco, Recife – PE, 2014. Citado na página 23.
- UMBACH, D. On inference for a mixture of a poisson and a degenerate distribution. **Communications in Statistics-Theory and Methods**, v. 10, n. 3, p. 299–306, 1981. Citado na página 21.
- UNDP. **Atlas do desenvolvimento Humano no Brasil (ranking)**. 2018. Disponível em: <<http://www.atlasbrasil.org.br>>. Acesso em: 05/06/2018. Citado na página 52.

CÁLCULO DAS DERIVADAS

Assim como descrito na [subseção 4.1.1](#), os intervalos de confiança assintóticos para os parâmetros p_{ij} , $\mu_{ZMP_{ij}}$ e p_{0ij} , do modelo (3.2), são calculados a partir do método Delta. Aqui vamos apresentar mais detalhes sobre o cálculo destes intervalos, considerando o modelo (3.2) com apenas uma covariável X .

A.1 Parâmetro p

A partir da relação $E(p_{ij}|b_i) = \frac{\omega_{ij}}{1-e^{-\mu_{ij}}}$, definimos a função h como

$$h(\beta, \gamma, \sigma, \lambda) = \int_{-\infty}^{\infty} \frac{e^{\gamma_0 + \gamma_1 x_{ij}}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i,$$

sendo h diferenciável. Logo, $\mathbf{D} = \left(\frac{\partial h}{\partial \beta_0}, \frac{\partial h}{\partial \beta_1}, \frac{\partial h}{\partial \gamma_0}, \frac{\partial h}{\partial \gamma_1}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \lambda} \right)'$, em que

$$\frac{\partial h}{\partial \beta_0} = - \int_{-\infty}^{\infty} \frac{e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\} + \gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(e^{\gamma_0 + \gamma_1 x_{ij}} + 1)(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})^2} \pi_b(b_i; \mu, \sigma, \lambda) db_i,$$

$$\frac{\partial h}{\partial \beta_1} = -x_{ij} \int_{-\infty}^{\infty} \frac{e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\} + \gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(e^{\gamma_0 + \gamma_1 x_{ij}} + 1)(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})^2} \pi_b(b_i; \mu, \sigma, \lambda) db_i,$$

$$\frac{\partial h}{\partial \gamma_0} = \int_{-\infty}^{\infty} \frac{e^{\exp\{\beta_0 + \beta_1 x_{ij} + b_i\} + \gamma_0 + \gamma_1 x_{ij}}}{(e^{\gamma_0 + \gamma_1 x_{ij}} + 1)^2 (e^{\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}} - 1)} \pi_b(b_i; \mu, \sigma, \lambda) db_i,$$

e

$$\frac{\partial h}{\partial \gamma_1} = x_{ij} \int_{-\infty}^{\infty} \frac{e^{\exp\{\beta_0 + \beta_1 x_{ij} + b_i\} + \gamma_0 + \gamma_1 x_{ij}}}{(e^{\gamma_0 + \gamma_1 x_{ij}} + 1)^2 (e^{\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}} - 1)} \pi_b(b_i; \mu, \sigma, \lambda) db_i.$$

A.2 Parâmetro μ_{ZMP}

A partir da relação $E(Y_{ij}|b_i) = \frac{\omega_{ij}\mu_{ij}}{1-e^{-\mu_{ij}}}$, definimos a função h como

$$h(\beta, \gamma, \sigma, \lambda) = \int_{-\infty}^{\infty} \frac{e^{\gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i,$$

sendo h diferenciável. Logo, $D = \left(\frac{\partial h}{\partial \beta_0}, \frac{\partial h}{\partial \beta_1}, \frac{\partial h}{\partial \gamma_0}, \frac{\partial h}{\partial \gamma_1}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \lambda} \right)'$, em que

$$\begin{aligned} \frac{\partial h}{\partial \beta_0} &= \int_{-\infty}^{\infty} \frac{e^{\gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i \\ &\quad - \int_{-\infty}^{\infty} \frac{e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}} + \gamma_0 + \gamma_1 x_{ij} + 2\beta_0 + 2\beta_1 x_{ij} + 2b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})^2} \pi_b(b_i; \mu, \sigma, \lambda) db_i, \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial h}{\partial \beta_1} &= x_{ij} \frac{e^{\gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i \\ &\quad - x_{ij} \int_{-\infty}^{\infty} \frac{e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}} + \gamma_0 + \gamma_1 x_{ij} + 2\beta_0 + 2\beta_1 x_{ij} + 2b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})^2} \pi_b(b_i; \mu, \sigma, \lambda) db_i, \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial h}{\partial \gamma_0} &= \frac{e^{\gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i \\ &\quad - \int_{-\infty}^{\infty} \frac{e^{2\gamma_0 + 2\gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})^2 (1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i, \end{aligned}$$

e

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial h}{\partial \gamma_1} &= x_{ij} \frac{e^{\gamma_0 + \gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})(1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i \\ &\quad - x_{ij} \int_{-\infty}^{\infty} \frac{e^{2\gamma_0 + 2\gamma_1 x_{ij} + \beta_0 + \beta_1 x_{ij} + b_i}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})^2 (1 - e^{-\exp\{\beta_0 + \beta_1 x_{ij} + b_i\}})} \pi_b(b_i; \mu, \sigma, \lambda) db_i. \end{aligned}$$

