Modelos para dados binários longitudinais com superdispersão

Afrânio M. C. Vieira[1], Clarice G. B. Demétrio[2], Geert Molenberghs[3] and Geert Verbeke[4]

[1] Departamento de Estatística, Universidade de Brasília, Brasília, DF
[2] Departamento de Ciências Exatas, ESALQ - Universidade de São Paulo, Piracicaba, SP
[3] I-BioStat, Universiteit Hasselt, Diepenbeek, Bélgica
[4] I-BioStat, Katholieke Universiteit Leuven, Leuven, Bélgica

1 Introdução

Resultados experimentais e observacionais medidos na forma binária (seja presença ou ausência de alguma característica em uma unidade observacional) são bastante recorrentes em diversas áreas da Ciência e os modelos lineares generalizados (GLM) trouxeram unidade e flexibilidade para a análise deste tipo de dados. A generalidade dessa classe de modelos, sua simplicidade conceitual, sofisticação e eficiência computacional tornaram-na referência metodológica para a análise de dados binários no final do século 20.

Para medidas binárias, é natural assumir a distribuição Bernoulli como referência probabilística na construção de modelos. Entretanto, a relação limitada entre a esperança e a variância dessa distribuição e a ausência de um parâmetro para quantificar a variabilidade extra-Bernoulli, torna necessária a utilização de modelos específicos para acomodar o efeito de sobredispersão. Um desses modelos acomoda a correlação entre medidas binárias de um mesmo elemento[2]. Se as diversas medidas feitas sobre um elemento forem longitudinais, a classe de modelos lineares generalizados mistos[1] (GLMM) pode ser utilizada, na qual a estrutura de dependência temporal é acomodada adicionando-se efeitos aleatórios ao preditor linear.

Entretanto, a superdispersão e a estrutura longitudinal são, usualmente, modeladas separadamente. Situações clínicas e agronômicas para as quais diversos elementos sob agrupamentos não-especificados são medidos longitudinalmente, não são contemplados pelos modelos separados. Objetivou-se com esse trabalho propor dois modelos combinados, que estendem os GLMM para acomodar essas duas fontes de variação, através de dois efeitos aleatórios.

2 GLM, Modelos para sobredispersão e GLMM

2.1 Modelos Lineares Generalizados

Os GLM[5] são uma importante extensão de vários modelos e métodos conhecidos como os modelos lineares, modelos log-lineares, alguns modelos paramétricos de sobrevivência, regressão Poisson e regressão logística, dentre outros. Esta classe de modelos pode ser descrita da seguinte forma:

$$\begin{aligned} Y_i &\sim f(y_i;\theta_i,\phi), \quad i=1,\dots,n \\ E(Y_i)=\mu_i &= g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta}), \end{aligned} \tag{1}$$

sendo Y_i uma variável aleatória com distribuição probabilística pertencente a família exponencial na forma $f(y_i;\theta_i,\phi) = \exp\left\{a(\phi)^{-1}\left[y_i\theta_i - b(\theta_i)\right] + c(y_i,\phi)\right\}$; $\mathbf{x}_i^T\boldsymbol{\beta} = \eta_i$ é um preditor linear e $g^{-1}(\cdot)$ é uma função de ligação monotônica e diferenciável, que liga a média da variável aleatória ao preditor linear η_i.

Um caso particular dos GLM é quando se assume que $Y \sim$ Bernoulli(π) e função de ligação $g(\mu) = \log[\pi/(1-\pi)]$, conhecida como função logit. Na forma da família exponencial tem-se que $\theta = \log[\pi/(1-\pi)]$, $\phi = 1$ e $V(\mu) = \pi(1-\pi)$. Definindo-se um preditor linear η em particular, tem-se, então, o conhecido modelo de regressão logística. Outra alternativa é considerar a função de ligação probit, definida por $g(\mu) = \Phi^{-1}(\mu)$ sendo Φ a função de distribuição acumulada da distribuição normal padronizada.

