Modelos Lineares Generalizados
para Análise de Dados com
Medidas Repetidas

MARIA KELLY VENEZUELA

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

Área de Concentração: **Estatística**

Orientadora: **Profa. Dra. Denise Aparecida Botter**

São Paulo, julho de 2003.

Modelos Lineares Generalizados para Análise de Dados com Medidas Repetidas

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por Maria Kelly Venezuela
e aprovada pela comissão julgadora.

São Paulo, 04 de julho de 2003.

Banca Examinadora:

- Profa. Dra. Denise Aparecida Botter (orientadora) - IME/USP
- Prof. Dr. Rinaldo Artes - IME/USP
- Profa. Dra. Antonieta D´Alcântara de Queiroz Peres - UNIRIO

*“A mente que se abre a uma nova idéia,
jamais volta ao seu tamanho original”.
(Albert Einstein)*

*Aos meus pais e
ao meu marido.*

Agradecimentos

Durante todo o tempo que me dediquei a este trabalho, contei com a ajuda e compreensão de muitas pessoas, entre as quais agradeço especialmente:

Às professoras **Denise Aparecida Botter** e **Mônica Carneiro Sandoval**, pela orientação, confiança e apoio dedicados a mim e a este trabalho.

Aos meus pais **Batista** e **Terezinha**, pelo amor, educação e paciência que nunca hesitaram em me oferecer. Ao meu irmão **Sandro**, em quem sempre me espelhei com relação aos estudos.

Ao meu marido **Gilberto**, pela compreensão, companheirismo e pelo apoio nos momentos mais difíceis.

Aos meus queridos amigos **Bianca**, **Fabio**, **Jacqueline** e **Marcos**, pelas valiosas discussões e, principalmente, pela amizade tornando os momentos de convívio no IME muito mais agradáveis e divertidos.

A todos os meus amigos e parentes que sempre compreenderam a minha ausência ou cansaço em algumas confraternizações.

À FAPESP, pela bolsa de mestrado concedida (Processo 01/00098-3), com a qual pude me dedicar com mais afinco a este trabalho. E pelo apoio financeiro do projeto temático da FAPESP “Modelos de Regressão e Aplicações” (Processo 99/10611-8) e do projeto do PRONEX “Métodos Estatísticos em Modelos de Regressão” (Convênio No. 76.97.1081.00).

Resumo

Neste trabalho, apresentamos as equações de estimação generalizadas desenvolvidas por Liang e Zeger (1986), sob a ótica da teoria de funções de estimação apresentada por Godambe (1991). Essas equações de estimação são obtidas para os modelos lineares generalizados (MLGs) considerando medidas repetidas. Apresentamos também um processo iterativo para estimação dos parâmetros de regressão, assim como testes de hipóteses para esses parâmetros. Para a análise de resíduos, generalizamos para dados com medidas repetidas algumas técnicas de diagnóstico usuais em MLGs. O gráfico de probabilidade meio-normal com envelope simulado é uma proposta para avaliarmos a adequação do ajuste do modelo. Para a construção desse gráfico, simulamos respostas correlacionadas por meio de algoritmos que descrevemos neste trabalho. Por fim, realizamos aplicações a conjuntos de dados reais.

Abstract

In this work, we consider the generalized estimation equations developed by Liang and Zeger (1986) focusing the theory of estimating functions presented by Godambe (1991). These estimation equations are an extension of generalized linear models (GLMs) to the analysis of repeated measurements. We present an iterative procedure to estimate the regression parameters as well as hypothesis testing of these parameters. For the residual analysis, we generalize to repeated measurements some diagnostic methods available for GLMs. The half-normal probability plot with a simulated envelope is useful for diagnosing model inadequacy and detecting outliers. To obtain this plot, we consider an algorithm for generating a set of nonnegatively correlated variables having a specified correlation structure. Finally, the theory is applied to real data sets.

Sumário

1	Introdução	1
1.1	Funções de Estimação	4
2	Equações de Estimação para Modelos Lineares Generalizados com Medidas Repetidas	10
2.1	Equação de Estimação Independente	11
2.2	Equação de Estimação Generalizada	13
2.3	Estimação de β , α e ϕ	18
2.4	Testes de Hipóteses	22
3	Técnicas de Diagnóstico	24
3.1	Pontos Alavanca, Influentes e Aberrantes	25
3.2	Envelope Simulado	29
3.3	Geração de variáveis correlacionadas	31
3.3.1	Variáveis Binárias	32
3.3.2	Variáveis Poisson e Gama	38
4	Aplicações	44
4.1	Aplicação 1	44
4.2	Aplicação 2	51
4.3	Aplicação 3	55

5	Conclusões e Estudos Futuros	60
A	Conjuntos de Dados	62
B	Rotinas Desenvolvidas no S-Plus	68
B.1	Aplicação 1	68
B.2	Aplicação 2	75
B.3	Aplicação 3	83

Lista de Tabelas

3.1	Valores referentes ao exemplo para gerar variáveis binárias.	37
3.2	Valores referentes ao exemplo para gerar variáveis Poisson ou gama. . .	42
4.1	Estimativas e erros padrão dos parâmetros do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.	48
4.2	Variação nas estimativas dos parâmetros do modelo ajustado com estrutura de correlação AR-1 e sem as unidades experimentais 1, 33 e 39.	50
4.3	Estimativas e erros padrão dos parâmetros do modelo de regressão logística ajustado com a estrutura de correlação não-estruturada.	52
4.4	Estimativas e erros padrão dos parâmetros do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.	56
A.1	Conjunto de dados referentes à Aplicação 1.	62
A.2	Conjunto de dados referentes à Aplicação 2.	64
A.3	Conjunto de dados referentes à Aplicação 3.	67

Lista de Figuras

4.1	Distância de Cook do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.	46
4.2	Resíduo padronizado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.	46
4.3	Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.	47
4.4	Distância de Cook do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.	49
4.5	Resíduo padronizado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.	49
4.6	Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.	50
4.7	Distância de Cook do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.	53
4.8	Resíduo padronizado do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.	54

4.9	Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.	54
4.10	Medida h_{ij} do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.	57
4.11	Medida $\mathbf{h}_i$ do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.	58
4.12	Distância de Cook do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.	58
4.13	Resíduo padronizado do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.	59
4.14	Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1. . .	59
B.1	Formato do <i>data.frame</i> Pratica criado no software S-Plus 4.5.	69
B.2	Formato do <i>data.frame</i> Depressao criado no software S-Plus 4.5.	76
B.3	Formato do <i>data.frame</i> MontPoisson criado no software S-Plus 4.5. . .	83

Capítulo 1

Introdução

Experimentos com medidas repetidas são freqüentes em quase todos os campos científicos, tais como, agricultura, economia, educação, engenharia e medicina, por exemplo. Nesses estudos, as unidades experimentais podem ser pessoas, animais, plantas, algum produto em teste, entre outros. Além disto, podem estar classificadas em diferentes grupos segundo um ou mais fatores, tais como, sexo, região, marca, etc. Quando cada unidade experimental é submetida a diferentes condições experimentais, aqui nomeadas de *tratamentos*, geram as *medidas repetidas* (valores de uma ou mais variáveis respostas, categorizadas ou contínuas). Sendo assim, temos independência entre unidades experimentais e dependência entre as observações de uma mesma unidade experimental.

Em estudos com medidas repetidas, os tratamentos podem ser aplicados às unidades experimentais de forma aleatória quando estes não devem seguir uma ordem de aplicação, como por exemplo, verificar o efeito de três diferentes pastas de dentes no índice de placa bacteriana em um mesmo indivíduo. Porém, existem experimentos nos quais os tratamentos devem seguir uma ordem de aplicação, como por exemplo, avaliar mensalmente a taxa de glicose de um paciente, medir o peso de um animal a cada semana durante um mês ou acompanhar a evolução do efeito de uma droga diminuindo a concentração da dosagem da mesma em cada aplicação. Esses exemplos caracterizam o

estudo longitudinal, no qual não se pode aleatorizar a ordem dos tratamentos para cada unidade experimental.

Esses estudos, embora tenham como principal desvantagem o acompanhamento das unidades experimentais durante a realização do estudo, têm como principais vantagens os seguintes fatos:

- requerem menos unidades experimentais do que planejamentos completamente casualizados;
- proporcionam condições mais adequadas para o estudo de covariáveis que possam ter influência na variável resposta;
- melhoram, em geral, a precisão de contrastes associados às diferenças entre os valores médios da variável resposta de diferentes tratamentos e
- permitem o estudo da mudança do comportamento da resposta média da unidade experimental nos diferentes tratamentos (incorpora informação sobre a variação individual na análise).

Para análise de dados com medidas repetidas, existe uma grande variedade de técnicas quando a variável resposta segue uma distribuição normal. Entre elas, podemos destacar a análise de variância com medidas repetidas (Neter et al., 1996), análise uni/multivariada de perfis, análise de curvas de crescimento (Singer e Andrade, 1986) e modelos normais com efeitos aleatórios (Laird e Ware, 1982 e Andreoni, 1989).

O fato de existir uma versão multivariada da distribuição normal com boas propriedades do ponto de vista inferencial facilita o estudo desses modelos. Porém, abandonando a suposição de normalidade da variável resposta, uma série de dificuldades pode surgir devido à escassez de distribuições multivariadas alternativas com tais propriedades.

Justamente por esse motivo, não são muitas as técnicas existentes na literatura para se trabalhar com distribuições não-normais levando em consideração a dependência

entre observações da mesma unidade experimental. Modelos de regressão com respostas binárias são apresentados, por exemplo, por Stiratelli, Laird e Ware (1984), Anderson e Aitkin (1985), Ochi e Prentice (1984) e Koch et al. (1977). Ashby et al. (1992) apresentam um método para análise de dados categorizados correlacionados.

Liang e Zeger (1986) propõem analisar dados com medidas repetidas utilizando modelos lineares generalizados (MLGs) desenvolvidos por Nelder e Weddeburn (1972). Eles obtêm estimativas consistentes para os parâmetros de regressão, desde que o modelo de regressão esteja corretamente especificado, utilizando equações de estimação generalizadas e tratam os parâmetros de correlação como parâmetros de perturbação. Prentice e Zhao (1991) utilizam equações de estimação para obter estimativas consistentes dos parâmetros de regressão e de correlação e, neste caso, é necessário que tanto o modelo de regressão como a estrutura de correlação estejam corretamente especificados. Na literatura, a proposta de Liang e Zeger também é conhecida como EEG1 e a proposta de Prentice e Zhao, como EEG2.

Em nosso trabalho, vamos apresentar a proposta de Liang e Zeger (1986) supondo que todas as unidades experimentais foram observadas em todos os tratamentos.

No Capítulo 2 apresentamos as equações de estimação generalizadas para os parâmetros de interesse, sob a ótica da teoria de funções de estimação (Godambe, 1991). Inicialmente, adotamos a suposição de que as observações repetidas de uma mesma unidade experimental são independentes e, em seguida, incorporamos a informação sobre a estrutura de dependência entre observações da mesma unidade experimental. Um processo iterativo para estimação dos parâmetros envolvidos nas equações de estimação, assim como testes de hipóteses para os parâmetros de regressão, também são apresentados nesse capítulo. No Capítulo 3, generalizamos algumas medidas de diagnóstico usuais em MLGs para dados com medidas repetidas. Para avaliarmos a adequação do ajuste do modelo, construímos o gráfico de probabilidade meio-normal com envelope simulado. Para a construção desse gráfico, é necessário simular variáveis

correlacionadas e, para isso, descrevemos os algoritmos para gerar variáveis com distribuições normal, binomial, Poisson e gama correlacionadas. Finalmente, o Capítulo 4 é dedicado a aplicações práticas, nas quais utilizamos a teoria descrita no Capítulo 2, para a estimação dos parâmetros do modelo, e as técnicas de diagnóstico do Capítulo 3, para detectar possíveis pontos influentes e/ou aberrantes e avaliar a adequação do ajuste do modelo. Nesse capítulo, utilizamos o pacote computacional S-Plus 4.5 (Venables e Ripley, 1999) para os ajustes dos modelos e as análises de diagnóstico. No Apêndice A estão os conjuntos de dados utilizados no Capítulo 4 e, no Apêndice B, as rotinas desenvolvidas no pacote computacional S-Plus 4.5.

A seguir, introduzimos alguns conceitos básicos sobre funções de estimação.

1.1 Funções de Estimação

Nesta seção, introduzimos alguns conceitos sobre funções de estimação baseados em Artes (1997) e Jørgensen e Labouriau (1994).

Uma função de estimação é uma função dos dados ($\mathbf{y}$) e dos parâmetros de interesse ($\boldsymbol{\theta}$). Um ponto importante no estudo dessas funções é o estabelecimento de condições que garantam que os estimadores dos parâmetros envolvidos possuam boas propriedades. Em geral, desejamos que os estimadores sejam consistentes e tenham distribuição assintótica conhecida.

Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um espaço de probabilidade, com $\mathcal{X} \in \mathbf{R}$ e $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbf{R}^p\}$, para algum $p \in \mathbf{N}$ (p é um valor fixado referente à dimensão do espaço paramétrico Θ). Por definição, uma função $\boldsymbol{\psi} : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^p$, é uma função de estimação se para cada $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi}(\cdot; \boldsymbol{\theta}) = (\psi_1, \dots, \psi_p)^T$ é uma variável aleatória.

Assumindo a existência de uma amostra de n vetores aleatórios independentes $\mathbf{y}_i = (y_{i1}, \dots, y_{it})^T$, $i = 1, \dots, n$, e que a cada vetor esteja associada uma função de estimação

ψ_i , estendemos o conceito de função de estimação para a amostra por

$$\Psi_n(\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n \psi_i(\mathbf{y}_i; \boldsymbol{\theta}),$$

com dimensão $(p \times 1)$, sendo $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ um vetor $(N \times 1)$, $N = nt$.

A seguir, apresentamos algumas propriedades de funções de estimação denotando $\Psi_n(\mathbf{y}; \boldsymbol{\theta})$ por $\Psi_n(\boldsymbol{\theta})$, para simplificar a notação.

As funções de estimação $\Psi_n(\boldsymbol{\theta})$ e $\Phi_n(\boldsymbol{\theta})$ são *funções de estimação equivalentes* se

$$\Psi_n(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})\Phi_n(\boldsymbol{\theta}),$$

sendo $\mathbf{C}(\boldsymbol{\theta})$ uma matriz quadrada não singular e não estocástica. Isto é, $\Psi_n(\boldsymbol{\theta})$ e $\Phi_n(\boldsymbol{\theta})$ têm as mesmas raízes.

A função de estimação $\Psi_n(\boldsymbol{\theta})$ é uma *função de estimação não viciada* se

$$\mathbf{E}_{\boldsymbol{\theta}}(\Psi_n(\boldsymbol{\theta})) = \mathbf{0}.$$

Se todas as funções de estimação ψ_i , $i = 1, \dots, n$, são não viciadas, então a função de estimação Ψ_n baseada na amostra de tamanho n também será não viciada.

A *matriz de variabilidade* de uma função de estimação não viciada é dada por

$$\mathbf{V}_{\Psi}(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}(\Psi_n(\boldsymbol{\theta})\Psi_n^\top(\boldsymbol{\theta})),$$

e a *matriz de sensibilidade* de uma função de estimação não viciada é dada por

$$\mathbf{S}_{\Psi}(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial \boldsymbol{\theta}^\top} \Psi_n(\boldsymbol{\theta})\right),$$

e ambas têm dimensão $(p \times p)$.

Uma função de estimação $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)^\top : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^p$ é dita uma *função de estimação regular* quando as seguintes condições são satisfeitas para todo $\boldsymbol{\theta} \in \Theta$ e para $i, j = 1, \dots, p$:

- i. $E_{\boldsymbol{\theta}}(\boldsymbol{\psi}(\boldsymbol{\theta})) = \mathbf{0}$, isto é, a função é não viciada;
- ii. a derivada parcial de $\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\theta})$ com respeito a $\boldsymbol{\theta}$ existe quase certamente para todo $\mathbf{y} \in \mathcal{X}$;
- iii. é possível permutar o sinal de integração e diferenciação da seguinte forma:

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} \boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\theta}) dP_{\boldsymbol{\theta}} = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} [\boldsymbol{\psi}(\mathbf{y}; \boldsymbol{\theta})] dP_{\boldsymbol{\theta}}$$

- iv. $E_{\boldsymbol{\theta}}(\psi_i(\boldsymbol{\theta})\psi_j(\boldsymbol{\theta})) \in \mathbf{R}$ e $\mathbf{V}_{\boldsymbol{\psi}}(\boldsymbol{\theta})$ é positiva definida;
- v. $E_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial \theta_i} \psi_i(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} \psi_j(\boldsymbol{\theta})\right) \in \mathbf{R}$, com $l, k = 1, \dots, p$ e $\mathbf{S}_{\boldsymbol{\psi}}(\boldsymbol{\theta})$ é não singular.

Seja $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ uma função de estimação regular. Definimos a *matriz de informação de Godambe* de $\boldsymbol{\theta}$ associada a $\boldsymbol{\Psi}_n$ por

$$\mathbf{J}_{\boldsymbol{\Psi}}(\boldsymbol{\theta}) = \mathbf{S}_{\boldsymbol{\Psi}}(\boldsymbol{\theta})^{\top} \mathbf{V}_{\boldsymbol{\Psi}}^{-1}(\boldsymbol{\theta}) \mathbf{S}_{\boldsymbol{\Psi}}(\boldsymbol{\theta}),$$

com dimensão $(p \times p)$.

A matriz de informação de Godambe $\mathbf{J}_{\boldsymbol{\Psi}}(\boldsymbol{\theta})$ é igual a matriz de informação de Fisher se e somente se $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ é equivalente à função escore. Por exemplo, se $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ é a função escore e, portanto, satisfaz as condições para ser uma função de estimação regular, então $\mathbf{S}_{\boldsymbol{\Psi}}(\boldsymbol{\theta}) = -\mathbf{V}_{\boldsymbol{\Psi}}(\boldsymbol{\theta})$, fazendo com que sua matriz de informação de Godambe coincida com a matriz de informação de Fisher. De maneira geral, para todas as funções de estimação regulares que não sejam necessariamente funções escores, a informação de Godambe desempenha o papel da informação de Fisher.

Um conceito importante desenvolvido por Godambe (1960) é o de otimalidade de uma função de estimação regular. No caso de θ ser unidimensional, podemos definir uma função de estimação ótima como aquela cujas raízes possuem variância assintótica mínima. Esse conceito pode ser estendido para o caso multidimensional por meio da

introdução de alguma ordenação das matrizes de covariância assintóticas (Chandrasekar e Kale, 1984).

