Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC
Departamento de Estatística

Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros†

Dissertação de Mestrado

Alejandro Guillermo Monzón Montoya
alejo_monzon@hotmail.com

Orientador: **Prof. Dr. Víctor Hugo Lachos Dávila**

20 de março de 2009
Campinas - SP

† Este trabalho contou com apoio financeiro da CAPES

INFERÊNCIA E DIAGNÓSTICO EM MODELOS PARA DADOS DE CONTAGEM COM EXCESSO DE ZEROS

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Alejandro Guillermo Monzón Montoya** e aprovada pela comissão julgadora.

Campinas, 20 de março de 2009.

Prof. Dr. **Víctor Hugo Lachos Dávila**
Orientador

Banca Examinadora:

Prof. Dr. Víctor Hugo Lachos Dávila IMECC - UNICAMP
Prof. Dr. Mário de Castro Andrade Filho ICMC - USP
Profª. Drª. Hildete Prisco Pinheiro IMECC - UNICAMP
Prof. Dr. Edwin Moises Marcos Ortega ESALQ - USP
Profª. Drª. Mariana Rodrigues Motta IMECC - UNICAMP

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, **UNICAMP**, como requisito parcial para obtenção do Título de **MESTRE em ESTATÍSTICA**.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Bibliotecária: Maria Júlia Milani Rodrigues - CRB8a 2116

Monzón Montoya, Alejandro Guillermo
M769i Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros / Alejandro Guillermo Monzón Montoya – Campinas, [S.P.:s.n.], 2009.

Orientador: Victor Hugo Lachos Dávila

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Dados de contagem. 2. Análise de regressão. 3. Influência local. 4. Resíduos. I. Lachos Dávila, Victor Hugo. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Inference and diagnostic in zero-inflated count data models

Palavras-chave em inglês (keywords): 1. Count data. 2. Regression analysis. 3. Local influence. 4. Residues.

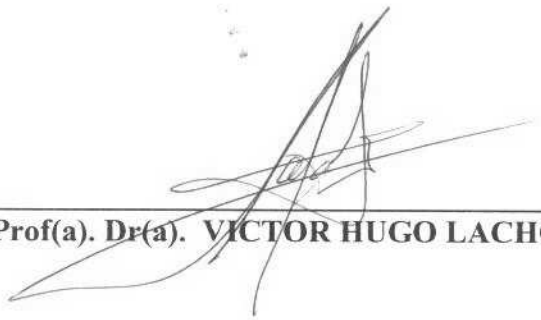
Área de concentração: Modelos de regressão
Titulação: Mestre em Estatística

Banca examinadora: 1. Prof. Dr. Victor Hugo Lachos Dávila (IMECC-UNICAMP)
2. Prof. Dr. Mário de Castro Andrade Filho (ICMC-USP)
3. Prof^a. Dr^a. Hildete Prisco Pinheiro (IMECC-UNICAMP)
4. Prof. Dr. Edwin Moises Marcos Ortega (ESALQ-USP)
5. Prof^a. Dr^a. Mariana Rodrigues Motta (IMECC-UNICAMP)

Data da defesa: 20/03/2009
Programa de pós-graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 20 de março de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). VICTOR HUGO LACHOS DÁVILA

Mário de Castro Andrade Filho

Prof(a). Dr(a). MÁRIO DE CASTRO ANDRADE FILHO

Hildete Pinheiro

Prof(a). Dr(a). HILDETE PRISCO PINHEIRO

DEDICATÓRIA

A **DEUS**, *pela presença constante em minha vida e por ter me dado força e esperança nos momentos difíceis desta caminhada.*

Ao **MANOLITO**, *meu campeão, que com seus escassos cinco anos me deu tantas lições de vida.*
À **ALESSANDRITA**, *minha princesa, força constante em minha vida.*
À **JENNY**, *minha outra metade.*

Aos meus pais, **OLINDA** *(in Memoriam) e* **LEANDRO**, *e minha irmã* **LIDIA**, *porque sempre me apoiarem em minhas escolhas.*

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido durante o período de março de 2007 a fevereiro de 2009, sem o qual não seria possível a realização de meus estudos de Mestrado em Estatística.

A meu orientador e amigo, Prof. Victor Hugo Lachos Dávila, por sua ajuda constante, não só na parte acadêmica.

Aos professores Mário de Castro, Hildete Prisco Pinheiro e Mariana Rodrigues Motta, pelas correções e sugestões.

Aos professores do departamento de Estatística do IMECC, especialmente à Professora Marina Vachkovskaia, ao Prof. Jesús Enrique García e ao Prof. Mauricio Zevallos.

Aos funcionários do IMECC, particularmente à Tânia, Cidinha e Ednaldo, a atenção, o carinho e a amizade.

Aos colegas e amigos do IMECC, pela amizade, incentivo e por fazer-me sentir como em casa, em especial a Cristiano, Simoni, Lúcia, Rodrigo Basso, Bruno, Rafael e o Aldo.

Ao Brasil, minha gratidão eterna.

Resumo

Em análise de dados, muitas vezes encontramos dados de contagem onde a quantidade de zeros excede aquela esperada sob uma determinada distribuição, tal que não é possível fazer uso dos modelos de regressão usuais. Além disso, o excesso de zeros pode fazer com que exista sobredispersão nos dados. Neste trabalho são apresentados quatro tipos de modelos para dados de contagem inflacionados de zeros: o modelo Binomial (ZIB), o modelo Poisson (ZIP), o modelo binomial negativa (ZINB) e o modelo beta-binomial (ZIBB). Usa-se o algoritmo EM para obter estimativas de máxima verossimilhança dos parâmetros do modelo e usando a função de log-verossimilhança dos dados completos obtemos medidas de influência local baseadas na metodologia proposta por Zhu e Lee (2001) e Lee e Xu (2004). Também propomos como construir resíduos para os modelos ZIB e ZIP. Finalmente, as metodologias descritas são ilustradas pela análise de dados reais.

Palavras-chave: Dados de contagem; Análise de regressão; Influência local; Resíduos.

Abstract

When analyzing count data sometimes a high frequency of extra zeros is observed and the usual regression analysis is not applicable. This feature may be accounted for by over-dispersion in the data set. In this work, four types of models for zero inflated count data are presented: viz., the zero-inflated Binomial (ZIB), the zero-inflated Poisson (ZIP), the zero-inflated Negative Binomial (ZINB) and the zero-inflated Beta-Binomial (ZIBB) regression models. We use the EM algorithm to obtain maximum likelihood estimates of the parameter of the proposed models and by using the complete data likelihood function we develop local influence measures following the approach of Zhu and Lee (2001) and Lee and Xu (2004). We also discuss the calculation of residuals for the ZIB and ZIP regression models with the aim of identifying atypical observations and/or model misspecification. Finally, results obtained for two real data sets are reported, illustrating the usefulness of the proposed methodology.

Key-words: Count data; Regression analysis; Local influence; Residues.

Sumário

Lista de Figuras

Introdução

Recentemente existe um interesse crescente em modelos de mistura para dados de contagem inflacionados de zeros. Nestes modelos, chamados modelos de regressão para dados inflacionados de zero (ZI), mistura-se uma distribuição degenerada com ponto de massa unitária e um modelo de regressão simples baseado em uma distribuição padrão. Por exemplo, modelos de regressão inflacionados de zero relativos ao modelo Poisson (modelo ZIP) foram estudados por Lambert (1992), Hall (2000), e outros; modelos binomial negativa inflacionados de zero (ZINB) aparecem em Ridout et al. (2001) e outros; e modelos binomial inflacionados de zero (ZIB) são discutidos por Hall (2000) e Vieira et al. (2000). Os modelos ZI para dados de contagem são úteis quando zeros são observados mais freqüentemente do que previsto pelo modelo. Sua forma de mistura conta para alguns dos zeros através da distribuição não-degenerada (por exemplo, Poisson) e algumas através da distribuição degenerada (zero).

Muitos aspectos desses modelos, porém, não estão completamente estudados, necessitando de estudos adicionais. Portanto, o presente trabalho tem como objetivos:

i) Fazer uma revisão da literatura sobre a modelagem de dados em que existe excesso de zeros;

ii) Implementar os métodos de estimação em programas computacionais;

iii) Implementar testes escore para a hipótese de excesso de zeros;

iv) Desenvolver métodos de diagnóstico de influência para validar as suposições dos modelos e detectar a presença de observações influentes;

v) Desenvolver métodos para análise residual.

Organização do trabalho

No Capítulo 1 revisamos os conceitos básicos dos modelos lineares generalizados e da análise de diagnóstico.

No Capítulo 2 estudamos os modelos para dados de contagem incluindo os modelos para dados inflacionados de zeros. Em seguida, dois processos de estimação usuais são abordados: o método de escore de Fisher e o algoritmo EM. Finalmente, desenvolvemos testes escore para a hipótese de excesso de zeros, para os modelos binomial e Poisson inflacionados de zeros.

No Capítulo 3 estendemos a análise de diagnóstico aos modelos inflacionados de zeros, desenvolvendo métodos de análise residual. Efetuamos também um estudo de diagnóstico de influência local e global usando a metodologia proposta por Zhu e Lee (2001).

No Capítulo 4 fazemos um estudo de simulação para estimar os parâmetros dos quatro modelos de regressão inflacionados de zeros aplicando o algoritmo EM; após disso, dois exemplos de aplicação são discutidos considerando conjuntos de dados reais. Abordamos a análise de diagnóstico para os modelos propostos, sendo traçados os gráficos de envelopes simulados para uma análise visual da qualidade do ajuste. Ainda, os gráficos de diagnóstico são traçados para identificar pontos influentes nos modelos de regressão inflacionados de zeros.

Finalmente, no Capítulo 5 são apresentadas algumas considerações finais e perspectivas para trabalhos futuros.

Capítulo 1

Modelos Lineares Generalizados

"Nossa recompensa está no esforço e não no resultado. Um esforço total é uma vitória completa."
Mahatma Gandhi

Os modelos lineares generalizados (MLG) são uma extensão do modelo linear geral (Stigler, 1981) para uma família mais geral e foram propostos por Nelder e Wedderburn (1972). Esta nova família unifica tanto modelos com variáveis resposta categóricas como numéricas; considera distribuições como binomial, Poisson, hipergeométrica, binomial negativa, entre outras, e não unicamente a distribuição normal. Como nos modelos de regressão linear, considera-se o suposto de independência para as observações. Porém, para estes modelos, diferente do que ocorre no modelo linear geral, a distribuição do componente aleatório não é necessariamente homocedástica, ou seja, não se requer homogeneidade das variâncias. Por exemplo, no modelo de regressão de Poisson a variância da variável resposta é determinada pelo valor esperado $\boldsymbol{\mu}$. Portanto, a variância pode variar à medida que varie o valor esperado, diferente do modelo clássico normal que tem dois parâmetros, a média e a variância (McCullagh e Nelder, 1991), que supõe variância constante.

Os modelos log-lineares, logito, probito, logístico e de regressão linear são algumas estruturas que formam parte desta família. Os modelos lineares generalizados são definidos pelos seguintes componentes:

a) Componente aleatório. Representado por um conjunto de variáveis aleatórias independentes $Y_1, Y_2, \ldots, Y_n$ provenientes de uma mesma distribuição pertencente à

família exponencial, definida por

$$f(y_i;\theta_i,\phi) = \exp\{\phi[y_i\theta_i - b(\theta_i)] + c(y_i,\phi)\}, \quad i = 1,2,\ldots,n, \tag{1.1}$$

em que b(·) e c(·) são funções reais diferenciáveis conhecidas. Para o modelo em (1.1) valem as seguintes relações:

$$E[Y_i] = \mu_i = b'(\theta_i) \quad \text{e} \quad Var[Y_i] = \phi^{-1}V_i,$$

sendo $\phi^{-1} > 0$ o parâmetro de dispersão, $V_i = V(\mu_i) = d\mu_i/d\theta_i$ é denominada função de variância e depende unicamente da média μ_i e o parâmetro θ_i é denominado de parâmetro natural ou canônico.

O parâmetro natural θ_i pode ser expresso como

$$\theta_i = \int V_i^{-1} d\mu_i = q(\mu_i),$$

sendo $q(\mu_i)$ uma função conhecida da média μ_i (veja Cordeiro e Demétrio, 2007).

b) Componente sistemático. As variáveis explicativas contribuem na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j = \boldsymbol{x}_i\boldsymbol{\beta}, \;\; i = 1,\ldots,n \quad \text{ou } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \tag{1.2}$$

sendo $\mathbf{X} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^\top$ a matriz de covariáveis ou variáveis explicativas do modelo, com $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ip})^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p)^\top$ $(p < n)$ é o vetor de parâmetros cujos valores são desconhecidos e precisam ser estimados, e $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_n)^\top$ o preditor linear. Se um parâmetro tem valor conhecido, o termo correspondente na estrutura linear é chamado *offset*.

c) Função de ligação. Relaciona o preditor linear $\boldsymbol{\eta}$ com o valor esperado da variável resposta, $E[\mathbf{Y}|\mathbf{X}] = \boldsymbol{\mu}$, ou seja, relaciona o componente aleatório ao componente sistemático através da função $h(\mu_i) = \mathbf{X}_i\boldsymbol{\beta}$, isto é,

$$h(\mu_i) = \eta_i, \quad i = 1,\ldots,n, \tag{1.3}$$

em que $h(\cdot)$ é uma função conhecida, monótona e diferenciável denominada função de ligação; então $\boldsymbol{\mu}_i = h^{-1}(\boldsymbol{\eta}_i)$, $i = 1,\ldots,n$. No caso particular do modelo linear, $\boldsymbol{\mu}$ e $\boldsymbol{\eta}$ podem assumir qualquer valor na reta real, e a função de ligação é a identidade $(\boldsymbol{\eta} = \boldsymbol{\mu})$, mas para a distribuição de Poisson, dado que $\boldsymbol{\mu} > 0$, uma função de ligação

adequada é a função logarítmica ($\boldsymbol{\eta} = \log(\boldsymbol{\mu})$); para o caso do modelo binomial, como existe a restrição de que o domínio da função de ligação é o intervalo $(0, 1)$, uma função adequada é a chamada função logito $\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right)$.

Assim, para a especificação do modelo, os parâmetros θ_i da família de distribuições definida em (1.1) não são de interesse direto (pois há um para cada observação), mas sim um conjunto menor de parâmetros $\beta_1, \ldots, \beta_p$ tais que uma combinação linear dos $\beta' s$ seja igual a alguma função do valor esperado de Y_i.

Portanto, segundo Cordeiro e Demétrio (2007), uma decisão importante na escolha do MLG é definir os termos do trinômio: (i) distribuição da variável resposta; (ii) matriz do modelo e (iii) função de ligação. Nesses termos, um MLG é definido por uma distribuição da família (1.1), uma estrutura linear (1.2) e uma função de ligação (1.3). Por exemplo, quando $\theta = \mu$ e a função de ligação é linear (identidade), obtém-se o modelo clássico de regressão como um caso particular. Os modelos log-lineares são deduzidos supondo $\theta = \log \mu$ com função de ligação logarítmica $\log \mu = \eta$. Torna-se clara, agora, a palavra "generalizado", significando uma distribuição mais ampla do que a normal para a variável resposta, e uma função não-linear em um conjunto linear de parâmetros conectando a média dessa variável com a parte determinística do modelo.

1.1 A Função de Log-Verossimilhança

Assumindo que cada componente da variável aleatória Y tem uma distribuição proveniente da família exponencial da forma denotada anteriormente, escrevemos a função de log-verossimilhança como

$$\ell(\boldsymbol{\theta}, \phi) = \log f_Y(\boldsymbol{y}; \boldsymbol{\theta}, \phi),$$

que é uma função de $\boldsymbol{\theta}$ e ϕ, dado um valor de $\boldsymbol{y}$. De acordo com (1.1),

$$\ell(\boldsymbol{\theta}, \phi) = \sum_{i=1}^{n} \{\phi_i [y_i \theta_i - b(\theta_i)] + c(y_i; \phi_i)\}. \tag{1.4}$$

A média e a variância de Y são obtidas a partir das seguintes relações de regularidade (Dobson, 2001):

$$E\left[\frac{\partial \ell}{\partial \theta}\right] = 0,$$

$$E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] + \left(E\left[\frac{\partial \ell}{\partial \theta}\right]\right)^2 = 0.$$

Destas relações obtém-se que $E[Y] = \mu = b'(\theta)$ e $Var[Y] = \phi^{-1}b''(\theta)$, sendo que $b''(\theta)$ é uma função que depende do parâmetro canônico θ e é chamada de função de variância.

1.2 Estatística Suficiente e Ligação Canônica

Cada distribuição tem uma função de ligação especial que está associada ao preditor linear $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, para o qual existe uma estatística suficiente com a mesma dimensão de $\boldsymbol{\beta}$. Estas ligações são chamadas canônicas, quando $\boldsymbol{\theta} = \boldsymbol{\eta}$, em que $\boldsymbol{\theta}$ é o parâmetro canônico.

Para as ligações canônicas, a estatística suficiente em notação vetorial é $\mathbf{X}^\top\mathbf{Y}$ (McCullagh e Nelder, 1991). Expressando o logaritmo da função de verossimilhança dos modelos lineares generalizados com respostas independentes da forma $\ell(\boldsymbol{\beta})$, temos que uma das vantagens de usar ligações canônicas é que garantem a concavidade de $\ell(\boldsymbol{\beta})$ e, portanto, obtém-se resultados assintóticos mais facilmente. A concavidade da função de log-verossimilhança garante a unicidade do estimador de máxima verossimilhança de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, quando este existe.

1.3 Estimação dos Parâmetros

Utilizam-se vários métodos para estimar os parâmetros. No entanto, nós usaremos o método de máxima verossimilhança, o qual fornece estimadores com as propriedades de consistência e eficiência assintótica (Cordeiro e Demétrio, 2007).

O vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança em (1.4). O logaritmo da função de verossimilhança como função apenas de $\boldsymbol{\beta}$ (considerando-se o parâmetro de dispersão ϕ conhecido) dado o vetor $\boldsymbol{y}$ é definido por $\ell(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}; \boldsymbol{y})$) e usando-se a expressão (1.1) tem-se

$$\ell(\boldsymbol{\beta}) = \phi \sum_{i=1}^{n} [y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i, \phi), \tag{1.5}$$

em que $\theta_i = q(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \sum_{r=1}^{p} x_{ir}\beta_r$. Da expressão (1.5) pode-se calcular, pela regra da cadeia, o vetor escore $\mathbf{U}(\boldsymbol{\beta}) = \dfrac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ de dimensão p, com elemento típico

$$U_r = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{i=1}^{n} \frac{d\ell_i}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta_r},$$

e sabendo-se que $\mu_i = b'(\theta_i)$ e $\dfrac{d\mu_i}{d\theta_i} = V_i$, tem-se

$$U_r = \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir}, \qquad r = 1, \ldots, p. \tag{1.6}$$

A estimativa de máxima verossimilhança $\widehat{\boldsymbol{\beta}}$ do vetor de parâmetros $\boldsymbol{\beta}$ é obtida igualando-se U_r a zero. Em geral, as equações $U_r = 0$, $r = 1, \ldots, p$, não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson.

O método iterativo de Newton-Raphson para a solução de uma equação $f(a) = 0$ é baseado na aproximação de Taylor para a função $f(a)$ na vizinhança do ponto a_0, ou seja,

$$f(a) = f(a_0) + (a - a_0) f'(a_0) = 0,$$

obtendo-se

$$a = a_0 - \frac{f(a_0)}{f'(a_0)},$$

ou, de uma forma mais geral,

$$a^{(m+1)} = a^{(m)} - \frac{f(a^{(m)})}{f'(a^{(m)})},$$

sendo $a^{(m+1)}$ o valor de a no passo $(m+1)$, $a^{(m)}$ o valor de a no passo m, $f(a^{(m)})$ a função $f(a)$ avaliada em $a^{(m)}$ e $f'(a^{(m)})$ a derivada da função $f(a)$ avaliada em $a^{(m)}$.

Considerando-se que se deseja obter a solução do sistema de equações $\mathbf{U}(\boldsymbol{\beta}) = \dfrac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$ e, usando-se a versão multivariada do método de Newton-Raphson, tem-se

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{J}(\boldsymbol{\beta}^{(m)})]^{-1} \mathbf{U}(\boldsymbol{\beta}^{(m)}),$$

sendo $\boldsymbol{\beta}^{(m)}$ e $\boldsymbol{\beta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $(m+1)$, respectivamente, $\mathbf{U}(\boldsymbol{\beta}^{(m)})$ o vetor escore avaliado no passo m, e $[\mathbf{J}(\boldsymbol{\beta}^{(m)})]^{-1}$ a inversa do negativo da matriz de derivadas parciais de segunda ordem de $\ell(\boldsymbol{\beta})$, com elementos $-\dfrac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_r \partial \beta_s}$, avaliada no passo m.

Quando as derivadas parciais de segunda ordem são avaliadas facilmente, o método de Newton-Raphson é bastante útil. Acontece, porém, que isso nem sempre ocorre e no caso dos MLG usa-se o método do escore de Fisher que, em geral, é mais simples (coincidindo com o método de Newton-Raphson no caso das funções de ligação canônicas). Esse método envolve a substituição da matriz de derivadas parciais de segunda ordem pela matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada, $\mathbf{J}$, pela matriz de informação esperada de Fisher, $\mathbf{K}$. Logo,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{K}(\boldsymbol{\beta}^{(m)})]^{-1} \mathbf{U}(\boldsymbol{\beta}^{(m)}), \tag{1.7}$$

sendo que $\mathbf{K}$ tem elementos típicos dados por

$$K_{r,s} = -E\left[\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial\beta_r \partial\beta_s}\right] = E\left[\frac{\partial \ell(\boldsymbol{\beta})}{\partial\beta_r}\frac{\partial \ell(\boldsymbol{\beta})}{\partial\beta_s}\right],$$

que é a matriz de covariâncias dos $U_r's$.

Multiplicando-se ambos os lados de (1.7) por $\mathbf{K}(\boldsymbol{\beta}^{(m)})$ tem-se

$$\mathbf{K}(\boldsymbol{\beta}^{(m)})\boldsymbol{\beta}^{(m+1)} = \mathbf{K}(\boldsymbol{\beta}^{(m)})\boldsymbol{\beta}^{(m)} + \mathbf{U}(\boldsymbol{\beta}^{(m)}). \tag{1.8}$$

O elemento típico K_{rs} de $\mathbf{K}$ é obtido de (1.6) como

$$K_{r,s} = E(U_r U_s) = \phi^2 \sum_{i=1}^{n} E(Y_i - \mu_i)^2 \frac{1}{V_i^2}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ir}x_{is}$$

ou

$$K_{r,s} = \phi \sum_{i=1}^{n} w_i x_{ir} x_{is},$$

sendo $w_i = \dfrac{1}{V_i}\left(\dfrac{d\mu_i}{d\eta_i}\right)^2$ denominado peso. Logo, a matriz de informação de Fisher para $\boldsymbol{\beta}$ tem a forma

$$\mathbf{K} = \phi \mathbf{X}^\top \mathbf{W} \mathbf{X},$$

sendo $\mathbf{W} = \text{diag}\{w_1, \ldots, w_n\}$ uma matriz diagonal de pesos que traz a informação sobre a distribuição e a função de ligação usadas e pode também incluir um termo para um peso *a priori*. No caso das funções de ligação canônicas tem-se $w_i = V_i$, pois $V_i = V(\mu_i) = \dfrac{d\mu_i}{d\eta_i}$.

O vetor escore $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$ com componentes em (1.6) pode, então, ser escrito na forma

$$\mathbf{U} = \phi \mathbf{X}^\top \mathbf{W} \mathbf{G}(\boldsymbol{y} - \boldsymbol{\mu}),$$

com $\mathbf{G} = \text{diag}\{d\eta_1/d\mu_1, \ldots, d\eta_n/d\mu_n\} = \text{diag}\{g'(\mu_1), \ldots, g'(\mu_n)\}$. Assim, a matriz diagonal $\mathbf{G}$ é formada pelas derivadas de primeira ordem da função de ligação.

Substituindo $\mathbf{K}$ e $\mathbf{U}$ em (1.8) e eliminando ϕ, tem-se

$$\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}\boldsymbol{\beta}^{(m+1)} = \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}\boldsymbol{\beta}^{(m)} + \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{G}^{(m)}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)}),$$

ou

$$\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}\boldsymbol{\beta}^{(m+1)} = \mathbf{X}^\top \mathbf{W}^{(m)} [\boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)})].$$

Definindo a variável dependente ajustada $\boldsymbol{z} = \boldsymbol{\eta} + \mathbf{G}(\boldsymbol{y} - \boldsymbol{\mu})$, temos que

$$\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}\boldsymbol{\beta}^{(m+1)} = \mathbf{X}^\top \mathbf{W}^{(m)} \boldsymbol{z}^{(m)};$$

logo

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \boldsymbol{z}^{(m)}. \tag{1.9}$$

A equação matricial (1.9) é válida para qualquer MLG e mostra que a solução das equações de máxima verossimilhança (MV) equivale a calcular repetidamente uma regressão linear ponderada de uma variável dependente ajustada $\boldsymbol{z}$ sobre a matriz $\mathbf{X}$ usando uma função de peso $\mathbf{W}$ que se modifica no processo iterativo. As funções de variância e de ligação tomam parte no processo iterativo através de $\mathbf{W}$ e $\boldsymbol{z}$. Note que $Cov(\boldsymbol{z}) = \mathbf{G} Cov(\mathbf{Y}) \mathbf{G} = \phi \mathbf{W}^{-1}$, isto é, os z_i não são correlacionados. É importante enfatizar que a equação iterativa (1.9) não depende do parâmetro de dispersão ϕ.

A variável dependente ajustada depende da derivada de primeira ordem da função de ligação. Quando a função de ligação é linear ($\boldsymbol{\eta} = \boldsymbol{\mu}$), isto é, a identidade, tem-se $\mathbf{W} = \mathbf{V}^{-1}$ em que $\mathbf{V} = \text{diag}\{V_1, \ldots, V_n\}$, $\mathbf{G} = \mathbf{I}$ e $\boldsymbol{z} = \boldsymbol{y}$, ou seja, a variável dependente ajustada reduz-se ao vetor de dados. Para o modelo normal linear ($\mathbf{V} = \mathbf{I}, \boldsymbol{\mu} = \boldsymbol{\eta}$), tornando $\mathbf{W}$ igual à matriz identidade de dimensão n, $\boldsymbol{z} = \boldsymbol{y}$ e de (1.9) obtém-se que a estimativa $\widehat{\boldsymbol{\beta}}$ reduz-se à conhecida expressão $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{y}$. Esse é o único caso em que $\widehat{\boldsymbol{\beta}}$ é calculado de forma explícita sem ser necessário um procedimento iterativo.

O processo iterativo consiste em especificar uma estimativa inicial e sucessivamente alterá-la até que a convergência seja obtida e, portanto, $\boldsymbol{\beta}^{(m+1)}$ aproxime-se de $\widehat{\boldsymbol{\beta}}$ quando m cresce. Note que cada observação pode ser considerada como uma estimativa do seu valor médio, isto é, $\mu_i^{(1)} = y_i$ e, portanto, calcula-se

$$\eta_i^{(1)} = h(\mu_i^{(1)}) = h(y_i) \quad \text{e} \quad w_i^{(1)} = \frac{1}{V(y_i)[h'(y_i)]^2}.$$

Usando-se $\boldsymbol{\eta}^{(1)}$ como variável resposta, $\mathbf{X}$ a matriz do modelo, e $\mathbf{W}^{(1)}$ a matriz diagonal de pesos com elementos $w_i^{(1)}$, obtém-se o vetor $\boldsymbol{\beta}^{(2)} = (\mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(1)} \boldsymbol{\eta}^{(1)}$. A seguir, o algoritmo de estimação, para $m = 2, \ldots, k$, sendo $k - 1$ o número necessário de iterações para convergência, na iteração m pode ser resumido nos seguintes passos:

(1) obter as estimativas

$$\eta_i^{(m)} = \sum_{r=1}^{p} x_{ir} \beta_r^{(m)} \quad \text{e} \quad \mu_i^{(m)} = h^{-1}(\eta_i^{(m)});$$

(2) obter a variável dependente ajustada

$$z_i^{(m)} = \eta_i^{(m)} + (y_i - \mu_i^{(m)}) g'(\mu_i^{(m)})$$

e os pesos

$$w_i^{(m)} = \frac{1}{V(\mu_i^{(m)})[g'(\mu_i^{(m)})]^2}, \quad i = 1, \ldots, n;$$

(3) calcular

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \boldsymbol{z}^{(m)},$$

voltar ao passo (1) com $\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m+1)}$ e repetir o processo até obter a convergência, definindo-se, então, $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(k-1)}$.

Dentre os muitos critérios existentes, um critério para verificar a convergência pode ser

$$\|\beta_r^{(m+1)} - \beta_r^{(m)}\| \leq 10^{-6} \tag{1.10}$$

onde $\|x\|$ indica a norma do vetor x.

Para pequenas amostras, o processo iterativo (1.9) pode divergir. O número de iterações até convergência depende inteiramente do valor inicial arbitrado para $\widehat{\boldsymbol{\beta}}$, embora, geralmente, o algoritmo convirja rapidamente. A desvantagem do método tradicional de Newton-Raphson com o uso da matriz observada de derivadas de segunda ordem é que, normalmente, não converge para determinados valores iniciais.

Vários programas estatísticos utilizam o algoritmo iterativo (1.9) para obter as estimativas de máxima verossimilhança $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ dos parâmetros lineares do MLG, entre os quais, S-PLUS, SAS, MATLAB e R. Neste trabalho será usado o programa R (R Development Core Team, 2008).

1.4 Propriedades e Distribuição de $\widehat{\beta}$

Cordeiro e Demétrio (2007) mencionam as seguintes propriedades do estimador $\widehat{\boldsymbol{\beta}}$:

i) O estimador $\hat{\boldsymbol{\beta}}$ é assintoticamente não viesado, isto é, para amostras grandes $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

ii) Denotando-se $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{U}$ tem-se que a matriz de variâncias e covariâncias de $\widehat{\boldsymbol{\beta}}$, para amostras grandes, é dada por

$$Cov(\widehat{\boldsymbol{\beta}}) = E[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top] = \mathbf{K}^{-1} E[\mathbf{U}\mathbf{U}^\top](\mathbf{K}^{-1})^\top = \mathbf{K}^{-1}\mathbf{K}\mathbf{K}^{-1} = \mathbf{K}^{-1},$$

pois $\mathbf{K}^{-1}$ é simétrica.

A matriz de covariâncias é consistentemente estimada por $\widehat{\mathbf{K}}^{-1} = \phi^{-1}(\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1}$, em que $\widehat{\mathbf{W}}$ é a matriz de pesos $\mathbf{W}$ avaliada em $\widehat{\boldsymbol{\beta}}$.

iii) Para amostras grandes, tem-se a aproximação

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{K} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx \chi^2_p \,, \tag{1.11}$$

sendo que " $\approx$ " significa aproximadamente. De forma equivalente,

$$\widehat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, \mathbf{K}^{-1}), \tag{1.12}$$

ou seja, $\widehat{\boldsymbol{\beta}}$ tem distribuição assintótica normal p-variada, que é a base para a construção de testes e intervalos de confiança para os parâmetros lineares de um MLG. Para modelos lineares com variáveis respostas com distribuição normal, (1.11) e (1.12) são distribuições exatas.

1.5 Diagnóstico de Influência

Em estudos de modelagem estatística, uma etapa importante corresponde à validação das pressuposições do modelo mediante estudos de sensibilidade. A análise de diagnósticos tem o objetivo de verificar possíveis afastamentos das suposições feitas para o modelo (parte sistemática e aleatória), verificar a existência de observações extremas com interferência desproporcional no ajuste e detectar observações influentes nas estimativas do modelo.

Pontos influentes são aqueles com influência desproporcional nas estimativas dos coeficientes, isto é, quando retirados do modelo mudam de forma substancial as estimativas ou mesmo a significância dos coeficientes.

O método mais conhecido para detectar tais pontos é o de deleção de pontos, que consiste em retirar um ponto e verificar as variações nas estimativas e outros resultados inferenciais.

A análise de resíduos busca detectar pontos extremos influentes na estimação dos modelos e também a adequação do modelo. Técnicas gráficas, como a construção de envelope simulado (Atkinson, 1985), auxiliam na busca em detectar pontos extremos na distribuição dos dados. Por último, análise de influência busca localizar observações influentes nas estimativas do modelo, feita através dos métodos de alavanca, influência global e local.

A análise de influência global, via exclusão de casos, em que o impacto de deletar uma (ou um grupo) observação na estimativa dos parâmetros é diretamente avaliada por métricas como a distância de Cook (Cook, 1977). Eliminação de casos é a ferramenta mais popular para avaliar o impacto individual de casos no processo de estimação, sendo considerada como uma técnica de influência global.