Um método de estimação para o vetor de parâmetros $\boldsymbol{\beta}$ é o método da máxima verossimilhança que, para o modelo (1), consiste em obter uma solução para o sistema de equações não-lineares, cujo j-ésima componente do vetor escore é

$$U_j = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{dl_i}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{a_i(\phi)}(y_i-\mu_i)\frac{1}{V(\mu_i)}\frac{d\mu_i}{d\eta_i}x_{ij} \quad j=1,\dots,p. \tag{2}$$

Nota-se que somente μ_i and $V(\mu_i)$ estão sendo utilizadas no processo de estimação, não sendo necessária a forma funcional completa da família exponencial. Para obter estimativas de Máxima Verossimilhança de $\boldsymbol{\beta}$, o algoritmo *Iterative Reweighted Least Square - IRLS* é utilizado.

2.2 Um Modelo Para Sobredispersão

Quando um GLM é ajustado a dados na forma de proporções (supondo distribuição binomial para $\mathbf{Y}$) e assumido que o preditor linear resultante está adequado, é esperado que a medida de *deviance* média seja próxima de 1, uma vez que, sob determinadas condições de regularidade, essa tem aproximadamente distribuição $\chi^2_{(n-p)}$, sendo n o número de observações e p o número de parâmetros estimados. Entretanto, algumas situações práticas em que esses modelos são utilizados, obtém-se uma *deviance* média muito maior do que 1, caracterizando assim a variação *extra-binomial* ou *sobredispersão*. Observando-se a função de variância para a distribuição binomial com denominador m, dada por $V(\mu) = m\pi(1-\pi)$, nota-se, que, uma vez estabelecida a esperança $m\pi$, tem-se automaticamente definida a variância esperada para essa distribuição. Em muitas situações práticas a variância observada pode ser bem maior. Dentre as possíveis causas da superdispersão, citamos[2]: má-especificação do preditor linear, presença de *outliers*, função de ligação inadequada, proporções com denominador reduzido, dependência longitudinal não ajustada, inflação de zeros e ocorrência de agrupamentos (*clusters*) que tornam a probabilidade de sucesso π não constante para todas as observações.

Foram propostos alguns métodos para modelar a superdispersão[2] e um desses métodos consiste em considerar uma abordagem em dois estágios. Seja $Y_i|P_i \sim \text{Bin}(m_i, P_i)$ uma variável aleatória com distribuição binomial, com parâmetro m_i e probabilidade de sucesso P_i, sendo que $P_i \sim \text{Beta}(\alpha_i, \beta_i)$, $i = 1, \ldots, N$. Portanto, a função de probabilidade condicional de $Y_i|P_i$ é $f_{Y_i|P_i}(y_i|p_i) = \binom{m_i}{y_i} p_i^{y_i}(1-p_i)^{m_i-y_i}$, $y_i = 0, 1, \ldots, m_i$, sendo a função densidade para P_i dada por $f_{P_i}(p_i) = p_i^{\alpha_i-1}(1-p_i)^{\beta_i-1}B(\alpha_i,\beta_i)^{-1}$, $0 \le p_i \le 1$ com $B(\alpha_i, \beta_i) = \int_0^1 x^{\alpha_i-1}(1-x)^{\beta_i-1}dx$ a função Beta. A função de probabilidade incondicional para Y_i, é dada por

$$f_{Y_i}(y_i) = \binom{m_i}{y_i} \frac{B(\alpha_i + y_i, m_i + \beta_i - y_i)}{B(\alpha_i, \beta_i)}, \qquad (3)$$

conhecida como distribuição *Beta-Binomial* e para qual obtém-se $\text{E}(Y_i) = m_i\alpha/(\alpha+\beta) = m_i\pi_i$ e $\text{Var}(Y_i) = m_i\pi_i(1-\pi_i)[1 + (m_i - 1)(\alpha_i + \beta_i + 1)^{-1}]$. Pode ser mostrado que a correlação entre duas medidas Y_{ij} e Y_{ik}, $j \neq k$, do mesmo elemento i é igual a $\rho = \text{Corr}(Y_{ij}, Y_{ik}) = (\alpha + \beta + 1)^{-1}$, descrito em Hinde e Demétrio (1998)[2]. Nota-se a semelhança dessa função de variância com aquela da distribuição binomial, sendo que o termo $1+(m_i-1)(\alpha_i+\beta_i+1)^{-1}$ é o termo que inflaciona a variância, dependente dos parâmetros α_i e β_i. Essa distribuição pode ser utilizada para modelar dados na forma de proporções com superdispersão. Porém, para dados binários, em que $m_i = 1$, a variância de (3) reduz-se à variância da distribuição Bernoulli e portanto, não é possível utilizar a função de probabilidade (3) para modelar medidas binárias com superdispersão. Dentre os métodos mais conhecidos para tratar a superdispersão, o único que pode ser utilizado em dados binários com repetições é baseado em modelos contendo efeitos fixos e aleatórios, descrito a seguir.

