

**Influência local em
modelos lineares generalizados mistos
com variável resposta discreta**

Alejandra Andrea Tapia Silva

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Prof.^a Dr.^a Viviana Giampaoli

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e
CONICYT (Chile)

São Paulo, 10 de dezembro de 2014

**Influência local em
modelos lineares generalizados mistos
com variável resposta discreta**

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 10/12/2014. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof.^a Dr.^a Viviana Giampaoli (orientadora) - IME-USP
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP
- Prof. Dr. Mário de Castro Andrade Filho - ICMC-USP
- Prof. Dr. Manuel Jesus Galea Rojas - PUC-Chile
- Prof. Dr. Germán Mauricio Ibacache Pulgar - UV-Chile

Agradecimentos

Primeiramente, agradeço a minha família e, em especial, a meus pais Verónica e Jorge por terem-me apoiado todo o tempo.

A Mauricio pelo apoio e ânimo constante tanto à distância como durante suas visitas aqui no Brasil.

A minha orientadora Viviana Giampaoli pelas sugestões, conselhos e ter confiado em mim ao longo da tese.

À Professora María del Pilar Díaz por sua importante participação nos inícios e ao longo do trabalho.

A minhas grandes amigas do IB-USP e a meus grandes amigos e amigas do IME-USP.

A CAPES e CONICYT (Chile) pelo apoio financeiro.

Resumo

Neste trabalho é proposta uma metodologia de análise de diagnóstico de influência local em modelos lineares generalizados mistos com variável resposta discreta. A medida de influência sugerida por Zhu e Lee (2003) é estudada e aplicada a dados agrupados. A análise foi desenvolvida para o modelo de regressão logística com efeitos mistos sob duas estratégias de perturbação: a perturbação multiplicativa e a perturbação multiplicativa apropriada da probabilidade de sucesso. Estas abordagens permitem uma avaliação indireta da perturbação na resposta. Também foi desenvolvida a análise de diagnóstico influência local no modelo de regressão Poisson com efeitos mistos para as estratégias de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada da taxa média de ocorrências. Para cada caso a curvatura normal foi derivada.

O desempenho da proposta como ferramenta de diagnóstico foi estudada por meio de simulações de Monte Carlo sob as diferentes estratégias de perturbação. Em particular, para o modelo de regressão logística com efeitos mistos se estudou também a influência na capacidade preditiva do modelo. Por outro lado, a metodologia foi ilustrada com aplicações a conjuntos de dados reais. Em todos os casos abordados obtiveram-se resultados muito satisfatórios.

Palavras-chave: modelos lineares generalizados mistos, variável resposta discreta, influência local, perturbação, perturbação apropriada.

Abstract

This work proposes a methodology of diagnostic analysis of local influence in generalized linear mixed models with discrete response variable. The measure of influence suggested for Zhu and Lee (2003) was studied and applied to clustered data. The diagnostic analysis was developed for mixed logistic regression model under two strategies of perturbation, the multiplicative perturbation and the appropriate multiplicative perturbation of the probability of successes. These approaches allow an indirect assesment of perturbation in the response. An diagnostic analysis of local influence in the Poisson mixed regression model was developed as well, for strategies of quadratic multiplicative perturbation and appropriate quadratic multiplicative perturbation on the average rate of occurrences. For each case the normal curvature was derived.

The performance of the proposal as diagnostic tool was studied by Monte Carlo simulations under different perturbation strategies. In particular, for mixed logistic regression models was also studied the influence in the predictive capacity of model. Moreover, the methodology was illustrated with applications to real data sets. In all the cases, there were very satisfactory results.

Keywords: generalized linear mixed models, discrete response variable, local influence, perturbation, appropriate perturbation.

Sumário

Lista de figuras	xi
Lista de tabelas	xix
1 Introdução	1
1.1 Organização do trabalho	3
2 Conceitos relacionados	5
2.1 Modelos lineares generalizados	5
2.1.1 Estimação de parâmetros	6
2.2 Modelos lineares generalizados mistos	7
2.2.1 Estimação de parâmetros	9
2.3 Análise de diagnóstico de influência local	11
2.4 Análise de diagnóstico de influência local nos modelos lineares generalizados mistos	15
2.4.1 Derivação da curvatura normal	18
2.4.2 Esquemas de perturbação	18
2.4.3 Aproximação das matrizes da curvatura normal	22
2.5 Esquema de perturbação apropriada	23
3 Influência local no modelo de regressão logística com efeitos mistos	25
3.1 Modelo de regressão logística com efeitos mistos	25
3.1.1 Medidas de desempenho da predição	26
3.2 Análise de diagnóstico de influência local	29
3.2.1 Derivação da curvatura normal	30
3.2.2 Esquema de perturbação multiplicativa da probabilidade de sucesso	30
3.2.3 Esquema de perturbação multiplicativa apropriada da probabilidade de sucesso	31
3.2.4 Aproximação das matrizes da curvatura normal	33

3.3	Estudos de simulação	35
3.3.1	Estudo de simulação 1	37
3.3.2	Estudo de simulação 2	51
3.4	Aplicações	55
3.4.1	Aplicação 1	55
3.4.2	Aplicação 2	64
3.4.3	Aplicação 3	74
3.5	Aspectos computacionais	83
3.5.1	Recursos computacionais	83
3.5.2	Problemas computacionais dos estudos de simulação	83
3.5.3	Tempos de execução dos estudos de simulação	84
4	Influência local no modelo de regressão Poisson com efeitos mistos	85
4.1	Modelo de regressão Poisson com efeitos mistos	85
4.2	Análise de diagnóstico de influência local	86
4.2.1	Derivação da curvatura normal	87
4.2.2	Esquema de perturbação multiplicativa quadrática da taxa média de ocorrências	87
4.2.3	Esquema de perturbação multiplicativa quadrática apropriada da taxa média de ocorrências	88
4.2.4	Aproximação das matrizes da curvatura normal	90
4.3	Estudos de simulação	91
4.3.1	Estudo de simulação 1	93
4.3.2	Estudo de simulação 2	104
4.4	Aplicações	108
4.4.1	Aplicação 1	108
4.4.2	Aplicação 2	117
4.4.3	Aplicação 3	125
4.5	Aspectos computacionais	132
4.5.1	Recursos computacionais	132
4.5.2	Problemas computacionais dos estudos de simulação	132
4.5.3	Tempos de execução dos estudos de simulação	132
5	Conclusões	135
5.1	Considerações finais	135

5.2	Sugestões para pesquisas futuras	136
A	Derivadas	139
A.1	Derivadas relacionadas ao Capítulo 3	139
A.2	Derivadas relacionadas ao Capítulo 4	142
B	Gráficos e tabelas	145
B.1	Gráficos e tabelas relacionados às aplicações do Capítulo 3	145
B.1.1	Aplicação 1	145
B.1.2	Aplicação 2	148
B.1.3	Aplicação 3	151
B.2	Gráficos e tabelas relacionados às aplicações do Capítulo 4	152
B.2.1	Aplicação 1	152
B.2.2	Aplicação 2	157
B.2.3	Aplicação 3	160
	Referências bibliográficas	163

Lista de figuras

2.1	Gráfico de influência.	13
2.2	Gráfico da linha projetada.	14
3.1	Gráfico da curva ROC.	28
3.2	Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1)(MRLEM)	39
3.3	Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)	40
3.4	Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)	41
3.5	Média, viés e REQM da estimativa do parâmetro σ para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM)	42
3.6	Média, viés e REQM da estimativa do parâmetro σ para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)	43
3.7	Média, viés e REQM da estimativa do parâmetro σ para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)	44
3.8	Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM)	45

3.9 Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM) 46

3.10 Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM) 47

3.11 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM) 48

3.12 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM) 49

3.13 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM) 50

3.14 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 2) (MRLEM) 52

3.15 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 2) (MRLEM) 53

3.16 Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 2) (MRLEM) 54

3.17 Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.31). 57

3.18 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.31). 59

3.19 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.31). 59

3.20 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.31). 60

3.21 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.31). 60

3.22 Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.32). 66

3.23 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.32). 67

3.24 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.32). 68

3.25 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.32). 68

3.26 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.32). 69

3.27 Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.33). 75

3.28 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.33). 77

3.29 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.33). 77

3.30 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.33). 78

3.31 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.33). 78

4.1 Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM) 95

4.2 Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM) 96

4.3 Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM) 97

4.4 Média, viés e REQM da estimativa do parâmetro σ para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM) 98

4.5	Média, viés e REQM da estimativa do parâmetro σ para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM)	99
4.6	Média, viés e REQM da estimativa do parâmetro σ para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM)	100
4.7	Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM)	101
4.8	Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM)	102
4.9	Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM)	103
4.10	Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 2) (MRPEM)	105
4.11	Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 2)(MRPEM)	106
4.12	Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 2) (MRPEM)	107
4.13	Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.32).	111
4.14	Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.32).	112
4.15	Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.32).	112

4.16 Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.33). 119

4.17 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.33). 120

4.18 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.33). 121

4.19 Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.34). 127

4.20 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.34). 128

4.21 Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.34). 128

B.1 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os terrários com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (3.31). 146

B.2 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os terrários com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (3.31). 147

B.3 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para as bandejas com $S = 10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (3.32). 149

B.4 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para as bandejas com $S = 10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (3.32). 150

B.5 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os grupos com $S = 10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (3.33). 151

B.6 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os grupos com $S = 10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (3.33). 151

B.7 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.32). 153

B.8 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.32). 155

B.9 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.33). 158

B.10 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.33). 159

B.11 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os indivíduos com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.34). 161

B.12 Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os indivíduos com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.34). 162

Lista de tabelas

3.1	Matriz de confusão.	29
3.2	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.31).	56
3.3	Predições BLUP do intercepto aleatório para os terrários pelos métodos AL e AQGH relacionados ao modelo (3.31).	57
3.4	Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.31).	62
3.5	Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.31).	63
3.6	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.32).	65
3.7	Predições BLUP do intercepto aleatório para as bandejas pelos métodos AL e AQGH relacionados ao modelo (3.32).	65
3.8	Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.32).	71
3.9	Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.32).	72
3.10	Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.32).	73
3.11	Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.32).	73

3.12	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.33).	75
3.13	Predições BLUP do intercepto aleatório para os grupos pelos métodos AL e AQGH relacionados ao modelo (3.33).	75
3.14	Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.33).	80
3.15	Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.33).	81
3.16	Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.33).	82
3.17	Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.33).	82
3.18	Número de réplicas adicionais para cada tamanho amostral q e valor da perturbação ω_{ij} nos estudos de simulação 1 e 2. (MRLEM)	84
3.19	Tempos de execução (em horas) dos estudos de simulação 1 e 2 para cada tamanho amostral q e valor da perturbação ω_{ij} . (MRLEM)	84
4.1	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.32).	109
4.2	Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.32).	110
4.3	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelo método AL relacionados ao modelo (4.32).	114
4.4	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelo método AQGH relacionados ao modelo (4.32).	115
4.5	Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.32).	116

4.6	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.33).	117
4.7	Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.33).	118
4.8	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.33).	122
4.9	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.33).	123
4.10	Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.33).	124
4.11	Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.33).	124
4.12	Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.34).	126
4.13	Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.34).	126
4.14	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.34).	129
4.15	Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.34).	130
4.16	Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.34).	131
4.17	Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.34).	131

4.18	Número de réplicas adicionais para cada tamanho amostral q e valor da perturbação ω_{ij}^2 nos estudos de simulação 1 e 2. (MRPEM)	132
4.19	Tempos de execução (em horas) dos estudos de simulação 1 e 2 para cada tamanho amostral q e valor da perturbação ω_{ij}^2 . (MRPEM)	133
B.1	Predições do intercepto aleatório para os terrários baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.31).	145
B.2	Predições do intercepto aleatório para as bandejas baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.32).	148
B.3	Predições do intercepto aleatório para os grupos baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.33).	151
B.4	Predições do intercepto aleatório para os pacientes baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.32).	152
B.5	Predições do intercepto aleatório para os pacientes baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.33).	157
B.6	Predições do intercepto aleatório para os indivíduos baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.34).	160

Capítulo 1

Introdução

Um modelo estatístico é uma ferramenta matemática extremamente útil para extrair e entender as características essenciais de um conjunto de dados. No entanto, este corresponde a uma descrição aproximada da realidade pelo que sua falta de exactidão pode levar a resultados pouco confiáveis. Depois da seleção e ajuste do modelo é necessário realizar o que se conhece como análise de diagnóstico. A análise de diagnóstico é uma etapa muito importante na modelagem estatística, pois permite verificar possíveis afastamentos das suposições do modelo e detectar observações que causam variações desproporcionais nas estimativas dos parâmetros. Com respeito a este último ponto, uma das metodologias com grande receptividade é a técnica de análise de diagnóstico de influência local apresentada por [Cook \(1986\)](#) para o modelo normal linear e abordada posteriormente por vários autores para outros modelos (vide, por exemplo, [Ibacache-Pulgar et al., 2012](#); [Osorio et al., 2009](#); [Villegas et al., 2013](#)). Basicamente, esta técnica consiste em estudar o comportamento de alguma medida de influência sob pequenas perturbações ou modificações nas suposições do modelo ou nos dados. Isto é, identificar a existência de observações que sob modificações modestas nas suposições do modelo ou nos dados causem variações importantes nas estimativas dos parâmetros. [Cook \(1986\)](#) utiliza a medida de influência denominada afastamento da verossimilhança que corresponde a duas vezes a diferença entre o logaritmo máximo da função de verossimilhança do modelo postulado e o logaritmo máximo da função de verossimilhança do modelo sob algum esquema de perturbação. Em seguida, mediante um enfoque geométrico analisa-se a medida afastamento da verossimilhança versus a perturbação e obtém-se informação em relação à influência do esquema de perturbação adotado.

Os modelos lineares generalizados mistos estendem os modelos lineares generalizados através da incorporação de efeitos aleatórios no preditor linear. Dessa forma, induz-se uma estrutura de corre-

lação entre as observações a qual é de fundamental importância em conjuntos de dados agrupados e provenientes de estudos longitudinais, entre outros. Neste contexto, aplicar diretamente a técnica de influência local de Cook (1986) é extremamente difícil pois o logaritmo da função de verossimilhança compreende integrais que não têm solução analítica e de grande dimensão. Para evitar essa dificuldade, Zhu e Lee (2003) tratam os efeitos aleatórios como dados faltantes e utilizam a medida de influência denominada função Q-afastamento que corresponde a duas vezes a diferença entre o valor esperado máximo do logaritmo da função de verossimilhança do modelo postulado e o valor esperado máximo do logaritmo da função de verossimilhança do modelo sob algum esquema de perturbação. Assim, analogamente a Cook (1986), utiliza-se o enfoque geométrico para analisar a medida função Q-afastamento versus a perturbação. Conforme com as características dos modelos lineares generalizados mistos, Zhu e Lee (2003) propõem os seguintes esquemas de perturbação: perturbação da ponderação das observações, perturbação da ponderação dos grupos, perturbação da matriz de variâncias e covariâncias da distribuição normal suposta para os efeitos aleatórios, perturbação simultânea da ponderação dos grupos e efeitos aleatórios, perturbação das variáveis explicativas e perturbação da variável resposta no caso contínuo. Contudo, o esquema de perturbação para a variável resposta no caso discreto não foi desenvolvido por Zhu e Lee (2003) e não se tem conhecimento de tere sido desenvolvido por outros pesquisadores para esta classe de modelos. Assim, o principal propósito deste trabalho consiste em propor estratégias de perturbação para a variável resposta discreta de alguns modelos lineares generalizados mistos.

Centrar-nos-emos no modelo de regressão logística com efeitos mistos e no modelo de regressão Poisson com efeitos mistos, os quais têm sido foco de trabalho de muitos pesquisadores (vide, por exemplo, Diggle *et al.*, 1996; Gibbons *et al.*, 2010; Goldstein *et al.*, 2002; Laird e Ware, 1982; Stiratelli *et al.*, 1984, entre outros). Assim, para o modelo de regressão logística com efeitos mistos propomos um esquema de perturbação indireto para a variável resposta binária dado pela perturbação do parâmetro de probabilidade do modelo. Para o modelo de regressão Poisson com efeitos mistos propomos um esquema de perturbação indireto para a variável resposta de contagem dado pela perturbação do parâmetro da taxa média de ocorrências do modelo. Além disso, baseados no trabalho de Chen *et al.* (2010), propomos as versões apropriadas dos esquemas de perturbação anteriormente mencionados.

Finalmente, para avaliar o desempenho das propostas realizamos estudos de simulação de Monte Carlo para diferentes cenários. Por outra parte, para ilustrar o desempenho destas mesmas realizamos aplicações a conjuntos de dados reais. Especificamente, para o modelo de regressão logística com efeitos mistos os conjuntos de dados são referentes ao comportamento reprodutivo da espécie de

aracnídeos opilião do Brasil (Werneck, 2012), ao estudo da germinação de sementes (Crowder, 1978) e ao estudo da infecção pelo HPV (*Human Papillomavirus*) em mulheres chilenas (Ferrecio *et al.*, 2011). Para o modelo de regressão Poisson com efeitos mistos, os conjuntos de dados são referentes a um ensaio clínico em pacientes com epilepsia (Thall e Vail, 1990), a um estudo em pacientes que padecem de enxaquecas (McKnight e Eeden, 1993) e a um estudo caso-controlado envolvendo pacientes com diabetes mellitus tipo I (Giampaoli *et al.*, 2014).

Tanto os estudos de simulação quanto as aplicações são desenvolvidos na linguagem de programação R, versão 3.0.1 (R Core Team, 2014).

1.1 Organização do trabalho

Este trabalho é organizado da seguinte maneira:

O Capítulo 2 trata de uma revisão dos conceitos mais importantes para o desenvolvimento do trabalho. Especificamente, apresentamos os modelos lineares generalizados e os modelos lineares generalizados mistos. Logo, apresentamos a técnica de análise de diagnóstico de influência local de Cook (1986) e sua abordagem para os modelos lineares generalizados mistos de Zhu e Lee (2003). Por último, expomos a metodologia para selecionar um esquema de perturbação apropriada, de acordo com Chen *et al.* (2010).

No Capítulo 3, apresentamos o modelo de regressão logística com efeitos mistos e desenvolvemos a análise de diagnóstico de influência local sob os esquemas de perturbação multiplicativa e multiplicativa apropriada. Em seguida, avaliamos o desempenho das análises com estudos de simulação para diferentes cenários e ilustramos estas mesmas mediante três aplicações a conjuntos de dados reais. Por último, apresentamos os aspectos computacionais relacionados aos estudos de simulação e às aplicações.

No Capítulo 4, abordamos o modelo de regressão Poisson com efeitos mistos e desenvolvemos a análise de diagnóstico de influência local sob os esquemas de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada. Posteriormente, avaliamos os desempenhos das análises com estudos de simulação para diferentes cenários e ilustramos estas mesmas mediante três aplicações a conjuntos de dados reais. Por último, apresentamos os aspectos computacionais relacionados aos estudos de simulação e às aplicações.

No Capítulo 5 apresentamos as conclusões e sugestões para pesquisas futuras.

Finalmente, detalhes relacionados aos Capítulos 3 e 4 podem ser encontrados nos Apêndices A e B e os programas computacionais correspondentes estão disponíveis para consulta.

Capítulo 2

Conceitos relacionados

Neste capítulo apresentamos uma revisão dos conceitos fundamentais para o desenvolvimento dos Capítulos 3 e 4. Especificamente, na Seção 2.1 explanamos sobre os modelos lineares generalizados (MLG) e na Seção 2.2 sobre os MLGM e os principais métodos de aproximação de integrais para a estimação de parâmetros por máxima verossimilhança. Na Seção 2.3 descrevemos a técnica de análise de diagnóstico de influência local e sua extensão para os MLGM y na Seção 2.5 a metodologia para selecionar um esquema de perturbação apropriada.

2.1 Modelos lineares generalizados

Com o objetivo de apresentar os MLG define-se o modelo linear clássico ou modelo normal linear por: (i) as respostas são independentes, (ii) cada resposta segue uma distribuição normal (iii) com média igual ao preditor linear e (iv) variância constante. No entanto, quando (ii) e (iv) não se cumpriam, durante muito tempo foram utilizadas técnicas de transformação nas respostas como, por exemplo, a transformação de [Box e Cox \(1964\)](#). Os autores [Nelder e Wedderburn \(1972\)](#) propuseram estender o modelo normal linear através dos MLG, isto é, supondo que cada resposta segue uma distribuição pertencente à família exponencial com média igual a uma função não linear do preditor linear e variância igual a uma função da média. Assim, os MLG incluem uma variedade de modelos além do modelo normal linear tais como o probito, logístico e log-lineares, entre outros (vide, por exemplo, [McCullagh e Nelder, 1989](#); [Paula, 2013](#)). Formalmente, a definição dos MLG é dada a seguir.

Considere Y_i , com $i = 1, \dots, n$, variáveis aleatórias independentes com função de densidade

pertencente à família exponencial dada por

$$p(y_i, \theta_i, \phi) = \exp[\phi(y_i\theta_i - a(\theta_i)) + c(y_i, \phi)], \quad (2.1)$$

em que θ_i é o parâmetro canônico, $a(\cdot)$ e $c(\cdot)$ são funções conhecidas e ϕ é o parâmetro de dispersão (conhecido ou desconhecido). Logo, sob condições usuais de regularidade (ver [Sen et al., 2010](#)) temos que

$$E\left(\frac{\partial \log p(y_i, \theta_i, \phi)}{\partial \theta_i}\right) = 0$$

e

$$E\left(\frac{\partial^2 \log p(y_i, \theta_i, \phi)}{\partial \theta_i^2}\right) = -E\left[\left(\frac{\partial \log p(y_i, \theta_i, \phi)}{\partial \theta_i}\right)^2\right],$$

o que implica que $E(Y_i) = \mu_i = \dot{a}(\theta_i)$, com $\dot{a}(\theta_i) = \partial a(\theta_i)/\partial \theta_i$ e $V(Y_i) = \ddot{a}(\theta_i)/\phi$, com $\ddot{a}(\theta_i) = \partial^2 a(\theta_i)/\partial \theta_i^2 = V(\mu_i)$, a função de variância. Note que $V(\mu_i) = \partial \mu_i/\partial \theta_i$ ($\theta_i = \int V(\mu_i)^{-1} d\mu_i$). Portanto, os MLG são definidos por (2.1) e pelo componente sistemático

$$g(\mu_i) = \eta_i, \quad (2.2)$$

em que $\eta_i = \mathbf{x}'_i \boldsymbol{\beta}$ é o preditor linear, com $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ o vetor de covariáveis de dimensão p , $\boldsymbol{\beta}$ o vetor de coeficientes de regressão fixos de dimensão p ($p < n$) e $g(\cdot)$ uma função conhecida, monótona e diferenciável denominada função de ligação. Em particular, se a função de ligação é igual à função que define o parâmetro canônico, ou seja, $g(\mu_i) = \theta_i$, a função é denominada função de ligação canônica. As distribuições mais conhecidas que pertencem a esta classe com suas principais características são:

- binomial com $\theta_i = \log[\mu_i/(1 - \mu_i)]$, $a(\theta_i) = \log[1 + \exp(\theta_i)]$, $\phi = n$ e $V(\mu_i) = \mu_i(1 - \mu_i)$ e $g(\mu_i) = \log[\mu_i/(n - \mu_i)]$;
- Poisson com $\theta_i = \log(\mu_i)$, $a(\theta_i) = \exp(\theta_i)$, $\phi = 1$ e $V(\mu_i) = \mu_i$ e $g(\mu_i) = \log(\mu_i)$;
- normal com $\theta_i = \mu_i$, $a(\theta_i) = \theta_i^2/2$, $\phi = \sigma^{-2}$ e $V(\mu_i) = 1$ e $g(\mu_i) = \mu_i$.

2.1.1 Estimação de parâmetros

A estimação de parâmetros de um MLG é usualmente realizada mediante o método de máxima verossimilhança descrito a seguir. Considere o modelo definido por (2.1) e (2.2), com ϕ conhecido, e defina por \mathbf{Y}_o o conjunto de dados observados em Y_i , com $i = 1, \dots, n$. Logo, o logaritmo da função

de verossimilhança do vetor de parâmetros β é dado por

$$L(\beta|\mathbf{Y}_o) = \sum_{i=1}^n \phi(y_i\theta_i - a(\theta_i)) + c(y_i, \phi),$$

e a estimativa de máxima verossimilhança é obtida tal que $\partial L(\beta|\mathbf{Y}_o)/\partial\beta = \mathbf{0}$, ou seja,

$$\sum_{i=1}^n \phi(y_i - \mu_i)V(\mu_i)^{-1} \left(\frac{\partial\mu_i}{\partial\eta_i} \right) \mathbf{x}_i = \mathbf{0}. \quad (2.3)$$

A equação em (2.3), em geral, não tem solução analítica e métodos iterativos tais como o de *Newton-Raphson* ou de *escore de Fisher* são necessários (vide, por exemplo, Paula, 2013). Cabe destacar que quando usarmos funções de ligação canônicas garante-se a concavidade de $L(\beta|\mathbf{Y}_o)$ e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a unicidade assintótica da estimativa de máxima verossimilhança quando esta existe. Por último, na literatura um grande número de aplicações no contexto dos MLG têm sido publicado. Algumas delas podem ser encontrados em McCullagh e Nelder (1989); McCulloch e Searle (2001), e em inúmeros artigos das diferentes áreas do conhecimento.

2.2 Modelos lineares generalizados mistos

Os MLG consideram situações em que as observações são independentes, porém existem muitas outras em que as observações são correlacionadas como consequência dos dados estarem agrupados ou serem provenientes de estudos longitudinais (vide, por exemplo, Demidenko, 2013; Jiang, 2007; Lee *et al.*, 2006; Verbeke e Molenberghs, 2000, entre outros). Logo, uma forma de lidar com a correlação é com a inclusão de variáveis aleatórias não observáveis denominadas efeitos aleatórios. Assim, considerando efeitos aleatórios no preditor linear dos MLG, obtêm-se os MLGM. Os principais elementos dos MLGM são: (i) condicionadas aos efeitos aleatórios, as respostas são independentes, (ii) cada resposta segue uma distribuição condicional pertencente à família exponencial e (iii) os efeitos aleatórios seguem uma distribuição normal. Formalmente, a definição dos MLGM é dada a seguir.

Considere Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$ variáveis aleatórias independentes para i variado, em que I corresponde ao número de grupos e n_i ao tamanho dos grupos, e \mathbf{b}_i um vetor aleatório de dimensão p_2 com distribuição normal de média $\mathbf{0}$ e matriz de variâncias e covariâncias $\Sigma = \Sigma(\gamma)$, em que γ é um vetor de componentes de variâncias e covariâncias desconhecidos de dimensão p_4 . Assuma que a distribuição condicional de y_{ij} dado \mathbf{b}_i pertence à família exponencial com função de

densidade dada por

$$p(y_{ij}, \theta_{ij}, \phi | \mathbf{b}_i) = \exp[\phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi)], \quad (2.4)$$

em que θ_{ij} é o parâmetro canônico, $a(\cdot)$ e $c(\cdot)$ são funções conhecidas e ϕ é o parâmetro de dispersão.

Logo, sob as condições usuais de regularidade temos que

$$E(Y_{ij} | \mathbf{b}_i) = \mu_{ij} = \dot{a}(\theta_{ij}),$$

com $\dot{a}(\theta_{ij}) = \partial a(\theta_{ij}) / \partial \theta_{ij}$ e

$$V(Y_{ij} | \mathbf{b}_i) = \frac{V(\mu_{ij})}{\phi} = \frac{\ddot{a}(\theta_{ij})}{\phi},$$

com $V(\mu_{ij}) = \ddot{a}(\theta_{ij}) = \partial^2 a(\theta_{ij}) / \partial \theta_{ij}^2$, a função de variância. Assim, os MLGM são definidos por (2.4) e pelo componente sistemático

$$g(\mu_{ij}) = \eta_{ij}, \quad (2.5)$$

em que g é a função de ligação e $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$ é o preditor linear, com $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp_1})'$ e $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ijp_2})'$ vetores de covariáveis de dimensões p_1 e p_2 , respectivamente, e $\boldsymbol{\beta}$ o vetor de coeficientes de regressão fixos de dimensão p_1 . Note que $\mu_{ij} = \dot{a}(\theta_{ij})$, o qual implica que $\theta_{ij} = \dot{a}^{-1}(\mu_{ij}) = k(\eta_{ij})$, com k uma função conhecida e diferenciável. Em seguida, com o intuito de realizar estimação por máxima verossimilhança, a função de densidade conjunta entre y_{ij} e \mathbf{b}_i pode ser expressa como

$$p(y_{ij}, \mathbf{b}_i) = p(y_{ij} | \mathbf{b}_i)p(\mathbf{b}_i)$$

e o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado \mathbf{Y}_o é dado por

$$\begin{aligned} L(\boldsymbol{\psi} | \mathbf{Y}_o) &= \sum_{i=1}^I \log \left(\int_{\mathbb{R}^{p_2}} \prod_{j=1}^{n_i} p(y_{ij} | \mathbf{b}_i) p(\mathbf{b}_i) d\mathbf{b}_i \right) \\ &= \sum_{i=1}^I \log \left\{ \int_{\mathbb{R}^{p_2}} \prod_{j=1}^{n_i} \exp[\phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi)] \right. \\ &\quad \left. \times \frac{1}{(2\pi)^{p_2/2}} \det(\boldsymbol{\Sigma})^{-1/2} \exp \left(-\frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i \right) d\mathbf{b}_i \right\}, \end{aligned} \quad (2.6)$$

em que $\boldsymbol{\psi} = (\phi, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ é o vetor de parâmetros desconhecidos de dimensão $(1 + p_1 + p_4)$, \mathbf{Y}_o é o conjunto de dados observados em Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$, e \det é a função matricial determinante. A integral múltipla dada em (2.6), geralmente, não tem solução analítica e portanto métodos de aproximação são necessários.

2.2.1 Estimação de parâmetros

Para a aproximação das integrais em (2.6) apresentamos os métodos AL e AQGH para o caso de uma variável de integração unidimensional e suas extensões para o caso multidimensional.

Método de aproximação de Laplace (AL)

O método de AL (Demidenko, 2013) tem por objetivo aproximar integrais da forma

$$\int_{\mathbb{R}} \exp(f(t)) dt, \quad (2.7)$$

com $f(t)$ uma função unimodal. A partir disso, a ideia básica consiste em aproximar $f(t)$ por uma expansão de Taylor de segunda ordem em torno de um valor \hat{t} para o qual $f(t)$ alcança seu máximo ($f'(\hat{t}) = 0$ e $f''(\hat{t}) < 0$), isto é,

$$f(t) \approx f(\hat{t}) + \frac{1}{2} f''(\hat{t})(t - \hat{t})^2. \quad (2.8)$$

Logo, substituindo (2.8) em (2.7) temos que

$$\begin{aligned} \int_{\mathbb{R}} \exp(f(t)) dt &\approx \int_{\mathbb{R}} \exp \left[f(\hat{t}) + \frac{1}{2} f''(\hat{t})(t - \hat{t})^2 \right] dt \\ &= \exp \left(f(\hat{t}) + \frac{1}{2} f''(\hat{t})\hat{t}^2 \right) \int_{\mathbb{R}} \exp \left(- f''(\hat{t})\hat{t}t + \frac{1}{2} f''(\hat{t})t^2 \right) dt, \end{aligned} \quad (2.9)$$

e pela propriedade da integral gaussiana (vide Demidenko, 2013)

$$\frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} \exp(at - bt^2) dt = \frac{1}{(2b)^{1/2}} \exp\left(\frac{a^2}{4b}\right)$$

a integral no lado direito de (2.9) fica dada por

$$\int_{\mathbb{R}} \exp \left(- f''(\hat{t})\hat{t}t + \frac{1}{2} f''(\hat{t})t^2 \right) dt = (2\pi)^{1/2} (-f''(\hat{t}))^{-1/2} \exp \left(- \frac{1}{2} f''(\hat{t})\hat{t}^2 \right).$$

Portanto,

$$\int_{\mathbb{R}} \exp(f(t)) dt \approx (2\pi)^{1/2} (-f''(\hat{t}))^{-1/2} \exp(f(\hat{t})). \quad (2.10)$$

A versão multidimensional do método para o caso de um vetor \mathbf{t} de dimensão r é dada por

$$\int_{\mathbb{R}^r} \exp(f(\mathbf{t})) d\mathbf{t} \approx (2\pi)^{r/2} \det(-f''(\hat{\mathbf{t}}))^{-1/2} \exp(f(\hat{\mathbf{t}})), \quad (2.11)$$

em que f'' denota a matriz hessiana (matriz das segundas derivadas).

Método de adaptativa quadratura de Gauss-Hermite (AQGH)

O método de AQGH é uma versão “melhorada” do método de aproximação numérica de quadratura de Gauss-Hermite (Naylor e Smith, 1982), o qual descrevemos inicialmente. Considere $u(t)$ uma função que assume valores no conjunto dos reais e $\exp(-t^2)$ uma função com conjunto imagem nos reais positivos. Então, o método de quadratura de Gauss-Hermite consiste em aproximar a integral do produto destas funções da seguinte maneira

$$\int_{\mathbb{R}} u(t) \exp(-t^2) dt \approx \sum_{k=1}^K u(t_k) \omega_k, \quad (2.12)$$

em que

$$\omega_k = \frac{2^{K-1} K! \sqrt{\pi}}{K^2 H_{k-1}(t_k)^2},$$

sendo K o número de pontos usados para a aproximação, t_k as soluções do polinômio de Hermite, isto é, $H_k(t_k) = 0$ para $H_k(t) = (-1)^k \exp(t^2) d^k/dt^k \exp(-t^2)$ e ω_k os pesos associados a $\exp(-t^2)$ (Davis e Rabinowitz, 1975). Em seguida, valores de t_k e ω_k podem ser encontrados em Abramowitz e Stegun (1972) até $k = 20$ e para $k > 20$ podem ser calculados através do algoritmo descrito por Press *et al.* (1992). Uma desvantagem de este método é que os pontos são escolhidos a partir da função $\exp(-t^2)$ independente de $u(t)$. Então, dependendo do suporte de $u(t)$, estes podem pertencer ou não a área de interesse. Por conseguinte, o método de AQGH tem por objetivo centrar e escalonar os pontos de modo que estes pertençam à maior área de $u(t)$ da seguinte maneira. Seja $g(t) = u(t) \exp(-t^2)$ e baseados no método de amostragem por importância (vide, por exemplo, Pinheiro e Bates, 1995) tendo função de importância igual à função de densidade normal de média \hat{t} e variância $\sigma_{\hat{t}}^2 = (-g''(\hat{t}))^{-1}$, a integral dada em (2.12) pode ser reescrita como

$$\int_{\mathbb{R}} \frac{u(t) \exp(-t^2)}{(2\pi\sigma_{\hat{t}}^2)^{-1/2} \exp\left[-\left(\frac{t-\hat{t}}{2^{1/2}\sigma_{\hat{t}}}\right)^2\right]} (2\pi\sigma_{\hat{t}}^2)^{-1/2} \exp\left[-\left(\frac{t-\hat{t}}{2^{1/2}\sigma_{\hat{t}}}\right)^2\right] dt, \quad (2.13)$$

e fazendo a transformação $z = \frac{t-\hat{t}}{2^{1/2}\sigma_{\hat{t}}}$, os novos pontos são $t = \hat{t} + 2^{1/2}\sigma_{\hat{t}}z$ tal que (2.13) fica igual a

$$\int_{\mathbb{R}} \frac{u(\hat{t} + 2^{1/2}\sigma_{\hat{t}}z) \exp[-(\hat{t} + 2^{1/2}\sigma_{\hat{t}}z)^2]}{2^{-1/2}\sigma_{\hat{t}}^{-1} \exp(-z^2)} \exp(-z^2) dz.$$

Logo, aplicando o método de quadratura de Gauss-Hermite temos

$$\int_{\mathbb{R}} h(z) \exp(-z^2) dz \approx \sum_{k=1}^K h(z_k) \omega_k, \quad (2.14)$$

com

$$h(z) = \frac{u(\hat{t} + 2^{1/2} \sigma_{\hat{t}} z) \exp[-(\hat{t} + 2^{1/2} \sigma_{\hat{t}} z)^2]}{2^{-1/2} \sigma_{\hat{t}}^{-1} \exp(-z^2)}.$$

Note que quando $K = 1$ a integral dada em (2.14) é dada por

$$\int_{\mathbb{R}} h(z) \exp(-z^2) dz \approx (2\pi)^{1/2} \sigma_{\hat{z}} h(\hat{z}) \exp(-\hat{z}^2), \quad (2.15)$$

que corresponde ao método de AL com $f(\hat{z}) = \log[h(\hat{z}) \exp(-\hat{z}^2)]$. A aproximação fornecida por este método se torna cada vez mais precisa à medida que aumenta o número de pontos. No entanto, com um maior número de pontos, maior é o custo computacional. Na prática, com $K \leq 7$, geralmente é suficiente e com $K = 1$ frequentemente obtém-se uma aproximação razoável.

A versão multidimensional de (2.14) é dada por

$$\int_{\mathbb{R}} h(\mathbf{z}) \exp(-\mathbf{z}'\mathbf{z}) d\mathbf{z} \approx \sum_{k_1=1}^K \dots \sum_{k_r=1}^K h(\mathbf{z}_k) \omega_k^*,$$

em que

$$h(\mathbf{z}) = \frac{u(\hat{\mathbf{t}} + 2^{1/2} \Sigma_{\hat{\mathbf{t}}}^{1/2} \mathbf{z}_k) \exp[-(\hat{\mathbf{t}} + 2^{1/2} \Sigma_{\hat{\mathbf{t}}}^{1/2} \mathbf{z}_k)'(\hat{\mathbf{t}} + 2^{1/2} \Sigma_{\hat{\mathbf{t}}}^{1/2} \mathbf{z}_k)]}{2^{-1/2} \det(\Sigma_{\hat{\mathbf{t}}})^{-1/2} \exp(\mathbf{z}'_k \mathbf{z}_k)},$$

com $\mathbf{z}_k = (z_{k_1}, \dots, z_{k_r})$, $\Sigma_{\hat{\mathbf{t}}} = (-g''(\hat{\mathbf{t}}))^{-1}$ e $\omega_k^* = \prod_{j=1}^r \omega_{k_j}$. Por último, quando $K = 1$ tem-se a aproximação de Laplace dada em (2.11).

