Modelo logístico multinível:
um enfoque em métodos de estimação e predição

Karin Ayumi Tamura

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DE TÍTULO DE MESTRE
EM
CIÊNCIAS

Área de Concentração: Estatística
Orientadora: Prof^a Dr^a Viviana Giampaoli

São Paulo, maio de 2007.

Modelo logístico multinível:
um enfoque em métodos de estimação e predição

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por Karin Ayumi Tamura
e aprovada pela comissão julgadora.

São Paulo, 25 de maio de 2007

Banca examinadora:

- Profa. Dra. Viviana Giampaoli (orientadora) - IME/USP
- Prof. Dr. Gilberto Alvarenga de Paula - IME/USP
- Prof. Dr. Francisco José Azevêdo Cysneiros - UFPE

Aos meus pais, Luiz e Yoshiko

Ao meu namorado, Alexandre

Agradecimentos

Primeiramente, agradeço a Deus e aos meus pais, Luiz e Yoshiko, por terem me preparado e me proporcionado a oportunidade de chegar até aqui.

Não posso me esquecer da minha querida orientadora, Viviana Giampaoli, que me direcionou para que este trabalho tenha sido finalizado com sucesso. Obrigada pela paciência, dedicação, amizade e principalmente por ser uma excelente profissional.

Agradeço aos meus irmãos, Larissa e Leandro, pelo apoio e compreensão nas noites e fins de semana que tive que ficar estudando e usando o computador.

Um agradecimento especial a algumas pessoas que me incentivaram e principalmente acreditaram no meu potencial: Rubens Stephan, Marcelo Sousa e Juan Carlos Ruilova.

Obrigada aos professores Gilberto e Francisco pelas enriquecedoras sugestões ao meu trabalho.

Aos meus amigos(as) da empresa e da USP. Alguns nomes não poderiam deixar de ser citados pela ajuda e amizade: Agnes, Regina, Catinha, Thelma, Tiago, Gustavo, Eduardo, Fernando, Guilherme, Michael, Ana Cristina, Rovane, Leonardo Barci. Peço desculpas por não citar todos os nomes, entretanto isso não significa que sejam pessoas menos importante na minha vida.

Finalmente, dedico meu trabalho a uma pessoa muito especial, Alexandre Noma, e a toda sua família. Sem ele ao meu lado eu jamais chegaria tão longe, pois ele é uma pessoa quem sempre me apoiou, me incentivou, me ajudou e principalmente compreendeu minha ausência neste período que tive que me dedicar ao mestrado. Obrigada por todo seu amor, saiba que este sentimento é que me motiva a querer ser uma pessoa cada vez melhor.

Sumário

Lista de Figuras	ii
Lista de Tabelas	iv
1 Introdução	1
1.1 Considerações iniciais	1
1.2 Organização dos capítulos	3
2 Revisão dos modelos multiníveis	5
2.1 Considerações iniciais	5
2.1.1 Modelo linear generalizado logístico	6
2.2 Modelos multiníveis	7
2.2.1 Modelos multiníveis com 2 níveis	9
2.2.2 Modelo multinível logístico com 2 níveis	15
2.3 Métodos de estimação	17
2.3.1 Estimação via QVP	18
2.3.2 Estimação via QGH	23
2.3.3 Aspectos computacionais	26
3 Predição da variável resposta	29
3.1 Introdução	29
3.2 Predição da variável resposta	30
3.2.1 Predição da variável resposta: modelo logístico multinível	32
3.2.2 Medidas de desempenho	34

4	Simulação	37
4.1	Esquema de simulação	38
4.2	Resultados	40
5	Aplicação	47
6	Comentários finais	53
A	Código do programa para simulação	57
B	Box-Plot do resíduo das simulações	73
B.1	Box-Plot do resíduo das simulações para $N=200$	73
B.2	Box-Plot do resíduo das simulações para $N=1000$	85
	Referências Bibliográficas	97

Lista de Figuras

2.1	Comparativo entre o modelo de regressão usual e o modelo com efeito aleatório.	10
B.1	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=0.1$.	73
B.2	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=0.1$.	74
B.3	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=0.1$.	75
B.4	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=0.1$.	76
B.5	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=1$.	77
B.6	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=1$.	78
B.7	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=1$.	79
B.8	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=1$.	80
B.9	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=2$.	81
B.10	Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=2$.	82

B.11 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=2$	83
B.12 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=2$	84
B.13 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=0.1$	85
B.14 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=0.1$	86
B.15 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=0.1$	87
B.16 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=0.1$	88
B.17 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=1$	89
B.18 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=1$	90
B.19 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=1$	91
B.20 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=1$	92
B.21 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=2$	93
B.22 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=2$	94
B.23 Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=2$	95
B.24 Box-plot da Distribuição dos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=2$	96

Lista de Tabelas

3.1	Tabela de classificação - observado versus predito.	35
4.1	Resultado das simulações para $N=200$	42
4.2	Resultado das simulações para $N=1000$	43
4.3	Medidas resumo da média dos resíduos absolutos na base de modelagem. . .	44
4.4	Medidas resumo da média dos resíduos absolutos na base de predição. . . .	45
5.1	Ajustes dos modelos usual, multinível via QVP e multinível via QGH. . . .	50
5.2	Resíduo de Pearson dos modelos usual, via QVP e via QGH.	50
5.3	Tabela de classificação do modelos usual, via QVP e via QGH na base de modelagem e predição.	51
5.4	Medidas de desempenho dos modelos usual, via QVP e via QGH.	51

Resumo

Modelo multinível é uma ferramenta estatística cada vez mais popular para análise de dados com estrutura hierárquica. O objetivo deste trabalho é propor um método para realizar a predição de observações de novos grupos usando modelos de regressão logística multinível com 2 níveis. Além disso, é apresentado e comparado dois métodos de estimação para o modelo multinível: Quase-verossimilhança Penalizada (QVP) e Quadratura de Gauss-Hermite (QGH).

A idéia central está baseada no trabalho de (Jiang e Lahiri,2006) no qual se propõe o uso do chamado melhor estimador empírico para o efeito aleatório. Através deste estimador, utilizou-se a parte fixa do modelo em conjunto com uma estimativa do desvio padrão do efeito aleatório para fazer a predição de observações de novos grupos, encontrando a probabilidade estimada dessa observação apresentar o evento de interesse, dadas suas características.

Palavras-chave: Modelos multiníveis, regressão logística, predição da variável resposta.

Abstract

Multilevel model is an statistical tool which is becoming more and more popular in data analysis with hierachical structure. The purpose of this dissertation is to present a method to make a prediction of new group observation in multilevel logistic regression models with 2 levels. Besides, were presented and compared two estimation methods for multilevel model: Penalized Quase-likelihood and Gauss-Hermite Quadrature.

The central idea is based on the paper of Jiang and Lahiri (2006), which is presented the empirical best estimator for the random effect. Through this estimator was used the fixed part of the model with an estimative of the standard deviation of the random effect to find the estimated probability of this observation presenting the target event, in accordance with its characteristic.

Key-words: Multilevel models, logistic regression, variable response prediction.

Capítulo 1

Introdução

1.1 Considerações iniciais

Recentemente, os modelos lineares hierárquicos ou multiníveis começaram a ser conhecidos e aplicados em diversos campos científicos, principalmente nas áreas da Educação e da Ciências Sociais.

Quando os dados são estruturados em hierarquias, as unidades de um mesmo nível pertencentes a uma unidade de nível mais alto são raramente independentes. Isso acontece porque essas unidades compartilham de um mesmo ambiente ou apresentam características semelhantes. Portanto, a suposição de independência é violada e assim passa a existir correlação entre unidades pertencentes a um nível mais alto. Ignorar esta relação, omitindo a importância de efeitos de grupo, pode tornar inválida muitas técnicas estatísticas tradicionais utilizadas para estudar a relação entre os dados.

As aplicações dos modelos lineares multiníveis têm crescido devido à larga abrangência de problemas com estruturas hierárquicas dos dados. Nos últimos 15 anos, muitas extensões dos modelos lineares multiníveis foram desenvolvidas, possibilitando a inclusão de

variáveis em ambos os níveis: individual e agregado. Por exemplo, em uma empresa, um pesquisador pode estar interessado em investigar como as características do local de trabalho influenciam na produtividade do trabalhador. Tais dados consistem em uma estrutura hierárquica, com as unidades de trabalhadores *aninhados* dentro das firmas. Assim, as variáveis poderão ser mensuradas em ambos os níveis: trabalhador e firma.

As análises hierárquicas também podem ser utilizadas nos estudos da área da educação, como por exemplo, na mensuração do rendimento escolar das crianças que compõem as turmas de alunos. Essas, por sua vez, compõem as escolas. E as escolas são dirigidas por órgãos administradores. Assim, pode-se ter quatro níveis hierárquicos associando as crianças ao 1º nível, as turmas de alunos ao 2º, as escolas ao 3º e os órgãos administradores ao 4º nível.

Desta forma, as análises baseadas em modelos multiníveis apresentam estimativas mais confiáveis, uma vez que não assumem erroneamente o pressuposto de independência entre as observações das unidades pertencentes a um mesmo grupo. Este argumento indica que uma análise de modelos de regressão tradicionais, que exige a suposição de independência entre as observações, não seria a mais adequada para essa estrutura de dados.

Neste tipo de modelo, como as unidades de cada nível são consideradas como amostras aleatórias de uma população, elas contribuem no modelo incorporando efeitos aleatórios. Estes efeitos aleatórios são responsáveis por coeficientes aleatórios que levam em conta a variabilidade intra-unidades, por exemplo, por meios da variabilidade nos interceptos e/ou da variabilidade nas inclinações. Esse tipo de modelo é citado por alguns autores como Modelo Condicional (Lee e Nelder,2004), (Liang,Diggle,Heagerty e Zeger,2002), Modelo Multinível (Fitzmaurice e Lard,1993), (Yu e Cumberland,Technical report) ou Modelo com variável latente, dependendo da formulação do modelo.

Um dos objetivos desta dissertação é apresentar e comparar dois métodos de estimação: Quase-Verossimilhança Penalizada (QVP) e Quadratura de Gauss-Hermite (QGH), no modelo de regressão logística multinível.

O outro objetivo é fazer a predição de novas observações utilizando modelos de regressão logística multinível. No modelo logístico usual ou tradicional, que é composto apenas pela parte fixa, fazer a predição é simples, pois através da função de ligação é possível obter a predição da variável resposta. Porém, para o modelo multinível, além do efeito fixo, também se tem o efeito aleatório que é estimado para o grupo da amostra de modelagem. Logo, a proposta deste trabalho é utilizar a parte fixa do modelo em conjunto com uma estimativa do desvio padrão do efeito aleatório para se fazer a predição de observações de novos grupos para regressão logística multinível e comparar seu resultado com o da regressão logística tradicional.

1.2 Organização dos capítulos

A presente dissertação encontra-se dividida em 6 capítulos.

No primeiro capítulo, como já visto, tem-se uma breve revisão da idéia básica sobre os dados com estrutura hierárquica, destacando o uso de modelos multiníveis.

O capítulo 2 apresenta brevemente os principais conceitos de modelos lineares generalizados e uma revisão dos modelos multiníveis com 2 níveis de hierarquia para variável resposta binária através de função logit, descrevendo os métodos de estimação via QVP e QGH.

O capítulo 3 trata da predição de observações de novos grupos para o modelo de regressão logística multinível com 2 níveis, considerando apenas o intercepto como efeito aleatório. Além disso, neste capítulo também são apresentadas as definições de algumas medidas de desempenho na classificação da variável resposta para o modelo com resposta binária.

O comportamento das estimativas dos parâmetros e o desempenho dos modelos de regressão logística usual e multinível via QVP e QGH, com base em simulações, é analisado no capítulo 4.

No capítulo 5, encontram-se os resultados referentes a uma aplicação na área de saúde,

considerando modelos de regressão logística usual e multinível via QVP e QGH.

Por fim, o último capítulo contém uma discussão sobre os resultados encontrados e sugestões para futuros trabalhos relacionados com este tema.

Capítulo 2

Revisão dos modelos multiníveis

2.1 Considerações iniciais

Neste capítulo, serão apresentados brevemente os modelos lineares generalizados e, em particular, o modelo para análise de dados com resposta binária com ligação logit, conhecido como regressão logística. Esse modelo pertence à classe dos Modelos Lineares Generalizados (MLG) com distribuição Binomial.

Suponha $Y_1, \dots, Y_n$ variáveis aleatórias independentes, cada uma com densidade na forma da família exponencial dada por

$$f(y; \theta_i, \phi) = \exp[\phi\{y\theta_i - b(\theta_i)\} + c(y, \phi)], \quad (2.1)$$

em que $E(Y_i) = \mu_i = b'(\theta_i)$, $Var(Y_i) = \phi^{-1}V_i$, $V_i = \frac{d\mu_i}{d\theta}$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão.

A função de variância desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição. Isto é, dada a função de variância, tem-se uma classe de distribuições correspondentes, e vice-versa.

Uma propriedade interessante envolvendo a distribuição de Y e a função de variância é a seguinte:

$$\sqrt{\phi}(Y - \mu) \xrightarrow{D} N(0, V(\mu)), \text{ quando } \phi \rightarrow \infty, \quad (2.2)$$

ou seja, para ϕ grande, Y segue uma distribuição aproximadamente normal de média μ e variância $\phi^{-1}V(\mu)$. Esse tipo de abordagem assintótica, diferente da usual em que n é grande, foi introduzida por (Jorgensen,1987).

Os modelos lineares generalizados são definidos por (2.1) e pela componente sistemática

$$g(\mu_i) = \eta_i, \quad (2.3)$$

em que $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^t$ representa os valores de p variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação.

A função de ligação canônica, ocorre quando o parâmetro natural (θ) e o preditor linear (η) coincidem, isto é, quando

$$\theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j \text{ com } i = 1, \dots, n.$$

Logo, obtém-se estimativas de máxima verossimilhança únicas para os parâmetros $\beta_1, \dots, \beta_p$ sempre que estas funções existirem, ou seja, as ligações canônicas garantem a concavidade de $L(\theta; y)$ e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente (Paula,2004).

2.1.1 Modelo linear generalizado logístico

Apresenta-se a seguir a distribuição Binomial pertencente à forma da família exponencial dada por (2.1).

Seja Y^* a proporção de sucessos de n ensaios independentes, cada um com probabilidade de ocorrência μ . Será assumido que $nY^* \sim N(n, \mu)$. A densidade de Y^* fica expressa na forma

$$\binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n - ny^*} = \exp \left\{ \log \binom{n}{ny^*} + ny^* \log \left(\frac{\mu}{1 - \mu} \right) + n \log(1 - \mu) \right\},$$

em que $0 < \mu, y^* < 1$. Obtém-se de (2.1), fazendo $\phi = n$, $\theta = \log(\mu/(1 + \mu))$, $b(\theta) = \log(1 + e^\theta)$ e $c(y^*, \phi) = \log\left(\frac{\phi}{\phi y^*}\right)$.

Neste caso particular, a função de variância, definida por $V(\mu) = \mu(1 - \mu)$, caracteriza a classe de distribuições binomiais com probabilidades de sucesso μ .

Ao se considerar a distribuição binomial na família exponencial, tem-se que a função de ligação canônica é

$$\text{logit} = \log\left(\frac{\mu}{1 - \mu}\right), \quad (2.4)$$

tem-se assim o chamado modelo logístico.

Há também outras duas ligações que são bastante utilizadas para modelos com resposta binária que não serão consideradas neste trabalho: probit e complemento log-log. A ligação probit produz estimativas muito próximas da logit e é definida por ser a inversa da distribuição acumulada da normal padrão. A ligação complemento log-log é indicada para distribuição com valores extremos ou distribuição não simétrica dos dados. Para mais detalhes, ver (Paula,2004) e (Hosmer e Lemeshow,2000).

2.2 Modelos multiníveis

Neste tópico, serão apresentados sucintamente os modelos multiníveis. Os Modelos Lineares Generalizados Mistos (MLGM) são uma extensão dos MLG, conforme (Demidenko,2004). Segundo (Renard,2002) os Modelos Multiníveis Lineares Generalizados (MMLG) são um caso particular do chamado MLGM. Entretanto, (Pinheiro e Bates,2000) afirmam que os modelos multiníveis são uma classe estendida dos MLGM, denotando cada nível como nível hierárquico de efeitos aleatórios. Neste trabalho, é considerando os modelos multiníveis com a mesma a definição adotada por (Pinheiro e Bates,2000). Mais detalhes sobre modelos multiníveis podem ser vistas em (Hox,1995) e (Bryk e Raudenbush,2002).

A importância de incluir efeitos aleatórios nos modelos tem sido estudada por alguns autores que consideram que os omitindo, poderia gerar perda de eficiência apesar do aumento nos erros padrão dos parâmetros estimados, conforme (Carroll,2003).

Considere um MLGM, que é equivalente a um MMLG com 1 nível, com y_{ij} sendo a variável resposta e os vetores $\mathbf{x}_i$ e $\mathbf{z}_i$ de variáveis explicativas, onde

- y_i : variável resposta com $i=1, \dots, n$;
- $\mathbf{x}_i$: covariáveis associadas aos efeitos fixos;
- $\mathbf{z}_i$: covariáveis associadas aos efeitos aleatórios.

O modelo em questão assume as seguintes suposições:

1. A média condicional é relacionada com o preditor linear $\eta_i^u = \mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u}$ pela função de ligação $g(\mu_i^u) = \eta_i^u$, com função inversa $h = g^{-1}$, em que $\boldsymbol{\beta}$ é um vetor p -dimensional de efeitos fixos. Denotando o vetor de observação por $y = (y_1, \dots, y_n)^t$ e a matriz com linhas $\mathbf{x}_i^t$ e $\mathbf{z}_i^t$ por $\mathbf{X}$ e $\mathbf{Z}$, respectivamente, a média condicional satisfaz

$$E(y|\mathbf{u}) = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}), \quad (2.5)$$

em que $\mathbf{u} \sim N(0, \mathbf{D})$ e $\mathbf{D}$ é a matriz de variância-covariância que depende de θ , vetor de componentes de variância desconhecido.