Sejam $\mathbf{Q}_i(\boldsymbol{\theta})$ matrizes não estocásticas de postos completos e $\mathbf{u}_i = \mathbf{u}_i(\mathbf{y}_i; \boldsymbol{\theta})$ vetores com média zero mutuamente independentes, $i = 1, \dots, n$. A classe das *funções de estimação aditivas* geradas por $\mathbf{u}_i$ é definida por

$$\mathcal{L}(\mathbf{u}) = \left\{ \boldsymbol{\Psi}_n(\boldsymbol{\theta}) \in \mathcal{R} : \boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{Q}_i(\boldsymbol{\theta}) \mathbf{u}_i(\mathbf{y}_i; \boldsymbol{\theta}) \right\},$$

sendo que $\mathcal{R}$ contém todas as funções regulares de $\boldsymbol{\theta}$ e $\mathbf{u} = (\mathbf{u}_1^\top, \dots, \mathbf{u}_n^\top)^\top$.

Crowder (1987) mostra que nessa classe $\mathcal{L}(\mathbf{u})$, a *função de estimação ótima* é dada por

$$\boldsymbol{\Psi}_n^*(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{Q}_i^*(\boldsymbol{\theta}) \mathbf{u}_i(\mathbf{y}_i; \boldsymbol{\theta}), \quad (1.1)$$

em que

$$\mathbf{Q}_i^*(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\theta}^\top} \right)^\top \text{Cov}_{\boldsymbol{\theta}}^{-1}(\mathbf{u}_i).$$

É importante ressaltar que se considerarmos $\mathbf{C}(\boldsymbol{\theta})\boldsymbol{\Psi}_n^*(\boldsymbol{\theta})$, sendo $\mathbf{C}(\boldsymbol{\theta})$ uma matriz não singular, essa nova função de estimação também será ótima.

O teorema descrito a seguir estabelece condições para a normalidade assintótica de estimadores obtidos a partir de funções de estimação regulares. Esse teorema trata do caso em que $\boldsymbol{\theta}$ é multidimensional e, além disso, os vetores resposta são mutuamente independentes, mas não identicamente distribuídos.

Teorema 1 *Considerando-se*

- a. $\mathbf{y}_i$, $i = 1, 2, \dots, n$, vetores aleatórios t -dimensionais independentes;
- b. $\boldsymbol{\psi}_i(\boldsymbol{\theta}) = (\psi_{i1}, \dots, \psi_{ip})^\top$, $i = 1, 2, \dots, n$, funções de estimação regulares;

c. $\Psi_n(\boldsymbol{\theta}) = \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta});$

d. para $\delta > 0$,

$$E_{\boldsymbol{\theta}} \left\{ \sup_{\|\mathbf{h}\| \leq \delta} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta} + \mathbf{h}) - \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta}) \right\| \right\} = \phi_{\delta_i},$$

$$\frac{1}{n} \sum_{i=1}^n \phi_{\delta_i} \rightarrow \phi_{\delta} \text{ e } \phi_{\delta} \rightarrow 0 \text{ quando } \delta \rightarrow 0;$$

e. para $\delta > 0$, existe ϵ , $0 < \epsilon < 1$, tal que

$$n^{-1-\epsilon} \sum_{i=1}^n E_{\boldsymbol{\theta}} \left\{ \sup_{\|\mathbf{h}\| \leq \delta} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta} + \mathbf{h}) - \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta}) - \phi_{\delta_i} \right\| \right\}^{1+\epsilon} \rightarrow \mathbf{0};$$

f. que para $\dot{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\psi}_i(\boldsymbol{\theta})$,

$$\frac{1}{n} \sum_{i=1}^n E \left\{ \dot{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) \right\} = \frac{1}{n} \sum_{i=1}^n \mathbf{S}_i(\boldsymbol{\theta}) \rightarrow \mathbf{S}(\boldsymbol{\theta}),$$

com $\mathbf{S}$ não singular e $\mathbf{S}_i = E \left\{ \dot{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) \right\}$;

g. que existe ϵ , $0 < \epsilon < 1$, tal que

$$n^{-1-\epsilon} \sum_{i=1}^n E \left| \dot{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) - \mathbf{S}_i(\boldsymbol{\theta}) \right|^{1+\epsilon} \rightarrow \mathbf{0};$$

h. que existe ϵ , $0 < \epsilon < 1$, tal que $\max_{1 \leq i \leq n} \max_{1 \leq j \leq t} E |\psi_{ij}|^{2+\epsilon} < \infty$;

i. $n^{-1} \sum_{i=1}^n \text{Cov}(\boldsymbol{\psi}_i) \rightarrow \mathbf{V}(\boldsymbol{\theta})$, positiva definida;

j. $\hat{\boldsymbol{\theta}}_n$ como a solução de $\boldsymbol{\Psi}(\mathbf{w}) = \mathbf{0}$, $\mathbf{w} \in \Theta$,

e sob condições que garantam a existência de uma seqüência de raízes de $\boldsymbol{\Psi}_n(\mathbf{w})$ que seja limitada em probabilidade ou restrita a um conjunto compacto q.c. quando n tende a infinito, vem que

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathcal{P}} \boldsymbol{\theta}.$$

Sob essas hipóteses,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_p(\boldsymbol{0}, \boldsymbol{J}_{\boldsymbol{\Psi}}^{-1}(\boldsymbol{\theta})),$$

em que

$$\begin{aligned} \boldsymbol{J}_{\boldsymbol{\Psi}}(\boldsymbol{\theta}) &= \left\{ \lim_{n \rightarrow \infty} \boldsymbol{S}_n^T(\boldsymbol{\theta}) \right\} \left\{ \lim_{n \rightarrow \infty} \boldsymbol{V}_n(\boldsymbol{\theta}) \right\}^{-1} \left\{ \lim_{n \rightarrow \infty} \boldsymbol{S}_n(\boldsymbol{\theta}) \right\} = \\ &= \left\{ \lim_{n \rightarrow \infty} \boldsymbol{S}_n^T(\boldsymbol{\theta}) \right\} \left\{ \lim_{n \rightarrow \infty} \boldsymbol{V}_n^{-1}(\boldsymbol{\theta}) \right\} \left\{ \lim_{n \rightarrow \infty} \boldsymbol{S}_n(\boldsymbol{\theta}) \right\} = \\ &= \lim_{n \rightarrow \infty} \left\{ \boldsymbol{S}_n^T(\boldsymbol{\theta}) \boldsymbol{V}_n^{-1}(\boldsymbol{\theta}) \boldsymbol{S}_n(\boldsymbol{\theta}) \right\}, \end{aligned}$$

com

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\boldsymbol{S}_{\psi_i}(\boldsymbol{\theta})}{n}, \quad \boldsymbol{S}_{\psi_i}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^T} \psi_i(y_i; \boldsymbol{\theta}) \right\},$$

e

$$\boldsymbol{V}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\boldsymbol{V}_{\psi_i}(\boldsymbol{\theta})}{n}, \quad \boldsymbol{V}_{\psi_i}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left\{ \boldsymbol{\Psi}_n(y_i; \boldsymbol{\theta}) \boldsymbol{\Psi}_n^T(y_i; \boldsymbol{\theta}) \right\},$$

desempenha o papel de uma matriz de informação de Godambe assintótica associada a $\boldsymbol{\Psi}_n$.

A prova e comentários sobre cada item do Teorema 1 estão descritos em Artes (1997, p.30).

Capítulo 2

Equações de Estimação para Modelos Lineares Generalizados com Medidas Repetidas

Neste capítulo, apresentamos as equações de estimação generalizadas desenvolvidas por Liang and Zeger (1986), utilizando a teoria de funções de estimação dada na Seção 1.1.

Sejam $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{it})^\top$ um vetor $(t \times 1)$ de respostas e $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{it})^\top$ uma matriz $(t \times p)$ de valores de covariáveis (ou matriz de planejamento) para a i -ésima unidade experimental, com $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^\top$, $i = 1, \dots, n$ e $j = 1, \dots, t$. Vamos assumir que a densidade marginal de y_{ij} pertence à família exponencial, isto é, sua densidade (ou função de probabilidade) é dada por

$$f(y_{ij}; \theta_{ij}, \phi) = \exp\left\{\phi[y_{ij}\theta_{ij} - b(\theta_{ij})] + c(y_{ij}, \phi)\right\}, \quad (2.1)$$

sendo $E(y_{ij}) = \mu_{ij} = b'(\theta_{ij})$, $\text{Var}(y_{ij}) = \phi^{-1}b''(\theta_{ij})$, $c(\cdot)$ uma função conhecida e ϕ^{-1} o parâmetro de dispersão conhecido. A segunda derivada de $b(\theta_{ij})$, $b''(\theta_{ij})$, é denominada função de variância e desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição.

As médias μ_{ij} são modeladas como nos MLGs, ou seja,

$$g(\mu_{ij}) = \eta_{ij}, \quad (2.2)$$

sendo $\eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta}$ o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $p < n$, o vetor de parâmetros desconhecidos a serem estimados e $g(\cdot)$ uma função monótona e diferenciável, denominada função de ligação.

2.1 Equação de Estimação Independente

Inicialmente, vamos considerar que as observações repetidas de uma mesma unidade experimental são independentes. Essa suposição resulta na equação de estimação independente (EEI) que definimos a seguir.

Consideremos a densidade (ou função de probabilidade) marginal de y_{ij} pertencente à família exponencial e o componente sistemático descritos, respectivamente, nas equações (2.1) e (2.2). Sejam $\mathbf{u}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$ vetores com média zero mutuamente independentes, com $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{it})^\top$ e $i = 1, \dots, n$. Assumindo independência entre as observações de uma mesma unidade experimental, os termos da função de estimação ótima definidos em (1.1) são, então, dados por

$$E_{\boldsymbol{\beta}} \left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\beta}^\top} \right)^\top = E_{\boldsymbol{\beta}} \left(\frac{\partial}{\partial \boldsymbol{\beta}^\top} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right)^\top = - \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \right)^\top = -\mathbf{X}_i^\top \boldsymbol{\Delta}_i \mathbf{A}_i = -\mathbf{D}_i^\top$$

e

$$\text{Cov}_{\boldsymbol{\beta}}(\mathbf{u}_i) = \text{Cov}(\mathbf{y}_i) = \text{diag}\{\text{Var}(y_{ij})\} = \phi^{-1} \text{diag}\{b''(\theta_{ij})\} = \phi^{-1} \mathbf{A}_i,$$

sendo

$$\mathbf{X}_i = \partial \boldsymbol{\eta}_i / \partial \boldsymbol{\beta} = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it})^\top, \quad \text{com dimensão } (t \times p),$$

$$\begin{aligned}\Delta_i &= \text{diag}\left\{\partial\theta_{ij}/\partial\eta_{ij}\right\}, \quad \text{com dimensão } (t \times t) \quad \text{e} \\ \mathbf{A}_i &= \text{diag}\left\{\partial\mu_{ij}/\partial\theta_{ij}\right\} = \text{diag}\left\{b''(\theta_{ij})\right\}, \quad \text{com dimensão } (t \times t),\end{aligned}$$

com $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{it})^\top$ e $j = 1, \dots, t$.

Notemos que $\mathbf{A}_i$ é uma matriz definida a partir da função de variância e Δ_i é uma matriz que depende da função de ligação utilizada na modelagem. Se a função de ligação é a canônica, $\theta_{ij} = \eta_{ij}$, então Δ_i será uma matriz identidade com dimensão $(t \times t)$.

A função de estimação ótima obtida dos termos acima é equivalente a

$$\boldsymbol{\Psi}_n^*(\boldsymbol{\beta}) = \phi \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad (2.3)$$

que é a função escore de um MLG. Esse resultado era esperado, uma vez que estamos considerando a suposição de independência entre as observações de uma mesma unidade experimental.

A equação obtida ao igualarmos uma função de estimação a zero é denominada *equação de estimação*. Logo, a *equação de estimação independente* de $\boldsymbol{\beta}$ é dada por

$$\boldsymbol{\Psi}_n^{I*}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \quad (2.4)$$

Sob condições gerais de regularidade (ver Teorema 1 na Seção 1.1), podemos demonstrar que a raiz da EEI, $\hat{\boldsymbol{\beta}}_{I*}$, é um estimador consistente de $\boldsymbol{\beta}$ e que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{I*} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{J}_{I*}^{-1}),$$

com

$$\mathbf{J}_{I*} = \lim_{n \rightarrow \infty} \frac{\mathbf{J}_{nI*}}{n},$$

sendo $\mathbf{J}_{nI^*}$ a matriz de informação de Godambe de $\boldsymbol{\beta}$ associada a $\boldsymbol{\Psi}_n^{I^*}$, dada por

$$\begin{aligned}\mathbf{J}_{nI^*} &= \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} = \\ &= \phi \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{A}_i^{-1} \mathbf{D}_i,\end{aligned}$$

pois

$$\begin{aligned}\mathbf{S}_i &= \mathbf{E}_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{D}_i^\top \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = \\ &= \mathbf{D}_i^\top \mathbf{A}_i^{-1} \mathbf{E}_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = -\mathbf{D}_i^\top \mathbf{A}_i^{-1} \mathbf{D}_i\end{aligned}$$

e

$$\begin{aligned}\mathbf{V}_i &= \mathbf{E}_{\boldsymbol{\beta}} \left[\mathbf{D}_i^\top \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \mathbf{A}_i^{-1} \mathbf{D}_i \right] = \\ &= \mathbf{D}_i^\top \mathbf{A}_i^{-1} \text{Cov}(\mathbf{y}_i) \mathbf{A}_i^{-1} \mathbf{D}_i = \phi^{-1} \mathbf{D}_i^\top \mathbf{A}_i^{-1} \mathbf{A}_i \mathbf{A}_i^{-1} \mathbf{D}_i = -\phi^{-1} \mathbf{S}_i.\end{aligned}$$

Um estimador consistente para a matriz de covariâncias de $\hat{\boldsymbol{\beta}}_{I^*}$, quando as observações de uma mesma unidade experimental são independentes, é dado por

$$\hat{\mathbf{J}}_{nI^*}^{-1} = \left\{ \phi \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\mathbf{A}}_i^{-1} \hat{\mathbf{D}}_i \right\}^{-1},$$

sendo todas as quantidades avaliadas em $\hat{\boldsymbol{\beta}}_{I^*}$. Notemos que, como $\boldsymbol{\Psi}_n^{I^*}(\boldsymbol{\beta})$ é uma função escore, a matriz de informação de Godambe de $\boldsymbol{\beta}$ é igual a matriz de informação de Fisher de $\boldsymbol{\beta}$.

2.2 Equação de Estimação Generalizada

Para obtermos a equação de estimação generalizada (EEG), introduziremos modificações em $\boldsymbol{\psi}_n^{I^*}(\cdot)$ dada em (2.4) de modo que a nova função de estimação incorpore

alguma informação sobre a estrutura de dependência dos dados. Os termos da função de estimação ótima definidos em (1.1) são dados por

$$E_{\beta} \left(\frac{\partial \mathbf{u}_i}{\partial \beta^T} \right)^T = E_{\beta} \left(\frac{\partial}{\partial \beta^T} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right)^T = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \beta^T} \right)^T = -\mathbf{X}_i^T \boldsymbol{\Delta}_i \mathbf{A}_i = -\mathbf{D}_i^T$$

e

$$\begin{aligned} \text{Cov}_{\beta}(\mathbf{u}_i) &= \text{Cov}(\mathbf{y}_i) = \text{diag}\{\text{Var}(y_{ij})^{1/2}\} \mathbf{R}_i \text{diag}\{\text{Var}(y_{ij})^{1/2}\} = \\ &= \phi^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2} = \boldsymbol{\Sigma}_i, \end{aligned}$$

sendo $\mathbf{R}_i$ a **verdadeira** matriz de correlação de $\mathbf{y}_i$, $i = 1, \dots, n$.

Logo, a equação de estimação de β quando consideramos a verdadeira matriz de correlação é equivalente a

$$\boldsymbol{\Psi}_n^{G*}(\beta) = \sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \quad (2.5)$$

Sob condições gerais de regularidade (ver Teorema 1 na Seção 1.1), podemos demonstrar que $\hat{\beta}_{G*}$, raiz de (2.5), é um estimador consistente de β e que

$$\sqrt{n}(\hat{\beta}_{G*} - \beta) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{J}_{G*}^{-1}).$$

com

$$\mathbf{J}_{G*} = \lim_{n \rightarrow \infty} \frac{\mathbf{J}_{nG*}}{n},$$

sendo $\mathbf{J}_{nG*}$ a matriz de informação de Godambe de β associada a $\boldsymbol{\Psi}_n^{G*}$, dada por

$$\begin{aligned} \mathbf{J}_{nG*} &= \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} = \\ &= \sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i, \end{aligned}$$

pois

$$\begin{aligned} \mathbf{S}_i &= \mathbb{E}_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = \\ &= \mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbb{E}_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = -\mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \end{aligned}$$

e

$$\begin{aligned} \mathbf{V}_i &= \mathbb{E}_{\boldsymbol{\beta}} \left[\mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \right] \\ &= \mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} \text{Cov}(\mathbf{y}_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i = \mathbf{D}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i = -\mathbf{S}_i. \end{aligned}$$

Neste caso, um estimador consistente para a matriz de covariâncias de $\hat{\boldsymbol{\beta}}_{G^*}$ é dado por

$$\hat{\mathbf{J}}_{nG^*}^{-1} = \left\{ \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{D}}_i \right\}^{-1}, \quad (2.6)$$

sendo todas as quantidades avaliadas em $\hat{\boldsymbol{\beta}}_{G^*}$. Na literatura, esse estimador recebe os nomes de *estimador “naive”* ou *“model-based”*.

Na prática, a função $\boldsymbol{\Psi}_n^{G^*}(\cdot)$ é pouco utilizada, pois, em geral, a verdadeira matriz de correlação, $\mathbf{R}_i$, é desconhecida. Para contornar esse problema, Liang and Zeger (1986) definiram $\mathbf{R}(\boldsymbol{\alpha})$ como sendo uma matriz simétrica $(t \times t)$ satisfazendo as condições para ser uma matriz de correlação, denominada *matriz de correlação de trabalho* em que $\boldsymbol{\alpha}$, um vetor $(s \times 1)$, caracteriza completamente $\mathbf{R}(\boldsymbol{\alpha})$. Notemos que $\mathbf{R}(\boldsymbol{\alpha})$ não precisa ser necessariamente a verdadeira matriz de correlação dos $\mathbf{y}_i$'s.

Com isso, a *equação de estimação generalizada* de $\boldsymbol{\beta}$ é dada por

$$\boldsymbol{\Psi}_n^G(\boldsymbol{\beta}) = \boldsymbol{\Psi}_n^G(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)) = \sum_{i=1}^n \mathbf{D}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (2.7)$$

sendo

$$\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\alpha}, \phi) = \phi^{-1} \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}, \quad (2.8)$$

e $\hat{\alpha}$ um estimador consistente de α .

A equação de estimação dada em (2.7) deixa de ser a ótima e, por isso, pode ser viciada. É necessário, então, que $\hat{\alpha}$ seja determinado de modo que o estimador de β , obtido a partir de (2.7), continue sendo consistente e assintoticamente normal.

Na prática, ϕ quase sempre é desconhecido. Logo, também é necessário propor um estimador para esse parâmetro.

O próximo teorema descreve algumas condições que os estimadores de α e ϕ devem satisfazer de modo que o estimador de β preserve as propriedades citadas acima.