2.3 Análise de diagnóstico de influência local

Análise de diagnóstico é uma etapa muito importante na modelagem estatística pois permite (i) verificar possíveis afastamentos das suposições estabelecidas sobre o modelo e (ii) detectar a existência de observações influentes com alguma interferência desproporcional nas estimativas dos parâmetros do modelo. Assim, para (i) uma técnica comumente utilizada corresponde à análise de resíduos que foi abordada por muitos autores, por exemplo, por [Atkinson \(1981, 1982\)](#); [Belsley et al. \(1980\)](#); [Chatterjee e Hadi \(1986\)](#); [Cook e Weisberg \(1982\)](#), para o modelo normal linear, por [Cordeiro \(2004\)](#); [Paula \(1995\)](#); [Pierce e Schafer \(1986\)](#); [Pregibon \(1981\)](#); [Williams \(1987\)](#) para os MLG e por [Langford e Lewis \(1998\)](#); [Zhu e Lee \(2003\)](#) para os MLGM. Em relação

a (ii) tem-se a técnica da matriz de alavanca proposta por Hoaglin e Welsh (1978) para o modelo normal linear e abordada, por exemplo, por Paula (1999) para os MLG. Por outro lado, também tem-se a técnica de eliminação de casos apresentada por Cook (1977) para o modelo normal linear e abordada, por exemplo, por Williams (1987) para os MLG e por Xiang *et al.* (2002); Xu *et al.* (2006) para os MLGM.

Contudo, uma das metodologias mais inovadoras e com grande receptividade entre os pesquisadores foi a técnica de influência local proposta por Cook (1986) para o modelo normal linear e aplicada, por exemplo, por Thomas e Cook (1989, 1990) para os MLG e estendida por Ouwens *et al.* (2001); Xiang *et al.* (2003); Zhu e Lee (2003) para os MLGM. Em particular, Ouwens *et al.* (2001); Xiang *et al.* (2002) desenvolveram esta técnica sob a perspectiva da contribuição à verossimilhança dos grupos e Zhu e Lee (2003) sob a abordagem da contribuição das observações. Por último, existem muitas outras aplicações desta técnica a outros modelos como, por exemplo, os trabalhos de Leiva *et al.* (2007); Russo *et al.* (2009), entre outros. A seguir apresentamos a técnica de influência local proposta por Cook (1986) e o enfoque para os MLGM proposto por Zhu e Lee (2003).

Seja $L(\boldsymbol{\psi})$ o logaritmo da função de verossimilhança do modelo postulado e $\hat{\boldsymbol{\psi}}$ o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\psi}$ sob $L(\boldsymbol{\psi})$. Seja uma modificação denominada vetor de perturbação, isto é, $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$ denotando por $L(\boldsymbol{\psi}, \boldsymbol{\omega})$ o logaritmo da função de verossimilhança do modelo perturbado e $\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$ o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\psi}$ sob $L(\boldsymbol{\psi}, \boldsymbol{\omega})$. Para avaliar a influência de $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$ Cook (1986) propõe a medida de influência denominada afastamento da verossimilhança definida por

$$LD(\boldsymbol{\omega}) = 2(L(\hat{\boldsymbol{\psi}}) - L(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}})), \quad (2.16)$$

em que $L(\hat{\boldsymbol{\psi}})$ e $L(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}})$ correspondem ao logaritmo da função de verossimilhança do modelo postulado avaliado em $\hat{\boldsymbol{\psi}}$ e $\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$, respectivamente. Observe que $LD(\boldsymbol{\omega}) \geq 0$, para todo $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$. Desta forma, um gráfico de $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contém informação essencial da influência do tipo de perturbação utilizado. Na literatura estatística este gráfico é denominado gráfico de influência e é representado pela superfície geométrica dada por $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}', LD(\boldsymbol{\omega}))'$, com $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$. Visto que analisar $\alpha(\boldsymbol{\omega})$ para todo $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$ é inviável, o autor a examina ao redor de um vetor de não perturbação $\boldsymbol{\omega}_0 \in \Omega \subset \mathbb{R}^q$ tal que $L(\boldsymbol{\psi}, \boldsymbol{\omega}_0) = L(\boldsymbol{\psi})$, para todo $\boldsymbol{\psi}$, da maneira descrita a seguir. Utilizando conceitos de geometria diferencial (vide Stoker, 1956) a técnica consiste em descrever o comportamento da superfície $\alpha(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$ mediante a análise da curvatura de curvas de $\alpha(\boldsymbol{\omega})$ que passam

por $\alpha(\boldsymbol{\omega}_0)$. Para $q = 2$, estas curvas correspondem às seções normais as quais estão formadas pela interseção entre $\alpha(\boldsymbol{\omega})$ e os denominados planos normais que são ortogonais ao plano tangente T_0 em $\boldsymbol{\omega}_0$. Para auxiliar o entendimento vide Figura 2.1. As curvaturas das seções normais são denominadas curvaturas normais e são utilizadas para descrever o comportamento do gráfico de influência em torno de $\boldsymbol{\omega}_0$. Nesse sentido, como $LD(\boldsymbol{\omega})$ atinge um mínimo local em $\boldsymbol{\omega}_0$, ou seja, $LD(\boldsymbol{\omega}_0) = 0$, é possível definir uma seção normal mediante uma linha reta em $\Omega \subset \mathbb{R}^q$ que passa por $\boldsymbol{\omega}_0$, isto é, $\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\mathbf{h}$, com $a \in \mathbb{R}$ e \mathbf{h} um vetor unitário diferente de zero em \mathbb{R}^q . Em consequência, a interseção entre a seção normal e o plano tangente gera a linha projetada na superfície $\alpha(\boldsymbol{\omega})$ passando por $\boldsymbol{\omega}_0$. Desse modo, cada direção \mathbf{h} especifica uma linha projetada, que por sua vez corresponde a uma seção normal. Em seguida, dada esta relação entre a linha projetada e a seção normal, o procedimento consiste em analisar a curvatura da linha projetada em um gráfico de $LD(\boldsymbol{\omega}_0 + a\mathbf{h})$ versus a , com $a \in \mathbb{R}$. Por exemplo, na Figura 2.2 ilustramos o comportamento de duas linhas projetadas nas direções \mathbf{h}_1 e \mathbf{h}_2 , isto é, $LD(\boldsymbol{\omega}_0 + a\mathbf{h}_1)$ e $LD(\boldsymbol{\omega}_0 + a\mathbf{h}_2)$. Logo, uma vez que $LD(\boldsymbol{\omega}_0) = 0$, temos que $LD(\boldsymbol{\omega}_0 + a\mathbf{h})$ atinge um mínimo local em $a = 0$, e a direção \mathbf{h}_2 é a que produz os maiores afastamentos da verossimilhança à medida que os valores de a se distanciam de 0 ou, equivalentemente, $LD(\boldsymbol{\omega}_0 + a\mathbf{h}_2)$ é a linha projetada de maior curvatura. Cook (1986)

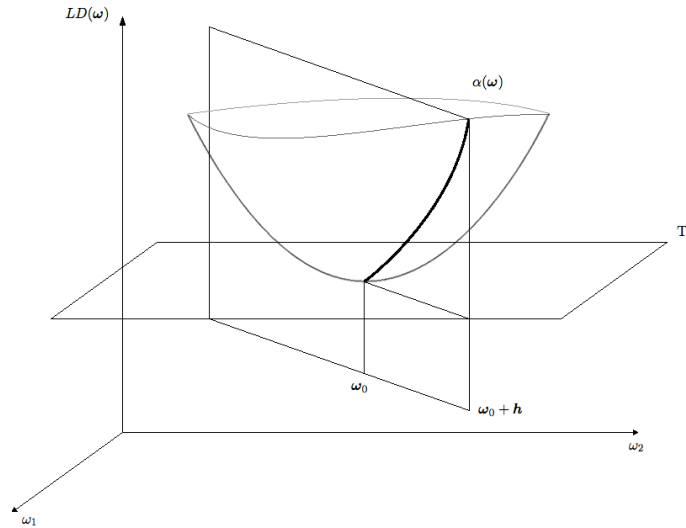


Figura 2.1: Gráfico de influência.

mostra que a curvatura normal da linha projetada na direção \mathbf{h} é dada por

$$C_{\mathbf{h}} = |\ddot{LD}(\boldsymbol{\omega}(a))|_{a=0}, \quad (2.17)$$

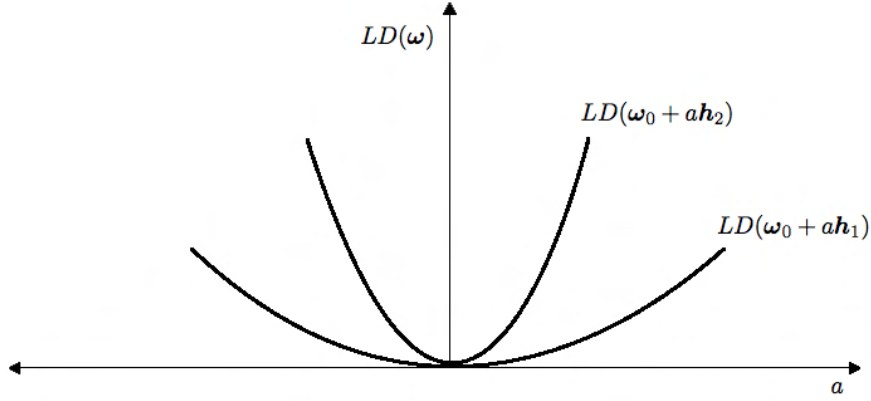


Figura 2.2: Gráfico da linha projetada.

em que $\ddot{LD}(\omega(a)) = \partial^2 LD(\omega(a))/\partial a^2$ e $|\cdot|$ denota o valor absoluto. Desenvolvendo as derivadas envolvidas em (2.17), a curvatura normal é igual a

$$C_{\mathbf{h}} = 2|\mathbf{h}'\ddot{L}_{\omega_0}\mathbf{h}|, \quad (2.18)$$

em que

$$\ddot{L}_{\omega_0} = \left. \frac{\partial^2 L(\hat{\psi}_{\omega})}{\partial \omega \partial \omega'} \right|_{\omega=\omega_0}. \quad (2.19)$$

Assim, $C_{\mathbf{h}}$ corresponde à curvatura normal da superfície $\alpha(\omega)$ em ω_0 na direção \mathbf{h} . Agora, para que a expressão dada em (2.18) seja simple de calcular, [Cook \(1986\)](#) mostra também que (2.19) pode ser expressa como

$$-\ddot{L}_{\omega_0} = \mathbf{\Delta}'(-\ddot{\mathbf{L}})^{-1}\mathbf{\Delta},$$

em que $-\ddot{\mathbf{L}}$ é a matriz de informação observada de Fisher, simétrica e positiva definida, de dimensão $(1 + p_1 + p_4) \times (1 + p_1 + p_4)$ dada por

$$-\ddot{\mathbf{L}} = -\left. \frac{\partial^2 L(\psi)}{\partial \psi \partial \psi'} \right|_{\psi=\hat{\psi}}$$

e $\mathbf{\Delta}$ é a matriz de perturbação de dimensão $(1 + p_1 + p_4) \times q$ dada por

$$\mathbf{\Delta} = \left. \frac{\partial^2 L(\psi, \omega)}{\partial \psi \partial \omega'} \right|_{\psi=\hat{\psi}, \omega=\omega_0}.$$

Assim, a curvatura normal da linha projetada na direção \mathbf{h} é igual a

$$C_{\mathbf{h}} = 2|\mathbf{h}'\mathbf{\Delta}'\ddot{\mathbf{L}}^{-1}\mathbf{\Delta}\mathbf{h}|. \quad (2.20)$$

Em conclusão, a expressão dada em (2.20) é denominada influência local da perturbação na direção \mathbf{h} sobre a estimativa de $\boldsymbol{\psi}$. Existem várias formas em que (2.20) pode ser estudada, cada uma correspondente à escolha do vetor \mathbf{h} . Cook (1986) sugere considerar a direção \mathbf{h}_{\max} que produz a linha projetada de maior curvatura normal $C_{\mathbf{h}_{\max}}$ com o intuito de obter a máxima mudança local em $LD(\boldsymbol{\omega}_0) = 0$. Especificamente, a curvatura normal $C_{\mathbf{h}_{\max}}$ equivale ao maior autovalor da matriz $\boldsymbol{\Delta}'(-\ddot{\mathbf{L}})^{-1}\boldsymbol{\Delta}$ e a direção \mathbf{h}_{\max} corresponde ao autovetor associado a esse maior autovalor. Assim, um gráfico de $|\mathbf{h}_{\max}|$ versus o índice das observações pode revelar quais pontos exercem uma influência desproporcional em torno de $LD(\boldsymbol{\omega}_0)$. Cabe mencionar que a análise da direção \mathbf{h}_{\max} pode revelar pontos que são simultaneamente influentes.

2.4 Análise de diagnóstico de influência local nos modelos lineares generalizados mistos

A técnica de análise de diagnóstico de influência local proposta por Cook (1986) tem sido aplicada a muitos modelos estatísticos. No entanto, para os MLGM, o logaritmo da função de verossimilhança envolve integrais de grande dimensão e que não têm solução analítica, pelo que aplicar diretamente esta técnica não é fácil. Por conseguinte, Zhu e Lee (2003) propõem abordar esta dificuldade baseados no trabalho de influência local para modelos com dados faltantes de Zhu e Lee (2001) da seguinte maneira. Considerando os efeitos aleatórios como dados faltantes, isto é, $\mathbf{Y}_f = \{\mathbf{b}_i : i = 1, \dots, I\}$, defina por $\mathbf{Y}_c = (\mathbf{Y}_o, \mathbf{Y}_f)$ o conjunto de dados completos. Logo, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$\begin{aligned} L(\boldsymbol{\psi}|\mathbf{Y}_c) &= \log \prod_{i=1}^I \prod_{j=1}^{n_i} p(y_{ij}, \mathbf{b}_i) \\ &= \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi) - \frac{1}{2}\mathbf{b}_i'\boldsymbol{\Sigma}^{-1}\mathbf{b}_i - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) \right], \end{aligned} \quad (2.21)$$

que é uma expressão relativamente simples para desenvolver análise de diagnóstico de influência local com base na esperança condicional de $L(\boldsymbol{\psi}|\mathbf{Y}_c)$. Deste modo, de maneira similar a (2.16), Zhu e Lee (2003) utilizam a medida de influência denominada função Q-afastamento dada por

$$f_Q(\boldsymbol{\omega}) = 2(Q(\hat{\boldsymbol{\psi}}|\hat{\boldsymbol{\psi}}) - Q(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}|\hat{\boldsymbol{\psi}})),$$

em que $\hat{\boldsymbol{\psi}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\psi}$ que maximiza

$$Q(\boldsymbol{\psi}|\hat{\boldsymbol{\psi}}) = E[L(\boldsymbol{\psi}|\mathbf{Y}_c)|\mathbf{Y}_o, \hat{\boldsymbol{\psi}}]$$

e $\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\psi}$ que maximiza

$$Q(\boldsymbol{\psi}, \boldsymbol{\omega}|\hat{\boldsymbol{\psi}}) = E[L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{Y}_o, \hat{\boldsymbol{\psi}}],$$

e cujas esperanças são calculadas em relação à função de densidade condicional $p(\mathbf{Y}_f|\mathbf{Y}_o, \hat{\boldsymbol{\psi}})$. Além disso, define-se um vetor de não perturbação $\boldsymbol{\omega}_0 \in \Omega \subset \mathbb{R}^q$ tal que $Q(\hat{\boldsymbol{\psi}}|\hat{\boldsymbol{\psi}}) = Q(\hat{\boldsymbol{\psi}}, \boldsymbol{\omega}_0|\hat{\boldsymbol{\psi}})$ e, conseqüentemente, a função $f_Q(\boldsymbol{\omega})$ atinge um mínimo global em $\boldsymbol{\omega}_0$, ou seja, $f_Q(\boldsymbol{\omega}_0) = 0$. Logo, analogamente a Cook (1986), a curvatura normal na direção \mathbf{h} é dada por

$$C_{f_Q, \mathbf{h}} = -2\mathbf{h}'\ddot{\mathbf{Q}}_{\boldsymbol{\omega}_0}\mathbf{h},$$

em que

$$\ddot{\mathbf{Q}}_{\boldsymbol{\omega}_0} = \left. \frac{\partial^2 Q(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}|\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}'} \right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (2.22)$$

Neste caso, mostra-se também que (2.22) pode ser expressa como

$$-\ddot{\mathbf{Q}}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}'_{\boldsymbol{\omega}_0} (-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}}))^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}, \quad (2.23)$$

em que $-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$ é uma matriz simétrica e semipositiva definida de dimensão $(1 + p_1 + p_4) \times (1 + p_1 + p_4)$ dada por

$$-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}}) = - \left. \frac{\partial^2 Q(\boldsymbol{\psi}|\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} \quad (2.24)$$

e $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ é a matriz de influência de dimensão $(1 + p_1 + p_4) \times q$ dada por

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \left. \frac{\partial^2 Q(\boldsymbol{\psi}, \boldsymbol{\omega}|\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\omega}'} \right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (2.25)$$

Assim, a curvatura normal na direção \mathbf{h} é igual a

$$C_{f_Q, \mathbf{h}} = 2\mathbf{h}'\boldsymbol{\Delta}'_{\boldsymbol{\omega}_0} (-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}}))^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0} \mathbf{h}. \quad (2.26)$$

De forma diferente a Cook (1986), Zhu e Lee (2003) estudam (2.26) na direção da i -ésima observação. Logo, uma vez que a matriz $-\ddot{\mathbf{Q}}_{\boldsymbol{\omega}_0} = 2\boldsymbol{\Delta}'_{\boldsymbol{\omega}_0} (-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}}))^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ pode ser expressa em termos

de sua decomposição espectral, ou seja,

$$-2\ddot{\mathbf{Q}}_{\omega_0} = \sum_{m=1}^M \lambda_m \mathbf{e}'_m \mathbf{e}_m,$$

em que $(\lambda_1, \mathbf{e}_1), \dots, (\lambda_M, \mathbf{e}_M)$ são pares de autovalores e autovetores tal que $\lambda_1 \geq \dots \geq \lambda_M > \lambda_{M+1} = \dots = \lambda_q = 0$ e $(\mathbf{e}_1, \dots, \mathbf{e}_M)$ é uma base ortonormal de \mathbb{R}^M , com $M = 1 + p_1 + p_4$, é possível definir a curvatura normal na direção da i -ésima observação ou curvatura normal individual por

$$C_{f_Q, \mathbf{h}_i} = \sum_{m=1}^M \lambda_m e_{m_i}^2,$$

em que e_{m_i} é a i -ésima componente de \mathbf{e}_m e \mathbf{h}_i é um vetor de dimensão q com a i -ésima componente igual a um e as restantes a zero. Por último, com o intuito de avaliar a magnitude das C_{f_Q, \mathbf{h}_i} [Zhu e Lee \(2003\)](#) estabelecem alguns critérios para identificar observações influentes através do seguinte teorema.

Teorema 2.4.1. *Seja $\hat{\boldsymbol{\psi}}$ o estimador de máxima verossimilhança de $\boldsymbol{\psi}$ e suponha $k(n)$ uma função crescente de n . Assuma que $-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})/k(n) \xrightarrow{P} \mathbf{I}_c$ e $\boldsymbol{\Delta}_{\omega_0} \boldsymbol{\Delta}'_{\omega_0}/k(n) \xrightarrow{P} \mathbf{B}$, com \mathbf{I}_c a matriz de informação de Fisher dos dados completos e \mathbf{B} uma matriz não negativa. Logo,*

$$\sum_{i=1}^q C_{f_Q, \mathbf{h}_i} = \text{tr}(2\ddot{\mathbf{Q}}_{\omega_0}) \xrightarrow{P} 2\text{tr}(\mathbf{I}_c^{-1} \mathbf{B}) = \lambda_0.$$

A implicação do Teorema 2.4.1 é que as curvaturas normais individuais C_{f_Q, \mathbf{h}_i} estão ao redor de λ_0/q . Na prática a maioria das C_{f_Q, \mathbf{h}_i} pertencem ao intervalo $[0, 3\lambda_0/q]$ e poucas são maiores que $3\lambda_0/q$. Do modo que, para quantificar que tão influentes são as observações [Zhu e Lee \(2003\)](#) introduzem a seguinte definição para k_0 valores inteiros positivos.

Definição 2.4.1. Se C_{f_Q, \mathbf{h}_i} pertence ao intervalo $[(k_0 - 1)\lambda_0/q, k_0\lambda_0/q)$, a i -ésima observação é denominada ponto influente de k_0 -ésima ordem.

Na prática, em geral são de interesse as observações que são pontos influentes de quarta ordem ou maior. Note que um gráfico de C_{f_Q, \mathbf{h}_i} versus o índice das observações permite identificar quais pontos são influentes. Cabe mencionar que a curvatura normal individual pode revelar pontos influentes que não são detectados pela análise do vetor \mathbf{h}_{\max} . Isto é, C_{f_Q, \mathbf{h}_i} pode determinar que um ponto seja influente sem que o i -ésimo componente de \mathbf{h}_{\max} indentifique tal ponto como influente.

2.4.1 Derivação da curvatura normal

Dado que as condições de regularidade permitem o intercâmbio da ordem entre a integração e a diferenciação, baseados em (2.21) as derivadas compreendidas na matriz $-\ddot{\mathbf{Q}}_{\psi}(\hat{\psi})$ são dadas por

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \phi^2} = -\sum_{i=1}^I \sum_{j=1}^{n_i} \frac{\partial^2 c(y_{ij}, \phi)}{\partial \phi^2},$$

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \phi \partial \beta'} = -\sum_{i=1}^I \sum_{j=1}^{n_i} \dot{r}_{ij} \mathbf{x}'_{ij},$$

com $\dot{r}_{ij} = (y_{ij} - \mu_{ij}) \dot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)$,

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \phi \partial \gamma'} = \mathbf{0},$$

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \beta \partial \beta'} = \phi \sum_{i=1}^I \sum_{j=1}^{n_i} \ddot{r}_{ij} \mathbf{x}_{ij} \mathbf{x}'_{ij},$$

com $\ddot{r}_{ij} = \ddot{a}(\theta_{ij}) (\dot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i))^2 - (y_{ij} - \mu_{ij}) \ddot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)$,

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \beta \partial \gamma'} = \mathbf{0}$$

e

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \gamma \partial \gamma'} = -\frac{I}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \right) \otimes \boldsymbol{\Sigma}^{-1}, \quad (2.27)$$

em que \otimes denota a função produto de Kronecker.

2.4.2 Esquemas de perturbação

Como a técnica da análise de diagnóstico de influência local consiste em estudar o comportamento de alguma medida de influência sob pequenas perturbações no modelo ou dados, [Zhu e Lee \(2003\)](#) propõem os esquemas de perturbação descritos a seguir.

Esquema 1: Perturbação da ponderação das observações

Se o objetivo é identificar observações influentes entre todas as observações, o esquema de perturbação consiste em atribuir uma ponderação à contribuição para a verossimilhança de cada uma delas. Deste modo, o logaritmo da função de verossimilhança de ψ dado os dados completos é

dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} \omega_{ij} [\phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi)] - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right\}. \quad (2.28)$$

Assim, de (2.28), o vetor de não perturbação é um vetor de uns de dimensão $q \times 1$, isto é, $\boldsymbol{\omega}_0 = \mathbf{1}_q$, com $q = \sum_{i=1}^I n_i$. Logo, mediante diferenciação tem-se

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \omega_{ij}} = y_{ij}\theta_{ij} - a(\theta_{ij}) + \frac{\partial c(y_{ij}, \phi)}{\partial \phi},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = \phi \dot{r}_{ij} \mathbf{x}_{ij},$$

com $\dot{r}_{ij} = (y_{ij} - \mu_{ij}) \dot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)$, e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma_l \partial \omega_{ij}} = 0, \quad l = 1, \dots, p_4.$$

Esquema 2: Perturbação da ponderação dos grupos

Se o objetivo é identificar grupos influentes o esquema de perturbação consiste em atribuir uma ponderação à soma das contribuições das log-verossimilhanças das observações pertencentes ao grupo. Deste modo, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \omega_i \left[\sum_{j=1}^{n_i} \phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi) \right] - \sum_{i=1}^I \left[\frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (2.29)$$

Assim, de (2.29), tem-se que $\omega_{i0} = 1$, para $i = 1, \dots, I$, e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \omega_i} = \sum_{j=1}^{n_i} (y_{ij}\theta_{ij} - a(\theta_{ij})) + \frac{\partial c(y_{ij}, \phi)}{\partial \phi},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} = \sum_{j=1}^{n_i} \phi \dot{r}_{ij} \mathbf{x}_{ij}$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma_l \partial \omega_i} = 0, \quad l = 1, \dots, p_4.$$

Esquema 3: Perturbação da matriz de variâncias e covariâncias dos efeitos aleatórios

Um importante suposto dos MLGM é que o vetor aleatório \mathbf{b}_i tem distribuição normal de média $\mathbf{0}$ e matriz de variâncias e covariâncias $\Sigma = \Sigma(\boldsymbol{\gamma})$, com $\boldsymbol{\gamma}$ um vetor de componentes de variâncias e covariâncias desconhecidos de dimensão p_4 , para $i = 1, \dots, I$. Logo, para estudar o efeito do afastamento da hipótese sobre a matriz de variâncias e covariâncias, o esquema de perturbação é dado por $\omega_i^{-1}\Sigma(\boldsymbol{\gamma})$, para $i = 1, \dots, I$, de modo que, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} [\phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi)] - \omega_i \frac{1}{2} \mathbf{b}_i' \Sigma^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\Sigma)) + \frac{1}{2} \log \omega_i \right\}. \quad (2.30)$$

Assim, de (2.30), tem-se que $\omega_{i0} = 1$, para $i = 1, \dots, I$, e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \omega_i} = 0,$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} = \mathbf{0}$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma}_l \partial \omega_i} = \frac{1}{2} \mathbf{b}_i' \Sigma^{-1} \dot{\Sigma}(l) \Sigma^{-1} \mathbf{b}_i,$$

com $\dot{\Sigma}(l) = (\partial \Sigma / \partial \gamma_l)$, para $l = 1, \dots, p_4$.

Esquema 4: Perturbação simultânea da ponderação dos grupos e efeitos aleatórios

Se o objetivo é identificar a influência de mudanças simultâneas na ponderação dos grupos e os efeitos aleatórios. Então, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \omega_i \left[\sum_{j=1}^{n_i} \phi(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi) - \frac{1}{2} \mathbf{b}_i' \Sigma^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\Sigma)) \right]. \quad (2.31)$$

Assim, de (2.31), tem-se que $\omega_{i0} = 1$, para $i = 1, \dots, I$, e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \omega_i} = \sum_{j=1}^{n_i} y_{ij} \theta_{ij} - a(\theta_{ij}) + \frac{\partial c(y_{ij}, \phi)}{\partial \phi},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} = \sum_{j=1}^{n_i} \phi \dot{r}_{ij} \mathbf{x}_{ij}$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma_l \partial \omega_i} = \frac{1}{2} \text{tr} [\boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\Sigma}}(l) \boldsymbol{\Sigma}^{-1} (\mathbf{b}_i \mathbf{b}_i' - \boldsymbol{\Sigma})], \quad l = 1, \dots, p_4.$$

Esquema 5: Perturbação das covariáveis no caso contínuo

Com o objetivo de perturbar o vetor \mathbf{x}_{ij} de covariáveis contínuas da matriz de delineamento \mathbf{X} do modelo, considere a matriz $\mathbf{C} = \text{diag}(c_1, \dots, c_{p_1})$ de fatores de escala e um vetor de perturbação $\boldsymbol{\omega}_{ij}$ de dimensão $p_1 \times 1$ tal que \mathbf{x}_{ij} é substituído por

$$\mathbf{x}_{ij}(\boldsymbol{\omega}_{ij}) = \mathbf{x}_{ij} + \mathbf{C} \boldsymbol{\omega}_{ij}. \quad (2.32)$$

Então, ao substituir (2.32) em θ_{ij} de (2.21), o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \phi(y_{ij} \theta_{ij}(\boldsymbol{\omega}_{ij}) - a(\theta_{ij}(\boldsymbol{\omega}_{ij}))) + c(y_{ij}, \phi) - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (2.33)$$

Logo,

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \boldsymbol{\omega}_{ij}} = \dot{r}_{ij} \mathbf{C} \boldsymbol{\beta},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_{ij}} = \phi \dot{r}_{ij} \mathbf{C} - \phi \ddot{r}_{ij} \mathbf{x}_{ij} \mathbf{C} \boldsymbol{\beta}',$$

com $\ddot{r}_{ij} = \ddot{a}(\theta_{ij}) (\dot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i))^2 - (y_{ij} - \mu_{ij}) \ddot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)$, e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma_l \partial \boldsymbol{\omega}_{ij}} = \mathbf{0}, \quad l = 1, \dots, p_4.$$

Esquema 6: Perturbação da variável resposta no caso contínuo

Com o objetivo de perturbar a variável resposta no caso contínuo, [Zhu e Lee \(2003\)](#) perturbam aditivamente y_{ij} , isto é, $y_{ij}(\boldsymbol{\omega}_{ij}) = y_{ij} + \boldsymbol{\omega}_{ij}$. Deste modo, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \phi(y_{ij}(\boldsymbol{\omega}_{ij}) \theta_{ij} - a(\theta_{ij})) + c(y_{ij}(\boldsymbol{\omega}_{ij}), \phi) - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right].$$

Assim,

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \phi \partial \boldsymbol{\omega}_{ij}} = \theta_{ij} + \frac{\partial^2 c(y_{ij}(\boldsymbol{\omega}_{ij}), \phi)}{\partial \phi \partial \boldsymbol{\omega}_{ij}},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = \phi \dot{k}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij}$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma_l \partial \omega_{ij}} = 0, \quad l = 1, \dots, p_4.$$

2.4.3 Aproximação das matrizes da curvatura normal

A esperança condicional que se encontra nos blocos das matrizes $-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ e $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$, em geral, envolve integrais intratáveis e de grande dimensão. Deste modo, torna-se necessária a realização de uma estratégia para sua obtenção. [Zhu e Lee \(2003\)](#) propõem amostrar $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ da função de densidade dada por

$$p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi}) \propto \exp \left[-\frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i + \sum_{j=1}^{n_i} \phi(y_{ij} k(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) - a(k(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i))) \right] \quad (2.34)$$

e aproximar estas matrizes por

$$-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}}) \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^S \left. \frac{\partial^2 L(\boldsymbol{\psi} | \mathbf{Y}_o, \mathbf{Y}_f^{(s)})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}} \quad (2.35)$$

e

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^S \left. \frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_o, \mathbf{Y}_f^{(s)})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\omega}'} \right|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}, \quad (2.36)$$

respectivamente, com $\mathbf{Y}_f^{(s)} = (\mathbf{b}_1^{(s)}, \dots, \mathbf{b}_I^{(s)})$ e M_0 as primeiras, por exemplo, 1000 observações. Agora, como a função de densidade dada em (2.34) é uma função complexa, [Zhu e Lee \(2003\)](#) utilizam o algoritmo Metropolis-Hastings ([Robert e Casella, 2004](#)) para realizar a amostragem da forma descrita a seguir.

O algoritmo é iniciado a partir de um valor arbitrário $\mathbf{b}_i^{(0)}$. Em seguida, na r -ésima iteração se realizam os seguintes passos:

- (1) Dado o valor atual de $\mathbf{b}_i^{(r-1)}$, gere um novo candidato como $\mathbf{b}_i \sim N_{p_2}(\mathbf{b}_i^{(r-1)}, \Gamma(\mathbf{0}))$, em que $\Gamma(\mathbf{0})$ é descrito com em [Xu et al. \(2006\)](#), isto é,

$$\Gamma(\mathbf{0}) = \Gamma(\mathbf{b}_i) |_{\mathbf{b}_i=\mathbf{0}} = \left(\boldsymbol{\Sigma}^{-1} + \phi \sum_{j=1}^{n_i} \ddot{a}(k(\eta_{ij})) \dot{k}^2(\eta_{ij}) \mathbf{z}_{ij} \mathbf{z}'_{ij} \right)^{-1} \Big|_{\mathbf{b}_i=\mathbf{0}}. \quad (2.37)$$

(2) Gere $u \sim U(0, 1)$. Se $u \leq \alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i)$ faça $\mathbf{b}_i^{(r)} = \mathbf{b}_i$, senão considere $\mathbf{b}_i^{(r)} = \mathbf{b}_i^{(r-1)}$, em que

$$\alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i) = \min \left(\frac{p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi})}{p(\mathbf{b}_i^{(r-1)} | \mathbf{Y}_o, \boldsymbol{\psi})}, 1 \right)$$

é a probabilidade de aceitar um novo candidato.

(3) Repita os passos (1) e (2) para $r + 1$.

Baseados nas observações amostradas, [Zhu e Lee \(2003\)](#) afirmam que

$$\hat{\mathbf{b}}_i = \frac{1}{S - M_0} \sum_{s=M_0+1}^S \mathbf{b}_i^{(s)},$$

é uma predição consistente de \mathbf{b}_i , para $i = 1, \dots, I$.

2.5 Esquema de perturbação apropriada

Como já foi mencionado, a identificação de observações influentes é uma etapa muito importante em análise de dados. No obstante, perturbar arbitrariamente o modelo ou os dados pode levar a conclusões pouco confiáveis com respeito à influência. Por exemplo, considerar tamanhos de grupos desbalanceados no esquema de perturbação dos grupos pode levar à identificação equívoca de grupos influentes entre todos os grupos. Dessa forma, os autores [Chen et al. \(2010\)](#) propõem a seguinte metodologia para selecionar um esquema de perturbação que seja apropriado.

Seja a função de densidade sob $L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)$ dada por $p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})$ e um vetor de não perturbação $\boldsymbol{\omega}_0 \in \Omega \subset \mathbb{R}^q$ tal que $p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega}_0) = p(\mathbf{Y}_c | \boldsymbol{\psi})$, para todo $\boldsymbol{\psi}$. Seja o vetor escore de $\boldsymbol{\omega}$ dado por $\mathbf{U}(\boldsymbol{\omega})$ com componentes

$$u_k(\boldsymbol{\omega}) = \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_k},$$

para $k = 1, \dots, q$, e a matriz de informação esperada de Fisher de $\boldsymbol{\omega}$ dada por $\mathbf{G}(\boldsymbol{\omega})$ com componentes

$$g_{kk'}(\boldsymbol{\omega}) = \mathbb{E} \left(\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_k} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_{k'}} \middle| \boldsymbol{\omega} \right), \quad (2.38)$$

para $k, k' = 1, \dots, q$, em que a esperança é calculada com respeito a $p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})$. Então, uma vez que $\mathbb{E}(\mathbf{U}(\boldsymbol{\omega})) = \mathbf{0}$ (vide, por exemplo, [Sen et al., 2010](#)), os componentes da diagonal de $\mathbf{G}(\boldsymbol{\omega})$ correspondem às variâncias dos componentes do vetor escore e estes indicam a quantidade de perturbação gerada pelos diferentes elementos de $\boldsymbol{\omega}$. Por outro lado, os componentes fora da diagonal correspondem às covariâncias dos componentes do vetor escore e estes indicam a associação entre

os diferentes elementos de $\boldsymbol{\omega}$. Desta forma, uma perturbação é apropriada se a matriz $\mathbf{G}(\boldsymbol{\omega})$ satisfaz as condições abaixo:

- (1) $\mathbf{G}(\boldsymbol{\omega})$ é de posto completo numa vizinhança de $\boldsymbol{\omega}_0$ com propósito de evitar qualquer componente redundante de $\boldsymbol{\omega}$.
- (2) Os componentes fora da diagonal devem ser tão pequenos quanto possível com o propósito de evitar uma forte associação entre eles e, em consequência, perturbações com efeitos ambíguos.
- (3) As diferenças entre os componentes da diagonal devem ser tão pequenas quanto possível tal que as quantidades geradas pelos elementos de $\boldsymbol{\omega}$ sejam semelhantes.

Uma matriz que satisfaz as condições anteriores é $\mathbf{G}(\boldsymbol{\omega})$ tal que $\mathbf{G}(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = a\mathbf{I}_q$, com $a > 0$ e \mathbf{I}_q a matriz identidade de dimensão q . Logo, para encontrar uma matriz que satisfaça (1), (2) e (3) pode-se escolher um novo vetor de perturbação $\boldsymbol{\omega}^*$ (vide [Chen et al., 2010](#)) dado por

$$\boldsymbol{\omega}^* = \boldsymbol{\omega}_0 + \mathbf{G}(\boldsymbol{\omega}_0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}_0), \quad (2.39)$$

tal que $\mathbf{G}(\boldsymbol{\omega}^*)$ avaliada em $\boldsymbol{\omega}_0$ seja igual a $a\mathbf{I}_q$. Este novo vetor de perturbação tem a característica de que seus elementos são ortonormais. Isto é, os elementos são não associados, ao menos localmente, e introduzem iguais quantidades de perturbação.

Geralmente, a esperança dada em (2.38) não tem solução analítica. Desta forma, [Chen et al. \(2010\)](#) utilizam o método clássico de integração de Monte Carlo (vide, por exemplo, [Robert e Casella, 2004](#)) para sua aproximação, isto é,

$$g_{kk'}(\boldsymbol{\omega}) \approx \frac{1}{T} \sum_{t=1}^T \frac{\partial \log p(\mathbf{Y}_o, \mathbf{b}_i^{(t)} | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_k} \frac{\partial \log p(\mathbf{Y}_o, \mathbf{b}_i^{(t)} | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_{k'}}, \quad (2.40)$$

em que $\{\mathbf{b}_i^{(t)}; t = 1, \dots, T\}$ é uma amostra aleatória gerada de uma distribuição normal de média $\mathbf{0}$ e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$. Neste trabalho para o cálculo de (2.40) utilizaremos as observações amostradas dadas por $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ geradas para a aproximação das matrizes da curvatura normal da forma descrita na seção 2.4.3. Assim, (2.40) fica dada por

$$g_{kk'}(\boldsymbol{\omega}) \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^S \frac{\partial \log p(\mathbf{Y}_o, \mathbf{b}_i^{(s)} | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_k} \frac{\partial \log p(\mathbf{Y}_o, \mathbf{b}_i^{(s)} | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_{k'}}, \quad (2.41)$$

com M_0 as primeiras, por exemplo, 1000 observações.

Capítulo 3

Influência local no modelo de regressão logística com efeitos mistos

Neste capítulo desenvolvemos a abordagem da análise de diagnóstico de influência local para os MLGM, apresentada no Capítulo 2, para o modelo de regressão logística com efeitos mistos (MRLEM) sob dois esquemas de perturbação indiretos para a variável resposta propostos por nós. Na Seção 3.1 descrevemos o MRLEM e na Seção 3.2 desenvolvemos a análise de diagnóstico de influência local sob cada um dos esquemas de perturbação propostos. Na Seção 3.3 avaliamos o desempenho das análises mediante estudos de estudos de simulação na Seção 3.4 ilustramos estas mesmas mediante aplicações a conjuntos de dados reais.