A análise de influência local, baseada em geometria diferencial, é efetuada comparando estimativas de parâmetros antes e depois de perturbar os dados ou as hipóteses do modelo (Cook, 1986). A metodologia de influência local é útil para verificar as suposições do modelo, assim como a identificação de dados aberrantes e/ou influentes, por meio de estudar o efeito de introduzir pequenas perturbações no modelo (ou dados) usando uma medida de influência apropriada. Esta etapa permite estudar suposições de modelos, como homogeneidade de variâncias.

Recentemente, inspirados pela idéia básica do algoritmo EM, Zhu e Lee (2001) propuseram um método unificado para análise de influência local em modelos estatísticos com dados faltantes, utilizando função de afastamento da verossimilhança completa.

Influência global

Para avaliar a influência de observações na estimativa do parâmetro $\boldsymbol{\theta}$ de dimensão $p\times 1$, algumas estatísticas são de uso comum. Uma, denominada LD_i, utiliza o afastamento da log-verossimilhança, dado por

$$LD_i(\boldsymbol{\theta}) = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{[i]})], \tag{1.13}$$

sendo $\ell(\boldsymbol{\theta})$ a função de log-verossimilhança, $\widehat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ com todos os dados da amostra e $\widehat{\boldsymbol{\theta}}_{[i]}$ o EMV de $\boldsymbol{\theta}$ com a exclusão da i-ésima observação. Note que $LD_i(\boldsymbol{\theta}) \geq 0$.

Uma outra medida de influência global, obtida através da eliminação de observações, é a medida de Cook (Cook e Weisberg, 1982) definida como

$$D_I = \frac{(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_I)^\top \mathbf{V} (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_I)}{c},$$

com $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_i$ representando, respectivamente, o EMV de $\boldsymbol{\theta}$ com todos os dados da amostra e com a eliminação do conjunto de observações I. A medida D_I mede a influência das observações do conjunto I na estimação de $\boldsymbol{\theta}$, segundo a métrica definida por $\mathbf{V}$ e c.

Enfoque de influência local de Cook

Dado um conjunto de observações, introduzimos perturbações no modelo através de um vetor $\boldsymbol{\omega}$ de dimensão $q \times 1$ em algum subconjunto aberto de $\mathbb{R}^q$. Geralmente $\boldsymbol{\omega}$ deve refletir qualquer esquema de perturbação bem definido. Por exemplo, $\boldsymbol{\omega}$ pode ser usado para introduzir uma pequena modificação nas variáveis explicativas ou para perturbar a matriz de covariâncias nos erros no modelo de regressão linear (veja Galea et al., 1997).

Supondo que o esquema de perturbação esteja bem definido, denota-se por $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ a função de log-verossimilhança correspondente ao modelo perturbado. Assume-se que existe um $\boldsymbol{\omega}_0$ tal que $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$, para todo $\boldsymbol{\theta}$. Para avaliar a influência da perturbação $\boldsymbol{\omega}$ sobre o estimador de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$, Cook (1986) sugere estudar o comportamento do afastamento da log-verossimilhança, definido por

$$LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_\omega)],$$

em torno de $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, em que $\widehat{\boldsymbol{\theta}}_\omega$ denota o estimador de máxima verossimilhança sob $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. Sob esta perspectiva, o gráfico de $LD(\boldsymbol{\omega})$ contém informação essencial sobre a influência do esquema de perturbação. A idéia consiste em analisar como a superfície $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, LD(\boldsymbol{\omega}))^\top$ desvia-se de seu plano tangente em $\boldsymbol{\omega}_0$. Essa análise pode ser feita estudando as curvaturas das seções normais à superfície $\alpha(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$, que são interseções de $\alpha(\boldsymbol{\omega})$ com planos contendo o vetor normal ao seu plano tangente em $\boldsymbol{\omega}_0$. Na literatura de geometria diferencial, as curvaturas dessas seções normais são denominadas *curvaturas normais*.

Cook (1986) mostra que a curvatura normal na direção $\mathbf{d}$ (com $\|\mathbf{d}\| = 1$) é dada por

$$C_d(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta}\mathbf{d}|,$$

em que $\ddot{\boldsymbol{L}} = -\mathbf{J} = -\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ é a matriz de informação observada sob o modelo postulado e $\boldsymbol{\Delta}$ é a matriz de segundas derivadas parciais com respeito a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}$,

$$\boldsymbol{\Delta} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$$

avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

O resultado na equação acima pode ser utilizado para avaliar a influência que o esquema de perturbações considerado exerce sobre as estimativas dos parâmetros do modelo. Segundo Cook (1986), a direção que produz a maior mudança local na estimativa dos parâmetros é dada por $\boldsymbol{d}_{\max}$, que corresponde ao autovetor associado ao maior autovalor (em módulo) de $\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta}$. O vetor $\boldsymbol{d}_{\max}$ é utilizado para identificar observações que podem estar controlando propriedades importantes na análise dos dados.

A metodologia de influência local de Cook (1986) apresenta alguns problemas, como a falta de invariância sob reparametrizações e a falta de uma medida (ponto de corte) para a tomada de decisão sobre pontos influentes.

Enfoque de influência Local de Zhu e Lee

Para introduzir a metodologia e algumas notações, segue um resumo do algoritmo EM. Seja $\mathbf{Y}_c = (\mathbf{Y}_o, \mathbf{Y}_m)$ um conjunto de dados completos com função densidade $p(\mathbf{Y}_c|\boldsymbol{\theta})$, parametrizada por um parâmetro $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, em que $\mathbf{Y}_o$ é o vetor de dados observados e $\mathbf{Y}_m$ é o vetor de dados faltantes, de uma amostra de tamanho n. De acordo com Zhu e Lee (2001), na maioria das aplicações estatísticas, a função de log-verossimilhança dos dados completos

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \log\{p(\mathbf{y}_c|\boldsymbol{\theta})\}, \tag{1.14}$$

geralmente tem forma mais simples que a função de log-verossimilhança dos dados observados

$$\ell_o(\boldsymbol{\theta}|\mathbf{y}_o) = \log\{p(\mathbf{y}_o|\boldsymbol{\theta})\} = \ell_o(\boldsymbol{\theta}).$$

O algoritmo EM consiste de dois passos: o passo E (Esperança) e o passo M (Maximização). No passo E calcula-se

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_o, \boldsymbol{\theta}^{(r)}\},$$

sendo que a esperança é tomada com respeito à distribuição condicional $p(\mathbf{y}_m|\mathbf{y}_o, \boldsymbol{\theta}^{(r)})$. No passo M, determina-se o valor $\boldsymbol{\theta}^{(r+1)}$ que maximiza $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$. Sob determinadas condições, a seqüência $\boldsymbol{\theta}^{(r)}$ obtida das iterações do algoritmo EM converge para o estimador de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$.

Considere um vetor de perturbação $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)^\top$ variando em uma região aberta $\boldsymbol{\Omega} \in \mathbb{R}^p$. Sejam $\ell_o(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_o)$ e $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ as funções de log-verossimilhança para os dados observados e para os dados completos, respectivamente, para o modelo perturbado. Assume-se que existe um $\boldsymbol{\omega}_0$ tal que $\ell_o(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_o) = \ell_o(\boldsymbol{\theta}|\mathbf{Y}_o)$ e $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{Y}_c)$, para todo $\boldsymbol{\theta}$. Cook (1986) considera a função de afastamento da verossimilhança

$$LD(\boldsymbol{\omega}) = 2[\ell_o(\widehat{\boldsymbol{\theta}}|\mathbf{y}_o) - \ell_o(\widehat{\boldsymbol{\theta}}_o(\boldsymbol{\omega})|\mathbf{y}_o)].$$

Em modelos mais complicados, nos quais a função de log-verossimilhança é intratável (por exemplo, na presença de integrais), a metodologia acima se torna inviável. Motivados por essa deficiência, Zhu e Lee (2001) propõem a função Q-afastamento como uma alternativa à função $LD(\boldsymbol{\omega})$:

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right],$$

em que $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ maximiza a função $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{Y}_o, \widehat{\boldsymbol{\theta}}]$. O gráfico de influência de $f_Q(\boldsymbol{\omega})$ é definido como $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$. Sendo $\boldsymbol{\theta}_0$ o verdadeiro valor do parâmetro

e se o maior autovalor da matriz semidefinida positiva $J_o^{-1/2} J_m J_o^{-1/2}$ converge para 0 em probabilidade, Zhu e Lee (2001) demonstram que

$$2\left[Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0|\widehat{\boldsymbol{\theta}})\right] \xrightarrow{L} \chi^2_p,$$

quando $n \to \infty$, em que $\xrightarrow{L}$ denota convergência em distribuição, J_o é a matriz de informação dos dados observados e J_m é a matriz de informação dos dados faltantes.

Seguindo a aproximação desenvolvida em Zhu e Lee (2001), a curvatura normal $C_{f_Q,\mathbf{d}}$ de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ próxima a $\boldsymbol{\omega}_0$ na direção de um vetor unitário $\mathbf{d}$ pode ser usada para resumir o comportamento local da função Q-afastamento. Pode ser mostrado que (veja Zhu e Lee, 2001)

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_o}\mathbf{d}, \quad -\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}^\top_{\boldsymbol{\omega}_o}\left[-\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\right]^{-1}\boldsymbol{\Delta}_{\boldsymbol{\omega}_0},$$

em que $\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ e $\Delta_{\boldsymbol{\omega}} = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

A curvatura normal conformal $B_{f_Q,\mathbf{d}}$ para $\boldsymbol{\omega}_0$ em uma direção unitária $\mathbf{d}$ é definida por

$$B_{f_Q,\mathbf{d}} = \frac{-2\mathbf{d}^\top \ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}\mathbf{d}}{tr(-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0})}.$$

Sob condições de regularidade, $\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$ é semidefinida positiva. Ainda, Zhu e Lee (2001) provam que $0 \leq B_{f_Q,\mathbf{d}} \leq 1$. Analogamente a Cook (1986), a expressão $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$ é a matriz fundamental para detectar observações influentes. Para tanto, considera-se a decomposição espectral de $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$ dada por

$$-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_o} = \sum_{k=1}^{q} \lambda_k \mathbf{e}_k \mathbf{e}_k',$$

em que $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_q, \mathbf{e}_q)$ são os pares (autovalores-autovetores) da matriz $-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_o}$ com $\lambda_1 \geq \ldots \geq \lambda_r, \lambda_{r+1} = \ldots = \lambda_q = 0$ e $\mathbf{e}_1, \ldots, \mathbf{e}_q$ são os vetores da base ortonormal associada. Lesaffre e Verbeke (1998) e Poon e Poon (1999) propuseram inspecionar todas $C_{f_Q,\mathbf{u}_j}$, onde $\mathbf{u}_j$ é um vetor de perturbação básica, com j-ésima entrada igual a 1 e restantes iguais a 0. Seja $\widehat{\lambda}_k = \lambda_k/(\lambda_1 + \ldots + \lambda_q)$. Como $tr(-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}) = \sum_{i=1}^{r}\lambda_i$, pode ser visto que

$$C_{f_Q,\mathbf{u}_j} = \sum_{i=1}^{r}\lambda_i e_{ij}^2, \quad B_{f_Q,\mathbf{u}_j} = \sum_{i=1}^{r}\widetilde{\lambda}_i e_{ij}^2.$$

Assim, tem-se que $B_{f_Q,\mathbf{e}_k} = \widetilde{\lambda}_k$. Segundo Zhu e Lee (2001), um autovetor $\mathbf{e}_i$ de $-2\ddot{Q}_{\boldsymbol{\omega}_o}$ é chamado m_0-influente se $B_{f_Q,\mathbf{e}_i} \geq m_0/r$. Considerando $\mathbf{e}_k^2 = (e_{k1}^2, \ldots, e_{kq}^2)$, define-se como

vetor de contribuição agregada de todos os autovetores m_0-influentes a soma ponderada

$$M(m_0) = \sum_{i:\widetilde{\lambda}_i \geq m_0/r} \widehat{\lambda}_i \mathbf{e}_i^2.$$

Em particular, quando $m_0 = 0$, $M(0) = \sum_{i=1}^{r} \widetilde{\lambda}_i \mathbf{e}_i^2$ e $M(0)_j = B_{f_Q,\mathbf{u}_j}$, para todo j e sua média é $\overline{M(0)} = 1/q$. Ver Zhu e Lee (2001) para outras propriedades teóricas de $B_{f_Q,\mathbf{u}_l}$, tal como a invariância sob reparametrização de $\boldsymbol{\theta}$.

Assim, a avaliação de casos influentes é baseada em $\{M(0)_l, l = 1, \ldots, q\}$. Zhu e Lee (2001) propõem como ponto de corte o valor $\overline{M(m_0)} + c^* DP(M(m_0))$, em que $DP(m_0)$ é o desvio padrão de $\{M(m_0)_l, l = 1, \ldots, q\}$ e c^* é uma constante arbitrária. Adicionalmente, Lee e Xu (2004) propõem usar $1/n + c^* DP(0)$ como um ponto de corte para considerar as observações influentes.

Por último, diagnósticos via exclusão de casos são também considerados neste trabalho. Medidas de diagnóstico são desenvolvidas para o caso em que o vetor $(\boldsymbol{y}_i^\top, \boldsymbol{x}_i^\top)$ é excluído. Na literatura, as medidas clássicas são a distância de Cook e o afastamento da verossimilhança. Baseado nessas idéias, Lee e Xu (2004) propõem as medidas análogas D_i^c e (LD_i^c) para a função Q. Essas medidas são, respectivamente, dadas por

$$D_i^c = (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top [-\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})](\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})$$

e

$$LD_i^c = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}_{(i)}|\widehat{\boldsymbol{\theta}}\right)\right],$$

em que $\widehat{\boldsymbol{\theta}}_{(i)}$ é o maximizador da função $Q_{(i)}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, n$. Neste trabalho, utilizamos $M(0)$ como diagnóstico para influência local e D_i^c e LD_i^c como medidas de influência global.

Técnicas Gráficas

A construção de envelopes simulados (Atkinson, 1985) é realizada basicamente através da distribuição de probabilidade de alguma função (resíduo ordinário, resíduo studentizado, formas quadráticas) conhecida do modelo proposto. Para cada observação do conjunto de dados, reamostra-se uma quantidade K fixa (por exemplo, 200, 300, 500) baseado na distribuição conhecida e toma-se a média e dois percentis (neste trabalho, adota-se sempre o quinto e o nonagésimo quinto) dessas reamostras, a fim de traçar linhas para cada uma das estatísticas mencionadas. Um bom ajuste dos dados à distribuição de referência ocorre quando todos os pontos da amostra estão contidos na banda de confiança (percentis 5 e 95).

1.5.1 Critérios de seleção de modelos

Alguns critérios comumente usados para a seleção de modelos são o critério de informação de Akaike (AIC) e o critério de informação bayesiano (BIC) propostos por Akaike (1973) e Schwarz (1978) respectivamente. Estes critérios são dados por

$$AIC = -2\ell(\widehat{\boldsymbol{\theta}}) + 2t$$

e

$$BIC = -2\ell(\widehat{\boldsymbol{\theta}}) + t\log(n),$$

onde $\widehat{\theta}$ é o estimador de máxima verossimilhança, ℓ é a função de log-verossimilhança, t é o número de parâmetros livres do modelo e n é o número de observações. A escolha do melhor modelo se faz considerando aquele que apresenta o menor valor dos critérios utilizados (AIC ou BIC).

Capítulo 2

Modelos para Dados de Contagem

2.1 Introdução

Os dados de contagem são um tipo de informação que aparecem com muita freqüência em pesquisas tanto nas Ciências Sociais quanto nas Ciências da Saúde, Engenharia, Economia, etc. Existem pesquisas com este tipo de variáveis em Demografia, Farmacologia, Biologia, Criminologia, Ciências Políticas e Ciências Econômicas, para citar alguns exemplos. Segundo Lindsay (1995), a contagem de dados define-se como o número de eventos que ocorrem numa mesma unidade de observação durante um intervalo temporal ou espacial definido. O conhecimento acerca da natureza de uma variável, assim como a identificação de suas características distribucionais são a base a partir da qual justifica-se a aplicação de um modelo estatístico determinado. Segundo Ridout *et al.* (2001), há duas estratégias para modelar um conjunto de dados: ajustar os dados ao modelo ou ajustar o modelo aos dados. A primeira das alternativas passa por adotar estratégias que façam possível a aplicação de procedimentos estatísticos a dados para os quais não foram concebidos. A estratégia mais freqüente neste caso é a transformação dos dados tendo como objetivo a aplicação do modelo linear geral. Por outro lado, segundo Lindsay (1995), ajustar o modelo aos dados mediante transformações não só incrementa problemas novos tal como viés na estimação ou dificuldade de interpretação dos resultados, mas também não resolve os problemas de ajuste às condições de aplicação do modelo linear geral. Neste sentido, tal como indicam Ridout *et al.* (2001), quando cumprem-se as suposições do modelo estatístico empregado, habitualmente os coeficientes descrevem corretamente a relação. No entanto,

quando não se cumprem ditas suposições, como por exemplo ao empregar regressão linear para analisar dados procedentes de distribuições não normais, as estimações podem não ser válidas, ou seja, podem identificar-se relações inexistentes (erros tipo I) ou subvalorizar relações existentes (erros tipo II). Uma técnica alternativa que vem sendo estudada para este tipo de problemas é a chamada regressão para dados inflacionados de zeros. Este tipo de modelo contempla a sobredispersão ou subdispersão que quase sempre existem nos dados pela excessiva presença de zeros, fato que não é levado em conta nos outros tipos de regressão.

A seguir definiremos os modelos utilizados neste trabalho.

2.1.1 O Modelo Binomial

Em muitos experimentos lidamos com situações em que um determinado item é analisado e dele obtemos como resposta "sucesso" ou "fracasso" (sucesso é o evento de interesse); portanto, a variável resposta é binária, isto é, admite apenas dois resultados. É comum encontrarmos situações práticas com esse tipo de variável resposta, nas quais cada realização do experimento é chamada ensaio.

A distribuição binomial é formada quando o número de ensaios (n) é fixo, o parâmetro π (a probabilidade de sucesso) é a mesma para cada ensaio e os ensaios são todos independentes. Uma variável aleatória Y tem distribuição binomial com parâmetros n e π, denotada por $bin(n, \pi)$, se sua função de probabilidade (fdp) é dada por

$$f(y; n, \pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \qquad \text{para} \quad y = 0, 1, \ldots, n,$$

com $0 < \pi < 1$.

A média e a variância da distribuição binomial são dadas por $E[Y] = n\pi$ e $V[Y] = n\pi(1-\pi)$.

O modelo binomial é bastante utilizado em ensaios do tipo dose-resposta (veja, Cordeiro e Demétrio, 2007). Ensaios do tipo dose-resposta são aqueles em que uma determinada droga é administrada em k diferentes doses, $d_1, \ldots, d_k$, respectivamente, a $n_1, \ldots, n_k$ indivíduos. Suponha que cada indivíduo responde, ou não, à droga, tal que a resposta é quantal (tudo ou nada, isto é, 1 ou 0), obtendo-se, após um período especificado, $y_1, \ldots, y_k$ indivíduos que respondem à droga. Por exemplo, quando um inseticida é aplicado a um determinado número de insetos, eles respondem (morrem), ou não (sobrevivem), à dose aplicada; quando uma droga benéfica é administrada a um grupo de pacientes, eles podem melhorar (sucesso), ou não (falha). Dados resultantes desse tipo

de ensaio podem ser considerados como provenientes de uma distribuição binomial com probabilidade π_i, que é a probabilidade de ocorrência (sucesso) do evento sob estudo, ou seja, o número de sucessos Y_i tem distribuição binomial $bin(n_i; \pi_i)$, $i = 1, \ldots, n$.

2.1.2 O Modelo Poisson

A distribuição de Poisson de parâmetro λ é uma distribuição de probabilidade discreta. A variável aleatória Y representa o número de ocorrências de um evento em um intervalo de tempo especificado e λ é o número médio de eventos que ocorrem neste intervalo. O intervalo pode ser de tempo, área, volume ou outra unidade. Segundo Jonhson e Kotz (1969), uma variável aleatória Y tem distribuição de Poisson com parâmetro λ, denotada por $P(\lambda)$, se sua função de probabilidade é dada por

$$f(y; \lambda) = P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y = 0, 1, 2, \ldots,$$

com $\lambda > 0$ o único parâmetro especificando a distribuição.
A esperança e a variância da distribuição de Poisson são dadas por $E[Y] = \lambda$ e $V[Y] = \lambda$.

Dados de contagem surgem de várias formas, podendo ser, por exemplo, o número de lagartas observadas na cultura de milho para se verificar a eficácia do milho geneticamente modificado no controle da praga. Também é usado o modelo Poisson nos ensaios de diluição para se estimar a concentração λ de um organismo (número por unidade de volume, de área, de peso, etc.) em uma amostra (veja Demétrio, 2002). Quando a contagem direta não é possível, mas a presença ou ausência do organismo em sub-amostras pode ser detectada pode-se, também, estimar λ. Em geral, registrar a presença ou ausência fica mais econômico do que fazer a contagem. Por exemplo, pode-se detectar se uma determinada bactéria está presente, ou não, em um líquido por um teste de cor, ou se um fungo está presente, ou não, em uma amostra de solo, plantando-se uma planta susceptível nesse solo e verificando se a planta apresenta sintomas da doença.

2.1.3 O Modelo Binomial Negativa

Seja um experimento estatístico com dois resultados possíveis: sucesso ou fracasso, em que sucesso ocorre com probabilidade p e fracasso ou falha ocorre com probabilidade $q = 1 - p$. Se o experimento é repetido indefinidamente e os ensaios são independentes, então a variável aleatória Y^* denotando o número de ensaios em que ocorre o k-ésimo sucesso tem uma distribuição Binomial Negativa com parâmetros k e p.

A função de probabilidades de Y^* é dada por

$$f(y^*) = \binom{y^*-1}{\phi-1} p^{\phi}(1-p)^{y^*-\phi}, \qquad y^* = \phi, \phi+1, \phi+2, \ldots.$$

A distribuição Binomial Negativa pode ser escrita de diversas maneiras dependendo da parametrização utilizada. Neste trabalho será utilizada a distribuição Binomial Negativa na notação de Nelder e Wedderburn (1972), em que $p = \dfrac{\mu}{\mu+\phi}$, $0 < p \leq 1$, o parâmetro ϕ^{-1} é o parâmetro de dispersão e é assumido que, $\phi \geq 0$. Fazendo $y = y^* - \phi$ temos a distribuição binomial negativa com parâmetros μ e ϕ, denotado por $NB(\mu, \phi)$, cuja função de probabilidade é dada por

$$f(y; \mu, \phi) = \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\mu}{\mu+\phi}\right)^{y} \left(\frac{\phi}{\mu+\phi}\right)^{\phi}, \qquad y = 0, 1, 2, \ldots, \tag{2.1}$$

em que $\Gamma(\cdot)$ é a função gama, i.e., $\Gamma(x) = \displaystyle\int_0^{\infty} t^{x-1}e^{-t}dt, \quad x > 0.$

Seja Y uma variável aleatória com distribuição binomial negativa com parâmetros μ e ϕ, e fdp dada em (2.1); então, o valor esperado de Y é dado por $E[Y] = \mu$ e a variância $V[Y] = \dfrac{\mu(\phi+\mu)}{\phi} = \mu + \dfrac{\mu^2}{\phi}$ (veja Paula, 2004).

Nota-se que a variância da distribuição Binomial Negativa tem um termo adicional positivo $\dfrac{\mu^2}{\phi}$, comparativamente com a variância da distribuição de Poisson, que, em muitos casos, ajuda a ajustar melhor um conjunto de dados onde existe sobredispersão. A distribuição Binomial Negativa aproxima-se à distribuição de Poisson quando ϕ^{-1} tende a 0 (Cameron e Trivedi, 1998).

Temos, por exemplo, o estudo da relação entre acidentes de caminhões e o desenho geométrico das seções da estrada (Miaou, 1994) e o estudo da distribuição Binomial Negativa na análise de fenômenos recorrentes (Navarro *et al.* 2001).

2.1.4 O Modelo Beta-Binomial

A distribuição Beta-Binomial resulta de uma mistura da distribuição Binomial e a distribuição Beta. Suponhamos que

a) dado π, Y tem uma distribuição Binomial, bin(n,π), e

b) π tem distribuição Beta.

Nesse caso temos que, dado π,

$$P(y|\pi) = \binom{n}{y} \pi^{y}(1-\pi)^{n-y}, \qquad y = 0, 1, \ldots, n, \tag{2.2}$$

e a função de densidade de probabilidade Beta com parâmetros $\alpha > 0$ e $\beta > 0$, denotada por Beta(α, β), é da forma

$$f(\pi; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1}, \quad 0 < \pi < 1, \tag{2.3}$$

A distribuição Beta tem média e variância, respectivamente, dadas por

$$E[\pi] = \frac{\alpha}{\alpha+\beta} \quad \text{e} \quad V[\pi] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Quando α e β excedem 1,0, a distribuição é unimodal, assimétrica à direita quando $\alpha < \beta$, assimétrica à esquerda com $\alpha > \beta$, e simétrica quando $\alpha = \beta$. Se $\alpha = \beta = 1$ ela se reduz à distribuição uniforme. De (2.2) e (2.3) segue a distribuição conjunta

$$\begin{aligned} f(y, \pi) &= P(y|\pi) f(\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1} \\ &= \binom{n}{y} \frac{\pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}}{B(\alpha,\beta)}, \qquad y = 0, 1, \ldots, n; \quad 0 < \pi < 1, \end{aligned}$$

em que $B(u, v) = \dfrac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$. Assim, a função de probabilidade marginal de Y é da forma

$$\begin{aligned} f(y; n, \alpha, \beta) &= \int_0^1 f(y, \pi) d\pi = \int_0^1 \binom{n}{y} \frac{\pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}}{B(\alpha,\beta)} d\pi \\ &= \binom{n}{y} \frac{B(\alpha+y, \beta+n-y)}{B(\alpha,\beta)} \int_0^1 \frac{\pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}}{B(\alpha+y, \beta+n-y)} d\pi \\ &= \binom{n}{y} \frac{B(\alpha+y, \beta+n-y)}{B(\alpha,\beta)}, \end{aligned}$$

que é a distribuição Beta-Binomial, com os dois primeiros momentos dados por

$$E[Y] = E[E[Y|\pi]] = E[n\pi] = \frac{n\alpha}{\alpha+\beta}$$

e

$$\begin{aligned} V[Y] &= E[V[Y|\pi]] + V[E[Y|\pi]] = E[n\pi(1-\pi)] + V[n\pi] \\ &= nE[\pi - \pi^2] + n^2 V[\pi]. \end{aligned}$$

Note que

$$E[\pi^2] = V[\pi] + (E[\pi])^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \left(\frac{\alpha}{\alpha+\beta}\right)^2;$$

logo,

$$\begin{aligned} V[Y] &= n\left[\frac{\alpha}{\alpha+\beta} - \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2\right] \\ &\quad + n^2\left[\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\ &= \frac{n\alpha}{\alpha+\beta}\left[1 - \frac{\beta}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha}{\alpha+\beta} + \frac{n\beta}{(\alpha+\beta)(\alpha+\beta+1)}\right] \\ &= \frac{n\alpha\beta(n+\alpha+\beta)}{(\alpha+\beta)^2(\alpha+\beta+1)}. \end{aligned}$$

Neste trabalho, por simplicidade utilizaremos uma parametrização alternativa da distribuição Beta. Assim, sejam $\mu = \dfrac{\alpha}{\alpha+\beta}$ e $\phi = \alpha + \beta$, i.e., $\alpha = \mu\phi$ e $\beta = (1-\mu)\phi$. Ou seja, π tem distribuição Beta$(\mu\phi, (1-\mu)\phi)$. Desta forma, após certa álgebra, a distribuição Beta-Binomial com parâmetros μ e ϕ, denotada por $BB(\mu, \phi)$, é da forma

$$f(y; n, \mu, \phi) = \binom{n}{y} \frac{B(\mu\phi + y, (1-\mu)\phi + n - y)}{B(\mu\phi, (1-\mu)\phi)},$$

em que $0 < \mu < 1$ e $\phi > 0$.

Com esta reparametrização, os primeiros dois momentos são

$$E[Y] = n\mu, \quad \text{e} \quad V[Y] = \frac{n\mu(1-\mu)(n+\phi)}{\phi+1}.$$

Quando $\phi^{-1} \to 0$, $V[\pi] \to 0$ e a distribuição Beta converge para uma distribuição degenerada em μ. Então, $V[Y] \to n\mu(1-\mu)$ e a distribuição Beta-Binomial converge para a distribuição $bin(n, \mu)$.

Altham (2002), por exemplo, faz um estudo utilizando o modelo beta-binomial para ajustar dados de tratamento de fertilização *in vitro* de 52 clínicas do Reino Unido e Slaton *et al.* (2000) fazem um estudo em toxicologia usando ensaios do tipo dose-resposta.

2.2 Modelos para Dados Inflacionados de Zeros

Os modelos para dados inflacionados de zeros são úteis para modelar dados resultantes do processamento de fabricação (veja Lambert, 1992) entre muitas outras aplicações. É assumido que com probabilidade p a única observação possível é 0, e com probabilidade $1-p$ é observada uma variável aleatória que descreve contagens de defeitos no estado imperfeito (estado em que o número de defeitos do produto segue uma distribuição de

probabilidade). Por exemplo, quando o equipamento de fabricação está alinhado corretamente (estado perfeito), pode não haver nenhum defeito. Senão, os defeitos podem ocorrer de acordo com uma distribuição do estado imperfeito. As contagens do número de defeitos no estado imperfeito poderiam seguir uma distribuição de Poisson, ou binomial negativa, ou outras distribuições, mas a maioria das pesquisas atuais usa a distribuição de Poisson.

Nesta seção nós propomos uma estrutura geral para os modelos inflacionados de zeros (ZI), que assume somente que a distribuição do estado imperfeito tem suporte nos inteiros não negativos e satisfaz apropriadas condições de regularidade.

Dados com excesso de zeros são vistos comumente em experimentos para melhorar a qualidade de fabricação de produtos eletrônicos, na investigação médica de pacientes de HIV com comportamentos de alto risco, no estudo agrícola do número de insetos por folha, entre outros.

Como foi visto anteriormente, a maioria das pesquisas atuais supõe que o estado imperfeito tem distribuição de Poisson, fato que exclui algumas outras distribuições úteis, tais como a distribuição binomial negativa. Nós introduzimos o conceito de modelos inflacionados de zeros generalizados que supõem que as contagens no estado imperfeito têm distribuição em uma classe mais ampla. Uma primeira classe interessante de distribuições consiste em distribuições contínuas com o suporte comum $[0,+\infty]$. Uma outra classe de distribuições são aquelas com o suporte nos inteiros não negativos. Nós estudamos a segunda das classes nesta dissertação. Resultados semelhantes para a primeira classe anterior poderiam ser derivados sem muita dificuldade, como por exemplo, Ospina (2008) faz um estudo dos modelos de regressão beta inflacionados.

Supor que o estado perfeito Y_0 é 0 com probabilidade 1 e o estado imperfeito Y_1 é uma variável aleatória que toma inteiros não negativos com a seguinte função de probabilidade

$$P(Y_1 = y) = g(y, \boldsymbol{\mu}), \qquad y = 0, 1, 2, \ldots,$$

em que $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$ é um vetor de parâmetros desconhecidos em um subconjunto aberto D do espaço euclideano p-dimensional $\mathbb{R}^p$. A distribuição dos $Y_1's$ é chamada a distribuição do estado imperfeito.

Consideremos a mistura de Y_0 e Y_1 com pesos p e $1-p$ com a distribuição Bernoulli(p) onde $0 \leq p < 1$ e assumamos que a função de probabilidade da mistura é $f\left(y, \begin{pmatrix} p \\ \boldsymbol{\mu} \end{pmatrix}\right)$.

Note que nós excluímos o caso que $p = 1$ pois é um caso trivial e não é geralmente de interesse. No entanto, nós gostaríamos de incluir o caso em que $p = 0$, em que somente o estado imperfeito existe. Por exemplo, nós podemos estar interessados no teste de

$p = 0$ contra $p > 0$, que testa se os dados são de um modelo inflacionado de zeros ou da distribuição do estado imperfeito.

Definamos $\boldsymbol{\theta} = \begin{pmatrix} p \\ \boldsymbol{\mu} \end{pmatrix}$ e $\boldsymbol{\Theta} = [0,1) \times D$.