Teorema 2 *Seja $\hat{\beta}_G$ a raiz de (2.7). Sob condições gerais de regularidade e assumindo que*

- a. $\hat{\alpha}(\beta, \phi)$ é um estimador $\sqrt{n}$ -consistente de α dados β e ϕ ;
- b. $\hat{\phi}(\beta)$ é um estimador $\sqrt{n}$ -consistente de ϕ dado β ; e
- c. $|\partial \hat{\alpha}(\beta, \phi)/\partial \phi| \leq \mathbf{H}(\mathbf{y}, \beta)$, sendo $\mathbf{H}(\mathbf{y}, \beta)$ uma função $Op(1)$,

então temos que $\hat{\beta}_G$ é um estimador consistente de β e

$$\sqrt{n}(\hat{\beta}_G - \beta) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{J}_G^{-1}),$$

com

$$\mathbf{J}_G = \lim_{n \rightarrow \infty} \frac{\mathbf{J}_{nG}}{n},$$

sendo $\mathbf{J}_{nG}$ a matriz de informação de Godambe de β associada a Ψ_n^G , dada por

$$\mathbf{J}_{nG} = \left\{ \sum_{i=1}^n \mathbf{s}_i \right\} \left\{ \sum_{i=1}^n \mathbf{v}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{s}_i \right\},$$

em que

$$\begin{aligned} \mathbf{S}_i &= E_{\beta} \left[\frac{\partial}{\partial \beta^T} \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = \\ &= \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} E_{\beta} \left[\frac{\partial}{\partial \beta^T} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] = -\mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \end{aligned}$$

e

$$\begin{aligned} \mathbf{V}_i &= E_{\beta} \left[\mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right] \\ &= \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \text{Cov}(\mathbf{y}_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i. \end{aligned}$$

A prova desse resultado encontra-se em Liang and Zeger (1986).

Vale enfatizar que os resultados assintóticos do Teorema 2 são válidos mesmo quando $\mathbf{R}(\boldsymbol{\alpha})$ não corresponde à verdadeira matriz de correlação de $\mathbf{y}_i$.

A matriz de covariâncias de $\hat{\boldsymbol{\beta}}_G$ pode ser consistentemente estimada por

$$\hat{\mathbf{J}}_{nG}^{-1} = \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1} \left\{ \sum_{i=1}^n \hat{\mathbf{D}}_i^T \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i \right\} \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1}. \quad (2.9)$$

A estimativa da expressão acima é obtida substituindo $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e ϕ por suas respectivas estimativas consistentes. Na literatura especializada esse estimador recebe os nomes de *estimador robusto*, *empírico* ou *sanduíche*. Quando $\mathbf{R}(\boldsymbol{\alpha})$ for a verdadeira matriz de correlação dos $\mathbf{y}_i$'s, então $\boldsymbol{\Omega}_i = \text{Cov}(\mathbf{y}_i)$ e o estimador robusto será igual ao estimador “naive” expresso em (2.6).

Considerando que o modelo de regressão está corretamente especificado, o estimador “naive” é consistente se a matriz de correlação de trabalho também está corretamente especificada. Já o estimador robusto é sempre consistente. Além disso, o estimador robusto é assintoticamente não viesado, mas pode ser altamente viesado quando o número de unidades experimentais (n) for pequeno. Quando $n < 20$, o estimador

“naive” pode ter melhores propriedades (Prentice, 1988) mesmo se $\mathbf{R}(\boldsymbol{\alpha})$ não estiver corretamente especificada.

Uma indicação da adequação de uma matriz $\mathbf{R}(\boldsymbol{\alpha})$ ao modelo ocorre quando as duas estimativas, “naive” e robusta, são próximas (Johnston, 1996).

Liang e Zeger (1986) afirmam que a equação de estimação generalizada se reduz à função escore quando assumimos que y_{ij} tem distribuição normal. Park (1993) mostra, para matriz de correlação não-estruturada, que isso ocorre apenas quando o conjunto de dados está completo, isto é, todas as unidades experimentais têm observações nos t tratamentos. Além disso, a densidade marginal de y_{ij} deve ser função de ϕ_j , com $j = 1, \dots, t$, e não apenas de ϕ como apresentamos na equação (2.1). O autor faz essa alteração para permitir heterocedasticidade de y_{ij} , pois, quando a densidade marginal de y_{ij} é normal e escrita conforme a equação (2.1), estamos supondo homocedasticidade, não importando a estrutura de correlação considerada no ajuste do modelo.

2.3 Estimação de $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e ϕ

O processo iterativo para calcular $\hat{\boldsymbol{\beta}}_G$ combina o método *scoring* de Fisher para estimar $\boldsymbol{\beta}$ com o método dos momentos para estimar $\boldsymbol{\alpha}$ e ϕ . Logo, expandindo a EEG dada na equação (2.7) em torno de um valor inicial $\boldsymbol{\beta}_G^{(0)}$, o processo iterativo para estimar $\boldsymbol{\beta}$ é dado por

$$\begin{aligned}\hat{\boldsymbol{\beta}}_G^{(m+1)} &= \hat{\boldsymbol{\beta}}_G^{(m)} - \left\{ \mathbf{E}_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} \boldsymbol{\Psi}_n^G(\hat{\boldsymbol{\beta}}_G^{(m)}) \right] \right\}^{-1} \boldsymbol{\Psi}_n^G(\hat{\boldsymbol{\beta}}_G^{(m)}) = \\ &= \hat{\boldsymbol{\beta}}_G^{(m)} + \left\{ \left[\sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i \right]^{-1} \left[\sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) \right] \right\}^{(m)}, \quad (2.10)\end{aligned}$$

sendo $m = 0, 1, 2, \dots$ o número de iterações. O índice m no lado direito das equações acima indica que as matrizes e os vetores são atualizados pelas estimativas de $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e ϕ da m -ésima iteração.

Lembrando que a matriz $\mathbf{D}_i$ é dada pela multiplicação das matrizes $\mathbf{A}_i \mathbf{\Delta}_i \mathbf{X}_i$, podemos reescrever a expressão (2.10) da seguinte forma:

$$\hat{\boldsymbol{\beta}}_G^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{z}_i \right] \right\}^{(m)}, \quad (2.11)$$

com $\hat{\mathbf{W}}_i = \hat{\mathbf{\Delta}}_i \hat{\mathbf{A}}_i \hat{\mathbf{\Omega}}_i^{-1} \hat{\mathbf{A}}_i \hat{\mathbf{\Delta}}_i$ e $\mathbf{z}_i = \hat{\boldsymbol{\eta}}_i + (\hat{\mathbf{A}}_i \hat{\mathbf{\Delta}}_i)^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)$.

Notemos que chegamos a um processo iterativo de mínimos quadrados ponderados em que $\mathbf{z}$ desempenha o papel de uma variável dependente modificada e $\mathbf{W}$ é uma matriz de pesos. Ambos mudam a cada passo do processo iterativo.

Liang e Zeger (1986) utilizam o método dos momentos para estimar os parâmetros de correlação $\boldsymbol{\alpha}$ e o parâmetro de escala ϕ , e os escrevem em função dos resíduos de Pearson. Com isso, para cada observação y_{ij} , o resíduo de Pearson obtido no passo m é dado por

$$\hat{r}_{ij}^{(m)} = \frac{y_{ij} - \hat{\mu}_{ij}^{(m)}}{\sqrt{\hat{a}_{ij}^{(m)}}},$$

sendo $\hat{a}_{ij}$ o j -ésimo elemento da diagonal principal de $\hat{\mathbf{A}}_i$.

A estimativa de ϕ obtida no m -ésimo passo do processo iterativo é dada por

$$\hat{\phi}^{(m)} = \left\{ \sum_{i=1}^n \left(\hat{r}_{ij}^{(m)} \right)^2 / (nt - p) \right\}^{-1}.$$

O procedimento da EEG para estimar $\boldsymbol{\beta}$ permite que a estrutura de correlação entre as observações da mesma unidade experimental seja especificada de diferentes formas. A seguir, apresentamos algumas estruturas usadas para definir $\mathbf{R}(\boldsymbol{\alpha})$. Notemos, entretanto, que para qualquer $\mathbf{R}(\boldsymbol{\alpha})$ dada, $\hat{\boldsymbol{\beta}}_G$ e $\hat{\mathbf{J}}_{nG}^{-1}$ serão consistentes e a eficiência cresce quanto mais próxima da verdadeira matriz de correlação estiver a escolha da matriz de correlação de trabalho.

1. A matriz de correlação padrão uniforme assume que $\text{Corr}(y_{ij}, y_{il}) = \alpha, \forall j \neq l$ e $1 \leq j, l \leq t$. Então, a estimativa de α no passo m é dada por

$$\hat{\alpha}^{(m)} = \hat{\phi}^{(m)} \sum_{i=1}^n \sum_{j>l}^t \hat{r}_{ij}^{(m)} \hat{r}_{il}^{(m)} / \left(\frac{1}{2} nt(t-1) - p \right).$$

2. A matriz de correlação 1-dependente especifica que $\text{Corr}(y_{ij}, y_{i(j+1)}) = \alpha_j, j = 1, 2, \dots, t-1$. Um estimador natural de α_j é dado por

$$\hat{\alpha}_j^{(m)} = \frac{\hat{\phi}^{(m)}}{n-p} \sum_{i=1}^n \hat{r}_{ij}^{(m)} \hat{r}_{i(j+1)}^{(m)}.$$

Como um caso especial, seja $\alpha_j = \alpha, j = 1, 2, \dots, t-1$. Então, α no passo m pode ser estimado por

$$\hat{\alpha}^{(m)} = \frac{1}{t-1} \sum_{j=1}^t \hat{\alpha}_j^{(m)}.$$

3. A matriz de correlação autoregressiva de primeira ordem, AR-1, especifica que $\text{Corr}(y_{ij}, y_{il}) = \alpha^{|j-l|}, 1 \leq j, l \leq t$. Para y_{ij} com distribuição Normal, $E(\hat{r}_{ij} \hat{r}_{il}) \cong \alpha^{|j-l|}$. Então, α no passo m pode ser estimado pelo coeficiente angular da regressão em que a variável dependente é $\log(\hat{r}_{ij}^{(m)} \hat{r}_{il}^{(m)})$ e a independente é $\log |j-l|$.

4. Quando a matriz de correlação é a não estruturada, a mesma pode ser estimada por

$$\mathbf{R}(\hat{\alpha}^{(m)}) = \frac{1}{n} \left[\hat{\phi} \sum_{i=1}^n \hat{\mathbf{A}}_i^{-1/2} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top \hat{\mathbf{A}}_i^{-1/2} \right]^{(m)}.$$

A seguir, apresentamos as etapas que nos mostram como na prática podemos estimar os parâmetros de $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e ϕ .

Etapas para estimação de β , α e ϕ .

1. Supondo independência entre as observações da mesma unidade experimental, utilizamos a equação (2.11) para estimar β . O processo iterativo pode ser iniciado considerando $\hat{\mu}_{ij}^{(0)} = y_{ij}$ ou, no caso, por exemplo, de uma regressão logística, $\hat{\mu}_{ij}^{(0)} = 0,5$. Notemos que a suposição de independência elimina os parâmetros α e ϕ do processo de estimação de β .
2. Definimos uma estrutura de correlação de trabalho para ser utilizada na modelagem dos dados. Se a estrutura for a independente, a estimativa de β é o vetor de valores obtidos na convergência do processo iterativo da etapa 1. Caso contrário, utilizamos novamente a equação (2.11), considerando como $\hat{\beta}_G^{(0)}$ a estimativa de β encontrada na etapa 1. Calculamos o resíduo de Pearson ($\hat{r}_{ij}^{(0)}$), o parâmetro de escala ($\hat{\phi}^{(0)}$) e os parâmetros de correlação ($\hat{\alpha}^{(0)}$). No passo seguinte do processo iterativo, calculamos os parâmetros de regressão ($\hat{\beta}_G^{(m)}$), o resíduo de Pearson ($\hat{r}_{ij}^{(m)}$), o parâmetro de escala ($\hat{\phi}^{(m)}$) e os parâmetros de correlação ($\hat{\alpha}^{(m)}$). Repetimos esse passo do processo até a convergência.

■

Em um dos exemplos apresentados em Liang e Zeger (1986), cujo modelo envolve um único parâmetro, os autores comparam a eficiência assintótica dos estimadores $\hat{\beta}_I$ e $\hat{\beta}_G(\text{AR-1})$, obtidos da equação (2.11) utilizando, respectivamente, a matriz de correlação de trabalho independente e a auto-regressiva de ordem 1. Os autores concluem que quanto mais forte for a verdadeira correlação entre as observações da mesma unidade experimental, mais eficiente é o estimador $\hat{\beta}_G(\text{AR-1})$. Quando essa correlação não é tão forte, no caso $|\alpha| \leq 0,2$, as eficiências obtidas para os estimadores $\hat{\beta}_I$ e $\hat{\beta}_G(\text{AR-1})$ são iguais.

Quando há ausência de algumas observações no vetor $\mathbf{y}$, as equações de estimação produzem estimadores não viciados dos parâmetros de regressão apenas se esses dados

ausentes são completamente aleatórios (Rubin, 1976). Uma aproximação geral para o cálculo da magnitude do vício dos estimadores obtidos pelas equações de estimação quando há dados incompletos é apresentado por Rotnitzky e Wypij (1994). Ziegler, Kastner e Blettner (1998) indicam diversas aproximações para resolver as equações de estimação generalizadas quando temos estruturas com dados incompletos.

2.4 Testes de Hipóteses

Apresentamos, a seguir, as estatísticas de Wald e escore generalizadas para testar hipóteses do tipo

$$H_0 : \gamma = \gamma_0 \quad \text{versus} \quad H_1 : \gamma \neq \gamma_0, \quad (2.12)$$

sendo γ um vetor $(q \times 1)$ referente a quaisquer q componentes do vetor β , $(1 \leq q \leq p)$. Ou seja, podemos particionar o vetor de coeficientes de regressão na forma $\beta = (\gamma^\top, \delta^\top)^\top$, em que γ contém os parâmetros que caracterizam a hipótese de interesse descrita acima.

A estatística de Wald generalizada é dada por

$$Q_W = (\hat{\gamma}_G - \gamma_0)^\top \hat{\mathbf{J}}_{n\gamma_G}^{-1} (\hat{\gamma}_G - \gamma_0), \quad (2.13)$$

sendo $\hat{\gamma}_G$ o vetor $(q \times 1)$ dos q primeiros componentes de $\hat{\beta}_G$, e $\hat{\mathbf{J}}_{n\gamma_G}^{-1}$ a submatriz $(q \times q)$ do estimador robusto $\hat{\mathbf{J}}_{nG}^{-1}$ expresso em (2.9).

A estatística escore generalizada é dada por

$$Q_S = \Psi_n^G(\gamma_0)^\top \hat{\mathbf{V}}_{n\gamma_0}^{-1} \Psi_n^G(\gamma_0), \quad (2.14)$$

sendo $\Psi_n^G(\gamma_0)$ o subvetor $(q \times 1)$ de Ψ_n^G e $\hat{\mathbf{V}}_{n\gamma_0}$ a submatriz $(q \times q)$ da matriz de variabilidade expressa no estimador robusto dado em (2.9), avaliados no vetor γ_0 .

Considerando que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{J}_G^{-1}),$$

quando $n \rightarrow \infty$, temos, sob a hipótese nula (2.12), que Q_W e Q_S têm distribuição assintótica χ_q^2 (qui-quadrado com q graus de liberdade).

Conforme já mencionamos, o estimador robusto $\hat{\mathbf{J}}_{nG}^{-1}$ pode apresentar um viés considerável se o número de unidades experimentais for pequeno. Neste caso, as estatísticas (2.13) e (2.14) que dependem desse estimador robusto podem ficar comprometidas. Rotnitzky e Jewell (1990) apresentam estatísticas de Wald e escore modificadas que dependem do estimador “naive” dado na equação (2.6). Os autores também apresentam uma técnica para verificar a adequação da matriz de correlação de trabalho considerada no modelo, utilizando essas estatísticas modificadas.

Cardoso (2000) apresenta testes de hipóteses restritos em desigualdades lineares usando equações de estimação generalizadas, os quais são convenientes se estivermos trabalhando com modelos de regressão com algum tipo de restrição nos parâmetros do modelo.

Capítulo 3

Técnicas de Diagnóstico

A análise de diagnóstico é uma etapa importante do ajuste de um modelo de regressão, pois nos auxilia na verificação de possíveis afastamentos das suposições feitas para o modelo e nos permite detectar observações extremas que interferem desproporcionalmente nos resultados do ajuste.

As técnicas de diagnóstico já se encontram muito bem definidas para os modelos lineares generalizados (Paula, 2001). Para esses modelos, os elementos da diagonal principal da matriz de projeção (matriz chapéu), a distância de Cook e os resíduos do modelo ajustado são algumas das medidas existentes na literatura, utilizadas para detectar observações influentes na matriz de covariáveis (pontos alavanca) ou no vetor de respostas e para detectar pontos aberrantes (“outliers”). Outro recurso disponível é o gráfico de probabilidade meio-normal com envelope simulado (Neter et al, 1996).

Tan, Qu e Kutner (1997) propõem essas técnicas de diagnóstico para modelos de regressão logística, levando em consideração a estrutura de dependência entre observações da mesma unidade experimental. Com base no trabalho desses autores, desenvolvemos uma generalização dessas técnicas para os MLG’s em que consideramos dados com medidas repetidas, a qual é apresentada na próxima seção.

Entre outros trabalhos que tratam de técnicas de diagnóstico para EEGs, temos Chang (2000) que apresenta um teste não-paramétrico para avaliar a aleatoriedade dos resíduos, Pan (2001) que apresenta medidas para escolher a melhor matriz de correlação de trabalho e para selecionar covariáveis, baseadas no critério de informação de Akaike (AIC) e, por fim, citamos Preisser e Qaqish (1996) que apresentam medidas para detectar observações e/ou unidades experimentais influentes em MLG's com medidas repetidas.

3.1 Pontos Alavanca, Influentes e Aberrantes

Na convergência do processo iterativo para estimar β dado em (2.11), temos

$$\hat{\beta}_G = \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{z}_i \right], \quad (3.1)$$

com $\hat{\mathbf{W}}_i = \hat{\Delta}_i \hat{\mathbf{A}}_i \hat{\Omega}_i^{-1} \hat{\mathbf{A}}_i \hat{\Delta}_i$ e $\mathbf{z}_i = \hat{\eta}_i + (\hat{\mathbf{A}}_i \hat{\Delta}_i)^{-1}(\mathbf{y}_i - \hat{\mu}_i)$.

Com o intuito de eliminar o somatório da equação (3.1) e simplificar essa expressão, definimos $\hat{\mathbf{W}}$ como uma matriz bloco diagonal dada por $\hat{\mathbf{W}} = \text{diag}(\hat{\mathbf{W}}_1, \dots, \hat{\mathbf{W}}_n)$, com dimensão $(N \times N)$, sendo $N = nt$. Neste caso, a equação (3.1) simplifica-se para

$$\hat{\beta}_G = \left(\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X} \right)^{-1} \mathbf{X}^\top \hat{\mathbf{W}} \mathbf{z}, \quad (3.2)$$

em que $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$ e $\mathbf{z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_n^\top)^\top$, com dimensões $(N \times p)$ e $(N \times 1)$, respectivamente.