3.1 Modelo de regressão logística com efeitos mistos

Sejam Y_{ij} , para $j = 1, \dots, n_i$ e $i = 1, \dots, I$, variáveis aleatórias binárias independentes para i variado que assumem o valor 1 com probabilidade de sucesso p_{ij} e o valor 0 com probabilidade de fracasso $1 - p_{ij}$. Seja também \mathbf{b}_i um vetor aleatório de dimensão p_2 com distribuição normal de média $\mathbf{0}$ e matriz de variâncias e covariâncias $\Sigma = \Sigma(\gamma)$, em que γ é um vetor de componentes de variâncias e covariâncias desconhecidas de dimensão p_4 . Assumindo que a distribuição condicional de y_{ij} dado \mathbf{b}_i pertence à família exponencial com função de distribuição de probabilidade dada por

$$p(y_{ij}|\mathbf{b}_i) = \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right], \quad (3.1)$$

podemos fazer a analogia com a expressão dada em (2.4). Assim, $\theta_{ij} = \log[p_{ij}/(1 - p_{ij})]$, $a(\theta_{ij}) = \log[1/(1 - p_{ij})]$ e $\phi = 1$. Para obter a esperança e a variância, expressamos $a(\theta_{ij})$ em termos de θ_{ij} ,

isto é, $a(\theta_{ij}) = \log[1 + \exp(\theta_{ij})]$ e calculamos $\dot{a}(\theta_{ij})$ e $\ddot{a}(\theta_{ij})$ obtendo $\dot{a}(\theta_{ij}) = \exp(\theta_{ij})/[1 + \exp(\theta_{ij})]$ e $\ddot{a}(\theta_{ij}) = \exp(\theta_{ij})/[1 + \exp(\theta_{ij})]^2$, respectivamente. Em seguida, substituindo $\theta_{ij} = \log[p_{ij}/(1 - p_{ij})]$ nas expressões anteriores, a esperança e a variância estão dadas por

$$E(Y_{ij}|\mathbf{b}_i) = \mu_{ij} = \dot{a}(\theta_{ij}) = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})} = p_{ij}$$

e

$$V(Y_{ij}|\mathbf{b}_i) = V(\mu_{ij}) = \ddot{a}(\theta_{ij}) = \frac{\exp(\theta_{ij})}{[1 + \exp(\theta_{ij})]^2} = p_{ij}(1 - p_{ij}),$$

respectivamente. Assim, o MRLEM é definido por (3.1) e pelo componente sistemático

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i, \quad (3.2)$$

em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp_1})'$ e $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ijp_2})'$ são vetores de covariáveis de dimensões p_1 e p_2 , respectivamente, e $\boldsymbol{\beta}$ é o vetor de coeficientes desconhecidos de dimensão p_1 . Note que de (3.2) obtemos $p_{ij} = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)/[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)]$, implicando que, $\theta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$ e, em consequência, a função k definida na Seção 2.2 é a identidade. Uma vez definido o MRLEM e considerando que os efeitos aleatórios não são observáveis, o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} L(\boldsymbol{\psi}|\mathbf{Y}_o) &= \sum_{i=1}^I \log \left\{ \int_{\mathbb{R}^{p_2}} \prod_{j=1}^{n_i} \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right] \right. \\ &\quad \left. \times \frac{1}{(2\pi)^{p_2/2}} \det(\boldsymbol{\Sigma})^{-1/2} \exp \left(-\frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i \right) d\mathbf{b}_i \right\}, \end{aligned} \quad (3.3)$$

em que $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ é o vetor de parâmetros desconhecidos de dimensão $(p_1 + p_4)$ e \mathbf{Y}_o o conjunto de dados observados Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$. Como já foi dito no Capítulo 2, a integral múltipla envolvida em (3.3) não tem solução analítica e métodos de aproximação de integrais como AL e AQGH podem ser utilizados no procedimento da obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\psi}$ denotada por $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\gamma}}')'$.

3.1.1 Medidas de desempenho da predição

Em muitas situações o objetivo da utilização de um MRLEM é a predição da variável resposta, ou seja, a classificação do valor observado da variável resposta entre os valores 1, sucesso, ou 0, fracasso. Para levar a cabo isto, é necessário definir uma forma de fazer a classificação e obter medidas que indiquem o desempenho ou qualidade da classificação. Neste sentido, a predição ou classificação do

valor observado de Y_{ij} , denotada por \widehat{Y}_{ij} , está baseada num ponto de corte definido a partir da probabilidade estimada pelo modelo, isto é, \widehat{p}_{ij} e a avaliação do desempenho da classificação pode ser feita mediante diferentes medidas (vide, por exemplo, [Fawcett, 2006](#); [Hosmer et al., 2013](#)) entre as quais estão a sensibilidade, especificidade e acurácia que são descritas a seguir.

Considere a predição dada por $\widehat{Y}_{ij} = 1$, se o valor observado de Y_{ij} for classificado como sucesso e por $\widehat{Y}_{ij} = 0$, se o valor observado de Y_{ij} for classificado como fracasso. Define-se a sensibilidade (Sen) como a probabilidade de classificação correta do sucesso, ou seja,

$$\text{Sen} = P(\widehat{Y}_{ij} = 1 | Y_{ij} = 1) = \frac{P(\widehat{Y}_{ij} = 1; Y_{ij} = 1)}{P(Y_{ij} = 1)};$$

a especificidade (Esp) como a probabilidade de classificação correta do fracasso, ou seja,

$$\text{Esp} = P(\widehat{Y}_{ij} = 0 | Y_{ij} = 0) = \frac{P(\widehat{Y}_{ij} = 0; Y_{ij} = 0)}{P(Y_{ij} = 0)}$$

e a acurácia (ACC) como a probabilidade de classificação correta, ou seja,

$$\text{ACC} = P(\widehat{Y}_{ij} = 1; Y_{ij} = 1) + P(\widehat{Y}_{ij} = 0; Y_{ij} = 0).$$

Então, para determinar qual deve ser o valor de \widehat{p}_{ij} para que o valor observado de Y_{ij} seja classificado como sucesso, é necessário um ponto de corte. Uma forma bastante utilizada para determinar o ponto de corte é através da curva ROC (*Receiver operating characteristic curve*) (vide, por exemplo, [Fawcett, 2006](#)) que consiste no gráfico entre a sensibilidade e o complementar da especificidade para todos os possíveis pontos de corte (vide Figura 3.1). Assim, baseado na suposição de que classificar o valor observado como sucesso dado que na realidade ele é fracasso e o valor observado como fracasso dado que ele é sucesso traz “prejuízos” equivalentes, escolhemos o ponto de corte que corresponde à combinação ótima da sensibilidade e especificidade. Isto é, o ponto que mais se aproxima do canto superior esquerdo da curva ROC, pois é este que conjuntamente maximiza a sensibilidade e especificidade. Deste modo, uma vez determinado o ponto de corte, o valor observado de Y_{ij} é classificado de acordo com a sua probabilidade estimada \widehat{p}_{ij} quando comparada com o ponto de corte. Logo, se a probabilidade estimada for maior que o ponto de corte, o valor observado de Y_{ij} é classificado como sucesso, ou seja, 1 e se for menor, como fracasso, ou seja 0. Os possíveis resultados na classificação dos valores observados são apresentados na matriz de confusão dada na Tabela 3.1 e estão definidos por:

- VP (verdadeiro positivo): quantidade de valores observados que o modelo classificou como sucesso quando de fato são sucessos;
- VN (verdadeiro negativo): quantidade de valores observados que o modelo classificou como fracasso quando de fato são fracassos;
- FP (falso positivo): quantidade de valores observados que o modelo classificou como sucesso quando de fato são fracassos;
- FN (falso negativo): quantidade de valores observados que o modelo classificou como fracassos quando de fato são sucessos;

com os seguintes totais marginais:

- VP+FP: quantidade de valores observados que são classificados como sucessos;
- FN+VN: quantidade de valores observados que são classificados como fracassos;
- VP+FN: quantidade de valores observados que de fato são sucessos;
- FP+VN: quantidade de valores observados que de fato são fracassos;
- VP+FP+FN+VN: quantidade total de resultados.

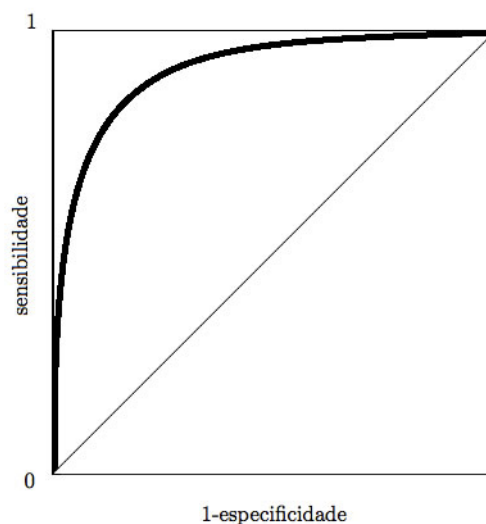


Figura 3.1: Gráfico da curva ROC.

Predição	Valor observado		Total
	$Y = 1$	$Y = 0$	
$\widehat{Y} = 1$	VP	FP	VP+FP
$\widehat{Y} = 0$	FN	VN	FN+VN
Total	VP+FN	FP+VN	VP+FP+FN+VN

Tabela 3.1: *Matriz de confusão.*

Por último, com os resultados da classificação é possível estimar as medidas de sensibilidade, especificidade e acurácia por

$$\text{Sen} = \frac{\text{VP}}{\text{VP} + \text{FN}},$$

$$\text{Esp} = \frac{\text{VN}}{\text{FP} + \text{VN}}$$

e

$$\text{ACC} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}};$$

respectivamente.

3.2 Análise de diagnóstico de influência local

Nesta seção realizamos análise de diagnóstico de influência local para o MRLEM sob dois esquemas perturbação indiretos para a variável resposta. Especificamente, desenvolvemos a metodologia com objetivo de derivar a curvatura normal sob cada um dos esquemas de perturbação. Começamos com a definição do logaritmo da função de verossimilhança do modelo postulado da seguinte maneira.

Baseados na Seção 2.4, seja $\mathbf{Y}_c = (\mathbf{Y}_o, \mathbf{Y}_f)$ o conjunto de dados completos em que $\mathbf{Y}_o = \{Y_{ij}; j = 1, \dots, n_i, i = 1, \dots, I\}$ é o conjunto de dados observados e $\mathbf{Y}_f = \{\mathbf{b}_i; i = 1, \dots, I\}$ o conjunto de dados faltantes. Logo, o logaritmo da função de verossimilhança conjunta de $\boldsymbol{\psi}$ dado os dados completos do MRLEM é dado por

$$L(\boldsymbol{\psi}|\mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} \left[y_{ij} \log \left(\frac{p_{ij}}{1-p_{ij}} \right) - \log \left(\frac{1}{1-p_{ij}} \right) \right] - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right\}. \quad (3.4)$$

Deste modo, a partir de (3.4) derivamos a curvatura normal sob cada um dos esquemas de perturbação.

3.2.1 Derivação da curvatura normal

Com intuito de derivar a curvatura normal começamos com a determinação da matriz $-\ddot{\mathbf{Q}}_{\psi}(\hat{\psi})$. Como ψ é de dimensão $(p_1 + p_4)$, temos que a matriz $-\ddot{\mathbf{Q}}_{\psi}(\hat{\psi})$ é de dimensão $(p_1 + p_4) \times (p_1 + p_4)$ e as derivadas compreendidas nesta são dadas por

$$-\frac{\partial^2 L(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^I \sum_{j=1}^{n_i} \frac{\exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij} \mathbf{x}'_{ij}}{[1 + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)]^2}_{(p_1 \times p_1)}, \quad (3.5)$$

$$-\frac{\partial^2 L(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = \mathbf{0}_{(p_1 \times p_4)}, \quad (3.6)$$

$$-\frac{\partial^2 L(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} = \mathbf{0}_{(p_4 \times p_1)} \quad (3.7)$$

e

$$-\frac{\partial^2 L(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\frac{I}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \right) \otimes \boldsymbol{\Sigma}^{-1}_{(p_4 \times p_4)}. \quad (3.8)$$

Para os detalhes dos cálculos de (3.5) a (3.8) vide Apêndice A.1.

Seguindo com a metodologia nas próximas seções definimos os esquemas de perturbação para a variável resposta e determinamos a matriz de influência para cada um deles.

3.2.2 Esquema de perturbação multiplicativa da probabilidade de sucesso

Com objetivo de identificar se existem observações influentes sob a perturbação da variável resposta binária, o esquema de perturbação consiste em modificar o valor da variável por meio de uma pequena perturbação na probabilidade de sucesso, isto é,

$$p_{ij}(\omega_{ij}) = p_{ij} \omega_{ij}, \quad \omega_{ij} \in (0, 1]. \quad (3.9)$$

Portanto, sob (3.9) o logaritmo da função de verossimilhança de ψ dado os dados completos do modelo perturbado fica dado por

$$L(\psi, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} \left[y_{ij} \log \left(\frac{p_{ij}(\omega_{ij})}{1 - p_{ij}(\omega_{ij})} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij})} \right) \right] - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right\}. \quad (3.10)$$

Logo, para o valor de não perturbação $\omega_{ij_0} = 1$, para $j = 1, \dots, n_i$ e $i = 1, \dots, I$, a matriz de influência dada por $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pm}}$ é uma matriz de dimensão $(p_1 + p_4) \times q$, com $q = \sum_{i=1}^I n_i$, e as derivadas

compreendidas nesta são dadas por

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = (y_{ij} - 1) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij}, \quad (3.11)$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0}, \quad (3.12)$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Para os detalhes dos cálculos de (3.11) e (3.12), vide Apêndice A.1. Note que para o caso em que $y_{ij} = 1$ a expressão dada em (3.11) é igual a 0. Assim, inicialmente a análise de diagnóstico é realizada para os casos em que a variável resposta assume valores iguais a 0. Posteriormente, para os casos em que a variável resposta assume valores iguais a 1, alternamos os valores desta e procedemos da mesma forma que no caso anterior.

3.2.3 Esquema de perturbação multiplicativa apropriada da probabilidade de sucesso

Para definir o esquema de perturbação multiplicativa apropriada da probabilidade de sucesso, consideramos o logaritmo da função de verossimilhança do modelo perturbado dado em (3.10) e obtemos a função de densidade associada, isto é,

$$p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega}) = \prod_{i=1}^I \prod_{j=1}^{n_i} \exp \left\{ \left[y_{ij} \log \left(\frac{p_{ij}(\omega_{ij})}{1 - p_{ij}(\omega_{ij})} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij})} \right) \right] - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right\}. \quad (3.13)$$

Desta forma, o vetor de não perturbação está dado por $\boldsymbol{\omega}_0 = \mathbf{1} \in \mathbb{R}^q$. Assim, baseados em (3.13), os elementos compreendidos na matriz $\mathbf{G}(\boldsymbol{\omega})$ definida na Seção 2.5 são dados por

$$g_{ij}(\boldsymbol{\omega}) = \mathbb{E} \left(\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} \middle| \boldsymbol{\omega} \right), \quad (3.14)$$

em que para $i = j$,

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} =$$

$$\begin{aligned} & \frac{y_{ij}[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)][1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) - 2\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)\omega_{ij}]}{\omega_{ij}^2[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) - \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)\omega_{ij}]^2} \\ & + \frac{\exp[2(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)]}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) - \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)\omega_{ij}]^2} \end{aligned} \quad (3.15)$$

e para $i \neq j$,

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0, \quad (3.16)$$

com $i, j = 1, \dots, q$. Para os detalhes dos cálculos de (3.15) e (3.16), vide Apêndice A.1. Logo, para o valor de não perturbação $\omega_{ij0} = 1$, as expressões dadas em (3.15) e (3.16) são iguais a

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = y_{ij}\{1 - \exp[2(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)]\} + \exp[2(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)] \quad (3.17)$$

e

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0, \quad (3.18)$$

respectivamente. No obstante, como a esperança de (3.17) não tem solução analítica é possível aproximar esta por

$$g_{ij}(\boldsymbol{\omega}_0) \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^S y_{ij}\{1 - \exp[2(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i^{(s)})]\} + \exp[2(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i^{(s)})], \quad (3.19)$$

para $i, j = 1, \dots, q$, com $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ uma amostra aleatória gerada mediante o algoritmo Metropolis-Hastings descrito na próxima seção.

Como $\mathbf{G}(\boldsymbol{\omega}_0) \neq a\mathbf{I}_q$, com $a > 0$ e \mathbf{I}_q a matriz identidade, escolhemos um novo vetor de perturbação dado por

$$\omega_{ij}^* = 1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\omega_{ij} - 1),$$

para $i, j = 1, \dots, q$. Então, o esquema de perturbação multiplicativa apropriada fica dado por

$$p_{ij}(\omega_{ij}^*) = p_{ij}\omega_{ij}^*, \quad \omega_{ij}^* \in (0, 1],$$

do modo que se verifica $\mathbf{G}(\boldsymbol{\omega}_0) = \mathbf{I}_q$, com $a = 1$. Logo, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos do modelo perturbado é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega}^* | \mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} \left[y_{ij} \log \left(\frac{p_{ij}(\omega_{ij}^*)}{1 - p_{ij}(\omega_{ij}^*)} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij}^*)} \right) \right] - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right\}. \quad (3.20)$$

Assim, para o valor de não perturbação $\omega_{ij0} = 1$, para $j = 1, \dots, n_i$ e $i = 1, \dots, I$, a matriz de influência dada por $\Delta_{\omega_{0pma}}$ é uma matriz de dimensão $(p_1 + p_4) \times q$, com $q = \sum_{i=1}^I n_i$, e as derivadas compreendidas nesta são dadas por

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = g_{ij}(\boldsymbol{\omega}_0)^{-1/2} (y_{ij} - 1) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij}, \quad (3.21)$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0}, \quad (3.22)$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Note que para o caso em que $y_{ij} = 1$ a expressão dada em (3.21) é igual a 0, assim o procedimento é o mesmo que para o esquema de perturbação multiplicativa. Para os detalhes dos cálculos de (3.21) e (3.22) vide Apêndice A.1.

3.2.4 Aproximação das matrizes da curvatura normal

Tendo em vista que a esperança condicional que se encontra nos blocos das matrizes $\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$, $\Delta_{\omega_{0pm}}$ e $\Delta_{\omega_{0pma}}$ é intratável e de grande dimensão, baseados na Seção 2.4.3, amostramos $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ da função de densidade dada por

$$p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi}) \propto \exp \left\{ -\frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i + \sum_{j=1}^{n_i} y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) - \log[1 + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)] \right\} \quad (3.23)$$

mediante o algoritmo Metropolis-Hastings como segue.

Inicia-se o algoritmo a partir de um valor arbitrário $\mathbf{b}_i^{(0)}$. Na r -ésima iteração realizam-se os seguintes passos:

(1) Dado o valor atual de $\mathbf{b}_i^{(r-1)}$, gere um novo candidato como $\mathbf{b}_i \sim N_{p_2}(\mathbf{b}_i^{(r-1)}, \Gamma(\mathbf{0}))$, em que

$$\Gamma(\mathbf{0}) = \Gamma(\mathbf{b}_i) |_{\mathbf{b}_i = \mathbf{0}} = \left\{ \boldsymbol{\Sigma}^{-1} + \sum_{j=1}^{n_i} \frac{\exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)}{[1 + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)]^2} \mathbf{z}_{ij} \mathbf{z}'_{ij} \right\}^{-1} \Big|_{\mathbf{b}_i = \mathbf{0}}.$$

(2) Gere $u \sim U(0, 1)$. Se $u \leq \alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i)$, faça $\mathbf{b}_i^{(r)} = \mathbf{b}_i$, senão faça $\mathbf{b}_i^{(r)} = \mathbf{b}_i^{(r-1)}$, com

$$\alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i) = \min \left(\frac{p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi})}{p(\mathbf{b}_i^{(r-1)} | \mathbf{Y}_o, \boldsymbol{\psi})}, 1 \right)$$

a probabilidade de aceitar um novo candidato.

(3) Repita os passos (1) e (2) para $r + 1$.

Uma vez amostradas as S observações descartamos as M_0 primeiras e calculamos as matrizes

$$-\ddot{Q}_\psi(\hat{\psi}) \approx \frac{1}{S - M_0} \begin{pmatrix} \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi|Y_o, Y_f^{(s)})}{\partial \beta \partial \beta'} \Big|_{\psi=\hat{\psi}} & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi|Y_o, Y_f^{(s)})}{\partial \beta \partial \gamma'} \Big|_{\psi=\hat{\psi}} \\ \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi|Y_o, Y_f^{(s)})}{\partial \gamma \partial \beta'} \Big|_{\psi=\hat{\psi}} & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi|Y_o, Y_f^{(s)})}{\partial \gamma \partial \gamma'} \Big|_{\psi=\hat{\psi}} \end{pmatrix}_{(p_1+p_4) \times (p_1+p_4)},$$

$$\Delta_{\omega_{0pm}} \approx \frac{1}{S - M_0} \begin{pmatrix} \dots & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi, \omega|Y_o, Y_f^{(s)})}{\partial \beta \partial \omega_{ij}} \Big|_{\psi=\hat{\psi}} & \dots \\ \dots & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi, \omega|Y_o, Y_f^{(s)})}{\partial \gamma \partial \omega_{ij}} \Big|_{\psi=\hat{\psi}} & \dots \end{pmatrix}_{(p_1+p_4) \times q}$$

e

$$\Delta_{\omega_{0pma}} \approx \frac{1}{S - M_0} \begin{pmatrix} \dots & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi, \omega|Y_o, Y_f^{(s)})}{\partial \beta \partial \omega_{ij}} \Big|_{\psi=\hat{\psi}} & \dots \\ \dots & \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi, \omega|Y_o, Y_f^{(s)})}{\partial \gamma \partial \omega_{ij}} \Big|_{\psi=\hat{\psi}} & \dots \end{pmatrix}_{(p_1+p_4) \times q}.$$

Em seguida, obtemos as matrizes

$$-2\ddot{Q}_{\omega_{0pm}} = 2\Delta'_{\omega_{0pm}} (-\ddot{Q}_\psi(\hat{\psi}))^{-1} \Delta_{\omega_{0pm}} \quad (3.24)$$

e

$$-2\ddot{Q}_{\omega_{0pma}} = 2\Delta'_{\omega_{0pma}} (-\ddot{Q}_\psi(\hat{\psi}))^{-1} \Delta_{\omega_{0pma}}, \quad (3.25)$$

e, por conseguinte, as curvaturas normais individuais

$$C_{fQ, h_{ijpm}} = \sum_{m=1}^M \lambda_{m_{pm}} e^2_{m_{ijpm}} \quad (3.26)$$

e

$$C_{fQ, h_{ijpma}} = \sum_{m=1}^M \lambda_{m_{pma}} e^2_{m_{ijpma}}, \quad (3.27)$$

para $i, j = 1, \dots, q$, respectivamente. Para realizar a análise de diagnóstico de influência local sob o esquema de perturbação multiplicativa e multiplicativa apropriada de acordo com a Definição 2.4.1 estabelecemos os correspondentes intervalos de influência por

$$\left[(k_0 - 1) \frac{\lambda_{0_{pm}}}{q}, k_0 \frac{\lambda_{0_{pm}}}{q} \right), \quad (3.28)$$

em que $\lambda_{0_{pm}} = \sum_{i,j=1}^q C_{f_Q, \mathbf{h}_{ij_{pm}}}$ e

$$\left[(k_0 - 1) \frac{\lambda_{0_{pma}}}{q}, k_0 \frac{\lambda_{0_{pma}}}{q} \right), \quad (3.29)$$

em que $\lambda_{0_{pma}} = \sum_{i,j=1}^q C_{f_Q, \mathbf{h}_{ij_{pma}}}$, para $k_0 = 1, 2, 3, \dots$, respectivamente.

3.3 Estudos de simulação

Com o objetivo de avaliar o desempenho da análise de diagnóstico de influência local sob os esquemas de perturbação propostos desenvolvemos dois estudos de simulação de Monte Carlo com cenários de $R = 100$ réplicas cada um (valor similar ao considerado por [Xu et al. \(2006\)](#), que utilizam $R=101$ réplicas). Cada cenário considera a seguinte estrutura do modelo e procedimento de simulação das réplicas.

Seja o MRLEM dado por

$$p(y_{ij}|b_i) = \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right]$$

e pelo componente sistemático

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (3.30)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, I$, em que x_{1ij} é uma variável indicadora de 0 e 1 gerada de tal forma que existam grupos com maior quantidade de uns e grupos com maior quantidade de zeros, x_{2ij} é uma variável aleatória com distribuição uniforme no intervalo $(0, 1)$, β_0 , β_1 e β_2 são os coeficientes de regressão fixos e b_i é o intercepto aleatório com distribuição normal de média 0 e variância σ^2 . Sejam também os verdadeiros valores dos parâmetros dados por $\beta_0 = 1$, $\beta_1 = 1$, $\beta_2 = 1$ e $\sigma^2 = 0,5$ ($\sigma = 0,71$) (valores iguais aos considerados por [Chen et al., 2010](#)). Para valores predeterminados de n_i e I , geramos os valores das variáveis x_{1ij} , x_{2ij} e b_i e, a partir de (3.30), calculamos as probabilidades p_{ij} . Geramos cada réplica da variável resposta y_{ij} a partir de uma distribuição Bernoulli de parâmetro p_{ij} (com os valores de p_{ij} calculados anteriormente), para $j = 1, \dots, n_i$ e $i = 1, \dots, I$. A este conjunto de dados o denominamos dados originais e o denotamos por O . Uma vez gerado o conjunto de dados O , geramos o conjunto de dados “perturbados” sob o esquema de perturbação multiplicativa da seguinte maneira. Escolhemos os grupos a perturbar

pelos grupos $i = 1$, grupo com uma maior quantidade de uns, e $i = 2$, grupo com uma maior quantidade de zeros. Do conjunto de dados originais voltamos a gerar os n_i valores de y_{ij} para o grupo $i = 1$ e os n_i valores de y_{ij} para o grupo $i = 2$, isto é, para um valor da perturbação ω_{ij} geramos os valores de y_{ij} a partir de uma distribuição Bernoulli de parâmetro $p_{ij}(\omega_{ij}) = p_{ij}\omega_{ij}$. A este novo conjunto de dados o denominamos dados sob o esquema de perturbação multiplicativa e o denotamos por PM. Da mesma forma, para o esquema de perturbação multiplicativa apropriada geramos os n_i novos valores de y_{ij} para o grupo $i = 1$ e os n_i valores de y_{ij} para o grupo $i = 2$. Isto é, geramos os valores de y_{ij} a partir de uma distribuição de parâmetro $p_{ij}(\omega_{ij}^*) = p_{ij}\omega_{ij}^*$, com $\omega_{ij}^* = [1 + g_{ij}(\omega_0)^{-1/2}(\omega_{ij} - 1)]$ em que $g_{ij}(\omega_0)$ é aproximada com 2000 valores adicionais de b_i (valor igual ao considerado por [Chen et al., 2010](#)) gerados de uma distribuição normal de média 0 e variância σ^2 . A este novo conjunto de dados o denominamos dados sob o esquema de perturbação multiplicativa apropriada e o denotamos por PMA.

Uma vez simulados os conjuntos de dados, desenvolvemos a análise de diagnóstico de influência local para cada um dos conjuntos PM e PMA. Primeiro, para cada um deles realizamos a estimação dos parâmetros por máxima verossimilhança com integração numérica pelos métodos AL e AQGH (com cinco pontos). Tanto pelo método AL quanto pelo método AQGH amostramos $S = 600$ observações de b_i , para $i = 1, \dots, I$, mediante o algoritmo Metropolis-Hastings descrito na Seção 3.2.4, e calculamos as matrizes $-\ddot{Q}_{\psi}(\hat{\psi})$, $\Delta_{\omega_{0_{pm}}}$ e $\Delta_{\omega_{0_{pma}}}$, com $M_0 = 100$, ou seja, com $S - M_0 = 500$ observações (valor um pouco maior do que o considerado por [Chen et al. \(2010\)](#) que utilizam $S - M_0 = 400$ observações). Posteriormente, calculamos as matrizes dadas por (3.24) e (3.25) e, por conseguinte, as curvaturas normais individuais dadas por (3.26) e (3.27), respectivamente. Por último, obtemos os intervalos de influência correspondentes e verificamos se as observações perturbadas para $Y_{ij} = 0$ estão sendo detectadas como influentes. Para verificar se as observações perturbadas com $Y_{ij} = 1$ estão sendo detectadas como influentes alternamos os valores da variável resposta e realizamos o mesmo procedimento descrito anteriormente. Com propósito de definir tipos de pontos influentes adotaremos a seguinte nomenclatura. Se a observação é detectada como influente e o valor da variável resposta é diferente do valor da variável resposta do conjunto de dados O, a observação é identificada como verdadeiro influente (VI). Se a observação é detectada como influente e o valor da variável resposta é igual ao valor da variável resposta do conjunto de dados O, a observação é identificada como falso influente (FI). Por outro lado, se a observação é detectada como influente independente do valor da variável resposta, a observação é identificada como influente (I). Por último, realizamos o procedimento para $R = 100$ réplicas e calculamos a proporção de VI, FI e I sob cada um dos esquemas de perturbação.

Para o estudo de simulação 1, para cada um dos conjuntos de dados O, PM e PMA, primeiramente, analisamos o comportamento das estimativas dos parâmetros mediante a média, viés e raiz do erro quadrático médio (REQM). A seguir, estudamos a capacidade preditiva do modelo mediante a média e o erro padrão das medidas de sensibilidade (Sen), especificidade (Esp) e acurácia (ACC). Por último, avaliamos o desempenho da análise de diagnóstico de influência local sob cada um dos esquemas de perturbação, com as proporções de VI, FI e I para os grupos 1 e 2 pelos métodos AL e AQGH para os valores de ω_{ij} iguais a 0,75; 0,85 e 0,95 para os seguintes cenários:

- $q = 90$, $I = 30$ e $n_i = 3$;
- $q = 360$, $I = 60$ e $n_i = 6$;
- $q = 1080$, $I = 120$ e $n_i = 9$.

Para o estudo de simulação 2 avaliamos o desempenho da análise de diagnóstico de influência local sob cada um dos esquemas de perturbação com as proporções de VI, FI e I para os grupos 1 e 2 no caso em que para um valor fixado de $q = \sum_{i=1}^I n_i$ consideram-se diferentes valores do número de grupos I e tamanho dos grupos n_i pelos métodos AL e AQGH para os valores de ω_{ij} iguais a 0,75; 0,85 e 0,95 para os seguintes cenários:

- $q = 90$: $I = 15$ e $n_i = 6$ e $I = 30$ e $n_i = 3$;
- $q = 180$: $I = 30$ e $n_i = 6$ e $I = 60$ e $n_i = 3$;
- $q = 270$: $I = 30$ e $n_i = 9$ e $I = 90$ e $n_i = 3$.

A seguir apresentamos os resultados para cada um deles

3.3.1 Estudo de simulação 1

Investigamos o comportamento das estimativas dos parâmetros β_0 , β_1 e β_2 mediante a média, viés e REQM. A Figura 3.2, correspondente a $\omega = 0,75$, mostra que para os parâmetros β_0 , β_1 e β_2 , à medida que q aumenta, a média tende ao verdadeiro valor 1 e o viés e REQM diminuem para cada um dos conjuntos de dados O, PM e PMA pelos métodos AL e AQGH. Além disso, cabe destacar que para $q = 90$ os parâmetros β_0 e β_1 são superestimados e o parâmetro β_2 é subestimado, com viés grande em valor absoluto para todos os casos. Com respeito ao comportamento das estimativas do parâmetro σ , a Figura 3.5, correspondente a $\omega = 0,75$, mostra que à medida que q aumenta a média tende a seu verdadeiro valor 0,71, e que a partir de $q = 360$ a média tende a ser subestimada, com viés e REQM grandes para todos os casos para cada um dos conjuntos de dados O, PM e

PMA pelos dois métodos. Assim, levando em conta todos os cenários, não se identifica nenhuma situação sobre qual esquema se produz maior impacto nas estimativas independente do método de integração numérica considerado. Observam-se os mesmos resultados para $\omega=0,85$ e $0,95$.

Investigamos também o comportamento das medidas Sen, Esp e ACC mediante a média e o erro padrão. A Figura 3.8, correspondente a $\omega=0,75$, mostra que a média de Sen, Esp e ACC para $q = 90$ encontra-se ao redor de 0,80, para $q = 360$ ao redor de 0,73 e para $q = 1080$ ao redor de 0,68 para os conjuntos de dados O, PM e PMA pelos dois métodos. Por outro lado, para todos os casos o erro padrão foi quase nulo tornando-se impercetível para alguns deles. Observam-se os mesmos resultados para $\omega=0,85$ e $0,95$ nas Figuras 3.9 e 3.10, respectivamente. Por fim, as medidas de Sen, Esp e ACC parecem não ser afetadas pela presença de dados perturbados independente do esquema de perturbação adotado.

Por último, com respeito ao desempenho da análise de diagnóstico de influência local sob os esquemas PM e PMA, temos que para todos os valores de ω para cada tipo de ponto influente, isto é, VI, FI e I as proporções, em geral, foram similares quando se compararam os grupos 1 e 2, os diferentes valores de q e os métodos AL e AQGH, pelo que expomos os resultados sem esses detalhes. A Figura 3.11, correspondente a $\omega = 0,75$, mostra que as proporções de VI foram maiores sob o esquema PMA que sob o esquema PM. Já as proporções de FI, em geral, foram maiores sob o esquema PM que sob o esquema PMA. Agora, ao se considerar os I as proporções foram similares entre os dois esquemas. Especificamente, a proporção máxima de VI foi de 0,34 sob o esquema PM e de 0,39 sob o esquema PMA; a proporção máxima de FI foi de 0,31 sob o esquema PM e de 0,42 sob o esquema PMA. Por último, a proporção máxima de I foi de 0,49 sob o esquema PM e de 0,48 sob o esquema PMA. Observam-se resultados similares para $\omega=0,85$ e $0,95$ nas Figuras 3.12 e 3.13, respectivamente. Assim, o valor da perturbação não parece ter impacto sobre as proporções de pontos influentes.

Em conclusão, a análise de diagnóstico de influência local para os VI levaram a um desempenho maior sob o esquema PMA. Ao contrário, para os FI levaram a um desempenho maior sob o esquema PM. Por último, para os I levaram a um desempenho similar entre os esquemas. Cabe destacar que as proporções de VI, em geral, foram maiores às proporções de FI, conforme esperado.