É fácil ver que

$$f(y;\boldsymbol{\theta}) = \begin{cases} p + (1-p)g(0,\boldsymbol{\mu}) & , \text{ para } y = 0, \\ (1-p)g(y,\boldsymbol{\mu}) & , \text{ para } y = 1,2,\ldots \end{cases} \tag{2.4}$$

A mistura é denotada como $Y \sim ZI(\boldsymbol{\theta}, g)$ ou $Y \sim ZI(\boldsymbol{\theta})$ se não há nenhuma necessidade de especificar a função g. Seja $\boldsymbol{\theta}^0 = \begin{pmatrix} p^0 \\ \boldsymbol{\mu}^0 \end{pmatrix}$ o parâmetro verdadeiro, onde $p^0 \in [0,1)$ e $\boldsymbol{\mu}^0 \in D$. Daqui para adiante nós supomos que as observações $y_1, \ldots, y_n$ são *iid* com a distribuição ZI($\boldsymbol{\theta}^0$). Como não conhecemos o parâmetro verdadeiro $\boldsymbol{\theta}^0$, somente podemos assumir que $y_1, \ldots, y_n$ são *iid* com a distribuição ZI($\boldsymbol{\theta}$). Uma de nossas tarefas principais é estimar o parâmetro $\boldsymbol{\theta}$, isto é, p e $\boldsymbol{\mu}$.

A função de verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\theta}) = L(\boldsymbol{\theta};\boldsymbol{y}) &= \prod_{i=1}^{n} f\left(y_i; \begin{pmatrix} p \\ \boldsymbol{\mu} \end{pmatrix}\right) \\ &= \prod_{y_i=0} [p + (1-p)g(0,\boldsymbol{\mu})] \times \prod_{y_i>0} [(1-p)g(y_i,\boldsymbol{\mu})] \\ &= [p+(1-p)g(0,\boldsymbol{\mu})]^{\sum \mathbb{I}_{\{y_i=0\}}} \times (1-p)^{\sum \mathbb{I}_{\{y_i>0\}}} \times \prod_{y_i>0} g(y_i,\boldsymbol{\mu}), \end{aligned} \tag{2.5}$$

em que $\mathbb{I}_{\{y_i=0\}}$ e $\mathbb{I}_{\{y_i>0\}}$ são funções indicadoras. A função indicadora (Casella e Berger, 2001) de um conjunto A, denotado por $\mathbb{I}_A(x)$ é a função

$$\mathbb{I}_A(x) = \begin{cases} 1 & , \quad x \in A \\ 0 & , \quad x \notin A \end{cases}$$

Utilizando (2.5) temos que a função de log-verossimilhança tem a forma

$$\begin{aligned} \ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\boldsymbol{y}) \;\; &= \;\; \log\left[L(\boldsymbol{\theta})\right] = \sum_{i=1}^{n} \mathbb{I}_{\{y_i=0\}} \log[p + (1-p)g(0,\boldsymbol{\mu})] \\ &\quad + \sum_{i=1}^{n} \mathbb{I}_{\{y_i>0\}} \log(1-p) + \sum_{i=1}^{n} \mathbb{I}_{\{y_i>0\}} \log\left[g(y_i,\boldsymbol{\mu})\right]. \end{aligned} \tag{2.6}$$

2.3 Modelos de Regressão para Dados Inflacionados de Zeros

Os modelos inflacionados de zeros têm sido usados para a modelagem de dados nas mais diversas áreas tais como Agricultura, Medicina, Sociología, Odontologia, Econometria entre outras. A pioneira no uso de modelos de regressão para contagens com distribuição inflacionada de zeros foi Lambert (1992). Ela propôs uma mistura finita das distribuições Bernoulli e Poisson para modelar o excesso de zeros em dados de contagem envolvendo defeitos de fabricação de placas de circuito impresso.

Um dos modelos mais utilizados em dados de contagens com excessos de zeros é o modelo Poisson Inflacionado de Zeros (ZIP). Neste contexto, Ridout *et al.* (1998) apresentam uma revisão de literatura e discutem uma metodologia geral para modelar dados de contagem inflacionados de zeros. Broek (1995) apresenta uma estatística escore para testar se o modelo ZIP se ajusta melhor a um conjunto de dados que uma distribuição de Poisson usual. Hall (2000) estuda os modelos binomial e Poisson inflacionados de zeros incluindo efeito aleatório, e alguns outros autores fazem uso do modelo ZIP nas mais diversas aplicações.

O modelo Binomial Negativa Inflacionado de Zeros (ZINB) também é uma boa alternativa para dados de contagens, visto que nos casos de dados de contagens com muitos zeros a sobredispersão pode acarretar maiores problemas e esse modelo pode ser mais adequado do que o modelo ZIP. Na literatura existem trabalhos recentes sobre o modelo ZINB. Dentre eles, Lewsey e Thomson (2004) mostram um estudo comparativo dos modelos ZINB e ZIP, e concluem em um estudo relativo a dentes com cáries que o modelo ZINB resulta em um melhor ajuste do que o modelo ZIP; Yau *et al.* (2003) ajustam um modelo de regressão ZINB para analisar dados referentes à recuperação de pacientes que se submeteram a uma cirurgia no fígado; Ridout *et al.* (2001) abordam neste contexto um teste escore do modelo ZIP contra uma alternativa ZINB.

Uma outra distribuição para dados de contagens com excesso de zeros, definida de maneira similar aos modelos ZIP e ZINB é a Binomial Inflacionada de zeros (ZIB), que pode ser encontrada em Hall (2000). Quando se trata de controlar a sobredispersão no caso de um modelo ZIB temos uma outra distribuição que é a Beta Binomial Inflacionada de Zeros (ZIBB).

Assumamos que D é um conjunto aberto na reta $\mathbb{R}$. Assumamos que as respostas

$y_1, y_2, \ldots, y_n$ são independentes e

$$y_i \sim ZI(\boldsymbol{\theta}_i),$$

em que $\boldsymbol{\theta}_i = \begin{pmatrix} p_i \\ \mu_i \end{pmatrix}$. Ou seja, a função massa de probabilidade dos $y_i's$ é

$$f(y_i;\boldsymbol{\theta}_i) = \begin{cases} p_i + (1-p_i)g(0,\mu_i) & , \quad y_i = 0 \\ (1-p_i)g(y_i,\mu_i) & , \quad y_i = 1,2,\ldots \end{cases} . \tag{2.7}$$

Como antes, nós assumimos que $p_i \in [0,1)$ e $\mu_i \in D$. Sejam h_1 e h_2 as funções de ligação para p_i e μ_i. Supomos que os parâmetros $\mu_1, \ldots, \mu_n$ e $p_1, \ldots, p_n$ se relacionam com as matrizes de covariáveis de posto completo $\mathbf{B}$ e $\mathbf{G}$, respectivamente, através das funções de ligação h_1 e h_2, de modo que

$$\begin{pmatrix} h_1(\mu_1) \\ \vdots \\ h_1(\mu_n) \end{pmatrix} = \boldsymbol{\eta} = \mathbf{B}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{B}_1\boldsymbol{\beta} \\ \vdots \\ \mathbf{B}_n\boldsymbol{\beta} \end{pmatrix}$$

e

$$\begin{pmatrix} h_2(p_1) \\ \vdots \\ h_2(p_n) \end{pmatrix} = \boldsymbol{\zeta} = \mathbf{G}\boldsymbol{\gamma} = \begin{pmatrix} \mathbf{G}_1\boldsymbol{\gamma} \\ \vdots \\ \mathbf{G}_n\boldsymbol{\gamma} \end{pmatrix},$$

em que $\mathbf{B}$ é uma matriz de valores fixos conhecidos de dimensão $(n \times p)$ e $\mathbf{G}$ é uma matriz de valores fixos conhecidos de dimensão $(n \times q)$. Assumamos que as funções inversas h_1^{-1} e h_2^{-1} existem. Portanto,

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} h_1^{-1}(\mathbf{B}_1\boldsymbol{\beta}) \\ \vdots \\ h_1^{-1}(\mathbf{B}_n\boldsymbol{\beta}) \end{pmatrix}$$

e

$$\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} h_2^{-1}(\mathbf{G}_1\boldsymbol{\gamma}) \\ \vdots \\ h_2^{-1}(\mathbf{G}_n\boldsymbol{\gamma}) \end{pmatrix}.$$

O modelo construído aqui é chamado um modelo de regressão inflacionado de zeros generalizado. As matrizes de covariáveis $\mathbf{B}$ e $\mathbf{G}$ podem ou não ser as mesmas. A função de verossimilhança é da forma

$$L(\boldsymbol{\theta}) = \prod_{y_i=0} \big[p_i + (1-p_i)g(0,\mu_i)\big] \times \prod_{y_i>0} \big[(1-p_i)g(y_i,\mu_i)\big].$$

Assim, a função de log-verossimilhança é

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\Big\{\mathbb{I}_{\{y_i=0\}}\log[p_i + (1-p_i)g(0,\mu_i)] + \mathbb{I}_{\{y_i>0\}}\log[(1-p_i)g(y_i,\mu_i)]\Big\},$$

Fazendo $u_i = \mathbb{I}_{\{y_i=0\}}$, temos que

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\Big\{u_i\log\Big[p_i + (1-p_i)g(0,\mu_i)\Big] + (1-u_i)\log(1-p_i) + (1-u_i)\log\Big[g(y_i,\mu_i)\Big]\Big\}.$$

Reparametrizando com $w_i = \dfrac{p_i}{1-p_i}$, temos que $p_i = \dfrac{w_i}{1+w_i}$ e $1-p_i = \dfrac{1}{1+w_i}$. Então, temos que a função de probabilidade definida em (2.7) é expressa como

$$f(y_i;\boldsymbol{\theta}_i) = \begin{cases} \dfrac{w_i + g(0,\mu_i)}{1+w_i} & , \quad y_i = 0 \\ \dfrac{g(y_i,\mu_i)}{1+w_i} & , \quad y_i = 1,2,\ldots \end{cases} \tag{2.8}$$

e a função de log-verossimilhança fica

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\left\{u_i\log\left[\frac{w_i+g(0,\mu_i)}{1+w_i}\right] + (1-u_i)\log\left(\frac{1}{1+w_i}\right) + (1-u_i)\log\Big[g(y_i,\mu_i)\Big]\right\} \\
&= \sum_{i=1}^{n}\Big\{u_i\log\Big[w_i+g(0,\mu_i)\Big] - u_i\log(1+w_i) - \log(1+w_i) + u_i\log(1+w_i) \\
&\quad +(1-u_i)\log\Big[g(y_i,\mu_i)\Big]\Big\} \\
&= \sum_{i=1}^{n}\Big\{u_i\log\Big[w_i+g(0,\mu_i)\Big] - \log(1+w_i) + (1-u_i)\log\Big[g(y_i,\mu_i)\Big]\Big\}. \qquad (2.9)
\end{aligned}$$

2.3.1 Modelo Binomial Inflacionado de Zeros (ZIB)

O modelo de regressão binomial com excesso de zeros tem a seguinte forma:

$$P(Y_i = y_i) = \begin{cases} p_i + (1-p_i)(1-\pi_i)^{n_i} & , \quad y_i = 0, \\ (1-p_i)\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i} & , \quad y_i = 1,2,\ldots,n_i, \end{cases}$$

onde $0 \leq p_i < 1$, $0 < \pi_i < 1$ e seus primeiros momentos são, respectivamente, dados por $E[Y_i] = (1-p_i)n_i\pi_i$ e $V[Y_i] = (1-p_i)n_i\pi_i[1-\pi_i(1-p_in_i)]$.

Os parâmetros $\boldsymbol{p} = (p_1,\ldots,p_n)^\top$ e $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_n)^\top$ são modelados usualmente pelas funções de ligação $\log\left(\dfrac{p_i}{1-p_i}\right) = \mathbf{G}_i\boldsymbol{\gamma}$ e $\log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \mathbf{B}_i\boldsymbol{\beta}$. Neste caso, $\phi = 1$.

Entre os diversos estudos realizados tem-se, por exemplo, Hall (2000), que estuda o modelo binomial inflacionado de zeros considerando efeito aleatório para dados de horticultura.

2.3.2 Modelo Poisson Inflacionado de Zeros (ZIP)

No caso do modelo de regressão de Poisson com excesso de zeros, a função de probabilidade com parâmetro λ_i é expressa na forma

$$P(Y_i = y_i) = \begin{cases} p_i + (1-p_i)e^{-\lambda_i} & , \quad y_i = 0, \\ (1-p_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & , \quad y_i = 1, 2, \ldots, \end{cases}$$

em que $0 \leq p_i < 1$, $\lambda_i > 0$ e a esperança e a variância neste modelo são, respectivamente, dadas por $E[Y_i] = (1-p_i)\lambda_i$ e $V[Y_i] = \lambda_i(1-p_i)(1+p_i\lambda_i)$.

Seus parâmetros $\boldsymbol{p} = (p_1, \ldots, p_n)^\top$ e $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^\top$ são modelados usualmente pelas funções de ligação $\log\left(\dfrac{p_i}{1-p_i}\right) = \mathbf{G}_i\boldsymbol{\gamma}$ e $\log(\lambda_i) = \mathbf{B}_i\boldsymbol{\beta}$, respectivamente. Também neste caso, $\phi = 1$.

O modelo Poisson inflacionado de zeros, sem covariáveis, é discutido por vários autores, entre eles, Xie *et al.* (2001), Cameron e Trivedi (1998) e Broek (1995), mas é Lambert (1992) quem primeiro apresenta este modelo associado ao uso de covariáveis em uma aplicação sobre o número de defeitos em um processo industrial. Em seguida, outros autores como Ridout *et al.* (1998), Hall (2000), Cheung (2002) e Famoye e Singh (2006) ajustam o modelo ZIP com covariáveis a dados provenientes de contagem aplicado a diversas áreas de pesquisa. Por exemplo, Lee *et al.* (2001) usam o modelo ZIP para estudar os ferimentos ocupacionais e Böhning *et al.* (1999) estudam prevenção de caries em Epidemiologia Dental.

2.3.3 Modelo Binomial Negativa Inflacionado de Zeros (ZINB)

A distribuição ZINB é o resultado de misturar uma distribuição binomial negativa e uma distribuição degenerada em zero, e é dada por

$$P(Y_i = y_i) = \begin{cases} p_i + (1-p_i)\left(\dfrac{\phi}{\mu_i+\phi}\right)^{\phi} & , y_i = 0, \\ (1-p_i)\dfrac{\Gamma(\phi+y_i)}{\Gamma(y_i+1)\Gamma(\phi)}\left(\dfrac{\mu_i}{\mu_i+\phi}\right)^{y_i}\left(\dfrac{\phi}{\mu_i+\phi}\right)^{\phi} & , y_i = 1, 2, \ldots, \end{cases}$$

em que $0 \leq p_i < 1$, $\mu_i > 0$ e $\phi > 0$. A esperança e a variancia da distribuição ZINB são, respectivamente, $E[Y_i] = (1-p_i)\mu_i$ e $V[Y_i] = (1-p_i)\left(1 + \dfrac{\mu_i}{\phi} + p_i\mu_i\right)\mu_i$. Temos que esta distribuição aproxima a ZIP e a binomial negativa quando $\phi \to \infty$ e $p_i \to 0$, respectivamente. Se ambos $1/\phi$ e $p_i \approx 0$ então a distribuicão ZINB é reduzida à distribuição de Poisson.

Os parâmetros $\boldsymbol{p} = (p_1, \ldots, p_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ são modelados usualmente pelas funções de ligação $\log\left(\dfrac{p_i}{1-p_i}\right) = \mathbf{G}_i\boldsymbol{\gamma}$ e $\log(\mu_i) = \mathbf{B}_i\boldsymbol{\beta}$, respectivamente.

Estudos diversos têm sido feitos sobre o modelo ZINB. Por exemplo, Ridout *et al.* (2001) propuseram um teste escore para testar o modelo ZIP contra uma alternativa de que o modelo correto é o ZINB e Yau *et al.* (2003) mostram um estudo de dados de contagens com sobredispersão e excesso de zeros, usando uma mistura do modelo de regressão binomial negativa inflacionado de zeros.

2.3.4 Modelo Beta–Binomial Inflacionado de Zeros (ZIBB)

A distribuição Beta–Binomial é uma extensão da distribuição Binomial, no sentido de que o parâmetro Binomial π não é constante para todo grupo observado (unidade amostral). O modelo original da distribuição Beta–Binomial foi proposto e obtido por Skellam (1948) a partir da pressuposição de que o número de sucessos y em n ensaios binomiais por grupo era independente para todo grupo; isto é, a probabilidade de se detectar y sucessos em n observações varia de grupo a grupo. Tal variabilidade era descrita por uma distribuição Beta com parâmetros α e β. A distribuição resultante é a Beta–Binomial. Uma formulação do modelo Beta–Binomial com excesso de zeros para $\alpha_i = \mu_i\phi$ e $\beta_i = (1-\mu_i)\phi$ (veja seção 2.1.4) é

$$P(Y_i = y_i) = \begin{cases} p_i + (1-p_i)\dfrac{B(\mu_i\phi, n_i + (1-\mu_i)\phi)}{B(\mu_i\phi, (1-\mu_i)\phi)}, & y_i = 0, \\ (1-p_i)\dbinom{n_i}{y_i}\dfrac{B(y_i + \mu_i\phi, n_i + (1-\mu_i)\phi - y_i)}{B(\mu_i\phi, (1-\mu_i)\phi)}, & y_i = 1, \ldots, n_i, \end{cases}$$

em que $0 \leq p_i < 1$ e dado que $\mu_i\phi > 0$ e $(1-\mu_i)\phi > 0$, tem-se que $0 < \mu_i < 1$ e $\phi > 0$. A esperança e a variância do modelo ZIBB são, respectivamente, $E[Y_i] = (1-p_i)n_i\mu_i$ e $V[Y_i] = (1-p_i)\left[\dfrac{n_i\mu_i(1-\mu_i)(n_i+\phi)}{\phi+1} + p_i n_i^2\mu_i^2\right]$.

Os parâmetros $\boldsymbol{p} = (p_1, \ldots, p_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ são modelados usualmente pelas funções de ligação $\log\left(\dfrac{p_i}{1-p_i}\right) = \mathbf{G}_i\boldsymbol{\gamma}$ e $\log\left(\dfrac{\mu_i}{1-\mu_i}\right) = \mathbf{B}_i\boldsymbol{\beta}$.

O modelo Beta-Binomial inflacionado de zeros é discutido por vários autores, entre eles, Deng e Paul (2005) em uma aplicação sobre as propriedades antiarrítmicas de uma droga administrada a 12 pacientes.

2.4 Estimação por Máxima Verossimilhança

Os modelos inflacionados de zeros podem ter seus parâmetros estimados utilizando o método de máxima verossimilhança.

Para escrever a função de verossimilhança, considera-se o caso particular do modelo binomial inflacionado de zeros, supondo que valores iguais a zero podem ocorrer devido a um processo binomial ou a um estado perfeito. Usando a equação (2.8), o modelo binomial inflacionado de zeros pode ser expresso como

$$f(y_i;\boldsymbol{\theta}_i) = \begin{cases} \dfrac{w_i + (1-\pi_i)^{n_i}}{1+w_i} & , \quad y_i = 0 \\ \dfrac{1}{1+w_i}\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i} & , \quad y_i = 1,2,\ldots,n_i \end{cases}$$

Segundo a equação (2.9), a função de log-verossimilhança para o modelo ZIB é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left\{ u_i \log\left[w_i + (1-\pi_i)^{n_i}\right] - \sum_{i=1}^{n}\log(1+w_i) \right.$$
$$\left. + \sum_{i=1}^{n}(1-u_i)\left[y_i \log\pi_i + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}\right]\right\},$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. A modelagem dos parâmetros $\boldsymbol{\pi}$ e $\boldsymbol{p}$ (reparametrizado como $\boldsymbol{w}$), envolve covariáveis associadas ao modelo através das funções de ligação

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i = \mathbf{B}_i\boldsymbol{\beta} \quad \text{e} \quad \log(w_i) = \zeta_i = \mathbf{G}_i\boldsymbol{\gamma},$$

$i = 1,\ldots,n$, sendo $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ os vetores dos parâmetros desconhecidos e, portanto, a função de log-verossimilhança para o modelo com covariáveis é dada por $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$. As estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ podem ser obtidas usando o método do escore, o método de Newton-Raphson ou o algoritmo EM.

Método do Escore de Fisher

O método do escore de Fisher é mais apropriado do que o método de Newton-Raphson, pois a derivada de segunda ordem de $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$ em relação a $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ pode ser simplificada usando as esperanças das derivadas de segunda ordem (Jansakul, 2001). A função escore é dada por

$$\mathbf{U}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{U}(\boldsymbol{\beta}) \\ \mathbf{U}(\boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}} \\ \dfrac{\partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}} \end{bmatrix}.$$

Assim, para $j = 1, \ldots, p$ temos para o caso binomial inflacionado de zeros que

$$U_j(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \frac{\partial \ell_i(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \pi_i} \frac{\partial \pi_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \Big\{ s_i B_{ij} \Big\}$$

em que

$$s_i = -\frac{u_i n_i \pi_i (1-\pi_i)^{n_i}}{w_i + (1-\pi_i)^{n_i}} + (1-u_i)(y_i - n_i \pi_i), \qquad i = 1, \ldots, n$$

e, para $r = 1, \ldots, q$, temos

$$U_r(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_r} = \sum_{i=1}^{n} \frac{\partial \ell_i(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial w_i} \frac{\partial w_i}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial \gamma_r} = \sum_{i=1}^{n} \Big\{ t_i G_{ir} \Big\}$$

em que

$$t_i = \frac{u_i w_i}{w_i + (1-\pi_i)^{n_i}} - \frac{w_i}{1+w_i}, \qquad i = 1, \ldots, n,$$

Se definimos os vetores $\boldsymbol{M}_\beta = (s_1, \ldots, s_n)^\top$ e $\boldsymbol{M}_\gamma = (t_1, \ldots, t_n)^\top$ podemos escrever

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{B}^\top \boldsymbol{M}_\beta \qquad \text{e} \qquad \mathbf{U}(\boldsymbol{\gamma}) = \mathbf{G}^\top \boldsymbol{M}_\gamma.$$

Definamos o vetor $\boldsymbol{M}$ de dimensão $(2n \times 1)$ por

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_\beta \\ \boldsymbol{M}_\gamma \end{bmatrix}, \tag{2.10}$$

e $\boldsymbol{X}$ a matriz de dimensão $2n \times (p+q)$ da forma

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{bmatrix}. \tag{2.11}$$

Então, a função escore pode ser escrita como

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{B}^\top \boldsymbol{M}_\beta \\ \mathbf{G}^\top \boldsymbol{M}_\gamma \end{bmatrix} = \boldsymbol{X}^\top \boldsymbol{M}.$$

Para obter a matriz de informação de Fisher calculamos os momentos das derivadas de segunda ordem da função de log-verossimilhança. A matriz de informação de Fisher, $\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\beta}, \boldsymbol{\gamma})$, pode ser subdividida em

$$\mathbf{K}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\gamma} \\ \mathbf{K}_{\gamma\beta} & \mathbf{K}_{\gamma\gamma} \end{bmatrix},$$

que é simétrica e cujos componentes sâo dados por

$$\mathbf{K}_{\beta\beta} = -E\left[\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right], \quad \mathbf{K}_{\beta\gamma} = -E\left[\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^\top}\right] \quad \text{e} \quad \mathbf{K}_{\gamma\gamma} = -E\left[\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right].$$

Assim, para $j = 1,\ldots,p$, $k = 1,\ldots,p$, $r = 1,\ldots,q$ e $s = 1,\ldots,q$ temos que, por exemplo para o modelo Binomial inflacionado de zeros,

$$\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_j\partial\beta_k} = -\sum_{i=1}^{n}\Big\{\Big[u_i\frac{n_i w_i(1-\pi_i)^{n_i} - n_i^2 w_i\pi_i(1-\pi_i)^{n_i-1} + n_i(1-\pi_i)^{2n_i}}{[w_i+(1-\pi_i)^{n_i}]^2} + n_i(1-u_i)\Big]\pi_i(1-\pi_i)B_{ik}B_{ij}\Big\},$$

$$\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\gamma_r\partial\gamma_s} = \sum_{i=1}^{n}\left\{\left[u_i\frac{(1-\pi_i)^{n_i}}{[w_i+(1-\pi_i)^{n_i}]^2} - \frac{1}{(1+w_i)^2}\right]w_i G_{is}G_{ir}\right\}$$

e

$$\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_j\partial\gamma_s} = \sum_{i=1}^{n}\left\{\left[\frac{u_i w_i n_i\pi_i(1-\pi_i)^{n_i}}{[w_i+(1-\pi_i)^{n_i}]^2}\right]G_{is}B_{ij}\right\}.$$

Utilizando o fato de que

$$E[u_i] = \frac{w_i+(1-\pi_i)^{n_i}}{1+w_i} \quad \text{e} \quad E[1-u_i] = \frac{1-(1-\pi_i)^{n_i}}{1+w_i},$$

tem-se que

$$K_{\beta\beta_{jk}} = \sum_{i=1}^{n}\Big\{p_i B_{ik}B_{ij}\Big\}, \quad j = 1,\ldots,p, \quad k = 1,\ldots,p,$$

em que para $i = 1,\ldots,n$

$$p_i = \frac{n_i w_i\pi_i(1-\pi_i) - n_i^2 w_i\pi_i^2(1-\pi_i)^{n_i} + n_i\pi_i(1-\pi_i)^{n_i+1}}{(1+w_i)[w_i+(1-\pi_i)^{n_i}]},$$

$$K_{\gamma\gamma_{rs}} = \sum_{i=1}^{n}\Big\{q_i G_{is}G_{ir}\Big\}, \quad r = 1,\ldots,q, \quad s = 1,\ldots,q,$$

em que para $i = 1,\ldots,n$

$$q_i = \frac{w_i^2 - w_i^2(1-\pi_i)^{n_i}}{(1+w_i)^2[w_i+(1-\pi_i)^{n_i}]}$$

e

$$K_{\beta\gamma_{js}} = \sum_{i=1}^{n}\Big\{r_i G_{is}B_{ij}\Big\}, \quad j = 1,\ldots,p, \quad s = 1,\ldots,q.$$

em que para $i = 1, \ldots, n$

$$r_i = -\frac{w_i n_i \pi_i (1-\pi_i)^{n_i}}{(1+w_i)[w_i + (1-\pi_i)^{n_i}]}.$$

Definindo as matrizes $\boldsymbol{W}_{\beta\beta} = \text{diag}\{p_1, \ldots, p_n\}$, $\boldsymbol{W}_{\gamma\gamma} = \text{diag}\{q_1, \ldots, q_n\}$ e $\boldsymbol{W}_{\beta\gamma} = \text{diag}\{r_1, \ldots, r_n\}$, os blocos da matriz de informação de Fisher podem ser escritos como $\mathbf{K}_{\beta\beta} = \mathbf{B}^\top \boldsymbol{W}_{\beta\beta}\mathbf{B}$, $\mathbf{K}_{\gamma\gamma} = \mathbf{G}^\top \boldsymbol{W}_{\gamma\gamma}\mathbf{G}$, $\mathbf{K}_{\beta\gamma} = \mathbf{B}^\top \boldsymbol{W}_{\beta\gamma}\mathbf{G}$ e $\mathbf{K}_{\gamma\beta} = \mathbf{K}_{\beta\gamma}^\top$.

Definindo a matriz $\boldsymbol{W}$ de dimensão $2n \times 2n$ por

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{\beta\beta} & \boldsymbol{W}_{\beta\gamma} \\ \boldsymbol{W}_{\gamma\beta} & \boldsymbol{W}_{\gamma\gamma} \end{bmatrix}, \tag{2.12}$$

e sendo $\boldsymbol{X}$ da forma dada em (2.11), temos que a matriz de informação de Fisher de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ pode ser escrita como

$$\mathbf{K}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}.$$

Conseqüentemente, as estimativas de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são obtidas iterativamente através do método do escore de Fisher por

$$\begin{bmatrix} \boldsymbol{\beta}^{(m+1)} \\ \boldsymbol{\gamma}^{(m+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(m)} \\ \boldsymbol{\gamma}^{(m)} \end{bmatrix} + \left[\mathbf{K}(\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)})\right]^{-1} \mathbf{U}\big(\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}\big),$$

em que m é o contador de iterações.

O processo inicia com valores iniciais adequados para os parâmetros usando algum critério de convergência tal como (1.10). Os erros padrões de $\widehat{\boldsymbol{\beta}}$ e de $\widehat{\boldsymbol{\gamma}}$ são obtidos diretamente de $\left[\mathbf{K}\big(\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}\big)\right]^{-1}$ na última iteração.

De acordo com Lambert (1992), para valores iniciais adequados, esses métodos convergem mais rapidamente do que o algoritmo EM, com a vantagem de fornecer estimativas para a matriz de covariâncias dos estimadores dos parâmetros. Entretanto, valores iniciais que estejam próximos da região de convergência nem sempre são simples de serem obtidos para o modelo ZIB, principalmente quando existem muitos parâmetros.

Algoritmo EM

O algoritmo EM, sistematizado por Dempster *et al.* (1977), pode ser implementado usando programas tais como GLIM, SAS ou R, com a vantagem de que os valores iniciais não influenciam diretamente no procedimento da análise. Esse algoritmo gera estimativas de

máxima verossimilhança que são geralmente as mesmas obtidas pelo método de Newton-Raphson. Hall (2000) fornece detalhes da aplicação do algoritmo EM para ajustar os modelos ZIB e ZIP com covariáveis.

Consideremos a variável não observada

$$Z_i = \begin{cases} 1 & , \text{ se } y_i = 0 \\ 0 & , \text{ se } y_i \sim g(.;\mu_i,\phi) \end{cases},$$

em que $g(.;\mu_i,\phi)$ pode ser a distribuição binomial, Poisson, binomial negativa ou beta-binomial. Pelo fato de que $f(y_i,z_i) = f(z_i)f(y_i|z_i) = (p_i)^{z_i}(1-p_i)^{1-z_i}\left(g(y_i;\mu_i,\phi)\right)^{1-z_i}$ e fazendo $g(y_i;\mu_i,\phi) = g(y_i),\ i = 1,\ldots,n$, temos que a função de log-verossimilhança dos dados completos definida em (1.14) é

$$\ell_c(\boldsymbol{\mu},\boldsymbol{p}|\boldsymbol{y},\boldsymbol{z}) = \sum_{i=1}^{n}\Big\{z_i\,\log(p_i) + (1-z_i)\,\log(1-p_i) + (1-z_i)\,\log\left[g(y_i)\right]\Big\}. \tag{2.13}$$

Sendo que a função de ligação para o parâmetro de excesso de zeros $\boldsymbol{p}$ é $\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{G}_i\boldsymbol{\gamma},\ \ i = 1,\ldots,n$, temos que $p_i = \frac{e^{\mathbf{G}_i\boldsymbol{\gamma}}}{1+e^{\mathbf{G}_i\boldsymbol{\gamma}}}$. Então, substituindo em (2.13), a função de log-verossimilhança dos dados completos pode ser expressa por

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{z}) &= \sum_{i=1}^{n}\Bigg\{z_i\log\left(\frac{e^{\mathbf{G}_i\boldsymbol{\gamma}}}{1+e^{\mathbf{G}_i\boldsymbol{\gamma}}}\right) + (1-z_i)\log\left(\frac{1}{1+e^{\mathbf{G}_i\boldsymbol{\gamma}}}\right) \\
&\quad +(1-z_i)\log\left[g(y_i)\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Big\{z_i\log(e^{\mathbf{G}_i\boldsymbol{\gamma}}) - \log(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}) + (1-z_i)\log\left[g(y_i)\right]\Big\} \\
&= \sum_{i=1}^{n}\Big\{z_i\mathbf{G}_i\boldsymbol{\gamma} - \log\left(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\right) + (1-z_i)\log\left[g(y_i)\right]\Big\}.
\end{aligned}$$

sendo $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top,\boldsymbol{\beta}^\top,\phi)^\top$ e onde para os modelos ZIB e ZIP o valor de $\phi = 1$. Portanto, a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{z})|\boldsymbol{y}]$ é dada por

$$\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = Q &= \sum_{i=1}^{n}\Big\{\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\left(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\right) + (1-\hat{z}_i)\log\left[g(y_i)\right]\Big\} \\
&= Q_a(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}) + Q_b(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}}),
\end{aligned} \tag{2.14}$$

sendo

$$\begin{aligned}
Q_a(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n}\Big\{\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\left(1+e^{G_i\boldsymbol{\gamma}}\right)\Big\}, \\
Q_b(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n}\Big\{(1-\hat{z}_i)\log\left[g(y_i)\right]\Big\}
\end{aligned} \tag{2.15}$$

e

$$\hat{z}_i = E\left[z_i | y_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \phi\right].$$

O algoritmo EM pode ser usado para maximizar Q, dado em (2.14), alternando entre o passo E, no qual os dados não observados, $\boldsymbol{z}$, são estimados através da esperança de $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \phi)$ dado $\boldsymbol{y}$ e o passo M, em que a função de log-verossimilhança completa é maximizada com relação a $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ e ϕ com $\boldsymbol{z}$ fixado.