2.3 Modelos Lineares Generalizados Mistos

Uma classe bastante flexível da família de modelos elemento-específico são os modelos lineares generalizados mistos (GLMM), proposto por Breslow e Clayton[1]. Trata-se de uma extensão direta dos GLM, que, dentre outras aplicações, permite acomodar a informação serial contida nas medidas longitudinais, ajustar diversas fontes de variação e acomodar a correlação existente entre observações.

Seja Y_{ij} a j-ésima medida longitudinal do i-ésimo elemento (*subject*), $j = 1, \ldots, n_i$ e $i = 1, \ldots, N$; logo, $\mathbf{Y}_i$ é um vetor n_i-dimensional com as medidas longitudinais do i-ésimo elemento. Seja $\mathbf{b}_i$ o vetor q-dimensional de efeitos aleatórios, independentes e com distribuição $N(\mathbf{0}, \mathbf{D})$, em que $\mathbf{0}$ é um vetor de zeros e $\mathbf{D}$ uma matriz de variância-covariância, geralmente estruturada e pré-especificada; $\mathbf{b}_i$ está associado a $q-1$ fatores e/ou covariáveis que compõe o vetor $\mathbf{z}_{ij}$, com dimensão $(q \times 1)$; $\boldsymbol{\beta}$ é o vetor p-dimensional de efeitos fixos, associados a p-1 fatores e/ou covariáveis que compõem o vetor $\mathbf{x}_{ij}$ com dimensão $(p \times 1)$. É assumido que as medidas Y_{ij} são condicionalmente independentes, dado $\mathbf{b}_i$, com função (densidade) de probabilidade pertencente à família exponencial de distribuições, ou seja,

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp\left\{\phi^{-1}\left[y_{ij}\theta_{ij} - \psi(\theta_{ij})\right] + c(y_{ij}, \phi)\right\}, \qquad (4)$$

sendo $\eta(\mu_{ij}) = \eta\left[\text{E}(Y_{ij}|\mathbf{b}_i)\right] = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i$, $\eta(\cdot)$ é uma função de ligação pré-especificada e ϕ é o parâmetro de dispersão. Se $\theta_{ij} = \eta(\mu_{ij})$, então η é uma função de ligação canônica. Denota-se também que $f(\mathbf{b}_i|\mathbf{D})$ é a função densidade dos efeitos aleatórios $\mathbf{b}_i$.

Assumindo um modelo probabilístico Bernoulli para as medidas longitudinais e utilizando a função de ligação logit, dada por $g(\cdot) = \exp(\cdot)/\{1 + \exp(\cdot)\}$, tem-se um modelo linear generalizado misto logit-normal-Bernoulli, definido por

$$\begin{aligned} Y_{ij}|\mathbf{b_i} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1 + \exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)} \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}). \end{aligned} \qquad (5)$$

Quando é adequado assumir a função de ligação probit, tem-se então um MLGM probit-normal-Bernoulli, dado por

$$\begin{aligned} Y_{ij}|\mathbf{b_i} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \Phi\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\big) \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}), \end{aligned} \tag{6}$$

sendo $\Phi(x) = \int_{-\infty}^{x}(2\pi)^{-1/2}\exp\left[(1/2)u^2\right]du$[1].

3 Modelos Combinando Efeitos Aleatórios Normal e Conjugado

Quando dados binários estão sujeitos à superdispersão e, ao mesmo tempo, são medidos longitudinalmente, uma extensão particular dos MLGM se faz necessária. Aqui serão combinadas as idéias da Seção ?? para produzir um modelo que permita acomodar essas duas fontes de variação simultaneamente, de forma similar àquela descrita em Molenberghs, Verbeke e Demétrio (2007) [3].