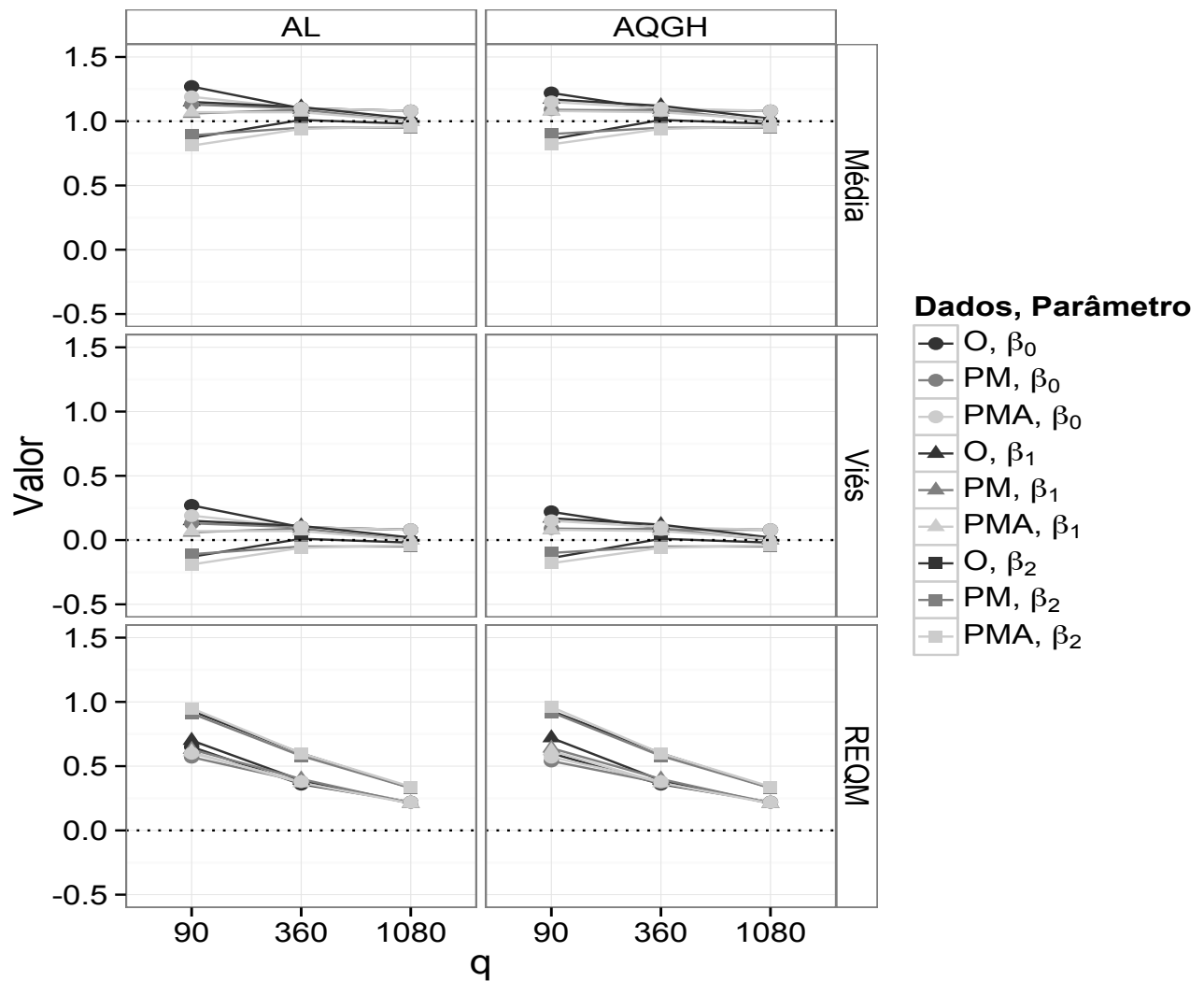


Figura 3.2: Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1)(MRLEM)

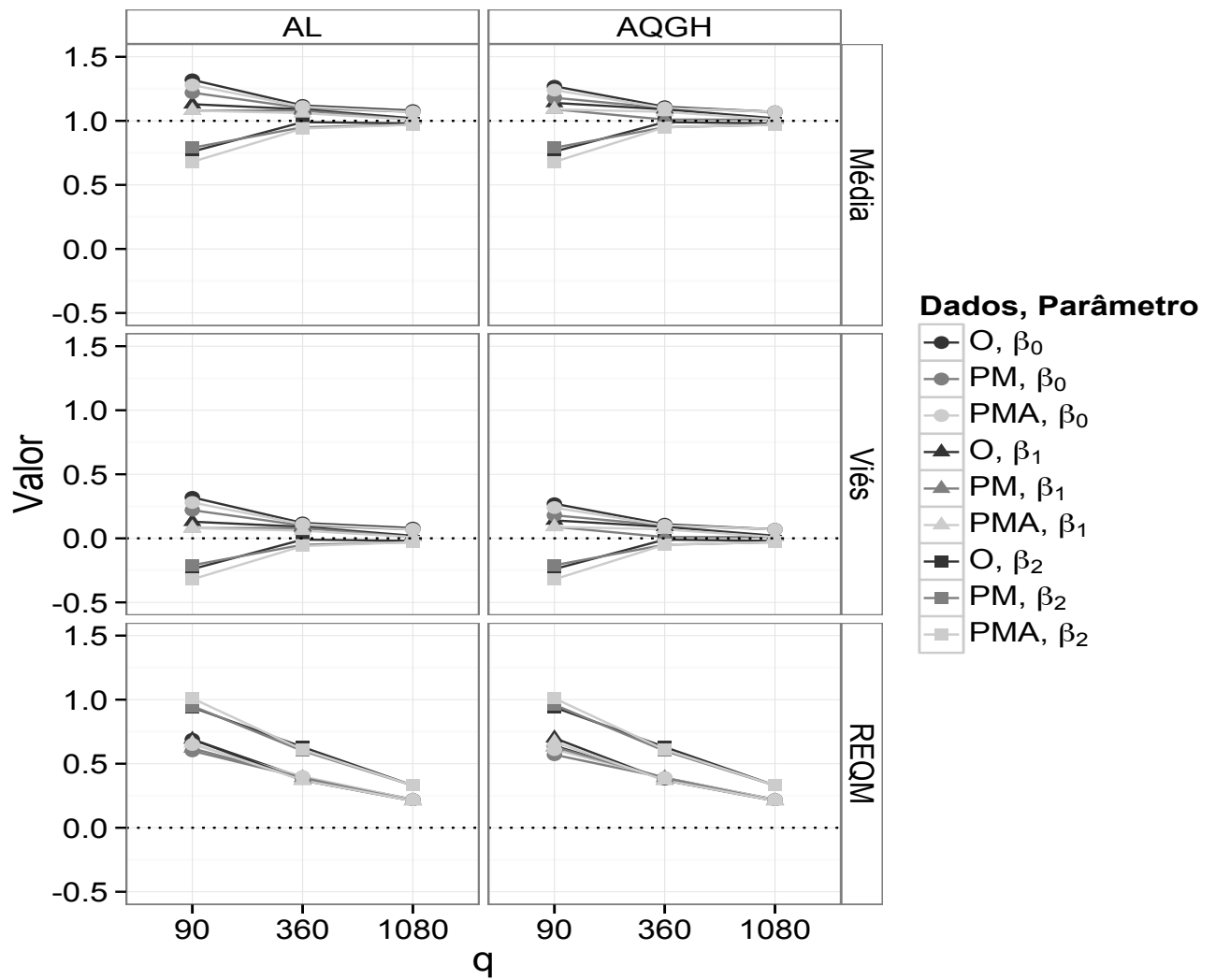


Figura 3.3: Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)

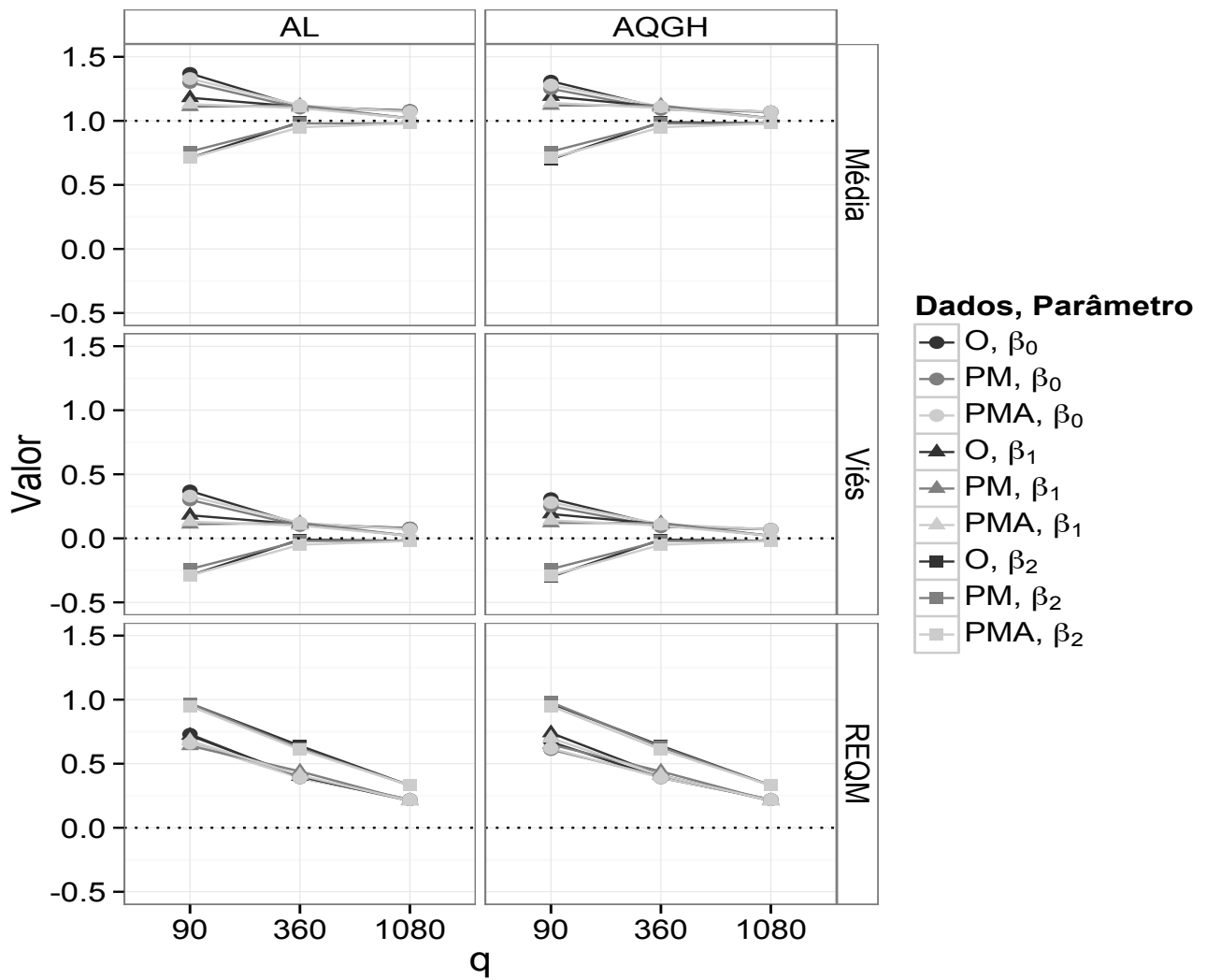


Figura 3.4: Média, viés e REQM das estimativas dos parâmetros β_0 , β_1 e β_2 para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)

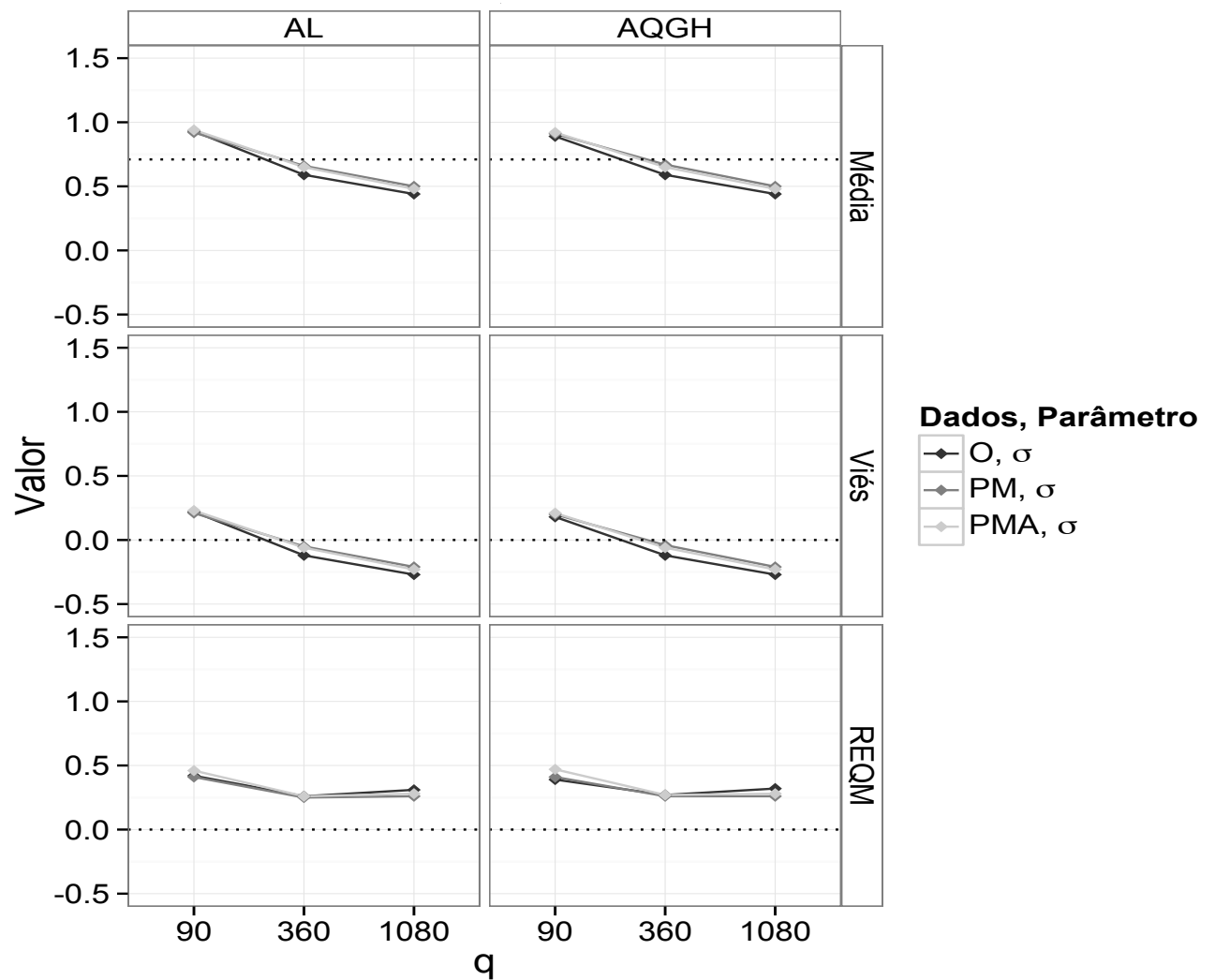


Figura 3.5: Média, viés e REQM da estimativa do parâmetro σ para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM)

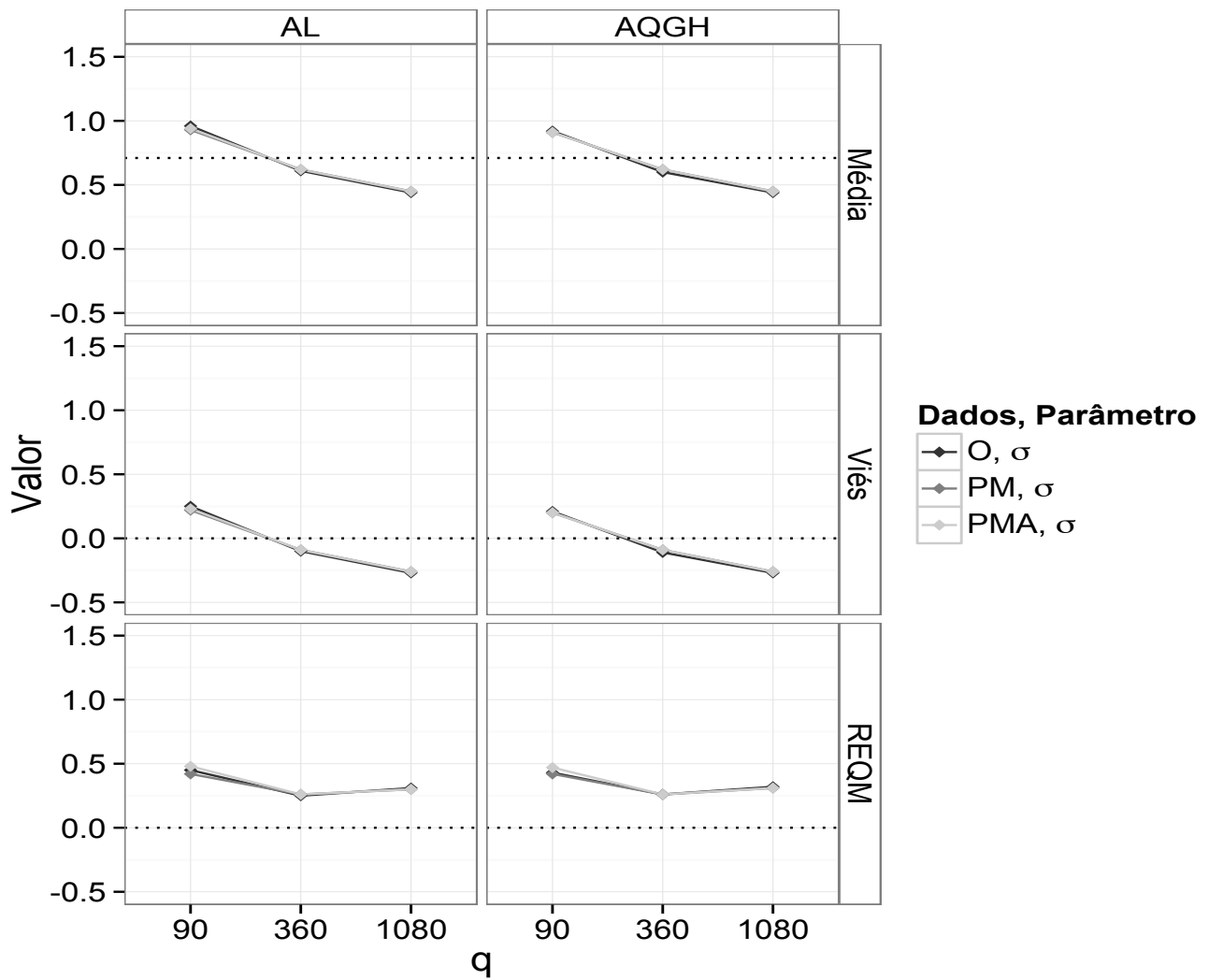


Figura 3.6: Média, viés e REQM da estimativa do parâmetro σ para os dados O , PM e PMA para diferentes valores de q pelos métodos AL e $AQGH$, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)

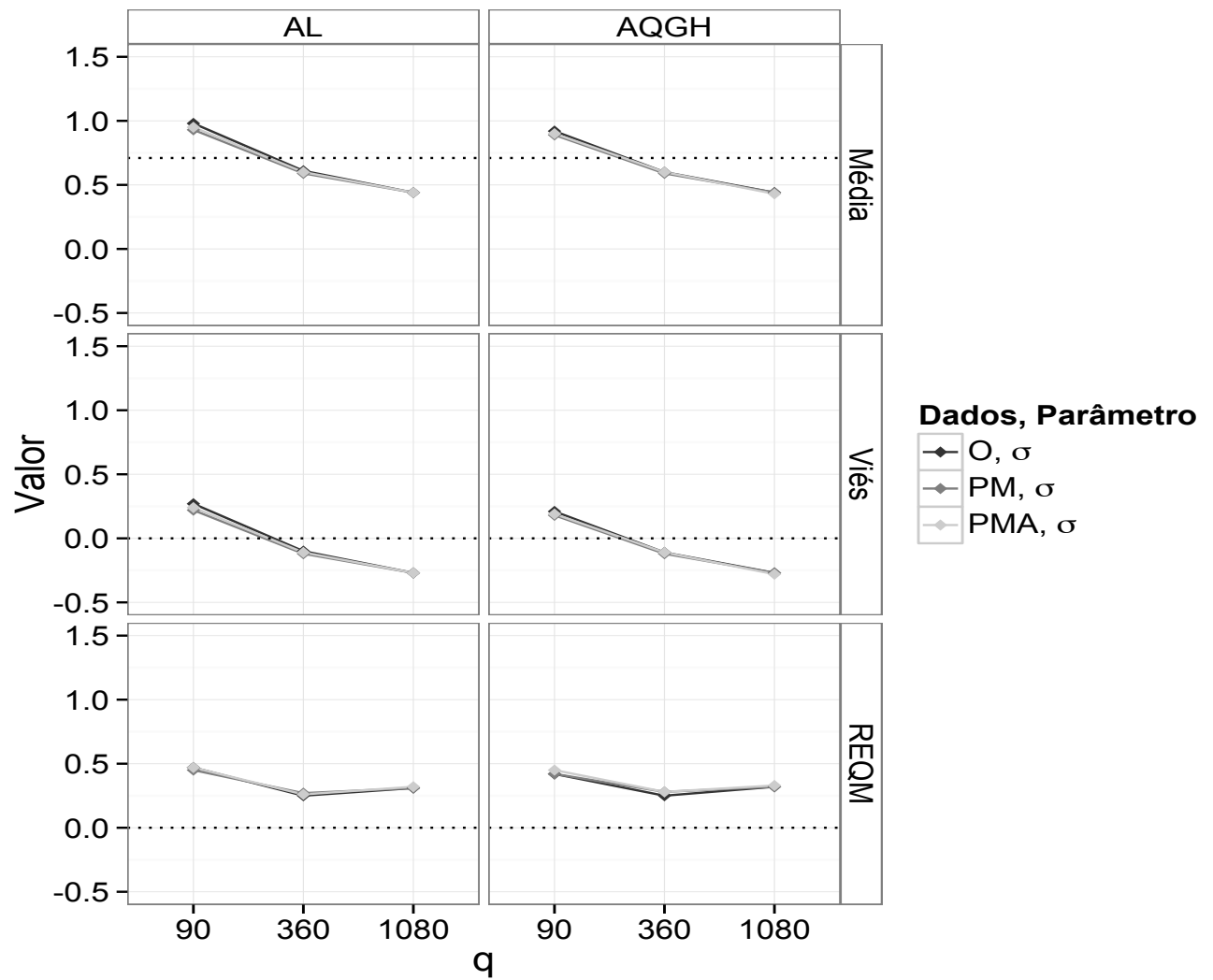


Figura 3.7: Média, viés e REQM da estimativa do parâmetro σ para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)

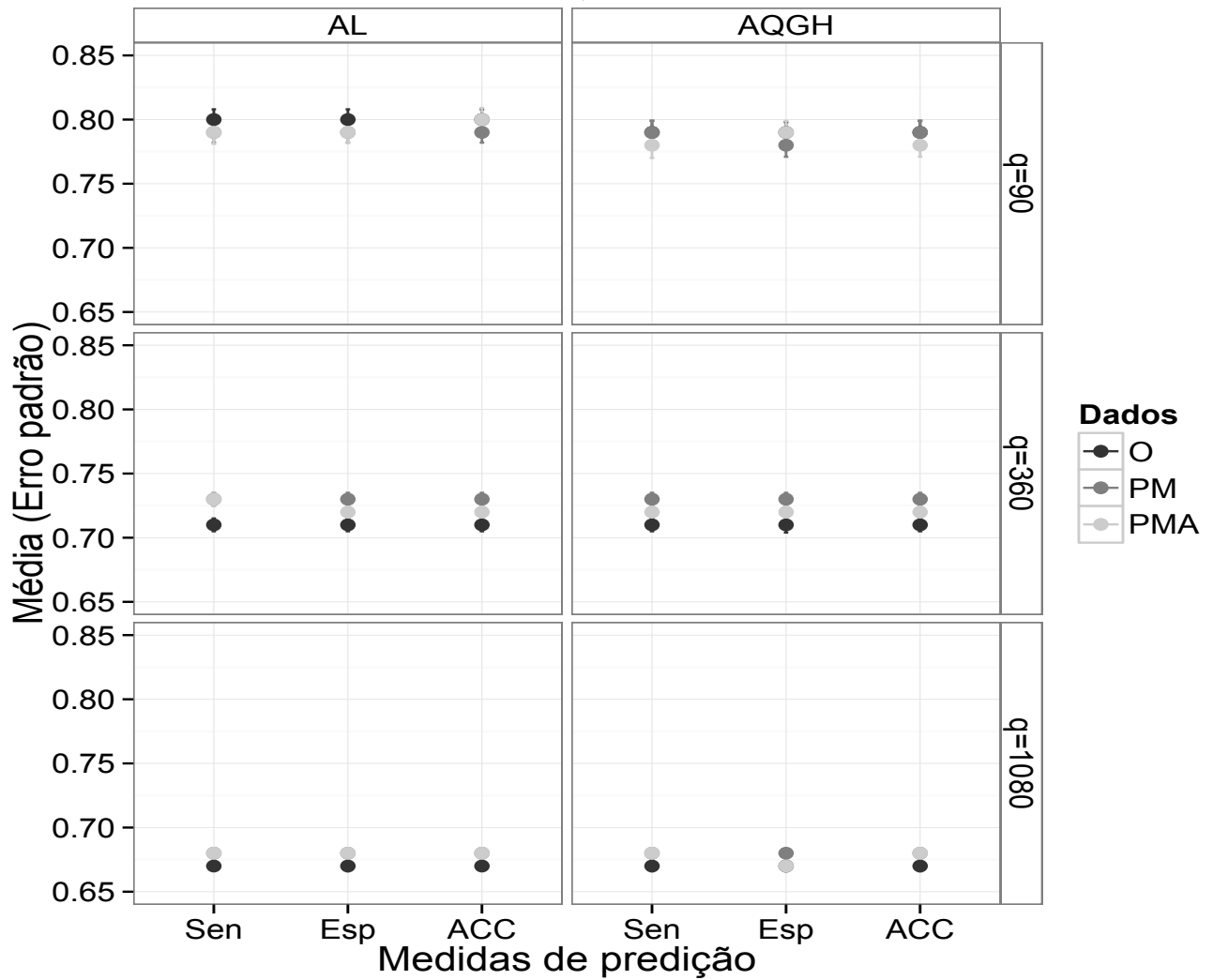


Figura 3.8: Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM)

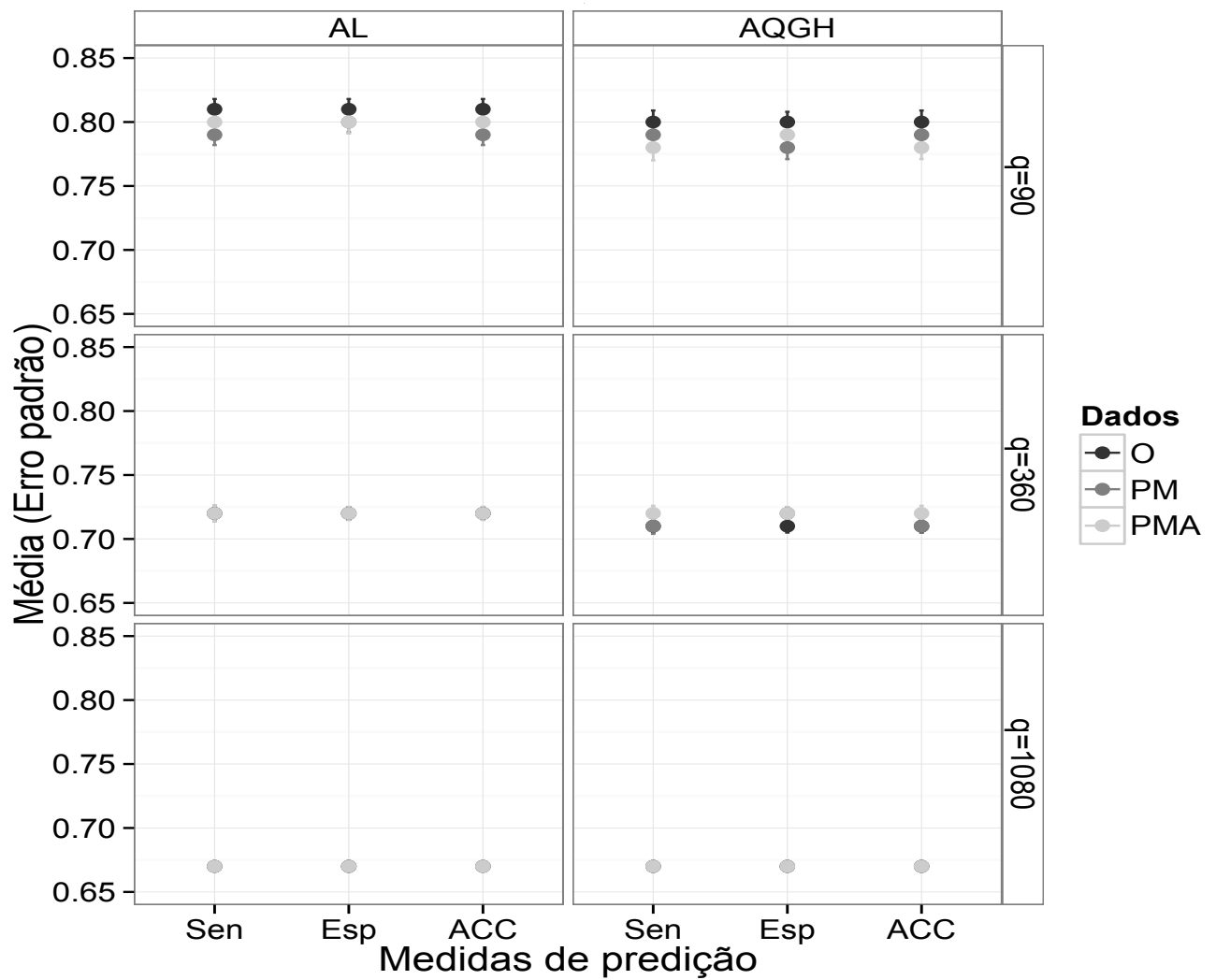


Figura 3.9: Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)

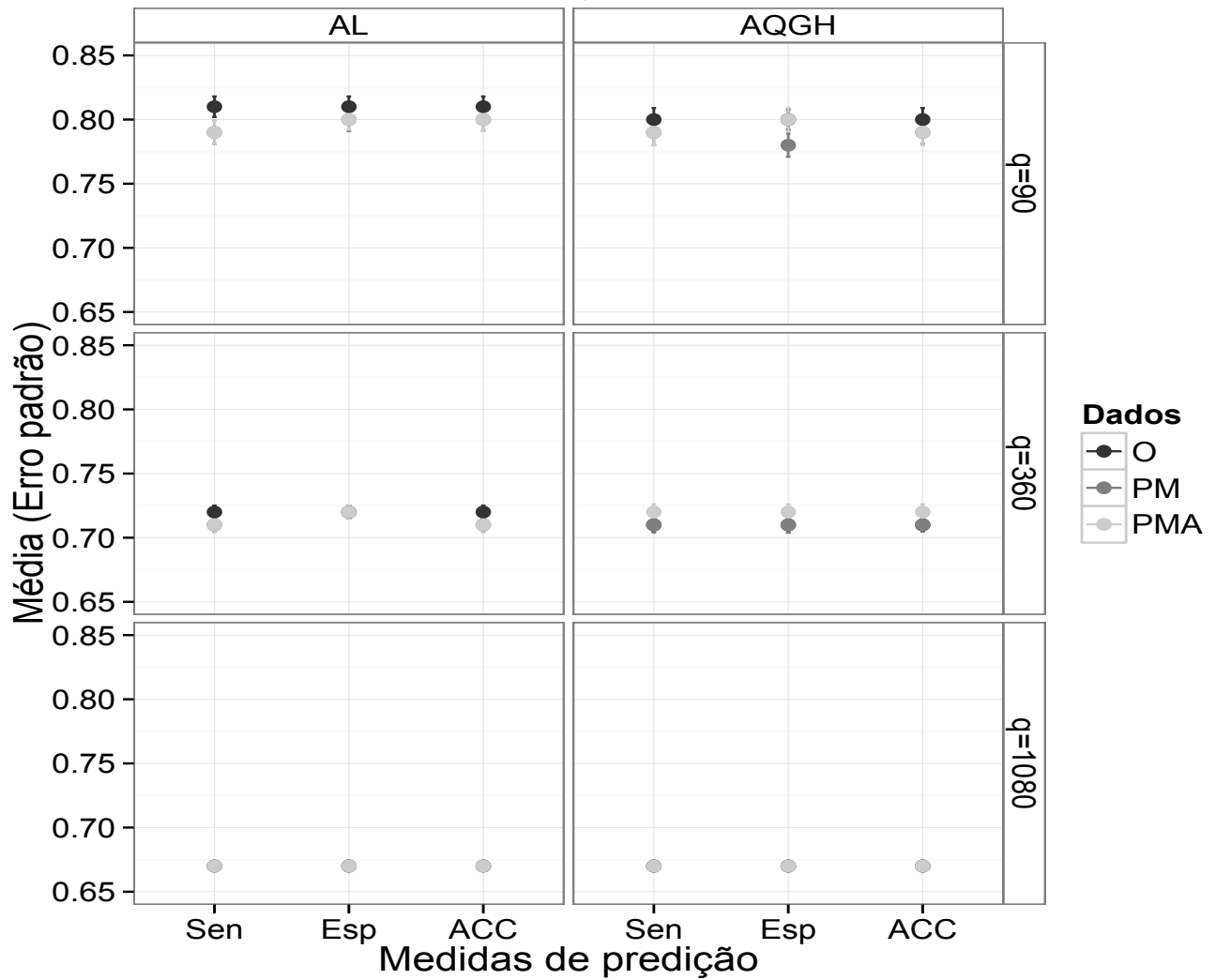


Figura 3.10: Média e erro padrão das medidas de Sen, Esp e ACC para os dados O, PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)

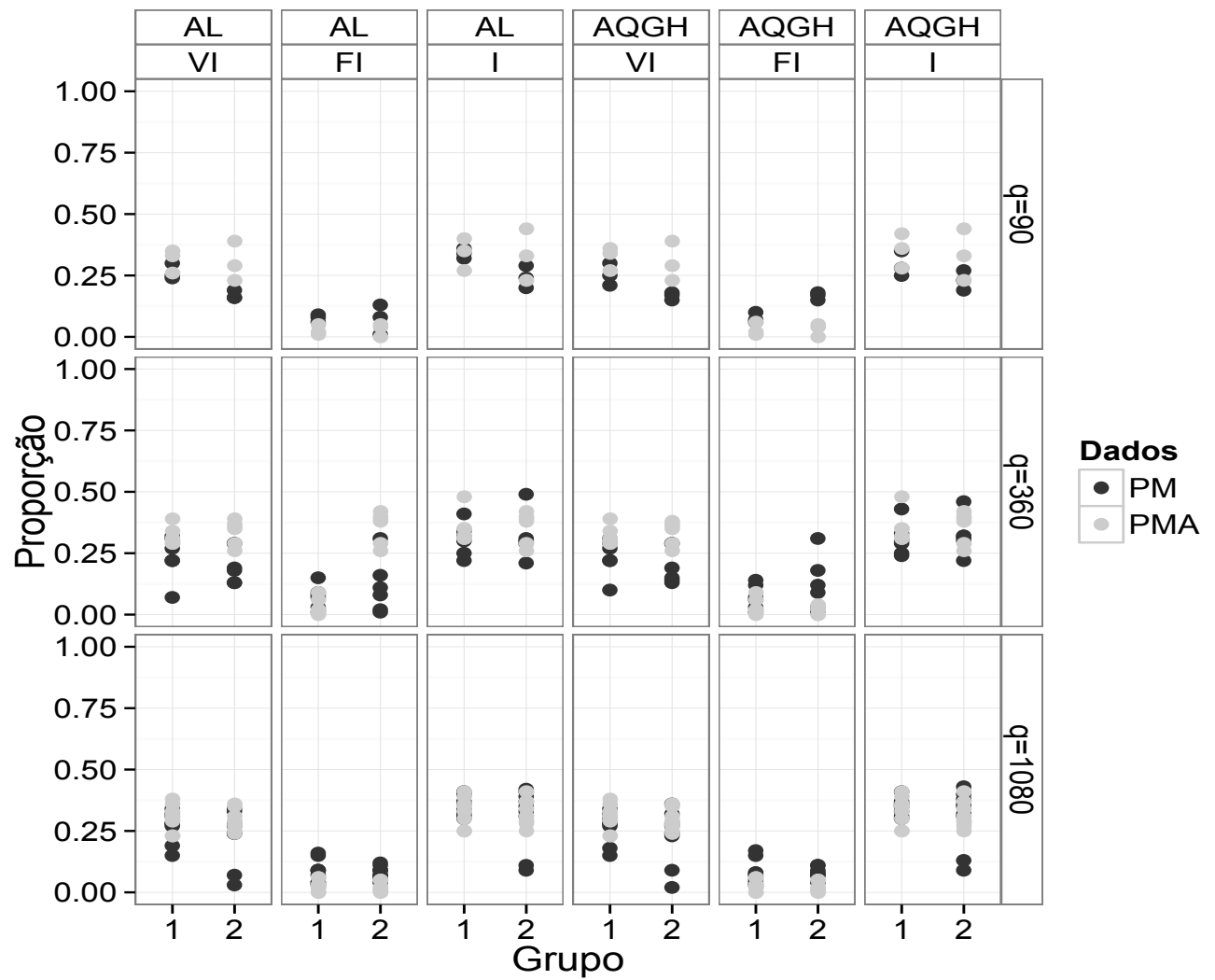


Figura 3.11: Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,75$. (Estudo de simulação 1) (MRLEM)

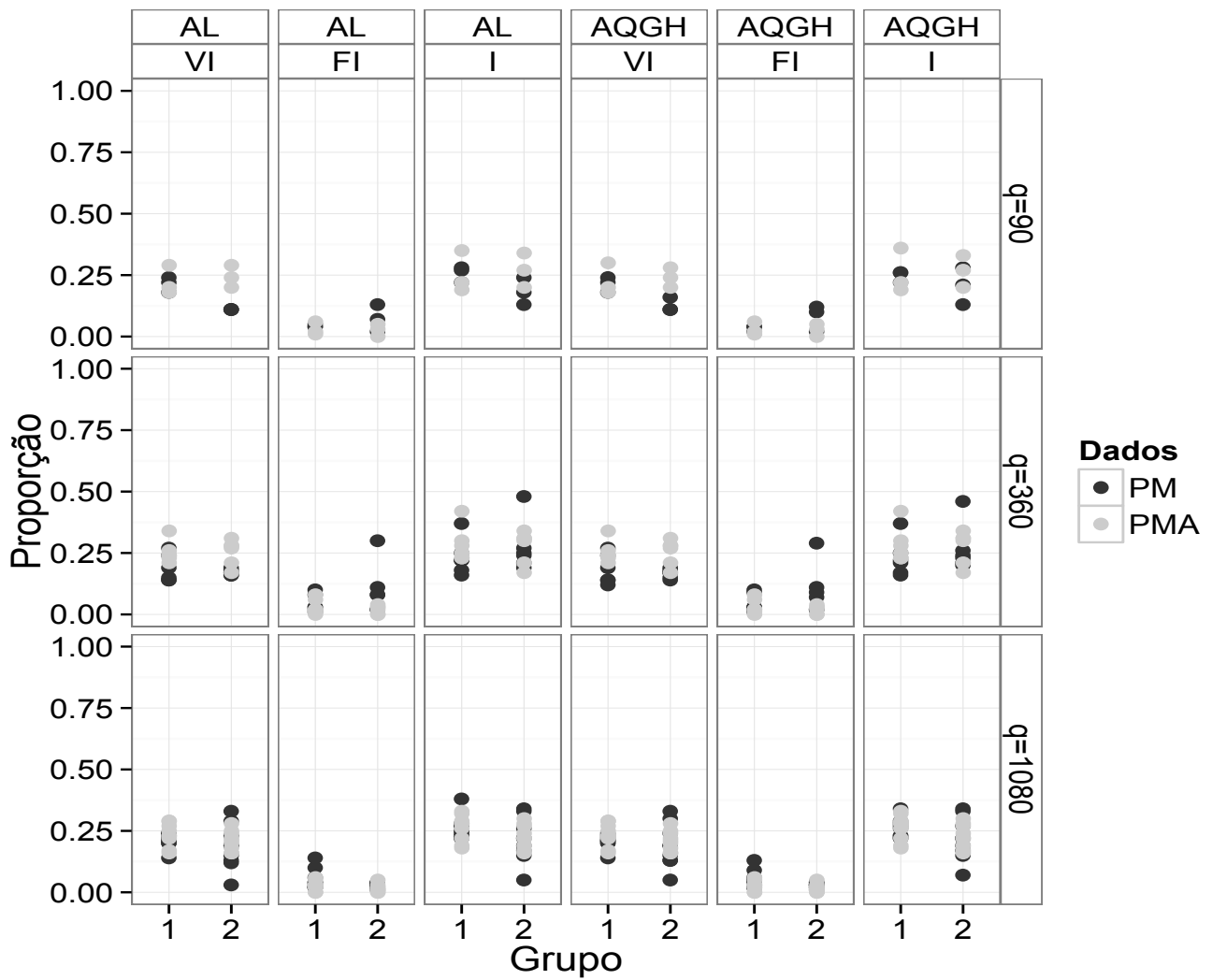


Figura 3.12: *Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 1) (MRLEM)*

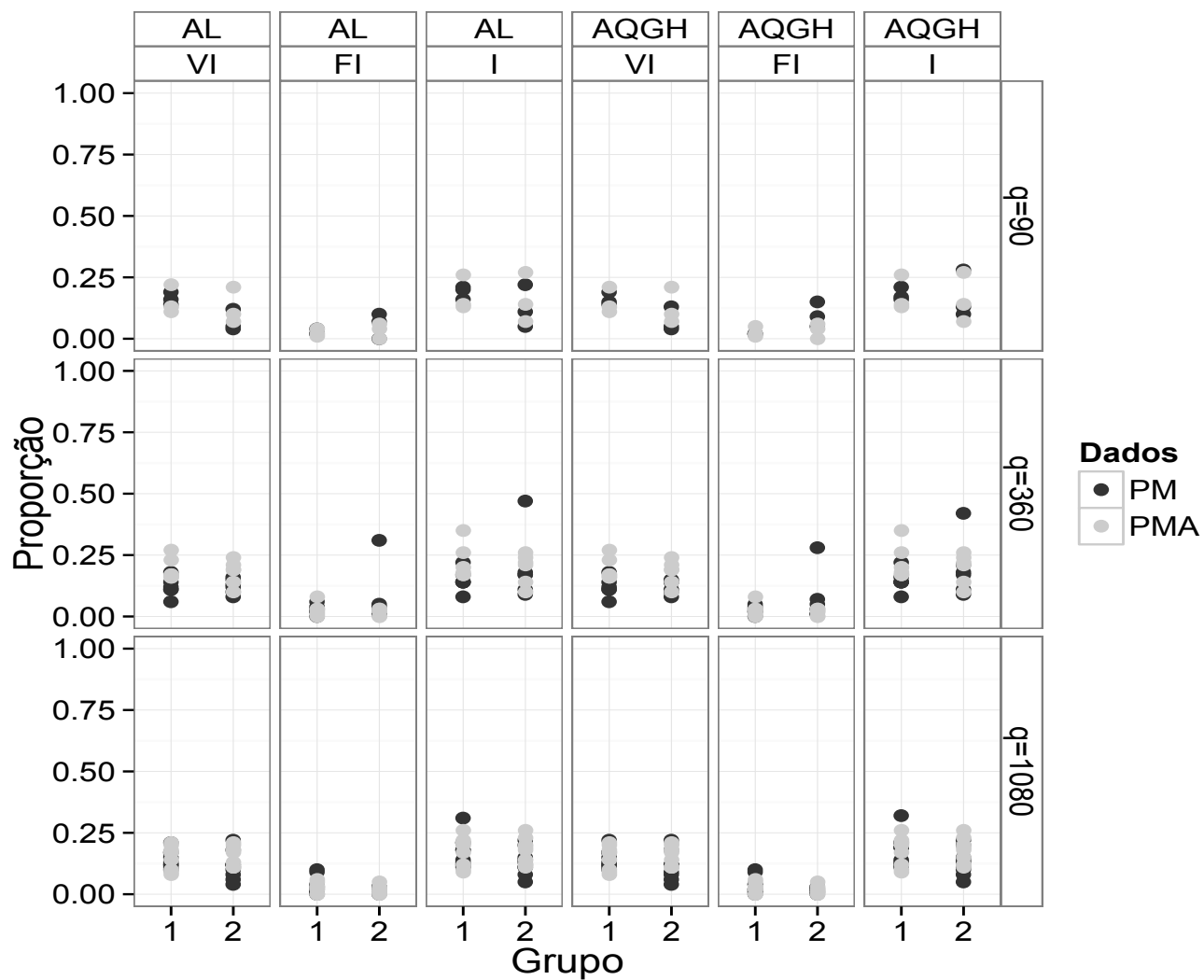


Figura 3.13: *Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA para diferentes valores de q pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 1) (MRLEM)*

3.3.2 Estudo de simulação 2

No estudo de simulação 2, avaliamos o desempenho da análise de diagnóstico de influência local sob cada um dos esquemas PM e PMA pelos métodos AL e AQGH, com as proporções de VI, FI e I no caso em que para um valor fixado de q consideram-se diferentes valores do número de grupos I e tamanho dos grupos n_i , para os valores da perturbação $\omega_{ij}=0,75; 0,85$ e $0,95$.