A expressão (2.15) mostra, no entanto, que a estimação de $(\boldsymbol{\beta}, \phi)$ e $\boldsymbol{\gamma}$ pode ser independentemente obtida, maximizando-se os dois termos de Q separadamente.

Para aplicar o algoritmo EM é necessário atribuir valores iniciais $(\boldsymbol{\gamma}^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)})$ e proceder iterativamente até a convergência.

Passos do algoritmo EM

Passo E. Estimar z_i da média condicional $\hat{z}_i^{(t)} = E\left[z_i | y_i, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\beta}^{(t)}, \phi^{(t)}\right]$, $i = 1, \ldots, n$, com as estimativas dos parâmetros de regressão. Essa esperança é determinada através do teorema de Bayes, obtendo-se

$$\begin{aligned} \hat{z}_i^{(t)} &= P(z_i | y_i, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\beta}^{(t)}, \phi^{(t)}) = \frac{P(z_i, y_i)}{P(y_i)} = \frac{P(z_i)P(y_i|z_i)}{P(y_i)} \\ &= \frac{p_i^{z_i}(1-p_i)^{1-z_i}\left[g(y_i)\right]^{1-z_i}}{[p_i + (1-p_i)g(0)]\mathbb{I}_{\{Y_i=0\}} + [(1-p_i)g(y_i)]\mathbb{I}_{\{Y_i>0\}}} \end{aligned}$$

e, então,

$$\hat{z}_i^{(t)} = \begin{cases} \left[1 + e^{-\mathbf{G}_i\boldsymbol{\gamma}}g(0)\right]^{-1} & , \quad \text{se} \quad y_i = 0, \\ 0 & , \quad \text{se} \quad y_i > 0 \end{cases}$$

Passo M para $\boldsymbol{\gamma}$. Deve-se maximizar a esperança condicional da função de log-verossimilhança dos dados completos (Q) para o parâmetro $\boldsymbol{\gamma}$, considerando o vetor estimado $\widehat{\boldsymbol{z}}^{(t)}$ obtido no passo E. Como Q foi dividido em dois termos, pode-se maximizar apenas o termo $Q_a(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})$. Isto pode ser realizado executando uma regressão binomial não ponderada de $\widehat{\boldsymbol{z}}^{(t)}$ na matriz $\mathbf{G}$ utilizando um denominador binomial de um para cada observação.

Passo M para $(\boldsymbol{\beta}, \phi)$. O vetor $(\boldsymbol{\beta}^{(t+1)}, \phi^{(t+1)})$ pode ser obtido maximizando $Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})$, que é o mesmo que maximizar a função de log-verossimilhança ponderada para o modelo discreto considerado (binomial, Poisson, binomial negativa ou beta-binomial) utilizando como pesos $1 - \hat{z}_i^{(t)}$, y_i como variável resposta e uma função de ligação.

Os valores dos $\hat{z}_i$ para cada um dos modelos utilizados são dados por

$$\hat{z}_i = \begin{cases} \left[1 + e^{-\mathbf{G}_i\boldsymbol{\gamma}}(1 + e^{\mathbf{B}_i\boldsymbol{\beta}})^{-n_i}\right]^{-1} & , \quad \text{se} \quad y_i = 0, \\ 0 & , \quad \text{se} \quad y_i > 0, \end{cases}$$

para o modelo ZIB;

$$\hat{z}_i = \begin{cases} \left[1 + e^{-\mathbf{G}_i\boldsymbol{\gamma} - e^{\mathbf{B}_i\boldsymbol{\beta}}}\right]^{-1} & , \quad \text{se} \quad y_i = 0, \\ 0 & , \quad \text{se} \quad y_i > 0, \end{cases}$$

para o modelo ZIP;

$$\hat{z}_i = \begin{cases} \left[1 + e^{-\mathbf{G}_i\boldsymbol{\gamma}} \left(\dfrac{\phi}{e^{\mathbf{B}_i\boldsymbol{\beta}} + \phi}\right)^{\phi}\right]^{-1} & , \quad \text{se} \quad y_i = 0, \\ 0 & , \quad \text{se} \quad y_i > 0, \end{cases}$$

para o modelo ZINB; e

$$\hat{z}_i = \begin{cases} \left[1 + e^{-\mathbf{G}_i\boldsymbol{\gamma}} \dfrac{B\left(\dfrac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i\boldsymbol{\beta}}}, n_i + \dfrac{\phi}{1 + e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)}{B\left(\dfrac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i\boldsymbol{\beta}}}, \dfrac{\phi}{1 + e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)}\right]^{-1} & , \quad \text{se} \quad y_i = 0, \\ 0 & , \quad \text{se} \quad y_i > 0, \end{cases}$$

para o modelo ZIBB, $i = 1, \ldots, n$.

2.5 Teste Escore nos Modelos Inflacionados de Zeros

A presença de muitos zeros não leva necessariamente a modelos inflacionados de zeros. Assim, nós devemos testar a hipótese de inflação de zeros antes que um modelo inflacionado de zeros seja ajustado. O teste da razão de verossimilhanças e o teste de Wald são largamente utilizados por pesquisadores para testar a bondade de ajuste de modelos. Em anos recentes, diferentes pesquisadores propuseram alguns testes baseados no teste qui-quadrado. Um teste escore para dados inflacionados de zeros é dado por Broek (1995) para testar se o número de zeros é demasiado grande para que uma distribuição de Poisson faça um bom ajuste dos dados. Uma das vantagens do teste escore de Broek (1995) é que para calcular a estatística escore não é necessário ajustar a distribuição ZIP, mas somente a distribuição de Poisson. Jansakul e Hinde (2002) estendem o trabalho de Broek

(1995) e propõem um teste para uma situação mais geral onde permite-se que a proporção de zeros nos dados dependa de covariáveis. Este teste escore é útil para determinar se o modelo Poisson é adequado contra a alternativa inflacionada de zeros. Neste trabalho nós desenvolvemos um teste escore para o modelo binomial inflacionado de zeros e para o modelo Poisson inflacionado de zeros. Novamente, não desenvolvemos testes escore para os modelos binomial negativa e beta-binomial inflacionado de zeros pela mesma dificuldade de encontrar a matriz de informação esperada de Fisher.

Notemos que na equação (2.4) é possível fazer $p < 0$ desde que $p \geq -\dfrac{g(0;\boldsymbol{\mu})}{1-g(0;\boldsymbol{\mu})}$ com igualdade para o truncamento à esquerda. Valores de $p > 0$ implicam a existência de muitos zeros (inflacionados de zeros), enquanto que $p < 0$ indicam que há poucos zeros (deflacionados de zeros). Neste trabalho nós estudaremos só o primeiro caso.

Sejam $y_1, \ldots, y_n$ uma amostra de observações independentes da distribuição (2.4). Escrevendo $w = \dfrac{p}{1-p}$ na equação (2.6) temos que

$$\begin{aligned}\ell(\boldsymbol{w}, \boldsymbol{\mu}|\boldsymbol{y}) &= \sum_{i=1}^{n} \ell_i(\boldsymbol{w}, \mu_i|y_i) = \sum_{i=1}^{n}\Big\{u_i \log\big[w + g(0;\mu_i)\big] \\ &\quad - \log(1+w) + (1-u_i)\log\big[g(y_i;\mu_i)\big]\Big\}, \end{aligned} \tag{2.16}$$

em que

$$u_i = \begin{cases} 1 \quad , & \text{se} \quad y_i = 0, \\ 0 \quad , & \text{se} \quad y_i > 0. \end{cases}$$

A estatística escore para testar $p = 0$ é baseada em

$$\psi = \frac{\partial \ell}{\partial w}\bigg|_{w=0} = \sum_{i=1}^{n}\left\{\frac{u_i}{w + g(0;\mu_i)} - \frac{1}{1+w}\right\}\bigg|_{w=0} = \sum_{i=1}^{n}\left\{\frac{u_i}{g(0;\mu_i)} - 1\right\}.$$

Seja

$$\mathbf{K}(\boldsymbol{\beta}, w) = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta w} \\ \mathbf{K}_{\beta w}^{\top} & \mathbf{K}_{ww} \end{bmatrix},$$

sendo que $\mathbf{K}_{\beta\beta}$, $\mathbf{K}_{\beta w}$ e $\mathbf{K}_{ww}$ são matrizes de ordem $p \times p$, $p \times 1$ e 1×1, respectivamente. Note que

$$\frac{\partial \ell_i}{\partial \mu_i} = \frac{u_i}{w + g(0;\mu_i)} \times \frac{\partial g(0;\mu_i)}{\partial \mu_i} + \frac{(1-u_i)}{g(y_i;\mu_i)} \times \frac{\partial g(y_i;\mu_i)}{\partial \mu_i},$$

$$\begin{aligned}\frac{\partial^2 \ell_i}{\partial \mu_i^2} &= \frac{u_i}{[w + g(0;\mu_i)]^2} \times \left(\frac{\partial^2 g(0;\mu_i)}{\partial \mu_i^2}\Big[w + g(0;\mu_i)\Big] - \left[\frac{\partial g(0;\mu_i)}{\partial \mu_i}\right]^2\right) \\ &\quad + \frac{(1-u_i)}{[g(y_i;\mu_i)]^2} \times \left(g(y_i;\mu_i) \times \frac{\partial^2 g(y_i;\mu_i)}{\partial \mu_i^2} - \left[\frac{\partial g(y_i;\mu_i)}{\partial \mu_i}\right]^2\right),\end{aligned}$$

$$\frac{\partial^2 \ell_i}{\partial w \partial \mu_i} = \frac{\partial}{\partial w}\left(\frac{\partial \ell_i}{\partial \mu_i}\right) = -\frac{u_i}{[w + g(0;\mu_i)]^2} \times \frac{\partial g(0;\mu_i)}{\partial \mu_i},$$

$$\frac{\partial \ell_i}{\partial w} = \frac{u_i}{w + g(0;\mu_i)} - \frac{1}{1+w}$$

e

$$\frac{\partial^2 \ell_i}{\partial w^2} = -\frac{u_i}{[w + g(0;\mu_i)]^2} + \frac{1}{(1+w)^2}.$$

Seja $\boldsymbol{P}$ uma matriz $n \times p$ com elemento (i,r) $\dfrac{\partial \mu_i}{\partial \beta_r}$, $\mathbf{1}$ o vetor unitário $n \times 1$ e $\mathbf{W}_1$ e $\mathbf{W}_2$ matrizes diagonais com o i-ésimo elemento diagonal

$$\begin{aligned}
W_{1i} &= E\left[-\frac{\partial^2 \ell_i}{\partial \mu_i^2}\right]\Bigg|_{w=0} \\
&= \left\{\frac{w + g(0;\mu_i)}{(1+w)[w+g(0;\mu_i)]^2}\left(\left[\frac{\partial g(0;\mu_i)}{\partial \mu_i}\right]^2 - \frac{\partial^2 g(0;\mu_i)}{\partial \mu_i^2}\Big[w + g(0;\mu_i)\Big]\right)\right. \\
&\quad \left. + \frac{1 - g(0;\mu_i)}{(1+w)[g(y_i;\mu_i)]^2} \times \left(-g(y_i;\mu_i) \times \frac{\partial^2 g(y_i;\mu_i)}{\partial \mu_i^2} + \left[\frac{\partial g(y_i;\mu_i)}{\partial \mu_i}\right]^2\right)\right\}\Bigg|_{w=0} \\
&= \frac{1}{g(0;\mu_i)}\left(\left[\frac{\partial g(0;\mu_i)}{\partial \mu_i}\right]^2 - g(0;\mu_i) \times \frac{\partial^2 g(0;\mu_i)}{\partial \mu_i^2}\right) \\
&\quad + \frac{1 - g(0;\mu_i)}{[g(y_i;\mu_i)]^2}\left(\left[\frac{\partial g(y_i;\mu_i)}{\partial \mu_i}\right]^2 - g(y_i;\mu_i) \times \frac{\partial^2 g(y_i;\mu_i)}{\partial \mu_i^2}\right)
\end{aligned}$$

e

$$\begin{aligned}
W_{2i} &= E\left[-\frac{\partial^2 \ell_i}{\partial w \partial \mu_i}\right]\Bigg|_{w=0} = \left\{\frac{w + g(0;\mu_i)}{(1+w)[w+g(0;\mu_i)]^2} \times \frac{\partial g(0;\mu_i)}{\partial \mu_i}\right\}\Bigg|_{w=0} \\
&= \frac{1}{g(0;\mu_i)} \times \frac{\partial g(0;\mu_i)}{\partial \mu_i},
\end{aligned}$$

respectivamente.

Então, $\mathbf{K}_{\beta\beta} = \boldsymbol{P}^\top \mathbf{W}_1 \boldsymbol{P}$, $\mathbf{K}_{\beta w} = \boldsymbol{P}^\top \mathbf{W}_2 \mathbf{1}$ e

$$\begin{aligned}
\mathbf{K}_{ww} &= \sum_{i=1}^{n} E\left[-\frac{\partial^2 \ell_i}{\partial w^2}\right]\Bigg|_{w=0} = \sum_{i=1}^{n}\left\{\frac{w + g(0;\mu_i)}{(1+w)[w+g(0;\mu_i)]^2} - \frac{1}{(1+w)^2}\right\}\Bigg|_{w=0} \\
&= \sum_{i=1}^{n}\left\{\frac{1}{g(0;\mu_i)} - 1\right\}
\end{aligned}$$

Daí, a variância assintótica de ψ é

$$V^2 = \mathbf{K}_{ww} - \mathbf{K}_{\beta w}^\top \mathbf{K}_{\beta\beta}^{-1} \mathbf{K}_{\beta w} = \mathbf{K}_{ww} - \mathbf{1}^\top \mathbf{W}_2 \boldsymbol{P}(\boldsymbol{P}^\top \mathbf{W}_1 \boldsymbol{P})^{-1} \boldsymbol{P}^\top \mathbf{W}_2 \mathbf{1}.$$

O teste estatístico escore para testar $H_0 : w = 0$ contra a alternativa $H_1 : w \neq 0$ é $S = \dfrac{\hat{\psi}^2}{\hat{V}^2}$, em que $\hat{\psi} = \psi(\hat{\mu})$, $\hat{V} = V(\hat{\mu})$ e $\hat{\mu}$ é o estimador de máxima verossimilhança de μ quando $w = 0$. Quando $n \to \infty$, a estatística S tem distribuição $\chi^2_{(1)}$. Para uma discussão da distribuição exata, veja Jansakul e Hinde (2002).

2.5.1 Casos Particulares

Teste escore para o modelo ZIB

Usando (2.16) temos que para o caso ZIB: $\mu = \pi$, $g(y_i;\pi_i) = \binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}$, $y_i = 1,\ldots,n_i$ e $g(0;\pi_i) = (1-\pi_i)^{n_i}$. Então,

$$\begin{aligned}\ell(w,\boldsymbol{\pi}|\boldsymbol{y}) &= \sum_{i=1}^{n}\ell_i(w,\pi_i|y_i) = \sum_{i=1}^{n}\Big\{u_i\log\left[w+(1-\pi_i)^{n_i}\right]-\log(1+w)\\ &\quad +(1-u_i)\left[y_i\log\pi_i+(n_i-y_i)\log(1-\pi_i)+\log\binom{n_i}{y_i}\right]. \qquad (2.17)\end{aligned}$$

Para $\log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \eta_i = \mathbf{B}_i\boldsymbol{\beta}$ temos que $\pi_i = \dfrac{e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}$ e $1-\pi_i = \dfrac{1}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}$.

A estatística escore para testar $H_0 : w = 0$ contra a alternativa $H_1 : w \neq 0$ é:

$$S = \mathbf{U}^\top(\widetilde{\boldsymbol{\beta}},0)[\mathbf{K}(\widetilde{\boldsymbol{\beta}},0)]^{-1}\mathbf{U}(\widetilde{\boldsymbol{\beta}},0),$$

em que $\widetilde{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança sob o modelo binomial. Derivando ℓ_i obtemos

$$\begin{aligned}\frac{\partial\ell_i}{\partial\beta_r} &= \frac{\partial\ell_i}{\partial\pi_i}\frac{\partial\pi_i}{\partial\beta_r} = \left\{\frac{-u_in_i(1-\pi_i)^{n_i-1}}{w+(1-\pi_i)^{n_i}}+(1-u_i)\left[\frac{y_i}{\pi_i}-\frac{n_i-y_i}{1-\pi_i}\right]\right\}\frac{e^{B_i\beta}B_{ir}}{(1+e^{B_i\beta})^2}\\ &= \left\{-\frac{u_in_i(1-\pi_i)^{n_i-1}}{w+(1-\pi_i)^{n_i}}+(1-u_i)\left[\frac{y_i-\pi_in_i}{\pi_i(1-\pi_i)}\right]\right\}\pi_i(1-\pi_i)B_{ir}\\ &= \left[-\frac{u_in_i\pi_i(1-\pi_i)^{n_i}}{w+(1-\pi_i)^{n_i}}+(1-u_i)(y_i-\pi_in_i)\right]B_{ir}.\end{aligned}$$

Então,

$$\begin{aligned}\frac{\partial\ell}{\partial\beta_r} &= \sum_{i=1}^{n}\frac{\partial\ell_i}{\partial\beta_r}\\ &= \sum_{i=1}^{n}\left\{\left[-\frac{u_in_i\pi_i(1-\pi_i)^{n_i}}{w+(1-\pi_i)^{n_i}}+(1-u_i)(y_i-\pi_in_i)\right]B_{ir}\right\},\ r=1,\ldots,p, \qquad (2.18)\end{aligned}$$

e

$$\frac{\partial \ell}{\partial w} = \sum_{i=1}^{n} \left\{ \frac{u_i}{w + (1-\pi_i)^{n_i}} - \frac{1}{1+w} \right\}. \tag{2.19}$$

Sob $H_0 : w = 0$, com $\widetilde{\boldsymbol{\beta}}$ e $\widetilde{\pi}_i$ estimadores de máxima verossimilhança sob a hipótese nula, (2.18) passa a ser:

$$\sum_{i=1}^{n} \Big\{ -u_i n_i \widetilde{\pi}_i + (1-u_i)(y_i - \widetilde{\pi}_i n_i) \Big\} B_{ir} = \sum_{i=1}^{n} (y_i - \widetilde{\pi}_i n_i) B_{ir} = 0, \quad r = 1, \ldots, p,$$

e (2.19) fica $\displaystyle\sum_{i=1}^{n} \left\{ \frac{u_i}{(1-\widetilde{\pi}_i)^{n_i}} - 1 \right\}$.

Então $\mathbf{U}^\top(\widetilde{\boldsymbol{\beta}}, 0) = \left(0, \ldots, 0, \displaystyle\sum_{i=1}^{n} \left\{ \frac{u_i}{(1-\widetilde{\pi}_i)^{n_i}} - 1 \right\} \right)$.

As segundas derivadas da função log-verossimilhança dos dados observados definida em (2.17) são dadas por

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \beta_s \partial \beta_r} &= \frac{\partial}{\partial \pi_i} \left\{ \left[-\frac{u_i n_i \pi_i (1-\pi_i)^{n_i}}{w + (1-\pi_i)^{n_i}} + (1-u_i)(y_i - \pi_i n_i) \right] B_{ir} \right\} \frac{\partial \pi_i}{\partial \beta_s} \\ &= \left\{ \frac{-u_i n_i w (1-\pi_i)^{n_i} - u_i n_i (1-\pi_i)^{2n_i} + u_i n_i^2 w \pi_i (1-\pi_i)^{n_i - 1}}{[w + (1-\pi_i)^{n_i}]^2} \right. \\ &\quad \left. -(1-u_i) n_i \right\} \pi_i (1-\pi_i) B_{ir} B_{is}, \; r = 1, \ldots, p, \quad s = 1, \ldots, p, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \beta_r \partial w} &= \frac{\partial}{\partial \pi_i} \left[\frac{u_i}{w + (1-\pi_i)^{n_i}} - \frac{1}{1+w} \right] \frac{\partial \pi_i}{\partial \beta_r} = \frac{u_i n_i (1-\pi_i)^{n_i - 1}}{[w + (1-\pi_i)^{n_i}]^2} \cdot \pi_i (1-\pi_i) B_{ir} \\ &= \frac{u_i n_i \pi_i (1-\pi_i)^{n_i}}{[w + (1-\pi_i)^{n_i}]^2} B_{ir}, \quad r = 1, \ldots, p \end{aligned}$$

e

$$\frac{\partial^2 \ell_i}{\partial w^2} = \frac{\partial}{\partial w} \left[\frac{u_i}{w + (1-\pi_i)^{n_i}} - \frac{1}{1+w} \right] = -\frac{u_i}{[w + (1-\pi_i)^{n_i}]^2} + \frac{1}{(1+w)^2}.$$

Usando o fato de que $E[u_i] = P[Y_i = 0] = \dfrac{w + (1-\pi_i)^{n_i}}{1+w}$ e $E[1-u_i] = P[Y_i > 0] = \dfrac{1 - (1-\pi_i)^{n_i}}{1+w}$, temos que a matriz de informação esperada, $\mathbf{K}(\boldsymbol{\beta}, w)$, tem entradas

$$\begin{aligned} K_{r,s} &= -E\left[\frac{\partial^2 \ell}{\partial \beta_s \partial \beta_r} \right] \\ &= \sum_{i=1}^{n} \left\{ \frac{n_i w \pi_i (1-\pi_i) + n_i \pi_i (1-\pi_i)^{n_i + 1} - n_i^2 w \pi_i^2 (1-\pi_i)^{n_i}}{[w + (1-\pi_i)^{n_i}](1+w)} \cdot B_{ir} B_{is} \right\}, \\ &\quad r = 1, \ldots, p \quad \text{e} \quad s = 1, \ldots, p, \end{aligned}$$

$$K_{r,p+1} = -E\left[\frac{\partial^2 \ell}{\partial \beta_r \partial w}\right] = \sum_{i=1}^{n}\left\{-\frac{n_i \pi_i (1-\pi_i)^{n_i}}{[w+(1-\pi_i)^{n_i}](1+w)} \cdot B_{ir}\right\}, \quad r = 1, \ldots, p$$

e

$$K_{p+1,p+1} = -E\left[\frac{\partial^2 \ell}{\partial w^2}\right] = \sum_{i=1}^{n}\left\{\frac{1-(1-\pi_i)^{n_i}}{[w+(1-\pi_i)^{n_i}](1+w)^2}\right\}.$$

Sob $H_0 : w = 0$, $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ tem entradas

$$\widetilde{K}_{r,s} = \sum_{i=1}^{n} n_i \widetilde{\pi}_i (1-\widetilde{\pi}_i) B_{ir} B_{is}, \quad r = 1, \ldots, p \quad \text{e} \quad s = 1, \ldots, p,$$

$$\widetilde{K}_{r,p+1} = -\sum_{i=1}^{n} n_i \widetilde{\pi}_i B_{ir}, \quad r = 1, \ldots, p$$

e

$$\widetilde{K}_{p+1,p+1} = \sum_{i=1}^{n}\left\{\frac{1}{(1-\widetilde{\pi}_i)^{n_i}} - 1\right\}.$$

Escrevamos $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ como $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0) = \begin{bmatrix} \widetilde{\mathbf{K}}_{11} & \widetilde{\mathbf{K}}_{12} \\ \widetilde{\mathbf{K}}_{21} & \widetilde{\mathbf{K}}_{22} \end{bmatrix}$, em que

$$\begin{aligned}
\widetilde{\mathbf{K}}_{11} &= \mathbf{B}^\top \operatorname{diag}\{n_i \widetilde{\pi}_i (1-\widetilde{\pi}_i)\}\, \mathbf{B}, \\
\widetilde{\mathbf{K}}_{12} &= -\mathbf{B}^\top \cdot (n_1 \widetilde{\pi}_1, \ldots, n_n \widetilde{\pi}_n)^\top, \\
\widetilde{\mathbf{K}}_{21} &= \widetilde{\mathbf{K}}_{12}^\top \quad \text{e} \\
\widetilde{\mathbf{K}}_{22} &= \sum_{i=1}^{n}\left\{\frac{1}{(1-\widetilde{\pi}_i)^{n_i}} - 1\right\}.
\end{aligned}$$

Denotemos a inversa de $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ como sendo $\mathbf{C}$ que pode ser particionada como

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}.$$

Devido à estrutura de $\mathbf{U}(\widetilde{\boldsymbol{\beta}}, 0)$ somente é necessário $\mathbf{C}_{22}$.

$$\begin{aligned}
\mathbf{C}_{22}^{-1} &= \widetilde{\mathbf{K}}_{22} - \widetilde{\mathbf{K}}_{21} \widetilde{\mathbf{K}}_{11}^{-1} \widetilde{\mathbf{K}}_{12} \\
&= \sum_{i=1}^{n}\left\{\frac{1}{(1-\widetilde{\pi}_i)^{n_i}} - 1\right\} - \widetilde{\boldsymbol{\mu}}^\top \mathbf{B}\left[\mathbf{B}^\top \operatorname{diag}\{n_i \widetilde{\pi}_i (1-\widetilde{\pi}_i)\} \mathbf{B}\right]^{-1} \mathbf{B}^\top \widetilde{\boldsymbol{\mu}},
\end{aligned}$$

em que $\widetilde{\boldsymbol{\mu}} = (n_1 \widetilde{\pi}_1, \ldots, n_n \widetilde{\pi}_n)^\top$.

Assim, temos que

$$S = \frac{\left[\sum_{i=1}^{n}\left(\frac{u_i}{(1-\widetilde{\pi}_i)^{n_i}} - 1\right)\right]^2}{\sum_{i=1}^{n}\left\{\frac{1}{(1-\widetilde{\pi}_i)^{n_i}} - 1\right\} - \widetilde{\boldsymbol{\mu}}^\top \mathbf{B}\left[\mathbf{B}^\top \operatorname{diag}\{n_i\widetilde{\pi}_i(1-\widetilde{\pi}_i)\}\mathbf{B}\right]^{-1}\mathbf{B}^\top\widetilde{\boldsymbol{\mu}}}. \tag{2.20}$$

Sob a hipótese nula esta estatística tem uma distribuição qui-quadrado com 1 grau de liberdade.

Teste escore para o modelo ZIP

Um teste escore para $H_0 : w = 0$ contra $H_1 : w \neq 0$ no modelo Poisson inflacionado de zeros tem a vantagem que não precisamos ajustar o modelo ZIP se não somente o modelo Poisson, que é a distribuição sob a hipótese nula.

Usando (2.16) temos que para o caso ZIP: $\theta = \lambda$, $g(y_i;\lambda_i) = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$, $y_i = 1, 2, \ldots$ e $g(0;\lambda_i) = e^{-\lambda_i}$. Logo,

$$\begin{aligned}\ell(w,\boldsymbol{\lambda};\boldsymbol{y}) &= \sum_{i=1}^{n}\ell_i(w,\lambda_i;y_i) = \sum_{i=1}^{n}\Big\{u_i\log\left[w + e^{-\lambda_i}\right] - \log(1+w)\\ &\quad +(1-u_i)\left[y_i\log\lambda_i - \lambda_i - \log(y_i!)\right].\end{aligned}$$

Para $\log\lambda_i = \eta_i = \mathbf{B}_i\boldsymbol{\beta}$ temos que $\lambda_i = e^{B_i\beta}$. A estatística escore para testar $H_0 : w = 0$ é

$$S = \mathbf{U}^\top(\tilde{\boldsymbol{\beta}}, 0)[\mathbf{K}(\tilde{\boldsymbol{\beta}}, 0)]^{-1}\mathbf{U}(\tilde{\boldsymbol{\beta}}, 0), \tag{2.21}$$

em que $\tilde{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança sob o modelo Poisson. Derivando ℓ_i obtemos

$$\frac{\partial\ell_i}{\partial\beta_r} = \frac{\partial\ell_i}{\partial\lambda_i}\frac{\partial\lambda_i}{\partial\beta_r} = \left[\frac{-u_ie^{-\lambda_i}}{w+e^{-\lambda_i}} + (1-u_i)\left(\frac{y_i}{\lambda_i} - 1\right)\right]\lambda_iB_{ir}.$$

Portanto,

$$\frac{\partial\ell}{\partial\beta_r} = \sum_{i=1}^{n}\frac{\partial\ell_i}{\partial\beta_r} = \sum_{i=1}^{n}\left\{-\frac{u_ie^{-\lambda_i}}{w+e^{-\lambda_i}}\lambda_iB_{ir} + (1-u_i)(y_i-\lambda_i)B_{ir}\right\},\ r = 1,\ldots,p, \tag{2.22}$$

e

$$\frac{\partial\ell}{\partial w} = \sum_{i=1}^{n}\left\{\frac{u_i}{w+e^{-\lambda_i}} - \frac{1}{1+w}\right\}. \tag{2.23}$$

Sob $H_0 : w = 0$, com $\widetilde{\boldsymbol{\beta}}$ e $\widetilde{\lambda}_i$ estimadores de máxima verossimilhança sob a hipótese nula, (2.22) fica

$$\sum_{i=1}^{n}\Big\{-u_i\widetilde{\lambda}_i + (1-u_i)(y_i - \widetilde{\lambda}_i)\Big\}B_{ir} = \sum_{i=1}^{n}(y_i - \widetilde{\lambda}_i)B_{ir} = 0, \quad r = 1, \ldots, p,$$

e (2.23) fica $\displaystyle\sum_{i=1}^{n}\left\{\frac{u_i}{e^{-\widetilde{\lambda}_i}} - 1\right\}$.

Então, $\mathbf{U}^\top(\widetilde{\boldsymbol{\beta}}, 0) = \left(0, \ldots, 0, \displaystyle\sum_{i=1}^{n}\left\{\frac{u_i}{e^{-\widetilde{\lambda}_i}} - 1\right\}\right)$.