Seja y_{ij} uma medida binária (codificado como 0 igual à não-ocorrência do evento e 1 igual à ocorrência do evento) do i-ésimo elemento (parcelas, pessoas, animais, equipamento, etc) no j-ésimo tempo; θ_{ij} é o parâmetro parâmetro de superdispersão para a (ij)-ésima observação. $\mathbf{b}_i$ é o vetor q-dimensional de efeitos aleatórios, associados ao vetor $\mathbf{z}_{ij}$ com q fatores e/ou covariáveis; o vetor de efeitos fixos $\boldsymbol{\beta}$ com dimensão p está associado ao vetor $\mathbf{x}_{ij}$ contendo p linhas referentes a fatores e/ou covariáveis. A probabilidade de ocorrer sucesso para a (ij)-ésima medida é $P(Y_{ij} = 1) = \pi_{ij}$.

3.1 Modelo Combinado Logit-Normal-Bernoulli-Beta

Assumindo o modelo probabilístico Bernoulli para as medidas binárias, a distribuição beta para o efeito aleatório que acomodará a superdispersão e a distribuição normal para o efeito aleatório que modelará a correlação entre as medidas longitudinais e considerando ainda a função de ligação logit, propõe-se o modelo logit-normal-Bernoulli-beta que se segue:

$$\begin{aligned} Y_{ij}|\mathbf{b_i} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \theta_{ij}\frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)} \\ \theta_{ij} &\sim \text{Beta}(\alpha, \beta) \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}), \end{aligned} \tag{7}$$

sendo que α e β são os parâmetros da distribuição beta e $\mathbf{D}$ é a matriz de variância-covariância referente a $\mathbf{b}_i$.

Uma forma de acomodar a superdispersão e a correlação presente nos dados longitudinais foi modificar o modelo (5), acrescentando um efeito aleatório multiplicativo θ_{ij} na expressão de π_{ij}. O método de estimação adotado foi o da máxima verossimilhança. A função de probabilidade condicional para $Y_{ij}|\theta_{ij}, \mathbf{b}_i$ é dada por

$$f(y_{ij}|\theta_{ij}, \mathbf{b}_i) = \left[\theta_{ij}\frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}\right]^{y_{ij}}\left[1-\theta_{ij}\frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}\right]^{1-y_{ij}}, \quad y_{ij} \in \mathbb{R} \tag{8}$$

sendo o parâmetro $\boldsymbol{\beta}$ fixo e desconhecido. A função densidade de probabilidade para o efeito aleatório $\mathbf{b}_i$ tem a forma da distribuição normal multivariada sendo $\mathbf{D}$ a matriz de variâncias e covariâncias contendo parâmetros fixos e desconhecidos. A função densidade de probabilidade para o efeito aleatório θ_{ij} tem a forma da distribuição Beta sendo α e β também parâmetros fixos e desconhecidos. A contribuição da (ij)-ésima observação para a função de verossimilhança, assumindo a independência entre θ_{ij} e $\mathbf{b}_i$, é dada por $f(y_{ij}, \theta_{ij}, \mathbf{b}_i) = f(y_{ij}|\theta_{ij}, \mathbf{b}_i)f(\mathbf{b}_i)f(\theta_{ij})$ e a contribuição do i-ésimo elemento para a função de verossimilhança, assumindo independência condicional[2] pode ser descrita por meio da função de densidade conjunta $f(\mathbf{y}_i, \boldsymbol{\theta}_i, \mathbf{b}_i) = \prod_{j=1}^{n_i} f(y_{ij}|\theta_{ij}, \mathbf{b}_i)f(\mathbf{b}_i)f(\theta_{ij})$. Para a estimação por máxima verossimilhança de $\boldsymbol{\beta}$ e $\mathbf{D}$, esta função foi integrada considerando os domínios de $\boldsymbol{\theta}_i$ e $\mathbf{b}_i$, obtendo-se

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i) = \int \prod_{j=1}^{n_i} f(y_{ij}|\theta_{ij}, \mathbf{b}_i)f(\mathbf{b}_i)f(\theta_{ij})d\mathbf{b}_i, \tag{9}$$

[1] Chama-se a atenção para o símbolo π, que nessa integral significa o número $\pi = 3,1415\ldots$; a probabilidade de sucesso Bernoulli π_{ij} estará sempre acompanhada de um ou dois sub-índices.