Nessa equação, $\hat{\beta}_G$ pode ser interpretado como a solução de mínimos quadrados da regressão normal linear de $\hat{\mathbf{W}}^{1/2} \mathbf{z}$ tendo como matriz de planejamento $\hat{\mathbf{W}}^{1/2} \mathbf{X}$. Nessa ótica, o resíduo ordinário, que é a diferença entre os valores observado e ajustado, fica então definido por

$$\mathbf{r}^* = \hat{\mathbf{W}}^{1/2} (\mathbf{z} - \hat{\eta}) = \hat{\mathbf{W}}^{1/2} (\hat{\mathbf{A}} \hat{\Delta})^{-1} (\mathbf{y} - \hat{\mu}),$$

em que $\hat{\mathbf{A}} = \text{diag}(\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_n)$ e $\hat{\mathbf{\Delta}} = \text{diag}(\hat{\mathbf{\Delta}}_1, \dots, \hat{\mathbf{\Delta}}_n)$, ambas com dimensão $(N \times N)$ e $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ e $\hat{\boldsymbol{\mu}} = (\hat{\boldsymbol{\mu}}_1^\top, \dots, \hat{\boldsymbol{\mu}}_n^\top)^\top$, ambos com dimensão $(N \times 1)$.

Se assumirmos que $\text{Cov}(\mathbf{z}) = (\hat{\mathbf{A}}\hat{\mathbf{\Delta}})^{-1} \text{Cov}(\mathbf{y}) (\hat{\mathbf{A}}\hat{\mathbf{\Delta}})^{-1} \cong \hat{\mathbf{W}}^{-1}$, temos que

$$\begin{aligned} \text{Cov}(\mathbf{r}^*) &= \text{Cov}(\hat{\mathbf{W}}^{1/2} \mathbf{z} - \hat{\mathbf{W}}^{1/2} \mathbf{X} \hat{\boldsymbol{\beta}}) = \\ &= \text{Cov}(\hat{\mathbf{W}}^{1/2} \mathbf{z} - \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}} \mathbf{z}) = \\ &= (\mathbf{I} - \mathbf{H}) \hat{\mathbf{W}}^{1/2} \text{Cov}(\mathbf{z}) \hat{\mathbf{W}}^{1/2} (\mathbf{I} - \mathbf{H}) \cong (\mathbf{I} - \mathbf{H}), \end{aligned}$$

sendo $\mathbf{I}$ a matriz identidade de dimensão $(N \times N)$ e $\mathbf{H}$ uma matriz bloco diagonal dada por $\mathbf{H} = \text{diag}(\mathbf{H}_1, \dots, \mathbf{H}_n)$, com

$$\mathbf{H}_i = \hat{\mathbf{W}}_i^{1/2} \mathbf{X}_i (\mathbf{X}_i^\top \hat{\mathbf{W}} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \hat{\mathbf{W}}_i^{1/2}, \quad i = 1, \dots, n. \quad (3.3)$$

A matriz $\mathbf{H}$ é simétrica ($\mathbf{H}^\top = \mathbf{H}$) e idempotente ($\mathbf{H}\mathbf{H} = \mathbf{H}$) e, por ser idempotente, temos que $r(\mathbf{H}) = \text{tr}(\mathbf{H}) = p$, sendo $r(\mathbf{H})$ o posto de $\mathbf{H}$.

Como os elementos de $\mathbf{r}^*$ possuem variâncias diferentes, o que dificulta compará-los entre si, definimos o resíduo padronizado associado à observação y_{ij} por

$$(\mathbf{r}_{SD})_{ij} = \frac{\mathbf{e}_{(ij)}^\top \hat{\mathbf{W}}_i^{1/2} (\hat{\mathbf{A}}_i \hat{\mathbf{\Delta}}_i)^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)}{\sqrt{1 - h_{ij}}}, \quad (3.4)$$

sendo $\mathbf{e}_{(ij)}$ um vetor de tamanho t com a posição referente à observação y_{ij} contendo o valor 1 e as demais posições contendo o valor zero e h_{ij} o j -ésimo elemento da diagonal principal de $\mathbf{H}_i$, $i = 1, \dots, n$ e $j = 1, \dots, t$.

O resíduo ordinário também pode ser escrito na forma $\mathbf{r}^* = (\mathbf{I} - \mathbf{H}) \hat{\mathbf{W}}^{1/2} \mathbf{z}$. Então, considerando que $\hat{\mathbf{W}}^{1/2} \mathbf{z}$ faz o papel do vetor resposta, podemos chamar $\mathbf{H}$ de matriz de projeção ortogonal (ou matriz chapéu) como na regressão normal linear em que $\hat{\mathbf{W}}$ é a identidade. Isso nos sugere a utilização dos elementos da diagonal principal de $\mathbf{H}$

para detectarmos a presença de pontos alavanca, assim como Paula (2001) fez para os MLGs e Tan, Qu e Kutner (1997) fizeram para os modelos de regressão logística com medidas repetidas.

Um ponto alavanca ocorre quando este possui um perfil diferente dos demais no que tange aos valores das variáveis explicativas. Logo, um valor alto de h_{ij} indica a influência de $\mathbf{x}_{ij}$ sobre o correspondente valor ajustado, $\hat{y}_{ij}$.

Supondo que todos os pontos exercem a mesma influência sobre os valores ajustados, podemos esperar que cada valor da diagonal principal de $\mathbf{H}$ esteja próximo de $\text{tr}(\mathbf{H})/N = p/N$. Dessa forma, os pontos para os quais $h_{ij} \geq 2p/N$, podem ser considerados de alavanca ou de alto *leverage*.

Analogamente, a i -ésima unidade experimental pode ser um ponto alavanca, se

$$\mathbf{h}_i = \frac{1}{t} \sum_{j=1}^t h_{ij} = \frac{\text{tr}(\mathbf{H}_i)}{t} \geq \frac{2p}{N}.$$

Notemos que os valores de h_{ij} dependem da matriz de pesos $\hat{\mathbf{W}}_i$, portanto observações com valores altos de h_{ij} nem sempre são pontos alavanca. Ou seja, é necessário examinar com cautela os pontos indicados como alavanca.

Esses resultados podem ser vistos com mais detalhes por meio do gráfico dos valores da diagonal principal da matriz de projeção, h_{ij} , $i = 1, \dots, n$ e $j = 1, \dots, t$, versus i que indica a ordem em que cada unidade experimental aparece no conjunto de dados (índice). Se o interesse for verificar se a unidade experimental é um ponto alavanca, então fazemos h_i versus i , $i = 1, \dots, n$.

Para detectar um ponto aberrante na análise gráfica, podemos utilizar o resíduo padronizado, $(r_{SD})_{ij}$, com $i = 1, \dots, n$ e $j = 1, \dots, t$, versus o índice i . Um ponto aberrante ocorre quando este possui perfil diferente dos demais no que tange aos valores da variável resposta e também apresenta valor baixo na matriz de projeção $\mathbf{H}$. Com isso, um mesmo ponto dificilmente é alavanca e aberrante.

Finalmente, um ponto influente ocorre quando este possui um perfil diferente dos demais no que tange aos valores da variável resposta, porém apresenta valor alto na matriz de projeção $\mathbf{H}$. Esse tipo de ponto tem grande peso na estimação dos parâmetros do modelo e para detectá-lo, a medida mais conhecida é a distância de Cook. Essa distância mede o afastamento entre a estimativa do vetor paramétrico utilizando todas as observações ($\hat{\beta}_G$) e sem a observação y_{ij} ($\hat{\beta}_{G(ij)}$), $i = 1, \dots, n$ e $j = 1, \dots, t$.

Como, em geral, não é possível obter uma forma fechada para $\hat{\beta}_{G(ij)}$, tem sido utilizada uma aproximação de um passo, que consiste em tomar a primeira iteração do processo iterativo pelo método de *scoring* de Fisher (dado em 2.10) quando o mesmo é iniciado em $\hat{\beta}_G$.

Essa aproximação, introduzida por Pregibon (1981), fica expressa para os MLGs com medidas repetidas na forma:

$$\hat{\beta}_{G(ij)}^{(1)} = \hat{\beta}_G - \frac{[\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X}]^{-1} [\mathbf{X}^\top \hat{\mathbf{W}}^{1/2} \mathbf{e}_{(ij)}] [\mathbf{e}_{(ij)}^\top \hat{\mathbf{W}}^{1/2} (\hat{\mathbf{A}} \hat{\Delta})^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}})]}{1 - h_{ij}}.$$

Logo, a distância de Cook, quando se exclui a observação y_{ij} , é definida por

$$(CD)_{ij} = \frac{1}{p} (\hat{\beta}_G - \hat{\beta}_{G(ij)})^\top \mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X} (\hat{\beta}_G - \hat{\beta}_{G(ij)}) = (r_{SD})_{ij}^2 \frac{h_{ij}}{p(1 - h_{ij})}. \quad (3.5)$$

Para visualizarmos graficamente, podemos fazer $(CD)_{ij}$, $i = 1, \dots, n$ e $j = 1, \dots, t$, versus o índice i e indicamos como um ponto influente aquele que possuir um valor alto na distância de Cook quando comparado com os demais pontos.

Como comentário geral, podemos alertar que todas as medidas apresentadas acima podem não ter acurácia se a estimativa da matriz de correlação de trabalho, $\mathbf{R}(\hat{\boldsymbol{\alpha}})$, não for próxima da verdadeira.

Assim como encontramos nossas medidas a partir do cálculo do resíduo ordinário, $\mathbf{r}^*$, fazendo uma analogia de $\hat{\beta}_G$ dada em (3.1) com a equação para estimar β no caso

da regressão normal linear, Preisser e Qaqish (1996) encontram suas medidas fazendo uma analogia de $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}_G$ com o preditor da regressão normal linear que é dado por

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{H}\mathbf{y}.$$

Logo, os autores definem $\mathbf{H}$, que é a matriz de projeção, por

$$\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}_G = \mathbf{X}(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}\mathbf{z} = \mathbf{H}\mathbf{z},$$

sendo $\mathbf{H}$, neste caso, uma matriz assimétrica e idempotente.

Seguindo essa proposta, o resíduo ordinário fica dado por $\mathbf{r}^* = \mathbf{z} - \hat{\boldsymbol{\eta}} = (\mathbf{I} - \mathbf{H})\mathbf{z}$, tendo como matriz de covariâncias uma expressão diferente de $(\mathbf{I} - \mathbf{H})$, fato esse que impossibilita a construção do gráfico de probabilidade meio-normal com envelope simulado, cujo procedimento consiste basicamente em gerar resíduos com média zero e matriz de covariâncias $(\mathbf{I} - \mathbf{H})$ (Paula, 2001, Seção 2.8.5).

3.2 Envelope Simulado

Para avaliarmos o ajuste de um modelo, utilizaremos o gráfico de probabilidade meio-normal com envelope simulado (Atkinson, 1985).

Em um gráfico de probabilidade meio-normal, dispomos o l -ésimo valor absoluto ordenado do resíduo padronizado, $l = 1, \dots, N$, versus o valor esperado da estatística de ordem, em valor absoluto, da distribuição normal padrão, $N(0,1)$, dado por

$$\Phi^{-1}\left(\frac{l + N - 1/8}{2N + 1/2}\right), \quad (3.6)$$

sendo $\Phi(\cdot)$ a função de distribuição acumulada da $N(0,1)$.

Esse gráfico pode ser utilizado mesmo que os resíduos não tenham distribuição de probabilidade normal (Neter et al, 1996). Quando isso ocorre, não esperamos que

os mesmos estejam próximos de uma reta que faz um ângulo de 45^0 com o eixo das abscissas. Portanto, utilizar os valores da equação (3.6) ou outros valores, como por exemplo, o índice l , $l = 1, \dots, N$, no eixo das abscissas nos fornece as mesmas conclusões sobre o gráfico.

Para identificarmos a adequação do ajuste do modelo, o gráfico de probabilidade meio-normal é combinado com um envelope simulado. A seguir, apresentamos um algoritmo para construir o gráfico de probabilidade meio-normal com envelope simulado.

Algoritmo: gráfico de probabilidade meio-normal com envelope simulado.

1. Para cada unidade experimental i , $i = 1, \dots, n$, simule um vetor de respostas de tamanho t , levando em consideração o vetor de médias e a matriz de covariâncias ajustados aos dados originais.
2. Ajuste, às respostas simuladas no passo 1, o mesmo modelo ajustado aos dados originais.
3. Calcule os resíduos padronizados conforme expresso na equação (3.4) e ordene os seus valores absolutos.
4. Repita os três primeiros passos mais 24 vezes. Aqui, definiremos $(r_{SD})_{lm}$ como sendo o l -ésimo valor absoluto ordenado do resíduo padronizado pertencente à m -ésima simulação, $l = 1, \dots, N$ e $m = 1, \dots, M$, com $M = 25$.
5. Determine o mínimo, a mediana e o máximo dos menores valores absolutos dos resíduos padronizados de todas as simulações, isto é, $(r_{SD})_{1m}$, $m = 1, \dots, 25$.
6. Repita o passo anterior para os segundos menores valores absolutos dos resíduos das simulações, $(r_{SD})_{2m}$; em seguida, os terceiros, $(r_{SD})_{3m}$, e assim sucessivamente, até os maiores valores absolutos dos resíduos das simulações, $(r_{SD})_{Nm}$, $m = 1, \dots, 25$. Ao final deste passo, teremos três vetores de tamanho N contendo

os mínimos, as medianas e os máximos dos resíduos padronizados, em valores absolutos.

7. Disponha no gráfico de probabilidade meio-normal, os valores mínimos, medianos e máximos versus os correspondentes valores esperados em (3.6). Una os pontos referentes aos valores mínimos. Faça o mesmo para os pontos referentes às medianas e aos valores máximos.

Grandes desvios dos pontos em torno da mediana dos valores simulados ou a ocorrência de pontos próximos dos limites ou fora da banda de simulação são indicadores de que o modelo não está apropriado. E se existirem, os pontos aberrantes devem aparecer no topo direito do gráfico separados dos demais pontos.

A sugestão de Atkinson (1985) é gerar $M=19$ vezes, pois, desse modo, há uma chance em vinte, ou 5%, de que o maior valor absoluto dos resíduos dos dados originais fique fora da banda de simulação quando o modelo ajustado estiver correto. O valor de $M=25$ simulações é sugestão de Tan, Qu e Kutner (1997), o qual adotamos em nosso trabalho.

Conforme mostramos nos sete passos do algoritmo anterior, a construção de um gráfico de probabilidade meio-normal com envelope simulado é muito simples desde que saibamos gerar variáveis correlacionadas.

Na próxima seção, descrevemos os algoritmos para gerar variáveis com distribuições normal, binomial, Poisson e gama correlacionadas.

3.3 Geração de variáveis correlacionadas

Nesta seção, apresentamos algoritmos para gerar variáveis com distribuições binomial, Poisson e gama correlacionadas. Para gerar distribuições normais multivariadas, uti-

lizamos o comando `rmvnorm` existente no software S-Plus, do qual daremos maiores detalhes no Apêndice B.

3.3.1 Variáveis Binárias

Para gerar variáveis com distribuição binomial correlacionadas, utilizamos o procedimento descrito por Park, Park e Shin (1996). Esse artigo fornece um algoritmo simples para gerar vetores aleatórios de dimensão arbitrária e não negativamente correlacionados, o que ocorre na maioria dos experimentos com medidas repetidas.

A idéia chave desse algoritmo consiste na propriedade de que qualquer variável aleatória com distribuição de Poisson pode ser expressa como uma convolução de outras variáveis aleatórias independentes, cada uma com distribuição de Poisson. As variáveis binárias são geradas com as correlações desejadas por compartilharem variáveis independentes com distribuição de Poisson.

Inicialmente, iremos considerar o caso de duas observações por unidade experimental ($t = 2$) para apresentar a idéia do algoritmo.

Seja $X(\lambda)$ uma variável aleatória com distribuição de Poisson de média $\lambda \geq 0$. Por convenção, $X(0) = 0$. As variáveis X 's com índices diferentes são consideradas mutuamente independentes.

Sejam U_1 e U_2 variáveis aleatórias definidas por

$$U_1 = X_1(\lambda_1 - \lambda_{12}) + X_3(\lambda_{12}) \quad \text{e} \quad U_2 = X_2(\lambda_2 - \lambda_{12}) + X_3(\lambda_{12}), \quad (3.7)$$

sendo λ_1, λ_2 e λ_{12} constantes não negativas. Como as variáveis X 's são independentes, então U_1 e U_2 têm distribuição de Poisson com médias λ_1 e λ_2 , respectivamente, e são não negativamente correlacionadas por causa do termo em comum $X_3(\lambda_{12})$. As variáveis binárias são definidas por

$$Z_1 = I_{\{0\}}(U_1) \quad \text{e} \quad Z_2 = I_{\{0\}}(U_2), \quad (3.8)$$

sendo $I_{\{ \cdot \}}$ a função indicadora tal que $I_A(y) = 1$, se $y \in A$ e $I_A(y) = 0$, se $y \notin A$.

As constantes λ_1 , λ_2 e λ_{12} são escolhidas de modo que $E(Z_1) = \mu_1$, $E(Z_2) = \mu_2$ e $\text{Corr}(Z_1, Z_2) = \alpha_{12} > 0$. Logo, $\lambda_j = -\ln(\mu_j)$, para $j = 1, 2$, pois

$$E(Z_j) = P(U_j = 0) = P(X_j = X_3 = 0) = e^{-\lambda_j}.$$

Por outro lado, temos que

$$\text{Var}(Z_j) = \mu_j(1 - \mu_j),$$

$$E(Z_1 Z_2) = P(U_1 = U_2 = 0) = P(X_1 = X_2 = X_3 = 0) = \mu_1 \mu_2 e^{\lambda_{12}}$$

e

$$\text{Cov}(Z_1, Z_2) = \mu_1 \mu_2 e^{\lambda_{12}} - \mu_1 \mu_2 = \mu_1 \mu_2 (e^{\lambda_{12}} - 1).$$

Com isso, o coeficiente de correlação entre Z_1 e Z_2 é

$$\alpha_{12} = \frac{\mu_1 \mu_2 (e^{\lambda_{12}} - 1)}{[\mu_1(1 - \mu_1)\mu_2(1 - \mu_2)]^{1/2}}.$$

Portanto, devemos escolher

$$\lambda_{12} = \ln \left\{ 1 + \alpha_{12} \left[\frac{(1 - \mu_1)(1 - \mu_2)}{\mu_1 \mu_2} \right]^{1/2} \right\}. \quad (3.9)$$

O caso em que $\lambda_{12} = 0$ corresponde à independência entre Z_1 e Z_2 .

Como $E(Z_1 Z_2) = P(Z_1 = Z_2 = 1) \leq P(Z_j = 1) = \mu_j$, então $\text{Cov}(Z_1, Z_2) \leq \mu_1(1 - \mu_2)$ e $\text{Cov}(Z_1, Z_2) \leq \mu_2(1 - \mu_1)$ e, portanto, α_{12} não varia livremente no intervalo $[-1, 1]$.