2. Dado um vetor u de efeitos aleatórios, y_{ij} é condicionalmente independente com média e variância dada respectivamente por

- $E(y_i|\mathbf{u}) = \mu_i^u$,
- $\text{Var}(y_i|\mathbf{u}) = \phi a_i v(\mu_i^u)$,

em que $v(\cdot)$ é a função de variância, a_i é uma constante conhecida e ϕ é o parâmetro de dispersão que pode ser conhecido ou não. Note que $\text{Var}(y_i|\mathbf{u})$ possui semelhança com a função de variância do modelo linear generalizado dada por $\text{Var}(y_{ij}) = \phi v(\mu_i^u)$ a menos de a_i . O parâmetro de dispersão considerado para distribuição Binomial vale $\phi=1$.

3. Marginalmente, os efeitos aleatórios são i.i.d. com distribuição $N(0, \sigma^2)$, sendo que a variância é desconhecida.

Analisando a inferência do modelo pelo método clássico, seguindo o procedimento proposto por (Goldstein,1991), (Breslow e Clayton,1993) apresentam o processo de estimação do vetor de parâmetros para os efeitos fixos (β) e para os efeitos aleatórios (u) de um modelo multinível com 2 níveis, utilizando o método da **Quase-Verossimilhança Penalizada (QVP)**. Além deste método de estimação, também é apresentado a estimação da função de log-verossimilhança pela **Quadratura de Gauss-Hermite (QGH)**, conforme (Brostrom,2003). Ambas as técnicas serão apresentadas ainda neste capítulo.

2.2.1 Modelos multiníveis com 2 níveis

O MMLG assume que existe um conjunto de dados hierárquicos com uma única variável resposta, medida no nível mais baixo da hierarquia, e as variáveis explicativas que podem ser observadas em todos os níveis existentes. Conceitualmente, o modelo pode ser visto como um sistema hierárquico de equações de regressão.

Na abordagem dos modelos lineares hierárquicos ou MMLG com 2 níveis, refere-se as unidades de nível 2 como observações no grupo e as unidades de nível 1 como grupo. A quantidade de grupos é denotada por I e o número de observações no grupo, que pode variar de grupo para grupo, é denotado por n_i para cada grupo i ($i = 1, 2, \dots, I$). Neste caso, os dados não precisam ser balanceados, uma vez que o número de observações pode variar de grupo para grupo.

Suponha um modelo linear de 2 níveis, dado por

$$Y_{ij} = a_j + b_j X_{ij} + \varepsilon_{ij}, \quad (2.6)$$

onde a_j , b_j são efeitos aleatórios, ε_{ij} são os erros da j -ésima observação no nível 2 do i -ésimo grupo do nível 1.

Na Figura 2.1, (a) ilustra o modelo e regressão usual; em (b), os interceptos a_i 's variam, mas os coeficientes de regressão b_i 's são considerados constantes e em (c), os interceptos a_i 's e os coeficientes de regressão b_i 's variam de grupo para grupo.

Figura 2.1: Comparativo entre o modelo de regressão usual e o modelo com efeito aleatório.

Suponha que Y é uma variável resposta e X é uma única variável explicativa. O modelo para a análise no nível 2 é definido da seguinte forma:

$$Y_{ij} = \beta_{0i} + \beta_{1i}X_{ij} + \varepsilon_{ij}, \quad (2.7)$$

com $j=1,2,\dots,n_i$ e $i=1,2,\dots,I$ em que:

- Y_{ij} é a variável resposta da j -ésima observação do nível 2 para o i -ésimo grupo do nível 1;
- X_{ij} é a variável explicativa em sua medida original, ou centrada na média geral $\bar{X}_{..}$, ou centrada na média de um grupo $\bar{X}_{i.}$, ou centrada em outro valor, medida no j -ésima observação e agrupado no i -ésimo grupo;
- β_{0i} é o intercepto aleatório para o i -ésimo grupo;
- β_{1i} é o coeficiente de regressão usual para o i -ésimo grupo, em que sua interpretação é dada pela mudança esperada na variável resposta quando X_{ij} aumenta em uma unidade;
- ε_{ij} é o erro aleatório usual associado à j -ésima observação agrupada no i -ésimo grupo em que os erros são independentes e normalmente distribuídos $N(0, \sigma^2)$.

A diferença com o modelo de regressão usual é que neste caso assume-se que cada

grupo é caracterizado por um diferente intercepto e também por um diferente coeficiente de regressão.

A interpretação dos parâmetros do modelo, particularmente os interceptos β_{0i} para todo i , depende da forma como as variáveis explicativas são consideradas no nível 2. Pode-se afirmar que:

- Quando a variável explicativa é considerada em sua medida original (X_{ij}), o intercepto β_{0i} é o valor esperado da variável resposta Y_{ij} quando X_{ij} for igual a zero;
- Quando a variável explicativa é considerada centrada na média geral ($\bar{X}_{..}$), o intercepto β_{0i} é o valor esperado da variável resposta Y_{ij} quando $\bar{X}_{..}$ for igual a zero;
- Quando a variável explicativa é considerada na média de um grupo ($\bar{X}_{i.}$), o intercepto β_{0i} é o valor esperado da variável resposta Y_{ij} quando $\bar{X}_{i.}$ for igual a zero;
- Quando a variável explicativa estiver centrada em qualquer outro valor específico, o intercepto β_{0i} é interpretado como o valor esperado da variável resposta Y_{ij} para esse valor específico.

Considerando a variável resposta Y_{ij} e a variável explicativa X_{ij} , no modelo para o grupo, têm-se diferentes coeficientes de interceptos para β_{0i} 's e inclinações β_{1i} 's com $i=1,2,\dots,I$, iguais aos da equação (2.7). Portanto, os coeficientes para o modelo do nível 1 (grupo) podem ser escritos como

$$\beta_{0i} = \gamma_{00} + u_{0i}, \quad (2.8)$$

$$\beta_{1i} = \gamma_{10} + u_{1i}, \quad (2.9)$$

em que u_{0i} 's e u_{1i} 's são os erros aleatórios independentes com as seguintes suposições:

$$u_{0i} \sim N(0, \sigma_{u0}^2), u_{1i} \sim N(0, \sigma_{u1}^2), Cov(u_{0i}, u_{1i}) = \sigma_{u01}. \quad (2.10)$$

em que:

- (i) u_{0i} 's e u_{1i} 's são independentes de ε_{ij} 's;
- (ii) u_{0i} é o efeito aleatório i -ésimo grupo no intercepto β_{0i} ;
- (iii) u_{1i} é o efeito aleatório associado ao i -ésimo grupo no coeficiente angular ou inclinação β_{1i} ;
- (iv) γ_{00} é o valor esperado dos interceptos nos grupos;
- (v) γ_{10} é o valor esperado das inclinações nos grupos;
- (vi) σ_{u0}^2 é a variância populacional dos interceptos;
- (vii) σ_{u1}^2 é a variância populacional das inclinações;
- (viii) σ_{u01} é a covariância entre β_{0i} e β_{1i} .

Portanto, a partir de (2.8) a (2.10) tem-se que

- $\beta_{0i} \sim N(\gamma_{00}, \sigma_{u0}^2)$,
- $\beta_{1i} \sim N(\gamma_{10}, \sigma_{u1}^2)$,
- $Cov(\beta_{0i}, \beta_{1i}) = \sigma_{u01}$.

Das equações de (2.8) e (2.9), nota-se que os interceptos e as inclinações variam de grupo para grupo para ajudar a levar em conta essas possíveis diferenças, mas os coeficientes γ_{00} e γ_{10} são constantes para todos os grupos.

Pode-se supor que a variação entre as unidades do nível 1 são devidas às variáveis explicativas Z_i , assim, o modelo no nível 1 fica da seguinte forma:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}Z_i + u_{0i}, \quad (2.11)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}Z_i + u_{1i}, \quad (2.12)$$

em que

- γ_{00} é o valor esperado dos interceptos quando $Z_i=0$;
- γ_{10} é o valor esperado das inclinações quando $Z_i=0$;
- γ_{01} é o coeficiente de regressão associado à variável explicativa do nível 1 relativo ao intercepto;
- γ_{11} é o coeficiente de regressão associado à variável explicativa do nível 1 relativo à inclinação;
- u_{0i} e u_{1i} são variáveis aleatórias com as mesmas suposições de (i) a (viii) feitas anteriormente;
- $\sigma_{u_0}^2$ é a variância populacional dos interceptos encontrada após a predição do coeficiente β_{0i} na presença da variável do nível 1, Z_i ;
- $\sigma_{u_1}^2$ é a variância populacional das inclinações encontrada após a estimação do coeficiente β_{1i} na presença da variável do nível 1, Z_i ;
- $\sigma_{u_{01}}$ é a covariância entre β_{0i} e β_{1i} .

No modelo para o nível 1, os β 's são tratados como variáveis respostas latentes em modelos de regressão com variável explicativa Z_i para a população de grupos. Pode-se considerar também Z_i em sua medida original, ou centrada na média geral, $(\bar{Z}_{..})$ ou centrada na média de cada grupo $(\bar{Z}_{i.})$, da mesma forma como visto anteriormente para a variável explicativa no nível 1 da equação(2.7). Substituindo as equações (2.11) e (2.12) em (2.7), tem-se o modelo combinado

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_i + \gamma_{10}X_{ij} + \gamma_{11}Z_iX_{ij} + u_{0i} + u_{1i} + \varepsilon_{ij}. \quad (2.13)$$

O modelo combinado (2.13) tem parte fixa ou determinística $\gamma_{00} + \gamma_{01}Z_i + \gamma_{10}X_{ij} + \gamma_{11}Z_iX_{ij}$, por conter todos os coeficientes fixos, e a parte aleatória ou estocástica é dada por $u_{0i} + u_{1i} + \varepsilon_{ij}$. Note que X_{ij} e Z_i são variáveis explicativas dos níveis 2 e 1, respectivamente,

e $Z_i X_{ij}$ é o termo de interação entre os níveis que aparecem no modelo como consequência de modelar a variação do coeficiente de inclinação β_{1i} da variável X_{ij} do nível 2 com a variável Z do nível 1. Então, o efeito de Z na relação entre X e Y é expressado como *interação de nível cruzado*.

Ainda, tem-se que os erros nas unidades de nível 2 (observações) não são independentes, ou seja, há uma dependência entre as observações agrupadas dentro de cada um dos grupos em termos de u_{0i} e u_{1i} . Além disso, as variâncias dos erros podem ser heterogêneas se u_{0i} e u_{1i} assumirem diferentes valores dentro do grupo.

O grau de dependência dentro dos grupos é medido pelo coeficiente de correlação intra-classe, $\rho(Y_{ij}; Y_{ij'})$. O coeficiente de correlação intra-classe pode ser estimado por meio dos modelos lineares multiníveis a partir do chamado modelo intercepto ou modelo nulo, ou seja, por um modelo que não inclui variáveis explicativas. Este modelo é obtido removendo-se todos os termos que contém as variáveis X ou Z no modelo (2.13). O modelo nulo é da forma

$$Y_{ij} = \gamma_{00} + u_{0i} + \varepsilon_{ij}. \quad (2.14)$$

A partir do modelo (2.14), pode-se obter a partição básica da variabilidade nos dados entre os dois níveis. A variância total de Y pode ser decomposta como a soma das variâncias dos níveis 1 e 2,

$$Var(Y_{ij}) = Var(u_{0i}) + Var(\varepsilon_{ij}) = \sigma_{u0}^2 + \sigma^2.$$

A covariância entre 2 observações diferentes $j \neq j'$ no mesmo grupo i é igual a variância da contribuição de u_{0i} que é compartilhada por estas observações, isto é,

$$cov(Y_{ij}; Y_{ij'}) = var(u_{0i}) = \sigma_{u0}^2 \quad (2.15)$$

e sua correlação intra-classe é dada por

$$\rho(Y_{ij}, Y_{ij'}) = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma^2}. \quad (2.16)$$

A correlação intra-classe é uma estimativa da proporção da variância populacional explicada pela estrutura do agrupamento. Assim, o modelo nulo (2.14) estabelece simplesmente que a correlação intra-classe estimada é igual a proporção estimada da variância no nível grupo, quando comparada com a variância total estimada.

2.2.2 Modelo multinível logístico com 2 níveis

Nesta seção, apresenta-se um caso particular do MMLG considerando 2 níveis para análise de dados com resposta binária (com a função de ligação logit). Neste caso considera-se uma única variável explicativa associada ao efeito fixo β e uma variável explicativa associada ao efeito aleatório u_i .

A resposta binária admite apenas dois resultados: sucesso (1) e fracasso (0). Usualmente, chama-se de sucesso o resultado que se pretende relacionar com as variáveis explicativas. Seja Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$, uma variável aleatória binária com distribuição Bernoulli de parâmetro μ_{ij} , ou seja, tal que as probabilidades são da forma

$$P(Y_{ij} = 0) = 1 - \mu_{ij} \text{ e } P(Y_{ij} = 1) = \mu_{ij}. \quad (2.17)$$

A média e a variância de Y_{ij} são dadas, respectivamente, por

$$E(Y_{ij}) = 1(\mu_{ij}) + 0(1 - \mu_{ij}) = \mu_{ij}$$

e

$$Var(Y_{ij}) = E(Y_{ij}^2) - E^2(Y_{ij}) = [1^2(\mu_{ij}) + 0(1 - \mu_{ij})] - \mu_{ij}^2 = \mu_{ij}(1 - \mu_{ij}).$$

Seja y_{ij} o valor da variável resposta da j -ésima observação do i -ésimo grupo e $u_i \sim N(0; D(\theta))$ com $D(\theta) = \sigma^2$. O MMLG de 2 níveis para resposta binária é definido pela distribuição de probabilidades (2.17) e pelo preditor linear

$$g(\mu_{ij}) = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i, \quad (2.18)$$

em que $g(\cdot)$ é a função de ligação logit (2.4). Seguem as suposições do modelo dado em (2.18):

1. Os y_{ij} 's são variáveis aleatórias independentes Bernoulli. Condicional ao intercepto aleatório u_i , tem-se que

$$P(y_{ij} = 1|u_i) = \mu_{ij},$$

com respostas binárias y_{ij} 's independentes.

2. No caso do modelo de regressão logística tem-se

$$\text{logit}[P(y_{ij} = 1|u_i)] = \log[\mu_{ij}/(1 - \mu_{ij})] = \mathbf{x}_{ij}^t \boldsymbol{\beta} + u_i,$$

em que $\mathbf{x}_{ij}$ é um vetor de covariáveis conhecidas, $\boldsymbol{\beta}$ é um vetor de coeficientes de regressão associado às respectivas covariáveis e u_i é o efeito aleatório do i -ésimo grupo.

3. Marginalmente, $u_1, u_2, \dots, u_I$ são i.i.d. $N(0, \sigma^2)$ com variância desconhecida.

A função de verossimilhança condicional tem a seguinte forma

$$L(\boldsymbol{\beta}|u) = \prod_{i=1}^I \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\beta}, u_i), \quad (2.19)$$

em que f é a função densidade de y_{ij} . A estratégia padrão usual para estimar os parâmetros do modelo $(\boldsymbol{\beta}; \theta)$ é integrar a verossimilhança condicional em relação ao efeito aleatório u_i ,

$$L(\boldsymbol{\beta}|D(\theta)) = \int_{u_i} L(\boldsymbol{\beta}|u) \phi(u_i) du_i \quad (2.20)$$

em que $\phi(\cdot)$ representa a densidade normal, onde $D(\theta) = \sigma^2$.

A função de verossimilhança (2.20) pode ser resolvida de várias formas aproximadas, por exemplo, como as soluções propostas por (Goldstein,1991), (Breslow e Clayton,1993), (Goldstein e Rasbash,1996), (Longford,1987) e (Longford,1994), que utilizam expansões da série de Taylor de 1ª ou 2ª ordem. Neste trabalho serão consideradas duas propostas: uma que estima os parâmetros via QVP (Breslow e Clayton,1993) e outra que aproxima a integral de (2.20) via QGH (Brostrom,2003).

Suponha agora que x_{ij} é uma variável explicativa do nível 2 associada ao efeito β_1 e $z_{ij} = 1$ uma variável explicativa associada ao efeito aleatório u_{0i} . O preditor linear (2.18) com $\mathbf{x}_{ij}^t = (1, x_{ij})$, $z_{ij}^t = 1$ e $g(\cdot) = \text{logit}(\cdot)$ é dado por

$$\text{Logit}(\mu_{ij}) = \beta_0 + \beta_1 x_{ij} + u_{0i}, \quad (2.21)$$

em que

$$\mu_{ij} = P(y_{ij} = 1) = \frac{\exp(\beta_0 + \beta_1 \mathbf{x}_{ij} + u_{0i})}{1 + \exp(\beta_0 + \beta_1 \mathbf{x}_{ij} + u_{0i})},$$

em que u_{0i} é o efeito aleatório do nível 1. Note também que os coeficientes β_0 e β_1 não dependem do grupo i , pois este é um modelo combinado. O modelo (2.21) pode ser, alternativamente, escrito como:

$$\text{Logit}(\mu_{ij}) = \beta_0 + \beta_1 \mathbf{x}_{ij} + u_{0i} = \beta_{0i} + \beta_{1i} \mathbf{x}_{ij},$$

em que $\beta_{0i} = \gamma_{00} + u_{0i}$, $\beta_{1i} = \gamma_{10}$, ou seja, $\gamma_{00} = \beta_0$ e $\gamma_{10} = \beta_1$. Além disso $u_{0i} \sim N(0, \sigma_{u0}^2)$.

Assumindo que, dado o vetor condicional de efeitos aleatórios u_{0i} , as respostas y_{ij} são independentes, a função de verossimilhança condicional para o modelo logístico (2.21) é dada por

$$L(\boldsymbol{\beta}|u) = \prod_{i=1}^I \prod_{j=1}^{n_i} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}}, \quad (2.22)$$

A função de verossimilhança conjunta para ajustar o modelo logístico (2.21) para I grupos independentes, cada um com n_i subunidades, é dada por

$$L(\boldsymbol{\beta}|\mathbf{D}(\theta)) = \prod_{i=1}^I \int \prod_{j=1}^{n_i} \mu_{ij}^{y_{ij}} (1 - \mu_{ij}) \phi(u_{0i}) du_{0i}, \quad (2.23)$$

em que $\phi(\cdot)$ representa a função densidade normal.