Em primeiro lugar, como no estudo de simulação 1, para cada tipo de ponto influente, isto é, VI, FI e I as proporções foram similares quando comparamos os grupos 1 e 2, os diferentes valores de q e os métodos AL e AQGH, de maneira que expomos os resultados sem esses detalhes.

A Figura 3.14, correspondente a $\omega = 0,75$, mostra que as proporções de VI para cada um dos esquemas PM e PMA foram similares no caso em que para um valor fixado de q consideram-se diferentes valores do número de grupos I e tamanho dos grupos n_i . As proporções de FI e I para cada um dos esquemas, em geral, foram levemente maiores no caso em que para um valor fixado de q consideram-se números de grupos I menores. Cabe destacar que para os I, com $q=90$ e 180 e números de grupos I menores observam-se proporções de até $0,67$, para cada um dos esquemas. Resultados similares observam-se nas Figuras 3.15 e 3.16 para $\omega=0,85$ e $0,95$, respectivamente.

Em conclusão, o desempenho da análise de diagnóstico de influência local sob cada um dos esquemas PM e PMA para um valor fixado de q aparentemente não é afetado pela relação número de grupos e tamanho dos grupos.

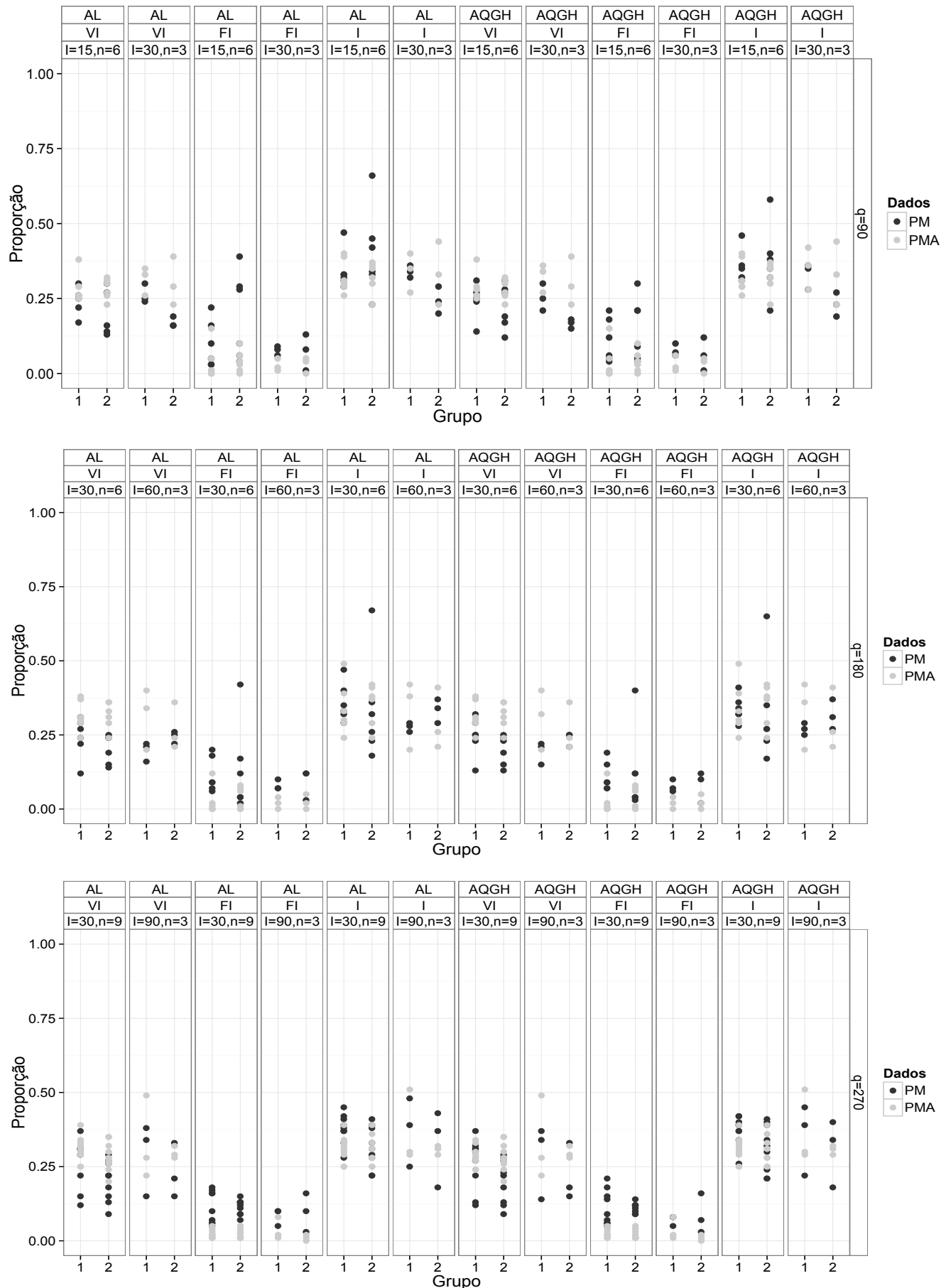


Figura 3.14: Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de l e n_i pelos métodos AL e AQQH, com $\omega = 0,75$. (Estudo de simulação 2) (MRLEM)

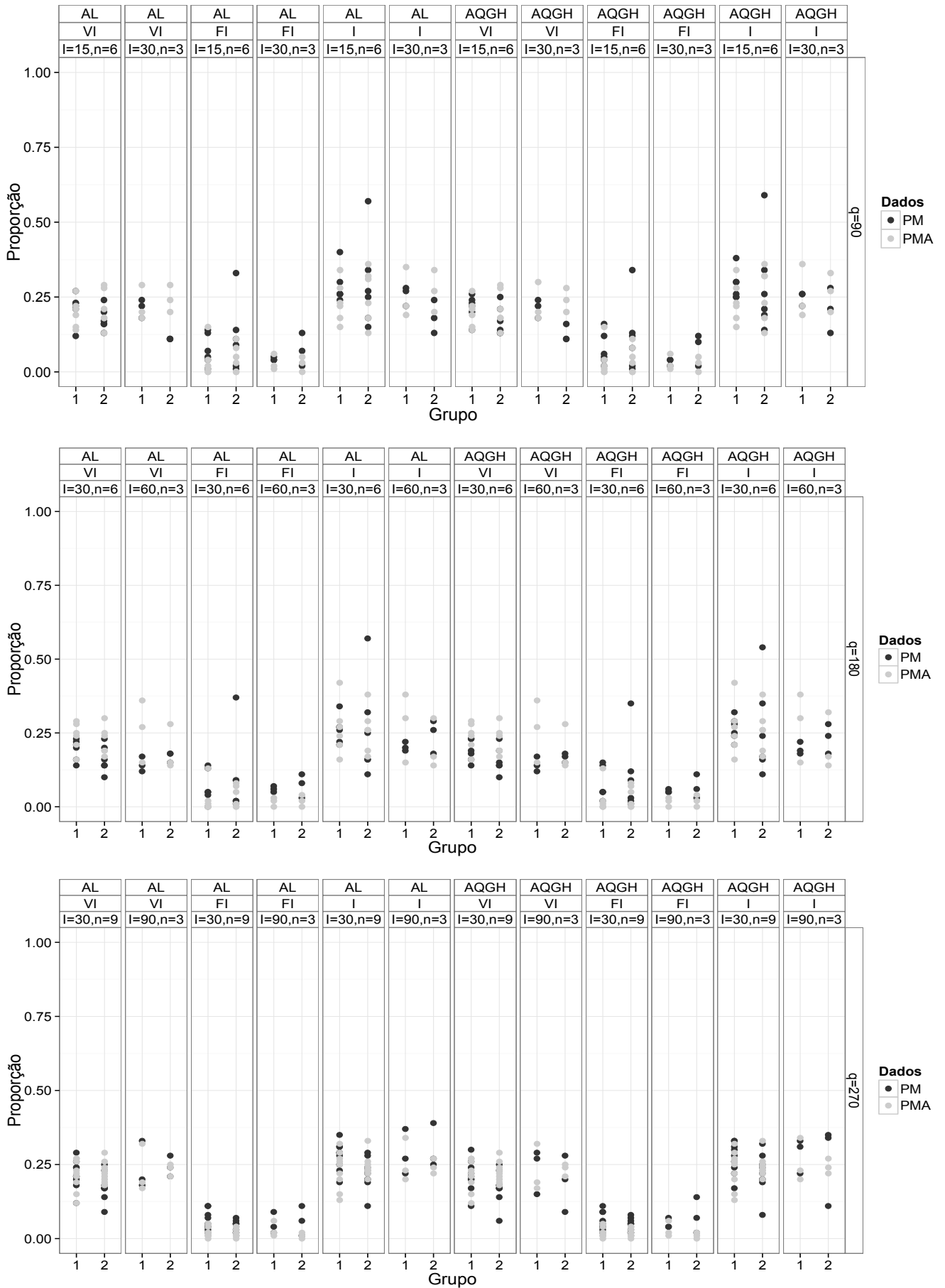


Figura 3.15: Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH, com $\omega = 0,85$. (Estudo de simulação 2) (MRLEM)

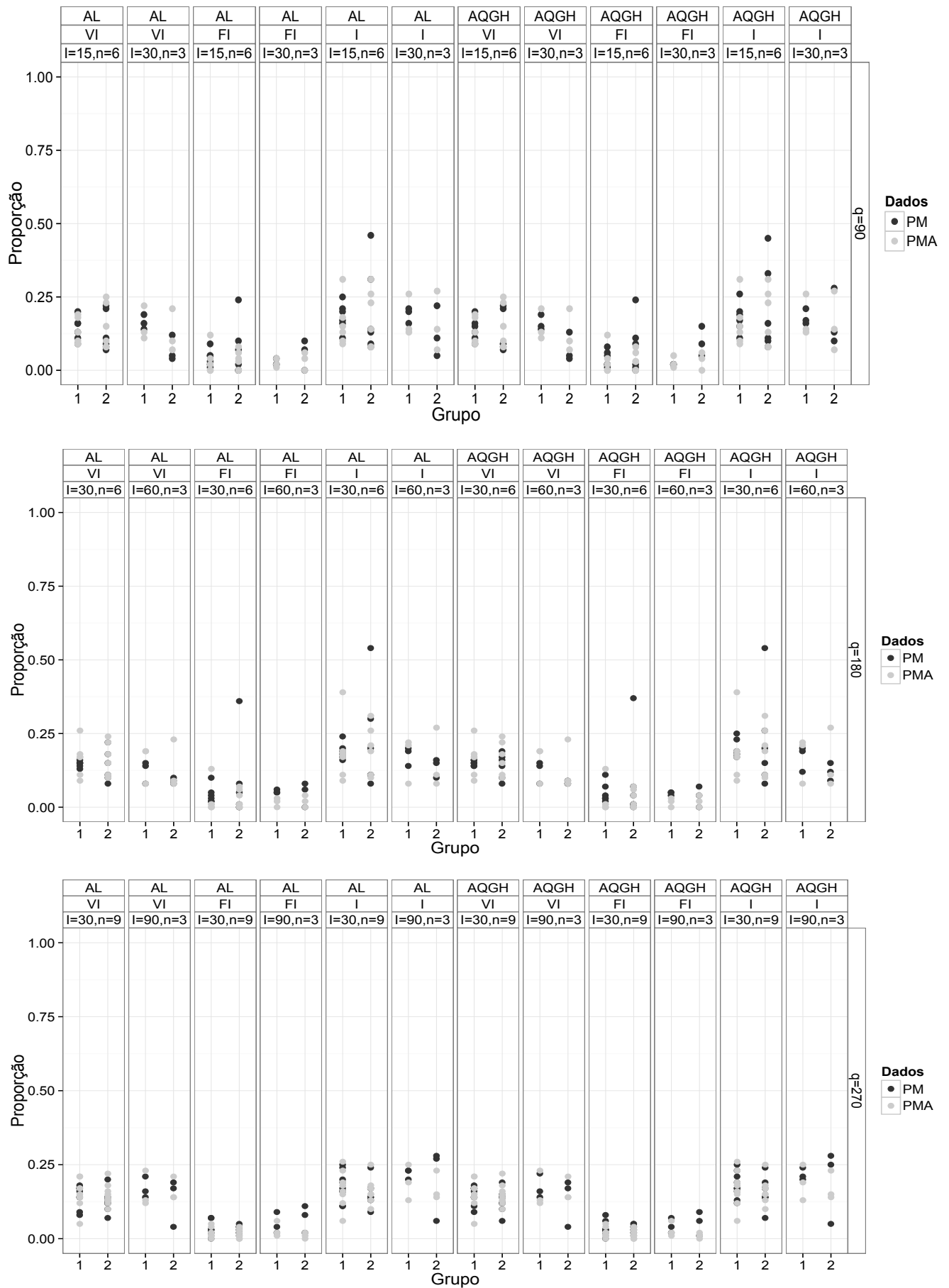


Figura 3.16: Proporções de VI, FI e I nos grupos 1 e 2 para os dados PM e PMA no caso em que para um valor fixado de q consideram-se diferentes valores de l e n_i pelos métodos AL e AQGH, com $\omega = 0,95$. (Estudo de simulação 2) (MRLEM)

3.4 Aplicações

Nesta seção ilustramos a metodologia desenvolvida mediante três aplicações a conjuntos de dados reais. Na aplicação 1 analisamos um conjunto de dados referentes ao comportamento reprodutivo da espécie de aracnídeos opilião do Brasil apresentado por [Werneck \(2012\)](#). Na aplicação 2 nos referimos a um conjunto de dados relacionados ao estudo da germinação de sementes apresentado por [Crowder \(1978\)](#) e abordado frequentemente na literatura. Por último, na aplicação 3 estudamos um conjunto de dados relacionados ao estudo da infecção pelo HPV (*Human papillomavirus*) em mulheres chilenas apresentado por [Ferrecio et al. \(2011\)](#).

3.4.1 Aplicação 1

Uma pesquisa desenvolvida no Instituto de Biociências da Universidade de São Paulo ([Werneck, 2012](#)) teve como objetivo estudar a competição intrasexual dos machos da espécie de aracnídeos opilião do Brasil. Para isto foram montados 14 terrários em que sete deles possuíam quatro ninhos (baixa disponibilidade de sítios reprodutivos) e os outros sete possuíam oito ninhos (alta disponibilidade de sítios reprodutivos). Em cada um dos 14 terrários foram colocados 12 machos e 12 fêmeas. A condição de baixa ou alta disponibilidade de ninhos está relacionada à competição. Assim, baixa disponibilidade de ninhos indica uma competição mais intensa entre os machos para conseguir ninho e alta disponibilidade de ninhos uma competição menos intensa. Por outro lado, a premissa é que o tamanho do macho é um indicativo de qualidade desse macho, isto é, machos maiores são de melhor qualidade e por isso serão escolhidos pelas fêmeas para se reproduzir e ter ovo. Na condição de baixa disponibilidade de ninhos, em que a competição é mais intensa, espera-se que o tamanho do macho seja ainda mais importante para se reproduzir e ter ovos do que em uma condição de alta disponibilidade de ninhos em que a competição é menos intensa. Cabe destacar que de acordo com o planejamento do estudo, tem-se um conjunto de dados desbalanceados.

Para este problema resulta adequado considerar um MRLEM cuja variável aleatória Y_{ij} assume o valor 1, com probabilidade p_{ij} , quando a fêmea coloca ovo e o valor 0, com probabilidade $1 - p_{ij}$, quando a fêmea não coloca ovo. Além disso, o efeito aleatório é definido pelos terrários e , portanto, pelo intercepto aleatório b_i para o qual assume-se uma distribuição normal de media 0 e variância σ^2 . Assim, o MRLEM é dado por

$$p(y_{ij}|b_i) = \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right]$$

e pelo seguinte componente sistemático

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (3.31)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 14$ ($q=63$), em que x_{1ij} é a variável indicadora da baixa (1) ou alta (0) disponibilidade de ninhos, x_{2ij} é a variável correspondente ao tamanho dos machos em milímetros e β_0, β_1 e β_2 são os coeficientes de regressão fixos. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \sigma^2)'$.

Primeiramente, realizamos o ajuste do modelo obtendo os resultados das estimativas de máxima verossimilhança dos parâmetros e das predições BLUP (*Best linear unbiased prediction*) (vide, por exemplo, McCulloch e Searle, 2001) do intercepto aleatório pelos métodos AL e AQGH (com cinco pontos) os quais encontram-se nas Tabelas 3.2 e 3.3, respectivamente. Na Tabela 3.2 observamos que pelos métodos AL e AQGH, com um nível de significância de 5%, não existem coeficientes significativos e que o desvio padrão do intercepto aleatório é próximo de 1. Na Tabela 3.3 observamos que as predições BLUP do intercepto aleatório são similares entre os métodos. Além disso, nas Figuras 3.17a e 3.17b, correspondentes aos gráficos normais de probabilidades com envelope simulado, observamos que os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório localizam-se totalmente no interior do envelope de modo que o ajuste do modelo parece ser razoável por ambos os métodos.

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-0,761	3,300	0,818
	β_1	0,426	0,715	0,551
	β_2	0,133	0,492	0,788
	σ	0,838		
AQGH	β_0	-0,774	3,337	0,817
	β_1	0,429	0,748	0,566
	β_2	0,136	0,498	0,785
	σ	0,916		

Tabela 3.2: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.31).

Em segundo lugar, para realizar a análise de diagnóstico de influência local com respeito aos valores da variável resposta iguais a 0, amostramos $\{b_i^{(s)}; s = 1, \dots, S, i = 1, \dots, 14\}$, com $S = 10.000$, da função de densidade dada por

$$p(b_i | \mathbf{Y}_o, \hat{\boldsymbol{\psi}}) \propto \exp\left\{-\frac{1}{2}b_i^2\hat{\sigma}^{-2} + \sum_{j=1}^{n_i} y_{ij}(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i) - \log[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i)]\right\},$$

Terrário	Método	
	AL	AQGH
1	-0,639	-0,713
2	-0,250	-0,275
3	-0,255	-0,289
4	-0,081	-0,093
5	0,746	0,830
6	0,502	0,573
7	0,181	0,197
8	-0,048	-0,057
9	-1,040	-1,163
10	0,298	0,321
11	0,225	0,261
12	-0,750	-0,825
13	0,660	0,739
14	0,302	0,325

Tabela 3.3: Predições BLUP do intercepto aleatório para os terrários pelos métodos AL e AQGH relacionados ao modelo (3.31).

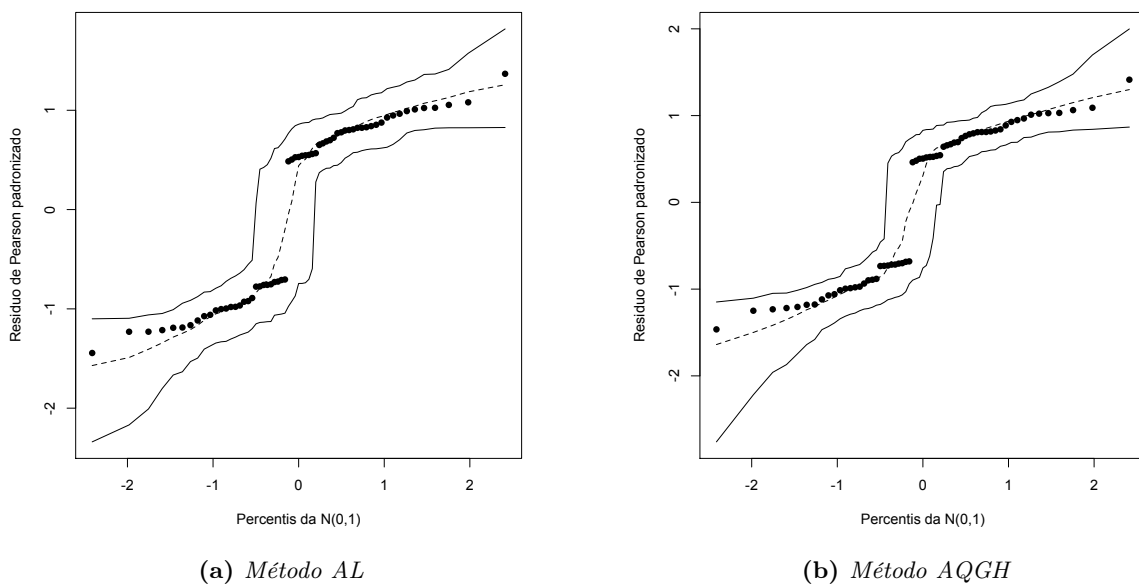


Figura 3.17: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.31).

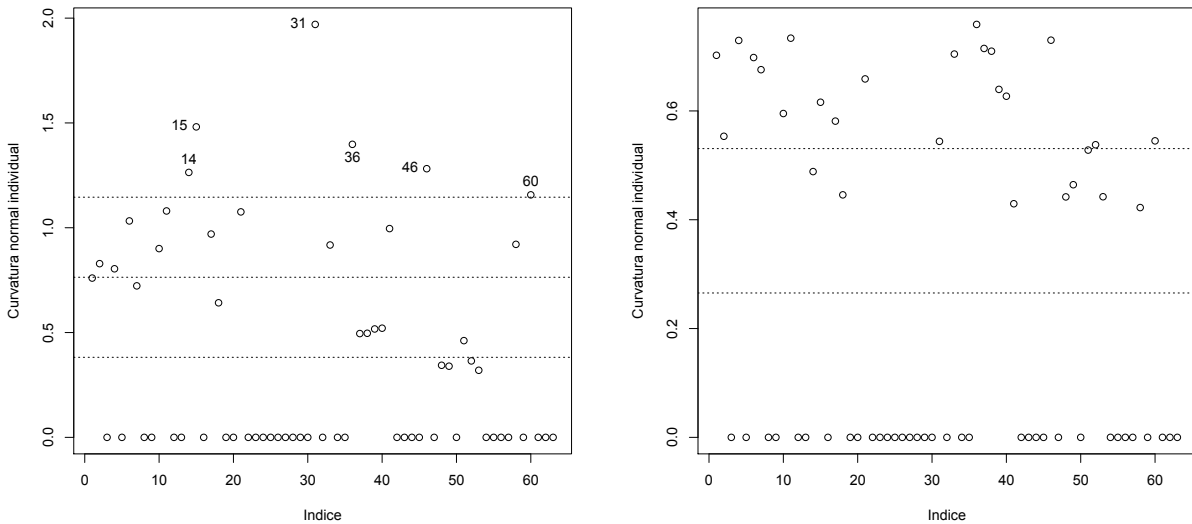
mediante o algoritmo Metropolis-Hastings definido na Seção 3.2.4, com as estimativas dos parâmetros obtidas tanto pelo método AL quanto pelo método AQGH. Os resultados das observações amostradas mostram-se nos histogramas e funções de densidades estimadas nas Figuras B.1 e B.2 do Apêndice B, respectivamente. Além disso, na Tabela B.1 observamos que as predições do intercepto aleatório baseadas na amostragem são similares entre os métodos e também similares às predições BLUP nos métodos respectivos (vide Tabela 3.3). A seguir, a partir das $S = 10.000$ observações, descartamos as $M_0 = 1.000$ primeiras e calculamos as matrizes $\ddot{Q}_\psi(\hat{\psi})$, $\Delta\omega_{0_{pm}}$ e $\Delta\omega_{0_{pma}}$. Posteriormente, calculamos as curvaturas normais individuais e os intervalos de influência sob os esquemas de perturbação multiplicativa e perturbação multiplicativa apropriada e realizamos os gráficos de diagnóstico correspondentes (vide Figura 3.18). Para realizar a análise de diagnóstico de influência local com respeito aos valores da variável resposta iguais a 1, alternamos os valores da variável resposta e realizamos o mesmo procedimento descrito anteriormente (vide Figura 3.19).

Sob o esquema de perturbação multiplicativa pelo método AL, para o caso em que a variável resposta assume valores iguais a 0, a Figura 3.18a mostra que as observações #14 e #15 do terrário 3, #31 do terrário 7, #36 do terrário 8, #46 do terrário 10 e #60 do terrário 14 foram detectadas como pontos influentes. Para o caso em que a variável resposta assume valores iguais a 1, a Figura 3.19a mostra que as observações #3 do terrário 1, #8 do terrário 2, #16 e #19 do terrário 4, #34 do terrário 8 e #50 do terrário 12 foram detectadas como pontos influentes. Por outro lado, sob o esquema de perturbação multiplicativa apropriada pelo método AL, as Figuras 3.18b e 3.19b mostram que não houve detecção de pontos influentes. O mesmos resultados obtiveram-se pelo método AQGH (vide Figura 3.20a, 3.21a, 3.20b e 3.21b).

Para avaliar a magnitude do impacto exercido sobre as estimativas dos parâmetros pelas observações influentes e conjuntos de observações influentes (quando estas pertencem ao mesmo terrário), isto é, #3 do terrário 1, #8 do terrário 2, {#14, #15} do terrário 3, {#16, #19} do terrário 4, #31 do terrário 7, {#34, #36} do terrário 8, #46 do terrário 10, #50 do terrário 12 e #60 do terrário 14, calculamos o desvio relativo percentual da seguinte maneira. Para $\psi = (\psi_1, \psi_2, \psi_3, \psi_4)' = (\beta_0, \beta_1, \beta_2, \sigma)'$ o desvio relativo percentual é dado por

$$DRP_j = \frac{\hat{\psi}_j - \hat{\psi}_j^*}{\hat{\psi}_j} \times 100\%,$$

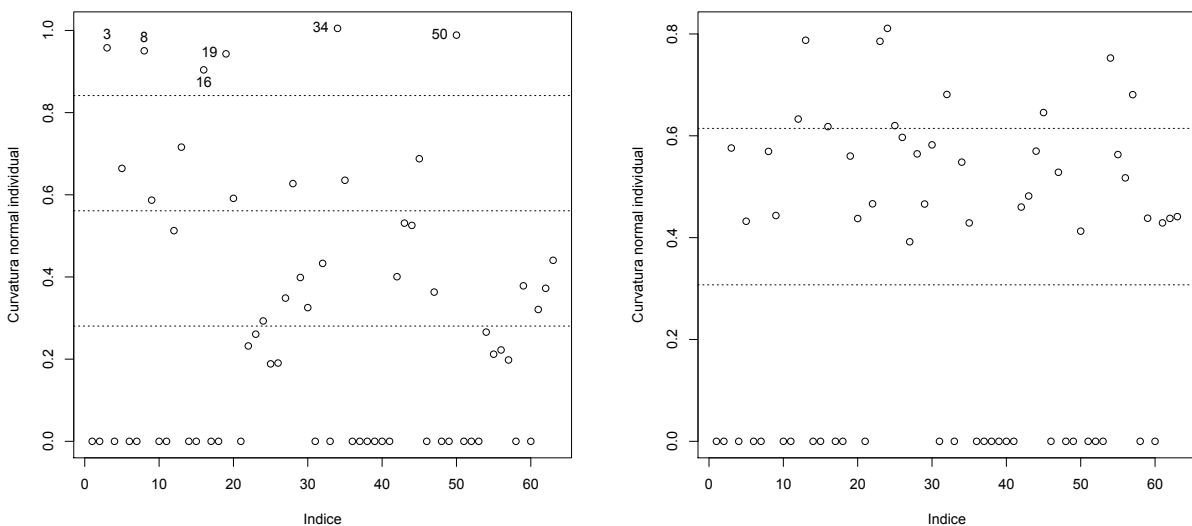
em que $\hat{\psi}_j$ é a estimativa obtida do ajuste do modelo com todas as observações e $\hat{\psi}_j^*$ é a estimativa obtida do ajuste do modelo excluindo as combinações de observações influentes e conjuntos de



(a) Esquema de perturbação multiplicativa

(b) Esquema de perturbação multiplicativa apropriada

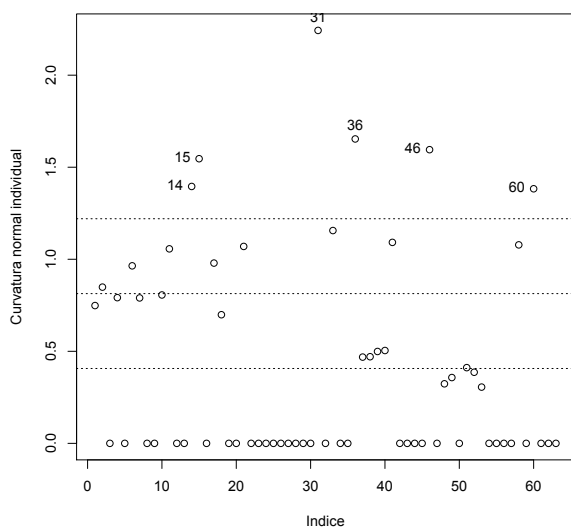
Figura 3.18: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.31).



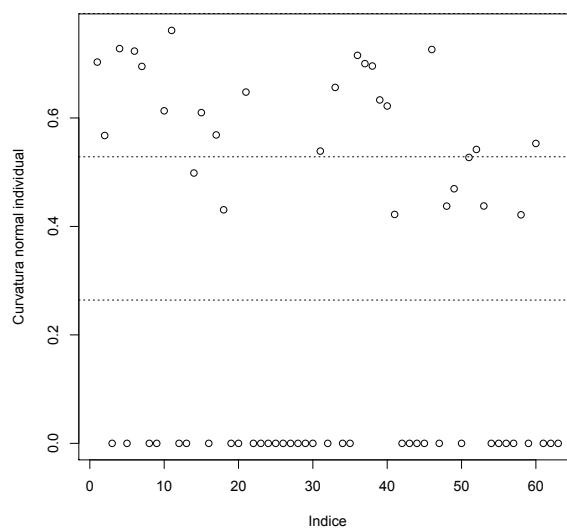
(a) Esquema de perturbação multiplicativa

(b) Esquema de perturbação multiplicativa apropriada

Figura 3.19: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.31).

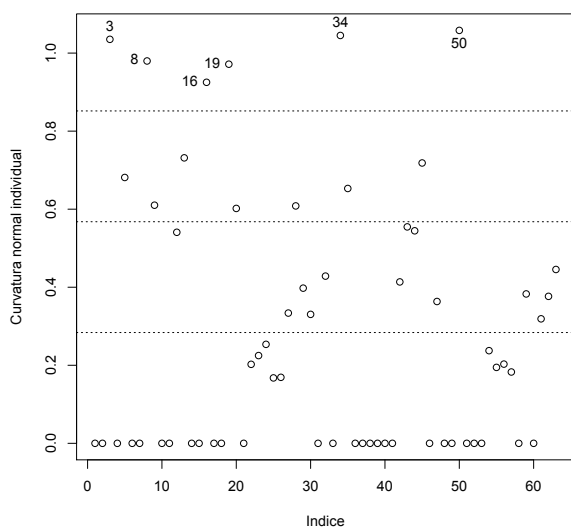


(a) Esquema de perturbação multiplicativa

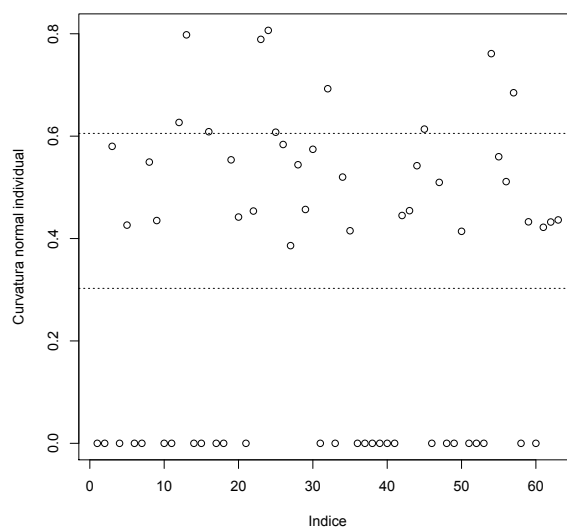


(b) Esquema de perturbação multiplicativa apropriada

Figura 3.20: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.31).



(a) Esquema de perturbação multiplicativa



(b) Esquema de perturbação multiplicativa apropriada

Figura 3.21: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.31).

observações influentes, para $j = 1, \dots, 4$. A Tabela 3.4 mostra as combinações nas quais encontramos os dois maiores totais de desvios relativos percentuais, considerando como total a soma dos DRP_j , para $j = 1, \dots, 4$. Observamos que pelo método AL, excluindo tanto a combinação $\{\#14, \#15\}$, $\{\#16, \#19\}$, $\#31$, $\{\#34, \#36\}$ e $\#50$ quanto a combinação $\#8$, $\{\#14, \#15\}$, $\{\#16, \#19\}$, $\#31$, $\{\#34, \#36\}$ e $\#50$, as estimativas de β_0 , β_1 , β_2 e σ apresentaram mudanças muito grandes, sendo estas muito sensíveis ao esquema de perturbação multiplicativa. Especificamente, o desvio relativo percentual, em valor absoluto, chegou até 758,280% para a primeira combinação e até 694,036% para a segunda combinação. No entanto, a pesar das grandes variações nas estimativas dos parâmetros não houve mudanças em relação à significância dos coeficientes. Resultados similares observamos pelo método AQGH.

Além disso, a Tabela 3.5 mostra os resultados das medidas de Sen, Esp e ACC para os dados com todas as observações e os dados excluindo as combinações escolhidas. Observamos que para os dados com todas as observações, as três medidas foram de 0,714, e que excluindo cada uma das combinações os valores de Sen, Esp e ACC aumentaram a 0,806, 0,833 e 0,818, respectivamente. Resultados idênticos observamos pelo método AQGH.

Em conclusão, os resultados da análise de diagnóstico de influência local sob o esquema de perturbação multiplicativa, pelos métodos AL e AQGH, levaram a grandes variações nas estimativas dos parâmetros mas não a mudanças inferenciais. Entretanto, a detecção e exclusão de observações influentes permitiram o melhoramento das medidas de predição do modelo.

Esquema de perturbação multiplicativa							
Método	Observações excluídas	Terrários	Parâmetro	Estimativa	Valor-p	DRP_j	
AL	Nenhuma	-	β_0	-0,761	0,818	-	
			β_1	0,426	0,551	-	
			β_2	0,133	0,788	-	
			σ	0,838	-	-	
	{#14, #15}, {#16, #19}, #31, {#34, #36}, #50	3, 4, 7, 8, 12	β_0	-6,529	0,289	-758,280	
			β_1	1,983	0,322	-365,278	
			β_2	0,969	0,277	-630,831	
			σ	2,073	-	-141,389	
	#8, {#14, #15}, {#16, #19}, #31, {#34, #36}, #50	2, 3, 4, 7, 8, 12	β_0	-6,041	0,227	-694,036	
			β_1	2,088	0,151	-389,915	
			β_2	0,884	0,235	-566,704	
			σ	2,112	-	-151,973	
	AQGH	Nenhuma	-	β_0	-0,774	0,817	-
				β_1	0,429	0,566	-
				β_2	0,136	0,785	-
				σ	0,916	-	-
{#14, #15}, {#16, #19}, #31, {#34, #36}, #50		3, 4, 7, 8, 12	β_0	-6,465	0,190	-735,592	
			β_1	1,879	0,190	-337,721	
			β_2	0,959	0,191	-606,623	
			σ	2,093	-	-127,897	
#8, {#14, #15}, {#16, #19}, #31, {#34, #36}, #50		2, 3, 4, 7, 8, 12	β_0	-5,973	0,231	-672,022	
			β_1	1,975	0,175	-360,095	
			β_2	0,873	0,239	-543,654	
			σ	2,129	-	-131,738	

Tabela 3.4: Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.31).