Logo, o mesmo deve satisfazer as seguintes desigualdades:

$$\alpha_{12} \leq \left(\frac{\mu_1(1 - \mu_2)}{\mu_2(1 - \mu_1)} \right)^{1/2} \quad \text{e} \quad \alpha_{12} \leq \left(\frac{\mu_2(1 - \mu_1)}{\mu_1(1 - \mu_2)} \right)^{1/2}.$$

Essas desigualdades levam à relação $\lambda_{12} \leq -\ln(\mu_j) = \lambda_j$, para $j = 1, 2$ e, por consequência, o procedimento gera variáveis binárias não negativamente correlacionadas.

A seguir, escrevemos um algoritmo para mostrar passo-a-passo como gerar variáveis binárias não negativamente correlacionadas com $t = 2$.

Algoritmo: gerar variáveis binárias com $t = 2$.

1. Dados μ_1 , μ_2 e α_{12} , calcule $\lambda_1 = -\ln(\mu_1)$, $\lambda_2 = -\ln(\mu_2)$ e λ_{12} conforme a expressão (3.9).
2. Gere, independentemente, as variáveis de Poisson X_1 de média $(\lambda_1 - \lambda_{12})$, X_2 de média $(\lambda_2 - \lambda_{12})$ e X_3 de média λ_{12} . Lembrar que, por convenção, $X(0) = 0$.
3. Calcule U_1 e U_2 conforme a equação (3.7) e defina

$$Z_j = \begin{cases} 1 & , \text{ se } U_j = 0 \\ 0 & , \text{ caso contrário} \end{cases} \quad , \quad \text{para } j = 1, 2.$$

Para o caso geral em que $t > 2$, consideremos t variáveis aleatórias $U_1, U_2, \dots, U_t$ com distribuição de Poisson que são somas parciais das variáveis aleatórias independentes com distribuição de Poisson $X_1(\gamma_1), X_2(\gamma_2), \dots, X_m(\gamma_m)$, para algum m inteiro não negativo, sendo $\gamma_1, \gamma_2, \dots, \gamma_m$ números reais. Alguns X 's podem aparecer simultaneamente em vários U 's.

O valor esperado e a estrutura de correlação das variáveis binárias $Z_1 = I_{\{0\}}(U_1)$, $Z_2 = I_{\{0\}}(U_2)$, ..., $Z_t = I_{\{0\}}(U_t)$ são obtidos controlando o padrão de somas parciais dos X 's em cada U e os valores de $\gamma_1, \gamma_2, \dots, \gamma_m$.

O algoritmo abaixo descreve como obter $m, \gamma_1, \gamma_2, \dots, \gamma_m$ e o padrão de somas parciais dados $E(Z_j) = \mu_j$ e $\text{Cov}(Z_j, Z_l) = \alpha_{jl}, \forall j \neq l, 1 \leq j, l \leq t$.

Algoritmo: gerar variáveis binárias com $t > 2$.

1. Faça $k = 0$ e para $j \leq l$ e $1 \leq j, l \leq t$, calcule

$$\lambda_{jl} = \ln \left\{ 1 + \alpha_{jl} \left[\frac{(1 - \mu_j)(1 - \mu_l)}{\mu_j \mu_l} \right]^{1/2} \right\}. \quad (3.10)$$

Observar que $\lambda_{jl} = \lambda_{lj}$.

2. Faça $k = k + 1$. Determine $T_k = \{\lambda_{jl} : \lambda_{jl} > 0, j \leq l \text{ e } 1 \leq j, l \leq t\}$.

Faça $\gamma_k = \lambda_{rs} = \min\{\lambda_{jl} : \lambda_{jl} \in T_k\}$, isto é, γ_k é o menor elemento do conjunto T_k . Se $\lambda_{rr} = 0$ ou $\lambda_{ss} = 0$, então pare porque o algoritmo falhou. Caso contrário, escolha um conjunto de índices S_k da seguinte maneira:

Seja $S_k^0 = \{r, s\}$ e para $j = 1, \dots, t$, faça

$$S_k^j = \begin{cases} S_k^{j-1} \cup \{j\} & , \text{ se } \lambda_{jl} > 0 \text{ para todo } l \in S_k^{j-1} \\ S_k^{j-1} & , \text{ caso contrário.} \end{cases}$$

Com isso, $S_k = S_k^t$.

3. Substitua λ_{jl} , $j, l \in S_k$, por $\lambda_{jl} - \gamma_k$. Se todo $\lambda_{jl} = 0$, então vá para o próximo passo; caso contrário, volte ao passo anterior.
4. Faça $m = k$ e para $j = 1, \dots, t$, obtenha

$$U_j = \sum_{k=1}^m X_k(\gamma_k) I_{S_k}(j) \quad \text{e} \quad Z_j = I_{\{0\}}(U_j).$$

Notemos que no passo 2, o menor γ_k é escolhido de modo que todas as variáveis Poisson tenham médias não negativas. É possível que λ_{rs} e S_k não sejam únicos no passo 2 e, nesse caso, podemos escolhê-los arbitrariamente.

Exemplo 1: Considere o caso que $t = 3$, $\hat{\mu}_1 = 0,9$, $\hat{\mu}_2 = 0,8$, $\hat{\mu}_3 = 0,7$ e as correlações entre as três variáveis são:

$$\mathbf{R}(\hat{\alpha}) = \begin{bmatrix} 1,0 & 0,1 & 0,5 \\ & 1,0 & 0,5 \\ & & 1,0 \end{bmatrix}.$$

A Tabela 3.1 nos mostra para cada k , $k = 1, \dots, 6$, uma matriz contendo os valores de λ_{jl} , $\forall j \leq l$ e $1 \leq j, l \leq 3$, o par de índices (r, s) e o conjunto S_k . Nas matrizes que contém os valores de λ_{jl} , os números em negrito e sublinhados são os γ_k 's e os números em negrito correspondem aos pares de índices (j, l) , tais que $j, l \in S_k$.

Os valores de λ_{jl} em $k = 1$ foram calculados por (3.10). Como todos os λ_{jl} 's são maiores do que zero, então $T_1 = \{0,105; 0,017; 0,104; 0,223; 0,152; 0,357\}$. Em T_1 , $\lambda_{12} = 0,017$ é o menor elemento, portanto, $\gamma_1 = 0,017$, $(r, s) = (1, 2)$ e $S_1 = \{1, 2, 3\}$. Atualize $\lambda_{jl} = \lambda_{jl} - \gamma_1$ para todo $j, l \in \{1, 2, 3\}$ e $j \leq l$. Os números resultantes são dados em $k = 2$. O conjunto T_2 de números positivos é dado por $T_2 = \{0,088; 0,087; 0,206; 0,135; 0,340\}$. O menor elemento de T_2 é $\gamma_2 = \lambda_{13} = 0,087$. Agora, $(r, s) = (1, 3)$ e, portanto, $S_2 = \{1, 3\}$. Atualize $\lambda_{jl} = \lambda_{jl} - \gamma_2$ para todo $j, l \in S_2$ e $j \leq l$. Esses resultados são dados em $k = 3$.

O algoritmo continua até que todos os λ_{jl} 's sejam iguais a zero e, no nosso exemplo, isso ocorreu em $k = 6$ definindo $m = 6$. Notemos que $\lambda_{rr} > 0$ e $\lambda_{ss} > 0$ em cada passo k . Por fim, utilizando γ_k e S_k , $k = 1, \dots, 6$, obtemos

$$\begin{aligned} U_1 &= X_1(0,017) + X_2(0,087) + X_3(0,002), \\ U_2 &= X_1(0,017) + X_4(0,135) + X_5(0,071), \\ U_3 &= X_1(0,017) + X_2(0,087) + X_4(0,135) + X_6(0,118). \end{aligned}$$

E, portanto, definimos as variáveis binárias por

$$Z_j = \begin{cases} 1 & , \text{ se } U_j = 0 \\ 0 & , \text{ caso contrário} \end{cases}, \text{ para } j = 1, 2, 3.$$

Tabela 3.1: Valores referentes ao exemplo para gerar variáveis binárias.

Valores de λ_{jl}				(r, s)	S_k
k	$j \leq l$ e $1 \leq j, l \leq 3$				
1	0,105	<u>0,017</u>	0,104	(1,2)	{1, 2, 3}
		0,223	0,152		
			0,357		
2	0,088	0,000	<u>0,087</u>	(1,3)	{1, 3}
		0,206	0,135		
			0,340		
3	<u>0,001</u>	0,000	0,000	(1,1)	{1}
		0,206	0,135		
			0,253		
4	0,000	0,000	0,000	(2,3)	{2, 3}
		0,206	<u>0,135</u>		
			0,253		
5	0,000	0,000	0,000	(2,2)	{2}
		<u>0,071</u>	0,000		
			0,118		
6	0,000	0,000	0,000	(3,3)	{3}
		0,000	0,000		
			<u>0,118</u>		

■

No nosso caso, o algoritmo para gerar variáveis binárias deve ser repetido para cada unidade experimental, isto é, n vezes. Fixada a unidade experimental i , faça $\hat{\mu}_j = \tilde{y}_{ij}$,

sendo $\tilde{y}_{ij}$ o valor esperado de y_{ij} dado pelo modelo ajustado, $j = 1, \dots, t$. Os valores de $\hat{\alpha}_{jl}$ são as estimativas da matriz de correlação de trabalho, $\mathbf{R}(\hat{\boldsymbol{\alpha}})$, obtida pelo modelo ajustado, a qual é a mesma para todas as n unidades experimentais.

3.3.2 Variáveis Poisson e Gama

Dados o vetor de médias $\boldsymbol{\mu} = (\mu_1, \dots, \mu_t)^\top$ e a matriz de correlação não negativa $\mathbf{R} = [\alpha_{jl}]$, $1 \leq j, l \leq t$, o algoritmo proposto por Park e Shin (1998) gera um vetor $\mathbf{U} = (U_1, \dots, U_t)^\top$, com média $\boldsymbol{\mu}$ e matriz de correlação $\mathbf{R}$, tal que

$$\mathbf{U} = \mathbf{TX},$$

sendo $\mathbf{T}$ uma matriz de dimensão $(t \times m)$ contendo zeros e uns, $\mathbf{X} = (X_1(\gamma_1), \dots, X_m(\gamma_m))^\top$ um vetor $(m \times 1)$, com $E(\mathbf{X}) = (\gamma_1, \dots, \gamma_m)^\top$, e m um inteiro não negativo também determinado pelo algoritmo.

A distribuição de cada componente do vetor $\mathbf{X}$ deve pertencer a uma classe $\mathcal{D}$ de distribuições fechadas para soma de acordo com a seguinte definição:

Definição 3.3.1 *Uma classe de funções de distribuição $\mathcal{D} = \{F_\gamma : \gamma > 0\}$ é fechada para soma se $F_{\gamma_1} \in \mathcal{D}$ e $F_{\gamma_2} \in \mathcal{D} \implies F_{\gamma_1} * F_{\gamma_2} \in \mathcal{D}$, sendo $F_{\gamma_1} * F_{\gamma_2}$ a convolução de F_{γ_1} e F_{γ_2} definida por $(F_{\gamma_1} * F_{\gamma_2})(x) = \int_{-\infty}^{\infty} F_{\gamma_1}(x - u) dF_{\gamma_2}(u)$.*

O fato da classe $\mathcal{D}$ ser fechada para soma é equivalente ao fato de que, para quaisquer distribuições $F_{\gamma_1}, F_{\gamma_2} \in \mathcal{D}$, a soma das variáveis aleatórias independentes com distribuição F_{γ_1} e F_{γ_2} tem função de distribuição $F_{\gamma_1 + \gamma_2}$. Um exemplo típico é a classe de todas as distribuições de Poisson. Se $X_1(\gamma_1)$ e $X_2(\gamma_2)$ são independentes com distribuição de Poisson de médias γ_1 e γ_2 , respectivamente, então $X_1(\gamma_1) + X_2(\gamma_2)$ tem distribuição de Poisson com média $\gamma_1 + \gamma_2$.

Nas classes das distribuições gama apenas algumas são fechadas para a soma. No nosso caso em que estamos considerando distribuições pertencentes à família exponencial, distribuições gama independentes são fechadas para soma. Por exemplo, se $X_1(\gamma_1) \sim G(\gamma_1, \phi)$ e $X_2(\gamma_2) \sim G(\gamma_2, \phi)$ são independentes com distribuição gama de médias γ_1 e γ_2 , respectivamente, e ϕ sendo o inverso do coeficiente de variação ao quadrado, então $X_1(\gamma_1) + X_2(\gamma_2) \sim G(\gamma_1 + \gamma_2, \phi)$ com média $\gamma_1 + \gamma_2$.

A seguir, descrevemos o algoritmo que determina a matriz $\mathbf{T}$, o vetor $\mathbf{X}$ e o inteiro não negativo m .

Algoritmo: gerar variáveis de Poisson ou gama.

1. Faça $k = 0$ e para $j \leq l$ e $1 \leq j, l \leq t$, calcule

$$\lambda_{jl} = \alpha_{jl} (\mu_j \mu_l)^{1/2}. \quad (3.11)$$

Observar que $\lambda_{jl} = \lambda_{lj}$.

2. Faça $k = k + 1$ e determine $T_k = \{\lambda_{jl} : \lambda_{jl} > 0, j \leq l \text{ e } 1 \leq j, l \leq t\}$. Faça $\gamma_k = \lambda_{rs} = \min\{\lambda_{jl} : \lambda_{jl} \in T_k\}$, isto é, γ_k é o menor elemento do conjunto T_k . Escolha um conjunto de índices S_k da seguinte maneira:

2.1 Sejam $B = \emptyset$ e $S_k^0 = \{r, s\}$.

2.2 Para $j = 1, \dots, t$, faça

$$S_k^j = \begin{cases} S_k^{j-1} \cup \{j\} & , \text{ se } \lambda_{jl} > 0 \text{ para todo } l \in S_k^{j-1} \\ S_k^{j-1} & , \text{ caso contrário.} \end{cases}$$

Com isso, $S_k = S_k^t$.

2.3 Seja $B = B \cup S_k$.

Se $l \notin B$, $l = 1, \dots, t$, e $\lambda_{rl} \lambda_{sl} > 0$, então $S_k^0 = \{r, s, l\}$ e volte ao passo anterior. Caso contrário, defina S_k conforme encontrado no passo 2.2.

Notemos que λ_{rs} e S_k podem não ser únicos no passo 2. Nesse caso, podemos escolher o primeiro λ_{rs} e S_k encontrados ou considerar todos os conjuntos encontrados.

3. Para todo $j, l \in S_k$, $j \leq l$, substitua λ_{jl} por $\lambda_{jl} - \gamma_k$. Se todo $\lambda_{jl} \leq 0$, então vá para o próximo passo; caso contrário, volte ao passo anterior.
4. Se algum valor de $\lambda_{jj} < 0$, então pare porque o algoritmo falhou. Caso contrário, faça $m = k$ e construa a matriz $\mathbf{T} = [T_{jk}]$ de dimensão $(t \times m)$ com

$$T_{jk} = \begin{cases} 1 & , \text{ se } j \in S_k \\ 0 & , \text{ se } j \notin S_k, \end{cases} \quad (3.12)$$

com $j = 1, \dots, t$ e $k = 1, \dots, m$.

A distribuição de $\mathbf{U} = \mathbf{TX}$ depende da distribuição dos elementos do vetor $\mathbf{X} = (X_1(\gamma_1), \dots, X_m(\gamma_m))^T$. Se $X_k(\gamma_k)$ tem distribuição de Poisson, então $\mathbf{U}$ terá distribuição de Poisson. Se a distribuição de $X_k(\gamma_k)$ é gama, então $\mathbf{U}$ terá distribuição gama.

No caso particular em que $\mu_1 = \mu_2 = \dots = \mu_t = \mu$ e $\alpha_{jl} = \alpha$, $\forall j \leq l$, é possível simplificar o algoritmo fazendo $\lambda_{jj} = \mu$ e $\lambda_{jl} = \alpha\mu$, $\forall j \leq l$. Com isso, temos que $m = t + 1$, $\gamma_1 = \alpha\mu$, $\gamma_k = \mu - \alpha\mu$, para $k = 2, 3, \dots, m$ e

$$U_j = X_1(\alpha\mu) + X_{j+1}(\mu - \alpha\mu), \quad j = 1, 2, \dots, t.$$

Exemplo 2: Considere o caso que $t = 4$, $\hat{\mu}_1 = 60$, $\hat{\mu}_2 = 50$, $\hat{\mu}_3 = 61$, $\hat{\mu}_4 = 31$ e as correlações entre as quatro variáveis são:

$$\mathbf{R}(\hat{\boldsymbol{\alpha}}) = \begin{bmatrix} 1,000 & 0,566 & 0,331 & 0,487 \\ & 1,000 & 0,543 & 0,381 \\ & & 1,000 & 0,230 \\ & & & 1,000 \end{bmatrix}.$$

A Tabela 3.2 nos mostra para cada k , $k = 1, \dots, 10$, uma matriz contendo os valores de λ_{jl} , $\forall j \leq l$ e $1 \leq j, l \leq 4$, o par de índices (r, s) e o conjunto S_k . Nas matrizes que contém os valores de λ_{jl} , os números em negrito e sublinhados são os γ_k 's e os números em negrito correspondem aos pares de índices (j, l) , tais que $j, l \in S_k$.

Os valores de λ_{jl} em $k = 1$ foram calculados por (3.11). Como todos $\lambda_{jl} > 0$, então $T_1 = \{60, 31, 20, 21, 50, 30, 15, 61, 10, 31\}$. Em T_1 , $\lambda_{34} = 10$ é o menor elemento, portanto, $\gamma_1 = 10$ e $(r, s) = (3, 4)$. Além disto, notemos que $S_1 = \{1, 2, 3, 4\}$. Atualize $\lambda_{jl} = \lambda_{jl} - \gamma_1$ para todo $j, l \in \{1, 2, 3, 4\}$, $j \leq l$. Os números resultantes são dados em $k = 2$. O conjunto T_2 de números positivos é dado por $T_2 = \{50, 21, 10, 11, 40, 20, 5, 51, 21\}$. O menor elemento de T_2 é $\gamma_2 = \lambda_{24} = 5$. Agora, $S_2 = \{1, 2, 4\}$. Atualizando os valores de λ_{jl} para todo $j, l \in S_2$, $j \leq l$, temos os resultados dados em $k = 3$.

O algoritmo continua até que todos os λ_{jl} 's sejam iguais a zero e, no nosso exemplo, isso ocorreu em $k = 10$ definindo $m = 10$. Notemos que não há nenhum $\lambda_{jj} < 0$, indicando que o algoritmo não falhou. Por fim, utilizando γ_k e S_k , $k = 1, \dots, 10$, obtemos

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

e

$$\mathbf{X} = (X_1(10), X_2(5), X_3(6), X_4(10), X_5(6), X_6(10), X_7(9), X_8(10), X_9(23), X_{10}(31))^T.$$

E, portanto, definimos $\mathbf{U} = \mathbf{TX}$ com o vetor de médias e as correlações desejadas no exemplo.