2.3 Métodos de estimação

Nesta seção, serão apresentados dois métodos de estimação dos parâmetros de interesse de um modelo logístico multinível.

Os métodos de inferência discutidos neste trabalho terão enfoque na maximização da função de verossimilhança, apesar da possibilidade do ajuste do MMLG ser resolvido via aproximação Bayesiana, vide (Molenberghs e Verbeke,2005). Existem também outras aproximações via Algoritmo de Monte Carlo, segundo (Demidenko,2004) e (McCulloch,1997), que não serão abordados neste trabalho.

Quando a variável resposta não é binária, existem outros métodos de inferência, como por exemplo, o REML (Restricted Maximum Likelihood) para variável resposta com distribuição Gaussiana, conforme (Pinheiro e Bates,2000).

2.3.1 Estimação via QVP

A Quase-Verossimilhança (QV) é muito utilizada por causa de sua eficiência na geração dos estimadores sem a necessidade de se fazer a suposição da distribuição da variável resposta. Este método de estimação não requer a especificação de uma distribuição, somente a média e a variância. Para um MMLG, a função a ser maximizada é a função de Quase-Verossimilhança modificada pela adição de um termo de penalização que depende do efeito aleatório, e por causa disso é chamada de Quase-Verossimilhança Penalizada (QVP). Esta função pode ser vista como um método de inferência aproximada, (Molenberghs e Verbeke,2005).

A função de quase-verossimilhança penalizada usada para estimar (α, θ) é definida por

$$\exp(ql(\alpha, \theta)) \propto |D|^{-\frac{1}{2}} \int \exp \left[-\frac{1}{2\phi} \sum_{i=1}^n d_i(y_i; \mu_i^u) - \frac{1}{2} u^t D^{-1} u \right] du, \quad (2.24)$$

em que

$$d_i(y_i; \mu_i^u) = -2 \int_y^\mu \frac{y-b}{a_i v(b)} db$$

denota o desvio pelo querado em que $ql(\alpha, \theta)$ representa a verdadeira log-verossimilhança dos dados. Se condicionalmente em u , as observações são escritas na forma da família exponencial com função de variância $v(\cdot)$, então a função desvio é conhecida como a diferença

escalar entre $2\phi[l(y; y, \phi) - l(y; \mu, \phi)]$, em que $l(y; \mu, \phi)$ é a verossimilhança de y , dado a média μ .

Se y tiver distribuição normal e a função de ligação for identidade, a equação (2.24) pode ser resolvida de forma fechada. Por outro lado, se a variável resposta y é discreta, a maximização desta expressão deve ser feita por procedimentos que incluam integrações numéricas a cada ciclo da iteração.

Os autores (Breslow e Clayton, 1993) apresentam o processo de estimação do vetor de parâmetros para os efeitos fixos e aleatórios de um modelo multinível de 2 níveis, utilizando o método da QVP, seguindo o procedimento proposto por (Goldstein, 1991) que se apresenta a seguir.

Considere o modelo de 2 níveis com I grupos tendo n_i observações dentro do i -ésimo grupo, $i = 1, \dots, I$, com os efeitos aleatórios (u_i) assumidos independentes entre os grupos. A j -ésima observação para o i -ésimo grupo pode ser escrita como

$$y_{ij} = \mu_{ij}^u + e_{ij} = h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i) + e_{ij}, \quad (2.25)$$

com $j = 1, \dots, n_i$ e $\text{Var}(e_{ij}) = a_j(\phi)v(\mu_{ij}^u)$, em que h é a inversa da função de ligação, $\mathbf{x}_{ij}$ e $\mathbf{z}_{ij}$ são os vetores de regressores associados aos efeitos fixos $(\boldsymbol{\beta})$ e aleatórios $(\mathbf{u}_i)$, respectivamente. Note que, quando é considerado somente o intercepto aleatório, $z_{ij}^t = 1$.

O objetivo é linearizar o modelo para associá-lo com o modelo teórico com distribuição normal, ou seja, a idéia é linearizar a função de ligação inversa, utilizando uma expansão da série de Taylor de 1ª ordem em torno dos valores de $\boldsymbol{\beta}$ e $\mathbf{u}_i$ inicialmente definidos a partir do método de Quase-Verossimilhança Penalizada de primeira ordem (QVP1).

Os passos para a linearização da função de ligação são apresentados abaixo.

(i) QVP1

- **Passo 1:** Expandir h sobre as estimativas atuais de $(\hat{\boldsymbol{\beta}}; \hat{\mathbf{u}}_i)$. A partir da equação (2.25) pode-se escrever a função:

$$f(\boldsymbol{\beta}; \mathbf{u}_i) = h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i), \quad (2.26)$$

as primeiras derivadas da função com relação a $\boldsymbol{\beta}$ e $\mathbf{u}$ são dados por $\frac{\partial f(\boldsymbol{\beta}; \mathbf{u}_i)}{\partial \boldsymbol{\beta}} = h'(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i) \mathbf{x}_{ij}^t$ e $\frac{\partial f(\boldsymbol{\beta}; \mathbf{u}_i)}{\partial \mathbf{u}} = h'(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i) \mathbf{z}_{ij}^t$, respectivamente. Então,

$$\begin{aligned} f(\boldsymbol{\beta}; \mathbf{u}_i) &\approx f(\hat{\boldsymbol{\beta}}; \tilde{\mathbf{u}}_i) + \left. \frac{\partial f(\boldsymbol{\beta}; \mathbf{u}_i)}{\partial \boldsymbol{\beta}} \right|_{(\hat{\boldsymbol{\beta}}; \tilde{\mathbf{u}}_i)} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \left. \frac{\partial f(\boldsymbol{\beta}; \mathbf{u}_i)}{\partial \mathbf{u}_i} \right|_{(\hat{\boldsymbol{\beta}}; \tilde{\mathbf{u}}_i)} (\mathbf{u}_i - \tilde{\mathbf{u}}_i) \\ &\approx \hat{\mu}_{ij}^u + h'(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + h'(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i) \\ &\approx \hat{\mu}_{ij}^u + h'(\hat{\eta}_{ij}^u) [\mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i)]. \end{aligned} \quad (2.27)$$

Substituindo em (2.25) obtém-se

$$y_{ij} \approx \hat{\mu}_{ij}^u + h'(\hat{\eta}_{ij}^u) [\mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i)] + e_{ij}. \quad (2.28)$$

- **Passo 2:** Aplicar a função g nos dados y_{ij} e expandi-la em série de Taylor de 1ª ordem em torno do valor estimado $\hat{\mu}_{ij}^u$.

A aproximação da série de Taylor de 1ª ordem da função g em torno do valor estimado é obtida da seguinte maneira

$$\begin{aligned} g(y_{ij}) &\approx g(\hat{\mu}_{ij}^u) + \left. \frac{\partial g(y_{ij})}{\partial y_{ij}} \right|_{\hat{\mu}_{ij}^u} (y_{ij} - \hat{\mu}_{ij}^u) \\ &\approx g(\hat{\mu}_{ij}^u) + g'(\hat{\mu}_{ij}^u) (y_{ij} - \hat{\mu}_{ij}^u) \\ &\approx \hat{\eta}_{ij}^u + g'(\hat{\mu}_{ij}^u) (y_{ij} - \hat{\mu}_{ij}^u), \end{aligned} \quad (2.29)$$

então, pode-se escrever (2.29) da seguinte forma

$$y_{ij}^* = \mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i + g'(\hat{\mu}_{ij}^u) (y_{ij} - \hat{\mu}_{ij}^u).$$

- **Passo 3:** Associar o modelo linearizado ao modelo teórico com distribuição normal. Denotando $g'(\hat{\mu}_{ij}^u) (y_{ij} - \hat{\mu}_{ij}^u)$ de e_{ij}^* e assumindo y_{ij}^* com distribuição normal, pode-se escrever o modelo multinível linear para y_{ij}^* da seguinte maneira

$$y_{ij}^* = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i + e_{ij}^*, \quad (2.30)$$

em que

$$\mathbf{u}_i \sim N(0, \mathbf{D}(\theta)) \text{ e } e_{ij}^* \sim N(0, [g'(\hat{\theta}_{ij}^u)]^2 a_i(\phi) v(\hat{\mu}_{ij}^u)), \quad (2.31)$$

em que

$$\mathbf{D}(\theta) = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} & \cdots & \sigma_{u0q} \\ \sigma_{u01} & \sigma_{u1}^2 & \cdots & \sigma_{u1q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u0q} & \sigma_{u1q} & \cdots & \sigma_{uq}^2 \end{bmatrix}.$$

As estimativas atualizadas de $(\boldsymbol{\beta}, \mathbf{u}_i, \theta)$ são obtidas resolvendo o modelo multinível linear definido por (2.30) e (2.31) simultaneamente, isto é, utilizando o algoritmo QVP1. Segundo (Breslow e Clayton, 1993), este algoritmo produz estimativas aproximadas das soluções das Equações de Máxima Verossimilhança Restrita (EMVR) para os coeficientes de regressão e para os componentes da variância, no caso do modelo linear Gaussiano. Para os outros modelos este algoritmo produz estimativas aproximadas das soluções das equações de máxima verossimilhança (EMV).

(ii) QVP2

Este algoritmo foi proposto por (Goldstein e Rasbash, 1996) e envolve uma aproximação da série de Taylor de 2ª ordem somente em termos do efeito aleatório u_i .

- **Passo 1:** Expandir h sobre as estimativas atuais de $(\hat{\boldsymbol{\beta}}; \tilde{\mathbf{u}}_i)$. A partir da equação (2.25), pode-se escrever a aproximação da série de Taylor de 2ª ordem da função em torno dos parâmetros estimados da seguinte forma

$$\begin{aligned}
f(\boldsymbol{\beta}, \mathbf{u}_i) &\approx f(\hat{\boldsymbol{\beta}}, \tilde{\mathbf{u}}_i) + \left. \frac{\partial f(\boldsymbol{\beta}, \mathbf{u}_i)}{\partial \boldsymbol{\beta}} \right|_{(\hat{\boldsymbol{\beta}}, \tilde{\mathbf{u}}_i)} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \left. \frac{\partial f(\boldsymbol{\beta}, \mathbf{u}_i)}{\partial \mathbf{u}_i} \right|_{(\hat{\boldsymbol{\beta}}, \tilde{\mathbf{u}}_i)} (\mathbf{u}_i - \tilde{\mathbf{u}}_i) \\
&+ \frac{1}{2} \left. \frac{\partial^2 f(\boldsymbol{\beta}, \mathbf{u}_i)}{\partial \mathbf{u}_i \partial \mathbf{u}_i^t} \right|_{(\hat{\boldsymbol{\beta}}, \tilde{\mathbf{u}}_i)} (\mathbf{u}_i - \tilde{\mathbf{u}}_i)(\mathbf{u}_i - \tilde{\mathbf{u}}_i)^t \\
&\approx \hat{\mu}_{ij} + h'(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + h'(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i) \\
&+ \frac{1}{2} h''(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i)(\mathbf{u}_i - \tilde{\mathbf{u}}_i)^t \mathbf{z}_{ij} \\
&\approx \hat{\mu}_{ij} + h'(\hat{\eta}_{ij}^u) [\mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i)] + \frac{1}{2} h''(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{z}_{ij}^2 (\mathbf{u}_i - \tilde{\mathbf{u}}_i)^2.
\end{aligned} \tag{2.32}$$

Os autores (Goldstein e Rasbash, 1996) sugerem ignorar o produto cruzado envolvendo diferentes componentes de $\mathbf{u}_i$, adicionando o termo quadrático como um *offsets*, isto é, um termo conhecido que está fora do processo de estimação. Substituindo (2.32) em (2.25) obtém-se

$$\begin{aligned}
y_{ij} &\approx \hat{\mu}_{ij} + h'(\hat{\eta}_{ij}^u) [\mathbf{x}_{ij}^t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{z}_{ij}^t (\mathbf{u}_i - \tilde{\mathbf{u}}_i)] \\
&+ \frac{1}{2} h''(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t \tilde{\mathbf{u}}_i) \mathbf{z}_{ij}^2 (\mathbf{u}_i - \tilde{\mathbf{u}}_i)^2 + e_{ij}.
\end{aligned}$$

- **Passo 2:** Aplicar a função g nos dados y_{ij} e expandi-la em série de Taylor de 2ª ordem em volta do valor estimado $\hat{\mu}_{ij}^u$. A aproximação da série de Taylor de 2ª ordem da função em torno do valor estimado é construída da seguinte maneira

$$\begin{aligned}
g(y_{ij}) &\approx g(\hat{\mu}_{ij}^u) + \left. \frac{\partial g(y_{ij})}{\partial y_{ij}} \right|_{(\hat{\mu}_{ij}^u)} (y_{ij} - \hat{\mu}_{ij}^u) + \frac{1}{2} \left. \frac{\partial^2 g(y_{ij})}{\partial y_{ij} \partial y_{ij}^t} \right|_{(\hat{\mu}_{ij}^u)} (y_{ij} - \hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u)^t \\
&\approx g(\hat{\mu}_{ij}^u) + g'(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u) + \frac{1}{2} g''(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u)^2 \\
&\approx \hat{\eta}_{ij}^u + g'(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u) + \frac{1}{2} g''(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u)^2
\end{aligned} \tag{2.33}$$

e então, pode-se escrever (2.33) da seguinte forma

$$y_{ij}^* \approx \hat{\eta}_{ij}^u + g'(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u) + \frac{1}{2} g''(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u)^2.$$

- **Passo 3:** Associar o modelo linearizado ao modelo teórico com distribuição normal. Denotando $g'(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u) + \frac{1}{2} g''(\hat{\mu}_{ij}^u)(y_{ij} - \hat{\mu}_{ij}^u)^2$ por e_{ij}^* e assumindo que y_{ij}^* é

distribuído normalmente, o modelo multinível linear para y_{ij}^* tem a mesma forma da equação (2.30), em que

$$\mathbf{u}_i \sim N(0, \mathbf{D}(\theta)) \quad \text{e} \quad e_{ij}^* \sim N(0, [g'(\hat{\mu}_{ij}^u)]a_i(\phi)v(\hat{\mu}_{ij}^u) - \frac{1}{2}[g''(\hat{\mu}_{ij}^u)]^2 a_i^2(\phi)v^2(\hat{\mu}_{ij}^u)). \quad (2.34)$$

As estimativas atualizadas de $(\boldsymbol{\beta}, \mathbf{u}_i, \theta)$ são obtidas resolvendo-se as equações (2.33) e (2.34) simultaneamente, que agora leva em consideração a aproximação da série de Taylor de 2ª ordem, isto é, utilizando o algoritmo QVP2. Este algoritmo também melhora as estimativas dos componentes da variância, mas (Goldstein,1999) adverte, que quando se trabalha com modelos complexos e/ou com conjuntos de dados pequenos, surgem problemas de convergência. Em função disso, geralmente, utiliza-se o QVP1.

2.3.2 Estimação via QGH

A integral descrita em (2.20) pode ser resolvida usando integração numérica ou quadratura de Gauss-Hermite (QGH).

A quadratura é o caso mais simples da avaliação de uma integral dada por

$$I = \int_a^b f(x)dx, \quad (2.35)$$

que é equivalente a resolver $I \equiv y(b)$, a equação diferencial $\frac{dy}{dx} = f(x)$ com $y(a) = 0$, com a e b sendo números reais.

O método de quadratura é baseado na avaliação da função f numa sequência de pontos, pertencentes ao intervalo de integração denotados por $x_0, x_1, \dots, x_N, x_{N+1}$, que são espaçados por uma constante h ,

$$x_i = x_0 + ih,$$

com $i = 0, 1, \dots, N, N + 1$. Logo, a função $f(x)$ tem valores conhecidos de x_i 's,

$$f(x_i) \equiv f_i.$$

Deseja-se integrar a função $f(x)$ entre o limite inferior a e o limite superior b , em que a e b são iguais aos valores dos x_i 's.

Esta função pode ser multiplicada por um peso w_i que pode ser interpretada como uma média p.m. desta função na forma

$$\int_{-\infty}^{\infty} f(x_i)w_i dx. \quad (2.36)$$

A função integral dada em (2.36) é aproximada pela soma de seus valores funcionais do conjunto de pontos igualmente espaçados e multiplicados por certos coeficientes (pesos).

Já definida o que é uma quadratura, pode-se definir o que é a quadratura de Gauss. Em muitas aplicações, a densidade normal é um fator explícito da integral e por isso ela é frequentemente usada na área de estatística, ver (Lange,1998), (Liu e Pierce,1994), (Nia,2006) e (Press,1994).

Seja $w(x)$ uma função de pesos fixados em um intervalo $[a, b]$ (que pode ser finito ou infinito). A integral

$$\int_a^b w(x)f(x)g(x)dx = (f, g)$$

é conhecida como o *produto interno* das funções $f(x)$ e $g(x)$ sob um intervalo $[a, b]$ com respeito ao peso $w(x)$.

As funções f e g são chamadas *ortogonais* sob o intervalo $[a, b]$ com respeito ao peso $w(x)$ se

$$\int_a^b w(x)f(x)g(x)dx = 0.$$

Para um dado peso $w(x)$, é possível definir uma sequência de polinômios $p_0(x), p_1(x), \dots, p_n(x)$ que são ortogonais e no qual cada $p_n(x)$ tem o exato grau n , isto é:

$$(p_m, p_n) = \int_a^b w(x)p_m(x)p_n(x)dx = 0,$$

com $m \neq n$.