As segundas derivadas são dadas por

$$\begin{aligned}\frac{\partial^2 \ell_i}{\partial\beta_s\partial\beta_r} &= \frac{\partial}{\partial\lambda_i}\left\{\left[-\frac{u_i e^{-\lambda_i}}{w + e^{-\lambda_i}}\lambda_i + (1-u_i)(y_i - \lambda_i)\right]B_{ir}\right\}\frac{\partial\lambda_i}{\partial\beta_s} \\ &= \left[\frac{u_i\lambda_i w e^{-\lambda_i} - u_i w e^{-\lambda_i} - u_i e^{-2\lambda_i}}{(w + e^{-\lambda_i})^2} - (1-u_i)\right]\lambda_i B_{ir}B_{is}, \\ &\quad r = 1, \ldots, p, \quad s = 1, \ldots, p,\end{aligned}$$

$$\frac{\partial^2 \ell_i}{\partial\beta_r\partial w} = \frac{\partial}{\partial\lambda_i}\left[\frac{u_i}{w + e^{-\lambda_i}} - \frac{1}{1+w}\right]\frac{\partial\lambda_i}{\partial\beta_r} = \left[\frac{u_i e^{-\lambda_i}}{(w + e^{-\lambda_i})^2}\right]\lambda_i B_{ir}, \quad r = 1, \ldots, p,$$

e

$$\frac{\partial^2 \ell_i}{\partial w^2} = \frac{\partial}{\partial w}\left[\frac{u_i}{w + e^{-\lambda_i}} - \frac{1}{1+w}\right] = -\frac{u_i}{(w + e^{-\lambda_i})^2} + \frac{1}{(1+w)^2}.$$

Usando o fato de que $E[u_i] = P[Y_i = 0] = \dfrac{w + e^{-\lambda_i}}{1+w}$ e $E[1-u_i] = P[Y_i > 0] = \dfrac{1 - e^{-\lambda_i}}{1+w}$, temos que a matriz de informação esperada, $\mathbf{K}(\boldsymbol{\beta}, w)$, tem entradas

$$\begin{aligned}K_{r,s} &= -E\left[\frac{\partial^2\ell}{\partial\beta_s\partial\beta_r}\right] = \sum_{i=1}^{n}\left\{\frac{-\lambda_i w e^{-\lambda_i} + w + e^{-\lambda_i}}{(w + e^{-\lambda_i})(1+w)}\cdot\lambda_i B_{ir}B_{is}\right\}, \\ &\quad r = 1, \ldots, p \quad \text{e} \quad s = 1, \ldots, p,\end{aligned}$$

$$K_{r,p+1} = -E\left[\frac{\partial^2\ell}{\partial\beta_r\partial w}\right] = \sum_{i=1}^{n}\left\{-\frac{e^{-\lambda_i}\lambda_i B_{ir}}{(w + e^{-\lambda_i})(1+w)}\right\}, \quad r = 1, \ldots, p,$$

e

$$K_{p+1,p+1} = -E\left[\frac{\partial^2\ell}{\partial w^2}\right] = \sum_{i=1}^{n}\left\{\frac{1 - e^{-\lambda_i}}{(w + e^{-\lambda_i})(1+w)^2}\right\}.$$

Sob $H_0 : w = 0$, $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ tem entradas

$$\widetilde{K}_{r,s} = \sum_{i=1}^{n}\widetilde{\lambda}_i B_{ir}B_{is}, \quad r = 1, \ldots, p \quad \text{e} \quad s = 1, \ldots, p,$$

$$\widetilde{K}_{r,p+1} = -\sum_{i=1}^{n} \widetilde{\lambda}_i B_{ir}, \quad r = 1, \ldots, p$$

e

$$\widetilde{K}_{p+1,p+1} = \sum_{i=1}^{n} \left\{ \frac{1}{e^{-\widetilde{\lambda}_i}} - 1 \right\}.$$

Escrevamos $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ como $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0) = \begin{bmatrix} \widetilde{\mathbf{K}}_{11} & \widetilde{\mathbf{K}}_{12} \\ \widetilde{\mathbf{K}}_{21} & \widetilde{\mathbf{K}}_{22} \end{bmatrix}$ em que

$$\begin{aligned} \widetilde{\mathbf{K}}_{11} &= \mathbf{B}^\top \operatorname{diag}\left\{\widetilde{\lambda}_i\right\} \mathbf{B}, \\ \widetilde{\mathbf{K}}_{12} &= -\mathbf{B}^\top \widetilde{\lambda}, \\ \widetilde{\mathbf{K}}_{21} &= -\widetilde{\lambda}^\top \mathbf{B} \quad \text{e} \\ \widetilde{\mathbf{K}}_{22} &= \sum_{i=1}^{n} \left\{ \frac{1}{e^{-\widetilde{\lambda}_i}} - 1 \right\}. \end{aligned}$$

Denotemos a inversa de $\mathbf{K}(\widetilde{\boldsymbol{\beta}}, 0)$ como sendo $\mathbf{C}$, que pode ser dividida como

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}.$$

Devido à estrutura de $\mathbf{U}(\widetilde{\boldsymbol{\beta}}, 0)$ somente é necessário usar $\mathbf{C}_{22}$, dada por

$$\begin{aligned} \mathbf{C}_{22}^{-1} &= \widetilde{\mathbf{K}}_{22} - \widetilde{\mathbf{K}}_{21} \widetilde{\mathbf{K}}_{11}^{-1} \widetilde{\mathbf{K}}_{12} \\ &= \sum_{i=1}^{n} \left\{ \frac{1}{e^{-\widetilde{\boldsymbol{\lambda}}_i}} - 1 \right\} - \widetilde{\boldsymbol{\lambda}}^\top \mathbf{B} \left[\mathbf{B}^\top \operatorname{diag}\{\widetilde{\boldsymbol{\lambda}}\} \mathbf{B} \right]^{-1} \mathbf{B}^\top \widetilde{\boldsymbol{\lambda}}. \end{aligned}$$

Substituindo em (2.21) temos que

$$S = \frac{\left[\sum_{i=1}^{n} \left(\frac{u_i}{e^{-\widetilde{\lambda}_i}} - 1 \right) \right]^2}{\sum_{i=1}^{n} \left\{ \frac{1}{e^{-\widetilde{\lambda}_i}} - 1 \right\} - \widetilde{\boldsymbol{\lambda}}^\top \mathbf{B} \left[\mathbf{B}^\top \operatorname{diag}\{\widetilde{\boldsymbol{\lambda}}\} \mathbf{B} \right]^{-1} \mathbf{B}^\top \widetilde{\boldsymbol{\lambda}}}. \tag{2.24}$$

Sob a hipótese nula esta estatística, S, tem uma distribuição qui-quadrado com 1 grau de liberdade.

Estudo de Simulação sobre o Poder do Teste Escore

Um pequeno estudo de simulação foi conduzido para examinar o nível de significância e o poder empírico da estatística escore para testar inflação de zeros. Duas situações foram

estudadas: uma para testar inflação de zeros no modelo $bin(10, \pi)$ para valores de π dados na Tabela 2.1, e a outra para testar inflação de zeros no modelo $P(\lambda)$ para os valores de λ dados na Tabela 2.2.

Tabela 2.1: Poder do teste escore S quando os dados são simulados do modelo $ZIB(p, \pi, 10)$, $\alpha = 0,05$, baseado em 10 000 réplicas. Os valores nas celas são as taxas de rejeição percentual usando a $\chi^2_{(1)}$ (sem parênteses) e a mistura de qui-quadrados $(\frac{1}{2}\chi^2_{(0)} + \frac{1}{2}\chi^2_{(1)})$ entre parênteses

		p			
n	π	0	0.1	0.2	0.3
50		5.41 (10.52)	8.71 (15.55)	17.77 (27.18)	31.21 (42.77)
100	0.1	5.38 (10.38)	14.22 (22.3)	36.00 (48.35)	58.50 (69.5)
200		5.70 (10.35)	27.3 (38.95)	65.25 (75.5)	89.25 (93.75)
50		4.98 (10.47)	29.79 (40.16)	70.12 (78.91)	91.01 (94.9)
100	0.2	4.84 (10.22)	54.67 (65.79)	94.69 (97.18)	99.84 (99.93)
200		6.03 (11.22)	84.85 (90.92)	99.9 (99.98)	100.0 (100.0)
50		4.11 (9.42)	49.93 (60.23)	90.19 (93.9)	98.89 (99.49)
100	0.25	5.17 (10.16)	81.22 (87.56)	99.46 (99.81)	99.99 (100.0)
200		5.55 (10.75)	97.88 (98.89)	100.0 (100.0)	100.0 (100.0)
50		3.75 (6.14)	72.70 (79.10)	97.90 (98.79)	99.83 (99.91)
100	0.3	3.81 (9.01)	94.80 (97.01)	99.98 (100.0)	100.0 (100.0)
200		4.41 (9.26)	99.86 (99.96)	100.0 (100.0)	100.0 (100.0)
50		4.49 (6.63)	86.96 (90.53)	99.38 (99.60)	99.95 (99.95)
100	0.35	3.9 (6.48)	98.95 (99.26)	100.0 (100.0)	100.0 (100.0)
200		4.08 (9.07)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
50		5.04 (7.13)	94.02 (95.40)	99.76 (99.83)	99.96 (99.98)
100	0.4	4.86 (6.97)	99.79 (99.87)	100.0 (100.0)	100.0 (100.0)
200		4.24 (6.80)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)

As covariáveis consideradas, $B1$ e $B2$, foram tomadas do pacote Zicounts em R e correspondem às variáveis *gender* e *age*, respectivamente. Amostras de tamanho $n = 50, 100, 200$ foram tomadas ou de uma distribuição binomial inflacionada de zeros, $ZIB(p, \pi, 10)$, ou de uma distribuição de Poisson inflacionada de zeros, $ZIP(p, \lambda)$ para o parâmetro de inflação $p = 0$ (para nível de significância), $0.1, 0.2, 0.3$. Cada exper-

imento para nível de significância ou para o poder do teste foi baseado em 10 000 réplicas. Seguindo a sugestão dada por Broek (1995) de que quando a contagem média é elevada e há poucos zeros a aproximação χ^2 pode ser pobre se n não é grande, usamos λ com valores 2, 2.5 e 3 para assegurar que alguns zeros sejam gerados. Segundo Jansakul e Hinde (2002), para um modelo constante de p, a distribuição apropriada é uma mistura igual de uma χ^2_0 (uma constante em zero) e uma distribuição χ^2_1, com os níveis descritivos dados por $\frac{1}{2}(P[\chi^2_1 \geq S])$. Nas tabelas 2.1 e 2.2 apresentamos, entre parêntesis, os valores dessas probabilidades.

Tabela 2.2: Poder do teste escore S quando os dados são simulados do modelo $ZIP(p, \lambda)$, $\alpha = 0,05$, baseado em 10 000 réplicas. Os valores nas celas são as taxas de rejeição percentual usando a $\chi^2_{(1)}$ (sem parênteses) e a mistura de qui-quadrados $(\frac{1}{2}\chi^2_{(0)} + \frac{1}{2}\chi^2_{(1)})$ entre parênteses

		p			
n	λ	0	0.1	0.2	0.3
50		4.85 (10.05)	22.67 (32.28)	58.51 (69.39)	85.63 (91.25)
100	2	4.86 (9.91)	42.77 (54.58)	89.07 (93.70)	99.23 (99.66)
200		5.16 (10.18)	72.18 (81.24)	99.54 (99.84)	100.0 (100.0)
50		4.38 (9.52)	39.33 (50.08)	82.22 (88.11)	96.92 (98.35)
100	2.5	4.95 (9.94)	66.66 (76.26)	98.38 (99.28)	100.0 (100.0)
200		5.05 (10.01)	92.29 (95.63)	99.58 (99.92)	100.0 (100.0)
50		3.54 (8.41)	57.49 (66.98)	93.52 (96.06)	99.48 (99.76)
100	3	4.80 (10.33)	84.68 (89.81)	99.85 (99.92)	100.0 (100.0)
200		5.33 (10.24)	98.81 (99.31)	100.0 (100.0)	100.0 (100.0)

Os resultados nas tabelas 2.1 e 2.2 indicam que os testes escore mantêm um nível nominal razoável, mas em alguns casos estes mostram algum comportamento conservativo. O poder dos testes para detectar inflação de zeros aumenta lentamente para λ e π pequenos. Em particular, para n pequeno (n=50) e π muito pequeno (π=0.1), o teste escore binomial mostra poder baixo para detectar inflação de zeros. Para valores grandes de λ (exemplo $\lambda = 3$) e não tão grandes de π (exemplo $\pi = 0.25$) o poder aumenta muito rápido e aproxima-se a 1.0 para $p = 0.2$. Note-se que em termos de poder, a mistura de chi-quadrados (Jansakul e Hinde, 2002) parece ser a melhor alternativa a ser considerada para a estatística escore.

Capítulo 3

Análise de Diagnóstico

Uma etapa importante na análise de um modelo ajustado consiste na verificação de possíveis afastamentos das suposições assumidas. Nesta direção é importante detectar a presença de pontos extremos no conjunto de dados e avaliar seu impacto nos resultados inferenciais. Uma análise de resíduos e, principalmente, os gráficos de resíduos podem ajudar na avaliação da estabilidade e robustez de resultados inferenciais, visto que os resíduos medem a discrepância entre o modelo ajustado e o conjunto de dados. Assim, é conveniente procurar uma definição para o resíduo que leve em consideração a contribuição que cada observação exerce sobre a adequação do modelo. Desta forma, a análise de resíduos num modelo estatístico pode ser baseada nos resíduos ordinários e suas versões padronizadas, resíduos construídos a partir dos componentes da função desvio ou em resíduos generalizados (Cox & Snell, 1968). Técnicas gráficas usando os resíduos do modelo ajustado são freqüentemente adotados na validação do modelo estatístico; por exemplo, Atkinson (1985) sugere que a construção de bandas de confiança (envelopes) simuladas podem ajudar a interpretar melhor o gráfico de probabilidade normal dos resíduos, de tal forma que, se o modelo estiver bem ajustado, a maioria dos pontos que representam os resíduos devem estar distribuídos aleatoriamente dentro dessas bandas. Outro aspecto importante na análise de diagnóstico é a detecção de observações influentes, ou seja, pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo.

3.1 Resíduos em Modelos de Regressão Inflacionados de Zeros

O resíduo pode ser definido como uma medida que objetiva identificar discrepâncias entre o modelo ajustado e os dados. Assim, é compreensível que a maioria dos resíduos esteja baseada na diferença $y_i - \widehat{E(y_i)}$. No entanto, respeitando o formato da distribuição de probabilidade da variável resposta, é mais interessante pensar no resíduo como uma função $r(y_i, \widehat{E(y_i)})$, definição geral de resíduos proposta por Cox & Snell (1968). Sob este ponto de vista propomos resíduos para os modelos de regressão binomial e Poisson inflacionados de zeros.

Resíduos Padronizados

Usando a definição geral de resíduos dada por Cox & Snell (1968), podemos propor resíduos padronizados para os modelos de regressão Binomial e Poisson inflacionado de zeros. Neste sentido, temos que o algoritmo escore de Fisher é expresso por

$$\begin{aligned}\boldsymbol{\theta}^{(m+1)} &= \boldsymbol{\theta}^{(m)} + (\boldsymbol{X}^\top \boldsymbol{W}^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}^{(m)} \\ &= (\boldsymbol{X}^\top \boldsymbol{W}^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(m)} \boldsymbol{z}^{(m)},\end{aligned}$$

em que $\boldsymbol{z}^{(m)} = \boldsymbol{X}\boldsymbol{\theta}^{(m)} + (\boldsymbol{W}^{(m)})^{-1}\boldsymbol{M}^{(m)}$ é um vetor da ordem $2n \times 1$. Após a convergência do processo iterativo do escore de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ temos

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{W}} \widehat{\boldsymbol{z}}, \tag{3.1}$$

sendo $\widehat{\boldsymbol{z}} = \widehat{\boldsymbol{\eta}} + \widehat{\boldsymbol{W}}^{-1}\widehat{\boldsymbol{M}}$ em que $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\theta}}$. Se definimos $\boldsymbol{z}^* = \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{z}}$ e $\boldsymbol{X}^* = \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}$, temos que

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^{*\top} \boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*\top} \boldsymbol{z}^*, \tag{3.2}$$

ou seja, $\widehat{\boldsymbol{\theta}}$ é a solução de mínimos quadrados de uma regressão linear de $\boldsymbol{z}^*$ contra as colunas de $\boldsymbol{X}^*$. Da equação (3.2) temos que

$$\begin{aligned}\widehat{\boldsymbol{z}^*} &= \boldsymbol{X}^*\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}(\boldsymbol{X}^\top\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{z}} \\ &= \widehat{\boldsymbol{H}}\boldsymbol{z}^*,\end{aligned}$$

em que $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}(\boldsymbol{X}^\top\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{W}}^{1/2}$ é uma matriz de projeção. Assim, de (3.1) os resíduos ordinários podem ser expressos por

$$\boldsymbol{e} = \boldsymbol{z}^* - \widehat{\boldsymbol{z}}^* = (\mathbb{I}_{2n} - \widehat{H})\boldsymbol{z}^*,$$

em que $\mathbb{I}_{2n}$ é a matriz identidade de dimensão $2n \times 2n$. Alternativamente, podemos escrever

$$\begin{aligned} \boldsymbol{e} = (\mathbb{I}_{2n} - \widehat{H})\boldsymbol{z}^* &= \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}(\boldsymbol{X}^\top\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{z}} \\ &= \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{W}}^{1/2}(\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{\eta}}) \\ &= \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{W}}^{-1}\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{W}}^{-1/2}\widehat{\boldsymbol{M}}. \end{aligned} \tag{3.3}$$

A matriz de covariâncias assintótica dos resíduos $\boldsymbol{e}$, com as quantidades avaliadas nos parâmetros verdadeiros é:

$$\begin{aligned} V(\boldsymbol{e}) &= V\left[(\mathbb{I}_{2n} - \boldsymbol{H})\boldsymbol{z}^*\right] \\ &= (\mathbb{I}_{2n} - \boldsymbol{H})V\left[\boldsymbol{z}^*\right](\mathbb{I}_{2n} - \boldsymbol{H})^\top \\ &= (\mathbb{I}_{2n} - \boldsymbol{H})V\left[\boldsymbol{W}^{1/2}\boldsymbol{z}\right](\mathbb{I}_{2n} - \boldsymbol{H}) \\ &= (\mathbb{I}_{2n} - \boldsymbol{H})\boldsymbol{W}^{1/2}Var\left[\boldsymbol{z}\right]\boldsymbol{W}^{1/2}(\mathbb{I}_{2n} - \boldsymbol{H}) \\ &= (\mathbb{I}_{2n} - \boldsymbol{H})\boldsymbol{W}^{1/2}\boldsymbol{W}^{-1}\boldsymbol{W}^{1/2}(\mathbb{I}_{2n} - \boldsymbol{H}) \\ &= (\mathbb{I}_{2n} - \boldsymbol{H}). \end{aligned}$$

Da convergência do processo iterativo do método escore de Fisher para $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$, obtemos a expressão (3.3). Então, os resíduos ordinários podem ser expressos como

$$\boldsymbol{r} = \widehat{\boldsymbol{W}}^{-1/2}\widehat{\boldsymbol{M}}.$$

Ou seja,

$$r_i = \frac{\widehat{M}_i}{\sqrt{\widehat{W}_{ii}}}.$$

Desta forma, podemos definir os resíduos padronizados para o modelo de regressão inflacionado de zeros com as quantidades avaliadas nos estimadores de máxima verossimilhança por

$$r_{p_i} = \frac{r_i}{\sqrt{1 - \widehat{h}_{ii}}} = \frac{\widehat{M}_i}{\sqrt{\widehat{W}_{ii}(1 - \widehat{h}_{ii})}}, \tag{3.4}$$

para $i = 1, \ldots, 2n$; $\widehat{M}_i$ é o i-ésimo elemento do vetor $\boldsymbol{M}$ definido em (2.10), $\widehat{W}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz definida em (2.12) e $\widehat{h}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz de projeção

$$\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}(\boldsymbol{X}^\top\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{W}}^{1/2}$$

em que as matrizes $\boldsymbol{X}$ e $\boldsymbol{W}$ foram definidos em (2.11) e (2.12), respectivamente.

Por construção, o vetor de resíduos padronizados tem duas partes: a primeira delas (os primeiros n resíduos) corresponde aos resíduos relacionados aos parâmetros da parte binomial ou Poisson ($r_{p_i}^{(1)}$) e a segunda (os últimos n resíduos) correspondem à parte de excesso de zeros ($r_{p_i}^{(2)}$). Adicionalmente, definimos o resíduo ponderado

$$r_{p_i}^{(*)} = \widehat{p}_i r_{p_i}^{(2)} + (1 - \widehat{p}_i) r_{p_i}^{(1)}, \tag{3.5}$$

em que $r_{p_i}^{(*)}$ é a média ponderada entre o resíduo $r_{p_i}^{(1)}$ e o resíduo $r_{p_i}^{(2)}$. Para mais detalhes pode-se consultar Ospina (2008).

Neste trabalho, nós apresentamos os resíduos padronizados e ponderados para os modelos binomial inflacionado de zeros e Poisson inflacionado de zeros. Nós tivemos a dificuldade de construir os resíduos para os modelos binomial negativa e beta-binomial inflacionado de zeros pela dificuldade de encontrar a matriz de informação esperada de Fisher, que precisa de simulação de Monte Carlo para uma parte da matriz de informação. Um estudo mais exaustivo fica como trabalho de pesquisa futura. Como uma alternativa, usaremos os resíduos de Pearson, que como veremos mostram-se bastante úteis em nossas aplicações.

3.2 Influência Local

A influência local é uma metodologia de diagnóstico que permite avaliar mudanças nos resultados da análise inferencial quando pequenas perturbações são impostas ao modelo e/ou aos dados em estudo. Se essas perturbações causarem efeitos desproporcionais, pode haver indício de que o modelo está mal ajustado ou que possa existir algum tipo de afastamento nas suposições feitas para o mesmo.

Exclusão de casos e diagnóstico de influência local têm sido amplamente aplicados a muitos modelos de regressão. Por exemplo, Christensen *et al.* (1992) discutem medidas de exclusão de casos para modelos mistos, Tsai e Wu (1992) apresentam influência local em modelos autoregressivos de primeira ordem, Ferrari e Cribari-Neto (2004) enfocam medidas de exclusão de casos para modelos de regressão Beta, Xiang *et al.* (2005) derivam medidas para influência local em modelos de regressão de misturas de Poisson com aplicações em saúde pública, Lee *et al.* (2004) estudam a sensitividade do teste escore para modelos de regressão inflacionados de zeros baseados no enfoque de influência local e recentemente Xie *et al.* (2008) desenvolvem diagnóstico de influência local para modelos de regressão Poisson generalizados inflacionados de zeros com efeitos mistos.

3.2.1 Influência Local nos Modelos de Regressão Inflacionados de Zeros

Nesta seção aplicamos a metodologia de Zhu e Lee (2001) para derivar medidas de influência local sob o esquema de perturbação de ponderação de casos para os modelos considerados. As primeiras e segundas derivadas da função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ (apresentada na seção 1.5) são essenciais para a implementação computacional da metodologia proposta.

Perturbação da Ponderação de casos

A ponderação de casos tem sido o esquema de perturbação mais amplamente difundido na análise de diagnóstico, já que pode ser interpretado como a perturbação na variância do i-ésimo caso. Seja $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ um vetor $n \times 1$ de ponderações. Sendo $\omega_i = 0$, temos que o i-ésimo ponto é eliminado e $\boldsymbol{\omega}_0 = (1, 1, \ldots, 1)^\top$ implica que todos os pontos são considerados.

Utilizando (2.14) e (2.15) temos que

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = Q_a(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}) + Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}}), \tag{3.6}$$

em que

$$Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \hat{z}_i \mathbf{G}_i \boldsymbol{\gamma} - \log\left(1 + e^{\mathbf{G}_i\boldsymbol{\gamma}}\right) + (1 - \hat{z}_i)\log[g(y_i)], \quad i = 1, \ldots, n.$$

Neste caso, a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ se decompõe em duas parcelas: uma que depende apenas de $\boldsymbol{\gamma}$ e outra que depende apenas de $(\boldsymbol{\beta}, \phi)$.

A função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ do modelo perturbado, considerando ponderação de casos, é dada por

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}) = \sum_{i=1}^{n} \left\{\omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})\right\} = \sum_{i=1}^{n} \left\{\omega_i Q_{ai}(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})\right\} + \sum_{i=1}^{n} \left\{\omega_i Q_{bi}(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})\right\}, \tag{3.7}$$

em que $0 \leq \omega_i \leq 1$, $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top)^\top$ nos modelos ZIB e ZIP e $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \phi)^\top$ nos modelos ZINB e ZIBB.

Derivando (3.7) com respeito a $\boldsymbol{\omega}^\top$ temos que

$$\frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = \left(Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}), \ldots, Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})\right),$$

logo, derivando com respeito a $\boldsymbol{\theta}$ a expressão anterior, temos que

$$\begin{aligned}\boldsymbol{\Delta} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} = \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top}\right) = \frac{\partial}{\partial \boldsymbol{\theta}} \left(Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}), \ldots, Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})\right) \\ &= \left(\frac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}, \ldots, \frac{\partial Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right).\end{aligned}$$

Observe que, para o esquema de perturbação da ponderação de casos, a matriz $\boldsymbol{\Delta}$ não depende do vetor $\boldsymbol{\omega}$.

Em geral, no modelo de regressão inflacionado de zeros temos que $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \phi)^\top$, em que $\boldsymbol{\gamma}$ é de dimensão $(q \times 1)$, $\boldsymbol{\beta}$ é de dimensão $(p \times 1)$ e ϕ é escalar; então $\boldsymbol{\Delta}$ é uma matriz de dimensão $(p+q+1) \times n$ dada por

$$\boldsymbol{\Delta} = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} = \begin{pmatrix} \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \gamma_1 \partial \omega_1}, & \cdots, & \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \gamma_1 \partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \gamma_q \partial \omega_1}, & \cdots, & \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \gamma_q \partial \omega_n} \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \beta_1 \partial \omega_1}, & \cdots, & \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \beta_1 \partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \beta_p \partial \omega_1}, & \cdots, & \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \beta_p \partial \omega_n} \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \phi \partial \omega_1}, & \cdots, & \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \phi \partial \omega_n} \end{pmatrix},$$

avaliada em $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Assim, é possível dividir $\boldsymbol{\Delta}$ na forma

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_\gamma \\ \boldsymbol{\Delta}_\beta \\ \boldsymbol{\Delta}_\phi \end{pmatrix},$$

em que

$$\begin{aligned} \boldsymbol{\Delta}_\gamma &= \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \\ \boldsymbol{\Delta}_\beta &= \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \\ \boldsymbol{\Delta}_\phi &= \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}. \end{aligned}$$

A matriz $\boldsymbol{\Delta}_\gamma$ é a mesma para os quatro modelos estudados neste trabalho e é dada por:

$$\boldsymbol{\Delta}_\gamma = \mathbf{G}^\top \operatorname{diag}\{\hat{z}_1 - p_1, \ldots, \hat{z}_n - p_n\}, \tag{3.8}$$

com $p_i = \dfrac{\boldsymbol{e}^{\mathbf{G}_i \boldsymbol{\gamma}}}{1 + \boldsymbol{e}^{\mathbf{G}_i \boldsymbol{\gamma}}}$.

A matriz hessiana

A fim de obter as medidas de diagnóstico para a influência local do esquema de perturbação de ponderação de casos, é necessário calcular a matriz hessiana que, em geral, é expressa como

$$\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{\gamma\gamma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{\beta\beta} & \ddot{Q}_{\beta\phi} \\ \mathbf{0} & \ddot{Q}_{\phi\beta} & \ddot{Q}_{\phi\phi} \end{pmatrix}, \tag{3.9}$$

em que $\ddot{Q}_{\gamma\gamma} = \dfrac{\partial^2 Q_1(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}$, $\ddot{Q}_{\beta\beta} = \dfrac{\partial^2 Q_2(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}$, $\ddot{Q}_{\beta\phi} = \dfrac{\partial^2 Q_2(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\phi}$, $\ddot{Q}_{\phi\beta} = \ddot{Q}_{\beta\phi}^\top$ e $\ddot{Q}_{\phi\phi} = \dfrac{\partial^2 Q_2(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}})}{\partial\phi^2}$ são matrizes de dimensões $(q\times q)$, $(p\times p)$, $(p\times 1)$, $(1\times p)$ e (1×1), respectivamente.

Note que para os modelos ZINB e ZIBB, a matriz hessiana é de dimensão $(p+q+1)\times(p+q+1)$ e para os modelos ZIB e ZIP é de dimensão $(p+q)\times(p+q)$ devido à ausência do parâmetro ϕ nesses dois últimos modelos.

A matriz $\ddot{Q}_{\gamma\gamma}$ é a mesma para os quatro modelos utilizados e pela equação (3.6) temos que

$$\begin{aligned} \ddot{Q}_{\gamma\gamma} &= \frac{\partial^2 Q_1(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top} = \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\frac{\partial Q_1(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\gamma}^\top}\right] = \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\sum_{i=1}^{n}\left\{\hat{z}_i\mathbf{G}_i - \frac{\boldsymbol{e}^{\mathbf{G}_i\boldsymbol{\gamma}}\mathbf{G}_i}{1+\boldsymbol{e}^{\mathbf{G}_i\gamma}}\right\}\right] \\ &= -\sum_{i=1}^{n}\frac{\boldsymbol{e}^{\mathbf{G}_i\boldsymbol{\gamma}}\mathbf{G}_i^\top\mathbf{G}_i}{(1+\boldsymbol{e}^{\mathbf{G}_i\gamma})^2} = -\sum_{i=1}^{n} p_i(1-p_i)\mathbf{G}_i^\top\mathbf{G}_i, \end{aligned}$$

lembrando (3.8). Escrevendo em forma matricial a expressão anterior temos que

$$\ddot{Q}_{\gamma\gamma} = -\mathbf{G}^\top \operatorname{diag}\left\{p_1(1-p_1),\ldots,p_n(1-p_n)\right\}\mathbf{G}.$$

Apresentamos em seguida a matriz $\boldsymbol{\Delta}_\beta$ para cada um dos quatro modelos utilizados, a matriz $\boldsymbol{\Delta}_\phi$ para os modelos ZINB e ZIBB e os blocos da matriz hessiana para cada um dos modelos.

Modelo ZIB:

Utilizando (3.6), temos que a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ para o modelo de regressão ZIB é expressa como

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\Bigg\{ &\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\left(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\right) + (1-\hat{z}_i)\Bigg[y_i\mathbf{B}_i\boldsymbol{\beta} - n_i\log\left(1+e^{\mathbf{B}_i\boldsymbol{\beta}}\right) \\ &+ \log\binom{n_i}{y_i}\Bigg]\Bigg\}. \end{aligned}$$

Daí,

$$\boldsymbol{\Delta}_{\beta} = \mathbf{B}^{\top} \operatorname{diag}\Big\{(1-\hat{z}_1)(y_1-n_1\pi_1), \ldots, (1-\hat{z}_n)(y_n-n_n\pi_n)\Big\},$$

em que $\pi_i = \dfrac{\boldsymbol{e}^{\mathbf{B}_i\boldsymbol{\beta}}}{1+\boldsymbol{e}^{\mathbf{B}_i\boldsymbol{\beta}}}$. Logo,

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{G}^{\top} \operatorname{diag}\Big\{\hat{z}_1 - p_1, \ldots, \hat{z}_n - p_n\Big\} \\ \mathbf{B}^{\top} \operatorname{diag}\Big\{(1-\hat{z}_1)(y_1-n_1\pi_1), \ldots, (1-\hat{z}_n)(y_n-n_n\pi_n)\Big\} \end{bmatrix},$$

avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\gamma}}_1, \ldots, \widehat{\boldsymbol{\gamma}}_q, \widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p)^{\top}$. Derivando $Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})$ em (3.6) duas vezes,

$$\begin{aligned} \ddot{\boldsymbol{Q}}_{\beta\beta} &= \frac{\partial}{\partial\boldsymbol{\beta}}\left[\frac{\partial Q_2(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}^{\top}}\right] = \frac{\partial}{\partial\boldsymbol{\beta}}\left[\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\left(y_i\mathbf{B}_i - \frac{n_i\boldsymbol{e}^{\mathbf{B}_i\beta}\mathbf{B}_i}{1+\boldsymbol{e}^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right\}\right] \\ &= -\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\frac{n_i\boldsymbol{e}^{\mathbf{B}_i\boldsymbol{\beta}}\boldsymbol{\beta}\mathbf{B}_i^{\top}\mathbf{B}_i}{(1+\boldsymbol{e}^{\mathbf{B}_i\beta})^2}\right\} = -\sum_{i=1}^{n}\Big\{(1-\hat{z}_i)n_i\pi_i(1-\pi_i)\mathbf{B}_i^{\top}\mathbf{B}_i\Big\}. \end{aligned} \tag{3.10}$$

Escrevendo de forma matricial a expressão anterior temos que

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = -\mathbf{B}^{\top} \operatorname{diag}\Big\{(1-\hat{z}_1)n_1\pi_1(1-\pi_1), \ldots, (1-\hat{z}_n)n_n\pi_n(1-\pi_n)\Big\}\mathbf{B}. \tag{3.11}$$

Portanto, a matriz hessiana para o modelo de regressão ZIB é

$$\ddot{\boldsymbol{Q}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\mathbf{G}^{\top} \operatorname{diag}\big\{p_1(1-p_1), \ldots, p_n(1-p_n)\big\}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \ddot{\boldsymbol{Q}}_{\beta\beta} \end{bmatrix},$$

sendo que $\ddot{\boldsymbol{Q}}_{\beta\beta}$ é expressa na equação (3.11).

Modelo ZIP

Utilizando (3.6), temos que a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ para o modelo de regressão ZIP é

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\left\{\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\Big(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\Big) + (1-\hat{z}_i)\Big(y_i\mathbf{B}_i\boldsymbol{\beta} - e^{\mathbf{B}_i\boldsymbol{\beta}} - \log(y_i!)\Big)\right\}.$$

Daí,

$$\boldsymbol{\Delta}_{\beta} = \mathbf{B}^{\top} \operatorname{diag}\Big\{(1-\hat{z}_1)(y_1-\lambda_1), \ldots, (1-\hat{z}_n)(y_n-\lambda_n)\Big\},$$

em que $\lambda_i = \boldsymbol{e}^{B_i\beta}$. Portanto,

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{G}^{\top} \operatorname{diag}\Big\{\hat{z}_1 - p_1, \ldots, \hat{z}_n - p_n\Big\} \\ \mathbf{B}^{\top} \operatorname{diag}\Big\{(1-\hat{z}_1)(y_1-\lambda_1), \ldots, (1-\hat{z}_n)(y_n-\lambda_n)\Big\} \end{bmatrix},$$

avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_q, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)^\top$. Derivando $Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})$ em (3.6) duas vezes,

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{\partial Q_2(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}^\top}\right] = \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sum_{i=1}^{n}\Big\{(1-\hat{z}_i)\left(y_i\mathbf{B}_i - e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i\right)\Big\}\right]$$
$$= -\sum_{i=1}^{n}\Big\{(1-\hat{z}_i)e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i^\top\mathbf{B}_i\Big\} = -\sum_{i=1}^{n}\Big\{(1-\hat{z}_i)\lambda_i\mathbf{B}_i^\top\mathbf{B}_i\Big\}. \tag{3.12}$$

Escrevendo de forma matricial a expressão anterior temos que

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = -\mathbf{B}^\top \operatorname{diag}\big\{(1-\hat{z}_1)\lambda_1, \ldots, (1-\hat{z}_n)\lambda_n\Big\}\mathbf{B}. \tag{3.13}$$

Assim, a matriz hessiana para o modelo de regressão ZIP é

$$\ddot{\boldsymbol{Q}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\mathbf{G}^\top \operatorname{diag}\big\{p_1(1-p_1), \ldots, p_n(1-p_n)\Big\}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \ddot{\boldsymbol{Q}}_{\beta\beta} \end{bmatrix},$$

sendo $\ddot{\boldsymbol{Q}}_{\beta\beta}$ vem de (3.13).