[2] Uma vez que a dependência longitudinal e a superdispersão foram levadas em conta na função de verossimilhança, as medidas são consideradas independentes.

e, portanto, a função de verossimilhança para $\boldsymbol{\beta}$ e $\mathbf{D}$ condicional ao valor de $\boldsymbol{\theta}$, para N elementos é descrita como

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}, \alpha, \beta|\boldsymbol{\theta}) &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \left[\theta_{ij} \frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)} \right]^{y_{ij}} \\ &\quad \times \left[1 - \theta_{ij} \frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)} \right]^{1-y_{ij}} \\ &\quad \times \frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\mathbf{D}|}} \exp\left(-\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right) \\ &\quad \times \frac{\theta_{ij}^{\alpha-1}(1-\theta_{ij})^{\beta-1}}{B(\alpha,\beta)} d\mathbf{b}_i. \end{aligned} \tag{10}$$

Essa função de verossimilhança continua condicionada ao efeito aleatório θ_{ij}. Para o caso Bernoulli, pode-se escrever que $f(y_{ij}|\theta_{ij}, \mathbf{b}_i) = f(y_{ij}=0|\theta_{ij}, \mathbf{b}_i)^{y_{ij}} f(y_{ij}=1|\theta_{ij}, \mathbf{b}_i)^{1-y_{ij}}$. Denotando $k_{ij} = [\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)][1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)]^{-1}$, tem-se que $f(y_{ij}=1|\mathbf{b}_i) = \int_0^1 f(y_{ij}=1|\theta_{ij}, \mathbf{b}_i) f(\theta_{ij}) d\theta_{ij} = \alpha k_{ij}(\alpha+\beta)^{-1}$ e, da mesma forma, calcula-se $f(y_{ij}=0|\mathbf{b}_i) = \int_0^1 f(y_{ij}=0|\theta_{ij}, \mathbf{b}_i) f(\theta_{ij}) d\theta_{ij} = \frac{(1-k_{ij})\alpha+\beta}{\alpha+\beta}$. Logo, a função de verossimilhança marginal, apenas condicionada a $\mathbf{b}_i$, é

$$L(\boldsymbol{\beta}, \mathbf{D}, \alpha, \beta) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \left(\frac{\alpha k_{ij}}{\alpha+\beta} \right)^{y_{ij}} \cdot \left[\frac{(1-k_{ij})\alpha+\beta}{\alpha+\beta} \right]^{1-y_{ij}} f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \tag{11}$$

$$= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \frac{1}{\alpha+\beta} (\alpha k_{ij})^{y_{ij}} \left[(1-k_{ij})\alpha+\beta\right]^{1-y_{ij}} f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i; \tag{12}$$

Essa função claramente não permite a obtenção de estimadores analíticos, o que demanda algoritmos iterativos como Newton-Raphson, Gauss-Newton, dentre outros. Dentro de cada ciclo iterativo é necessário que a integral sobre $\mathbf{b}_i$ seja solucionada. Como a integral em (12) é da forma $\int f(z)\phi(z)dz$ o algoritmo de quadratura de Gauss-Hermite adaptativa[3] pode ser aplicado para obter uma solução numérica da integral. Esse algoritmo, junto com alguns algoritmos de otimização já estão implementados no flexível e conveniente procedimento NLMIXED do sistema SAS, utilizado neste trabalho. Após os critérios de convergência terem sido atingidos, estimativas de máxima verossimilhança podem ser utilizadas para inferência estatística. Um comentário importante é sobre a identificabilidade dos parâmetros. Os parâmetros α e β não são identificáveis simultaneamente, quando o preditor linear contém um intercepto. Uma solução é estabelecer uma restrição do tipo $\beta/\alpha = c$, que soluciona o problema de identificabilidade.