Multiplicando-se cada $p_n(x)$ por uma constante apropriada, pode-se produzir um conjunto de polinômios $p_n^*(x)$ que são *ortonormais*, isto é:

$$(p_m^*, p_n^*) = \int_a^b w(x) p_m^*(x) p_n^*(x) dx = \delta_{mn} = \begin{cases} 0, & \text{se } m \neq n, \\ 1, & \text{se } m = n. \end{cases}$$

Se os n pontos distintos de $x_0, x_1, \dots, x_n$ do intervalo $[a, b]$ são especificados a priori, então é necessário definir os coeficientes $w_0, w_1, \dots, w_n$ tal que a regra

$$\int_{-\infty}^{\infty} w(x) f(x) dx = \sum_{i=1}^n w_i f(x_i) \quad (2.37)$$

será exata para todo polinômio da classe P_{n-1} , isto é, todas as combinações lineares da potência n com $1, x, x^2, \dots, x^{n-1}$. Se considerar os x 's e os w 's como $2n$ valores desconhecidos e determiná-los com cautela, tem-se um polinômio da classe P_{2n-1} , ou seja, todas as combinações lineares da potência $2n$ com $1, x, x^2, \dots, x^{2n-1}$.

Hermite é caso particular em que os parâmetros da quadratura gaussiana definida por (2.37) são dados por:

- $w(x) = e^{-x^2}$;
- o intervalo da integral é $(-\infty, +\infty)$.

Logo, a Quadratura de Gauss-Hermite é definida por

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx = \sum_{i=1}^n w_i f(x_i). \quad (2.38)$$

Quando a densidade normal não é um fator do integrando, o integrando original deve ser dividido e multiplicado por $\exp(-x^2)$. Deseja-se aplicar a QGH para aproximar uma integral na forma:

$$\int_{\mathbb{R}} h(x) dx,$$

em que $h(x) > 0$ é uma função suave e unimodal.

A quadratura de Gauss-Hermite pode ser reescrita em termos da densidade normal como:

$$\int_{\mathbb{R}} f(x)\phi(x; \mu, \sigma)dx \approx \sum_{i=1}^N w_i f(z_i) \quad (2.39)$$

em que $\phi(x; \mu, \sigma)$ é uma densidade normal arbitrária. Os pontos para avaliação são $z_i = \mu + \sqrt{2}\sigma x_i$ em que os pesos são modificados de w_i para $w_i/\sqrt{\pi}$, com $i = 1, \dots, N$ número de pontos de quadratura. Esta aproximação é chamada de Quadratura Clássica. Algumas vezes é necessário considerar muitos pontos para aproximar uma integral porque as abcissas são amostradas em uma região inapropriada.

Usando $\phi(x; \hat{\mu}, \hat{\sigma})$, a aproximação é dada por

$$\int g(x)dx \approx \sqrt{2}\hat{\sigma} \sum_{i=1}^N w_i^* g(\hat{\mu} + \sqrt{2}\hat{\sigma}x_i) \quad (2.40)$$

em que $w_i^* = w_i + \exp(x^2)$.

Quando a equação (2.40) é aplicada em somente um ponto, o resultado é a aproximação de Laplace, conforme (Liu e Pierce,1994). Na presente dissertação utilizou-se a abordagem proposta na equação (2.41) nos estudos de simulação e aplicação

$$\int g(x)dx \approx \sqrt{2\pi}\hat{\sigma}g(\hat{\mu}). \quad (2.41)$$

2.3.3 Aspectos computacionais

A função que permite o ajuste dos MMLG utilizando o método de Quase-Verossimilhança Penalizada no software R (Venables e Ripley,2002) é **glmmPQL** que requer a biblioteca *MASS*. Segue um esquema do que este comando faz:

- Ajusta um MMLG somente com o intercepto aleatório e uma covariável aleatória, usando Quase-Verossimilhança Penalizada;
- Aceita diferentes distribuições de respostas e funções de ligação dos MLG;
- Chama a função **lme** da biblioteca *nlme*;

- Permite diferentes estruturas de correlação;
- Usa o estimador **QVP1**.

A função que permite o ajuste dos MMLG utilizando o método de Quadratura de Gauss-Hermite no software R utiliza a função (Brostrom,2003) **glmmML**, que requer a biblioteca *glmmML*. Segue um esquema do que este comando faz:

- Ajusta um MMLG somente com o intercepto aleatório por Máxima Verossimilhança e integração numérica via QGH clássica (não permite covariável com efeito aleatório);
- Aceita somente as distribuições Binomial e Poisson;
- Aceita somente as funções de ligação complemento log-log e logit;
- Chama a função **optim** da biblioteca *MASS*;
- Não permite estruturas de correlação;
- Usa o método de Newton-Rapson para finalizar o procedimento de maximização.

Como já mencionado, o modelo de regressão logística multinível usualmente envolve integrais complexas com muitas dimensões. Existem outras técnicas para reduzir este problema, como por exemplo, podem ser utilizados dois Métodos de Monte Carlo: um deles é baseado na integração de Monte Carlo direta e o outro é baseado na versão estocástica da Quadratura de Gauss-Hermite, ver (Chen,2005).

Capítulo 3

Predição da variável resposta

3.1 Introdução

O objetivo principal deste trabalho é apresentar uma proposta para fazer a predição para observações de novos grupos, utilizando modelos multiníveis.

No modelo usual de regressão logística, que é composto apenas pela parte fixa, fazer a predição é simples, pois, através da função de ligação, é possível obter a predição de novos valores da variável resposta.

Entretanto, para o modelo multinível, além dos efeitos fixos, também existem os efeitos que são estimados para cada grupo da amostra utilizada para modelagem. Neste caso, quando é necessário fazer a predição de novos valores da variável resposta de uma amostra diferente da utilizada na modelagem, não se tem os coeficientes dos efeitos aleatórios. Consequentemente, não é possível fazer a predição da maneira usual.

Logo, a proposta deste trabalho é utilizar a parte fixa do modelo em conjunto com uma estimativa do desvio padrão do efeito aleatório para fazer a predição de observações de novos grupos através de uma proposta que será descrita na próxima seção deste capítulo.

Considere um MMLG com dois níveis, denotando y_{ij} como a variável resposta $y_{i1}, \dots, y_{in_i}$ que correspondem que a j -ésima observação do i -ésimo grupo, em que $i = 1, \dots, I$ e $j = 1, \dots, n_i$. A função densidade é dada por

$$f(y_{ij}|\alpha_i) = \exp \left[\left(\frac{a_{ij}}{\phi} \right) (y_{ij}\mu_{ij} - b(\mu_{ij})) + c \left(y_{ij}, \left(\frac{a_{ij}}{\phi} \right) \right) \right], \quad (3.1)$$

em que $b(\cdot)$ e $c(\cdot, \cdot)$ são funções associadas à família exponencial, ϕ é o parâmetro de dispersão e a_{ij} é um peso. Além disso, μ_{ij} está associado à função linear, tal que $a_{ij}=1$ quando a resposta de cada observação individual é considerada, $a_{ij}=l_{ij}$ é o tamanho do grupo quando a média é considerada como resposta e $a_{ij}=l_{ij}^{-1}$ quando a soma das respostas individuais é considerada. Além disso, μ_{ij} está associado à uma função linear

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i,$$

através de uma função de ligação $g(\cdot)$, isto é, $g(\mu_{ij}) = \eta_{ij}$, em que $h = g^{-1}$. Os valores $\mathbf{x}_{ij}$ e z_{ij} são conhecidos, $\boldsymbol{\beta}$ é um vetor desconhecido contendo os coeficientes da regressão e $\alpha_i = (\alpha_{ij})_{1 \leq j \leq n_i}$ é o efeito aleatório do i -ésimo grupo. Finalmente, suponha $\alpha_1, \dots, \alpha_n$ são independentes com densidade $f_{\vartheta}(\cdot)$, em que ϑ é um vetor de componentes de variância. Seja $\psi = (\boldsymbol{\beta}', \vartheta')$ representando o vetor de parâmetros desconhecidos e $\varphi = (\psi', \phi)$. Note que nos casos das distribuições *Binomial* e *Poisson*, o parâmetro de dispersão ϕ é conhecido.

3.2 Predição da variável resposta

A predição do efeito aleatório, em um contexto geral, tem uma longa história, iniciando com um problema de criação de animais. O mais conhecido método para predição do efeito aleatório é o BLUP (BLUP: Best Linear Unbiased Predictor), no caso do modelo linear Gaussiano (Pinheiro e Bates, 2000).

Define-se como o melhor preditor (BP: Best Predictor) para η o $\tilde{\eta}$, se o efeito aleatório é expressado no preditor linear η e as componentes de variância são conhecidas, quando $\tilde{\eta}$ é o η com α trocado por $E(\alpha|y)$. Logo, $\tilde{\eta}$ é o chamado melhor preditor (BP) de η no sentido de minimizar o quadrado médio do resíduo da predição (MSPE: Mean Squared Prediction Error) de um preditor linear que é definido por

$$MSPE(\hat{\varsigma}) = E(\hat{\varsigma} - \varsigma)^2,$$

mais detalhes em (Jiang e Lahiri, 2006).

Considere o problema da predição da seguinte função:

$$\varsigma = \varsigma(\boldsymbol{\beta}, \alpha_S),$$

em que S é um subconjunto de $1, \dots, I$, e $\alpha_S = (\alpha_i)_{i \in S}$. Seja $y_S = (y_i)_{i \in S}$, em que $y_i = (y_{ij})_{1 \leq j \leq n_i}$. Sob o modelo em questão, o BP de ς , no sentido de minimizar a MSPE, é definido como:

$$\begin{aligned} \hat{\varsigma} &= E(\varsigma|y) = E(\varsigma(\boldsymbol{\beta}, \alpha_S)|y_S) = \int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot f(\alpha_S|y_S) d\alpha_S \\ &= \int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot \left(\frac{f(\alpha_S, y_S)}{f(y_S)} \right) d\alpha_S = \int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot \left(\frac{f(y_S|\alpha_S) \cdot f_{\vartheta}(\alpha_S)}{f(y_S)} \right) d\alpha_S \\ &= \frac{\int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot f(y_S|\alpha_S) \cdot f_{\vartheta}(\alpha_S) d\alpha_S}{f(y_S)}, \end{aligned}$$

considerando

$$f(y_S) = \int f(y_S, \alpha_S) d\alpha_S = \int \frac{f(y_S, \alpha_S)}{f_{\vartheta}(\alpha_S)} f_{\vartheta}(\alpha_S) d\alpha_S = \int f(y_S|\alpha_S) f_{\vartheta}(\alpha_S) d\alpha_S,$$

tem-se que

$$= \frac{\int \varsigma \cdot f(y_S|\alpha_S) \cdot f_{\vartheta}(\alpha_S) d\alpha_S}{\int f(y_S|\alpha_S) \cdot f_{\vartheta}(\alpha_S) d\alpha_S}, \quad (3.2)$$

em que,

$$\begin{aligned} f(y_S|\alpha_S) &= \left\{ \prod_{i \in S} \prod_{j=1}^{n_i} f(y_{ij}|\alpha_i) \right\} \\ f_{\vartheta}(\alpha_S) &= \prod_{i \in S} f_{\vartheta}(\alpha_i). \end{aligned}$$

Substituindo em (3.2), obtém-se:

$$\frac{\int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot \exp \left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi} \right) \left[y_{ij} \mu_{ij} - b(\mu_{ij}) + c \left(y_{ij}, \frac{\phi}{a_{ij}} \right) \right] \right\} \right) \cdot \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i}{\int \exp \left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi} \right) \left[y_{ij} \mu_{ij} - b(\mu_{ij}) + c \left(y_{ij}, \frac{\phi}{a_{ij}} \right) \right] \right\} \right) \cdot \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i} \quad (3.3)$$

em que $\mu_{ij} = h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i)$. De (3.3), tem-se:

$$\frac{\int \varsigma(\boldsymbol{\beta}, \alpha_S) \cdot \exp \left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi} \right) \left[y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) - b(h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i)) + c \left(y_{ij}, \frac{\phi}{a_{ij}} \right) \right] \right\} \right) \exp \left(c \left(y_{ij}, \frac{\phi}{a_{ij}} \right) \right) \cdot \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i}{\int \exp \left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi} \right) \left[y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) - b(h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i)) + c \left(y_{ij}, \frac{\phi}{a_{ij}} \right) \right] \right\} \right) \cdot \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i}.$$

Note que $\exp\left(c\left(y_{ij}, \frac{\phi}{a_{ij}}\right)\right)$ não depende de α_i e ele existe tanto no numerador quanto no denominador. Logo, pode-se cancelar os respectivos termos da fração e a expressão anterior fica:

$$= \frac{\int \varsigma(\boldsymbol{\beta}, \alpha_S) \exp\left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi}\right) [y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) - b(h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i))] \right\}\right) \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i}{\int \exp\left(\sum_{i \in S} \sum_{j=1}^{n_i} \left\{ \left(\frac{a_{ij}}{\phi}\right) [y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) - b(h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i))] \right\}\right) \prod_{i \in S} f_{\vartheta}(\alpha_i) d\alpha_i}.$$

Assim, considerando $S=\{i\}$, tem-se que:

$$\hat{\varsigma} = \frac{\int \varsigma(\boldsymbol{\beta}, \alpha_i) \exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)) f_v(\alpha_i) d\alpha_i}{\int \exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)) f_v(\alpha_i) d\alpha_i}, \quad (3.4)$$

em que

$$S_i(\boldsymbol{\beta}, \alpha_i) = \sum_{j=1}^{n_i} a_{ij} (y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) - (\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i)) = \sum_{j=1}^{n_i} a_{ij} (y_{ij} \mu_{ij} - b(\mu_{ij})),$$

ou seja,

$$\hat{\varsigma} = \frac{E(\varsigma(\boldsymbol{\beta}, \alpha_i) \cdot \exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)))}{E(\exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)))}.$$

Note que o BP depende de y_S e ς . Desde que ς seja usualmente desconhecida, costuma-se trocar ς por um estimador consistente $\hat{\varsigma}$. Então, o resultado é chamado de melhor preditor empírico ou EBP (EBP: Empirical Best Predictor).

3.2.1 Predição da variável resposta: modelo logístico multinível

Para variável resposta binária, a proposta deste trabalho é modificar o EBP apresentado no artigo de (Jiang e Lahiri, 2006) para fazer a predição dos valores da variável resposta para observações dentro de um novo grupo. O EBP de $\hat{\alpha}_i = u_i(y_i, \hat{\psi})$, em que y_i é uma constante e $\psi = (\boldsymbol{\beta}^t, \sigma^t)$.

Considere o modelo logístico multinível com resposta binária y_{ij} , independentes, condicional ao efeito aleatório α_i , definido por

$$g(\mu_{ij}) = \text{logit}(P_{ij}) = \text{logit}(P(y_{ij} = 1 | \alpha_i)) = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \alpha_i,$$

em que $\text{logit}(P_{ij}) = \log(\frac{P_{ij}}{1-P_{ij}})$, $\mathbf{x}_{ij}$ um vetor de covariáveis desconhecidas e $\boldsymbol{\beta}$ um vetor conhecido de coeficientes de regressão, $\alpha_i \sim N(0, \sigma^2)$.

A distribuição *Binomial* pertence à família exponencial, podendo ser escrita na forma de (3.1). Para a distribuição *Binomial*, $b(\mu_{ij}) = \log(1 + \exp(\mu_{ij}))$, $\phi = n_i$ e $P_{ij} = h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i) = \frac{\exp\{\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i\}}{1 + \exp\{\mathbf{x}_{ij}^t \boldsymbol{\beta} + z_{ij}^t \alpha_i\}}$.

No artigo de (Jiang e Lahiri, 2006), é proposto como fazer a predição do efeito aleatório para um novo grupo por meio de (3.4). A nossa proposta é fazer a predição da probabilidade de uma observação, dentro de um novo grupo, ter resposta 1 (esta é a contribuição principal desta dissertação).

O interesse é particularizar a equação (3.4) para fazer a predição de um novo valor da variável resposta. Para isso, considere agora o caso particular em que: $a_{ij} = 1$, logo $\sum_{j=1}^{n_i} a_{ij} = n_i$, $y_i. = n_i./2$ e $z_{ij} = 1$ é o intercepto aleatório. Tem-se então

$$\varsigma(\boldsymbol{\beta}, \alpha_i) = P_{ij} = h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \alpha_i) = P_{ij} = \frac{\exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \sigma \xi)}{1 + \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \sigma \xi)},$$

com $\xi \sim N(0, 1)$.

Considerando que o BP para ς é $\hat{\varsigma}$ dado em (3.4), tem-se que

$$\hat{P}_{ij} = \frac{\int P_{ij}(\boldsymbol{\beta}, \alpha_i) \exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)) f_v(\alpha_i) d\alpha_i}{\int \exp(\phi^{-1} S_i(\boldsymbol{\beta}, \alpha_i)) f_v(\alpha_i) d\alpha_i} \quad (3.5)$$

com $S_i(\boldsymbol{\beta}, \alpha_i) = S_i(\boldsymbol{\beta}, \sigma \xi) = \sum_{j=1}^{n_i} [y_{ij} h(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \sigma \xi) - \log(1 + \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \sigma \xi))]$.

Substituindo σ e β_0 por $\hat{\sigma}$ e $\hat{\boldsymbol{\beta}}$, respectivamente, usando a igualdade $h_{ij}(\hat{\boldsymbol{\beta}}, \hat{\sigma} \xi) = \frac{\exp(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)}{1 + \exp(\mathbf{x}_{ij}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)}$ e fixando $j = k$, o melhor predito empírico (EBP) é dado por:

$$\begin{aligned} \hat{h}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\sigma} \xi) &= \frac{E\left(\frac{\exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)}{1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)} \cdot \exp(y_i \hat{\sigma} \xi - \sum_{k=1}^{n_i} \log(1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)))\right)}{E(\exp(y_i \hat{\sigma} \xi - \sum_{k=1}^{n_i} \log(1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi))))} \\ &= \frac{\exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}}) \cdot E\left(\frac{\exp(\hat{\sigma} \xi)}{1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)} \cdot \exp(y_i \hat{\sigma} \xi) \cdot \exp(-\sum_{k=1}^{n_i} \log(1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi)))\right)}{E(\exp(y_i \hat{\sigma} \xi - \sum_{k=1}^{n_i} \log(1 + \exp(\mathbf{x}_{ik}^t \hat{\boldsymbol{\beta}} + \hat{\sigma} \xi))))} \end{aligned} \quad (3.6)$$

em que as esperanças são consideradas a respeito de $\xi \sim N(0, 1)$.