Modelo ZINB

Utilizando (3.6), temos que a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ para o modelo de regressão ZINB é

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\Bigg\{\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\Big(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\Big) + (1-\hat{z}_i)\Big[\log\Gamma(\phi+y_i) + y_i\mathbf{B}_i\boldsymbol{\beta}$$
$$- (y_i+\phi)\log\left(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi\right) + \phi\log(\phi) - \log\Gamma(y_i+1) - \log\Gamma(\phi)\Big]\Bigg\}. \tag{3.14}$$

Daí,

$$\boldsymbol{\Delta}_\beta = \mathbf{B}^\top \operatorname{diag}\left\{\frac{(1-\hat{z}_1)\phi(y_1-\mu_1)}{\mu_1+\phi}, \ldots, \frac{(1-\hat{z}_n)\phi(y_n-\mu_n)}{\mu_n+\phi}\right\}$$

e

$$\boldsymbol{\Delta}_\phi = \Big((1-\hat{z}_1)b_1, \ldots, (1-\hat{z}_n)b_n\Big),$$

com $b_i = \psi(\phi+y_i) - \dfrac{y_i+\phi}{\mu_i+\phi} + \log(\phi) + 1 - \log(\mu_i+\phi) - \psi(\phi)$, $\mu_i = e^{\mathbf{B}_i\boldsymbol{\beta}}$ e $\psi(a)$ é a função digama, definida como $\psi(a) = \dfrac{\partial\log(\Gamma(a))}{\partial a}$. Então,

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{G}^\top \operatorname{diag}\{\hat{z}_1 - p_1, \ldots, \hat{z}_n - p_n\} \\ \mathbf{B}^\top \operatorname{diag}\left\{\dfrac{(1-\hat{z}_1)\phi(y_1-\mu_1)}{\mu_1+\phi}, \ldots, \dfrac{(1-\hat{z}_n)\phi(y_n-\mu_n)}{\mu_n+\phi}\right\} \\ \Big((1-\hat{z}_1)b_1, \ldots, (1-\hat{z}_n)b_n\Big) \end{bmatrix},$$

avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_q, \widehat{\beta}_1, \ldots, \widehat{\beta}_p, \widehat{\phi})^\top$. Derivando $Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})$ em (3.6) duas vezes obtemos

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = \frac{\partial}{\partial\boldsymbol{\beta}}\left[\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\Big[y_i\mathbf{B}_i - (y_i+\phi)\frac{e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i}{e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi}\Big]\right\}\right] = -\sum_{i=1}^{n}\left\{\frac{(1-\hat{z}_i)(y_i+\phi)\phi e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i^\top B_i}{(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi)^2}\right\}$$
$$= -\sum_{i=1}^{n}\left\{\frac{(1-\hat{z}_i)(y_i+\phi)\phi\mu_i\mathbf{B}_i^\top\mathbf{B}_i}{(\mu_i+\phi)^2}\right\}, \tag{3.15}$$

$$\ddot{Q}_{\phi\phi} = \frac{\partial}{\partial\phi}\left[\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\left[\psi(\phi+y_i) - \log\left(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi\right) - \frac{y_i+\phi}{e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi} + \log(\phi) + 1 - \psi(\phi)\right]\right\}\right]$$
$$= -\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\left[\psi\prime(\phi+y_i) - \frac{1}{e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi} - \frac{e^{\mathbf{B}_i\boldsymbol{\beta}}-y_i}{(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi)^2} + \frac{1}{\phi} - \psi\prime(\phi)\right]\right\}$$
$$= -\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\left[\psi\prime(\phi+y_i) - \frac{1}{\mu_i+\phi} - \frac{\mu_i-y_i}{(\mu_i+\phi)^2} + \frac{1}{\phi} - \psi\prime(\phi)\right]\right\}, \tag{3.16}$$

$$\ddot{\boldsymbol{Q}}_{\beta\phi} = \frac{\partial}{\partial\boldsymbol{\beta}}\left[\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\left[\psi(\phi+y_i) - \log\left(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi\right) - \frac{y_i+\phi}{e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi} + \log(\phi) + 1 - \psi(\phi)\right]\right\}\right]$$
$$= -\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\frac{(e^{\mathbf{B}_i\boldsymbol{\beta}}-y_i)e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i^\top}{(e^{\mathbf{B}_i\boldsymbol{\beta}}+\phi)^2}\right\} = -\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\frac{(\mu_i-y_i)\mu_i\mathbf{B}_i^\top}{(\mu_i+\phi)^2}\right\}, \tag{3.17}$$

em que $\psi\prime(a) = \dfrac{\partial\psi(a)}{\partial a}$.

Escrevendo de forma matricial as expressões (3.15) e (3.17) temos que

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = -\mathbf{B}^\top \operatorname{diag}\left\{\frac{(1-\hat{z}_1)(y_1+\phi)\phi\mu_1}{(\mu_1+\phi)^2}, \ldots, \frac{(1-\hat{z}_n)(y_n+\phi)\phi\mu_n}{(\mu_n+\phi)^2}\right\}\mathbf{B} \tag{3.18}$$

e

$$\ddot{\boldsymbol{Q}}_{\beta\phi} = -\mathbf{B}^\top\left(\frac{(1-\hat{z}_1)(\mu_1-y_1)\mu_1}{(\mu_1+\phi)^2}, \ldots, \frac{(1-\hat{z}_n)(\mu_n-y_n)\mu_n}{(\mu_n+\phi)^2}\right)^\top. \tag{3.19}$$

Portanto, a matriz hessiana para o modelo de regressão ZINB é

$$\ddot{\boldsymbol{Q}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\mathbf{G}^\top \operatorname{diag}\Big\{p_1(1-p_1), \ldots, p_n(1-p_n)\Big\}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\boldsymbol{Q}}_{\beta\beta} & \ddot{\boldsymbol{Q}}_{\beta\phi} \\ \mathbf{0} & \ddot{\boldsymbol{Q}}_{\beta\phi}^\top & \ddot{Q}_{\phi\phi} \end{bmatrix},$$

sendo que $\ddot{\boldsymbol{Q}}_{\beta\beta}$, $\ddot{\boldsymbol{Q}}_{\beta\phi}$ e $\ddot{\boldsymbol{Q}}_{\phi\phi}$ são encontradas nas equações (3.18), (3.19) e (3.16), respectivamente.

Modelo ZIBB

Utilizando (3.6), temos que a função $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ para o modelo de regressão ZIBB é

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\left\{\hat{z}_i\mathbf{G}_i\boldsymbol{\gamma} - \log\left(1+e^{\mathbf{G}_i\boldsymbol{\gamma}}\right) + (1-\hat{z}_i)\left[\log\binom{n_i}{y_i} + \log\Gamma\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right.\right.$$
$$+\log\Gamma\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i\right) + \log\Gamma(\phi) - \log\Gamma(n_i+\phi) - \log\Gamma\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)$$
$$\left.\left.-\log\Gamma\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right]\right\}. \tag{3.20}$$

Daí, $\boldsymbol{\Delta}_\beta = \mathbf{B}^\top \operatorname{diag}\left\{(1-\hat{z}_1)\phi\pi_1(1-\pi_1)c_1, \ldots, (1-\hat{z}_n)\phi\pi_n(1-\pi_n)c_n\right\}$, com $c_i = \psi(y_i + \phi\pi_i) - \psi(n_i + \phi(1-\pi_i) - y_i) + \psi(\phi(1-\pi_i)) - \psi(\phi\pi_i)$, $\pi_i = \dfrac{e^{B_i\beta}}{1+e^{B_i\beta}}$ e

$$\boldsymbol{\Delta}_\phi = \Big((1-\hat{z}_1)d_1, \ldots, (1-\hat{z}_n)d_n\Big),$$

com

$$d_i = \pi_i\left[\psi(y_i+\phi\pi_i) - \psi(\phi\pi_i)\right] + (1-\pi_i)\left[\psi(n_i+\phi(1-\pi_i)-y_i) - \psi(\phi(1-\pi_i))\right] + \psi(\phi) - \psi(n_i+\phi).$$

Então,

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{G}^\top \operatorname{diag}\{\hat{z}_1 - p_1, \ldots, \hat{z}_n - p_n\} \\ \mathbf{B}^\top \operatorname{diag}\left\{(1-\hat{z}_1)\phi\pi_1(1-\pi_1)c_1, \ldots, (1-\hat{z}_n)\phi\pi_n(1-\pi_n)c_n\right\} \\ \Big((1-\hat{z}_1)d_1, \ldots, (1-\hat{z}_n)d_n\Big) \end{bmatrix},$$

avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_q, \widehat{\beta}_1, \ldots, \widehat{\beta}_p, \widehat{\phi})^\top$. Derivando $Q_b(\boldsymbol{\beta}, \phi|\widehat{\boldsymbol{\theta}})$ em (3.6) duas vezes chegamos a

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = \frac{\partial}{\partial\boldsymbol{\beta}}\left(\sum_{i=1}^{n}\left\{(1-\hat{z}_i)\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}\mathbf{B}_i}{\left(1+e^{\mathbf{B}_i\boldsymbol{\beta}}\right)^2}\left[\psi\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right) - \psi\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i\right)\right.\right.\right.$$
$$\left.\left.\left.+\psi\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right) - \psi\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right]\right\}\right)$$
$$= \sum_{i=1}^{n}\left\{\frac{(1-\hat{z}_i)\phi^2 e^{2\mathbf{B}_i\boldsymbol{\beta}}}{\left(1+e^{\mathbf{B}_i\boldsymbol{\beta}}\right)^4}\left[\psi'\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right) + \psi'\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i\right) - \psi'\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right.\right.$$
$$\left.\left.-\psi'\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right]\mathbf{B}_i^\top\mathbf{B}_i + (1-\hat{z}_i)\phi\frac{e^{\mathbf{B}_i\boldsymbol{\beta}}\left(1-e^{\mathbf{B}_i\boldsymbol{\beta}}\right)}{\left(1+e^{\mathbf{B}_i\boldsymbol{\beta}}\right)^3}\left[\psi\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right.\right.$$
$$\left.\left.-\psi\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i\right) - \psi\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right) + \psi\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}}\right)\right]\mathbf{B}_i^\top\mathbf{B}_i\right\}$$

$$= \sum_{i=1}^{n} \Bigg\{ (1-\hat{z}_i) \Bigg(\phi^2 \pi_i^2 (1-\pi_i)^2 \Big[\psi\prime(y_i + \phi\pi_i) + \psi\prime\left(n_i + \phi(1-\pi_i) - y_i\right) - \psi\prime(\phi(1-\pi_i)) - \psi\prime(\phi\pi_i) \Big] + \phi\pi_i(1-\pi_i)(1-2\pi_i) \Big[\psi(y_i + \phi\pi_i) - \psi(n_i + \phi(1-\pi_i) - y_i) - \psi(\phi\pi_i) + \psi(\phi(1-\pi_i)) \Big] \Bigg) \mathbf{B}_i^\top \mathbf{B}_i \Bigg\}, \tag{3.21}$$

$$\begin{aligned}
\ddot{Q}_{\phi\phi} &= \frac{\partial}{\partial\phi} \left[\sum_{i=1}^{n} \left((1-\hat{z}_i) \left\{ \frac{e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \left[\psi\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] + \psi(\phi) \right.\right. \right. \\
&\quad \left.\left.\left. + \frac{1}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \left[\psi\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i \right) - \psi\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] - \psi(n_i+\phi) \right\} \right) \right] \\
&= \sum_{i=1}^{n} \left\{ (1-\hat{z}_i) \left(\frac{e^{2\mathbf{B}_i\boldsymbol{\beta}}}{(1+e^{\mathbf{B}_i\boldsymbol{\beta}})^2} \left[\psi\prime\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\prime\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] + \psi\prime(\phi) \right.\right. \\
&\quad \left.\left. + \frac{1}{(1+e^{\mathbf{B}_i\boldsymbol{\beta}})^2} \left[\psi\prime\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i \right) - \psi\prime\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] - \psi\prime(n_i+\phi) \right) \right\} \\
&= \sum_{i=1}^{n} \Bigg\{ (1-\hat{z}_i) \Bigg(\pi_i^2 \Big[\psi\prime(y_i + \phi\pi_i) - \psi\prime(\phi\pi_i) \Big] + \psi\prime(\phi) - \psi\prime(n_i+\phi) \\
&\quad + (1-\pi_i)^2 \Big[\psi\prime(n_i + \phi(1-\pi_i) - y_i) - \psi\prime(\phi(1-\pi_i)) \Big] \Bigg) \Bigg\},
\end{aligned} \tag{3.22}$$

$$\begin{aligned}
\ddot{\boldsymbol{Q}}_{\beta\phi} &= \frac{\partial}{\partial\boldsymbol{\beta}} \left[\sum_{i=1}^{n} \left((1-\hat{z}_i) \left\{ \frac{e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \left[\psi\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] \right.\right. \right. \\
&\quad \left.\left.\left. + \frac{1}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \left[\psi\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i \right) - \psi\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] + \psi(\phi) - \psi(n_i+\phi) \right\} \right) \right] \\
&= \sum_{i=1}^{n} \left\{ (1-\hat{z}_i) \left(\frac{\phi e^{2\mathbf{B}_i\boldsymbol{\beta}}}{(1+e^{\mathbf{B}_i\boldsymbol{\beta}})^3} \left[\psi\prime\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\prime\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] \right.\right. \\
&\quad + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{(1+e^{\mathbf{B}_i\boldsymbol{\beta}})^3} \left[\psi\prime\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\prime\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i \right) \right] \\
&\quad \left.\left. + \frac{e^{\mathbf{B}_i\boldsymbol{\beta}}}{(1+e^{\mathbf{B}_i\boldsymbol{\beta}})^2} \left[\psi\left(y_i + \frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) + \psi\left(\frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) - \psi\left(n_i + \frac{\phi}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} - y_i \right) - \psi\left(\frac{\phi e^{\mathbf{B}_i\boldsymbol{\beta}}}{1+e^{\mathbf{B}_i\boldsymbol{\beta}}} \right) \right] \right) \mathbf{B}_i^\top \right\} \\
&= \sum_{i=1}^{n} \Bigg\{ (1-\hat{z}_i) \Bigg(\phi\pi_i^2(1-\pi_i) \Big[\psi\prime\left(y_i + \phi\pi_i\right) - \psi\prime\left(\phi\pi_i\right) \Big] + \pi_i(1-\pi_i) \Big[\psi\left(y_i + \phi\pi_i\right) \\
&\quad + \psi\left(\phi(1-\pi_i)\right) - \psi\left(n_i + \phi(1-\pi_i) - y_i\right) - \psi\left(\phi\pi_i\right) \Big]
\end{aligned}$$

$$+ \phi\pi_i(1-\pi_i)^2\Big[\psi'(\phi(1-\pi_i)) - \psi'(n_i+\phi(1-\pi_i)-y_i)\Big]\Big)\mathbf{B}_i^\top\Big\}. \tag{3.23}$$

Escrevendo de forma matricial as expressões (3.21) e (3.23) temos que

$$\ddot{\boldsymbol{Q}}_{\beta\beta} = \mathbf{B}^\top \boldsymbol{T}_1 \mathbf{B} = \mathbf{B}^\top \operatorname{diag}\{t_1,\ldots,t_n\}\,\mathbf{B}, \tag{3.24}$$

sendo que

$$\begin{aligned} t_i = (1-\hat{z}_i)\Big\{&\phi^2\pi_i^2(1-\pi_i)^2\Big[\psi'(y_i+\phi\pi_i)+\psi'(n_i+\phi(1-\pi_i)-y_i)-\psi'(\phi\pi_i)\\ &-\psi'(\phi(1-\pi_i))\Big]+\phi\pi_i(1-\pi_i)(1-2\pi_i)\Big[\psi(y_i+\phi\pi_i)-\psi(n_i+\phi(1-\pi_i)-y_i)\\ &-\psi(\phi\pi_i)+\psi(\phi(1-\pi_i))\Big]\Big\} \quad \text{e} \end{aligned}$$

$$\ddot{\boldsymbol{Q}}_{\beta\phi} = \mathbf{B}^\top \boldsymbol{T}_2 = \mathbf{B}^\top (\tau_1,\ldots,\tau_n)^\top, \tag{3.25}$$

com

$$\begin{aligned} \tau_i = (1-\hat{z}_i)\Big\{&\phi\pi_i^2(1-\pi_i)\Big[\psi'(y_i+\phi\pi_i)-\psi'(\phi\pi_i)\Big]+\pi_i(1-\pi_i)\Big[\psi(y_i+\phi\pi_i)\\ &+\psi(\phi(1-\pi_i))-\psi(n_i+\phi(1-\pi_i)-y_i)-\psi(\phi\pi_i)\Big]\\ &+\phi\pi_i(1-\pi_i)^2\Big[\psi'(\phi(1-\pi_i))-\psi'(n_i+\phi(1-\pi_i)-y_i)\Big]\Big\}, \quad i=1,\ldots,n \end{aligned}$$

Portanto, a matriz hessiana para o modelo de regressão ZIBB é

$$\ddot{\boldsymbol{Q}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\mathbf{G}^\top \operatorname{diag}\Big\{p_1(1-p_1),\ldots,p_n(1-p_n)\Big\}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^\top\boldsymbol{T}_1\mathbf{B} & \mathbf{B}^\top\boldsymbol{T}_2 \\ \mathbf{0} & \boldsymbol{T}_2^\top\mathbf{B} & \ddot{Q}_{\phi\phi} \end{bmatrix},$$

sendo que $\mathbf{B}^\top\boldsymbol{T}_1\mathbf{B}$, $\mathbf{B}^\top\boldsymbol{T}_2$ e $\ddot{Q}_{\phi\phi}$ são expressas nas equações (3.24), (3.25) e (3.22), respectivamente.

Influência global

Um dos métodos de diagnóstico empregados em modelos de regressão utiliza a exclusão de casos que consiste em comparar os estimadores de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{[i]}$. Estes métodos são utilizados com freqüência para diagnosticar globalmente possíveis observações influentes e podem ser facilmente adaptados aos modelos para dados inflacionados de zeros,

dado que podemos calcular os estimadores de máxima verossimilhança com e sem a i-ésima observação.

Neste trabalho utilizaremos duas medidas de diagnóstico de influência global, uma chamada de LDi (afastamento da verossimilhança) definida em (1.13), e outra chamada de distância de Cook generalizada, definida como

$$D_i = (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{[i]})^\top(-\ddot{\boldsymbol{Q}}(\widehat{\boldsymbol{\theta}}))(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{[i]}),$$

em que $-\ddot{\boldsymbol{Q}}$ é matriz hessiana definida em (3.9), $\widehat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ com todos os dados da amostra e $\widehat{\boldsymbol{\theta}}_{[i]}$ é o EMV de $\boldsymbol{\theta}$ com a exclusão da i-ésima observação.

Capítulo 4

Aplicações

Este capítulo tem como objetivo ilustrar a metodologia desenvolvida nesta dissertação.

4.1 Estudo de Simulação

Utilizando dados simulados gerados pelos programas listados no Apêndice A.1.2, aplicamos o algoritmo EM (Apêndice A.1.1) para estimar os parâmetros dos quatro modelos de regressão inflacionados de zeros discutidos nos capítulos anteriores. São eles ZIB, ZIP, ZINB e ZIBB. Na simulação consideramos um modelo de regressão inflacionado de zeros com a seguinte estrutura:

$$\begin{aligned} g(\mu_i) &= \beta_1 B_1 + \beta_2 B_2 \\ \text{e} \qquad h(p_i) &= \gamma_1 G_1 + \gamma_2 G_2, \end{aligned} \tag{4.1}$$

$i = 1, \ldots, n$, em que h é a função de ligação logito nos quatro modelos e g é a função de ligação logito nos modelos ZIB e ZIBB e a função de ligação log-linear nos modelos ZIP e ZINB. Desta forma, a variável resposta tem distribuição $y_i \sim ZIB(p_i, \mu_i, n_i)$ com $n_i = 25$ e $i = 1, \ldots, n$ para o modelo binomial inflacionado de zeros, $y_i \sim ZIP(p_i, \mu_i)$ com $i = 1, \ldots, n$ para o modelo Poisson inflacionado de zeros, $y_i \sim ZINB(p_i, \mu_i, \phi)$ com $i = 1, \ldots, n$ para o modelo binomial negativo inflacionado de zeros e $y_i \sim ZIBB(p_i, \mu_i, \phi, n_i)$ com $n_i = 25$ e $i = 1, \ldots, n$ para o modelo beta-binomial inflacionado de zeros. Nos modelos ZINB e ZIBB, ϕ é o parâmetro de dispersão. Os valores das covariáveis B_1 e B_2, que nesta simulação foram consideradas $B_1 = G_1$ e $B_2 = G_2$, foram tomados do pacote Zicounts em R e correspondem às variáveis *gender* e *age*, respectivamente. O tamanho de amostra considerado foi $n = 200$ e foram simuladas 10 000 amostras Monte Carlo.

Nas Tabelas 4.1–4.4 apresentamos os verdadeiros valores dos parâmetros de regressão, as médias dos estimadores dos parâmetros nas 10 000 replicações, juntamente com a média dos erros padrões empíricos obtidos através da matriz de informação observada em cada uma das 10 000 amostras e também o desvio padrão amostral das 10 000 estimativas obtidas dos parâmetros. Além disso, para uma das amostras implementamos o gráfico de envelope para os resíduos padronizados e ponderados do modelo binomial inflacionado de zeros ($y_i \sim ZIB(p_i, \mu_i, 25)$, com $i = 1, \ldots, 200$) e do modelo Poisson inflacionado de zeros ($y_i \sim ZIP(p_i, \mu_i)$, com $i = 1, \ldots, 200$), cujos programas são dados no Apêndice A.1.3, e são mostrados nas Figuras 4.1 e 4.2. De acordo com as Tabelas 4.1–4.4, percebe-se que o algoritmo EM se comporta bem em todos os casos, as estimativas dos parâmetros diferem um pouco dos valores verdadeiros. Em relação ao erro padrão, tanto o empírico como o teórico, em todos os modelos os mesmos são muito similares, o que demonstra a eficiência da matriz de informação observada.

Tabela 4.1: Estimativas por máxima verossimilhança com erros padrões (EP) para os dados simulados do modelo $ZIB(p_i, \mu_i, 25)$ considerando 10 000 simulações.

Parâmetro	γ_1	γ_2	β_1	β_2
Valor verdadeiro	-0,5	0,23	0,3	0,1
Média das estimativas	-0,503	0,235	0,299	0,100
Desvio padrão das estimativas	0,385	0,047	0,150	0,018
Média dos EP das estimativas	0,378	0,045	0,148	0,018

Tabela 4.2: Estimativas por máxima verossimilhança com erros padrões (EP) para os dados simulados do modelo $ZIP(p_i, \mu_i)$, considerando 10 000 simulações.

Parâmetro	γ_1	γ_2	β_1	β_2
Valor verdadeiro	-0,35	-0,1	0,5	0,25
Média das estimativas	-0,347	-0,102	0,501	0,250
Desvio padrão das estimativas	0,324	0,036	0,068	0,008
Média dos EP das estimativas	0,322	0,035	0,067	0,008

Tabela 4.3: Estimativas por máxima verossimilhança com erros padrões (EP) para os dados simulados do modelo $ZINB(p_i, \mu_i, \phi)$, considerando 10 000 simulações.

Parâmetro	γ_1	γ_2	β_1	β_2	ϕ
Valor verdadeiro	-0,35	-0,1	0,5	0,25	1
Média das estimativas	-0,350	-0,104	0,502	0,249	1.060
Desvio padrão das estimativas	0,423	0,053	0,204	0,025	0,235
Média dos EP das estimativas	0,459	0,057	0,224	0,029	0,236

Tabela 4.4: Estimativas por máxima verossimilhança com erros padrões (EP) para os dados simulados do modelo $ZIBB(p_i, \mu_i, \phi, 25)$, considerando 10 000 simulações.

Parâmetro	γ_1	γ_2	β_1	β_2	ϕ
Valor verdadeiro	-0,5	0,3	0,3	0,1	1
Média das estimativas	-0,517	0,301	0,297	0,099	1,108
Desvio padrão das estimativas	0,455	0,056	0,669	0,089	0,445
Média dos EP das estimativas	0,451	0,063	0,644	0,095	0,514

4.2 Aplicações a Dados Reais

Em seguida são apresentados dois conjuntos de dados reais com os quais exemplificamos os modelos propostos na Seção 2.3.

4.2.1 Exemplo 1

O controle biológico aplicado (CBA) é um ramo importante da Entomologia, que busca métodos para controle de pragas sem prejudicar o meio ambiente e o ecossistema, utilizando inimigos naturais da praga (parasitóides e/ou predadores). Especificamente no Brasil, uma praga que vem sendo combatida através de CBA é a broca-da-cana (*Diatraea saccharalis*), que é a principal praga da cana-de-açúcar neste país. O conjunto de dados é proveniente de um experimento completamente casualizado com 10 repetições realizado no Laboratório de Biologia de Departamento de Entomologia da ESALQ/USP. A espécie de parasita *Trichogramma galloi*, tendo número variável de fêmeas (2, 4, 8, 16, 32, 64 e 128), foi colocada para parasitar 128 ovos de *Anagasta kuehniella*, um hospedeiro alternativo ao

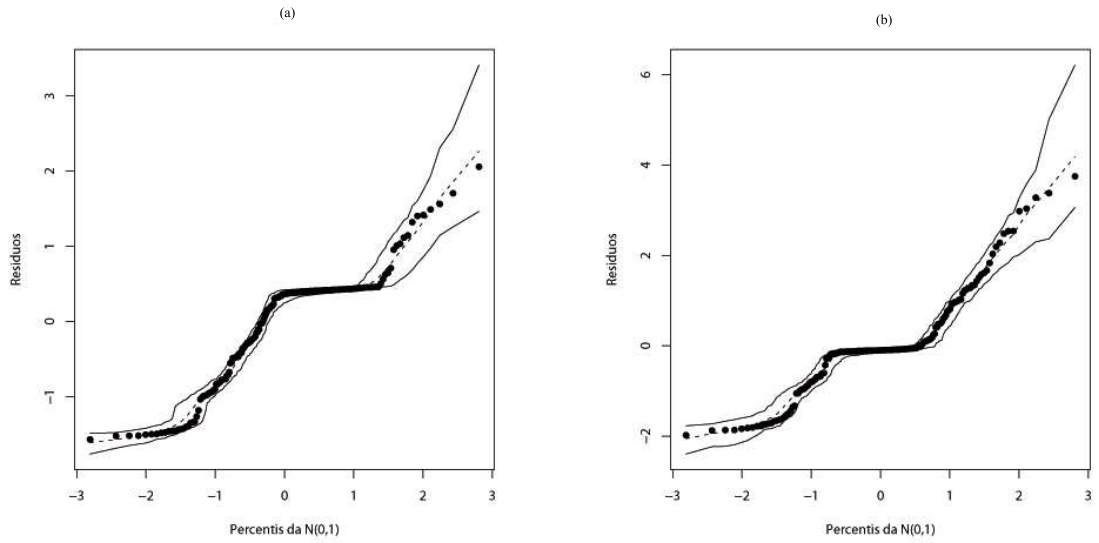


Figura 4.1: Gráficos normais de probabilidades com envelopes simulados para os resíduos ponderados (a) e padronizados (b) do modelo Binomial inflacionado de zeros (ZIB).

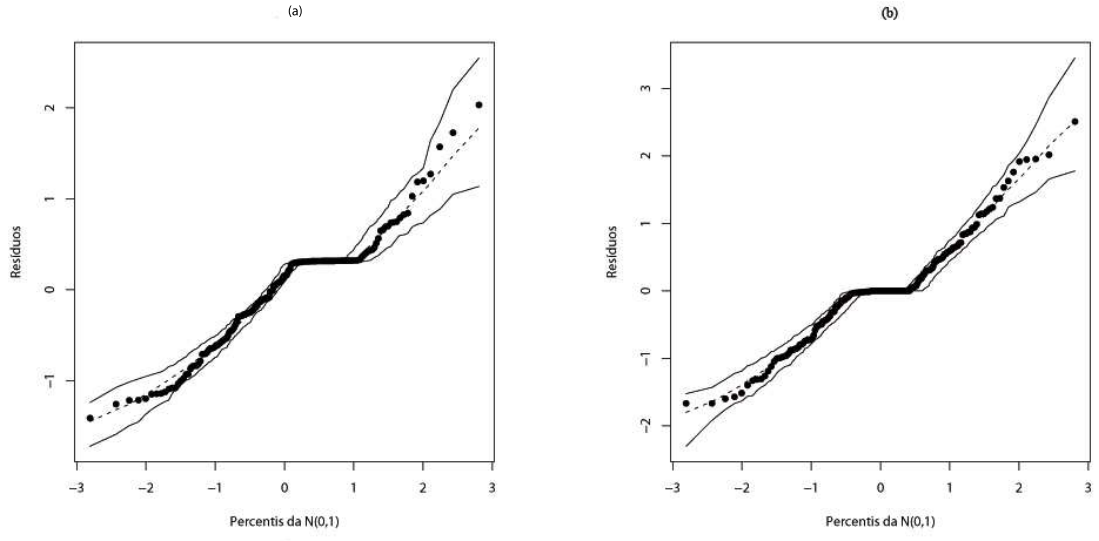


Figura 4.2: Gráficos normais de probabilidades com envelopes simulados para os residuos ponderados (a) e padronizados (b) do modelo Poisson inflacionado de zeros (ZIP).

D. saccharalis. A variável resposta observada foi o número de ovos parasitados dentre os 128 ovos do *A. kuehniella* (y), sendo que o interesse do pesquisador estava em determinar o número necessário de fêmeas para produzir o número máximo de ovos parasitados. Ini-

cialmente, foi realizada uma análise descritiva das variáveis envolvidas no problema. Na Tabela 4.5 estão apresentadas algumas medidas resumo. Na Figura 4.3 observamos que a distribuição do número de ovos parasitados apresenta duas modas e há uma concentração de observações em zero (53%), o que nos faz pensar que poderia ser melhor realizar um ajuste com o modelo de regressão binomial inflacionado de zeros em vez do modelo de regressão binomial padrão.

Tabela 4.5: Medidas resumo das variáveis número de ovos parasitados (y) e $\log_2$(número de fêmeas) - Exemplo 1.

Medida	y	$\log_2$(número de fêmeas)
Mínimo	0	1
Q_1	0	2
Mediana	0	4
Média	27,871	4
Q_3	55,750	6
Máximo	120	7
Desvio padrão	36,035	2,014

Para determinar se a freqüência do excesso de zeros poderia fazer que exista falta de ajuste do modelo binomial padrão, o teste escore para inflação de zeros é considerado a seguir.

Baseados nos dados coletados e considerando o modelo ZIB, foi calculado o valor de S usando a expressão (2.20) para testar a hipótese $H_0 : p = 0$ contra $H_1 : p > 0$ a fim de determinar se há excesso de zeros, dado que a proporcão de zeros estimados pelo modelo é um criterio importante para a qualidade do ajuste. Portanto, além de analisar os resíduos, a observação da proporção de zeros estimados pelo modelo ($\widehat{p}$) e da proporção na amostra pode ser relevante na escolha do melhor modelo.

A estatística do escore fornece $S = 283,66$ (com nível descritivo$< 0,0001$). O percentil da $\chi^2_{(1)}$, considerando o nível de 5% de significância, tem valor crítico do teste igual a 3,84 e, observa-se que o valor obtido é bem maior do que o valor crítico, levando à rejeição da hipótese nula $H_0 : p = 0$. Pode-se concluir que o modelo binomial inflacionado de zeros é mais apropriado do que o modelo binomial padrão, que é a distribuição sob a hipótese nula para estes dados.