A.3 Parâmetro p_0

A partir da relação $p_0 = 1 - \omega_{ij}$, definimos a função h como

$$h(\gamma) = 1 - \frac{e^{\gamma_0 + \gamma_1 x_{ij}}}{1 + e^{\gamma_0 + \gamma_1 x_{ij}}},$$

sendo h diferenciável. Logo, $D = \left(\frac{\partial h}{\partial \gamma_0}, \frac{\partial h}{\partial \gamma_1} \right)'$, em que

$$\frac{\partial h}{\partial \gamma_0} = -\frac{e^{\gamma_0 + \gamma_1 x_{ij}}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})^2} \quad e \quad \frac{\partial h}{\partial \gamma_1} = -x_{ij} \frac{e^{\gamma_0 + \gamma_1 x_{ij}}}{(1 + e^{\gamma_0 + \gamma_1 x_{ij}})^2}.$$

AMOSTRAS DA DENSIDADE A POSTERIORI CONJUNTA

A [Figura 11](#), [Figura 10](#), [Figura 12](#) e [Figura 13](#) exibem os histogramas das amostras geradas a partir do algoritmo de Metropolis-Hastings das densidades *a posteriori* dos parâmetros dos modelos ZMP com efeito aleatório Normal e ZMP com efeito aleatório LGG, considerando a aplicação sobre as notificações de óbitos infantis nos municípios do Estado da Bahia nos anos de 2014, 2015 e 2016.

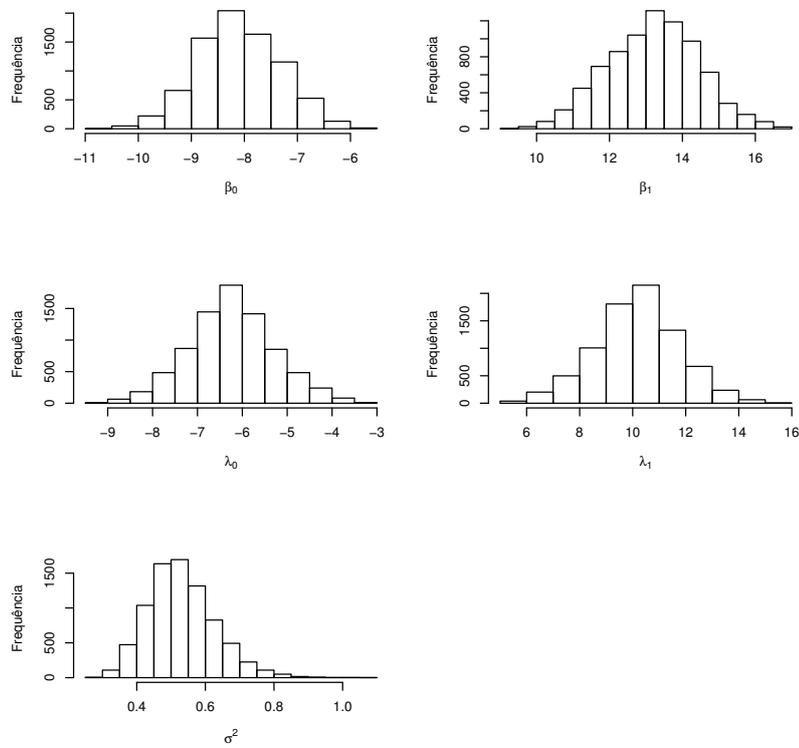


Figura 10 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados completos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório Normal.