Logo, a probabilidade estimada da j -ésima observação do i -ésimo grupo (novo) apresentar o evento 1, a partir de um modelo logístico multinível, é dada por (3.6).

Note que, como o objetivo é fazer a predição da variável resposta, não existe o valor verdadeiro de y_i . e por isso foi assumido que o mesmo vale $n_{i.}/2$, ou seja, como não se sabe quais observações dentro do grupo apresentaram o evento 1, foi assumido com probabilidade 50% que o evento 1 pode ocorrer. Foram feitos estudos com diferentes valores para y_i . e, através de estudos de simulação, detectou-se que esta informação influencia somente na escala da variável resposta, isto é, caso fosse ordenado as probabilidades estimadas através da equação (3.6), independente do valor de y_i , a posição dentre as probabilidades estimadas das observações se mantêm exatamente a mesma.

3.2.2 Medidas de desempenho

Quando se constrói um modelo com o objetivo de fazer predição, é muito importante avaliar o seu desempenho, tanto na base de modelagem (ou construção) quanto na base de predição (ou validação). A seguir, serão apresentadas algumas medidas de desempenho comuns na literatura para avaliar a qualidade da predição dos modelos de regressão logística.

Na Tabela 3.1, pode-se observar o que será chamada de tabela de classificação do modelo. Através dela, é possível analisar a classificação correta através da resposta binária observada versus a resposta binária estimada no ponto de corte. O ponto de corte é usualmente definido como aquele que maximiza o acerto de cada um dos eventos da variável resposta. Quando a variável resposta é 0, o evento é classificado como negativo e, quando vale 1, como positivo. Abaixo, segue a definição das caselas desta tabela:

- VP (verdadeiro-positivo): quantidade de observações que o modelo classificou como positivo quando, de fato, elas são positivas;
- VN (verdadeiro-negativo): quantidade de observações que o modelo classificou como

negativo quando, de fato, elas são negativas;

- *FP* (falso-positivo): quantidade de observações que o modelo classificou como positivo quando, de fato, elas são negativas;
- *FN* (falso-negativo): quantidade de observações que o modelo classificou como negativo quando, de fato, elas são positivas.

		Valor Predito		Total
		0	1	
Valor Observado	0	VN	FP	VN+FP
	1	FN	VP	FN+VP
Total		VN+FN	FP+VP	VN+FN+ FP+VP

Tabela 3.1: Tabela de classificação - observado versus predito.

A medida *acurácia* mede o acerto geral do modelo e é dada por:

$$Acuracia = \frac{VP+VN}{VN+FP+FN+VP}.$$

A *sensibilidade* mede o percentual de acerto dentro do evento de interesse, ou seja, indica o quão bom é o modelo para identificar as observações com resposta 1. Enquanto a *especificidade* mede o acerto dentro do evento complementar, isto é, indica o quão bom é o modelo para identificar as observações com resposta 0. Segue abaixo como são calculados esses dois índices:

$$Sensibilidade = \frac{VP}{FN+VP} \quad e \quad Especificidade = \frac{VN}{VN+FP}.$$

Observe que essas medidas variam entre 0 e 1 de tal forma que um modelo com boa predição tem valores próximos de 1. Na prática, é muito difícil encontrar modelos com acerto igual a 1, comumente modelos com índices iguais ou superiores a 0,8 já são considerados muito bons.

Uma outra medida muito utilizada é a estatística não paramétrica de Komolgorov-Smirnov (KS). Esta estatística pode ser calculada através da combinação da *sensibilidade* com a *especificidade* da seguinte maneira:

$$KS = |Sensibilidade - (1 - Especificidade)|. \quad (3.7)$$

Note que KS pode variar entre 0 e 1 e que quanto maior o seu valor, melhor o poder discriminatório. Um KS inferior a 0,2 é considerado como predição não satisfatória. Valores de KS entre 0,2 e 0,6 são indicadores aceitáveis de predição, sendo que, quanto maior o KS, melhor o desempenho do modelo. Na prática, é muito raro encontrar valores de KS superiores à 0,6.

Todas as medidas aqui apresentadas serão utilizadas na avaliação dos modelos de regressão logística usual e multinível nos próximos capítulos.

Capítulo 4

Simulação

Neste capítulo, serão comparados os comportamentos dos estimadores da regressão logística multinível, utilizando os métodos de estimação QVP e QGH, e da regressão logística usual. Neste caso avaliou-se:

- Se o valor estimado de β pela regressão logística usual e multinível são próximos do valor verdadeiro selecionado ($\beta = 1$);
- Se o valor do desvio padrão de α ($DP(\alpha)$) para regressão logística multinível é próximo do valor verdadeiro para os métodos QVP e QGH.

Também é de interesse comparar os modelos de regressão logística usual e multinível em relação às medidas de desempenho *acurácia*, *sensibilidade*, *especificidade* e *KS*, tanto na base de modelagem quanto na base de predição.

Segue o modelo multinível ajustado no estudo de simulação:

Y é uma variável resposta binária, X é uma única variável explicativa com distribuição normal de média zero e variância 1. O modelo para a análise no nível 2 é definido da seguinte forma:

$$\text{logit}[P(Y_{ij} = 1|\alpha_i)] = \beta_{0i} + \beta_1 X_{ij}, \quad (4.1)$$

com $j=1,2,\dots,g_i$ e $i=1,2,\dots,N/2$, em que:

- β_{0i} é o intercepto aleatório para o i -ésimo grupo;
- β_1 é o coeficiente de regressão usual,

em que $\beta_{0i} = \gamma_{00} + u_{0i}$. Além disso, $u_{0i} \sim N(0, [DP(\alpha)]^2)$, em que $u_{0i} = \alpha_i$.

Foram fixados os seguinte valores para as simulações:

- N : quantidade amostral, assumindo os valores 200 e 1000;
- g : quantidade de observações dentro do grupo, assumindo os valores 2, 10, 25 e 50;
- $DP(\alpha)$: valor do desvio padrão do efeito aleatório, assumindo os valores 0.1, 1 e 2.

4.1 Esquema de simulação

O esquema da simulação foi desenvolvido no software R, e seu código está descrito no Apêndice A.

A seguir, são descritos os passos simulação:

1. Gera-se aleatoriamente os valores da covariável $X \sim N(0, 1)$, por grupo, contento g observações;
2. Gera-se aleatoriamente um valor do efeito aleatório α_i para cada grupo com $X \sim N(0, DP(\alpha))$, replicando o valor gerado para as g observações dentro do grupo;
3. Usando o valor fixado $\beta=1$, o valor gerado x_{ij} no Passo 1 e o valor gerado α_i no Passo 2, para cada um dos casos, calcula-se a probabilidade através da função logit, dada por:

$$prob = \frac{\exp\{1 * x_{ij} + \alpha_i\}}{1 + \exp\{1 * x_{ij} + \alpha_i\}} \quad (4.2)$$

4. Gera-se aleatoriamente a variável resposta com distribuição Binomial usando a probabilidade calculada no passo 3. Após atribuído 0 ou 1 às obsevações, calcula-se o % de respostas 1 dentre as $N/2$ primeiras observações;
5. Separa-se a base em duas partes. Os primeiros $N/2$ casos foram utilizados para ajustar os modelos de regressão logística usual, regressão logística multinível ajustado via QVP e

via QGH, através dos comandos **lm**, **glmmPQL** e **glmmML**, respectivamente. O restante da base foi usado para a predição da variável resposta, ou seja, a probabilidade da variável resposta ser 1 a predição da regressão logística usual, foi usada a função logito, descrita no passo 3, usando o valor do β estimado e o valor da covariável X em cada caso. Para os modelo de regressão logística multinível foi utilizada a expressão (3.6) proposta no Capítulo 3, que considera, além do valor do β estimado e da covariável X , o desvio padrão do efeito aleatório estimado para cada caso;

6. Ao final da rodada de cada simulação, calcula-se, para cada observação da amostra, o módulo da diferença entre a probabilidade verdadeira e a probabilidade estimada, tanto para a base de modelagem quanto para a de predição;
7. Através do percentual de casos igual a 1, encontrado no passo 4, atribui-se valor 1 para as probabilidades que ficaram acima desse percentual e zero, caso contrário. Neste caso, classificam-se as probabilidade estimadas nos eventos 0 ou 1, conforme ponto de corte estipulado no Passo 4;
8. O procedimento descrito até o passo 7 foi repetido 1000 vezes.

Após terminado de rodar as 1000 simulações, os passos seguintes foram:

- i. Obter a média aritmética dos β 's estimados para as três metodologias e também uma média dos $DP(\alpha)$'s estimados para o modelo multinível;
- ii. Construir a tabela de classificação da variável resposta observada versus predito, para os três casos, tanto na base de modelagem quanto na base de predição;
- iii. Através desta tabela, calcula-se os índices de desempenho *acurácia*, *sensibilidade*, *especificidade* e *KS*. Os resultados podem ser encontrados nas Tabelas 4.1 e 4.2, para $N=200$ e $N=1000$, respectivamente;
- iv. Os resíduos, tanto na base de modelagem quanto na base de predição, foram obtidos no Passo 6 para cada um dos modelos. Para avaliar sua distribuição, gera-se medidas resumo (mínimo, máximo, primeiro quartil, mediana, média, terceiro quartil e máximo), conforme Tabela 4.3 e 4.4 para a base de modelagem e a de predição, respectivamente. Maiores

detalhes sobre a distribuição dos resíduos podem ser encontrados no Apêndice B, através dos *Box-plots*.

Nos estudos de simulação realizados, verificou-se falta de convergência quando assumiu-se $DP(\alpha) = 2$ para o método de QGH, raros problemas de convergência quando utilizado o método QVP e nenhum problema para método usual.

4.2 Resultados

Interpretando as Tabelas 4.1 e 4.2, tem-se as seguintes conclusões:

- Quanto maior o $DP(\alpha)$, a média da estimativa do β está mais afastada do seu valor verdadeiro para o modelo usual, para ambos valores de N , para qualquer tamanho de grupo. O modelo de regressão logística usual apresentou estimativas mais viesadas de β que o modelo multinível, para $DP(\alpha)=1$ e $DP(\alpha)=2$, e este viés aumenta quando o $DP(\alpha)$ aumenta. As estimativas obtidas via QVP em geral foram menos viesadas que as obtidas via QGH. As exceções são quando $N=1000$ com $DP(\alpha)=1$ para todos os tamanhos de grupos.
- No caso em que $DP(\alpha) = 0.1$, os dois métodos de ajuste conduzem a uma sobrestimação do desvio padrão de α , que diminui a medida que o tamanho do grupo aumenta, sendo que o método QVP leva a uma sobrestimação mais leve do que a QGH. O contrário acontece no caso em que $DP(\alpha) = 1$, com exceção de $g=2$ no qual ambos os métodos subestimam o $DP(\alpha)$. É importante salientar que, no método de QGH, não se obteve convergência quando $DP(\alpha) = 2$.
- Quando foi analisada a base de modelagem, os índices de desempenho do modelo usual, são sempre inferiores em relação aos índices do modelo multinível ajustado com as metodologias QVP e QGH. Dentre os modelos multiníveis, nota-se que o modelo ajustado via QVP possui esses índices superiores ao modelo ajustado via

QGH quando o $DP(\alpha)=1$ para qualquer tamanho de grupo e quando $DP(\alpha)=0.1$ para o tamanho do grupo diferente de 2.

- Quando analisada a tabela de classificação na base de predição, os índices de desempenho do modelo usual e do modelo multinível via QVP e via QGH são bem próximos. Esses índices superam o modelo usual timidamente quando o tamanho do grupo aumenta em relação ao tamanho da amostra $N/2$ e também quando o $DP(\alpha)$ aumenta. Comparando os índices de desempenho entre os métodos QVP e QGH, nota-se que, de maneira geral, o método QGH parece ser levemente superior ao QVP, porém os valores continuam a ser bem próximos.

Analisando a Tabela 4.3, que contém as medidas resumo dos resíduos na base de modelagem, nota-se, de maneira geral que, o modelo multinível ajustado via QVP possui resíduos menores do que o ajustado via QGH, com exceção de quando $g = 2$, para $N=200$, e $g = 2$ com $DP(\alpha) = 0.1$, para $N=1000$. Para o modelo multinível estimado via QVP, observa-se claramente que, a medida que o tamanho do grupo e o $DP(\alpha)$ aumentam, mais próximo de zero fica a distribuição dos resíduos. Esse mesmo padrão é encontrado no modelo multinível ajustado via QGH com $DP(\alpha)=1$. Para a regressão logística usual, verifica-se que a distribuição dos resíduos parece não depender do tamanho do grupo, porém depende do $DP(\alpha)$ e, neste caso, quanto menor o $DP(\alpha)$ mais próximo de zero fica a distribuição dos resíduos.

Através dos valores presentes na Tabela 4.4, verifica-se de maneira geral que o modelo usual apresentou resíduos próximos aos do modelo multinível, sendo que, a medida que aumenta o $DP(\alpha)$, mais distante de zero fica a distribuição dos resíduos. Para os casos em que $g = 2$ e $g = 10$ os valores dos resíduos são menores para a regressão logística usual, seguido pelo modelo multinível estimado via QGH e, em seguida, pela regressão via QVP. Nota-se que os resíduos, para o modelo multinível, são menores para $DP(\alpha)=2$ com $g = 25$ e $g = 50$, lembrando que não se obteve convergência para o modelo ajustado via QGH para $DP(\alpha)=2$.

		DP(α) 0.1			DP(α) 1			DP(α) 2		
g	Método de Estimação	Usual	QVP	QGH	Usual	QVP	QGH	Usual	QVP	QGH
2	Média de $\hat{\beta}$	1.052	1.136	1.106	0.877	1.086	1.028	0.646	1.143	-
	Média de $D\hat{P}(\alpha)$	-	0.530	0.316	-	1.213	0.829	-	2.080	-
	Base de Modelagem									
	Acurácia	66.881	74.232	70.911	63.480	81.913	67.488	58.915	89.578	-
	Sensibilidade	66.299	73.650	70.132	61.961	81.385	66.095	58.742	89.762	-
	Especificidade	67.473	74.825	71.704	65.044	82.457	68.922	59.088	89.393	-
	KS	33.772	48.475	41.836	27.005	63.842	35.017	17.830	79.155	-
	Base de Predição									
	Acurácia	67.132	66.788	66.993	64.194	63.368	63.828	60.306	59.195	-
	Sensibilidade	66.394	65.992	66.137	62.847	61.527	62.222	60.230	58.842	-
	Especificidade	67.878	67.592	67.858	65.546	65.217	65.442	60.382	59.549	-
	KS	34.272	33.584	33.995	28.393	26.744	27.664	20.612	18.391	-
10	Média de $\hat{\beta}$	1.024	1.036	1.046	0.852	0.975	1.027	0.633	0.978	-
	Média de $D\hat{P}(\alpha)$	-	0.212	0.217	-	0.940	0.976	-	1.903	-
	Base de Modelagem									
	Acurácia	66.059	67.906	69.021	61.295	73.556	70.816	54.225	80.765	-
	Sensibilidade	66.142	68.145	69.198	60.956	73.422	70.564	54.175	80.847	-
	Especificidade	65.977	67.667	68.844	61.636	73.691	71.069	54.276	80.683	-
	KS	32.119	35.812	38.042	22.592	47.113	41.633	8.451	61.530	-
	Base de Predição									
	Acurácia	67.286	67.207	67.207	64.242	63.980	64.046	58.778	59.090	-
	Sensibilidade	67.489	67.411	67.415	64.142	63.740	63.832	58.958	59.048	-
	Especificidade	67.083	67.003	66.999	64.343	64.220	64.261	58.598	59.132	-
	KS	34.572	34.414	34.414	28.485	27.960	28.093	17.556	18.180	-
25	Média de $\hat{\beta}$	1.019	1.026	1.033	0.853	0.999	1.029	0.647	1.012	-
	Média de $D\hat{P}(\alpha)$	-	0.156	0.169	-	0.952	0.988	-	1.938	-
	Base de Modelagem									
	Acurácia	66.022	67.233	67.722	58.897	71.947	68.460	48.880	78.181	-
	Sensibilidade	66.020	67.239	67.699	58.993	72.113	68.083	48.072	78.336	-
	Especificidade	66.024	67.227	67.745	58.801	71.781	68.836	49.692	78.026	-
	KS	32.044	34.466	35.444	17.794	43.894	36.919	2.236	56.362	-
	Base de Predição									
	Acurácia	67.260	67.220	67.205	63.045	63.315	63.413	57.358	58.943	-
	Sensibilidade	67.376	67.309	67.289	63.202	63.621	63.753	56.554	58.332	-
	Especificidade	67.145	67.131	67.121	62.887	63.007	63.071	58.160	59.553	-
	KS	34.521	34.440	34.410	26.089	26.628	26.824	14.714	17.885	-
50	Média de $\hat{\beta}$	1.023	1.028	1.031	0.892	1.016	1.030	0.673	1.042	-
	Média de $D\hat{P}(\alpha)$	-	0.111	0.131	-	0.822	0.860	-	1.889	-
	Base de Modelagem									
	Acurácia	66.072	66.943	67.486	56.273	69.571	65.324	42.806	74.192	-
	Sensibilidade	65.957	66.749	67.197	56.068	69.600	65.182	42.050	74.193	-
	Especificidade	66.187	67.137	67.776	56.477	69.541	65.466	43.569	74.191	-
	KS	32.144	33.886	34.973	12.545	39.141	30.648	14.381	48.384	-
	Base de Predição									
	Acurácia	67.318	67.284	67.289	61.665	62.084	62.160	56.017	57.564	-
	Sensibilidade	67.296	67.248	67.232	61.325	61.877	61.950	55.394	57.016	-
	Especificidade	67.340	67.320	67.346	61.998	62.286	62.365	56.649	58.121	-
	KS	34.636	34.568	34.578	23.323	24.163	24.315	12.043	15.137	-

Tabela 4.1: Resultado das simulações para $N=200$.