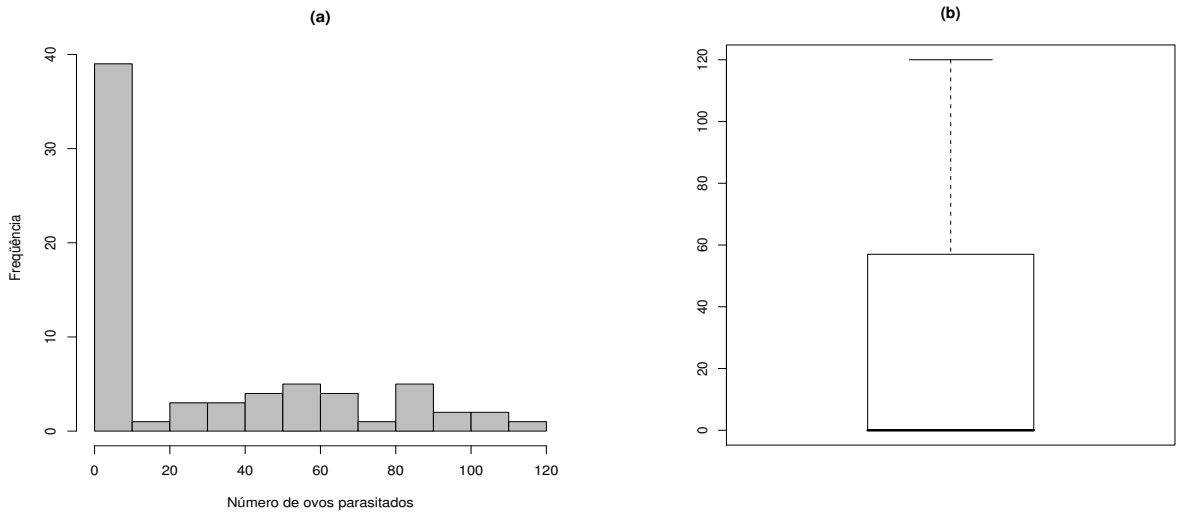


Figura 4.3: Histograma (a) e gráfico de caixas (b) do número de ovos parasitados - Exemplo 1.

Assumimos um modelo de regressão binomial inflacionado de zeros, em que $y_i \sim ZIB(p_i, \pi_i, n_i)$, $i = 1, \ldots, 70$, são variáveis aleatórias independentes, tais que

$$\begin{aligned} logit(p_i) &= \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 x_i^3 + \gamma_4 x_i^4 + \gamma_5 x_i^5, \\ logit(\pi_i) &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5, \end{aligned}$$

em que x representa $\log_2$(Número de fêmeas). Borgatto (2004) faz um ajuste destes dados utilizando modelos de quarto grau. Para a estimação dos parâmetros é utilizado o algoritmo EM implementado em R (Apêndice A.1.1).

Utilizando o critério de informação de Akaike (AIC), selecionamos o melhor modelo que se ajusta aos dados (modelo mais parcimonioso) utilizando o método passo a passo. O modelo finalmente selecionado foi:

$$\begin{aligned} logit(p_i) &= \gamma_0, \\ logit(\pi_i) &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3, \end{aligned} \tag{4.2}$$

com valor de AIC igual a 803,455. Na Tabela 4.6 apresentamos para este modelo as estimativas por máxima verossimilhança com seus respectivos erros padrões.

Com o objetivo de verificar possíveis afastamentos das suposições feitas para o modelo, construímos os gráficos dos resíduos r_{p_i} e $r^*_{p_i}$, definidos em (3.4) e (3.5), respectivamente, contra o índice das observações. Na Figura 4.4 apresentamos os gráficos dos resíduos padronizados (Figura 4.4(a)) e dos resíduos ponderados (Figura 4.4(b)). Em relação a

estes gráficos podemos destacar a presença das observações 18, 57 e 66 como possíveis pontos aberrantes.

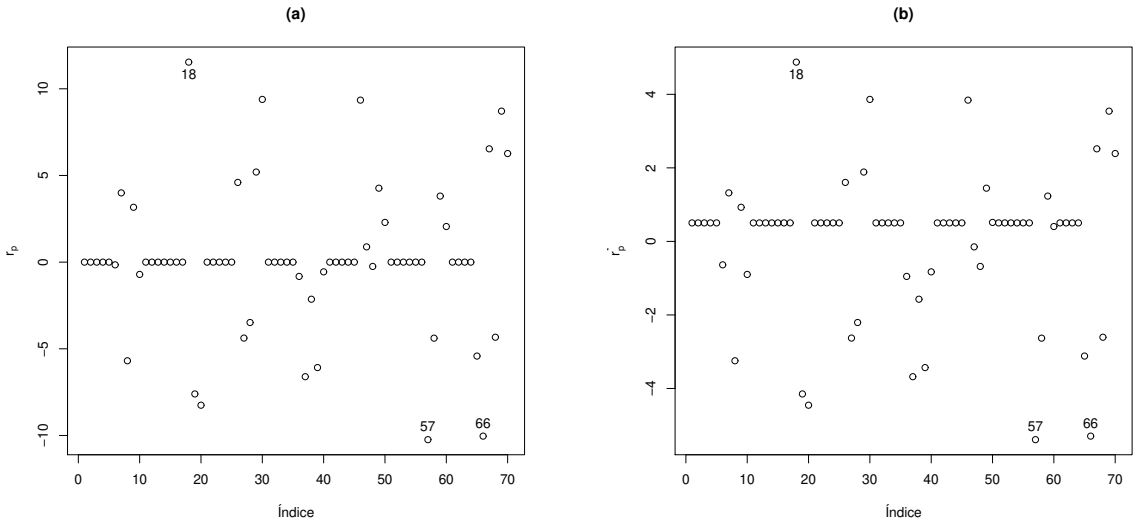


Figura 4.4: Gráficos dos resíduos padronizados (a) e ponderados (b) do modelo binomial inflacionado de zeros - Exemplo 1.

Com relação aos gráficos normais de probabilidades (Figura 4.5) notamos que há fortes indícios de afastamento da suposição de que o modelo de regressão binomial inflacionado de zeros seja o mais adequado para os dados, uma vez que, em geral, a maioria dos resíduos não pertencem às bandas de confiança dadas pelos envelopes simulados. Na Figura 4.7(a) apresentamos o envelope baseado nos resíduos de Pearson, que mostram a mesma tendência.

Tabela 4.6: Estimativas de máxima verossimilhança para o modelo (4.2) selecionado com seus erros padrões, segundo o modelo de regressão ZIB - Exemplo 1.

Parâmetro	Estimativa	Erro Padrão	z	Nível descritivo (p)
γ_0	0,114	0,239	0,478	0,633
β_0	2,365	0,248	9,542	$< 0,0001$
β_1	-3,940	0,264	-14,897	$< 0,0001$
β_2	1,255	0,075	16,629	$< 0,0001$
β_3	-0,106	0,006	-17,145	$< 0,0001$

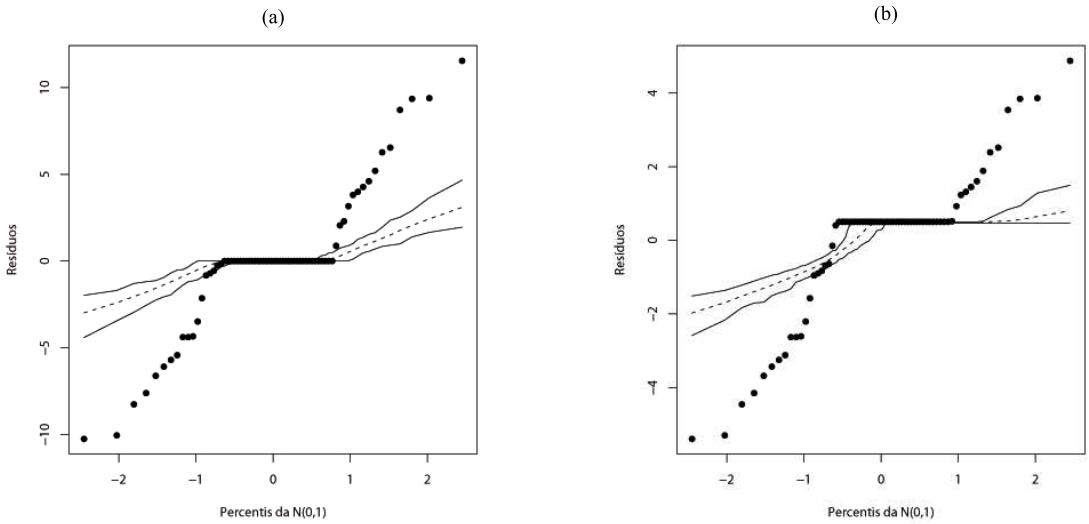


Figura 4.5: Gráficos normais de probabilidades com envelopes simulados para os resíduos padronizados (a) e ponderados (b) do modelo ZIB - Exemplo 1.

Utilizando o modelo beta-binomial inflacionado de zeros fazemos uma seleção do melhor modelo sendo finalmente selecionado o modelo (4.2). Os resultados do ajuste são apresentados na Tabela 4.7. Na Tabela 4.8 são apresentadas comparações das estimativas sob o modelo ZIB, o modelo ZIBB e o modelo de regressão binomial padrão. É claro que existe uma melhora no ajuste quando consideramos o modelo ZIBB.

Tabela 4.7: Estimativas de máxima verossimilhança para o modelo (4.2) selecionado com seus erros padrões, segundo o modelo de regressão ZIBB - Exemplo 1.

Parâmetro	Estimativa	Erro Padrão	z	Nível descritivo (p)
γ_0	0,107	0,240	0,447	0,655
β_0	2,687	1,523	1,764	0,078
β_1	-4,445	1,344	-3,307	$< 0,001$
β_2	1,408	0,349	4,032	$< 0,001$
β_3	-0,119	0,027	-4,341	$< 0,001$
ϕ	7,518	2,450	3,069	0,002

A Figura 4.6 mostra o gráfico de caixas dos resíduos de Pearson. A inspeção destes resíduos mostra que eles são tipicamente menores em magnitude para o modelo ZIBB.

Tabela 4.8: Resultados do AIC e a soma dos quadrados dos resíduos de Pearson nos modelos de regressão binomial padrão, ZIB e ZIBB - Exemplo 1.

Modelo	Binomial	ZIB	ZIBB
AIC	4375,200	803,455	398,131
g.l.	66	65	64
SQ_R_Pearson	3603,350	97,459	76,704
p	0	0,006	0,133

Além disso, nenhuma observação discrepante foi observada. Finalmente, temos os gráficos normais de probabilidades dos resíduos de Pearson (Figura 4.7), onde notamos que há indícios de que o modelo de regressão beta-binomial inflacionado de zeros seja adequado para os dados, uma vez que, em geral, os resíduos encontram-se entre os limites da banda de confiança (Figura 4.7(b)). Entretanto, na Figura 4.7(a), notamos que há indícios de afastamento da suposição de que o modelo de regressão binomial inflacionado de zeros seja adequado para os dados.

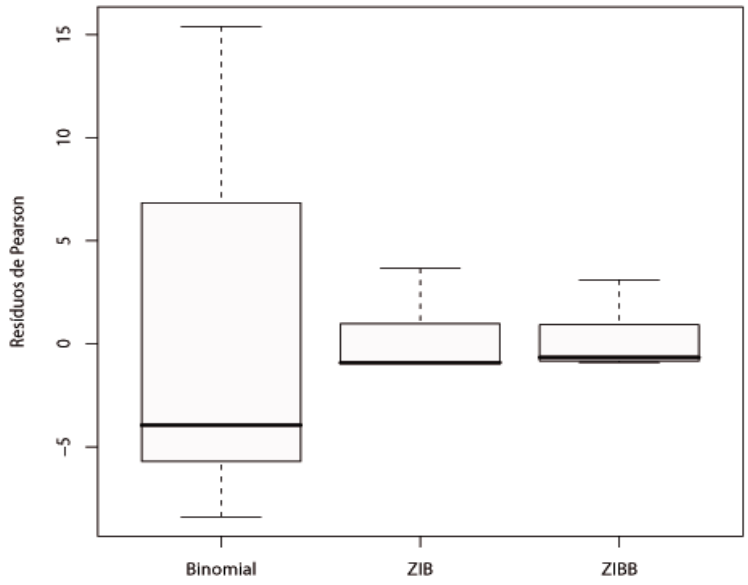


Figura 4.6: Dados de número de ovos parasitados. Gráficos de caixas dos resíduos de Pearson de acordo com diferentes modelos ajustados - Exemplo 1.

Com o intuito de verificar a existência de possíveis pontos influentes nos dados aplicamos as técnicas de influência global (afastamento da verossimilhança e distância de

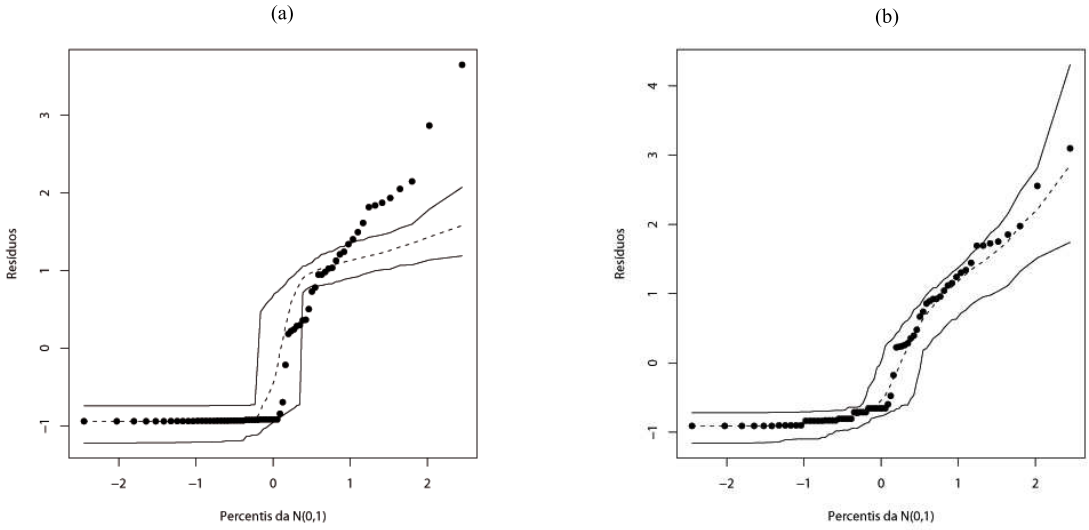


Figura 4.7: Gráficos normais de probabilidades com envelopes simulados para os resíduos de Pearson do modelo ZIB (a) e o modelo ZIBB (b) - Exemplo 1.

Cook) e influência local sob o esquema de perturbação da ponderação de casos (veja Seção 3.2). As observações influentes na estimação do vetor de parâmetros $\boldsymbol{\theta}$ detectadas pela distância de Cook (veja Figura 4.8(b)) e pelo esquema da ponderação de casos (veja Figura 4.8(c)), no modelo ZIBB são as observações 20 e 66. As observações influentes detectadas no modelo ZIBB (20 e 66) são retiradas uma a uma e conjuntamente, com a finalidade de verificar o impacto causado no ajuste do modelo. Para avaliar a magnitude do impacto exercido pelas observações utilizamos o desvio relativo percentual (DRP) dado por

$$DRP = \frac{\widehat{\boldsymbol{\theta}}^{*} - \widehat{\boldsymbol{\theta}}}{\widehat{\boldsymbol{\theta}}} \times 100\% \tag{4.3}$$

sendo $\widehat{\boldsymbol{\theta}}^{*}$ o estimador do parâmetro $\boldsymbol{\theta}$ encontrado ao retirar uma ou várias (conforme o caso) observações atípicas.

Na Tabela 4.9 encontram-se os valores das estimativas de máxima verossimilhança, os respectivos desvios padrões (entre parênteses) e o desvio relativo percentual DRP das estimativas ao retirar uma a uma e coletivamente as observações apontadas como possivelmente influentes. Note que a estimativa do parâmetro γ_0 é a que sofre o maior impacto em todos os casos avaliados; no entanto, o valor de $\widehat{p}$ permanece quase inalterado. Além disso, as maiores variações percentuais ocorrem para as estimativas de todos os parâmetros quando a observação 20 é retirada individualmente e quando as observações 20 e 66 são retiradas coletivamente.

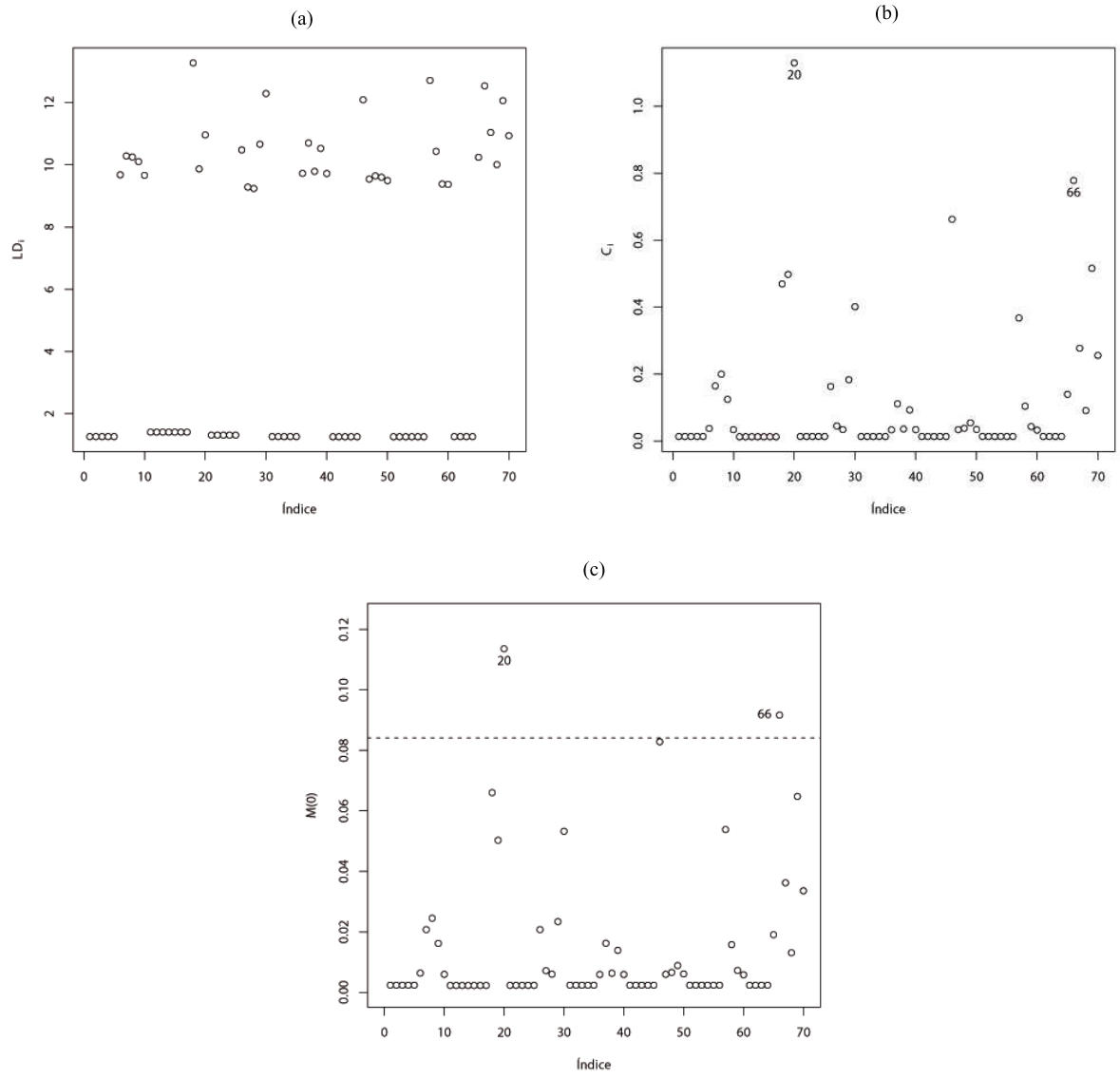


Figura 4.8: Análise de diagnóstico para o modelo ZIBB segundo o afastamento da verossimilhança (a), distância de Cook (b) e perturbação da ponderação de casos (c) - Exemplo 1.

Na Tabela 4.10 apresentamos uma comparação entre os três modelos (binomial padrão, ZIB e ZIBB) após a eliminação das observações influentes comuns a eles (20 e 66). Fazendo a comparação dos parâmetros comuns (β_0, β_1, β_2 e β_3) aos três modelos, temos que a eliminação das observações influentes 20 e 66 produz maiores mudanças nas estimativas do modelo Binomial padrão do que nas estimativas dos modelos ZIB e ZIBB (veja Tabela 4.10); ou seja, os modelos ZIB e ZIBB parecem produzir estimativas para os parâmetros mais robustas contra os pontos discrepantes do que o modelo binomial padrão; entretanto,

Tabela 4.9: Estimativas de máxima verossimilhança, erro padrão e DRP dos parâmetros do modelo ZIBB com a amostra completa e tirando as observações influentes - Exemplo 1.

Observações eliminadas	Estimativas					
	$\widehat{\gamma}_0^*$	$\widehat{\beta}_0^*$	$\widehat{\beta}_1^*$	$\widehat{\beta}_2^*$	$\widehat{\beta}_3^*$	$\widehat{\phi}^*$
Nenhuma	0,107	2,687	-4,445	1,408	-0,119	7,518
	(0,240)	(1,523)	(1,344)	(0,349)	(0,027)	(2,450)
20	0,143	2,359	-3,887	1,238	-0,105	8,490
	(0,242)	(1,380)	(1,238)	(0,324)	(0,026)	(2,860)
	33%	-12%	-13%	-12%	-12%	13%
66	0,140	2,574	-4,292	1,353	-0,113	8,175
	(0,243)	(1,385)	(1,230)	(0,322)	(0,025)	(2,690)
	30%	-4%	-3%	-4%	-5%	9%
20,66	0,176	2,243	-3,728	1,181	-0,099	9,426
	(0,246)	(1,388)	(1,245)	(0,327)	(0,026)	(3,069)
	64%	-17%	-16%	-16%	-17%	25%

comparando-se apenas os modelos ZIB e ZIBB notamos menores variações para o modelo ZIBB nas estimativas dos parâmetros.

Tabela 4.10: Mudanças (em %) nas estimativas dos parâmetros dos modelos ajustados depois de excluídas as observações 20 e 66 - Exemplo 1.

Parâmetro	Binomial	ZIB	ZIBB
γ_0		54,65	58,26
β_0	110,23	5,73	5,97
β_1	13,23	7,06	6,61
β_2	12,64	6,17	5,75
β_3	12,13	4,92	4,58
ϕ			19,73

4.2.2 Exemplo 2

Como parte de um estudo sociológico foram selecionados 316 alunos de duas escolas urbanas dos Estados Unidos. A variável resposta de interesse neste estudo é o número de dias de ausência na escola (y) e as covariáveis consideradas foram escola de procedência (E, dois níveis), sexo (dois níveis), notas em matemática (Mat) e notas em língua e artes (LA). Estatísticas da variável resposta encontram-se na Tabela 4.11 (veja também a Figura 4.9).

Tabela 4.11: Medidas resumo das variáveis número de dias de ausência, nota em Matemática e nota em Língua e artes - Exemplo 2.

Medida	Dias de ausência	Nota em Matemática	Nota em Língua e artes
Mínimo	0	1,007	1,007
Q_1	1	37,725	40,150
Mediana	3	48,679	50
Média	5,810	48,751	50,064
Q_3	8	61,044	61,044
Máximo	45	98.993	98.993
Desvio padrão	7,449	17,881	17,939

Na Figura 4.9 observa-se a existência de muitos zeros (dos alunos que nunca faltaram à escola). Além disso, nota-se na Figura 4.9 que o comportamento da variável resposta poderia ser representado adequadamente por uma distribuição de Poisson ou por uma distribuição binomial negativa com excesso de zeros.

Na Figura 4.10 observa-se uma comparação de medianas da variável resposta (número de dias de ausência) segundo categorias das variáveis explicativas (escola de procedência e sexo), onde parece que existem diferenças significativas nas medianas da variável resposta nas categorias da variável escola de procedência e não há para as categorias da variável sexo. Para determinar se a freqüência do excesso de zeros pode ser de fato responsável pela falta de ajuste do modelo Poisson padrão, recorremos ao teste escore para inflação de zeros desenvolvido na Seção 2.5.

Baseados na Figura 4.9 e considerando o modelo ZIP, foi calculado o valor de S usando a estatística de teste definida em (2.24) para testar a hipótese $H_0 : p = 0$ contra a alternativa $H_1 : p > 0$ a fim de determinar se existe indícios de excesso de zeros, dado que a proporção de zeros estimados pelo modelo ZIP é um critério importante para a qualidade do ajuste.

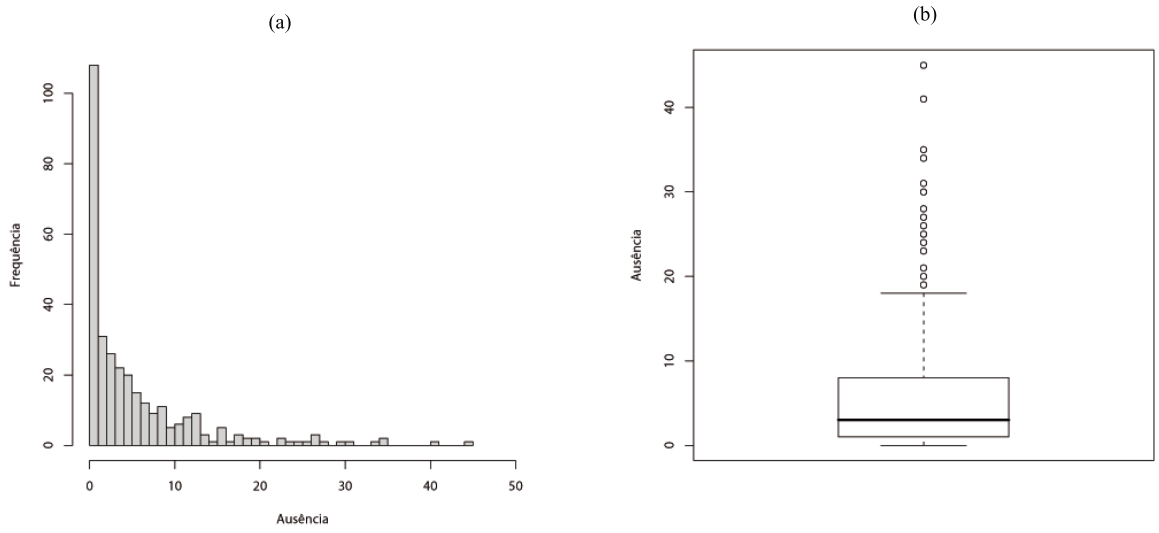


Figura 4.9: Dados de número de dias de ausência na escola. Histograma (a) e gráfico de caixas (b) do número de dias de ausência na escola - Exemplo 2.

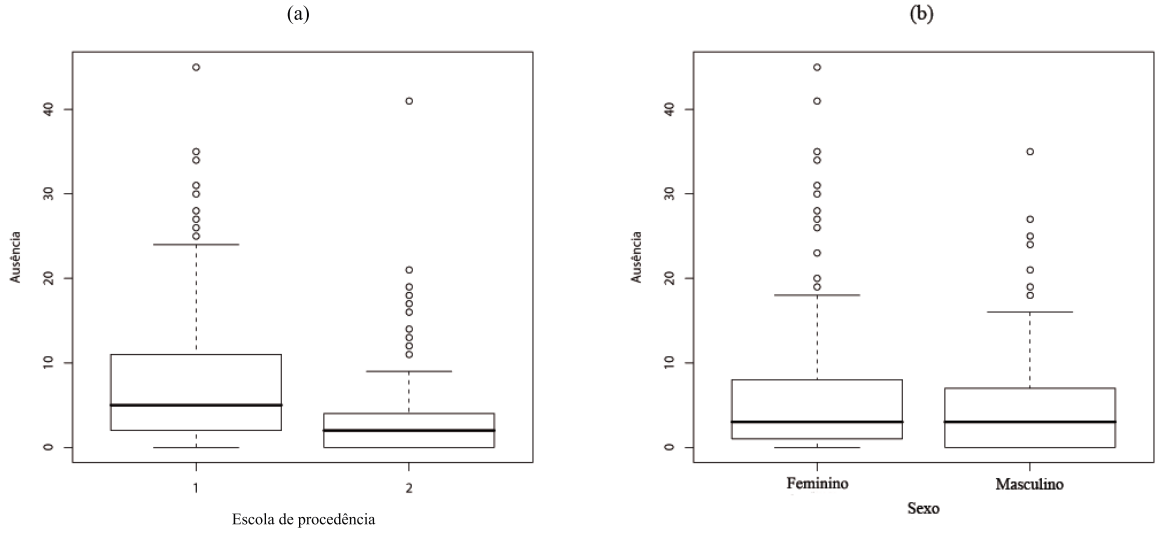


Figura 4.10: Gráfico de caixas do número de dias de ausência na escola segundo a escola de procedência e o sexo dos alunos - Exemplo 2.

Portanto, além de analisar os resíduos, a observação da proporção de zeros estimados pelo modelo ($\hat{p}$) e da proporção na amostra pode ser relevante na escolha do melhor modelo.

A estatística escore forneceu $S = 38,0347$ (com nível descritivo $< 0,0001$) de modo que rejeitamos a hipótese nula $H_0 : p = 0$. Pode-se concluir que o modelo Poisson Inflacionado

de Zeros (ZIP) é mais apropriado do que o modelo Poisson padrão.

Nesta aplicação, o modelo ZIP, apresentado na Seção 2.3.2, é dado por

$$log(\lambda_i) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 B_3 + \beta_4 B_4 \tag{4.4}$$

para a parte não inflacionada, e

$$logit(p_i) = \gamma_0 + \gamma_1 B_1 + \gamma_2 B_2 + \gamma_3 B_3 + \gamma_4 B_4 \tag{4.5}$$

para a parte inflacionada, em que B_1, B_2, B_3 e B_4 correspondem às covariáveis escola, nota em matemática, nota em língua e artes e sexo, respectivamente.

Para os parâmetros desse modelo, as estimativas de máxima verossimilhança foram obtidas usando o algoritmo EM (veja programa na Seção A.1.1 do Apêndice), cujos resultados são dados na Tabela 4.12.

Tabela 4.12: Dados de número de dias de ausência na escola. Estimativas de máxima verossimilhança com seus respectivos erros padrões usando o modelo ZIP - Exemplo2.

Parâmetro	Estimativa	Erro Padrão	z	Nível descritivo (p)
γ_0	-4,541	0,736	-6,171	$< 0,0001$
γ_1	0,816	0,337	2,423	0,015
γ_2	0,014	0,012	1,185	0,236
γ_3	0,011	0,012	0,956	0,339
γ_4	0,948	0,318	2,985	0,003
β_0	3,107	0,087	35,625	$< 0,0001$
β_1	-0,671	0,057	-11,800	$< 0,0001$
β_2	0,003	0,002	1,323	0,186
β_3	-0,004	0,002	-2,211	0,027
β_4	-0,276	0,049	-5,661	$< 0,0001$

Selecionamos o modelo que melhor se ajusta aos dados (modelo mais parcimonioso) utilizando o método passo a passo. O modelo finalmente selecionado é dado por

$$\begin{aligned} logit(p_i) &= \gamma_0 + \gamma_1 B_1 + \gamma_2 B_2 + \gamma_4 B_4, \\ log(\lambda_i) &= \beta_0 + \beta_1 B_1 + \beta_3 B_3 + \beta_4 B_4, \end{aligned} \tag{4.6}$$

com valor de AIC igual a 2556,651. Na Tabela 4.13 apresentamos para este modelo as estimativas por máxima verossimilhança com seus respectivos erros padrões.

Tabela 4.13: Estimativas de máxima verossimilhança para o modelo (4.6) selecionado com seus erros padrões, segundo o modelo de regressão ZIP - Exemplo 2.

Parâmetro	Estimativa	Erro Padrão	z	Nível descritivo (p)
γ_0	-4,354	0,704	-6,187	$< 0,0001$
γ_1	0,875	0,330	2,651	0,008
γ_2	0,021	0,009	2,198	0,028
γ_4	0,896	0,312	2,871	0,004
β_0	3,124	0.086	36,261	$< 0,0001$
β_1	-0,664	0,057	-11,721	$< 0,0001$
β_3	-0,003	0,001	-1,819	0,069
β_4	-0,271	0,049	-5,581	$< 0,0001$

Com relação aos gráficos normais de probabilidades (Figura 4.11) notamos que há fortes indícios de afastamento da suposição de que o modelo de regressão Poisson inflacionado de zeros seja adequado para estes dados, uma vez que a maioria dos resíduos não pertence à banda de confiança.