3.2 Modelo Combinado Probit-Normal-Bernoulli-Beta

O modelo probit-normal-Bernoulli-beta é, essencialmente, igual ao modelo (7), apenas diferindo pela função de ligação adotada. É descrito como

$$\begin{aligned} Y_{ij}|\mathbf{b_i} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \theta_{ij}\Phi\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right) \\ \theta_{ij} &\sim \text{Beta}(\alpha,\beta) \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}). \end{aligned} \tag{13}$$

O processo de estimação para o modelo (13) modifica a função de probabilidade condicional para $Y_{ij}|\theta_{ij}, \mathbf{b}_i$, sendo descrita como

$$f(y_{ij}|\theta_{ij}, \mathbf{b}_i) = \left(\theta_{ij}k_{ij}\right)^{y_{ij}} \left(1-k_{ij}\right)^{1-y_{ij}}, \tag{14}$$

sendo $k_{ij} = \Phi(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)$ e $\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp[-t^2/2]dt$. Utilizando os resultados do modelo Logit-normal-Bernoulli-Beta, tem-se a função densidade de probabilidade de Y_{ij} condicional a $\mathbf{b}_i$, dada por $f(y_{ij}|\mathbf{b}_i) = \frac{1}{\alpha+\beta}(\alpha k_{ij})^{y_{ij}} \left[(1-k_{ij})\alpha+\beta\right]^{1-y_{ij}}$ e dessa forma, a função de verossimilhança marginal, condicional a $\mathbf{b}_i$, para N elementos é dada por

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}, \alpha, \beta) &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \frac{\left[\Phi(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)\alpha\right]^{y_{ij}}}{\alpha+\beta} \left\{[1-\Phi(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)]\alpha+\beta\right\}^{1-y_{ij}} \\ &\quad \times (2\pi)^{-q/2}|\mathbf{D}|^{-1/2} \exp\left\{ -\frac{1}{2}\mathbf{b}_i^T\mathbf{D}\mathbf{b}_i \right\} d\mathbf{b}_i. \end{aligned} \tag{15}$$

[3]Esse algoritmo está revisado na seção **??**.

Para essa função de verossimilhança, pode-se adotar o mesmo método utilizado para a otimização de (12), utilizando um algoritmo para integração numérica em conjunto com um algoritmo de otimização iterativo, adotando alguma restrição sobre os parâmetros α e β.

Entretanto, para esse modelo, também é possível obter uma solução analítica para a integral em (15), o que permite obter uma expressão para função densidade de probabilidade marginal de $\mathbf{Y}_i$ [7].

Com esse resultado, a expressão da função de probabilidade marginal de $\mathbf{Y}_i = \mathbf{1}$ passa a ser

$$f(\mathbf{y}_i = \mathbf{1}) = \Phi(\mathbf{X}_i\boldsymbol{\beta}; \mathbf{L}_i^{-1}), \quad \mathbf{L}_i = \mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T. \tag{16}$$

Essa expressão é a função de probabilidade do modelo MLGM probit-Bernoulli-normal. Tem-se, portanto, como a função de probabilidade para um modelo probit-normal-Bernoulli-beta a função

$$f(\mathbf{y}_i = \mathbf{1}) = \left(\frac{\alpha}{\alpha+\beta}\right)^{n_i} \Phi(\mathbf{X}_i\boldsymbol{\beta}; \mathbf{L}_i^{-1}), \quad \mathbf{L}_i = \mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T. \tag{17}$$

Essa expressão pode ser utilizada diretamente no processo de estimação pelo método de máxima verossimilhança. Uma aproximação da função distribuição Gaussiana acumulada para a distribuição logística pode ser utilizada para aproximar as estimativas do modelo probit para o modelo logit, por meio da multiplicação do preditor linear com a constante $c = 16\sqrt{3}/15\pi$, o que permite o cálculo de medidas como a razão das chances e seus intervalos de confiança (ZEGER; LIANG; ALBERT, 1988).