Esquema de perturbação multiplicativa					
Método	Observações excluídas	Terrários	Sen	Esp	ACC
AL	Nenhuma	-	0,714	0,714	0,714
	{#14, #15}, {#16, #19}, #31, {#34, #36}, #50	3, 4, 7, 8, 12	0,806	0,833	0,818
	#8, {#14, #15}, {#16, #19}, #31, {#34, #36}, #50	2, 3, 4, 7, 8, 12	0,806	0,833	0,818
AQGH	Nenhuma	-	0,714	0,714	0,714
	{#14, #15}, {#16, #19}, #31, {#34, #36}, #50	3, 4, 7, 8, 12	0,806	0,833	0,818
	#8, {#14, #15}, {#16, #19}, #31, {#34, #36}, #50	2, 3, 4, 7, 8, 12	0,806	0,833	0,818

Tabela 3.5: Medidas de Sen, Esp e ACC para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.31).

3.4.2 Aplicação 2

Um conjunto de dados correspondentes ao estudo da germinação de sementes em 21 bandejas foram apresentados por Crowder (1978). Os tratamentos foram os tipos de semente *Orobanchae aegyptiaco* 75 (OA75) e *Orobanchae aegyptiaco* 73 (OA73) e dois meios de germinação provenientes de extratos de raiz feijão (F) e pepino (P). Das 21 bandejas cinco foram de semente OA75 e extrato P, seis de semente OA75 e extrato P, cinco de semente OA73 e extrato F e seis de semente OA73 e extrato P. Para cada bandeja o número de sementes foi diferente, constituindo assim um conjunto de dados desbalanceados.

Consideramos um MRLEM cuja variável aleatória Y_{ij} assume o valor 1, com probabilidade p_{ij} , quando a semente germina e o valor 0, com probabilidade $1 - p_{ij}$, quando a semente não germina. O efeito aleatório é definido pelas bandejas e , portanto, pelo intercepto aleatório b_i para o qual assume-se uma distribuição normal de media 0 e variância σ^2 . Logo, o MRLEM é dado por

$$p(y_{ij}|b_i) = \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right],$$

em que

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (3.32)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 21$ ($q=831$), com x_{1ij} a variável indicadora do tipo de semente, isto é, 0 para OA75 e 1 OA73, x_{2ij} a variável indicadora do tipo de extrato, isto é, 0 para F e 1 para P e β_0, β_1 e β_2 os coeficientes de regressão. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \sigma^2)'$.

As estimativas dos parâmetros e as predições BLUP do intercepto aleatório pelos métodos AL e AQGH são apresentadas nas Tabelas 3.6 e 3.7, respectivamente. Observamos que com um nível de significância de 5% o coeficiente da covariável extrato é significativo e que o desvio padrão do intercepto aleatório é próximo de 0,295, por ambos os métodos. Na Tabela 3.7 observamos que as predições BLUP do intercepto aleatório são similares entre os métodos AL e AQGH. Observamos também mediante as Figuras 3.22a e 3.22b que os resíduos de Pearson padronizados localizam-se totalmente no interior do envelope do modo que o ajuste do modelo parece ser razoável, por ambos os métodos.

Para a análise de diagnóstico de influência local com respeito aos valores da variável resposta

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-1,072	0,428	0,012
	β_1	-0,346	0,209	0,098
	β_2	1,029	0,204	<0,001
	σ	0,293		
AQGH	β_0	-1,071	0,429	0,013
	β_1	-0,347	0,210	0,098
	β_2	1,029	0,204	<0,001
	σ	0,295		

Tabela 3.6: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.32).

Bandeja	Método	
	AL	AQGH
1	-0,278	-0,280
2	-0,077	-0,078
3	-0,320	-0,322
4	0,224	0,225
5	0,059	0,059
6	0,082	0,084
7	0,163	0,164
8	0,290	0,291
9	-0,060	-0,060
10	-0,191	-0,192
11	0,102	0,104
12	0,184	0,186
13	0,015	0,015
14	-0,061	-0,061
15	0,381	0,384
16	-0,104	-0,105
17	-0,265	-0,268
18	-0,069	-0,069
19	-0,115	-0,116
20	0,116	0,117
21	-0,076	-0,076

Tabela 3.7: Predições BLUP do intercepto aleatório para as bandejas pelos métodos AL e AQGH relacionados ao modelo (3.32).

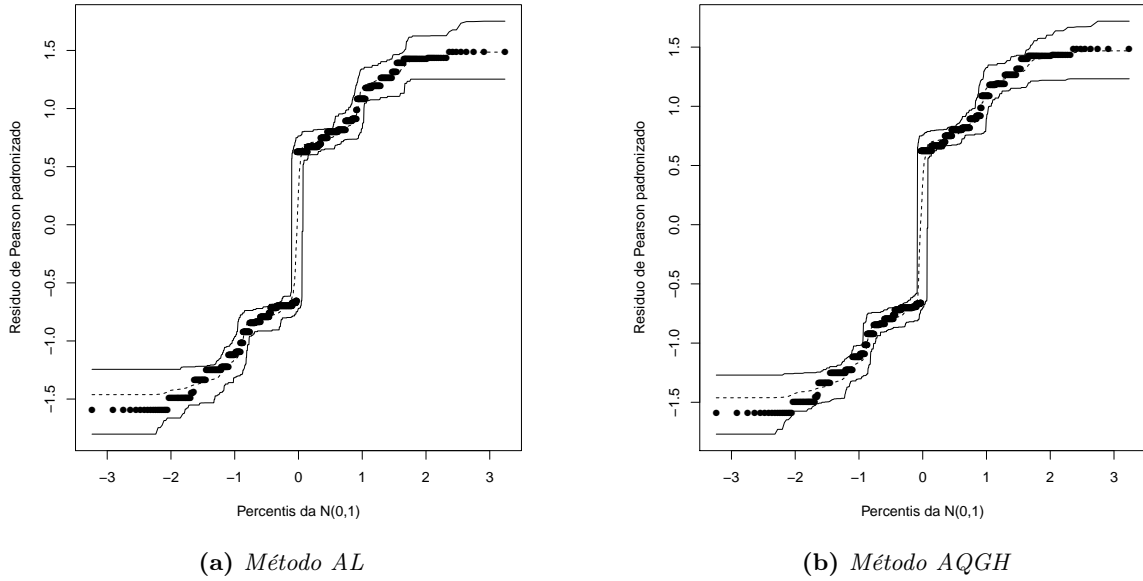


Figura 3.22: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as previsões BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.32).

iguais a 0, amostramos $\{b_i^{(s)}; s = 1, \dots, S, i = 1, \dots, 21\}$, com $S = 10.000$ da função de densidade

$$p(b_i | \mathbf{Y}_o, \hat{\boldsymbol{\psi}}) \propto \exp \left\{ -\frac{1}{2} b_i^2 \hat{\sigma}^{-2} + \sum_{j=1}^{n_i} y_{ij} (\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i) - \log[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i)] \right\},$$

mediante o algoritmo Metropolis-Hastings, com as estimativas dos parâmetros obtidas por ambos os métodos. Os resultados mostram-se nas Figuras B.3 e B.4 do Apêndice B para os métodos AL e AQGH, respectivamente. Além disso, na Tabela B.2 observamos que as previsões do intercepto aleatório baseadas na amostragem são similares entre os métodos. No entanto, ao comparar estas com as previsões BLUP nos métodos respectivos (vide Tabela 3.7) são diferentes. Baseados nas $S = 10.000$ observações, descartamos as $M_0 = 1.000$ primeiras e calculamos as matrizes $\ddot{Q}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$, $\Delta\omega_{0_{pm}}$ e $\Delta\omega_{0_{pma}}$. A seguir obtemos para cada um dos esquemas as curvaturas normais individuais e os intervalos de influência com os quais realizamos os gráficos de diagnóstico respectivos (vide Figura 3.23 e 3.25). Para os valores da variável resposta iguais a 1, alternamos os valores da variável resposta e realizamos o mesmo procedimento descrito anteriormente (vide Figura 3.24 e 3.26).

Sob o esquema de perturbação multiplicativa, pelo método AL, para o caso em que a variável resposta assume valores iguais a 0, a Figura 3.23a mostra que as observações #332 a #352 da bandeja 7, #408 a #424 da bandeja 8 e #522 a #554 da bandeja 10 foram detectadas como pontos influentes. Para o caso em que a variável resposta assume valores iguais a 1, a Figura 3.24a mostra

que as observações #40 a #62 da bandeja 2 e #102 a #124 da bandeja 3 foram detectadas como pontos influentes.

Sob o esquema de perturbação multiplicativa apropriada, pelo método AL, para o caso em que a variável resposta assume valores iguais a 0, a Figura 3.23b mostra que não houve detecção de pontos influentes e para o caso em que a variável resposta assume valores iguais a 1, a Figura 3.24b mostra que as observações #568 a #575 da bandeja 12, #584 a #593 da bandeja 13, #614 a #621 da bandeja 14 e #642 a #664 da bandeja 15 foram detectadas como pontos influentes.

Resultados idênticos observamos pelo método AQGH (vide Figura 3.25a, 3.26a, 3.25b e 3.26b).

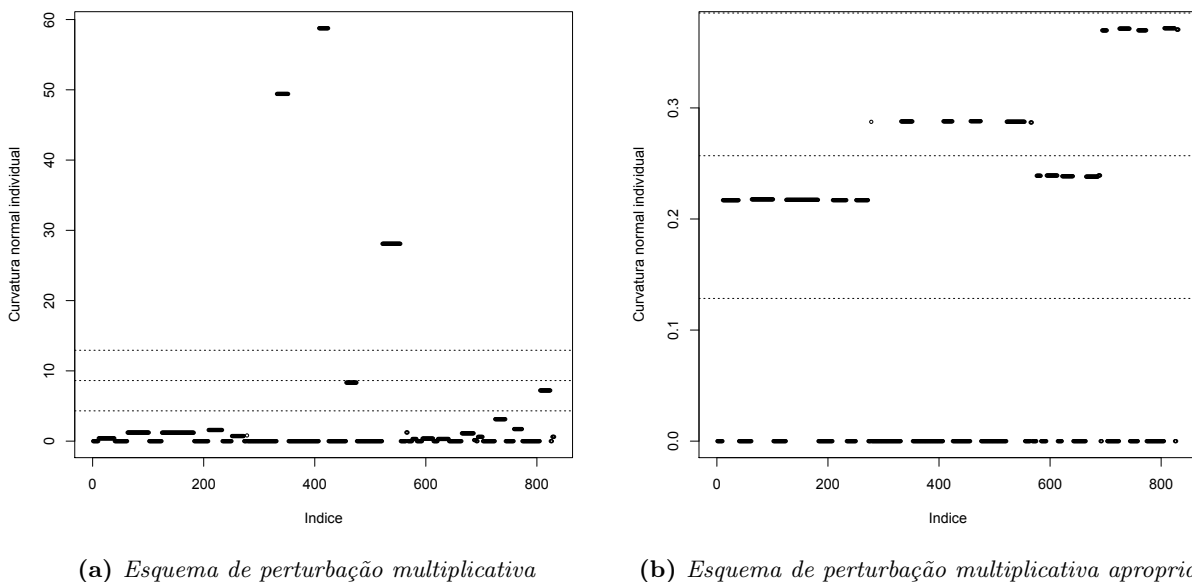
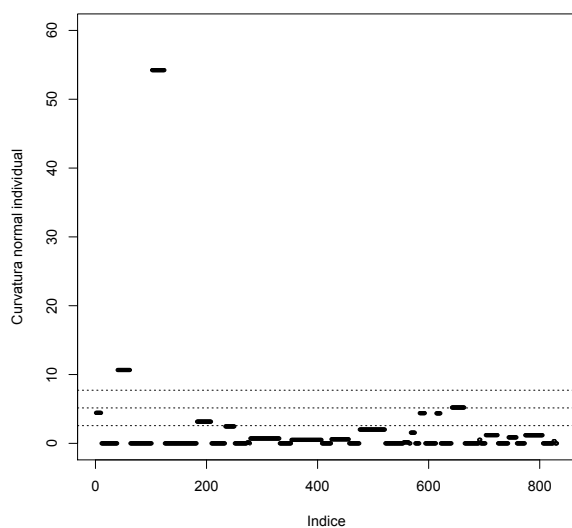
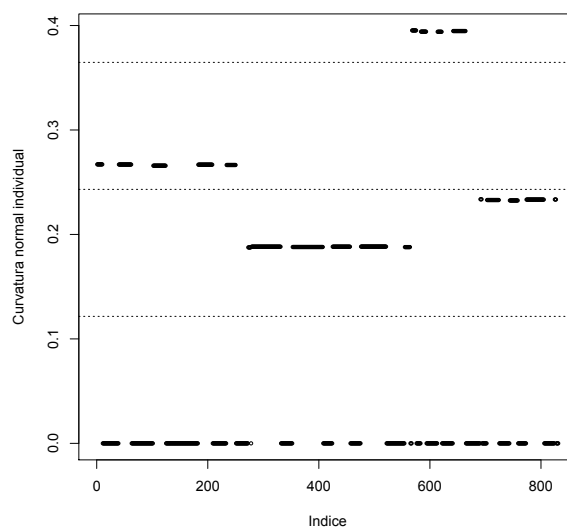


Figura 3.23: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.32).

De maneira similar à Aplicação 1, calculamos o desvio relativo percentual excluindo as combinações de conjuntos de observações influentes. Nas Tabelas 3.8 e 3.9 encontram-se as combinações nas quais se teve os dois maiores totais de desvios relativos percentuais para os esquemas de perturbação multiplicativa e multiplicativa apropriada, respectivamente. Observamos que para o esquema de perturbação multiplicativa, pelo método AL, ao excluir as duas combinações escolhidas, as estimativas de β_0 , β_1 , β_2 e σ apresentaram uma mudança muito grande sendo estas muito sensíveis ao esquema de perturbação multiplicativa. Observamos também que excluindo os conjuntos de observações destacadas em negrito, o coeficiente β_0 deixou de ser significativo e o coeficiente β_1

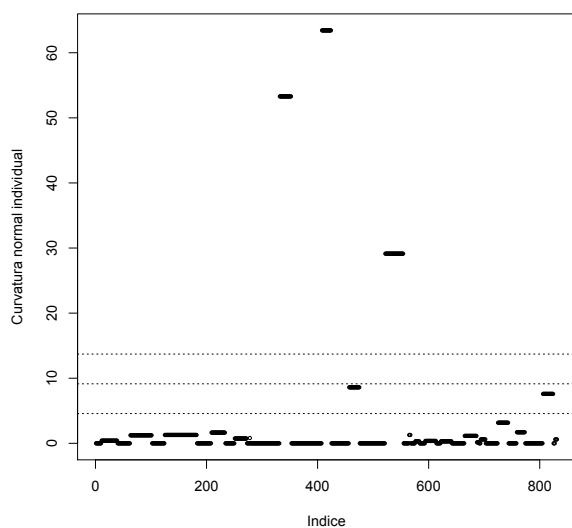


(a) Esquema de perturbação multiplicativa

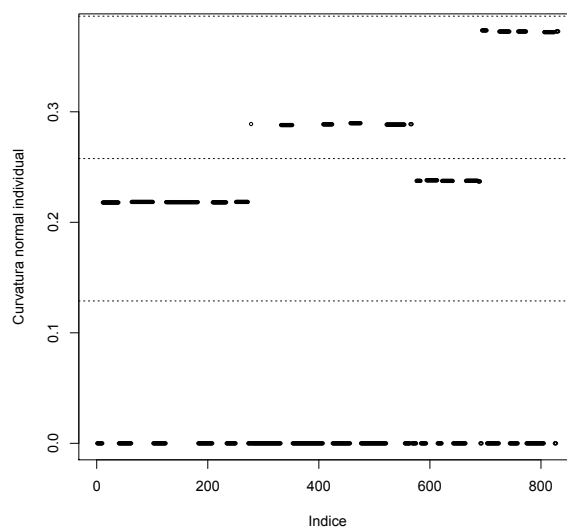


(b) Esquema de perturbação multiplicativa apropriada

Figura 3.24: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.32).

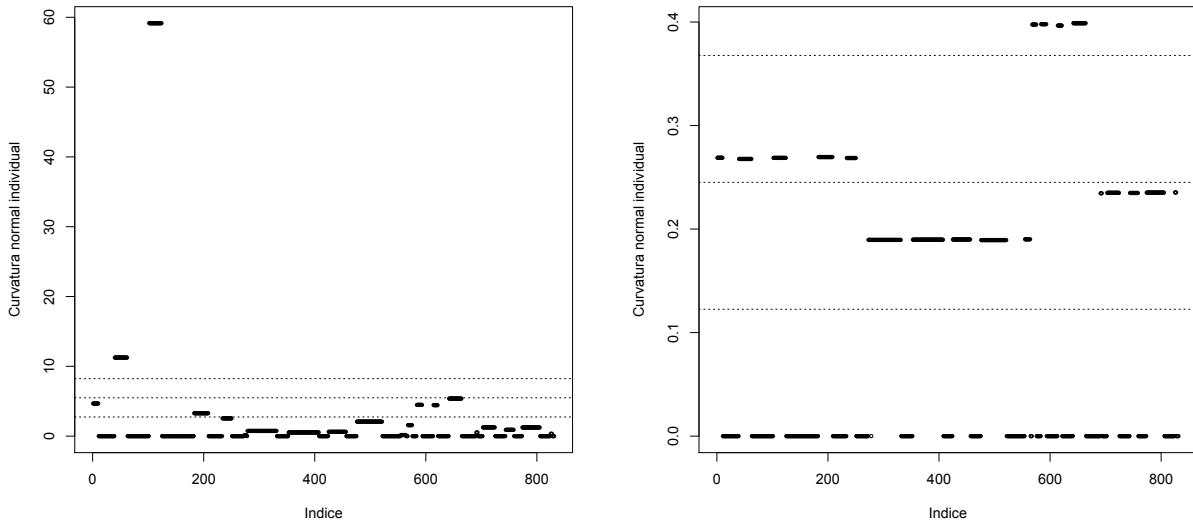


(a) Esquema de perturbação multiplicativa



(b) Esquema de perturbação multiplicativa apropriada

Figura 3.25: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.32).



(a) Esquema de perturbação multiplicativa

(b) Esquema de perturbação multiplicativa apropriada

Figura 3.26: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.32).

ficou muito próximo do nível de significância de 5%. Resultados similares observamos pelo método AQGH.

Com respeito ao esquema de perturbação multiplicativa apropriada, pelo método AL, observamos que ao excluir as combinações destacadas em negrito, as estimativas de β_1 , β_2 e σ apresentaram uma mudança muito grande sendo estas muito sensíveis a este esquema. Além disso, observamos que para as duas combinações o valor-p associado a β_1 foi menor a 0,001 tornando-se o coeficiente da covariável tipo de semente significativa. Resultados similares observamos pelo método AQGH.

Por último, nas Tabelas 3.10 e 3.11 encontram-se os resultados das medidas de Sen, Esp e ACC para os dados com todas as observações e os dados excluindo as combinações mencionadas anteriormente para os esquemas de perturbação multiplicativa e multiplicativa apropriada, respectivamente. Observamos que para os dados com todas as observações, as medidas de Sen, Esp e ACC foram de 0,644, 0,629 e 0,637, respectivamente. Para o esquema de perturbação multiplicativa, pelo método AL, observamos que excluindo as combinações escolhidas as medidas de Sen, Esp e ACC aumentaram chegando até 0,783, 0,699 e 0,744, respectivamente. Resultados idênticos observamos pelo método AQGH.

Para o esquema de perturbação multiplicativa apropriada, pelo método AL, observamos que excluindo as combinações escolhidas as medidas de Sen, Esp e ACC aumentaram levemente chegando

até 0,680, 0,676 e 0,678, respectivamente. Resultados idênticos observamos pelo método AQGH.

Em conclusão, os resultados da análise de diagnóstico de influência local, tanto para o esquema de perturbação multiplicativa quanto para o esquema de perturbação multiplicativa apropriada, pelos métodos AL e AQGH, conduziram a alterações grandes nas estimativas dos parâmetros. No entanto, para o esquema de perturbação multiplicativa apropriada, estes resultados foram importantes pois levaram a mudanças relacionadas à significância do coeficiente de regressão da covariável tipo de semente. Para finalizar, a detecção e exclusão de observações influentes implicaram um aumento das medidas de predição do modelo, o qual foi maior sob o esquema de perturbação multiplicativa do que sob o esquema de perturbação multiplicativa apropriada.

Esquema de perturbação multiplicativa							
Método	Observações excluídas	Bandejas	Parâmetro	Estimativa	Valor-p	DRP_j	
AL	Nenhuma	-	β_0	-1,072	0,012	-	
			β_1	-0,346	0,098	-	
			β_2	1,029	<0,001	-	
			σ	0,293	-	-	
	{#40,...,#62}, {#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	2, 3, 7, 8, 10	β_0	-2,910	0,146	-171,449	
			β_1	-1,178	0,199	-240,451	
			β_2	3,138	<0,001	-204,923	
			σ	1,948	-	-564,981	
	{#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	3, 7, 8, 10	β_0	-1,438	0,369	-34,191	
			β_1	-1,435	0,055	-314,900	
			β_2	2,541	<0,001	-146,981	
			σ	1,573	-	-436,666	
	AQGH	Nenhuma	-	β_0	-1,071	0,013	-
				β_1	-0,347	0,098	-
				β_2	1,029	<0,001	-
				σ	0,295	-	-
{#40,...,#62}, {#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}		2, 3, 7, 8, 10	β_0	-2,904	0,149	-171,247	
			β_1	-1,164	0,207	-235,892	
			β_2	3,120	<0,001	-203,318	
			σ	1,960	-	-564,050	
{#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}		3, 7, 8, 10	β_0	-1,440	0,371	-34,488	
			β_1	-1,428	0,057	-311,948	
			β_2	2,534	<0,001	-146,359	
			σ	1,583	-	-436,480	

Tabela 3.8: Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.32).

Esquema de perturbação multiplicativa apropriada						
Método	Observações excluídas	Bandejas	Parâmetro	Estimativa	Valor-p	DRP_j
AL	Nenhuma	-	β_0	-1,072	0,012	-
			β_1	-0,346	0,098	-
			β_2	1,029	<0,001	-
			σ	0,293	-	-
	{#568,...,#575}, {#584,...,#593}, {#614,...,#621}, {#642,...,#664}	12, 13, 14, 15	β_0	-1,309	0,024	-22,075
			β_1	-1,576	<0,001	-355,572
			β_2	2,020	<0,001	-96,283
			σ	0,480	-	-63,851
	{#584,...,#593}, {#614,...,#621}, {#642,...,#664}	13, 14, 15	β_0	-1,202	0,051	-12,101
			β_1	-1,243	<0,001	-259,229
			β_2	1,729	<0,001	-67,991
			σ	0,527	-	-79,990
AQGH	Nenhuma	-	β_0	-1,071	0,013	-
			β_1	-0,347	0,098	-
			β_2	1,029	<0,001	-
			σ	0,295	-	-
	{#568,...,#575}, {#584,...,#593}, {#614,...,#621}, {#642,...,#664}	12, 13, 14, 15	β_0	-1,309	0,024	-22,278
			β_1	-1,578	<0,001	-355,146
			β_2	2,021	<0,001	-96,489
			σ	0,482	-	-63,392
	{#584,...,#593}, {#614,...,#621}, {#642,...,#664}	13, 14, 15	β_0	-1,202	0,052	-12,241
			β_1	-1,245	<0,001	-259,113
			β_2	1,730	<0,001	-68,188
			σ	0,532	-	-80,137

Tabela 3.9: Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.32).

Esquema de perturbação multiplicativa					
Método	Observações excluídas	Bandejas	Sen	Esp	ACC
AL	Nenhuma	-	0,644	0,629	0,637
	{#40,...,#62}, {#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	2, 3, 7, 8, 10	0,783	0,699	0,744
	{#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	3, 7, 8, 10	0,738	0,699	0,744
AQGH	Nenhuma	-	0,644	0,629	0,637
	{#40,...,#62}, {#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	2, 3, 7, 8, 10	0,783	0,699	0,744
	{#102,...,#124}, {#332,...,#352}, {#408,...,#424}, {#522,...,#554}	3, 7, 8, 10	0,738	0,699	0,744

Tabela 3.10: Medidas de *Sen*, *Esp* e *ACC* para o esquema de perturbação multiplicativa pelos métodos *AL* e *AQGH* relacionados ao modelo (3.32).

Esquema de perturbação multiplicativa apropriada					
Método	Observações excluídas	Bandejas	Sen	Esp	ACC
AL	Nenhuma	-	0,644	0,629	0,637
	{#568,...,#575}, {#584,...,#593}, {#614,...,#621}, {#642,...,#664}	12, 13, 14, 15	0,680	0,676	0,678
	{#584,...,#593}, {#614,...,#621}, {#642,...,#664}	13, 14, 15	0,666	0,676	0,671
AQGH	Nenhuma	-	0,644	0,629	0,637
	{#568,...,#575}, {#584,...,#593}, {#614,...,#621}, {#642,...,#664}	12, 13, 14, 15	0,680	0,676	0,678
	{#584,...,#593}, {#614,...,#621}, {#642,...,#664}	13, 14, 15	0,666	0,676	0,671

Tabela 3.11: Medidas de *Sen*, *Esp* e *ACC* para o esquema de perturbação multiplicativa apropriada pelos métodos *AL* e *AQGH* relacionados ao modelo (3.32).

3.4.3 Aplicação 3

Um conjunto de dados referentes ao estudo da infecção pelo HPV (*Human Papillomavirus*) em mulheres chilenas de 17 anos ou mais da cidade de Santiago no ano 2006 foi apresentado por Ferreccio *et al.* (2011). Com objetivo de explicar a infecção pelo HPV o estudo considerou as seguintes variáveis: HPV: presença (1) ou ausência (0) do vírus, escolaridade: 1 (menor ou igual a 3 anos), 2 (de 4 a 7 anos), 3 (de 8 a 11 anos) e 4 (maior ou igual a 12 anos), uso de preservativo: nunca (0), às vezes (1) e sempre (2), idade (anos), idade da primeira relação sexual e número de parceiros sexuais (anos).

Para este problema vamos supor um MRLEM cuja variável aleatória Y_{ij} assume o valor 1, com probabilidade p_{ij} , quando há presença do vírus e o valor 0, com probabilidade $1 - p_{ij}$, quando não há presença do vírus. Propomos o intercepto aleatório b_i a partir dos quartis da variável número de parceiros sexuais/(idade - idade da primeira relação sexual) que pode ser considerada como “quantificadora da atividade sexual”, constituindo assim um conjunto de dados desbalanceados, para a qual assume-se uma distribuição normal de media 0 e variância σ^2 . Logo, o MRLEM é dado por

$$p(y_{ij}|b_i) = \exp \left[y_{ij} \log \left(\frac{p_{ij}}{1 - p_{ij}} \right) - \log \left(\frac{1}{1 - p_{ij}} \right) \right],$$

em que

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (3.33)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 4$ ($q=512$), com x_{1ij} a variável indicadora da escolaridade, x_{2ij} a variável indicadora do uso de preservativo e β_0 , β_1 e β_2 os coeficientes de regressão. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \sigma^2)'$.

Nas Tabelas 3.12 e 3.13 observamos que os resultados das estimativas de máxima verossimilhança dos parâmetros e das predições BLUP do intercepto aleatório, respectivamente, são iguais entre os métodos AL e AQGH. Observamos também que com um nível de significância de 5% não existem coeficientes significativos e que o desvio padrão do intercepto aleatório é de 0,216. Além disso, as Figuras 3.27a e 3.27b mostram que não há indícios de afastamentos sérios do modelo adotado, por ambos os métodos.

Realizamos a análise de diagnóstico de influência local da mesma forma que nas aplicações anteriores. Os resultados da amostragem encontram-se nas Figuras B.5 e B.6 do Apêndice B para os métodos AL e AQGH, respectivamente. Além disso, na Tabela B.3 observamos que as predições do intercepto aleatório baseadas na amostragem são similares entre os métodos, mas diferentes às

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-2,140	0,482	<0,001
	β_1	-0,065	0,180	0,716
	β_2	-0,867	0,572	0,129
	σ	0,216		
AQGH	β_0	-2,140	0,482	<0,001
	β_1	-0,065	0,180	0,716
	β_2	-0,867	0,572	0,129
	σ	0,216		

Tabela 3.12: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (3.33).

Grupo	Método	
	AL	AQGH
1	-0,129	-0,129
2	-0,007	-0,007
3	-0,047	-0,047
4	0,199	0,199

Tabela 3.13: Predições BLUP do intercepto aleatório para os grupos pelos métodos AL e AQGH relacionados ao modelo (3.33).

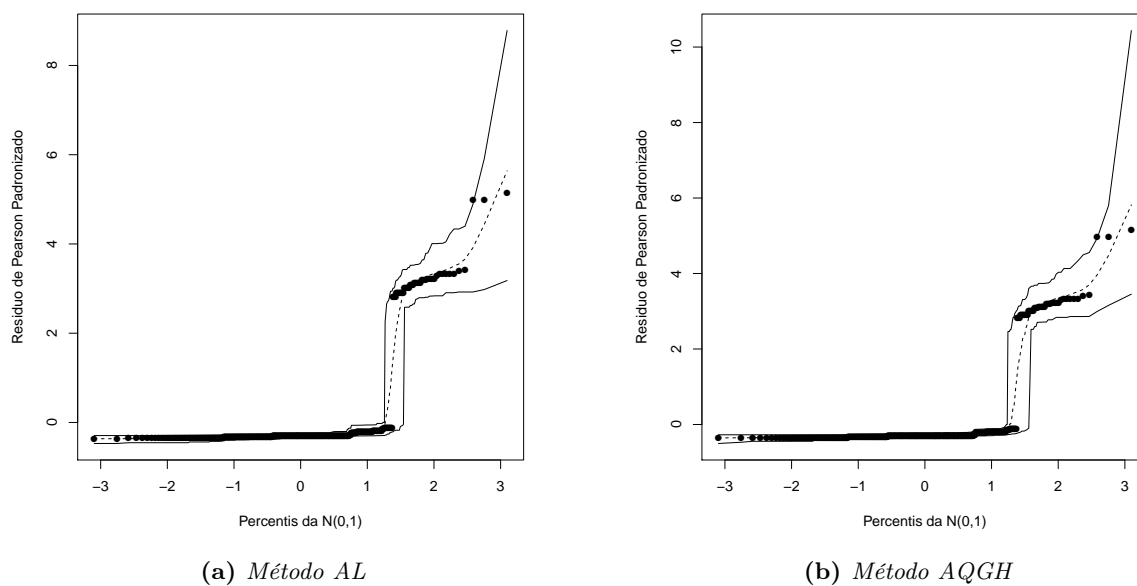


Figura 3.27: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (3.33).

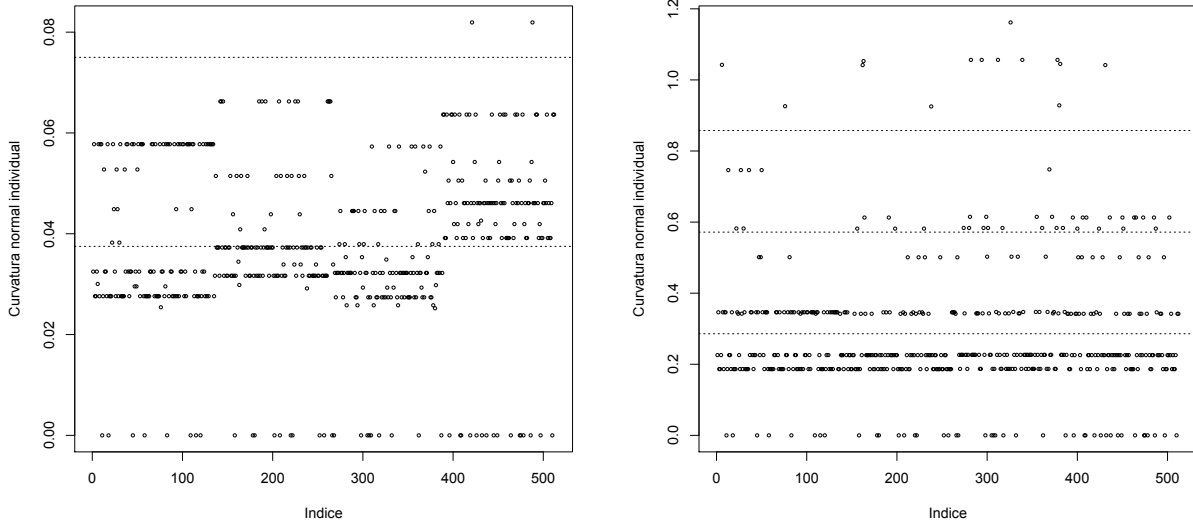
predições BLUP nos métodos respectivos (vide Tabela 3.13).

Sob o esquema de perturbação multiplicativa, pelo método AL, para o caso em que a variável resposta assume valores iguais a 0, a Figura 3.28a mostra que não houve detecção de pontos influentes. Entretanto, para o caso em que a variável resposta assume valores iguais a 1, a Figura 3.29a mostra que todas as observações foram detectadas como pontos influentes, ou seja, #11, #18, #45, #58, #83, #109, #115 e #120 do grupo 1 #158, #178, #180, #202, #208, #220, #222, #252, #257, #266 e #268 do grupo 2, #298, #305, #307, #309, #318, #332, #362 e #387 do grupo 3, #396, #408, #409, #419, #426, #432, #437, #445, #449, #464, #474, #475, #478, #486, #497 e #510 do grupo 4. Por facilidade expressaremos as observações mencionadas como {#11-#120} do grupo 1, {#158-#268} do grupo 2, {#298-#387} do grupo 3 e {#396-#510} do grupo 4, respectivamente. Resultados idênticos observamos pelo método AQGH (vide Figura 3.30a, 3.31a, 3.30b e 3.31b).

Sob o esquema de perturbação multiplicativa apropriada, pelo método AL, para o caso em que a variável resposta assume valores iguais a 0, a Figura 3.28b mostra que as observações #6 e #76 do grupo 1, #162, #163 e #238 do grupo 2, #282, #294, #312, #326, #339, #378, #380 e #381 do grupo 3 e #431 do grupo 4 foram detectadas como pontos influentes. De maneira similar ao outro esquema, para o caso em que a variável resposta assume valores iguais a 1, a Figura 3.29b mostra que todas as observações foram detectadas como pontos influentes. Por facilidade expressaremos as observações mencionadas como {#6-#120} do grupo 1, {#158-#268} do grupo 2, {#282-#387} do grupo 3 e {#396-#510} do grupo 4, respectivamente. Resultados idênticos observamos pelo método AQGH (vide Figura 3.30a, 3.31a, 3.30b e 3.31b).

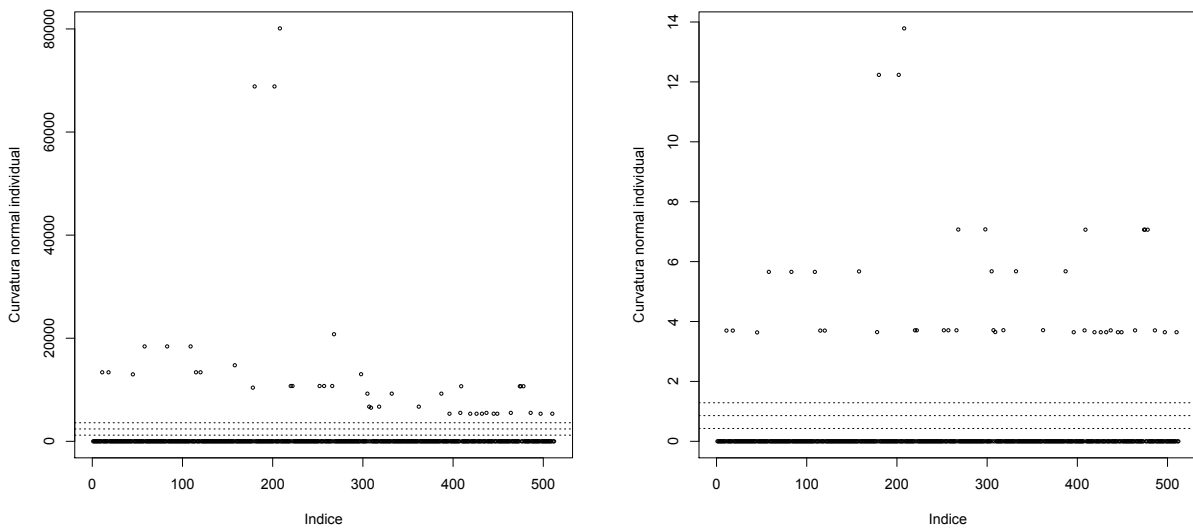
Avaliamos a magnitude do impacto exercido sobre as estimativas dos parâmetros pelas combinações de conjuntos de observações influentes calculando o desvio relativo percentual. Expomos os resultados onde se teve os dois maiores totais de desvios relativos percentuais sem considerar aqueles que excluam todas as observações com variável resposta igual a 1. Os resultados encontram-se nas Tabelas 3.14 e 3.15. Observamos que sob os dois esquemas houve alterações nas estimativas dos parâmetros β_0 , β_1 , β_2 e σ , evidenciados pelos grandes valores dos desvios relativos percentuais. Por outro lado, mesmo que os esquemas de perturbação indicaram conjuntos de observações influentes diferentes os valores das estimativas resultaram similares. É de ressaltar que em ambos os esquemas, no caso do método AL, após a exclusão das observações influentes destacadas em negrito, o intercepto deixou de ser significativo e o valor-p associado a β_1 resultou ser menor a 0,05 tornando-se assim o coeficiente da covariável escolaridade significativa.