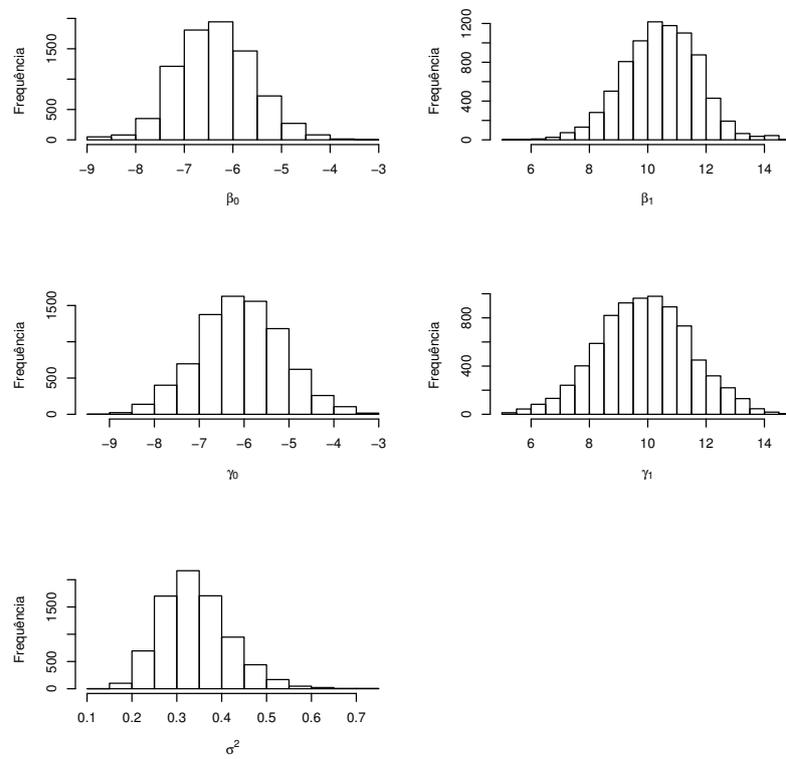


Figura 11 – Histogramas das amostras das distribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados incompletos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório Normal.

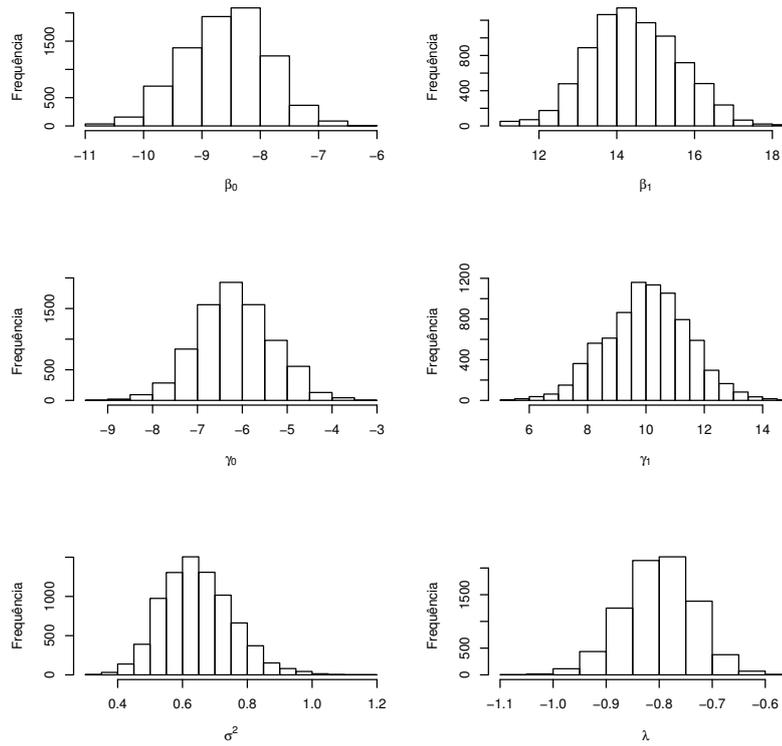


Figura 12 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados completos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório LGG.

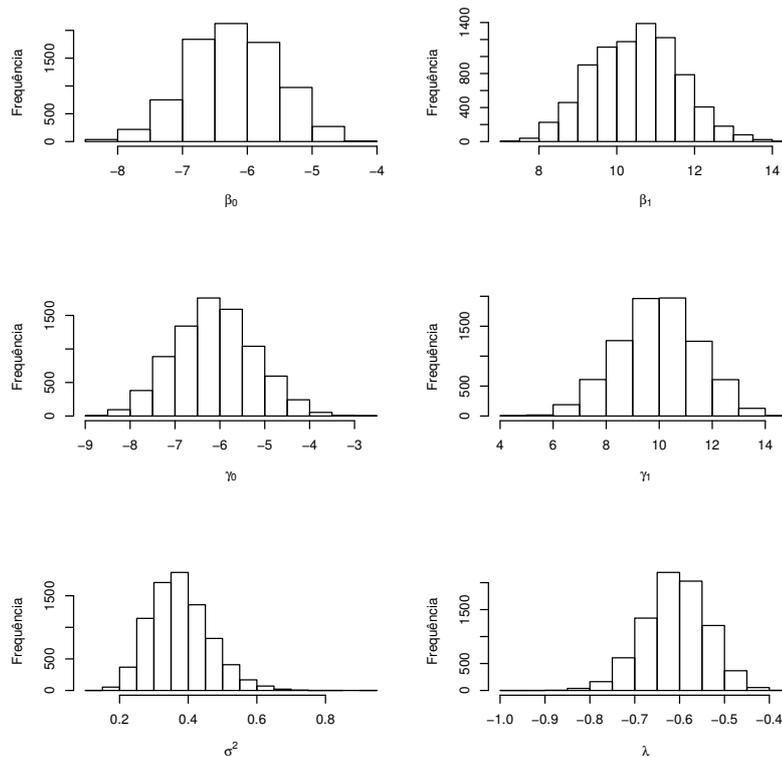


Figura 13 – Histogramas das amostras das ditribuições a posteriori dos parâmetros β_0 , β_1 , γ_0 , γ_1 e σ^2 para os dados incompletos de notificações de óbitos infantis considerando o modelo ZMP com efeito aleatório LGG.