4 Aplicação: um estudo clínico em dermatologia

Um estudo aleatorizado multi-centro, duplo cego e com grupos paralelos foi realizado com o objetivo de comparar dois medicamentos via oral (denominados aqui *A* e *B*) para tratamento da *toenail dermatophite onychomycosis* (TDO). Essa dermatite é uma infecção de unha bastante comum e de difícil tratamento. Os compostos antifúngicos tradicionais para tratamento de TDO necessitam que o paciente administre a medicação até que a(s) unha(s) se desenvolva(m) saudavelmente, por completo. Os novos medicamentos têm reduzido esse período de administração do medicamento para três meses. Neste estudo, foram comparadas duas drogas quanto à sua eficácia e segurança durante o tratamento, por 12 semanas de terapia contínua. Este estudo foi descrito em De Backer et al. (1996) e seu conjunto de dados explorado em Molenberghs e Verbeke (2005).

Foram estudados 378 pacientes associados a 36 centros de atendimento de saúde. Os pacientes foram acompanhados por 12 semanas durante o tratamento e, posteriormente, acompanhados até um total de 48 semanas (12 meses). As medidas foram tomadas no início do tratamento, a cada mês durante o tratamento e a cada 3 meses após o tratamento, o que resultou de um máximo de 7 medidas para cada paciente. Antes de iniciar o tratamento, o médico responsável selecionou uma unha para ser acompanhada durante o período. O estudo quantitativo se limitou aos pacientes cujas unhas selecionadas para acompanhamento correspondem a uma dos primeiros pododáctilos ("dedão do pé"). Dessa forma, o número de pacientes para os tratamentos A e B reduziu para 294. Os tratamentos foram alocados aos pacientes por processo aleatório. Foram medidos 7 pontos no tempo para 76% dos pacientes. Uma das medidas de interesse foi o grau de severidade codificados como 0 (não-severo) e 1 (severo). As questões de interesse são (i) como o percentual de severidade decresce com o tempo e (ii) se os dois tratamentos diferem ou não. A Figura **??** apresenta a evolução do percentual de severidade para os dois tratamentos, ao longo do tempo. Nota-se a clara influência do tempo no percentual de severidade para ambos os tratamentos mas não é claro se os tratamentos diferem significativamente. Como o objetivo da pesquisa era verificar se os dois tratamentos estudados diferem ou não, foi utilizado os seguinte preditor linear:

$$\text{Logit}(\mu_{ij}) = \begin{cases} (\beta_0 + b_i) + \beta_1 t_{ij}, & \text{Tratamento A} \\ (\beta_2 + b_i) + \beta_3 t_{ij}, & \text{Tratamento B} \end{cases} \tag{18}$$

5 Resultados

A análise dos dados de TDO foi feita utilizando-se os modelos baseados na função de ligação probit, com os resultados apresentados na Tabela a seguir. Como no caso logit, foram ajustados os modelos MLG Bernoulli, beta-binomial, MLGM Bernoulli-normal e o modelo combinado Bernoulli-beta-normal. Foi assumida para os modelos beta-binomial e o combinado a restrição $\alpha + \beta = c$ para eliminar o problema da identificabilidade, uma vez que a restrição $\beta/\alpha = c$ não levou a uma convergência dos algoritmos. Foram obtidos resultados diferentes do modelo que considerou a função de ligação logit: com base na estatística de -2×log-verossimilhança e no critério de Akaike, os modelos beta-binomial e o modelo combinado foram bastante superiores aos modelos MLG e MLGM. O modelo beta-binomial foi considerado apenas com os coeficientes angulares dos modelos, uma vez que a presença dos interceptos e do parâmetro de superdispersão α, simultaneamente no modelo, gerava forte correlação entre os parâmetros estimados e, consequentemente, instabilidade numérica no processo iterativo de estimação. Nota-se também que os modelos

Tabela 1: Resultados dos ajustes dos modelos MLG, beta-binomial, MLGM e o modelo combinado, assumindo função de ligação probit