Finalmente, nas Tabelas 3.16 e 3.17 encontram-se os resultados das medidas de Sen, Esp e



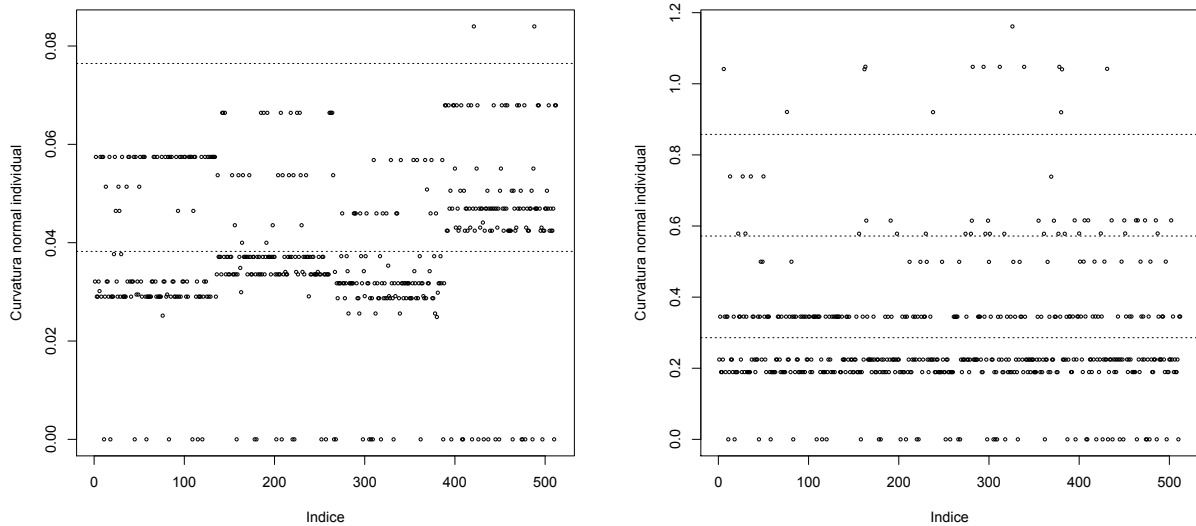
(a) Esquema de perturbação multiplicativa (b) Esquema de perturbação multiplicativa apropriada

Figura 3.28: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 0 relacionados ao modelo (3.33).



(a) Esquema de perturbação multiplicativa (b) Esquema de perturbação multiplicativa apropriada

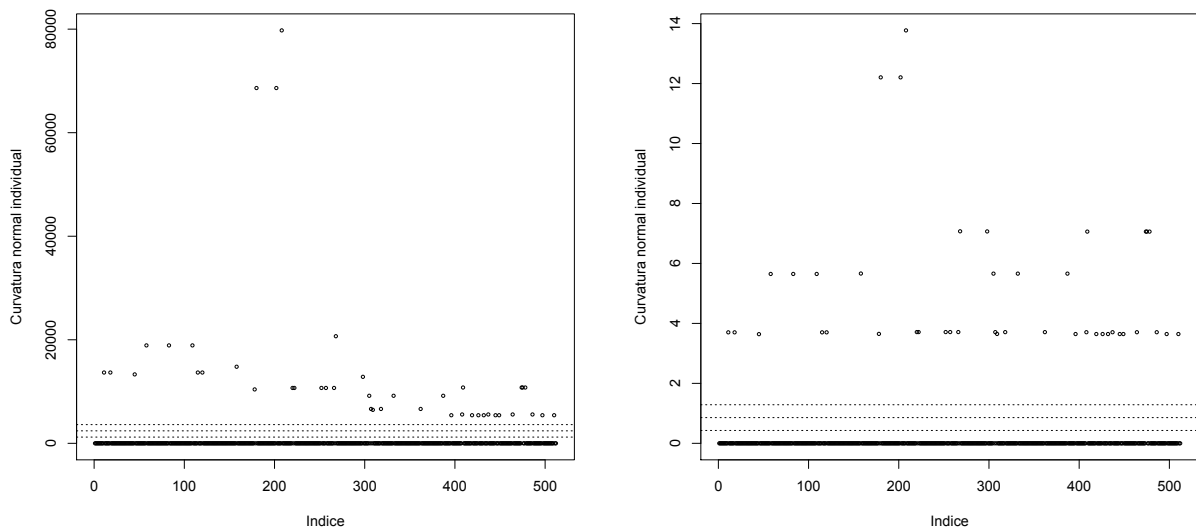
Figura 3.29: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AL para valores da variável resposta iguais a 1 relacionados ao modelo (3.33).



(a) Esquema de perturbação multiplicativa

(b) Esquema de perturbação multiplicativa apropriada

Figura 3.30: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 0 relacionados ao modelo (3.33).



(a) Esquema de perturbação multiplicativa

(b) Esquema de perturbação multiplicativa apropriada

Figura 3.31: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa e o (b) esquema de perturbação multiplicativa apropriada pelo método AQGH para valores da variável resposta iguais a 1 relacionados ao modelo (3.33).

ACC para os esquemas de perturbação multiplicativa e multiplicativa apropriada, respectivamente. Observamos que para os dados com todas as observações as medidas de Sen, Esp e ACC foram de 0,581, 0,687 e 0,678, respectivamente, tanto pelo método AL quanto pelo AQGH. Observamos também que para os dois esquemas em ambos os métodos, ao excluir as observações influentes as medidas de Sen, Esp e ACC aumentaram chegando até 0,875, 0,872 e 0,868, respectivamente.

Em conclusão, os resultados da análise de diagnóstico de influência local para os esquemas de perturbação multiplicativa e multiplicativa apropriada, pelos métodos AL e AQGH, levaram a alterações muito grandes nas estimativas dos parâmetros. No entanto, unicamente pelo método AL, obtiveram-se mudanças inferenciais em relação ao coeficiente da covariável escolaridade. Por último, em todos os casos houve uma melhora das medidas de predição do modelo ao excluir as observações detectadas como influentes.

Esquema de perturbação multiplicativa							
Método	Observações	Grupos	Parâmetro	Estimativa	Valor-p	DRP_j	
AL	Nenhuma	-	β_0	-2,140	<0,001	-	
			β_1	-0,065	0,716	-	
			β_2	-0,867	0,129	-	
			σ	0,216	-	-	
	{#11-#120}, {#158-#268}, {#396-#510}	1, 2, 4	β_0	-8,666	0,175	-304,916	
			β_1	-0,904	0,046	-1284,387	
			β_2	-15,936	0,995	-1737,980	
			σ	7,315	-	-3284,130	
	{#11-#120 }, {#158-#268}	1, 2	β_0	-4,039	0,020	-88,733	
			β_1	-0,274	0,308	-320,003	
			β_2	-16,331	0,992	-1783,481	
			σ	2,901	-	-1242,143	
	AQGH	Nenhuma	-	β_0	-2,140	<0,001	-
				β_1	-0,065	0,716	-
				β_2	-0,867	0,129	-
				σ	0,216	-	-
{#11-#120}, {#158-#268}, {#298-#387}		1, 2, 3	β_0	-9,399	0,065	-339,205	
			β_1	0,111	0,754	-270,176	
			β_2	-16,044	0,991	-1750,661	
			σ	6,670	-	-2985,024	
{#11-#120 }, {#158-#268}		1, 2	β_0	-3,854	0,021	-88,072	
			β_1	-0,274	0,308	-319,146	
			β_2	-16,331	0,992	-1783,713	
			σ	2,783	-	-1187,067	

Tabela 3.14: Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa pelos métodos AL e AQGH relacionados ao modelo (3.33).

Esquema de perturbação multiplicativa apropriada							
Método	Observações	Grupos	Parâmetro	Estimativa	Valor-p	DRP_j	
AL	Nenhuma	-	β_0	-2,140	<0,001	-	
			β_1	-0,065	0,716	-	
			β_2	-0,867	0,129	-	
			σ	0,216	-	-	
	{#6-#120}, {#158-#268}, {#396-#510}	1, 2, 4	β_0	-8,666	0,175	-304,916	
			β_1	-0,904	0,046	-1284,387	
			β_2	-16,122	0,995	-1759,472	
			σ	7,315	-	-3284,131	
	{#6-#120}, {#158-#268}	1, 2	β_0	-4,039	0,020	-88,733	
			β_1	-0,274	0,308	-320,003	
			β_2	-16,486	0,992	-1801,416	
			σ	2,901	-	-1242,143	
	AQGH	Nenhuma	-	β_0	-2,140	<0,001	-
				β_1	-0,065	0,716	-
				β_2	-0,867	0,129	-
				σ	0,216	-	-
{#6-#120}, {#158-#268}, {#282-#387}		1, 2, 3	β_0	-9,399	0,065	-339,205	
			β_1	0,111	0,754	-270,176	
			β_2	-16,505	0,993	-1803,854	
			σ	6,670	-	-2985,024	
{#6-#120}, {#158-#268}		1, 2	β_0	-3,854	0,021	-88,072	
			β_1	-0,274	0,308	-319,146	
			β_2	-16,486	0,992	-1801,650	
			σ	2,783	-	-1187,067	

Tabela 3.15: Estimativas de máxima verossimilhança, valores-p e desvios relativos percentuais para o esquema de perturbação multiplicativa apropriada pelos métodos AL e AQGH relacionados ao modelo (3.33).

Esquema de perturbação multiplicativa					
Método	Observações	Grupos	Sen	Esp	ACC
AL	Nenhuma	-	0,581	0,687	0,678
	{#11-#120}, {#158-#268}, {#396-#510}	1, 2, 4	0,875	0,842	0,843
	{#11-#120}, {#158-#268}	1, 2	0,792	0,802	0,801
AQGH	Nenhuma	-	0,581	0,687	0,678
	{#11-#120}, {#158-#268}, {#298-#387}	1, 2, 3	0,750	0,872	0,868
	{#11-#120}, {#158-#268}	1, 2	0,792	0,802	0,801

Tabela 3.16: Medidas de *Sen*, *Esp* e *ACC* para o esquema de perturbação multiplicativa pelos métodos *AL* e *AQGH* relacionados ao modelo (3.33).

Esquema de perturbação multiplicativa apropriada					
Método	Observações	Grupos	Sen	Esp	ACC
AL	Nenhuma	-	0,581	0,687	0,678
	{#6-#120}, {#158-#268}, {#396-#510}	1, 2, 4	0,875	0,840	0,841
	{#6-#120}, {#158-#268}	1, 2	0,792	0,800	0,799
AQGH	Nenhuma	-	0,581	0,687	0,678
	{#6-#120}, {#158-#268}, {#282-#387}	1, 2, 4	0,750	0,868	0,864
	{#6-#120}, {#158-#268}	1, 2	0,792	0,800	0,799

Tabela 3.17: Medidas de *Sen*, *Esp* e *ACC* para o esquema de perturbação multiplicativa apropriada pelos métodos *AL* e *AQGH* relacionados ao modelo (3.33).

3.5 Aspectos computacionais

3.5.1 Recursos computacionais

Os estudos de simulação e as aplicações foram desenvolvidas na linguagem de programação R (R Core Team, 2014), versão 3.0.1, disponível gratuitamente em <http://www.r-project.org/>, com o Cluster Intel Linux Puma do Laboratório de Computação Científica Avançada (LCCA) da Universidade de São Paulo (<http://www.usp.br/lcca/>). R é uma linguagem de programação livre, disponível sob os termos da GNU (<http://www.gnu.org/>), para o análise estatístico e geração de gráficos, que pode-se utilizar numa ampla variedade de plataformas Unix e sistemas similares (incluindo FreeBSD e Linux), Windows e Mac OS X.

Para os estudos de simulação e as aplicações utilizamos as bibliotecas **lme4** com a função **glmer** para a estimação de parâmetros por máxima verossimilhança pelos métodos de aproximação de integrais AL e AQGH (com cinco pontos), **PresenceAbsence** para os cálculos das medidas de sensibilidade, especificidade e acurácia e **matrixcalc** para as operações matriciais.

3.5.2 Problemas computacionais dos estudos de simulação

Inicialmente os estudos de simulação apresentaram dois tipos de problemas. O primeiro relacionado com a falta de convergência (falsa ou singular) no procedimento da estimação de parâmetros por máxima verossimilhança pelos métodos de integração AL e AQGH (com cinco pontos). O segundo com um valor da estimativa de σ^2 próximo a zero que originava que $\frac{1}{S-M_0} \sum_{s=M_0+1}^S \frac{\partial^2 L(\psi | \mathbf{Y}_o, \mathbf{Y}_f^{(s)})}{\partial \gamma \partial \gamma'} \Big|_{\psi=\hat{\psi}}$ fosse muito grande e que, em consequência, a matriz $(-\ddot{Q}_\psi(\hat{\psi}))^{-1}$ se tornasse computacionalmente singular.

Para lidar com estes problemas, os estudos de simulação foram desenvolvidos considerando aquelas réplicas que não apresentavam falta de convergência (falsa ou singular) e que cumpriam com as condições $\hat{\beta} \in (-1, 3)$ e $\hat{\sigma} \in (0, 01; 3\sigma)$. A Tabela 3.18 mostra o número de réplicas adicionais que tiveram que ser geradas para completar as 100 réplicas para cada tamanho amostral q e valor da perturbação ω_{ij} nos estudos de simulação 1 e 2. Pode-se observar que em ambos os estudos, para cada valor da perturbação ω_{ij} , à medida que o tamanho amostral q aumenta o número de réplicas adicionais diminui e, portanto, diminuem os problemas de convergência e da obtenção de um valor da estimativa de σ^2 próximo a zero. Além disso, para o estudo de simulação 2, pode-se observar que o número de réplicas adicionais é maior no caso em que para um valor fixado de q considera-se um número de grupos I maior.

Número de réplicas adicionais						
Estudo de simulação	q	I	n_i	ω_{ij}		
				0,75	0,85	0,95
1	90	30	3	424	408	331
	360	90	6	67	75	75
	1080	120	9	19	21	21
2	90	15	6	215	273	273
		30	3	424	408	331
	180	30	6	93	97	105
		60	3	138	135	136
	270	30	9	37	39	42
		90	3	99	99	98

Tabela 3.18: Número de réplicas adicionais para cada tamanho amostral q e valor da perturbação ω_{ij} nos estudos de simulação 1 e 2. (MRLEM)

3.5.3 Tempos de execução dos estudos de simulação

Com respeito aos tempos de execução dos cenários dos estudos de simulação 1 e 2, a Tabela 3.19 mostra os resultados em horas para cada tamanho amostral q e valor da perturbação ω_{ij} . Pode-se observar que em ambos os estudos, para cada valor da perturbação ω_{ij} , à medida que o tamanho amostral q aumenta o tempo de execução aumenta, conforme esperado. Além disso, para o estudo de simulação 2, pode-se observar que o tempo de execução foi maior no caso em que para um valor fixado de q considera-se um número de grupos I maior.

Tempos de execução (em horas)						
Estudo de simulação	q	I	n_i	ω_{ij}		
				0,75	0,85	0,95
1	90	30	3	51,85	51,77	51,33
	360	60	6	122,22	122,85	121,82
	1080	120	9	341,72	342,78	329,45
2	90	15	6	25,2	28,7	28,48
		30	3	51,85	51,77	51,33
	180	30	6	50,95	57,57	50,53
		60	3	92,48	98,58	91,77
	270	30	9	64,67	65,03	57,25
		90	3	152,35	153,12	144,28

Tabela 3.19: Tempos de execução (em horas) dos estudos de simulação 1 e 2 para cada tamanho amostral q e valor da perturbação ω_{ij} . (MRLEM)

Capítulo 4

Influência local no modelo de regressão Poisson com efeitos mistos

Neste capítulo apresentamos análise de diagnóstico de influência local no modelo de regressão Poisson com efeitos mistos (MRPEM) sob dois esquemas de perturbação indiretos para a variável resposta propostos por nós. Na Seção 4.1 descrevemos o MRPEM, na Seção 4.2 desenvolvemos a metodologia proposta sob cada um dos esquemas de perturbação e na Seção 4.3 implementamos dois estudos de simulação para a avaliação desta. Por último, na Seção 4.4 ilustramos a análise de diagnóstico com três aplicações a conjuntos de dados reais.

4.1 Modelo de regressão Poisson com efeitos mistos

Sejam Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$ variáveis aleatórias de contagem independentes para i variado com distribuição Poisson de parâmetro λ_{ij} , com $\lambda_{ij} > 0$. Seja também \mathbf{b}_i um vetor aleatório de dimensão p_2 com distribuição normal de média $\mathbf{0}$ e matriz de variâncias e covariâncias $\Sigma = \Sigma(\boldsymbol{\gamma})$, em que $\boldsymbol{\gamma}$ é um vetor de componentes de variâncias e covariâncias desconhecidos de dimensão p_4 . Assume-se que a distribuição condicional de y_{ij} dado \mathbf{b}_i pertence à família exponencial com função de densidade dada por

$$p(y_{ij}|\mathbf{b}_i) = \exp(y_{ij} \log(\lambda_{ij}) - \lambda_{ij} - \log(y_{ij}!)). \quad (4.1)$$

Desta forma, fazendo a analogia com (2.4), temos $\theta_{ij} = \log(\lambda_{ij})$, $a(\theta_{ij}) = \lambda_{ij}$, $c(y_{ij}, \phi) = -\log(y_{ij}!)$ e $\phi = 1$. Logo, a esperança e a variância são

$$E(Y_{ij}|\mathbf{b}_i) = \mu_{ij} = \lambda_{ij}$$

e

$$V(Y_{ij}|\mathbf{b}_i) = V(\mu_{ij}) = \lambda_{ij},$$

respectivamente. Assim, o MRPEM é definido por (4.1) e pelo componente sistemático

$$\log(\lambda_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i, \quad (4.2)$$

em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp_1})'$ e $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ijp_2})'$ são vetores de covariáveis de dimensões p_1 e p_2 , respectivamente, e $\boldsymbol{\beta}$ é o vetor de coeficientes de regressão fixos de dimensão p_1 . Note que de (4.2) obtemos $\lambda_{ij} = \exp(\eta_{ij})$, implicando que, $\theta_{ij} = \eta_{ij}$ e, em consequência, a função k definida na Seção 2.2 é a identidade. O logaritmo da função de verossimilhança do modelo é dado por

$$\begin{aligned} L(\boldsymbol{\psi}|\mathbf{Y}_o) &= \sum_{i=1}^I \log \left[\int_{\mathbb{R}^{p_2}} \prod_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}) - \lambda_{ij} - \log(y_{ij}!)) \right. \\ &\quad \left. \times \frac{1}{(2\pi)^{p_2/2}} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{b}'_i\boldsymbol{\Sigma}^{-1}\mathbf{b}_i\right) d\mathbf{b}_i \right], \end{aligned} \quad (4.3)$$

em que $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ é o vetor de parâmetros desconhecidos de dimensão $(p_1 + p_4)$ e \mathbf{Y}_o o conjunto de dados observados em Y_{ij} , com $j = 1, \dots, n_i$ e $i = 1, \dots, I$. A integral múltipla compreendida em (4.3) não tem solução analítica e métodos como AL e AQGH podem ser utilizados no procedimento da obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\psi}$ denotada por $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\gamma}}')'$.

4.2 Análise de diagnóstico de influência local

Com o objetivo de desenvolver análise de diagnóstico de influência local no modelo de regressão Poisson com efeitos mistos sob dois esquemas de perturbação indiretos para a variável resposta, definimos a expressão dada em (4.3) a partir do enfoque de Zhu e Lee (2003) da seguinte maneira.

Sejam $\mathbf{Y}_f = \{\mathbf{b}_i; i = 1, \dots, I\}$ o conjunto de dados faltantes e $\mathbf{Y}_c = (\mathbf{Y}_o, \mathbf{Y}_f)$ o conjunto de dados completos. Logo, o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos é

dado por

$$L(\boldsymbol{\psi}|\mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}) - \lambda_{ij} - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (4.4)$$

Desta forma, a partir de (4.4) derivamos a curvatura normal sob cada um dos esquemas de perturbação.

4.2.1 Derivação da curvatura normal

Como $\boldsymbol{\psi}$ é de dimensão $(p_1 + p_4)$, temos que a matriz $-\ddot{\mathbf{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ é de dimensão $(p_1 + p_4) \times (p_1 + p_4)$ e as derivadas compreendidas nesta são dadas por

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^I \sum_{j=1}^{n_i} \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij} \mathbf{x}'_{ij(p_1 \times p_1)}, \quad (4.5)$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = \mathbf{0}_{(p_1 \times p_4)}, \quad (4.6)$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} = \mathbf{0}_{(p_4 \times p_1)} \quad (4.7)$$

e

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\frac{I}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \right) \otimes \boldsymbol{\Sigma}^{-1}_{(p_4 \times p_4)}. \quad (4.8)$$

Para os detalhes dos cálculos de (4.5) a (4.8) vide Apêndice A.2.

Seguindo com a metodologia, nas próximas seções definimos os esquemas de perturbação indiretos para a variável resposta e determinamos a matriz de influência para cada um deles.

4.2.2 Esquema de perturbação multiplicativa quadrática da taxa média de ocorrências

Com objetivo de estudar a sensibilidade das estimativas dos parâmetros sob a perturbação da variável resposta de contagem o esquema de perturbação numa primeira instância consistiu em introduzir uma pequena perturbação de forma multiplicativa na taxa média de ocorrências, isto é, $\lambda_{ij}(\omega_{ij}) = \lambda_{ij} \omega_{ij}$. No entanto, a versão apropriada não foi factível pois o fator $g_{ij}(\boldsymbol{\omega}_0)$ (vide Seção 2.5) podia assumir o valor igual a zero. O mesmo aconteceu para o esquema de perturbação aditivo, isto é, $\lambda_{ij}(\omega_{ij}) = \lambda_{ij} + \omega_{ij}$. Deste modo, com o intuito de fazer possível a comparação com a versão apropriada, o esquema de perturbação é dado por uma pequena perturbação quadrática de forma

multiplicativa na taxa média de ocorrências, isto é,

$$\lambda_{ij}(\omega_{ij}) = \lambda_{ij}\omega_{ij}^2, \quad \omega_{ij} > 0. \quad (4.9)$$

Logo, sob (4.9) o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos do modelo perturbado é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}(\omega_{ij})) - \lambda_{ij}(\omega_{ij}) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (4.10)$$

Assim, para $\omega_{ij0} = 1$, com $j = 1, \dots, n_i$ e $i = 1, \dots, I$, a matriz de influência dada por $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0_{pmq}}}$ é uma matriz de dimensão $(p_1 + p_4) \times q$, com $q = \sum_{i=1}^I n_i$, e as derivadas compreendidas nesta são

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = -2 \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij}, \quad (4.11)$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0}, \quad (4.12)$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Para os detalhes dos cálculos de (4.11) e (4.12), vide Apêndice A.2.

4.2.3 Esquema de perturbação multiplicativa quadrática apropriada da taxa média de ocorrências

Para definir o esquema de perturbação multiplicativa quadrática apropriada da taxa média de ocorrências, consideramos o logaritmo da função de verossimilhança do modelo perturbado dado em (4.10) e obtemos a função de densidade associada, isto é,

$$p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega}) = \prod_{i=1}^I \left[\prod_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}(\omega_{ij})) - \lambda_{ij}(\omega_{ij}) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (4.13)$$

Desta forma, o vetor de não perturbação é dado por $\boldsymbol{\omega}_0 = \mathbf{1} \in \mathbb{R}^q$. Logo, baseados em (4.13) os elementos compreendidos na matriz $G(\boldsymbol{\omega})$ são

$$g_{ij}(\boldsymbol{\omega}) = \mathbb{E} \left(\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} \middle| \boldsymbol{\omega} \right), \quad (4.14)$$

em que para $i = j$,

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 2 \left[\frac{y_{ij}}{\omega_{ij}^2} + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \right] \quad (4.15)$$

e para $i \neq j$,

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0, \quad (4.16)$$

com $i, j = 1, \dots, q$. Para os detalhes dos cálculos de (4.15) e (4.16), vide o Apêndice A.2. Para o valor de não perturbação $\omega_{ij0} = 1$, as expressões dadas em (4.14) e (4.16) são iguais a

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 2[y_{ij} + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i)] \quad (4.17)$$

e

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0, \quad (4.18)$$

respectivamente. No obstante, como a esperança de (4.17) não tem solução analítica, é possível aproximar esta por

$$g_{ij}(\boldsymbol{\omega}_0) \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^S 2[y_{ij} + \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i^{(s)})], \quad (4.19)$$

para $i, j = 1, \dots, q$, com $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ uma amostra aleatória gerada mediante o algoritmo Metropolis-Hastings descrito na próxima seção.

Como $G(\boldsymbol{\omega}_0) \neq a\mathbf{I}_q$, com $a > 0$ e \mathbf{I}_q a matriz identidade de dimensão q , escolhemos um novo vetor de perturbação dado por

$$\omega_{ij}^* = 1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\omega_{ij} - 1),$$

para $i, j = 1, \dots, q$, de modo que, o esquema de perturbação multiplicativa quadrática apropriada é dado por

$$\lambda_{ij}(\omega_{ij}^*) = \lambda_{ij} \omega_{ij}^{*2}, \quad \omega_{ij}^{*2} > 0, \quad (4.20)$$

com o qual verifica-se $G(\boldsymbol{\omega}_0) = \mathbf{I}_q$, com $a = 1$. Em seguida, a partir de (4.20) o logaritmo da função

de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos do modelo perturbado é dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}(\omega_{ij}^*)) - \lambda_{ij}(\omega_{ij}^*) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right]. \quad (4.21)$$

Assim, para $\omega_{ij0} = 1$, com $j = 1, \dots, n_i$ e $i = 1, \dots, I$, a matriz de influência dada por $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pmqa}}$ é uma matriz de dimensão $(p_1 + p_4) \times q$, com $q = \sum_{i=1}^I n_i$, e as derivadas compreendidas nesta são

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = -2g_{ij}(\boldsymbol{\omega}_0)^{-1/2} \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{x}_{ij}, \quad (4.22)$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0}, \quad (4.23)$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Para os detalhes dos cálculos de (4.22) e (4.23), vide Apêndice A.2.

4.2.4 Aproximação das matrizes da curvatura normal

Para aproximar as matrizes $-\ddot{Q}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$, $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pmq}}$ e $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pmqa}}$, amostramos $\{\mathbf{b}_i^{(s)}; s = 1, \dots, S, i = 1, \dots, I\}$ da função de densidade dada por

$$p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi}) \propto \exp \left[-\frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i + \sum_{j=1}^{n_i} y_{ij} (\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) - \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \right], \quad (4.24)$$

utilizando o algoritmo Metropolis-Hastings como segue.

Dado um valor arbitrário $\mathbf{b}_i^{(0)}$, na r -ésima iteração realize os seguintes passos:

1. Dado o valor atual de $\mathbf{b}_i^{(r-1)}$, gere um novo candidato como $\mathbf{b}_i \sim N_{p_2}(\mathbf{b}_i^{(r-1)}, \Gamma(\mathbf{0}))$, em que

$$\Gamma(\mathbf{0}) = \Gamma(\mathbf{b}_i) |_{\mathbf{b}_i=\mathbf{0}} = \left[\boldsymbol{\Sigma}^{-1} + \sum_{j=1}^{n_i} \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i) \mathbf{z}_{ij} \mathbf{z}'_{ij} \right]^{-1} \Big|_{\mathbf{b}_i=\mathbf{0}}.$$

2. Gere $u \sim U(0, 1)$. Se $u \leq \alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i)$, faça $\mathbf{b}_i^{(r)} = \mathbf{b}_i$, senão faça $\mathbf{b}_i^{(r)} = \mathbf{b}_i^{(r-1)}$, com

$$\alpha(\mathbf{b}_i^{(r-1)}, \mathbf{b}_i) = \min \left(\frac{p(\mathbf{b}_i | \mathbf{Y}_o, \boldsymbol{\psi})}{p(\mathbf{b}_i^{(r-1)} | \mathbf{Y}_o, \boldsymbol{\psi})}, 1 \right),$$

a probabilidade de aceitar um novo candidato.

3. Repita os passos (1) e (2) para $r + 1$.

Uma vez amostradas as S observações, descartamos as M_0 primeiras e calculamos $-\ddot{Q}_\psi(\hat{\psi})$, $\Delta\omega_{0pmq}$ e $\Delta\omega_{0pmqa}$. Logo, obtemos as matrizes

$$-2\ddot{Q}_{\omega_{0pmq}} = 2\Delta'_{\omega_{0pmq}} (-\ddot{Q}_\psi(\hat{\psi}))^{-1} \Delta\omega_{0pmq} \quad (4.25)$$

e

$$-2\ddot{Q}_{\omega_{0pmqa}} = 2\Delta'_{\omega_{0pmqa}} (-\ddot{Q}_\psi(\hat{\psi}))^{-1} \Delta\omega_{0pmqa}, \quad (4.26)$$

e, por conseguinte, as curvaturas normais individuais

$$C_{f_Q, \mathbf{h}_{ijpmq}} = \sum_{m=1}^M \lambda_{m_{pmq}} e_{m_{ijpmq}}^2 \quad (4.27)$$

e

$$C_{f_Q, \mathbf{h}_{ijpmqa}} = \sum_{m=1}^M \lambda_{m_{pmqa}} e_{m_{ijpmqa}}^2, \quad (4.28)$$

para $i, j = 1, \dots, q$, respectivamente. Em seguida, de acordo com a Definição 2.4.1, os intervalos de influência são dados por

$$\left[(k_0 - 1) \frac{\lambda_{0_{pmq}}}{q}, k_0 \frac{\lambda_{0_{pmq}}}{q} \right), \quad (4.29)$$

em que $\lambda_{0_{pmq}} = \sum_{i,j=1}^q C_{f_Q, \mathbf{h}_{ijpmq}}$ e

$$\left[(k_0 - 1) \frac{\lambda_{0_{pmqa}}}{q}, k_0 \frac{\lambda_{0_{pmqa}}}{q} \right), \quad (4.30)$$

em que $\lambda_{0_{pmqa}} = \sum_{i,j=1}^q C_{f_Q, \mathbf{h}_{ijpmqa}}$, para $k_0 = 1, 2, 3, \dots$, respectivamente.

4.3 Estudos de simulação

Com o objetivo de avaliar o desempenho da análise de diagnóstico de influência local sob os esquemas de perturbação propostos, implementamos dois estudos de simulação de Monte Carlo com diferentes cenários de $R = 100$ réplicas cada um. Cada cenário considera a seguinte estrutura do modelo e procedimento de simulação das réplicas.

Seja o modelo de regressão Poisson com efeitos mistos dado por $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$, em que

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1ij} + b_i, \quad (4.31)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, I$, com x_{1ij} uma variável aleatória com distribuição uniforme no intervalo $(0, 1)$, β_0 e β_1 os coeficientes de regressão e b_i o intercepto aleatório com distribuição normal de média 0 e variância σ^2 . Sejam também os verdadeiros valores dos parâmetros dados por $\beta_0 = 0,5$, $\beta_1 = 0,5$ e $\sigma^2 = 0,5$ ($\sigma = 0,71$) (valores iguais aos considerados por [Zhu e Lee \(2003\)](#)). Desta forma, para valores de n_i e I geramos as variáveis x_{1ij} e b_i e, com (4.31) calculamos as taxas médias de ocorrências λ_{ij} . Baseados nestes últimos valores geramos cada réplica da variável resposta y_{ij} a partir de uma distribuição Poisson com taxa média de ocorrências λ_{ij} , para $j = 1, \dots, n_i$ e $i = 1, \dots, I$. A este conjunto de dados o denominamos dados originais e o denotamos por O. Depois disto geramos os conjuntos de dados “perturbados” sob o esquema de perturbação multiplicativa quadrática e perturbação multiplicativa quadrática apropriada como segue. Começamos escolhendo o grupo a perturbar como o grupo i no qual se encontra o valor máximo das taxas médias de ocorrências. Do conjunto de dados originais voltamos a gerar os n_i valores da variável resposta do grupo i , isto é, para um valor da perturbação ω_{ij} , geramos y_{ij} a partir de uma distribuição Poisson com taxa média de ocorrências $\lambda_{ij}(\omega_{ij}) = \lambda_{ij}\omega_{ij}^2$, para $j = 1, \dots, n_i$. A este nuevo conjunto de dados o denominamos dados sob o esquema de perturbação multiplicativa quadrática e o denotamos por PMQ. Do mesmo modo, para o outro esquema voltamos a gerar os n_i valores do grupo i , isto é, geramos y_{ij} a partir de uma distribuição Poisson com taxa média de ocorrências $\lambda_{ij}(\omega_{ij}^*) = \lambda_{ij}\omega_{ij}^{*2}$, em que $\omega_{ij}^* = [1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\omega_{ij} - 1)]$, para $j = 1, \dots, n_i$, com $g_{ij}(\boldsymbol{\omega}_0)$ aproximado a partir de 2000 valores adicionais de b_i gerados de uma distribuição normal de média 0 e variância $\hat{\sigma}^2$. A este novo conjunto de dados o denominamos dados sob o esquema de perturbação multiplicativa quadrática apropriada e o denotamos por PMQA.

Uma vez simulados os conjuntos de dados O, PMQ e PMQA, realizamos para cada um deles a estimação dos parâmetros por máxima verossimilhança mediante os métodos AL e AQGH (com cinco pontos). Para cada um dos conjuntos PMQ e PMQA e baseados nas estimativas correspondentes, amostramos $S = 600$ observações de b_i , para $i = 1, \dots, I$, mediante o algoritmo Metropolis-Hastings e calculamos as matrizes $-\ddot{Q}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$, $\Delta_{\boldsymbol{\omega}_{0pmq}}$ e $\Delta_{\boldsymbol{\omega}_{0pmqa}}$, com $M_0 = 100$. Em seguida, calculamos as curvaturas normais individuais e os intervalos de influência sob cada um dos esquemas de perturbação e verificamos se as observações perturbadas estão sendo detectadas como influentes. Por último, realizamos este procedimento para $R = 100$ réplicas e calculamos as proporções de VI, FI

e I (vide Seção 3.3) sob cada um dos esquemas de perturbação.

O estudo de simulação 1 tem por objetivo avaliar o desempenho da análise de diagnóstico de influência local sob os esquemas de perturbação com as proporções de VI, FI e I pelos métodos AL e AQGH para cada um dos valores da perturbação $\omega^2=0,8; 1,2$ e $1,5$ para os seguintes cenários:

- $q=90, I=30$ e $n_i=3$;
- $q=360, I=60$ e $n_i=6$;
- $q=1080, I=120$ e $n_i=9$.

O estudo de simulação 2 tem por objetivo avaliar o desempenho da análise de diagnóstico de influência local sob os esquemas de perturbação com as proporções de VI, FI e I no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i pelos métodos AL e AQGH para cada um dos valores da perturbação $\omega^2=0,8; 1,2$ e $1,5$ para os seguintes cenários:

- $q=90: I=15$ e $n_i=6$ e $I=30$ e $n_i=3$;
- $q=360: I=30$ e $n_i=12$ e $I=60$ e $n_i=6$;
- $q=1080: I=60$ e $n_i=18$ e $I=120$ e $n_i=9$.

Em seguida apresentamos os resultados para cada um deles.

4.3.1 Estudo de simulação 1

Iniciamos a análise observando o comportamento das estimativas dos parâmetros mediante a média, viés e REQM. A Figura 4.1, correspondente a $\omega^2=0,8$, mostra que para cada um dos parâmetros β_0 e β_1 , à medida que q aumenta a média tende ao verdadeiro valor $0,5$ e o viés e REQM diminuem para os conjuntos de dados O, PMQ e PMQA. Além disso, para $q = 90$, o parâmetro β_0 é superestimado e o parâmetro β_1 é subestimado, com viés grande em valor absoluto, para ambos os casos. Por outro lado, para cada um dos parâmetros a média, viés e REQM são muito parecidas (superpondo-se os pontos e as linhas) entre os conjuntos de dados PMQ e PMQA à medida que q aumenta pelos dois métodos, pelo que não se observa um comportamento padrão sobre qual esquema produz maior impacto nas estimativas. Os mesmos resultados observam-se para $\omega^2=1,2$ e $1,5$ nas Figuras 4.2 e 4.3, respectivamente.

Com respeito ao comportamento das estimativas do parâmetro σ , a Figura 4.4 correspondente a $\omega^2=0,8$, mostra que à medida que q aumenta a média apresenta uma leve tendência ao verdadeiro

valor 0,71 e, portanto, o viés e REQM diminuem levemente para os conjuntos de dados O, PMQ e PMQA. Cabe destacar que para todos os valores de q , o parâmetro é subestimado e o viés é grande em valor absoluto. Por outro lado, a média, viés e REQM são muito parecidas (superpondo-se os pontos e as linhas) entre os conjuntos de dados PMQ e PMQA à medida que q aumenta pelos dois métodos, pelo que também para este parâmetro não se observa um comportamento padrão sobre qual esquema produz maior impacto nas estimativas.

Em relação aos resultados das proporções de VI, FI e I, a Figura 4.7, correspondente a $\omega^2=0,8$, mostra que para $q=90$ não houve detecção de VI, FI e, portanto, de I sob os esquemas PMQ e PMQA pelos métodos AL e AQGH. Para $q=360$ e 1080, as proporções de VI, FI e I foram maiores para o esquema PMQ que PMQA e às vezes maiores para $q=1080$ do que para $q=360$; pelos dois métodos. Especificamente, para $q=360$, sob o esquema PMQ a proporção máxima de VI está ao redor de 0,54, a proporção máxima de FI é de 0,11 e a proporção máxima de I está ao redor de 0,65. Sob o esquema PMQA, a proporção máxima de VI está ao redor de 0,23, a proporção máxima de FI é de 0,03 e a proporção máxima de I está ao redor de 0,26. Para $q=1080$, sob o esquema PMQ a proporção máxima de VI é de 0,90, de FI é de 0,08 e de I é de 0,98. Sob o esquema PMQA, a proporção máxima de VI está ao redor de 0,46, a proporção máxima de FI é de 0,05 e, por último, a proporção máxima de I está ao redor de 0,47. Os mesmos resultados observam-se para $\omega^2=1,2$ e 1,5 nas Figuras 4.8 e 4.9, respectivamente, porém com proporções muito maiores.

Em conclusão, para cada valor de q , onde houve detecção, as proporções de VI e I foram maiores para o esquema PMQ (de até 0,96) do que para o esquema PMQA (de até 0,75) e as proporções de FI foram similares (de até 0,13) entre os esquemas PMQ e PMQA; pelos dois métodos e para todos os valores da perturbação. Por último, conforme esperado as proporções de VI foram muito maiores às proporções de FI.