Como observação final, em $k = 4$, encontramos $\lambda_{13} = \lambda_{44} = 10$ como sendo o menor elemento de T_4 . Se a escolha de $\gamma_4 = \lambda_{13}$ apresentasse falha ao final do procedimento, isto é, algum valor $\lambda_{jj} < 0$, então poderíamos retornar em $k = 4$ e fazer $\gamma_4 = \lambda_{44}$.

Quando encontramos mais do que uma escolha para λ_{rs} e para S_k , podemos tentar todas as escolhas até encontrar uma que o algoritmo não falhe ao final do procedimento.

Tabela 3.2: Valores referentes ao exemplo para gerar variáveis Poisson ou gama.

Valores de λ_{jl}					(r,s)	S_k
k	$j \leq l$ e $1 \leq j,l \leq 4$					
1	60	31	20	21	(3,4)	{1, 2, 3, 4}
		50	30	15		
			61	<u>10</u>		
				31		
2	50	21	10	11	(2,4)	{1, 2, 4}
		40	20	<u>5</u>		
			51	0		
				21		
3	45	16	10	<u>6</u>	(1,4)	{1, 4}
		35	20	0		
			51	0		
				16		
4	39	16	<u>10</u>	0	(1,3)	{1, 2, 3}
		35	20	0		
			51	0		
				10		
5	29	<u>6</u>	0	0	(1,2)	{1, 2}
		25	10	0		
			41	0		
				10		

Tabela 3.2: Continuação.

Valores de λ_{jl}					(r, s)	S_k
k	$j \leq l$ e $1 \leq j, l \leq 4$					
6	23	0	0	0	(2,3)	{2, 3}
		19	10	0		
			41	0		
				10		
7	23	0	0	0	(2,2)	{2}
		9	0	0		
			31	0		
				10		
8	23	0	0	0	(4,4)	{4}
		0	0	0		
			31	0		
				10		
9	23	0	0	0	(1,1)	{1}
		0	0	0		
			31	0		
				0		
10	0	0	0	0	(3,3)	{3}
		0	0	0		
			31	0		
				0		

Capítulo 4

Aplicações

Neste capítulo, analisamos três conjuntos de dados utilizando o método de equações de estimação generalizadas detalhado no Capítulo 2 e as técnicas de diagnóstico apresentadas no Capítulo 3.

4.1 Aplicação 1

Nesta aplicação, cujos dados foram obtidos do CEA/IME-USP (Lima e Sañudo, 1997), cada um dos 40 voluntários realizou, na etapa Prática, uma mesma tarefa 80 vezes, divididas em 8 blocos de 10 tentativas. Em cada tentativa, foi observada a diferença de tempo entre o instante em que o voluntário recebeu um estímulo e o instante de sua reação motora (em ms).

O objetivo deste estudo era avaliar o processo de aprendizagem da tarefa desenvolvida pelos voluntários, analisando a variável Erro Absoluto. Essa variável foi definida, para cada bloco, como sendo a média dos valores absolutos das diferenças de tempo observados nas tentativas e foi ajustada via um modelo considerando distribuição normal com função de ligação identidade (canônica). Os parâmetros envolvidos no modelo linear foram: β_0 , coeficiente linear ou intercepto e β_1 , coeficiente angular

avaliando o efeito de bloco. Esse modelo foi ajustado utilizando a técnica de EEG e como matriz de correlação de trabalho as estruturas padrão uniforme, AR-1 e não estruturada (ver Seção 2.3).

O ajuste com a estrutura AR-1 foi escolhido como o mais adequado para explicar a correlação entre as respostas da mesma unidade experimental. Isso se deve ao fato de que esse ajuste apresentou os menores valores dos resíduos padronizados quando comparados aos ajustes com as demais estruturas de correlação.

Aplicando aos dados as técnicas de diagnóstico apresentadas no Capítulo 3, calculamos a distância de Cook e o resíduo padronizado cujos valores estão apresentados na Figura 4.1 e na Figura 4.2, respectivamente. Na primeira figura, temos que a distância de Cook das observações do primeiro bloco referentes às unidades experimentais 1, 33 e 39 são maiores do que as demais observações, indicando serem possíveis pontos influentes. Na Figura 4.2, temos que o resíduo padronizado das observações do primeiro bloco novamente referentes às unidades experimentais 1, 33 e 39 são maiores do que as demais observações, indicando serem também possíveis pontos aberrantes. Notemos que nosso exemplo não contém covariáveis e por esse motivo, não utilizamos a matriz de projeção para detectar pontos alavanca.

Na Tabela A.1, que apresenta o conjunto de dados utilizado nesta aplicação, podemos observar que os valores do erro absoluto para as observações do primeiro bloco referentes às unidades experimentais 1, 33 e 39 foram, respectivamente, 150,5, 163,5 e 130,2. Essas respostas realmente não apresentam comportamento similar às demais observações do primeiro bloco cuja média é 45,0 (em ms).

Entretanto, no gráfico de probabilidade meio-normal com envelope simulado apresentado na Figura 4.3, observamos pontos fora da banda de confiança, indicando que o modelo ajustado com distribuição normal e ligação identidade é inadequado à variável resposta Erro Absoluto.

Figura 4.1: Distância de Cook do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.

Figura 4.2: Resíduo padronizado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.

Figura 4.3: Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável Erro Absoluto.

Ajustamos novamente o mesmo modelo para a variável resposta logaritmo do Erro Absoluto. Nesse novo ajuste, a estrutura AR-1 foi escolhida como a mais adequada para explicar a correlação entre as respostas da mesma unidade experimental, por apresentar os menores valores dos resíduos padronizados quando comparados aos ajustes com as demais estruturas de correlação.

A Tabela 4.1 apresenta os resultados do ajuste do modelo para o logaritmo do Erro Absoluto com matriz de correlação de trabalho AR-1. Por meio da estatística de Wald proposta em (2.13), detectamos que há efeito de bloco (nível descritivo = $p < 0,001$), ao nível de 5% de significância.

Na Figura 4.4, temos que as observações do primeiro bloco referentes às unidades

Tabela 4.1: Estimativas e erros padrão dos parâmetros do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.

		Erro Padrão		
	Parâmetro	Estimativa	Robusto	“Naive”
β_0	(Intercepto)	3,850	0,067	0,068
β_1	(Bloco)	-0,051	0,010	0,013
ϕ^{-1}	(Dispersão)	0,173		
α	(Correlação)	0,531		

experimentais 1, 33 e 39 continuam se destacando das demais com relação aos valores da distância de Cook, indicando serem possíveis pontos influentes. Na Figura 4.5, não detectamos pontos com resíduos padronizado muito distintos dos demais.

O gráfico de probabilidade meio-normal com envelope simulado apresentado na Figura 4.6 não nos mostra pontos fora da banda de confiança e, portanto, concluímos que o ajuste do modelo de regressão normal é adequado à variável logaritmo do Erro Absoluto.

Para avaliarmos a influência que as observações das unidades experimentais 1, 33 e 39 têm sobre as estimativas dos parâmetros de regressão, dispersão e correlação, realizamos uma análise confirmatória. Essa análise avalia descritivamente o quanto variam as estimativas dos parâmetros na ausência de ponto(s) com comportamento(s) distinto(s) dos demais por meio da seguinte medida de variação:

$$\frac{\hat{\theta}_{-ponto(s)} - \hat{\theta}}{\hat{\theta}} \quad (4.1)$$

sendo $\hat{\theta}_{-ponto(s)}$ a estimativa de θ sem o(s) ponto(s) com característica(s) distinta(s) dos demais e $\hat{\theta}$ a estimativa de θ com todos os pontos no modelo.

Na Tabela 4.2, apresentamos as variações nas estimativas dos coeficientes de regres-

Figura 4.4: Distância de Cook do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.

Figura 4.5: Resíduo padronizado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.

Figura 4.6: Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão normal linear ajustado com estrutura de correlação AR-1 para a variável logaritmo do Erro Absoluto.

Tabela 4.2: Variação nas estimativas dos parâmetros do modelo ajustado com estrutura de correlação AR-1 e sem as unidades experimentais 1, 33 e 39.

		Erro Padrão		Medida de	
Parâmetro	Estimativa	Robusto	“Naive”	Variação	
β_0	(Intercepto)	3,775	0,055	0,069	-1,9%
β_1	(Bloco)	-0,041	0,009	0,013	-19,6%
ϕ^{-1}	(Dispersão)	0,162			-6,4%
α	(Correlação)	0,542			2,1%

são, dispersão e correlação após a eliminação das unidades experimentais 1, 33 e 39. Como estamos trabalhando com $n = 40$, isto implica que cada unidade experimental deve influenciar em 2,5% nas estimativas dos parâmetros. No nosso caso, com a exclusão dessas três unidades experimentais, podemos considerar uma influência razoável uma variação em até 7,5%. De maneira geral, verificamos que a presença das três unidades experimentais superestima apenas o coeficiente angular em quase 20%. Apesar disso, as conclusões dos ajustes com e sem as unidades experimentais 1, 33 e 39 não se alteram.

Notemos que não é conveniente realizarmos uma análise confirmatória com a exclusão apenas dos três pontos considerados influentes, já que as equações de estimação consideradas em nosso trabalho produzem estimadores viciados dos parâmetros de regressão quando dados ausentes em uma ou mais unidades experimentais não são completamente aleatórios (ver mais detalhes na Seção 2.3).

Além dos modelos ajustados anteriormente que utilizam a distribuição normal, ajustamos diversos modelos com a distribuição gama alternando as possíveis funções de ligação e matrizes de correlação de trabalho. Porém, nenhum deles se mostrou adequado aos dados pelo gráfico de probabilidade meio-normal com envelope simulado.

4.2 Aplicação 2

Este estudo visa comparar a eficiência de uma droga no controle de distúrbios emocionais em 172 pacientes do sexo feminino. As pacientes foram avaliadas no momento inicial sem medicamento (visita 1) e nas visitas 2 e 3, com medicamento.

A variável resposta (Depressão) analisada neste estudo foi definida a partir da ausência ou presença de depressão no momento da visita. Isto é, a variável Depressão é igual a 1 se a paciente teve ausência de sintomas de depressão ou 0, caso contrário. A variável Visita foi considerada como fator com três níveis: Visita 1, Visita 2 e Visita

3. Pelo fato de termos uma variável resposta dicotômica, modelamos os dados via uma regressão logística levando em consideração a estrutura de correlação existente entre as observações de uma mesma paciente.

Os parâmetros envolvidos nesse modelo foram β_j (logaritmo da chance da paciente ter ausência de depressão na visita j), $j = 1, 2, 3$. Esse modelo foi ajustado utilizando as estruturas de correlação padrão uniforme e não-estruturada. A estrutura AR-1 não foi aplicada aos dados, pois no estudo são consideradas apenas três visitas.

O ajuste com a matriz de correlação de trabalho não-estruturada foi escolhido para explicar a correlação entre as observações da mesma unidade experimental por apresentar os menores resíduos padronizados. Além disso, o algoritmo descrito na Seção 3.3.1 falhou para o ajuste com estrutura de correlação padrão uniforme, impossibilitando a construção do gráfico de probabilidade meio-normal com envelope simulado para esse estrutura.

A Tabela 4.3 apresenta os resultados do ajuste do modelo para a matriz de correlação de trabalho não-estruturada. Por meio da estatística de Wald proposta em (2.13), ao nível de 5% de significância, rejeitamos as hipóteses de que cada parâmetro de regressão: β_1 , β_2 e β_3 , é igual a zero (nível descritivo = $p < 0,001$, $p = 0,006$ e $p < 0,001$, respectivamente).

Tabela 4.3: Estimativas e erros padrão dos parâmetros do modelo de regressão logística ajustado com a estrutura de correlação não-estruturada.

		Erro Padrão		
Parâmetro		Estimativa	Robusto	“Naive”
β_1	(Visita 1)	-1,401	0,191	0,191
β_2	(Visita 2)	0,425	0,156	0,156
β_3	(Visita 3)	0,624	0,160	0,160

A matriz de correlação de trabalho não-estruturada ajustada aos dados foi:

$$\mathbf{R}(\hat{\alpha}) = \begin{bmatrix} 1 & 0,342 & 0,271 \\ & 1 & 0,581 \\ & & 1 \end{bmatrix}.$$

Para detectar possíveis pontos influentes e/ou aberrantes, calculamos a distância de Cook e o resíduo padronizado cujos valores estão apresentados na Figura 4.7 e na Figura 4.8, respectivamente. Em ambas figuras, não detectamos pontos com valores muito distintos dos demais.

A Figura 4.9 apresenta o gráfico de probabilidade meio-normal com envelope simulado. Nesse gráfico, não observamos pontos fora da banda de confiança (simulada pelo algoritmo apresentado na Seção 3.3.1). Logo, podemos concluir que o ajuste do modelo é adequado à variável resposta em questão.

Figura 4.7: Distância de Cook do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.

Figura 4.8: Resíduo padronizado do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.

Figura 4.9: Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão logística ajustado com estrutura de correlação não-estruturada.

4.3 Aplicação 3

Neste estudo, cujos dados foram extraídos de Montgomery (2001, p.215), o objetivo era comparar três drogas utilizadas em quimioterapia. Para isso, cada droga foi aplicada em 10 ratos com leucemia. Após o tratamento, foram observados o número de glóbulos brancos (NGB), o número de glóbulos vermelhos (NGV) e o número de células cancerígenas (NCC), em quatro momentos diferentes.

A variável resposta foi definida pelo número de células cancerígenas e ajustada utilizando a distribuição de Poisson considerando a ligação log (canônica) e as covariáveis NGB e NGV. Os parâmetros envolvidos nesse modelo foram:

β_k , coeficiente linear que avalia o efeito da droga k , $k = 1, \dots, 3$;

β_4 , variação do logaritmo do número de células cancerígenas no acréscimo de um glóbulo branco, quando fixadas as demais variáveis da matriz de planejamento e

β_5 , variação do logaritmo do número de células cancerígenas no acréscimo de um glóbulo vermelho, quando fixadas as demais variáveis da matriz de planejamento.

Esse modelo foi ajustado utilizando como matriz de correlação de trabalho as estruturas padrão uniforme, AR-1 e não-estruturada (com a qual não convergiu os parâmetros de regressão após 50 passos do processo iterativo). O ajuste com a matriz de correlação de trabalho AR-1 foi escolhido para explicar a correlação entre as observações da mesma unidade experimental por apresentar os menores resíduos padronizados. Além disso, o algoritmo descrito na Seção 3.3.1 falhou para o ajuste com estrutura de correlação padrão uniforme, impossibilitando a construção do gráfico de probabilidade meio-normal com envelope simulado para esse estrutura.

A Tabela 4.4 apresenta os resultados do ajuste do modelo para a matriz de correlação de trabalho AR-1. Por meio da estatística de Wald proposta em (2.13), ao nível

Tabela 4.4: Estimativas e erros padrão dos parâmetros do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

		Erro Padrão		
Parâmetro		Estimativa	Robusto	“Naive”
β_1	(Droga 1)	3,0120	0,0778	0,0315
β_2	(Droga 2)	3,2315	0,0976	0,0891
β_3	(Droga 3)	3,1363	0,1540	0,1075
β_4	(NGB)	-0,0305	0,0051	0,0045
β_5	(NGV)	0,0221	0,0065	0,0073
α	(Correlação)	0,9227		

de 5% de significância, rejeitamos a hipótese de que cada parâmetro de regressão β_k , $k = 1, \dots, 5$, é igual a zero (nível descritivo = $p < 0,001$, para todos os parâmetros).

Aplicando aos dados as técnicas de diagnóstico apresentadas no Capítulo 3, calculamos as medidas h_{ij} e $\mathbf{h}_i$ para avaliar, respectivamente, se a observação (i, j) ou a unidade experimental i é um ponto alavanca, $i = 1, \dots, 30$ e $j = 1, \dots, 4$. Na Figura 4.10, que apresenta os valores de h_{ij} , temos 6 pontos se destacando como alavanca dos quais três são referentes à Droga 1, $(1, 4)$, $(3, 1)$ e $(6, 1)$, dois à Droga 2, $(16, 1)$ e $(16, 4)$, e apenas um à Droga 3, $(23, 4)$. Porém, observando a Figura 4.11 que apresenta os valores de $\mathbf{h}_i$, nenhuma unidade experimental se destaca como ponto alavanca. Apenas a unidade experimental 16 se aproxima de $2p/N = 0,083$ por apresentar duas observações, $(16, 1)$ e $(16, 4)$, como pontos alavanca no gráfico anterior.

Para detectar possíveis pontos influentes e/ou aberrantes, calculamos a distância de Cook e o resíduo padronizado cujos valores estão apresentados na Figura 4.12 e na Figura 4.13, respectivamente. Em ambas as figuras, não detectamos pontos com valores muito distintos dos demais. A Figura 4.14 apresenta o gráfico de probabilidade meio-

normal com envelope simulado. Nesse gráfico, não observamos pontos fora da banda de confiança (simulada pelo algoritmo apresentado na Seção 3.3.2). Logo, podemos concluir que o ajuste do modelo é adequado à variável resposta em questão.

Apesar dos pontos destacados como alavanca na Figura 4.10, não iremos realizar uma análise confirmatória pois nenhum deles se destacaram como influentes e nenhuma das unidades experimentais envolvidas se destacaram como ponto alavanca.

Figura 4.10: Medida h_{ij} do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

Figura 4.11: Medida h_i do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

Figura 4.12: Distância de Cook do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

Figura 4.13: Resíduo padronizado do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

Figura 4.14: Gráfico de probabilidade meio-normal com envelope simulado do modelo de regressão de Poisson ajustado com estrutura de correlação AR-1.

Capítulo 5

Conclusões e Estudos Futuros

Neste trabalho, apresentamos as equações de estimação generalizadas (EEGs) propostas por Liang e Zeger (1986) e mostramos como construí-las por meio da teoria de funções de estimação. Além disso, nos dedicamos às técnicas de diagnóstico, generalizando para MLGs com estrutura de dependência entre observações da mesma unidade experimental, medidas usuais em MLGs tais como: diagonal da matriz de projeção, distância de Cook e resíduo padronizado.

Outro ponto importante deste trabalho é que apresentamos alguns algoritmos para simular variáveis correlacionadas, os quais são úteis para construirmos o gráfico de probabilidade meio-normal com envelope simulado quando temos dados com medidas repetidas. A Aplicação 1 do Capítulo 4 mostra claramente a importância desse gráfico como técnica de diagnóstico.

As rotinas desenvolvidas no S-Plus 4.5 para as aplicações do Capítulo 4 estão escritas no caso geral e podem ser utilizadas para outros bancos de dados.