		Bernoulli	Beta-binomial
Efeito	Parâmetro	Estimativa (e.p.)	Estimativa (e.p.)
Intercepto Tratamento A	β_0	−0,3678 (0,06493)	—
Coef. Angular Tratamento A	β_1	−0,09647 (0,01276)	-0,1125 (0,01136)
Intercepto Tratamento B	β_2	−0,3661 (0,06592)	—
Coef. Angular Tratamento B	β_3	−0,1335 (0,01502)	-0,1537 (0,01320)
Desvio Padrão Ef. Aleatórios	σ	—	—
Parâmetro Superdispersão	α	—	7,0493 (0,3507)
−2 log-verossimilhança		1815,1	-6965
Critério AIC		1823,1	-6959
		Bernoulli-normal	Combinado
Efeito	Parâmetro	Estimativa (e.p.)	Estimativa (e.p.)
Intercepto Tratamento A	β_0	−0,9175 (0,2296)	−1,0714 (0,6774)
Coef. Angular Tratamento A	β_1	−0,1961 (0,02117)	−0,8335 (0,1631)
Intercepto Tratamento B	β_2	−0,9970 (0,2361)	−1,8020 (0,4939)
Coef. Angular Tratamento B	β_3	−0,2723 (0,02626)	−1,2039 (0,2099)
Desvio Padrão Ef. Aleatório	σ	2,1173 (0,1975)	8,0835 (1,1202)
Parâmetro Superdispersão	α	—	7,8882 (0,2189)
−2 log-verossimilhança		1271,7	-7523
Critério AIC		1281,7	-7511

alternativos ao modelo MLG Bernoulli foram todos mais informativos, indicando a necessidade de se modelar a variação extra-Bernoulli. Porém, a utilização do modelo combinado mostra-se mais geral e abrangente, acomodando tanto o efeito de superdispersão quanto a dependência temporal dos dados binários. Tanto o parâmetro de superdispersão α quanto o desvio padrão do efeito aleatório mostraram-se significativos e com valores mais razoáveis, quando comparado ao modelo com função de ligação logit.

A aproximação do modelo probit para o modelo logit, em que a constante $c = 16\sqrt{3}/15\pi$ é multiplicada pelo preditor linear, permite que inferências sobre a razão das chances sejam estabelecidas. A significância dos coeficientes angulares já indicam que ambos os tratamentos levam à redução da severidade ao longo dos 12 meses. A diferença entre os coeficientes angulares do tratamento A e B foi testada e foi concluído que os tratamentos diferem na intensidade de redução da severidade de ocorrência do TDO (P=0,0255) e, portanto, o tratamento B é mais rápido na redução da severidade. Calculando a razão das chances, tem-se que, para o tratamento A, a chance de redução da severidade de TDO é de $1/\exp(-1,4173) = 4.13$ vezes, para cada mês após o início do tratamento. Para o tratamento B a chance de redução da severidade é de 7,8 vezes, para cada mês que passa, após o início do tratamento.

A aplicação clínica aqui apresentada foi a motivação para o estudo desses modelos mas diversas situações são também contempladas por esses modelos. Outras aplicações agronômicas serão desenvolvidas em um futuro próximo.

Referências

[1] BRESLOW, N.E; CLAYTON, D.G. Approximate inference in generalized linear mixed models. JASA, Boston, V. 88, p. 9-25, 1993.

[2] HINDE, J.; DEMÉTRIO, C.G.B. Overdispersion: models and estimation. Computational Statistics and Data Analysis, Amsterdam, v. 27, n. 2, p. 151-70, 1988.

[3] MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B. An extended random-effect approach to modeling repeated, overdispersion count data. Lifetime Data Analysis, Hinghan, v. 13, n. 14, 2007.

[4] MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B.; VIEIRA, A.M.C. A family of generalized linear models for repeated measures with normal and conjugate random effects. Statistical Science, 2009. *accepted for publication*

[5] NELDER,J.A.; WEDDERBURN. Generalised linear models. JRSS, A, London, v. 135. n.3, p. 370-84, 1972.

[6] VERBEKE, G.; MOLENBERGHS, G. Linear Mixed for Longitudinal Data. New York: Springer, 2000.

[7] VIEIRA, A.M.C. Modelagem simultanea de média e dispersão e aplicações na pesquisa agronômica. Piracicaba: ESALQ-Universidade de São Paulo (Tese). 2008.