		DP(α) 0.1			DP(α) 1			DP(α) 2		
g	Método de Estimação	Usual	QVP	QGH	Usual	QVP	QGH	Usual	QVP	QGH
2	Média de $\hat{\beta}$	1.007	1.025	1.026	0.839	0.934	0.948	0.618	0.945	-
	Média de $D\hat{P}(\alpha)$	-	0.366	0.194	-	1.094	0.743	-	1.837	-
	Base de Modelagem									
	Acurácia	67.241	72.018	68.259	64.740	83.535	65.016	59.879	81.785	-
	Sensibilidade	67.300	72.080	68.307	64.462	83.327	64.934	60.030	81.760	-
	Especificidade	67.182	71.957	68.212	65.018	83.743	65.099	59.728	81.811	-
	KS	34.482	44.037	36.519	29.480	67.070	30.033	19.758	63.571	-
	Base de Predição									
	Acurácia	67.402	67.332	67.384	64.825	64.444	64.718	61.008	60.729	-
	Sensibilidade	67.478	67.448	67.478	64.597	64.215	64.500	61.141	60.798	-
10	Média de $\hat{\beta}$	1.007	1.010	1.017	0.844	0.951	1.007	0.618	0.916	-
	Média de $D\hat{P}(\alpha)$	-	0.145	0.141	-	0.953	0.986	-	1.882	-
	Base de Modelagem									
	Acurácia	67.245	68.087	68.689	64.476	75.154	72.708	60.931	90.998	-
	Sensibilidade	67.197	68.058	68.665	64.391	75.058	72.640	60.766	90.931	-
	Especificidade	67.293	68.116	68.714	64.561	75.250	72.777	61.097	91.064	-
	KS	34.490	36.174	37.379	28.952	50.308	45.417	21.863	81.995	-
	Base de Predição									
	Acurácia	67.437	67.442	67.445	65.041	64.662	64.670	61.148	60.305	-
	Sensibilidade	67.490	67.480	67.478	64.831	64.473	64.481	61.085	60.217	-
25	Média de $\hat{\beta}$	1.006	1.008	1.013	0.840	0.969	1.000	0.617	0.969	-
	Média de $D\hat{P}(\alpha)$	-	0.123	0.125	-	0.954	0.987	-	1.883	-
	Base de Modelagem									
	Acurácia	67.212	67.805	68.274	63.688	73.420	70.312	57.930	79.987	-
	Sensibilidade	67.061	67.658	68.185	63.608	73.380	70.338	58.053	79.903	-
	Especificidade	67.364	67.951	68.363	63.768	73.460	70.286	57.807	80.070	-
	KS	34.425	35.609	36.548	27.376	46.840	40.624	15.860	59.973	-
	Base de Predição									
	Acurácia	67.324	67.317	67.313	64.548	64.384	64.401	60.334	60.665	-
	Sensibilidade	67.239	67.192	67.187	64.462	64.254	64.276	60.400	60.587	-
50	Média de $\hat{\beta}$	0.861	0.862	0.865	0.845	0.985	1.003	0.624	0.984	-
	Média de $D\hat{P}(\alpha)$	-	0.084	0.088	-	0.956	0.979	-	1.906	-
	Base de Modelagem									
	Acurácia	57.600	57.914	58.018	62.678	72.646	69.292	55.307	79.229	-
	Sensibilidade	67.197	67.504	67.560	62.520	72.638	69.161	54.999	79.226	-
	Especificidade	50.396	50.716	50.855	62.837	72.655	69.423	55.617	79.232	-
	KS	17.593	18.220	18.415	25.357	45.293	38.584	10.616	58.458	-
	Base de Predição									
	Acurácia	57.709	57.693	57.693	64.245	64.305	64.326	59.392	60.338	-
	Sensibilidade	67.258	67.243	67.230	64.097	64.148	64.167	59.066	60.063	-

Tabela 4.2: Resultado das simulações para $N=1000$.

			Método Usual	Método Via QVP	Método via QGH
N	g	DP(α)	Min. 1Qu. 2.Qu. Mean 3Qu. Max.	Min. 1Qu. 2.Qu. Mean 3Qu. Max.	Min. 1Qu. 2.Qu. Mean 3Qu. Max.
200	2	0.1	.000 0.013 0.028 0.035 0.049 0.216	.000 0.019 0.043 0.073 0.092 0.794	.000 0.018 0.039 0.053 0.074 0.406
	10		.000 0.013 0.028 0.036 0.050 0.282	.000 0.016 0.035 0.047 0.063 0.405	.000 0.019 0.042 0.061 0.083 0.512
	25		.000 0.013 0.028 0.036 0.051 0.277	.000 0.015 0.034 0.044 0.061 0.399	.000 0.019 0.045 0.066 0.092 0.517
	50		.000 0.013 0.028 0.035 0.050 0.274	.000 0.014 0.031 0.041 0.057 0.345	.000 0.019 0.044 0.067 0.095 0.477
	2	1	.000 0.065 0.135 0.155 0.224 0.759	.000 0.057 0.125 0.155 0.222 0.870	.000 0.060 0.130 0.158 0.225 0.869
	10		.000 0.065 0.136 0.155 0.227 0.736	.000 0.036 0.080 0.098 0.141 0.641	.000 0.038 0.084 0.112 0.153 0.875
	25		.000 0.065 0.136 0.157 0.230 0.687	.000 0.026 0.058 0.073 0.105 0.465	.000 0.031 0.069 0.102 0.132 0.740
	50		.000 0.058 0.121 0.145 0.210 0.642	.000 0.019 0.045 0.056 0.083 0.487	.000 0.030 0.068 0.100 0.126 0.826
	2	2	.000 0.133 0.250 0.257 0.367 0.872	.000 0.053 0.128 0.171 0.247 0.944	- - - - -
	10		.000 0.133 0.251 0.257 0.369 0.833	.000 0.027 0.069 0.093 0.134 0.749	- - - - -
	25		.000 0.131 0.250 0.257 0.375 0.838	.000 0.018 0.047 0.065 0.094 0.636	- - - - -
	50		.000 0.123 0.247 0.254 0.379 0.831	.000 0.014 0.037 0.050 0.071 0.624	- - - - -
1000	2	0.1	.000 0.009 0.019 0.022 0.032 0.135	.000 0.012 0.027 0.042 0.052 0.598	.000 0.012 0.025 0.034 0.046 0.270
	10		.000 0.009 0.019 0.022 0.032 0.142	.000 0.010 0.021 0.026 0.037 0.273	.000 0.013 0.028 0.042 0.055 0.431
	25		.000 0.009 0.019 0.022 0.032 0.144	.000 0.010 0.021 0.026 0.036 0.251	.000 0.013 0.030 0.049 0.064 0.472
	50		.000 0.004 0.015 0.019 0.029 0.146	.000 0.005 0.016 0.021 0.032 0.240	.000 0.007 0.024 0.045 0.059 0.431
	2	1	.000 0.065 0.134 0.153 0.221 0.739	.000 0.051 0.110 0.136 0.195 0.853	.000 0.063 0.134 0.164 0.235 0.849
	10		.000 0.065 0.134 0.153 0.220 0.832	.000 0.032 0.072 0.090 0.129 0.610	.000 0.034 0.077 0.105 0.143 0.923
	25		.000 0.065 0.134 0.153 0.222 0.691	.000 0.023 0.051 0.064 0.092 0.445	.000 0.026 0.059 0.092 0.113 0.892
	50		.000 0.065 0.134 0.153 0.223 0.715	.000 0.017 0.039 0.049 0.070 0.353	.000 0.022 0.050 0.087 0.099 0.847
	2	2	.000 0.137 0.251 0.257 0.364 0.867	.000 0.051 0.116 0.159 0.228 0.912	- - - - -
	10		.000 0.136 0.252 0.258 0.366 0.858	.000 0.026 0.063 0.087 0.126 0.674	- - - - -
	25		.000 0.135 0.251 0.257 0.366 0.855	.000 0.016 0.042 0.058 0.084 0.593	- - - - -
	50		.000 0.134 0.251 0.257 0.368 0.872	.000 0.012 0.030 0.043 0.062 0.436	- - - - -

Tabela 4.3: Medidas resumo da média dos resíduos absolutos na base de modelagem.

			Método Usual					Método Via QVP					Método via QGH							
N	g	DP(α)	Min.	1Qu.	2.Qu.	Mean	3Qu.	Max.	Min.	1Qu.	2.Qu.	Mean	3Qu.	Max.	Min.	1Qu.	2.Qu.	Mean	3Qu.	Max.
200	2	0.1	.000	0.013	0.028	0.035	0.050	0.211	.000	0.015	0.033	0.043	0.060	0.470	.000	0.014	0.029	0.037	0.052	0.251
	10		.000	0.013	0.028	0.036	0.050	0.266	.000	0.014	0.030	0.038	0.053	0.266	.000	0.014	0.030	0.038	0.053	0.269
	25		.000	0.013	0.028	0.036	0.051	0.273	.000	0.014	0.030	0.037	0.052	0.336	.000	0.014	0.030	0.037	0.053	0.332
	50		.000	0.013	0.029	0.036	0.051	0.261	.000	0.014	0.030	0.037	0.052	0.261	.000	0.014	0.030	0.037	0.053	0.269
	2	1	.000	0.064	0.136	0.156	0.228	0.780	.000	0.071	0.146	0.163	0.236	0.738	.000	0.068	0.140	0.159	0.231	0.772
	10		.000	0.065	0.137	0.156	0.228	0.748	.000	0.066	0.139	0.159	0.230	0.721	.000	0.065	0.138	0.158	0.230	0.736
	25		.000	0.066	0.138	0.156	0.227	0.705	.000	0.064	0.135	0.156	0.227	0.713	.000	0.063	0.134	0.156	0.227	0.719
	50		.000	0.070	0.145	0.164	0.239	0.679	.000	0.068	0.141	0.164	0.240	0.669	.000	0.068	0.142	0.165	0.240	0.667
	2	2	.000	0.133	0.252	0.259	0.371	0.938	.000	0.137	0.261	0.265	0.379	0.989	- - - - -					
	10		.000	0.133	0.253	0.259	0.371	0.905	.000	0.120	0.242	0.258	0.373	0.940	- - - - -					
	25		.000	0.133	0.251	0.258	0.371	0.912	.000	0.116	0.234	0.256	0.369	0.940	- - - - -					
	50		.000	0.135	0.255	0.264	0.377	0.930	.000	0.120	0.237	0.261	0.377	0.944	- - - - -					
1000	2	0.1	.000	0.009	0.019	0.022	0.032	0.150	.000	0.010	0.022	0.027	0.039	0.219	.000	0.009	0.019	0.023	0.033	0.151
	10		.000	0.009	0.019	0.022	0.032	0.132	.000	0.009	0.019	0.023	0.033	0.164	.000	0.009	0.019	0.023	0.033	0.160
	25		.000	0.009	0.019	0.022	0.032	0.130	.000	0.009	0.019	0.023	0.033	0.130	.000	0.009	0.019	0.023	0.033	0.130
	50		.000	0.004	0.015	0.019	0.029	0.129	.000	0.004	0.015	0.020	0.030	0.129	.000	0.004	0.015	0.020	0.030	0.126
	2	1	.000	0.066	0.134	0.154	0.221	0.758	.000	0.073	0.146	0.161	0.232	0.694	.000	0.068	0.138	0.156	0.225	0.741
	10		.000	0.064	0.133	0.153	0.221	0.724	.000	0.065	0.135	0.155	0.225	0.748	.000	0.063	0.133	0.155	0.225	0.760
	25		.000	0.066	0.136	0.154	0.223	0.801	.000	0.064	0.133	0.155	0.224	0.828	.000	0.063	0.132	0.155	0.224	0.833
	50		.000	0.064	0.134	0.153	0.221	0.730	.000	0.061	0.130	0.153	0.222	0.740	.000	0.060	0.130	0.153	0.222	0.744
	2	2	.000	0.138	0.251	0.258	0.365	0.871	.000	0.144	0.262	0.264	0.373	0.849	- - - - -					
	10		.000	0.136	0.252	0.258	0.366	0.910	.000	0.123	0.238	0.256	0.368	0.924	- - - - -					
	25		.000	0.138	0.254	0.259	0.367	0.865	.000	0.118	0.233	0.255	0.368	0.932	- - - - -					
	50		.000	0.134	0.251	0.257	0.366	0.861	.000	0.112	0.229	0.253	0.367	0.938	- - - - -					

Tabela 4.4: Medidas resumo da média dos resíduos absolutos na base de predição.

Capítulo 5

Aplicação

O estudo longitudinal foi constituído inicialmente de uma *coorte* de 652 crianças de famílias residentes em áreas urbanas da zona da mata meridional do estado de Pernambuco, das quais apenas 477 delas tinham informações completas desde o nascimento até os 2, 4, 6, 9, 12, 15 e 18 meses de vida.

A coleta dos dados sobre as crianças foi realizada em 6 maternidades, localizadas nos municípios da zona da mata meridional de Pernambuco, no período de setembro de 1997 a agosto de 1998.

A avaliação nutricional foi realizada nas primeiras 24 horas do nascimento. O acompanhamento da morbidade e do aleitamento materno foi realizado através de visitas domiciliares. As medidas antropométricas foram realizadas aos 2, 4, 6, 9, 12, 15 e 18 meses de vida.

As variáveis que foram acompanhadas ao longo dos 18 meses como diarreia, o aleitamento materno e outro tipo de alimentação foram organizadas de modo a fornecer apenas informações relevantes aos meses citados acima, juntamente com as medidas antropométricas.

Para a avaliação da condição nutricional, sabe-se que o peso é considerado o melhor indicador de crescimento da criança por representar todas as mudanças que ocorrem no

corpo e é muito sensível, pois modifica-se muito rapidamente em situações de desnutrição.

Neste trabalho considerou-se a medida antropométrica avaliada através do escore Z HAZ (altura por idade) que é utilizado no diagnóstico de déficit linear e reflete a desnutrição passada ou crônica.

O sistema de desvio-padrão ou escore Z utiliza a altura/idade da criança avaliada em números de desvios-padrão do valor mediano da população de referência. Um escore Z negativo indica que a medida da criança encontra-se abaixo do padrão de referência.

A classificação de referência para o escore Z, segundo a literatura, é:

1. Desnutrição grave: $Z < 3DP$;
2. Desnutrição moderada: $3DP \leq Z < 2DP$;
3. Desnutrição leve: $2DP \leq Z < 1DP$;
4. Eutrófica: $Z \geq 1DP$.

O objetivo da aplicação, neste banco de dados, é comparar os três métodos de estimação: regressão logística usual, regressão logística multinível utilizado o estimador de QVP e aproximação via QGH em relação à predição da variável resposta.

Para a modelagem, foi utilizado uma amostra de 50% dos grupos, sendo que cada grupo representa uma criança. A outra metade foi utilizada para fazer a predição de valores da variável resposta.

Após diversas tentativas de modelagem, o modelo que ficou melhor ajustado contém as variáveis **Anos de estudo da mãe** e **Peso da criança ao nascer**.

A variável resposta do modelo é o escore Z HAZ (estatura por idade) binário e vale 1 se a criança possui desnutrição grave e 0 caso contrário.

A variável **Anos de estudo da mãe** foi codificada da seguinte maneira: 0 se a mãe tem 5 anos de estudo ou mais e 1 caso contrário.

Segue o modelo multinível ajustado:

HAZ é a variável resposta, X_{1ij} é a variável explicativa **Anos de Estudo da mãe** e X_{2ij} é a variável explicativa **Peso ao nascer**. O modelo para a análise no nível 2 é

definido da seguinte forma:

$$\text{logit}[P(HAZ_{ij} = 1|\alpha_i)] = \beta_{0i} + \beta_1 X_{1ij} + \beta_2 X_{2ij}, \quad (5.1)$$

para $j=1,2,\dots,7$ e $i=1,2,\dots,237$, em que:

- β_{0i} é o intercepto aleatório para a i -ésima criança;
- β_1 é o coeficiente de regressão usual associado à covariável X_{1ij} ;
- β_2 é o coeficiente de regressão usual associado à covariável X_{2ij} ,

em que $\beta_{0i} = \gamma_{00} + u_{0i}$. Além disso, $u_{0i} \sim N(0, \sigma_{u0}^2)$, em que $u_{0i} = \alpha_i$.

Para o modelo de regressão usual, não foi assumido que as observações são agrupadas, conforme segue:

$$\text{logit}[P(HAZ_{ij} = 1)] = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}, \quad (5.2)$$

para $j=1,2,\dots,7$ e $i=1,2,\dots,237$, em que:

- β_0 é o intercepto usual;
- β_1 é o coeficiente de regressão usual associado à covariável X_{1ij} ;
- β_2 é o coeficiente de regressão usual associado à covariável X_{2ij} .

Conforme a Tabela 5.1, para os três modelos ajustados, a interpretação dos coeficientes das variáveis ficaram da seguinte maneira:

- **Anos de estudo da mãe:** a mãe que tem 5 anos estudo ou mais tem menor chance de ter uma criança com HAZ classificada como desnutrição grave.
- **Peso ao nascer:** quanto maior o peso ao nascer, menor é a chance da criança ser classificada com desnutrição grave.

Através do resíduos de Pearson Tabela 5.2, é possível verificar que todos os métodos de estimação conduziram a resíduos fora do intervalo $[-2,2]$,destacando que uma análise de

Variável	Estimativa	Erro Padrão	P-valor	Erro Padrão/Estimativa
Método Usual				
Intercepto	6.23	0.47	<0.00	7.54%
Anos de estudo da mãe	0.86	0.12	<0.00	13.95%
Peso da criança ao nascer	-2.44	0.16	<0.00	-6.56%
Método via QVP				
Intercepto	9.61	1.34	<0.00	13.89%
Anos de estudo da mãe	1.39	0.36	<0.00	25.89%
Peso da criança ao nascer	-3.83	0.43	<0.00	-11.20%
Variância do Intercepto Aleatório (σ)	2.34			
Método via QGH				
Intercepto	12.16	1.74	<0.00	14.28%
Anos de estudo da mãe	1.77	0.46	<0.00	26.00%
Peso da criança ao nascer	-4.85	0.58	<0.00	-12.00%
Variância do Intercepto Aleatório (σ)	2.79			

Tabela 5.1: Ajustes dos modelos usual, multinível via QVP e multinível via QGH.