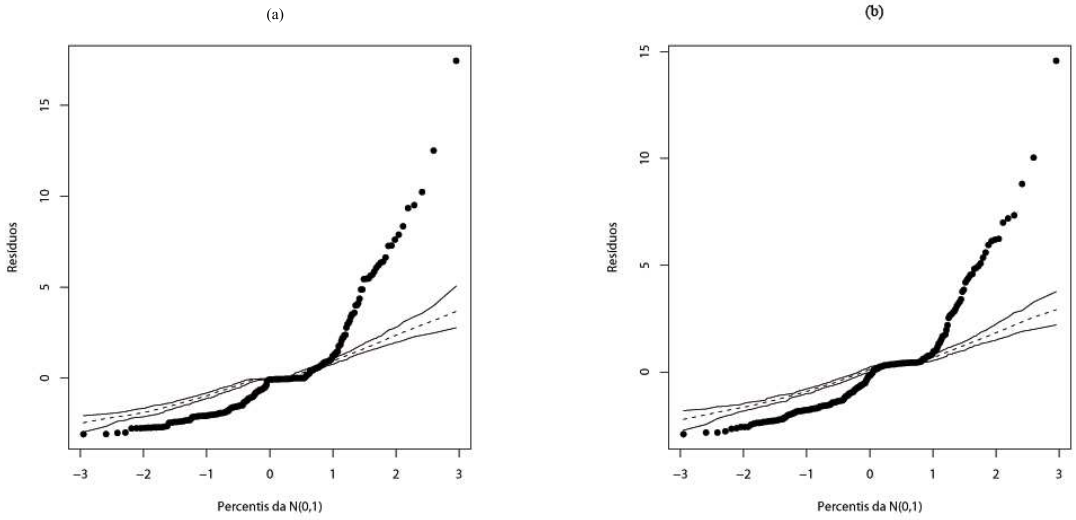


Figura 4.11: Gráficos normais de probabilidades com envelopes simulados para os resíduos padronizados (a) e ponderados (b) do modelo ZIP - Exemplo 2.

Com o objetivo de verificar possíveis pontos aberrantes para o modelo construímos os gráficos dos resíduos r_{p_i} e $r^*_{p_i}$, definidos em (3.4) e (3.5), respectivamente, contra os

índices das observações. Na Figura 4.12 apresentamos os gráficos dos resíduos padronizados (Figura 4.12(a)) e dos resíduos ponderados (Figura 4.12(b)). Em relação a estes gráficos podemos destacar a presença da observação 308 como possível ponto aberrante.

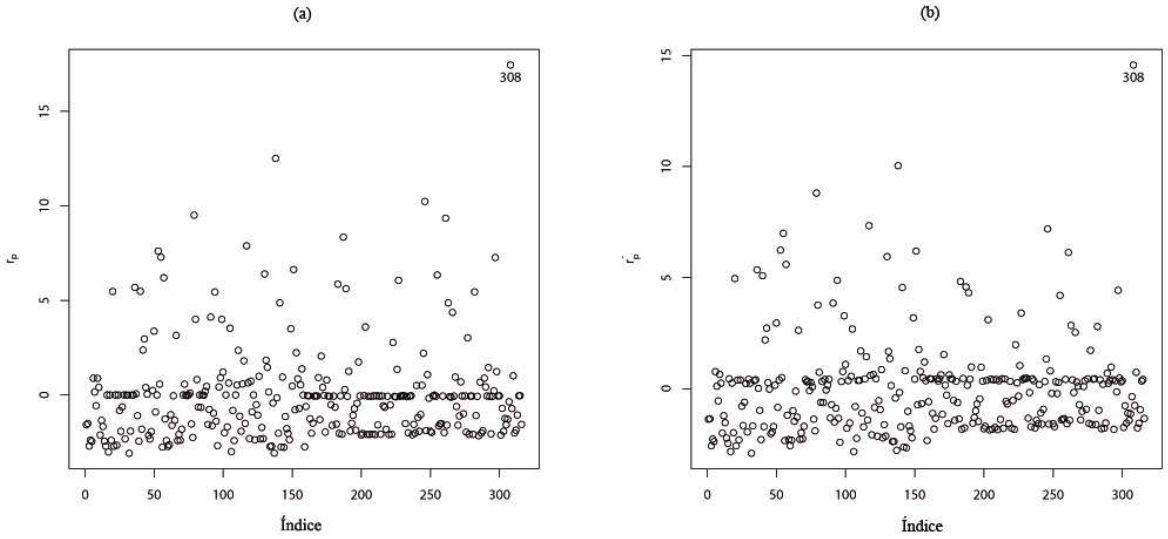


Figura 4.12: Gráficos de resíduos padronizados (a) e ponderados (b) do modelo Poisson inflacionado de zeros - Exemplo 2.

Utilizando o modelo binomial negativa inflacionado de zeros (ZINB) fazemos uma seleção do melhor modelo utilizando o método passo a passo. O modelo finalmente selecionado foi

$$\begin{aligned} logit(p_i) &= \gamma_2 B_2, \\ log(\mu_i) &= \beta_0 + \beta_1 B_1 + \beta_3 B_3 + \beta_4 B_4, \end{aligned} \tag{4.7}$$

com valor de AIC igual a 1746,248. Na Tabela 4.14 apresentamos as estimativas por máxima verossimilhança com seus respectivos erros padrões.

Na Tabela 4.15 é apresentado um resumo das estimativas para o modelo ZIP e para o modelo ZINB. Com o intuito de comparação são também mostrados os resultados da regressão Poisson padrão. É claro que um melhor ajuste é obtido com o modelo binomial negativa inflacionado de zeros. Esta observação pode ser corroborada pelos gráficos normais de probabilidades usando os resíduos de Pearson (Figura 4.13), onde notamos que não há indícios de afastamento da suposição de que o modelo de regressão binomial negativa inflacionado de zeros seja adequado para os dados.

Finalmente, a Figura 4.14 mostra o gráfico de caixas dos resíduos de Pearson para cada um dos três modelos. A inspeção destes resíduos mostra que eles são freqüentemente

menores em magnitude para o modelo ZINB.

Aplicamos agora as técnicas de influência global (afastamento da verossimilhança e distância de Cook) e influência local sob o esquema de perturbação da ponderação de casos (veja seções 3.2.1 e 3.2, respectivamente) objetivando verificar a existência de possíveis pontos influentes nos dados.

No modelo ZINB, a observação 308 foi detectada sendo influente na estimação do vetor de parâmetros $\boldsymbol{\theta}$ pelo afastamento da verossimilhança (veja Figura 4.15(a)), pela distância de Cook (veja Figura 4.15(b)) e no esquema da perturbação da ponderação de casos (veja Figura 4.15(c)). A observação influente (308) foi retirada com a finalidade de verificar o impacto causado no ajuste do modelo. Para avaliar a magnitude do impacto exercido pela observação influente utilizamos o Desvio Relativo Percentual (DRP) dado pela expressão (4.3). Na Tabela 4.16 encontram-se os valores das estimativas de máxima verossimilhança, os respectivos desvios padrões (entre parênteses) e o desvio relativo percentual DRP dos

Tabela 4.14: Estimativas de máxima verossimilhança para o modelo (4.7) selecionado com seus erros padrões, segundo o modelo de regressão ZINB - Exemplo 2.

Parâmetro	Estimativa	Erro Padrão	z	Nível descritivo (p)
γ_2	-2,062	2,613	-0,789	0,430
β_0	3,384	0,029	115,735	$< 0,0001$
β_1	-0,762	0,019	-39,848	$< 0,0001$
β_3	-0,008	0,001	-17,021	$< 0,0001$
β_4	-0,392	0,017	-23,629	$< 0,0001$
ϕ	0,873	0,089	9,838	$< 0,0001$

Tabela 4.15: Resultados do AIC e a soma dos quadrados dos resíduos de Pearson nos modelos de regressão Poisson padrão, ZIP e ZINB - Exemplo 2.

Modelo	Poisson	ZIP	ZINB
AIC	2879,692	2556,651	1746,248
g.l.	312	308	310
SQ_R_Pearson	2494,001	1246,923	368,151
p	$< 0,0001$	$< 0,0001$	0,013

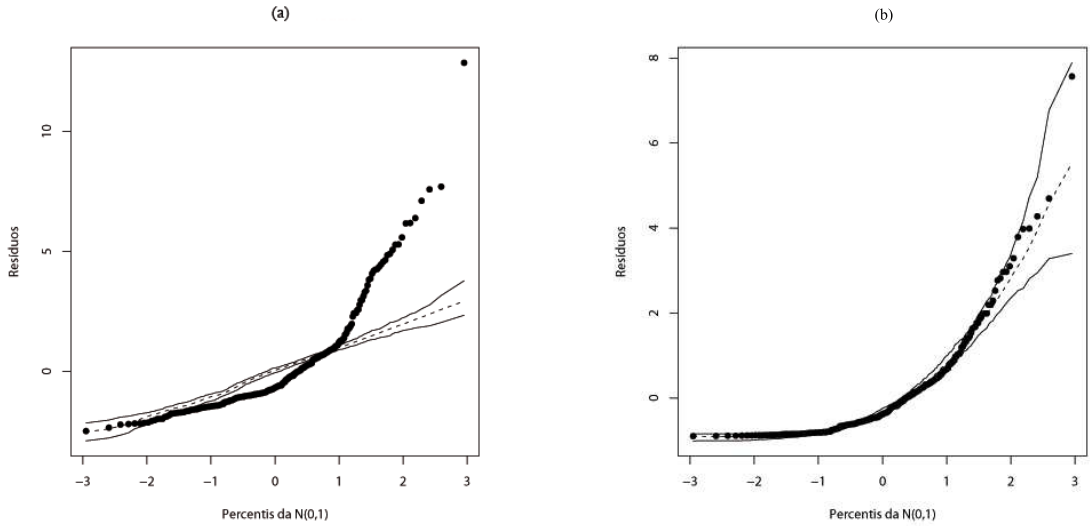


Figura 4.13: Gráficos normais de probabilidades com envelopes simulados para os resíduos de Pearson do modelo ZIP (a) e o modelo ZINB (b) - Exemplo 2.

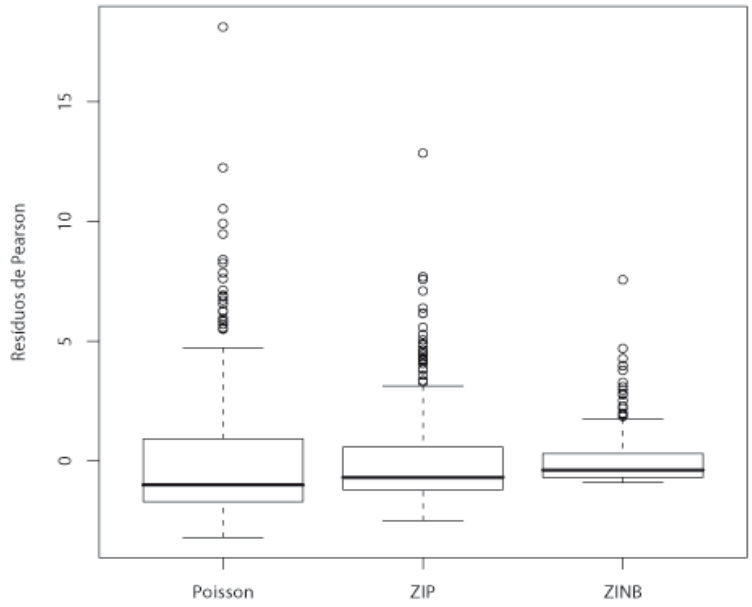


Figura 4.14: Dados de número de dias de ausência na escola. Gráficos de caixas dos resíduos de Pearson de acordo com diferentes modelos ajustados - Exemplo 2.

parâmetros estimados ao retirar a observação 308 considerada como influente. Note que as estimativas dos parâmetros não sofreram grande impacto. Como podemos perceber, a maior variação percentual ocorre para a estimativa de β_4.

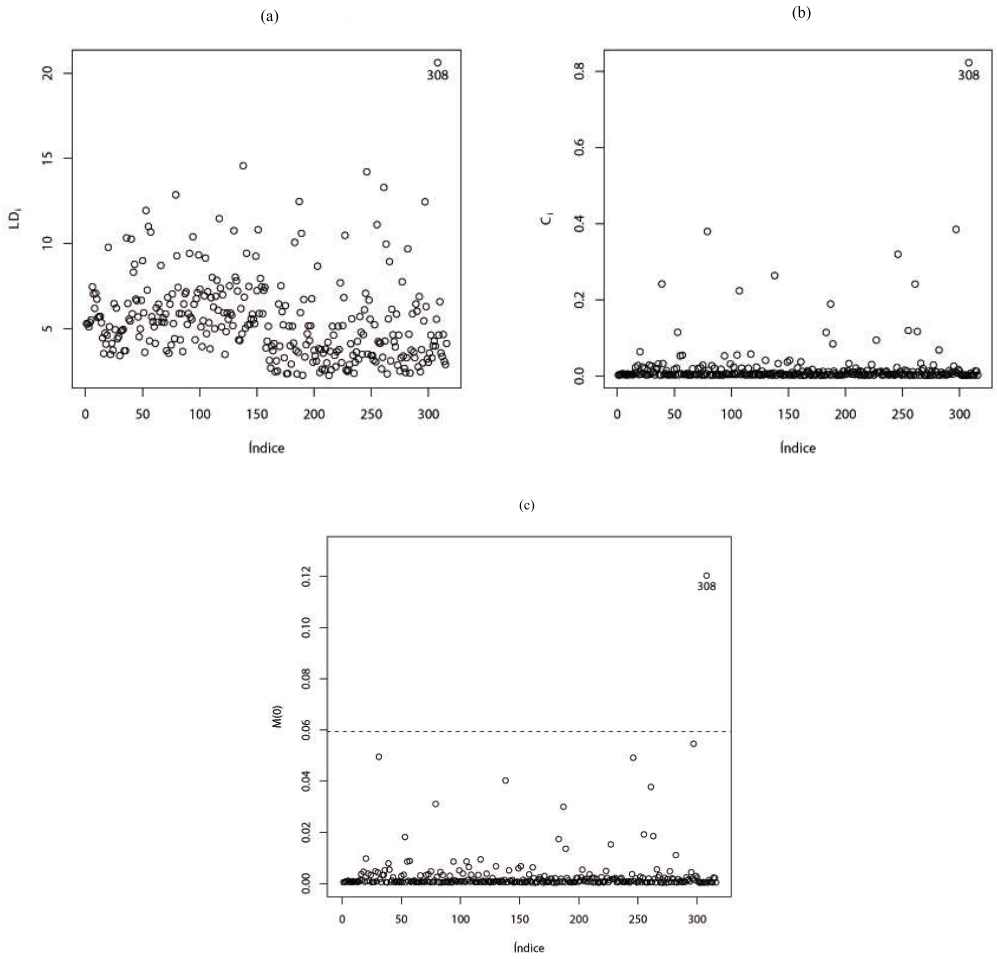


Figura 4.15: Análise de diagnóstico para o modelo ZINB segundo o afastamento da verossimilhança (a), distância de Cook (b) e perturbação da ponderação de casos (c) - Exemplo 2.

Tabela 4.16: Estimativas de máxima verossimilhança, erro padrão e DRP dos parâmetros do modelo ZINB com a amostra completa e tirando a observação influente - Exemplo 2.

Observações eliminadas	Estimativas					
	$\widehat{\gamma}_2^*$	$\widehat{\beta}_0^*$	$\widehat{\beta}_1^*$	$\widehat{\beta}_3^*$	$\widehat{\beta}_4^*$	$\widehat{\phi}^*$
Nenhuma	-2,062	3,384	-0,762	-0,008	-0,392	0,872
	(2,613)	(0,029)	(0,019)	(0,001)	(0,017)	(0,089)
308	-1,953	3,402	-0,831	-0,007	-0,334	0,911
	(2,371)	(0,029)	(0,021)	(0,001)	(0,016)	(0,096)
	-5%	1%	9%	-7%	-15%	4%

Capítulo 5

Conclusões

5.1 Considerações finais

Neste trabalho foram estudados quatro modelos para dados de contagem com excesso de zeros. Neste sentido, propomos modelos de regressão em que a distribuição da variável resposta é uma mistura entre uma distribuição discreta (binomial, Poisson, binomial negativa ou beta-binomial) e uma distribuição de Bernoulli degenerada em zero. Os parâmetros que modelam a probabilidade de ocorrer zero são representados em uma estrutura de regressão. Para cada modelo apresentado propusemos a modelagem dos parâmetros através de preditores lineares usando funções de ligação adequadas. Desenvolvemos estimação por máxima verossimilhança, bem como a obtenção de expressões matriciais para o vetor escore e a matriz de informação de Fisher, implementamos o algoritmo EM para a estimativa dos parâmetros de regressão dos quatro modelos usados, efetuamos testes de hipóteses e seleção de modelos, entre outros tópicos. Para validar as suposições e identificar observações influentes dos modelos de regressão inflacionados de zeros, desenvolvemos métodos de diagnóstico. Para isto, criamos resíduos padronizados e ponderados. Adicionalmente, desenvolvemos medidas de influência local baseadas na curvatura normal conforme e derivamos expressões matriciais sob o esquema de perturbação de ponderação de casos. Em dois exemplos com dados reais ilustramos a potencialidade da metodologia de influência local no sentido de detectar as observações que podem afetar os resultados inferenciais obtidos para o modelo ajustado aos dados. Finalmente, sugerimos aos usuários interessados em modelar dados na forma de contagens que utilizem modelos de regressão inflacionados de

zeros sempre que muitos zeros estejam presentes no conjunto de dados.

5.2 Trabalhos futuros

Como foco de trabalhos futuros sugerimos os seguintes temas de pesquisa:

1. Elaborar os gráficos de envelope simulado para os resíduos dos modelos binomial negativa e beta-binomial inflacionados de zeros utilizando a matriz de informação observada.

2. Estudar métodos de estimação restrita para os modelos de regressão inflacionados de zeros.

3. Analisar e implementar a correção por viés para os modelos de regressão inflacionados de zeros.

4. Estender os resultados desta dissertação para os modelos de regressão inflacionados de zeros com efeito aleatório.

Apêndice A

Apêndices

A.1 Programas implementados no software R

A.1.1 Algoritmo EM para os modelos Inflacionados de Zeros

Modelo ZIB

```
EM.ZIB<-function(bb,ni,B,G){
  n <- nrow(B)
  aa <- matrix(0,n,1)
  aa[bb>0] <- 1
  beta <- as.vector(glm(cbind(bb,ni-bb)~B-1,family=binomial)$coeff)
  gama <- as.vector(glm(cbind(aa,1-aa)~G-1,family=binomial)$coeff)
  teta <- c(beta,gama)
  criterio <- sqrt(teta%*%teta)
  cont <- 0
  while(criterio>0.00000001){
    cont <- cont+1
    z <- (1+exp(-G%*%gama)*(1+exp(B%*%beta))^(-ni))^(-1)
    z[bb>0] <- 0
    beta <- as.vector(glm(cbind(bb,ni-bb)~B-1,family=binomial,
            weights=1-z)$coeff)
    gama <- as.vector(glm(cbind(z,1-z)~G-1,family=binomial)$coeff)
    param <- teta
    teta <- c(beta,gama)
```

```
    criterio <- sqrt((teta-param)%*%(teta-param))
    print(cont)
    }
  return(teta)
  }
```

Modelo ZIP

```
EM.ZIP<-function(bb,B,G){
  n <- nrow(bb)
  aa <- matrix(0,n,1)
  aa[bb>0] <- 1
  beta <- as.vector(glm(bb~B-1,family=poisson)$coeff)
  gama <- as.vector(glm(cbind(aa,1-aa)~G-1,family=binomial)$coeff)
  teta <- c(beta,gama)
  criterio <- sqrt(teta%*%teta)
  cont <- 0
  while(criterio>0.00000001){
    cont <- cont+1
    z <- (1+exp(-G%*%gama-exp(B%*%beta)))^(-1)
    z[bb>0] <- 0
    beta <- as.vector(glm(bb~B-1,family=poisson,weights=1-z)$coeff)
    gama <- as.vector(glm(z~G-1,family=binomial)$coeff)
    param <- teta
    teta <- c(beta,gama)
    criterio <- sqrt((teta-param)%*%(teta-param))
    print(cont)
    }
  return(teta)
  }
```

Modelo ZINB

```
library(MASS)
EM.ZINB <- function(bb,B,G){
  n <- nrow(B)
```

```
  aa <- matrix(0,n,1)
  aa[bb>0] <- 1
  beta <- as.vector(glm.nb(bb~B-1)$coeff)
  gama <- as.vector(glm(cbind(aa,1-aa)~G-1,family=binomial)$coeff)
  phi <- glm.nb(bb~B-1)$theta
  teta <- c(beta,gama,phi)
  criterio <- sqrt(teta%*%teta)
  cont <- 0
  while(criterio>0.00000001){
    cont <- cont+1
    z <- (1+exp(-G%*%gama)*(phi/as.real(exp(B%*%beta)+phi))^(phi))^(-1)
    z[bb>0] <- 0
    auxpar <- glm.nb(bb~B-1,weights=1-z)
    beta <- as.vector(auxpar$coeff)
    gama <- as.vector(glm(cbind(z,1-z)~G-1,family=binomial)$coeff)
    phi <- auxpar$theta
    param <- teta
    teta <- c(beta,gama,phi)
    criterio <- sqrt((teta-param)%*%(teta-param))
    print(cont)
    }
  return(teta)
  }
```

Modelo ZIBB

```
maxLOGbeta<-function(theta,bb,ni,z,B){
  p <- ncol(B)
  q <- ncol(G)
  k <- p+q+1
  n <- nrow(B)
  sum <- 0
  beta <- theta[1:p]
  phi <- theta[p+1]
  aux1 <- lchoose(ni,bb)+lgamma(bb+phi*exp(B%*%beta)/(1+exp(B%*%beta)))
```

```
  aux2 <- lgamma(ni+phi/(1+exp(B%*%beta))-bb)+lgamma(phi)-lgamma(ni+phi)
  aux3 <- -lgamma(phi*exp(B%*%beta)/(1+exp(B%*%beta)))
          -lgamma(phi/(1+exp(B%*%beta)))
  aux <- -sum((1-z)*(aux1+aux2+aux3))
  return(aux)
  }

EM.ZIBB<-function(bb,ni,B,G){
  n <- nrow(B)
  aa <- matrix(0,n,1)
  aa[bb>0] <- 1
  beta <- as.vector(glm(cbind(bb,ni-bb)~B-1,family=binomial)$coeff)
  gama <- as.vector(glm(cbind(aa,1-aa)~G-1,family=binomial)$coeff)
  phi <- 0.5
  p <- ncol(B)
  q <- ncol(G)
  k <- p+q+1
  theta <- c(beta,phi)
  teta <- c(beta,gama,phi)
  criterio <- sqrt(teta%*%teta)
  cont <- 0
  datos <- data.frame(B,ni,bb)
  while(criterio>0.00000001){
    cont <- cont+1
    aux <- exp(-G%*%gama)*gamma(ni+phi/(1+exp(B%*%beta)))*gamma(phi)
    z <- (1+aux/(gamma(ni+phi)*gamma(phi/(1+exp(B%*%beta)))))^(-1)
    z[bb>0] <- 0
    beta.phi<-optim(theta,maxLOGbeta,gr=NULL,method="L-BFGS-B",
              lower=c(rep(-10,p),0.1),upper=c(rep(10,p),100),
              control=list(maxit=2000),hessian=FALSE,bb,ni,z,B)$par
    beta<-c(beta.phi[1:p])
    phi<-beta.phi[p+1]
    gama<-as.vector(glm(z~G-1,family=binomial)$coeff)
    param <- teta
```

```
    teta <- c(beta,gama,phi)
    criterio <- sqrt((teta-param)%*%(teta-param))
    print(cont)
    }
  return(teta)
  }
```

A.1.2 Programas geradores de dados inflacionados de zeros

Programa para gerar dados ZIB

```
Gera.amostraZIB<-function(ni,B,G,beta,gama){
  n<-nrow(B)
  amostra<-matrix(0,n,1)
  for(i in 1:n){
    pi<-exp(B[i,]%*%beta)/(1+exp(B[i,]%*%beta))
    p<-exp(G[i,]%*%gama)/(1+exp(G[i,]%*%gama))
    amostra[i]<-if(rbinom(1,1,p)==1) 0 else rbinom(1,ni[i],pi)
  }
  return(amostra)
}
```

Programa para gerar dados ZIP

```
Gera.amostraZIP<-function(B,G,beta,gama){
  n<-nrow(B)
  amostra<-matrix(0,n,1)
  for(i in 1:n){
    lambda<-exp(B[i,]%*%beta)
    p<-exp(G[i,]%*%gama)/(1+exp(G[i,]%*%gama))
    amostra[i]<-if(rbinom(1,1,p)==1) 0 else rpois(1,lambda)
  }
  return(amostra)
}
```

Programa para gerar dados ZINB

```
Gera.amostraZINB<-function(B,G,beta,gama,phi){
  n<-nrow(B)
  amostra<-matrix(0,n,1)
  for(i in 1:n){
    mu<-exp(B[i,]%*%beta)
    p<-exp(G[i,]%*%gama)/(1+exp(G[i,]%*%gama))
    amostra[i]<-if(rbinom(1,1,p)==1) 0 else rnbinom(1,size=phi,mu=mu)
  }
  return(amostra)
}
```

Programa para gerar dados ZIBB

```
library(emdbook)
Gera.amostraZIBB<-function(ni,B,G,beta,gama,phi){
  n<-nrow(B)
  amostra<-matrix(0,n,1)
  for(i in 1:n){
    pi<-exp(B[i,]%*%beta)/(1+exp(B[i,]%*%beta))
    p<-exp(G[i,]%*%gama)/(1+exp(G[i,]%*%gama))
    amostra[i]<-if(rbinom(1,1,p)==1) 0 else rbetabinom(1,pi,size=ni[i],
                theta=phi)
  }
  return(amostra)
}
```

A.1.3 Programas geradores de envelopes para modelos de regressão inflacionados de zeros

Programa para gerar envelope para o modelo ZIB

```
n <- nrow(B)                    # bb: vetor de dados.\\
p1<-ncol(B)                   # B, G: matrizes de planejamento
q1<-ncol(G) # B modela a media e G o parametro zero inflado
theta<-EM.ZIB(bb,ni,B,G) # beta, Gama: parametros de regressao
```

```
residuos<-r.ZIB(bb,ni,B,G,theta) # Calculo dos residuos.
e<-matrix(0,n,100) beta<-theta[1:p1]; gama<-theta[(p1+1):(p1+q1)]
for(i in 1:100){
    nresp <- Est.amostraZIB(ni,B,G,beta,gama)
    tethasim<-EM.ZIB(nresp,ni,B,G)
    fit <- r.ZIB(nresp,ni,B,G,tethasim)
    e[,i] <- sort(fit)}
e1 <- numeric(n)
e2 <- numeric(n) for(i in 1:n){
    eo <- sort(e[i,])
    e1[i] <- (eo[2]+eo[3])/2
    e2[i] <- (eo[97]+eo[98])/2}
med<-apply(e,1,mean); faixa <- range(residuos,e1,e2); par(pty="s")
qqnorm(residuos,main="Residuo ponderado",xlab="Percentis da
        N(0,1)",sub="(a)", ylab="Residuos", ylim=faixa, pch=16)
par(new=T)
qqnorm(e1,main="",axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(e2,main="",axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1)
par(new=T)
qqnorm(med,main="",axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=2)
```

Referências Bibliográficas

[1] Altham, P.M. (2002). *An aspect of discrete data analysis: fitting a beta-binomial distribution to the hospitals data*. Centre for the Mathematical Sciences, Cambridge. http://www.statslab.cam.ac.uk/∼pat

[2] Atkinson, A.C. (1985). *Plots, Transformations and Regressions*. Oxford University Press, Londres.

[3] Atkinson, A.C. (1987). *Plots, transformations and regression: an introduction to graphical methods of diagnostics regression analysis*, 2nd ed. Oxford science publications, Oxford.

[4] Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L. & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Statist. Soc.*, **A, 162**, 195-209.

[5] Borgatto, A. (2004). *Modelos para proporções com superdispersão e excesso de zeros - um procedimento bayesiano*. Tese de doutorado. ESALQ-USP, Piracicaba, São Paulo.

[6] Broek, J.V. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**, 738-743.

[7] Casella, G. & Berger, R.L. (2001). *Statistical Inference*, 2nd ed. Duxbury, California.

[8] Cameron, A.C. & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

[9] Cheung, Y.B. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine*, **21**, 1461-1469.

[10] Christensen, R., Pearson, L. & Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, **34**, 38-45.

[11] Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 5-18.

[12] Cook, R.D. (1986). Assessment of local influence (with discussion). *J. R. Statist. Soc.*, **B, 48**, 133-169.

[13] Cook, R.D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, Nova York.

[14] Cordeiro, G.M. & Demétrio C.G.B. (2007). Modelos Lineares Generalizados. Minicurso para o 12^o SEAGRO e a 52^a Reunião Anual da RBRAS, UFSM, Santa María.

[15] Cox, D. & Snell, E. (1968). A general definition of residuals. *J. R. Statist. Soc.*, **B, 30**, 248-275.

[16] Demétrio C.G.B. (2002). *Modelos Lineares Generalizados em Experimentação Agronômica*. ESALQ/USP, Piracicaba.

[17] Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Statist. Soc.*, **B, 39**, 1, 1-38.

[18] Deng, D. & Paul, S.R. (2000). Score tests for zero-inflation in generalized linear models. *The Canadian Journal of Statistics*, **27**, 3, 563-570.

[19] Deng, D. & Paul, S.R. (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statistica Sinica*. **15**, 257-276.

[20] Dobson, A.J. (2001). *An Introduction to Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC, Londres.

[21] Famoye, F. & Singh, K.P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, **4**, 117-130.

[22] Ferrari, S. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**, 7, 799-815.

[23] Galea-Rojas, M., Paula, G.A. & Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71-79.

[24] Hall, D. (2000). Zero-inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, **56**, 4, 1030-1039.

[25] Jansakul, N. (2001). *Some aspects of modelling overdispersed and zero-inflated count data*. Thesis (Ph.D.) - University of Exeter.

[26] Jansakul, N. & Hinde, J.P. (2002). Score Tests for Zero-Inflated Poisson Models *Computational Statistics & Data Analysis*, **40**, 75-96.

[27] Johnson, N.L. & Kotz, S. (1969). *Discrete Distributions: Distributions in Statistics*. John Wiley & Sons, Nova York.

[28] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.

[29] Lee, A., Wang, K. & Yau, K. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal*, **43**, 963-975.

[30] Lee, A., Xiang, L. & Fung, W. (2004). Sensitivity of score tests for zero-inflation in count data. *Statistics in Medicine*, **23**, 2757-2769.

[31] Lee, S. & Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, **45**, 321-341.

[32] Lewsey, J.D. & Thomson, W.M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dent Oral Epidemiol*, **32**, 183-189.

[33] Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications. NSF-CMBS Regional Conference Series in Probability and Statistics*. Vol. 5. Institute of Mathematical Statistics, Hayward, California.

[34] Miaou, S.P. (1994). The relationship between truck accidents and geometric design of road sections. Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, **26**, 471-482.

[35] Navarro, A. (2001). Negative binomial distribution versus Poisson in the analysis of recurrent phenomena. *Gac Sanit*, **15**, 5:447-452.

[36] Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Statist. Soc.*, **A, 135**, 3, 370-384.

[37] Ospina, R. (2008). *Modelos de regressão beta inflacionados*. Tese de doutorado. IME-USP, São Paulo.

[38] Paula, G.A. (2004). *Modelos de regressão com Apoio Computacional*. IME-USP, São Paulo.

[39] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[40] Ridout, M., Hinde, J. & Demétrio C.G.B. (2001). A score test for testing zero inflated Poisson regression model against zero inflated negative binomial alternatives. *Biometrics*, **57**, 219-223.

[41] Ridout, M., Demétrio C.G.B. & Hinde, J. (1998). Models for count data with many zeros. International Biometric Conference, Cape Town.

[42] Skellam, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *J. R. Statist. Soc.*, **B, 10**, 257-261.

[43] Slaton, T., Piegorsch, W. & Durham S. (2000). Estimation and Testing with Overdispersed Proportions Using the Beta-Logistic Regression Model of Heckman and Willis. *Biometrics*, **56**, 125-133.

[44] Tsai, C. & Wu, X. (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics and Probability Letters*, **14**, 247-252.

[45] Xiang, L., Yau, K., Lee, A. & Fung, W. (2005). Influence diagnostics for two-component Poisson mixture regression models: applications in public health. *Statistics in Medicine*, **24**, 3053-3071.

[46] Xie, M., He, B. & Goh, T.N. (2001). Zero-inflated Poisson model in statistical process control. *Computational Statistics & Data Analysis*, **38**, 191-201.

[47] Xie, F., Wei, B. & Lin, J. (2008). Assessing influence for pharmaceutical data in zero-inflated generalized Poisson mixed models. *Statistics in Medicine*, **27**, 3656-3673.

[48] Yau, K., Wang, K. & Lee, A. (2003). Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biometrical Journal*, **45**, 4, 437-452.

[49] Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *J. R. Statist. Soc.*, **B, 63**, 111-126.