Para estudos futuros, ainda há bastante a ser explorado, principalmente quando se trata de técnicas de diagnóstico para análise de dados com medidas repetidas. Entre eles citamos:

- Estudar as propriedades das medidas de diagnóstico para cada uma das distribuições pertencentes à família exponencial.
- Estender, para dados com medidas repetidas, os métodos de seleção de modelos existentes para os MLGs, como por exemplo, o método *stepwise*.
- Apresentar algoritmos para simular variáveis correlacionadas a partir de outras distribuições pertencentes à família exponencial que não foram citadas neste trabalho. Uma sugestão seria trabalhar com cópulas (Joe, 1997) que é um método para construir distribuições multivariadas a partir das distribuições marginais das observações e da estrutura de dependência entre elas.

Apêndice A

Conjuntos de Dados

Neste apêndice, apresentamos os conjuntos de dados referentes às aplicações do Capítulo 4.

Tabela A.1: Conjunto de dados referentes à Aplicação 1.

Voluntário	Bloco							
	1	2	3	4	5	6	7	8
1	150,5	51,2	24,9	23,1	34,4	24,3	19,7	33,4
2	27,7	36,5	19,9	16,4	27,2	34,4	14,6	18,8
3	57,7	62,6	58,2	38,1	26,3	23,2	31,2	28,0
4	39,3	60,4	53,6	80,3	35,7	51,4	38,7	28,0
5	88,2	22,7	17,9	34,2	14,0	30,7	42,7	36,8
6	89,9	77,1	47,9	115,9	94,1	43,0	64,4	63,4
7	25,0	49,2	31,4	29,1	44,9	38,7	37,4	37,4
8	89,9	85,7	42,9	63,8	55,9	36,7	52,8	68,0
9	31,2	58,7	38,2	43,7	29,3	37,4	38,9	36,3
10	21,7	48,3	28,4	40,8	31,8	34,1	39,5	41,1

Voluntário	Bloco							
	1	2	3	4	5	6	7	8
11	37,4	34,1	47,7	23,3	62,4	27,8	31,3	17,6
12	26,1	35,4	57,4	37,9	35,5	30,2	32,9	53,2
13	55,1	56,7	48,9	35,1	24,4	43,6	36,5	21,3
14	37,7	31,4	42,9	49,8	41,6	40,9	24,8	33,9
15	50,0	22,5	20,3	32,1	23,1	20,7	18,7	33,9
16	37,2	53,5	35,3	46,3	38,9	31,3	32,9	32,0
17	27,1	20,2	31,0	35,2	34,3	11,8	18,6	44,8
18	27,4	22,9	22,2	16,5	13,8	13,7	18,9	18,5
19	40,9	40,2	43,1	32,3	38,9	46,8	43,3	29,3
20	43,7	41,6	49,9	21,3	25,9	53,8	33,3	25,6
21	82,5	49,5	49,4	29,1	33,1	33,4	36,9	42,4
22	58,1	49,8	35,2	70,7	37,9	49,0	44,4	26,7
23	73,0	81,5	81,6	68,9	82,4	81,4	94,2	41,7
24	51,6	44,6	44,7	39,8	32,1	40,3	31,5	37,2
25	47,5	40,6	43,0	37,3	38,0	21,8	29,2	40,0
26	39,9	23,8	23,0	22,1	35,1	21,1	18,4	18,8
27	35,2	19,7	50,8	31,8	19,5	34,5	22,3	20,5
28	35,5	34,1	35,2	29,1	66,2	52,7	55,5	52,6
29	37,0	37,8	22,5	25,4	25,3	19,2	17,9	14,4
30	38,2	50,0	40,7	46,5	47,1	25,5	18,7	29,4
31	37,4	43,9	61,4	41,0	55,0	29,4	53,8	41,3
32	29,1	26,3	35,4	35,4	18,0	29,9	24,9	26,6
33	163,5	89,0	61,7	62,5	34,2	41,0	42,9	34,7
34	28,3	68,3	68,6	54,5	49,1	68,8	33,1	46,8
35	41,9	55,5	50,6	40,4	73,6	40,6	70,7	51,0
36	45,7	54,0	23,9	41,3	41,2	29,6	41,1	30,5
37	40,1	25,6	18,2	20,7	20,4	24,2	26,6	18,6
38	44,5	45,2	49,1	39,0	39,0	40,9	44,9	25,5
39	130,2	66,3	44,2	65,3	45,2	32,7	29,7	51,3
40	47,6	42,6	29,4	23,9	20,2	30,0	35,7	26,0

Tabela A.2: Conjunto de dados referentes à Aplicação 2.

Visita				Visita			
Paciente	1	2	3	Paciente	1	2	3
1	1	1	1	26	0	1	1
2	1	1	1	27	0	1	1
3	0	1	1	28	0	1	1
4	0	1	1	29	0	1	1
5	0	1	1	30	0	1	1
6	0	1	1	31	0	0	0
7	1	1	1	32	0	0	0
8	0	1	1	33	0	1	1
9	1	1	1	34	1	1	1
10	0	0	0	35	0	0	0
11	1	1	1	36	1	1	1
12	0	1	1	37	0	0	0
13	0	0	0	38	0	1	1
14	1	1	0	39	0	1	1
15	0	1	1	40	1	1	1
16	0	0	0	41	1	1	1
17	0	0	0	42	1	1	1
18	0	0	0	43	0	1	1
19	0	0	0	44	1	1	1
20	0	0	1	45	0	1	1
21	1	1	1	46	1	1	1
22	1	1	1	47	0	0	1
23	1	1	1	48	0	0	0
24	1	1	1	49	0	0	1
25	0	0	1	50	0	1	1

Visita				Visita			
Paciente	1	2	3	Paciente	1	2	3
51	0	1	1	81	0	1	1
52	0	1	1	82	0	1	0
53	0	1	1	83	0	0	0
54	0	0	0	84	0	0	0
55	0	0	0	85	0	1	1
56	0	1	0	86	0	0	0
57	1	1	1	87	0	0	0
58	0	1	1	88	0	0	0
59	0	0	1	89	1	1	1
60	0	1	0	90	0	1	1
61	0	0	0	91	0	1	1
62	0	0	0	92	0	0	0
63	0	0	0	93	0	1	1
64	1	1	1	94	0	0	1
65	0	0	0	95	0	0	1
66	0	0	0	96	0	0	0
67	1	1	1	97	0	1	1
68	0	0	0	98	0	1	0
69	0	0	0	99	0	1	1
70	0	0	1	100	0	0	0
71	1	1	1	101	0	1	0
72	0	1	1	102	0	0	0
73	0	0	1	103	0	0	0
74	0	1	1	104	0	0	0
75	0	0	0	105	1	1	1
76	0	1	1	106	0	1	0
77	0	0	1	107	0	1	1
78	0	1	1	108	0	1	0
79	1	1	1	109	0	1	1
80	0	1	1	110	0	1	1

Visita				Visita			
Paciente	1	2	3	Paciente	1	2	3
111	0	1	1	142	0	0	0
112	0	1	1	143	0	0	0
113	1	1	1	144	1	0	0
114	0	1	1	145	0	0	0
115	1	0	1	146	0	0	0
116	0	0	1	147	0	1	1
117	0	1	1	148	0	0	1
118	0	0	1	149	0	1	1
119	0	1	1	150	0	1	1
120	0	0	0	151	1	1	0
121	1	1	1	152	0	1	1
122	0	0	0	153	1	1	1
123	0	1	1	154	0	1	1
124	0	1	1	155	0	1	1
125	0	0	1	156	0	1	1
126	0	0	0	157	0	1	0
127	0	1	1	158	0	0	0
128	0	0	1	159	0	0	1
129	0	0	0	160	0	1	1
130	0	1	1	161	0	0	1
131	0	0	0	162	0	0	0
132	0	0	1	163	0	1	1
133	0	1	1	164	0	0	1
134	0	1	1	165	0	1	0
135	0	1	0	166	0	0	0
136	0	1	1	167	0	1	1
137	0	1	1	168	0	0	1
138	0	1	0	169	0	0	0
139	1	1	1	170	1	1	1
140	0	1	1	171	1	1	1
141	1	1	1	172	0	0	0

Tabela A.3: Conjunto de dados referentes à Aplicação 3.

Rato	Droga	NGB				NGV				NCC			
		1	2	3	4	1	2	3	4	1	2	3	4
1	1	15	18	19	24	2	3	2	5	14	14	12	11
2	1	8	11	14	14	2	4	4	5	17	18	18	16
3	1	4	5	6	4	7	5	4	4	23	20	19	19
4	1	16	14	14	12	3	4	4	2	13	12	12	11
5	1	6	4	4	4	7	6	5	2	24	20	20	19
6	1	22	20	21	18	4	3	3	2	12	12	10	9
7	1	18	17	17	16	5	3	5	2	16	16	14	12
8	1	4	7	4	4	8	7	4	4	28	26	26	26
9	1	14	12	12	10	3	4	4	5	14	13	12	10
10	1	10	10	10	10	3	4	5	2	16	15	15	14
11	2	14	14	16	17	6	6	7	6	16	15	15	14
12	2	7	7	6	5	4	4	4	2	36	32	30	29
13	2	9	8	9	11	8	8	7	4	18	16	17	15
14	2	21	20	20	20	3	3	4	3	14	13	13	12
15	2	18	17	17	17	4	4	2	2	19	19	18	17
16	2	3	6	6	2	10	10	8	7	38	38	37	37
17	2	8	9	9	8	3	3	2	2	18	18	17	16
18	2	29	30	29	29	6	6	5	4	8	8	7	6
19	2	8	8	8	7	9	9	8	8	19	19	18	17
20	2	5	4	4	3	8	7	7	7	36	35	30	29
21	3	16	17	17	18	2	3	4	2	15	16	17	15
22	3	13	11	12	12	6	4	5	4	17	16	16	18
23	3	7	8	6	5	3	2	2	3	28	25	27	31
24	3	9	8	9	9	4	5	3	3	29	30	32	30
25	3	18	19	21	20	3	2	5	4	11	12	12	13
26	3	23	25	24	24	5	5	4	4	8	10	9	8
27	3	27	28	27	30	7	6	6	4	7	8	8	7
28	3	30	32	33	35	6	7	8	7	4	5	5	4
29	3	17	19	20	21	4	3	3	2	14	13	13	12
30	3	12	12	13	11	3	5	4	5	17	15	16	16

Apêndice B

Rotinas Desenvolvidas no S-Plus

Neste apêndice, apresentamos os programas desenvolvidos no software S-Plus 4.5 para as aplicações do Capítulo 4.

B.1 Aplicação 1

A Figura B.1 nos mostra o formato do *data.frame* criado no software S-Plus 4.5 contendo os dados da Tabela A.1.

Para obter os resultados apresentados na Seção 4.1, os programas devem ser executados na mesma ordem em que são exibidos a seguir.

Programa para ajuste do modelo normal linear com estrutura de correlação AR-1

Este programa ajusta um modelo normal linear para a resposta Erro Absoluto com estrutura de correlação AR-1 utilizando a técnica de EEGs. O mesmo programa foi utilizado para o ajuste da resposta logaritmo do Erro Absoluto, alterando `erro` para `log(erro)` na função `gee`.

	1	2	3
	sujeito	bloco	erro
1	1.00	1.00	150.50
2	1.00	2.00	51.20
3	1.00	3.00	24.90
4	1.00	4.00	23.10
5	1.00	5.00	34.40
6	1.00	6.00	24.30
7	1.00	7.00	19.70
8	1.00	8.00	33.40
9	2.00	1.00	27.70
10	2.00	2.00	36.50
11	2.00	3.00	19.90

Figura B.1: Formato do *data.frame* Pratica criado no software S-Plus 4.5.

```
original <- Pratica #Pratica contém os dados da Tabela A.1
cluster <- as.vector(original[,1])
attach(original)
library(gee)
orig.gee.all <- gee(erro ~ bloco , sujeito, family=gaussian,
data = original, na.action = na.omit, corstr="AR-M", M=1)
summary(orig.gee.all)
n <- 40 #Número de sujeitos no conjunto de dados
t<-as.vector(rep(8,n)) #t=8 para todos os sujeitos
```

Programa para calcular a distância de Cook e o resíduo padronizado para modelos normais lineares

```
library(Matrix)
X <- model.matrix(orig.gee.all)
y <- orig.gee.all$y
beta <- coef(orig.gee.all)
R <- orig.gee.all$work
```

```

mi <- fitted(orig.gee.all)
N <- nrow(X)
p <- ncol(X)

#Calculo do Residuo de Pearson
r<-(y-mi)

#Calculo de phi
invphi<-as.numeric(sum(r^2)/(N-p))
phi<-1/invphi

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(1,N)
C <- A

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}
Omega <- invphi*Omega
invOmega <- phi*invOmega

#Matriz H e W
W <- C%*%invOmega%*%C

```

```

H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  auto<-eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors%*%
  sqrt(diag(auto$values))%*%t(auto$vectors)
  l <- l+t[i]
  i <- i+1
}
H <- raizW%*%X%*%H%*%t(X)%*%raizW
h <- diag(H)

```

```

#Ponto Alavanca por UE
hue<-as.vector(rep(0,n))
haux <- matrix(h,n,t[1],byrow=T)
for (i in 1:n) hue[i] <- sum(haux[i,])/t[i]

```

```

#Residuo Padronizado
rsd <- as.vector(rep(0,N))
part.rsd <- raizW%*%solve(C)%*%(y-mi)
for (l in 1:N)
{
  e <- as.vector(rep(0,N))
  e[l] <- 1
  rsd[l] <- t(e)%*%part.rsd/sqrt(1-h[l])
}

```

```

#Distancia de Cook
cd <- as.vector(rep(0,N))
for (l in 1:N) cd[l] <- (rsd[l])^2*h[l]/((1-h[l])*p)

```

```
#Construção dos gráficos de cd e rsd
plot(cluster,cd,xlab="Unidade Experimental", ylab="Distancia de Cook",
pch=16)
plot(cluster,rsd,xlab="Unidade Experimental", ylab="Residuo Padronizado",
pch=16)
```

Programa para gerar respostas normais correlacionadas

```
orig.fit <- fitted(orig.gee.all)
media <- as.vector(tapply(orig.fit, list(original[,2]), mean))
dp <- as.vector(sqrt(tapply(y, list(original[,2]),var)))
repl <- 25 #Número de réplicas indicada por Tan, Qu e Kutner
random.y <- array(dim=c(N,repl)) #Matriz com as respostas simuladas
for(i in 1:repl)
{
  grupo <- rmvnorm(n, mean=media, cov=covy, d=8)
  random.y[,i] <- c(t(grupo))
}
```

Programa para construir o gráfico de probabilidade meio-normal com envelope simulado para modelos normais lineares

```
orig.res <- rsd
ABSorig.res <- abs(orig.res)
SORTorig.res <- sort(ABSorig.res)
original2 <- cbind(original,random.y)
attach(original2)
random.res <- array(dim=c(N,repl))
for(k in 1:repl)
{
  temp.gee.all <- gee(random.y[,k] ~ bloco , sujeito, family=gaussian,
data = original2, corstr="AR-M",M=1)
  y <- temp.gee.all$y
  beta <- coef(temp.gee.all)
  R <- temp.gee.all$work
```

```

mi <- fitted(temp.gee.all)

#Calculo do Residuo de Pearson
r <- (y-mi)

#Calculo de phi
invphi <- as.numeric(sum(r^2)/(N-p))
phi <- 1/invphi

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(1,N)
C <- A

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}
Omega <- invphi*Omega
invOmega <- phi*invOmega

#Matriz H e W
W <- C%*%invOmega%*%C
H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)

```

```

i <- 1
l <- 1
while (l<N)
{
  auto<-eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors
  %*%sqrt(diag(auto$values))%*%t(auto$vectors)
  l <- l+t[i]
  i <- i+1
}
H <- raizW%*%X%*%H%*%t(X)%*%raizW
h <- diag(H)

random.rsd <- as.vector(rep(0,N))
part.rsd <- raizW%*%solve(C)%*%(y-mi)
for (l in 1:N)
{
  e <- as.vector(rep(0,N))
  e[l] <- 1
  random.rsd[l] <- t(e)%*%part.rsd*sqrt(1/(1-h[l]))
}
random.res[,k] <- random.rsd
}

ABSrandom.res <- abs(random.res)
SORTrandom.res <- array(dim=c(N,repl))
for(k in 1:repl) SORTrandom.res[,k] <- sort(ABSrandom.res[,k])

descritiva <- array(dim=c(N,3))
for(k in 1:N)
{
  descritiva[k,1] <- min(SORTrandom.res[k,])
  descritiva[k,2] <- median(SORTrandom.res[k,])
  descritiva[k,3] <- max(SORTrandom.res[k,])
}

```

```

}
Z<-array(dim=c(N,1))

for(i in 1:N) Z[i] <- qnorm((i+N-1/8)/(2*N+1/2))

final <- cbind(Z,descritiva,SORTorig.res)
faixa <- range(final[,5],final[,2],final[,4])
par(mfrow=c(1,1))
par(pty="s")
plot(final[,1],final[,5],xlab="Valor Esperado da Estatística de Ordem
Meio-Normal", ylab="Valor Absoluto Ordenado do Resíduo Padronizado",
ylim=faixa, pch=16)
par(new=T)
lines(final[,1],final[,2])
lines(final[,1],final[,3],lty=2)
lines(final[,1],final[,4])

```

B.2 Aplicação 2

A Figura B.2 nos mostra o formato do *data.frame* criado no software S-Plus 4.5 contendo os dados da Tabela A.2.

Para obter os resultados apresentados na Seção 4.2, os programas devem ser executados na mesma ordem em que são exibidos a seguir.