Medidas Resumo	Min	1o Qt	Mediana	Média	3o Qt	Max
Método Usual	-2.89	-0.63	-0.35	-0.01	0.69	4.66
Método via QVP	-3.57	-0.27	-0.17	-0.05	0.28	4.19
Método via QGH	-2.51	-0.24	-0.04	-0.18	0.37	2.76

Tabela 5.2: Resíduo de Pearson dos modelos usual, via QVP e via QGH.

diagnóstico mais detalhada posterior seria necessária. Apesar disso, nota-se que o modelo multinível apresentou resíduos menores em relação a regressão logística usual.

Pela Tabela 5.3, pode-se observar a tabela de classificação dos modelos na base de modelagem e predição. Através destes valores, foram calculadas as medidas de desempenho apresentadas no Capítulo 3 nas base de modelagem e de predição, respectivamente, para os três ajustes dos modelos.

Observando as medidas de desempenho na base de modelagem, conforme Tabela 5.4, nota-se que, de maneira geral, o método QVP superou os demais, para os índices de sensibilidade, acurácia e KS. Entretanto o método de QGH tem desempenho melhor quando foi observado o índice especificidade. A regressão logística usual mostrou um resultado inferior quando foi comparado todos os índices em relação aos métodos QVP e QGH.

Analisando as medidas de desempenho na base de predição, conforme Tabela 5.4, verifica-se que os três modelos conduzem a uma classificação similar. Logo, apesar do

Base de Modelagem					Base de Predição				
Método Usual									
		Predito		Total			Predito		Total
		0	1				0	1	
Observado	0	817	323	1140	Observado	0	802	366	1168
	1	177	342	519		1	157	355	512
		994	665	1659			959	721	1680
Método via QVP									
		Predito		Total			Predito		Total
		0	1				0	1	
Observado	0	984	153	1137	Observado	0	802	366	1168
	1	45	474	519		1	157	355	512
		1029	657	1659			959	721	1680
Método via QGH									
		Predito		Total			Predito		Total
		0	1				0	1	
Observado	0	1017	123	1140	Observado	0	802	366	1168
	1	173	346	519		1	157	355	512
		1190	469	1659			959	721	1680

Tabela 5.3: Tabela de classificação do modelos usual, via QVP e via QGH na base de modelagem e predição.

Medidas de Desempenho	Método Usual	Método via QVP	Método via QGH
Base de Modelagem			
Sensibilidade	65.9	91.33	66.67
Especificidade	71.67	86.54	89.21
Acurácia	69.86	88.04	82.16
KS	37.56	77.87	55.88
Base de Predição			
Sensibilidade	69.34	69.34	69.34
Especificidade	68.66	68.66	68.66
Acurácia	68.87	68.87	68.87
KS	38.00	38.00	38.00

Tabela 5.4: Medidas de desempenho dos modelos usual, via QVP e via QGH.

modelo multinível, na base de modelagem, ajustado via QVP e QGH ter acerto significativamente superior ao modelo de regressão logística usual, na base de predição, os 3

métodos acabam sendo equivalentes. Isso ocorre devido ao fato do tamanho do grupo ser pequeno em relação ao tamanho da amostra e também pelo fato de somente variáveis no nível do grupo (criança) compor o modelo. No estudo de simulação, apresentado no capítulo anterior, mostrou-se que a predição do modelo multinível é eficaz quando se tem o desvio padrão do efeito aleatório e o tamanho do grupo em relação ao da amostra for *grande*.

Conforme citado anteriormente o método multinível apresentou melhores resultados em relação a metodologia usual, mostrando-se assim que há um ganho expressivo em ajustar o modelo multinível. Dentre os métodos de estimação para o ajuste do modelo multinível, o método de QVP que apresentou medidas de desempenho muito superiores à metodologia de QGH.

Pode-se concluir desta maneira que a quantidade de anos de estudos da mãe e o peso da criança ao nascer são fatores que influenciam na desnutrição grave em vários estágios da vida da criança. Entretanto, em termos de predição, neste caso específico, não se encontrou resultados significativos para utilizar modelo multinível, ou seja, dado que nasça uma nova criança a probabilidade dela apresentar desnutrição grave nos respectivos períodos após seu nascimento produz resultados similares à metodologia usual de predição da variável resposta.

Capítulo 6

Comentários finais

Os modelos multiníveis para resposta binária são úteis para análises de dados binários quando as observações estão agrupadas.

Na presente dissertação, estudou-se métodos para obter as estimativas dos parâmetros de interesse no modelo multinível de 2 níveis para resposta binária com função de ligação logit, considerando as estimações via Quase-Verossimilhança Penalizada (QVP) e via Quadratura de Gauss-Hermite (QGH). Para a estimação QVP, utilizou-se o procedimento **glmmPQL** e, para a QGH, o procedimento **glmmML** para se obter as estimativas dos parâmetros do modelo estudado. Ambos os modelos foram ajustados no software estatístico R.

Apresentou-se também uma proposta para realizar a predição da resposta neste tipo de modelo. A investigação da predição da variável resposta binária foi realizada no presente trabalho através de simulações para avaliar o desempenho dos estimadores na base de modelagem. O resultado mostra que ignorar o agrupamento entre as respostas, quando o desvio padrão do efeito aleatório e o tamanho do grupo são grandes, causa maior viés na estimativa do parâmetro de regressão. Esse resultado pode ser verificado quando comparado o modelo de regressão logística usual com as duas metodologias de estimação do modelo multinível apresentadas.

Além disso, na base de modelagem, foi possível constatar que as estimativas obtidas via QVP, em geral, foram menos viesadas que as obtidas via QGH e, na base de predição, os três métodos convergiram à uma classificação similar. Nela, observou-se uma pequena superação da proposta apresentada neste trabalho quando o tamanho do grupo aumenta em relação ao tamanho da amostra e quando o desvio padrão do efeito aleatório aumenta. Desta forma, mostrou-se que quando utilizado somente uma estimativa do desvio padrão do efeito aleatório é possível fazer a predição de observações de novos grupos. Entretanto, o ganho encontrado não foi tão expressivo quanto o apresentado no ajuste do modelo. Isso se deve ao fato de não existir o valor o efeito aleatório para cada grupo quando se faz a predição de variável resposta de um novo grupo.

O objetivo da aplicação tratada neste trabalho baseou-se na predição da variável resposta de observações, ao longo do tempo, de uma criança possuir desnutrição grave. Entre as duas propostas de ajuste de modelos multiníveis, notou-se que as medidas de desempenho da estimação via QVP é superior ao modelo ajustado via QGH. Para fazer a predição da variável resposta para novas observações de uma criança, verificou-se que as medidas de desempenho dos modelos ajustados convergem aos mesmos resultados. Isto ocorreu porque o tamanho do grupo em relação ao tamanho da amostra não foi suficientemente grande para que as medidas de desempenho dos multiníveis superassem o modelo de regressão logística usual, apesar do desvio padrão do efeito aleatório ter um valor consideravelmente grande.

Para predizer a variável resposta de observações de novos grupo, os modelos multiníveis são recomendados quando o tamanho do grupo e o desvio padrão do efeito aleatório aumentam. Em resumo, pode-se concluir que incorporar os efeitos aleatórios do grupo é importante no ajuste (estimação) quando as respostas estão agrupadas. Entretanto, não ter o valor do efeito aleatório para um novo grupo faz com que o modelo multinível tenha resultados similares ao da regressão logística usual, em termos de predição.

Como continuidade desta pesquisa, são indicadas as seguintes sugestões:

- Considerar modelo multinível com 3 ou mais níveis;
- Desenvolver métodos de estimação que considerem mais de uma covariável aleatória;
- Estudar em detalhe a falta de convergência para o método de estimação via QGH, em particular, Laplace quando o desvio padrão do efeito aleatória aumenta;
- Por fim, para o método de predição em modelos multiníveis, fica como sugestão desenvolver uma equação para a predição da variável resposta considerando mais covariáveis aleatórias e níveis maiores do que 2.

Apêndice A

Código do programa para simulação

```
#####  
#                                                                 #  
#           Bibliotecas necessárias                               #  
#                                                                 #  
#####  
  
library(MASS)  
library(nlme)  
library(glmmML)  
  
#####  
#                                                                 #  
#           Parâmetros para Simulação                           #  
#                                                                 #  
#####  
  
#Entrar com a semente para x  
semente<-0  
  
#Entrar com a quantidade de observações  
n<-200  
  
#Entrar com o número de simulações  
simulacao<-1000
```

```

#Desvios padrão do efeito aleatório "a"
#efeito aleatório sempre com média zero
sd.a<-0.1

#Média e desvio padrão da covariável "x"
mean.x<-0
sd.x<-1

#Valor de beta
beta<-1

#Cria índice, identificando cada observação
tam.grupo<-2
index.n<-cbind(c(1:(n/2)))
total.grupo<-(n/2)/tam.grupo

#Identifica observações dentro do grupo
for(i in 1:total.grupo) {
  for (j in 1:tam.grupo) {
    k <- (i-1) * tam.grupo + j
    index.n[k] <- i
  }
}

#####
#
#           Declaração e inicialização das variáveis
#
#####

#Valores gerados a partir da distribuição Binomial, com probabilidade p
bin<-array(0,c(n,simulacao))
bin.predicao<-array(0,c(n/2,simulacao))
bin.construcao<-array(0,c(n/2,simulacao))
covariavel<-array(0,c(n/2,1))

#Probabilidade do evento gerado a partir da binomial para base de predicao
p.evento<-array(0,c(1,simulacao))
p <- array(0,c(n,simulacao))

```

```

p.construc<- array(0,c(n/2,simulacao))
p.valid<- array(0,c(n/2,simulacao))

#Valor do efeito fixo da Logística Misto
ef.fixo.pql<-array(0,c(n/2,1))
hi.pql<-array(0,c(n/2,simulacao))
ef.fixo.ml<-array(0,c(n/2,1))
hi.ml<-array(0,c(n/2,simulacao))

#Variáveis relacionadas ao modelo usual, identificado como ".u"
beta.est.u<-0
beta.est.r.u<-array(0,c(simulacao,1))
dp.beta.est.u<-0
vies.beta.u<-0
contagem0.u.c<-0
contagem1.u.c<-0
contagem0.u.p<-0
contagem1.u.p<-0

#Variáveis relacionadas ao modelo PQL, identificado como ".pql"
var.est.pql<-0
beta.est.pql<-0
beta.est.r.pql<-array(0,c(simulacao,1))
var.est.r.pql<-array(0,c(simulacao,1))
dp.beta.est.pql<-0
dp.var.est.pql<-0
vies.beta.pql<-0
vies.var.pql<-0
contagem0.pql.c<-0
contagem1.pql.c<-0
contagem0.pql.p<-0
contagem1.pql.p<-0

#Variáveis relacionadas ao modelo ML, identificado como ".ml"
var.est.ml<-0
beta.est.ml<-0
beta.est.r.ml<-array(0,c(simulacao,1))
var.est.r.ml<-array(0,c(simulacao,1))
dp.beta.est.ml<-0

```

```

dp.var.est.ml<-0
vies.beta.ml<-0
vies.var.ml<-0
contagem0.ml.c<-0
contagem1.ml.c<-0
contagem0.ml.p<-0
contagem1.ml.p<-0

#Probabilidade estimada pela Logística usual
prob.usual.construc<-array(0,c(n/2,simulacao))
prob.usual.valid<-array(0,c(n/2,simulacao))
predicao.usual<-array(0,c(n/2,1))

#Valor predicao da Logística PQL
predicao.pql<-array(0,c(n/2,1))
prob.pql.construc<-array(0,c(n/2,simulacao))

#Valor predicao da Logística ML
predicao.ml<-array(0,c(n/2,1))
prob.ml.construc<-array(0,c(n/2,simulacao))

#Matriz que guarda diferença absoluta para base de construção e predição
abs.prob.u.construc<-array(0,c(n/2,simulacao))
abs.prob.pql.construc<-array(0,c(n/2,simulacao))
abs.prob.ml.construc<-array(0,c(n/2,simulacao))
abs.prob.u.valid<-array(0,c(n/2,simulacao))
abs.prob.pql.valid<-array(0,c(n/2,simulacao))
abs.prob.ml.valid<-array(0,c(n/2,simulacao))

#####
#                                                                 #
#                               Simulação do dados                #
#                                                                 #
#####

for(r in 1:simulacao){
  a<-matrix(0,tam.grupo,2*total.grupo)
  x<-matrix(0,tam.grupo,2*total.grupo)

```

```

#fixa a semente através do comando "set.seed"
set.seed(r)
for(j in 1:(2*total.grupo)){
  #gera valores de a por grupo
  a[,j]<-rnorm(1,0,sd.a)
}

#fixa a semente através do comando "set.seed"
set.seed(r+semente)
for(j in 1:(2*total.grupo)){
  x[,j]<-rnorm(tam.grupo,mean.x,sd.x)#gera valores de x por grupo
}

#converte a matriz x em dados
a1<-data.frame(a)

#converte a matriz x em dados
x1<-data.frame(x)

#empilha cada coluna em um vetor
vec1<-stack(a1)
vec2<-stack(x1)
a<-vec1[,1]
x<-vec2[,1]

for(j in 1:n){
  p[j,r]<-exp(x[j]*beta + a[j])/(1 + exp(x[j]*beta + a[j]))
  bin[j,r]<-rbinom(1,1,p[j,r])
}

for(j in 1:(n/2)){ #seleciona variável resposta da base de predição
  bin.predicao[j,r]<-bin[j+(n/2),r]
  p.construc[j,r]<-p[j,r]
  p.valid [j,r]<-p[j+(n/2),r]
}

#Roda os modelos Logística Usual e Misto (PQL e ML)
for(j in 1:(n/2)){
  covariavel[j]<-x[j]
  bin.construcao[j,r]<-bin[j,r]

```



```

}

#Calcula o ponto de corte (baseado na base de modelagem ou construção)
p.evento[1,r]<-sum(bin.construcao[,r])/(n/2)

VAR<-matrix(0,n/2,3)
VAR[,1]<-bin.construcao[,r]
VAR[,2]<-covariavel
VAR[,3]<-index.n
#converte as informações um conjunto de dados
dados<-data.frame(VAR)
#mostra os nomes das variaveis
banco.dados<-names(dados)

#Ajusta a regressão logística usual
fit.u<-glm( X1 ~ X2-1 , family = binomial, data=dados)
beta.est.r.u[r]<-coef(fit.u)

#Ajusta a regressão logística multinível com Quase-Verossimilhança Penalizada
fit.pql<-glmmPQL( X1 ~ X2-1 , random =~1| X3, family=binomial,data=dados)
var.est.r.pql[r]<-as.numeric(VarCorr(fit.pql)[1,2])
beta.est.r.pql[r]<-fixef(fit.pql)

#Ajusta a regressão logística multinível com Quadratura de Gauss-Hermite
fit.ml<-glmmML( X1 ~ X2-1 , cluster = X3, family=binomial(), start.sigma=sd.a,
               start.coef=1, data=dados )
var.est.r.ml[r]<-as.numeric(fit.ml[5])
beta.est.r.ml[r]<-coef(fit.ml)

#Calcula o valor predito para o modelo multinível na base de modelagem (construção)
for(k in 1 : total.grupo){
  for(i in 1:(n/2)){
    if (index.n[i]==k){
      prob.ml.construc[i,r]<-exp(beta.est.r.ml[r]*x[(i)]+
                                fit.ml[14]$posterior.modes[k])/(1+exp(beta.est.r.ml[r]*x[(i)]+
                                fit.ml[14]$posterior.modes[k]))
      prob.pql.construc[i,r]<-exp(beta.est.r.pql[r]*x[(i)]+
                                raneff(fit.pql)[k,1])/(1+exp(beta.est.r.pql[r]*x[(i)]+
                                raneff(fit.pql)[k,1]))
    }
  }
}

```

```

    }
  }
}

for(j in 1:(n/2)){
  #Calcula o valor predito para o modelo usual na base de modelagem e predição
  prob.usual.construc[j,r]<-exp(beta.est.r.u[r]*x[(j))]/(1+exp(beta.est.r.u[r]*x[(j))))
  prob.usual.valid[j,r]<-exp(beta.est.r.u[r]*x[(j+(n/2))])/(1+exp(beta.est.r.u[r]*x[(j+(n/2))]))

  #Calcula a parte fixa do modelo multinível que será utilizado na predição da variável resposta
  ef.fixo.pql[j]<-beta.est.r.pql[r]*x[(j+(n/2))]
  ef.fixo.ml[j]<-beta.est.r.ml[r]*x[(j+(n/2))]
}

beta.est.u<-mean(beta.est.r.u)
dp.beta.est.u<-sd(beta.est.r.u)

beta.est.pql<-mean(beta.est.r.pql)
var.est.pql<-mean(var.est.r.pql)
dp.beta.est.pql<-sd(beta.est.r.pql)
dp.var.est.pql<-sd(var.est.r.pql)

beta.est.ml<-mean(beta.est.r.ml)
var.est.ml<-mean(var.est.r.ml)
dp.beta.est.ml<-sd(beta.est.r.ml)
dp.var.est.ml<-sd(var.est.r.ml)

vies.beta.u<-beta.est.u-beta
vies.beta.pql<-beta.est.pql-beta
vies.var.pql<-var.est.pql-sd.a
vies.beta.ml<-beta.est.ml-beta
vies.var.ml<-var.est.ml-sd.a

#Predição da variável resposta do modelo logístico multinível (apresentado no Capítulo 3)

ni<-tam.grupo
yi.<-ni/2

# Função densidade da normal:((1/(sqrt(2*pi)*sigma))*exp(-((z -mu)^2/(2*sigma^2))))

```

```

sigma<-1
mu<-0

#Variáveis auxiliares para predição do modelo multinível
parte1<-array(0,c(n/2,1))
parte2<-array(0,c(n/2,1))

#Para modelo ajustado via Quase-Verossimilhança Penalizada
for(i in 1: ((n/2)/tam.grupo)){
  ii <- (i-1)*tam.grupo

  #expressão para o somatoria
  texto_somatoria <- "("
  for (qq in 1:tam.grupo)
    texto_somatoria <- paste(texto_somatoria,"+log(1+exp(ef.fixo.pql[ii+",as.character(qq),"
      +var.est.r.pql[r]*z)))")
  texto_somatoria <- paste(texto_somatoria,")")

  for(k in 1:tam.grupo) {
    j <- ii+k
    texto_integrand <- paste("function(z)
      ((exp((yi.+1) * var.est.r.pql[r] * z -(",
      texto_somatoria, ")
      ) / (1+exp(ef.fixo.pql[j]+var.est.r.pql[r]*z)))
      * ( ( 1/( sqrt(2*pi)*sigma ) ) * exp(-( (z - mu)^2 / (2*sigma^2) ) ) ) ) ")

    integrand <- eval(parse(text=texto_integrand))

    aux1<-integrate(integrand, lower = -Inf, upper =Inf)
    parte1[j]<-aux1$value

    texto_integrand2 <- paste("function(z)
      ( exp(yi.*var.est.r.pql[r]*z - (",
      texto_somatoria, ")
      ) * ( ( 1/(sqrt(2*pi)*sigma) ) * exp( -( (z - mu)^2/(2*sigma^2) ) ) ) ) ) ")
    integrand2 <- eval(parse(text=texto_integrand2))
    aux2<-integrate(integrand2, lower = -Inf, upper =Inf)
    parte2[j]<-aux2$value
  }
}

```

```

        hi.pql[j,r]<-exp(ef.fixo.pql[j])*(parte1[j]/parte2[j])
    }
}

#Para modelo ajustado via Quadratura de Gauss-Hermite
for(i in 1: ((n/2)/tam.grupo)) {
    ii <- (i-1)*tam.grupo

    #expressão para o somatoria
    texto_somatoria <- "("
    for (qq in 1:tam.grupo)
        texto_somatoria <- paste(texto_somatoria,"+log(1+exp(ef.fixo.ml[ii+",as.character(qq),"
                                +var.est.r.ml[r]*z)))")
    texto_somatoria <- paste(texto_somatoria,")")

    for(k in 1:tam.grupo) {
        j <- ii+k
        texto_integrand <- paste("function(z)
            ( ( exp( (yi.+1) * var.est.r.ml[r] * z - (",
                texto_somatoria, ")
            ) / ( 1+exp(ef.fixo.ml[j]+var.est.r.ml[r]*z) ) )
            * ( ( 1/( sqrt(2*pi)*sigma ) ) * exp( -( (z - mu)^2 / (2*sigma^2) ) ) ) ) )"

        integrand <- eval(parse(text=texto_integrand))

        aux1<-integrate(integrand, lower = -Inf, upper =Inf)
        parte1[j]<-aux1$value

        texto_integrand2 <- paste("function(z)
            ( ( exp( yi.*var.est.r.ml[r]*z - (",
                texto_somatoria, ")
            ) * ( ( 1/(sqrt(2*pi)*sigma) ) * exp( -( (z - mu)^2/(2*sigma^2) ) ) ) ) )"
        integrand2 <- eval(parse(text=texto_integrand2))
        aux2<-integrate(integrand2, lower = -Inf, upper =Inf)
        parte2[j]<-aux2$value

        hi.ml[j,r]<-exp(ef.fixo.ml[j])*(parte1[j]/parte2[j])
    }
}

```

```

#Classificação das probabilidades em zero ou um
construc.usual<-array(0,c(n/2,1))
construc.pql<-array(0,c(n/2,1))
construc.ml<-array(0,c(n/2,1))

predicao.usual<-array(0,c(n/2,1))
predicao.pql<-array(0,c(n/2,1))
predicao.ml<-array(0,c(n/2,1))

for (j in 1:(n/2)){
  if (prob.usual.valid[j,r]>=p.evento[1,r])
    predicao.usual[j]<-1
  if (hi.pql[j,r] >= p.evento[1,r])
    predicao.pql[j]<-1
  if (hi.ml[j,r] >= p.evento[1,r])
    predicao.ml[j]<-1
  if (prob.usual.construc[j,r]>=p.evento[1,r])
    construc.usual[j]<-1
  if (prob.pql.construc[j,r] >= p.evento[1,r])
    construc.pql[j]<-1
  if (prob.ml.construc[j,r] >= p.evento[1,r])
    construc.ml[j]<-1
}

#Tabela de classificação para base de construção
for (j in 1:(n/2)){
  #Contagem do acerto da Logística Usual, acumulando para cada simulação
  if (construc.usual[j]==0 & bin.construcao[j,r]==0)
    contagem0.u.c<-contagem0.u.c+1
  if (construc.usual[j]==1 & bin.construcao[j,r]==1)
    contagem1.u.c<-contagem1.u.c + 1
  #Contagem do acerto da Logística PQL, acumulando para cada simulação
  if (construc.pql[j]==0 & bin.construcao[j,r]==0)
    contagem0.pql.c<-contagem0.pql.c + 1
  if (construc.pql[j]==1 & bin.construcao[j,r]==1)
    contagem1.pql.c<-contagem1.pql.c + 1
  #Contagem do acerto da Logística ML, acumulando para cada simulação
  if (construc.ml[j]==0 & bin.construcao[j,r]==0)

```

```

        contagem0.ml.c<-contagem0.ml.c + 1
    if (construc.ml[j]==1 & bin.construcao[j,r]==1)
        contagem1.ml.c<-contagem1.ml.c + 1
}

#Tabela de classificação para base de predição
for (j in 1:(n/2)){
    #contagem do contagem do acerto da Logística Usual, acumulando para cada simulação
    if (predicao.usual[j]==0 & bin.predicao[j,r]==0)
        contagem0.u.p<-contagem0.u.p+1
    if (predicao.usual[j]==1 & bin.predicao[j,r]==1)
        contagem1.u.p<-contagem1.u.p + 1
    #contagem do contagem do acerto da Logística PQL, acumulando para cada simulação
    if (predicao.pql[j]==0 & bin.predicao[j,r]==0)
        contagem0.pql.p<-contagem0.pql.p + 1
    if (predicao.pql[j]==1 & bin.predicao[j,r]==1)
        contagem1.pql.p<-contagem1.pql.p + 1
    #contagem do contagem do acerto da Logística ML, acumulando para cada simulação
    if (predicao.ml[j]==0 & bin.predicao[j,r]==0)
        contagem0.ml.p<-contagem0.ml.p + 1
    if (predicao.ml[j]==1 & bin.predicao[j,r]==1)
        contagem1.ml.p<-contagem1.ml.p + 1
}

#Compararando probabilidade estimada versus predito
for (j in 1:(n/2)){
    #Base construção
    abs.prob.u.construc[j,r]<-abs(p.construc[j,r]-prob.usual.construc[j,r])
    abs.prob.pql.construc[j,r]<-abs(p.construc[j,r]-prob.pql.construc[j,r])
    abs.prob.ml.construc[j,r]<-abs(p.construc[j,r]-prob.ml.construc[j,r])

    #Base predição
    abs.prob.u.valid[j,r]<-abs(p.valid[j,r]-prob.usual.valid[j,r])
    abs.prob.pql.valid[j,r]<-abs(p.valid[j,r]-hi.pql[j,r])
    abs.prob.ml.valid[j,r]<-abs(p.valid[j,r]-hi.ml[j,r])
}
print(r)
}

```

```
#####
#
#                                CÁLCULO DAS MEDIDAS DE DESEMPENHO                                #
#
#####

#Medidas de desempenho dos modelos na base de construção ".c"

acuracia.u.c<-((contagem0.u.c+contagem1.u.c)      /((n/2)*simulacao))*100
acuracia.pql.c<-((contagem0.pql.c+contagem1.pql.c)/((n/2)*simulacao))*100
acuracia.ml.c<-((contagem0.ml.c+contagem1.ml.c)   /((n/2)*simulacao))*100

sensibilidade.u.c<-(contagem1.u.c      /(sum(bin.construcao)))*100
sensibilidade.pql.c<-(contagem1.pql.c/(sum(bin.construcao)))*100
sensibilidade.ml.c<-(contagem1.ml.c   /(sum(bin.construcao)))*100

especificidade.u.c<-(contagem0.u.c      /(((n/2)*simulacao-sum(bin.construcao))))*100
especificidade.pql.c<-(contagem0.pql.c/(((n/2)*simulacao-sum(bin.construcao))))*100
especificidade.ml.c<-(contagem0.ml.c   /(((n/2)*simulacao-sum(bin.construcao))))*100

KS.u.c<-abs(sensibilidade.u.c-      (100-especificidade.u.c))
KS.pql.c<-abs(sensibilidade.pql.c-(100-especificidade.pql.c))
KS.ml.c<-abs(sensibilidade.ml.c-   (100-especificidade.ml.c))

#Medidas de desempenho dos modelos na base de predição ".p"

acuracia.u.p<-((contagem0.u.p+contagem1.u.p)/((n/2)*simulacao))*100
acuracia.pql.p<-((contagem0.pql.p+contagem1.pql.p)/((n/2)*simulacao))*100
acuracia.ml.p<-((contagem0.ml.p+contagem1.ml.p)/((n/2)*simulacao))*100

sensibilidade.u.p<-(contagem1.u.p/(sum(bin.predicao)))*100
sensibilidade.pql.p<-(contagem1.pql.p/(sum(bin.predicao)))*100
sensibilidade.ml.p<-(contagem1.ml.p/(sum(bin.predicao)))*100

especificidade.u.p<-(contagem0.u.p/(((n/2)*simulacao-sum(bin.predicao))))*100
especificidade.pql.p<-(contagem0.pql.p/(((n/2)*simulacao-sum(bin.predicao))))*100
especificidade.ml.p<-(contagem0.ml.p/(((n/2)*simulacao-sum(bin.predicao))))*100

KS.u.p<-abs(sensibilidade.u.p-(100-especificidade.u.p))
KS.pql.p<-abs(sensibilidade.pql.p-(100-especificidade.pql.p))
```

```

KS.ml.p<-abs(sensibilidade.ml.p-(100-especificidade.ml.p))

#Empilhando os resíduos absolutos em um único vetor
a<-stack(data.frame(abs.prob.u.construc))
banco.dados<-names(a)
b<-stack(data.frame(abs.prob.u.valid))
banco.dados<-names(b)
c<-stack(data.frame(abs.prob.pql.construc))
banco.dados<-names(c)
d<-stack(data.frame(abs.prob.pql.valid))
banco.dados<-names(d)
e<-stack(data.frame(abs.prob.ml.construc))
banco.dados<-names(e)
f<-stack(data.frame(abs.prob.ml.valid))
banco.dados<-names(f)

#####
#                                                                 #
#                               OUTPUT DA SIMULAÇÃO              #
#                                                                 #
#####

#Mostra quais foram os parâmetros de entrada
semente
n
simulacao
tam.grupo
mean.x
sd.x
sd.a
beta

# Tabela de classificação na base de predição
print("Logística Usual")
round(beta.est.u,3)
round(vies.beta.u,3)

round(acuracia.u.p,3)
round(sensibilidade.u.p,3)

```



```

round(especificidade.u.p,3)
round(KS.u.p,3)

print("Logística PQL")
round(beta.est.pql,3)
round(vies.beta.pql,3)
round(var.est.pql,3)
round(vies.var.pql,3)

round(acuracia.pql.p,3)
round(sensibilidade.pql.p,3)
round(especificidade.pql.p,3)
round(KS.pql.p,3)

print("Logística ML")
round(beta.est.ml,3)
round(vies.beta.ml,3)
round(var.est.ml,3)
round(vies.var.ml,3)

round(acuracia.ml.p,3)
round(sensibilidade.ml.p,3)
round(especificidade.ml.p,3)
round(KS.ml.p,3)

#Box-plot da diferença absoluta entre valor predido e observado
boxplot(a$values,c$values,e$values,b$values,d$values,f$values,ylim=c(0,1),main="Erro absoluto")

# Medidas resumo para a diferença absoluta entre valor predido e observado
#Usual.construção
round(summary(a$values),3)
#Usual.predição
round(summary(b$values),3)
#PQL.construção
round(summary(c$values),3)
#PQL.predição
round(summary(d$values),3)
#ML.construção
round(summary(e$values),3)

```

```

#ML.predição
round(summary(f$values),3)

# Tabela de classificação na base de construção
print("Logística Usual")
round(acuracia.u.c,3)
round(sensibilidade.u.c,3)
round(especificidade.u.c,3)
round(KS.u.c,3)

print("Logística PQL")
round(acuracia.pql.c,3)
round(sensibilidade.pql.c,3)
round(especificidade.pql.c,3)
round(KS.pql.c,3)

print("Logística ML")
round(acuracia.ml.c,3)
round(sensibilidade.ml.c,3)
round(especificidade.ml.c,3)
round(KS.ml.c,3)

```


Apêndice B

Box-Plot do resíduo das simulações

B.1 Box-Plot do resíduo das simulações para $N=200$

Figura B.1: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=0.1$.

Figura B.2: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=0.1$.

Figura B.3: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=0.1$.

Figura B.4: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=0.1$.

Figura B.5: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=1$.

Figura B.6: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=1$.

Figura B.7: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=1$.

Figura B.8: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=1$.

Figura B.9: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=2$ e $DP(\alpha)=2$.

Figura B.10: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=10$ e $DP(\alpha)=2$.

Figura B.11: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=25$ e $DP(\alpha)=2$.

Figura B.12: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=200$, $g=50$ e $DP(\alpha)=2$.

B.2 Box-Plot do resíduo das simulações para $N=1000$

Figura B.13: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=0.1$.

Figura B.14: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=0.1$.

Figura B.15: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=0.1$.

Figura B.16: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=0.1$.

Figura B.17: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=1$.

Figura B.18: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=1$.

Figura B.19: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=1$.

Figura B.20: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=1$.

Figura B.21: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=2$ e $DP(\alpha)=2$.

Figura B.22: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=10$ e $DP(\alpha)=2$.

Figura B.23: Box-plot da Distribuição dos Resíduos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=25$ e $DP(\alpha)=2$.

Figura B.24: Box-plot da Distribuição dos Absolutos nas bases de modelagem e predição, com $N=1000$, $g=50$ e $DP(\alpha)=2$.

Referências Bibliográficas

- [Breslow e Clayton,1993] Breslow, N. E., Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. v.88, p.9-25.
- [Bryk e Raudenbush,2002] Bryk, A. and Raudenbush, S. (2002). *Hierarchical Linear Models*. Newbury Park, California, Sage.
- [Brostrom,2003] Brostrom, G. (2003). Generalized linear models with random intercepts. SE-90187 Umea, Sweden.
- [Caroll,2003] Carroll, R. (2003). Variances are not always nuisance parameters. *Biometrics*.v.59, p.211-220.
- [Chen,2005] Chen, Y. (2005). Computationally efficient Monte Carlo EM algorithms for generalized linear mixed models. *Journal of Statistical Computation and Simulation*. v.76, No.9, p.817-828.
- [Davis e Rabinowitz,1984] Davis, P. and Rabinowitz, P. (1984). *Methods of Numerical Integration*. Academic Press, second edition.
- [Demidenko,2004] Demidenko, E. (2004). *Mixed Models. Theory and Applications*. John Wiley & Sons, first edition.
- [Goldstein,1991] Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*. v.78, p.45-51.

- [Fitzmaurice e Lard,1993] Fitzmaurice, G. and Lard, N. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80(1):141 - 51.
- [Goldstein,1999] Goldstein, H. (1999). *Multilevel Statistical Models*. London: Institute of Education. University of London, sencond edition.
- [Goldstein e Rasbash,1996] Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society. Serie A*, v.150, p.505-513.
- [Hosmer e Lemeshow,2000] Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wile, sencond edition.
- [Hox,1995] Hox.J (1995). *Applied Multilevel Analysis*. Amsterdam:T-T Pubilicates.
- [Jiang e Lahiri,2006] Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. *Test*, No. 1, p.1-96.
- [Jorgensen,1987] Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society* v.49, p.127-162.
- [Lange,1998] Lange, K (1998). *Numerical Analysis for Statistician*. Springer, first edition.
- [Lee e Nelder,2004] Lee, Y. and Nelder, J. (2004). Conditional and marginal models: Another view. *Statistical Science*, 19(2):219 - 238.
- [Liang,Diggle,Heagerty e Zeger,2002] Liang, K., Diggle P., Heagerty P. and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press, second edition.
- [Liu e Pierce,1994] Liu, Q. and Pierce, A. (1994). A note on Gauss-Hermite Quadrature. *Biometrika*, 81(2): 624-9.

- [Longford,1987] Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*. v.74, p.817-827.
- [Longford,1994] Longford, N. T. (1994). Logistic regression with random coefficients. *J. Comput. Statist.Data Anal.* v.17, p.1-15.
- [McCulloch,1997] McCulloch, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*. 92(437):162-170.
- [Molenberghs e Verbeke,2005] Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, first edition.
- [Nia,2006] Nia V. (2006). *Gauss-Hermite Quadrature: Numerical or Statistical Method?*. Institute of Mathematics, Ecole Polytechnique Fédéral de Lausanne.
- [Paula,2004] Paula, G. A. (2004). *Modelos de regressão com apoio computacional*. São Paulo:IME/USP.
- [Pinheiro e Bates,2000] Pinheiro, J. and Bates, D.(2000). *Mixed-effects Models in S and S-PLUS*. New York: Springer-Verlag.
- [Press,1994] Press, W.H. (1994). *Numerical recipes in C: the art of scientific computing*. New York, NY, USA : Cambridge University Press, second edition. *Journal of Computational and Graphical Statistics*.* 9.1, p141.
- [Renard,2002] Renard, D. (2002). *Topics in modeling multilevel and longitudinal data*. [http:// www.luc.ac.be/censtat/pub/dr.pdf](http://www.luc.ac.be/censtat/pub/dr.pdf).
- [Venables e Ripley,2002] Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S-PLUS*. Springer,fourth edition.

[Yu e Cumberland, Technical report] Yu, H. and Cumberland, W. *Logistic regression models for analysis of multistage survey data*. Technical report, University of California.