Programa para ajuste da regressão logística com correlação não-estruturada

```

original <- Depressao #Depressao contém os dados da Tabela A.2
cluster <- as.vector(original[,1])
attach(original)
library(gee)
orig.gee.all<-gee(y ~ (-1) + visita, paciente, family = binomial(link =
logit), data = original, corstr="unstructured", scale.fix = T,
scale.value = 1)
summary(orig.gee.all)

```


	1	2	3
	paciente	visita	y
1	1.00	1	1.00
2	1.00	2	1.00
3	1.00	3	1.00
4	2.00	1	1.00
5	2.00	2	1.00
6	2.00	3	1.00
7	3.00	1	0.00
8	3.00	2	1.00
9	3.00	3	1.00
10	4.00	1	0.00

Figura B.2: Formato do *data.frame* Depressao criado no software S-Plus 4.5.

```
n <- 172 #Número de pacientes no exemplo
t<-as.vector(rep(3,n)) #t=3 para todas as pacientes
```

Programa para calcular a distância de Cook e o resíduo padronizado para modelos de regressão logística

```
library(Matrix)
X <- model.matrix(orig.gee.all)
y <- orig.gee.all$y
beta <- coef(orig.gee.all)
R <- orig.gee.all$work
mi <- fitted(orig.gee.all)
N <- nrow(X)
p <- ncol(X)

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(mi*(1-mi),N)
C <- A
```

```

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}

#Matriz H e W
W <- C%*%invOmega%*%C
H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  auto <- eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors
  %*%sqrt(diag(auto$values))%*%t(auto$vectors)
  l <- l+t[i]
  i <- i+1
}
H <- raizW%*%X%*%H%*%t(X)%*%raizW
h <- diag(H)

#Ponto Alavanca por UE

```

```

hue <- as.vector(rep(0,n))
haux <- matrix(h,n,t[1],byrow=T)
for (i in 1:n) hue[i] <- sum(haux[i,])/t[i]

#Residuo Padronizado
rsd <- as.vector(rep(0,N))
part.rsd <- raizW%%solve(C)%%(y-mi)
for (l in 1:N)
{
e <- as.vector(rep(0,N))
e[l] <- 1
rsd[l] <- t(e)%%part.rsd/sqrt(1-h[l])
}

#Distancia de Cook
cd <- as.vector(rep(0,N))
for (l in 1:N) cd[l] <- (rsd[l])^2*h[l]/((1-h[l])*p)

#Construção dos gráficos de cd e rsd
plot(cluster,cd,xlab="Unidade Experimental", ylab="Distancia de Cook",
pch=16)
plot(cluster,rsd,xlab="Unidade Experimental", ylab="Residuo Padronizado",
pch=16)

```

Programa para gerar respostas binárias correlacionadas

```

repl <- 25 #Numero de replicas indica por Tan, Qu and Kutner
random.y <- array(dim=c(N,repl))
mi.fit <- matrix(mi,n,byrow=T)
indice <- matrix(c(1:t[1]),t[1],t[1],byrow=T)
cont <- 1
for (i in 1:n)
{
flag <- T
k <- 0

```

```

TS <- matrix(c(1:t[1]),t[1])
gama <- 0
media <- mi.fit[i,]
lambda <- log(matrix(1,t[i],t[i])+sqrt(diag((1-media)/media))
%*%R%*%sqrt(diag((1-media)/media)))
lambda <- round(lambda,6)
while (flag)
{
k <- k+1
gama.min <- min(lambda[lambda>0])
rs <- indice[lambda==gama.min]
S <- unique(rs)
for (j in 1:length(S))
for (l in j:length(S))
if (lambda[S[j],S[l]]==gama.min) rs <- c(S[j],S[l])
S <- sort(unique(rs))

for (j in 1:t[1])
if (all(lambda[j,S]>0)) S <- c(S,j)
S <- sort(unique(S))

for (j in 1:length(S))
for (l in 1:length(S))
lambda[S[j],S[l]] <- lambda[S[j],S[l]]-gama.min

ts <- rep(0,t[1])
for (j in 1:t[1])
if (is.element(j,S)) ts[j] <- 1

TS <- cbind(TS,ts)
gama <- c(gama,gama.min)

if (any(lambda)==0) flag <- F
}

```

```

TS <- TS[,-1]
gama <- gama[-1]
m <- k

for(j in 1:repl)
{
x <- rpois(1,gama[1])
for (k in 2:m) x <- cbind(x,rpois(1,gama[k]))
U <- TS%*%t(x)
Z <- as.vector(rep(0,t[i]))
Z[U==0] <- 1
random.y[cont:(cont+t[i]-1),j] <- Z
}

cont <- cont+t[i]
}

```

Programa para construir o gráfico de probabilidade meio-normal com envelope simulado para modelos de regressão logística

```

orig.res <- rsd
ABSorig.res <- abs(orig.res)
SORTorig.res <- sort(ABSorig.res)
original2 <- cbind(original,random.y)
attach(original2)
random.res <- array(dim=c(N,repl))
for(k in 1:repl)
{
temp.gee.all <- gee(random.y[,k] ~ (-1) + visita, paciente,
family = binomial(link = logit), data = original2, corstr="unstructured",
scale.fix = T, scale.value = 1)
summary(temp.gee.all)
}

```

```

y <- temp.gee.all$y
beta <- coef(temp.gee.all)
R <- temp.gee.all$work
mi <- fitted(temp.gee.all)

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(mi*(1-mi),N)
C <- A

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}

#Matriz H e W
W <- C%*%invOmega%*%C
H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  auto <- eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors
  %*%sqrt(diag(auto$values))%*%t(auto$vectors)

```

```

l <- l+t[i]
i <- i+1
}

H <- raizW%*%X%*%H%*%t(X)%*%raizW
h <- diag(H)

#Residuo Padronizado
random.rsd <- as.vector(rep(0,N))
part.rsd <- raizW%*%solve(C)%*%(y-mi)
for (l in 1:N)
{
e <- as.vector(rep(0,N))
e[l] <- 1
random.rsd[l] <- t(e)%*%part.rsd/sqrt(1-h[l])
}

random.res[,k] <- random.rsd
}

ABSrandom.res <- abs(random.res)
SORTrandom.res <- array(dim=c(N,repl))
for(k in 1:repl) SORTrandom.res[,k] <- sort(ABSrandom.res[,k])

descritiva <- array(dim=c(N,3))
for(k in 1:N)
{
descritiva[k,1] <- min(SORTrandom.res[k,])
descritiva[k,2] <- median(SORTrandom.res[k,])
descritiva[k,3] <- max(SORTrandom.res[k,])
}
Z<-array(dim=c(N,1))

for(i in 1:N) Z[i] <- qnorm((i+N-1/8)/(2*N+1/2))

```

```

final <- cbind(Z,descritiva,SORTorig.res)
faixa <- range(final[,5],final[,2],final[,4])
par(mfrow=c(1,1))
par(pty="s")
plot(final[,1],final[,5],xlab="Valor Esperado da Estatística de Ordem
Meio-Normal", ylab="Valor Absoluto Ordenado do Residuo Padronizado",
ylim=faixa, pch=16)
par(new=T)
lines(final[,1],final[,2])
lines(final[,1],final[,3],lty=2)
lines(final[,1],final[,4])

```

B.3 Aplicação 3

A Figura B.3 nos mostra o formato do *data.frame* criado no software S-Plus 4.5 contendo os dados da Tabela A.3.

	1	2	3	4	5	6
	paciente	tempo	droga	NGB	NGV	NCC
1	1.00	1	1	15.00	2.00	14.00
2	1.00	2	1	18.00	3.00	14.00
3	1.00	3	1	19.00	2.00	12.00
4	1.00	4	1	24.00	5.00	11.00
5	2.00	1	1	8.00	2.00	17.00
6	2.00	2	1	11.00	4.00	18.00
7	2.00	3	1	14.00	4.00	18.00
8	2.00	4	1	14.00	5.00	16.00
9	3.00	1	1	4.00	7.00	23.00
10	3.00	2	1	5.00	5.00	20.00
11	3.00	3	1	6.00	4.00	10.00

Figura B.3: Formato do *data.frame* MontPoisson criado no software S-Plus 4.5.

Para obter os resultados apresentados na Seção 4.3, os programas devem ser executados na mesma ordem em que são exibidos a seguir.

Programa para ajuste da regressão de Poisson com estrutura de correlação AR-1

```
original <- MontPoisson #MontPoisson contém os dados da Tabela A.3
cluster <- as.vector(original[,1])
attach(original)
library(gee)
orig.gee.all<-gee(NCC ~ (-1) + droga + NGB + NGV, paciente, family =
poisson(link = log), data = original, corstr="AR-M", M=1, scale.fix = T,
scale.value = 1)
summary(orig.gee.all)
n <- 30 #Numero de pacientes no exemplo
t<-as.vector(rep(4,n)) #t=4 para todos os pacientes
```

Programa para calcular a distância de Cook e o resíduo padronizado para modelos de regressão de Poisson

```
library(Matrix)
X <- model.matrix(orig.gee.all)
y <- orig.gee.all$y
beta <- coef(orig.gee.all)
R <- orig.gee.all$work
mi <- fitted(orig.gee.all)
N <- nrow(X)
p <- ncol(X)

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(mi,N)
C <- A

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
```

```

invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}

#Matriz H e W
W <- C%*%invOmega%*%C
H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  auto <- eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors
  %*%sqrt(diag(auto$values))%*%t(auto$vectors)
  l <- l+t[i]
  i <- i+1
}
H <- raizW%*%X%*%H%*%t(X)%*%raizW
h <- diag(H)

#Ponto Alavanca por UE
hue <- as.vector(rep(0,n))
haux <- matrix(h,n,t[1],byrow=T)
for (i in 1:n) hue[i] <- sum(haux[i,])/t[i]

```

```

#Residuo Padronizado
rsd <- as.vector(rep(0,N))
part.rsd <- raizW%%solve(C)%%(y-mi)
for (l in 1:N)
{
e <- as.vector(rep(0,N))
e[l] <- 1
rsd[l] <- t(e)%%part.rsd/sqrt(1-h[l])
}

#Distancia de Cook
cd <- as.vector(rep(0,N))
for (l in 1:N) cd[l] <- (rsd[l])^2*h[l]/((1-h[l])*p)

#Construção dos gráficos de h, hue, cd e rsd
plot(cluster,h,xlab="Unidade Experimental", ylab="H por observacao",
pch=16)
cut <- 2*p/N
abline(cut,0,lty=2)
plot(hue,xlab="Unidade Experimental", ylab="H por unidade experimental",
pch=16)
abline(cut,0,lty=2)
plot(cluster,cd,xlab="Unidade Experimental", ylab="Distancia de Cook",
pch=16)
plot(cluster,rsd,xlab="Unidade Experimental", ylab="Residuo Padronizado",
pch=16)

```

Programa para gerar respostas de Poisson correlacionadas

```

repl <- 25 #Numero de replicas indica por Tan, Qu and Kutner
random.y <- array(dim=c(N,repl))
mi.fit <- matrix(mi,n,byrow=T)
indice <- matrix(c(1:t[1]),t[1],t[1],byrow=T)
cont <- 1

```

```

for (i in 1:n)
{
flag <- T
k <- 0
TS <- matrix(c(1:t[1]),t[1])
gama <- 0
media <- mi.fit[i,]
lambda<-sqrt(diag(media))*%R%*sqrt(diag(media))
lambda <- round(lambda,6)
while (flag)
{
k <- k+1
gama.min <- min(lambda[lambda>0])
rs <- indice[lambda==gama.min]
S <- unique(rs)
for (j in 1:length(S))
for (l in j:length(S))
if (lambda[S[j],S[l]]==gama.min) rs <- c(S[j],S[l])
S <- sort(unique(rs))

for (j in 1:t[1])
if (all(lambda[j,S]>0)) S <- c(S,j)
S <- sort(unique(S))

for (j in 1:length(S))
for (l in 1:length(S))
lambda[S[j],S[l]] <- lambda[S[j],S[l]]-gama.min

ts <- rep(0,t[1])
for (j in 1:t[1])
if (is.element(j,S)) ts[j] <- 1

TS <- cbind(TS,ts)
gama <- c(gama,gama.min)

```

```

if (any(lambda)==0) flag <- F
}

TS <- TS[,-1]
gama <- gama[-1]
m <- k

for(j in 1:repl)
{
x <- rpois(1,gama[1])
for (k in 2:m) x <- cbind(x,rpois(1,gama[k]))
U <- TS%*%t(x)
random.y[cont:(cont+t[i]-1),j] <- U
}

cont <- cont+t[i]
}

```

Programa para construir o gráfico de probabilidade meio-normal com envelope simulado para modelos de regressão de Poisson

```

orig.res <- rsd
ABSorig.res <- abs(orig.res)
SORTorig.res <- sort(ABSorig.res)
original2 <- cbind(original,random.y)
attach(original2)
random.res <- array(dim=c(N,repl))
for(k in 1:repl)
{
temp.gee.all<-gee(random.y[,k] ~ (-1) + droga + NGB + NGV, paciente,
family = poisson(link = log), data = original2, corstr="AR-M", M=1,
scale.fix = T, scale.value = 1)
summary(temp.gee.all)
}

```

```

y <- temp.gee.all$y
beta <- coef(temp.gee.all)
R <- temp.gee.all$work
mi <- fitted(temp.gee.all)

#Matriz C <- A * Delta (Ligação canonica -> Delta=Identidade)
A <- diag(mi,N)
C <- A

#Matriz Omega - variancia e covariancia de y
Omega <- matrix(0,N,N)
invOmega <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  Omega[l:(l+t[i]-1),l:(l+t[i]-1)] <- sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  %*%R%*%sqrt(A[l:(l+t[i]-1),l:(l+t[i]-1)])
  invOmega[l:(l+t[i]-1),l:(l+t[i]-1)] <-
  solve(Omega[l:(l+t[i]-1),l:(l+t[i]-1)])
  l <- l+t[i]
  i <- i+1
}

#Matriz H e W
W <- C%*%invOmega%*%C
H <- solve(t(X)%*%W%*%X)
raizW <- matrix(0,N,N)
i <- 1
l <- 1
while (l<N)
{
  auto <- eigen.Hermitian(W[l:(l+t[i]-1),l:(l+t[i]-1)])
  raizW[l:(l+t[i]-1),l:(l+t[i]-1)] <- auto$vectors

```

```

%%sqrt(diag(auto$values))%%t(auto$vectors)
l <- l+t[i]
i <- i+1
}

H <- raizW%%X%%H%%t(X)%%raizW
h <- diag(H)

#Residuo Padronizado
random.rsd <- as.vector(rep(0,N))
part.rsd <- raizW%%solve(C)%%(y-mi)
for (l in 1:N)
{
e <- as.vector(rep(0,N))
e[l] <- 1
random.rsd[l] <- t(e)%%part.rsd/sqrt(1-h[l])
}

random.res[,k] <- random.rsd
}

ABSrandom.res <- abs(random.res)
SORTrandom.res <- array(dim=c(N,repl))
for(k in 1:repl) SORTrandom.res[,k] <- sort(ABSrandom.res[,k])

descritiva <- array(dim=c(N,3))
for(k in 1:N)
{
descritiva[k,1] <- min(SORTrandom.res[k,])
descritiva[k,2] <- median(SORTrandom.res[k,])
descritiva[k,3] <- max(SORTrandom.res[k,])
}
Z<-array(dim=c(N,1))

```

```
for(i in 1:N) Z[i] <- qnorm((i+N-1/8)/(2*N+1/2))

final <- cbind(Z,descritiva,SORTorig.res)
faixa <- range(final[,5],final[,2],final[,4])
par(mfrow=c(1,1))
par(pty="s")
plot(final[,1],final[,5],xlab="Valor Esperado da Estatística de Ordem
Meio-Normal", ylab="Valor Absoluto Ordenado do Residuo Padronizado",
ylim=faixa, pch=16)
par(new=T)
lines(final[,1],final[,2])
lines(final[,1],final[,3],lty=2)
lines(final[,1],final[,4])
```


Referências Bibliográficas

- [1] Anderson, D. A. and Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B* **47**: 203-10.
- [2] Andreoni, S. (1989). *Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo*. São Paulo: IME-USP. Dissertação de Mestrado. 142p.
- [3] Artes, R. (1997). *Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão*. São Paulo: IME-USP. Tese de Doutorado. 130p.
- [4] Ashby, M., Neuhaus, J. M., Hauck, W. W., Bacchetti, P., Heilbron, D. C., Jewell, N. P., Segal, M. R. and Fusaro, R. E. (1992). An annotated bibliography of methods for analysing correlated categorical data. *Statistics in Medicine* **11**: 67-99.
- [5] Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford Statistical Science Series, Oxford.
- [6] Cardoso, J. (2000). *Teste para Hipóteses Restritas em Desigualdades Lineares usando Equações de Estimação Generalizadas*. São Paulo: IME-USP. Tese de Doutorado. 195p.
- [7] Chandrasekar, B. and Kale, B. K. (1984). Unbiased statistical estimation functions in presence of nuisance parameter. *Journal of Statistical Planning and Inference*, **9**: 45-54.
- [8] Chang, Y.-C. (2000). Residuals analysis of the generalized linear models for longitudinal data. *Statistics in Medicine*, **19**: 1277-93.

- [9] Crowder, M. (1987). On linear and quadratic estimating function. *Biometrika*, **74**: 591-7.
- [10] Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**: 1208-11.
- [11] Godambe, V. P. (1991). *Estimating Functions*. Oxford, Oxford University Press (Oxford Statistical Science series – 7).
- [12] Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman & Hall.
- [13] Johnston, G. (1996). *Repeated measures analysis with discrete data using the SAS System*. SAS Institute Inc. Cary, NC.
<http://academic.son.wisc.edu/rdsu/pdf/gee.pdf>
- [14] Jørgensen, B. and Labouriau, R. S. (1994). *Exponential Families and Theoretical Inference*. Lecture Notes. Department of Statistical. University of British Columbia.
- [15] Koch, G. C., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehmann, R. B. (1977). A general methodology for the analysis of repeated measurements of categorical data. *Biometrics*, **33**: 133-58.
- [16] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**: 963-74.
- [17] Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal analysis using generalized linear models. *Biometrika*, **73**: 13-22.
- [18] Lima, A. C. P. e Sañudo, A. (1997). *Transferência entre Tarefas Sincronizatórias com Diferentes Níveis de Complexidade*. São Paulo: IME-USP. RAE-CEA-9602.
- [19] Montgomery, D. C., Myers, R. H. and Vining, G. (2001). *Generalized Linear Models: with Applications in Engineering and the Sciences*. New York: John Wiley & Sons.
- [20] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**: 370-84.

- [21] Neter, J., Kutner, M. H., Nachstheim, C. J. and Wasserman, W. (1996). *Applied Linear Statistical Models*. 4.ed. Chicago: IE McGraw Hill.
- [22] Ochi, Y. and Prentice, R. L. (1984). Likelihood inference in correlated probit regression. *Biometrika*, **71**: 531-43.
- [23] Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, **57**: 120-5.
- [24] Park, T. (1993). A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Statistics in Medicine*, **12**: 1723-32.
- [25] Park, C. G., Park, T. and Shin, D. W. (1996). A simple method for generating correlated binary variates. *The American Statistician*, **50**: 306-10.
- [26] Park, C. G. and Shin, D. W. (1998). An algorithm for generating correlated random variables in a class of infinitely divisible distributions. *J. Statist. Comput. Simul.*, **61**: 127-39.
- [27] Paula, G. A. (2001). *Modelos de regressão com apoio computacional*. Notas de aulas. Departamento de Estatística. Universidade de São Paulo.
<http://www.ime.usp.br/~giapaula>
- [28] Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika*, **83**: 551-62.
- [29] Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**: 705-24.
- [30] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**: 1033-48.
- [31] Prentice, R. L. and Zhao, L. P. (1991). Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. *Biometrics* **47**: 825-39.
- [32] Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**: 485-97.

- [33] Rotnitzky, A. and Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics* **50**: 1163-70.
- [34] Rubin, R. M. (1976). Inference and missing data. *Biometrika* **63**: 581-92.
- [35] Singer, J. M. e Andrade, D. F. (1986). *Análise de dados longitudinais*. São Paulo: Campinas. VII Simpósio Nacional de Probabilidade e Estatística. 106p.
- [36] Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, **40**: 961-71.
- [37] Tan, M., Qu, Y. and Kutner, M. H. (1997). Model diagnostics for marginal regression analysis of correlated binary data. *Commun. Statist. - Simula.*, **26**: 539-58.
- [38] Venables, W. N. e Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*. 3.ed. New York: Springer-Verlag.
- [39] Ziegler, A., Kastner, C. and Blettner, M. (1998). The Generalised Estimating Equations: An Annotated Bibliography. *Biometrical Journal*, **40** 2: 115-39.