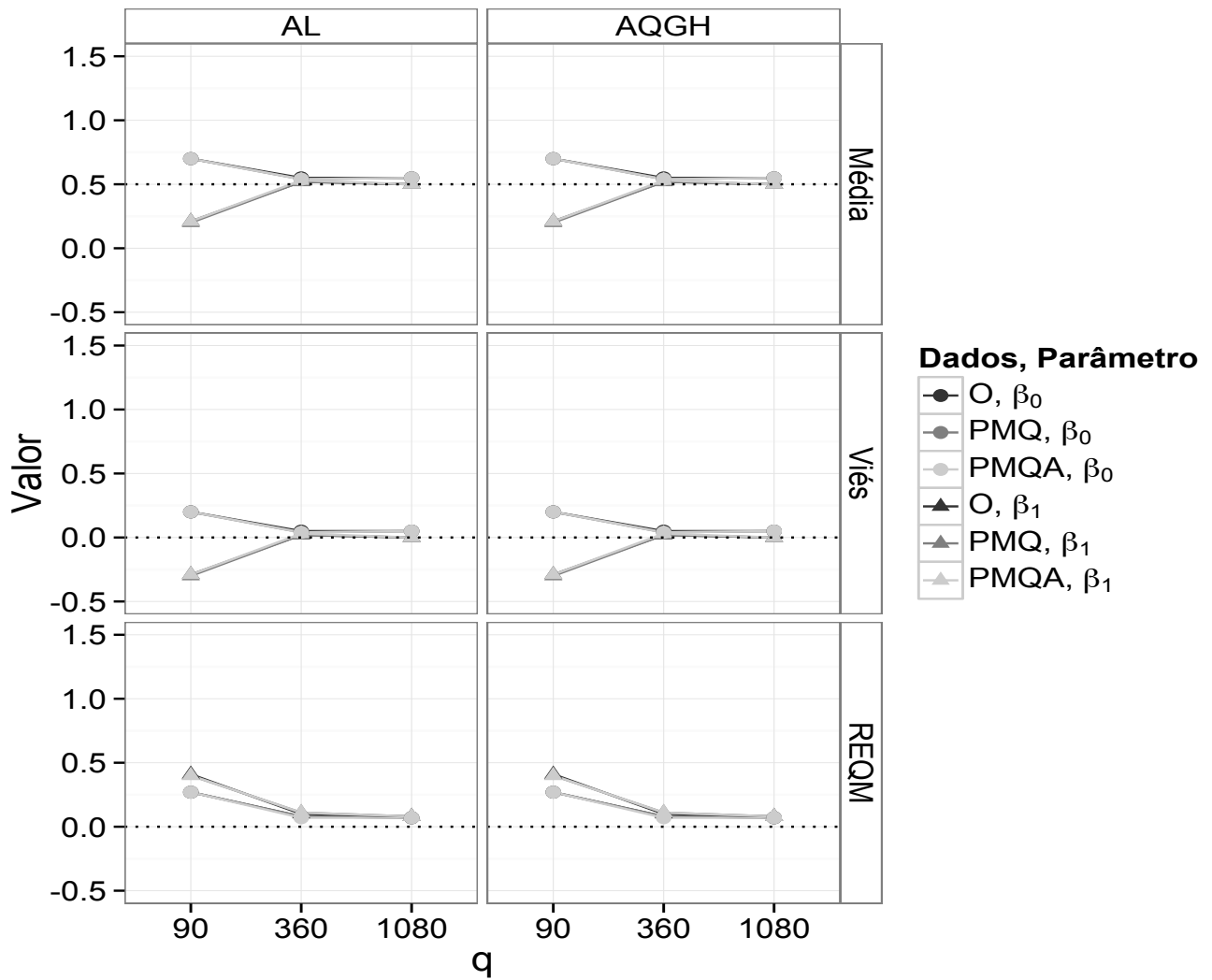


Figura 4.1: Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O , PMQ e $PMQA$ para diferentes valores de q pelos métodos AL e $AQGH$, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM)

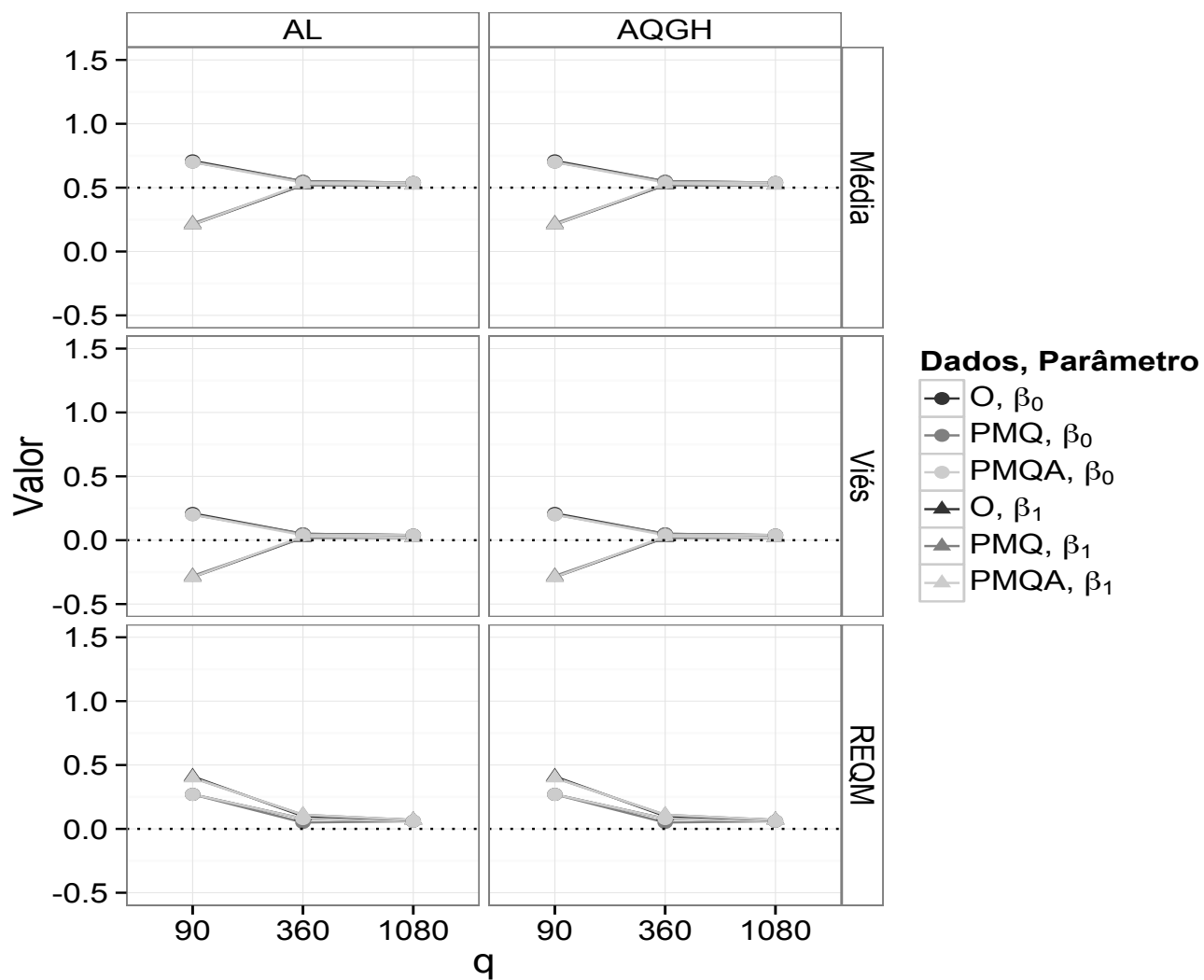


Figura 4.2: Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM)

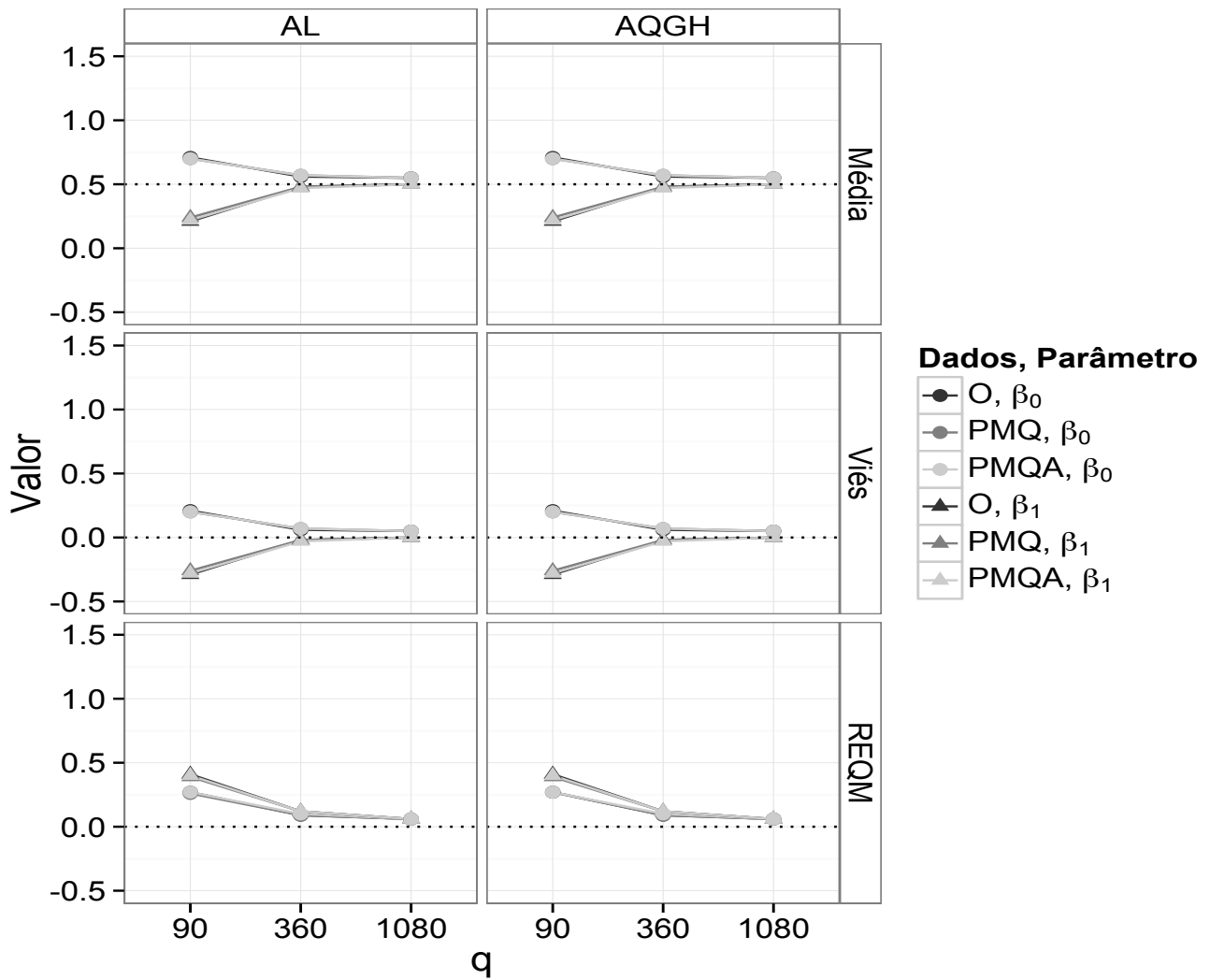


Figura 4.3: Média, viés e REQM das estimativas dos parâmetros β_0 e β_1 para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM)

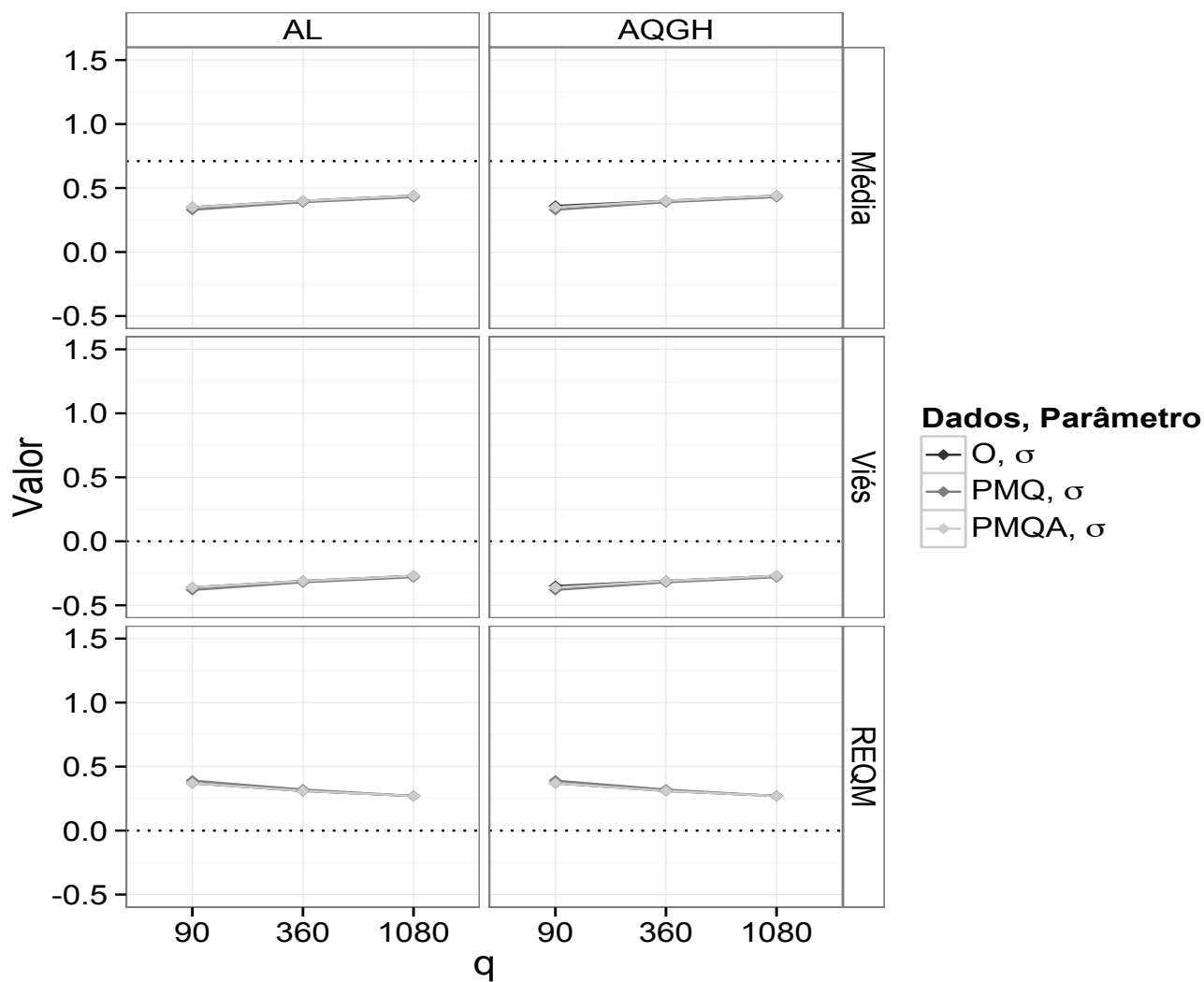


Figura 4.4: Média, viés e REQM da estimativa do parâmetro σ para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM)

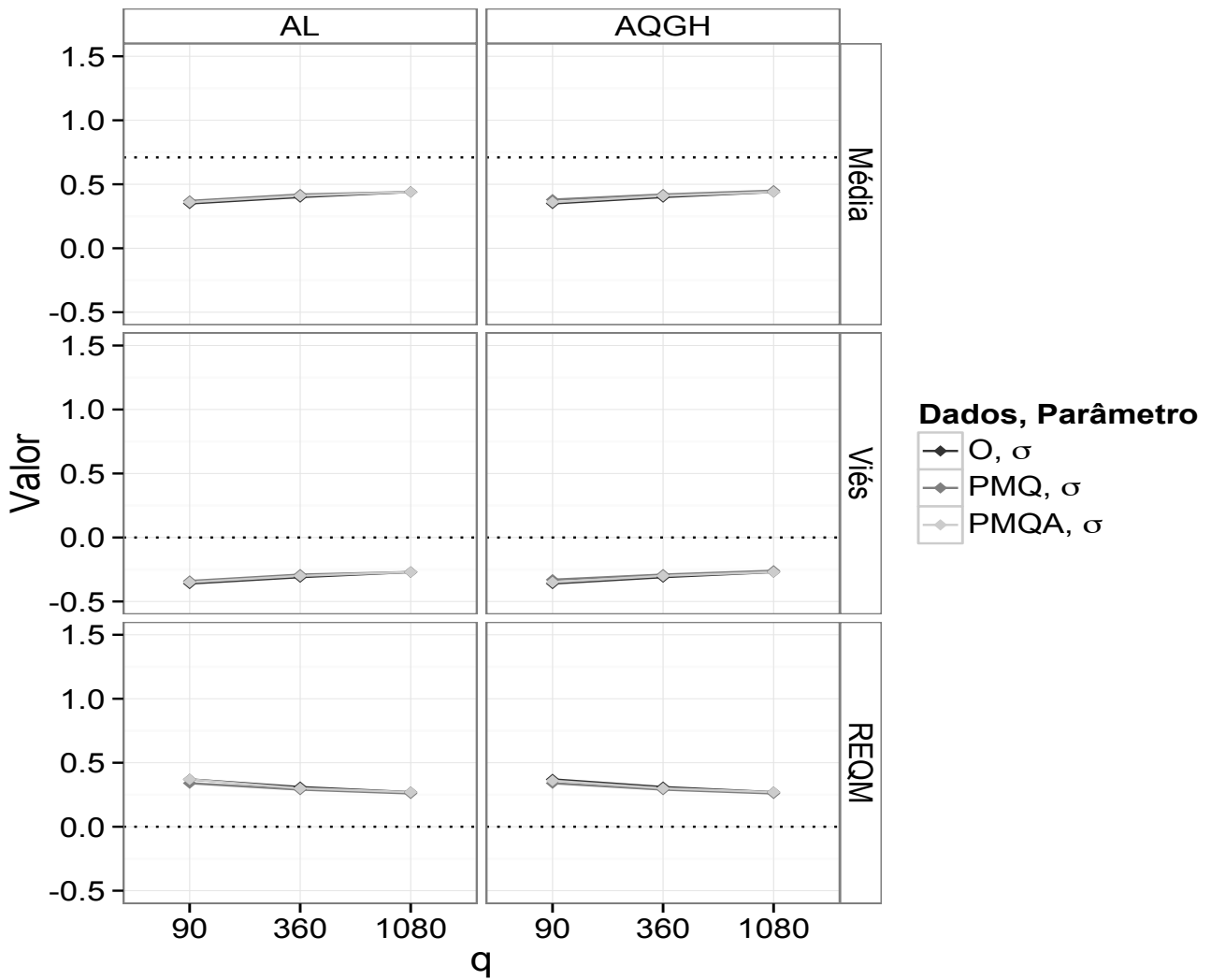


Figura 4.5: Média, viés e REQM da estimativa do parâmetro σ para os dados O , PMQ e $PMQA$ para diferentes valores de q pelos métodos AL e $AQGH$, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM)

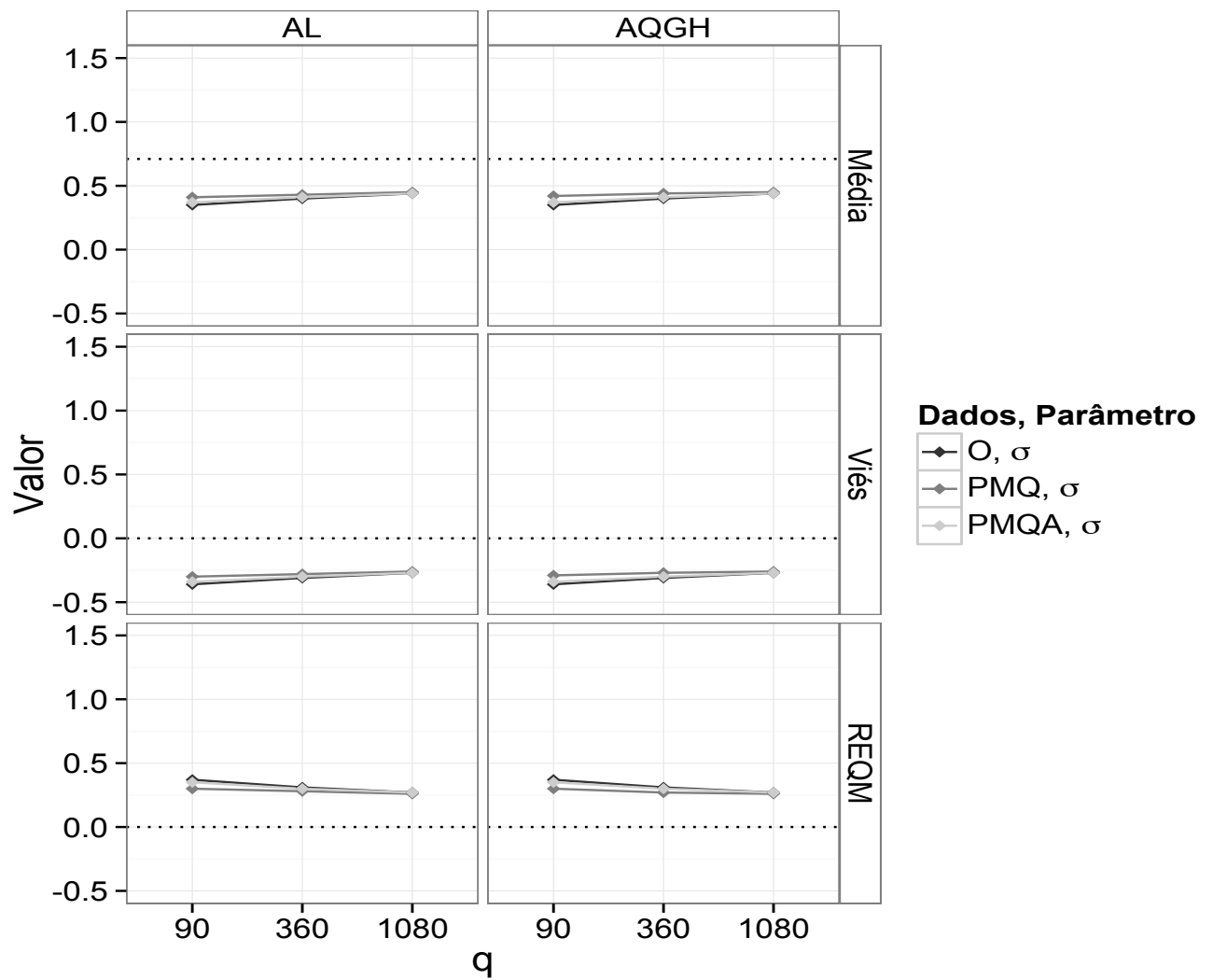


Figura 4.6: Média, viés e REQM da estimativa do parâmetro σ para os dados O, PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM)

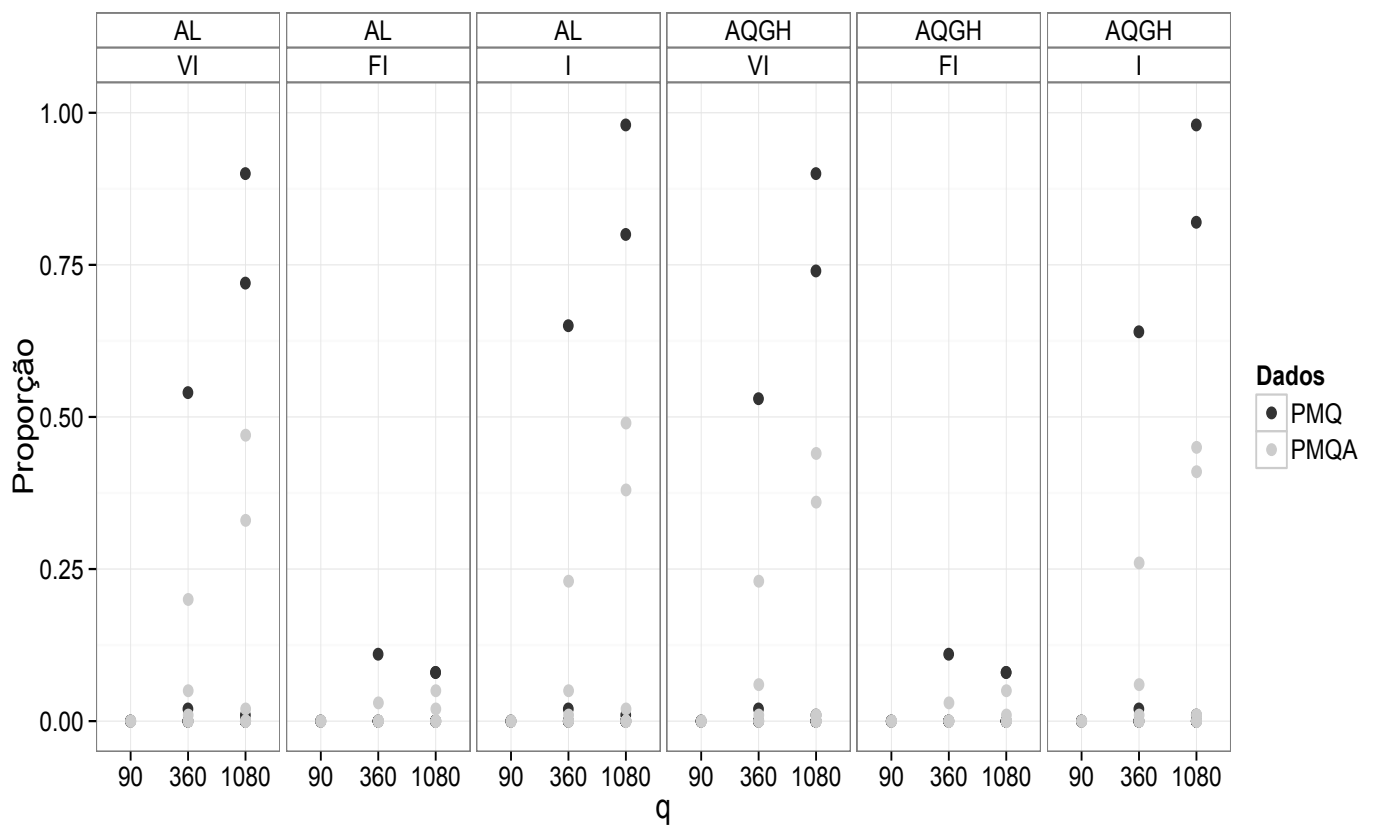


Figura 4.7: Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 1) (MRPEM)

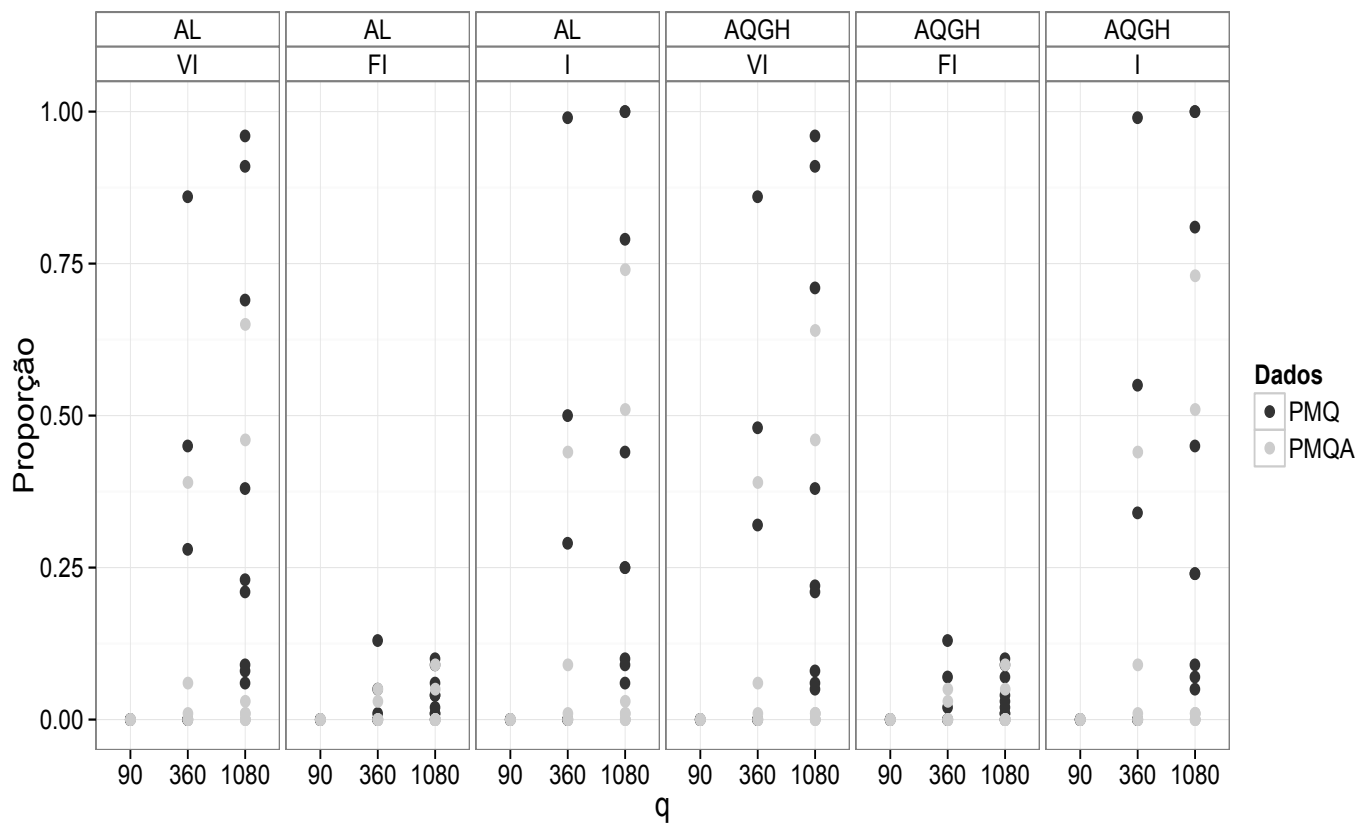


Figura 4.8: Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 1) (MRPEM)

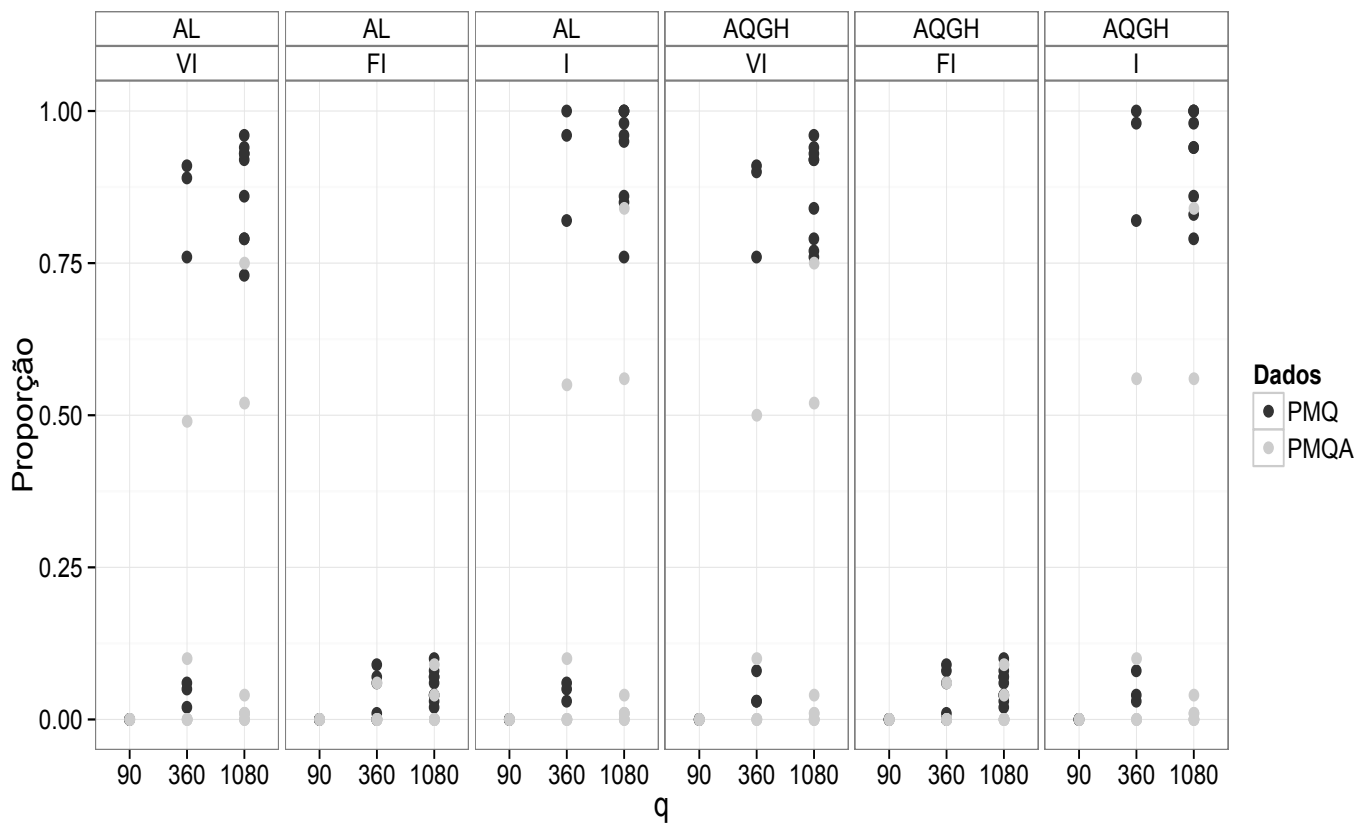


Figura 4.9: Proporções de VI, FI e I para os dados PMQ e PMQA para diferentes valores de q pelos métodos AL e AQGH, com $\omega^2=1,5$. (Estudo de simulação 1) (MRPEM)

4.3.2 Estudo de simulação 2

Referente aos resultados das proporções de VI, FI e I no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i , a Figura 4.10, correspondente a $\omega^2 = 0,8$, mostra que sob cada um dos esquemas PMQ e PMQA pelo métodos AL e AQGH, para $q=90$ as proporções de VI e I foram levemente maiores no caso em que para um valor fixado de q o número de grupos I for menor, e para $q=360$ e 1080 , as proporções de VI e I foram maiores no caso em que para um valor fixado de q o número de grupos I for maior. Por outro lado, para todos os valores de q as proporções de FI foram similares no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i .

As Figuras 4.11 e 4.12 correspondentes a $\omega^2 = 1,2$ e $1,5$, respectivamente, mostram que sob cada um dos esquemas PMQ e PMQA pelos métodos AL e AQGH, para $q=90$ as proporções de VI e I foram maiores no caso em que para um valor fixado de q o número de grupos I for menor, e para $q=360$ e 1080 , as proporções de VI e I foram similares no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i . Por outro lado, para todos os valores de q as proporções de FI foram similares no caso em que para um valor fixado de q consideram-se diferentes valores de I e n_i .

Em conclusão, as proporções de VI, FI e I sob cada um dos esquemas PM e PMA pelo métodos AL e AQGH, em geral, não parecem ser afetadas pela relação número de grupos I e tamanho dos grupos n_i .

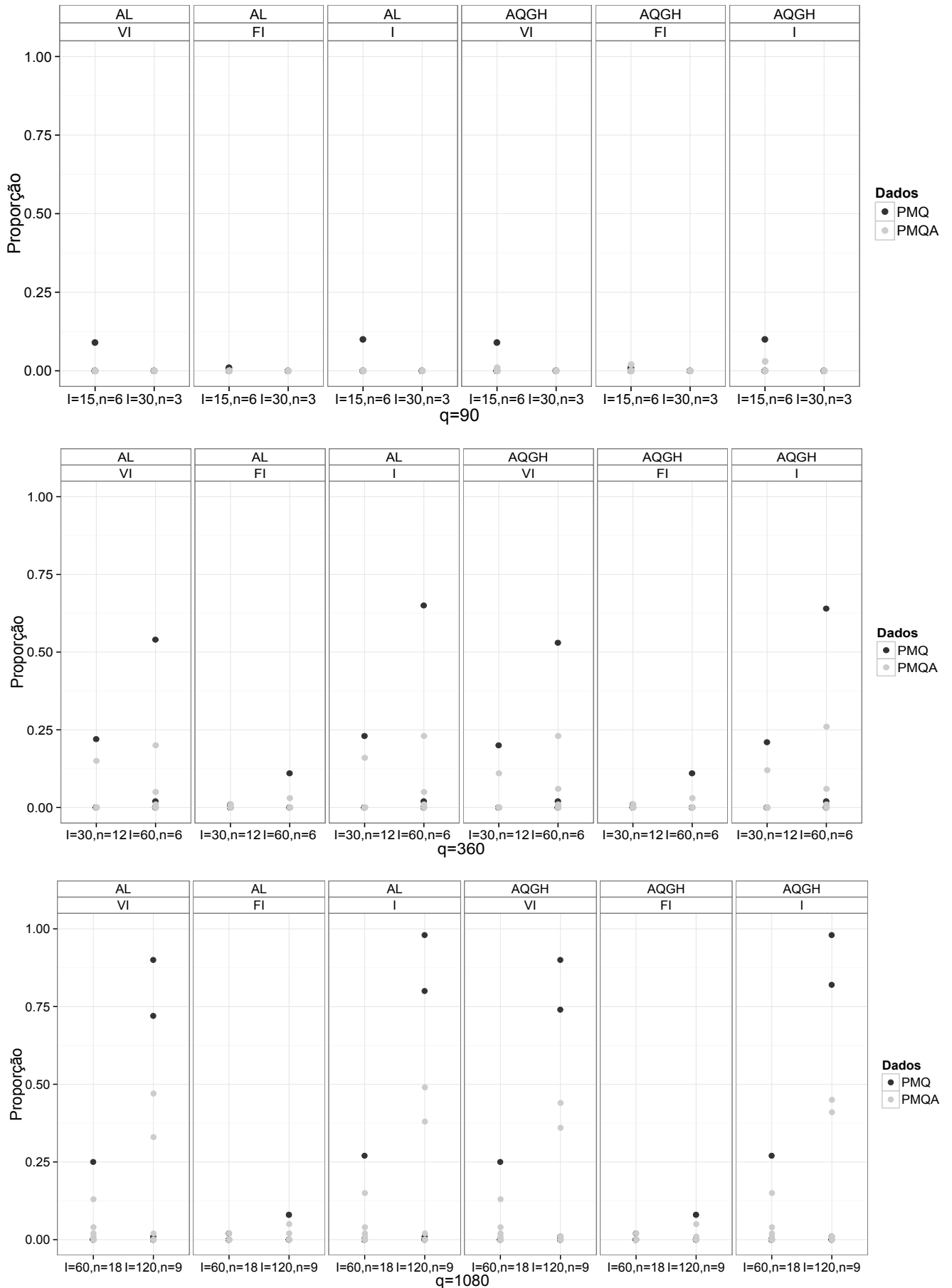


Figura 4.10: Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de l e n_i pelos métodos AL e AQGH, com $\omega^2=0,8$. (Estudo de simulação 2) (MRPEM)

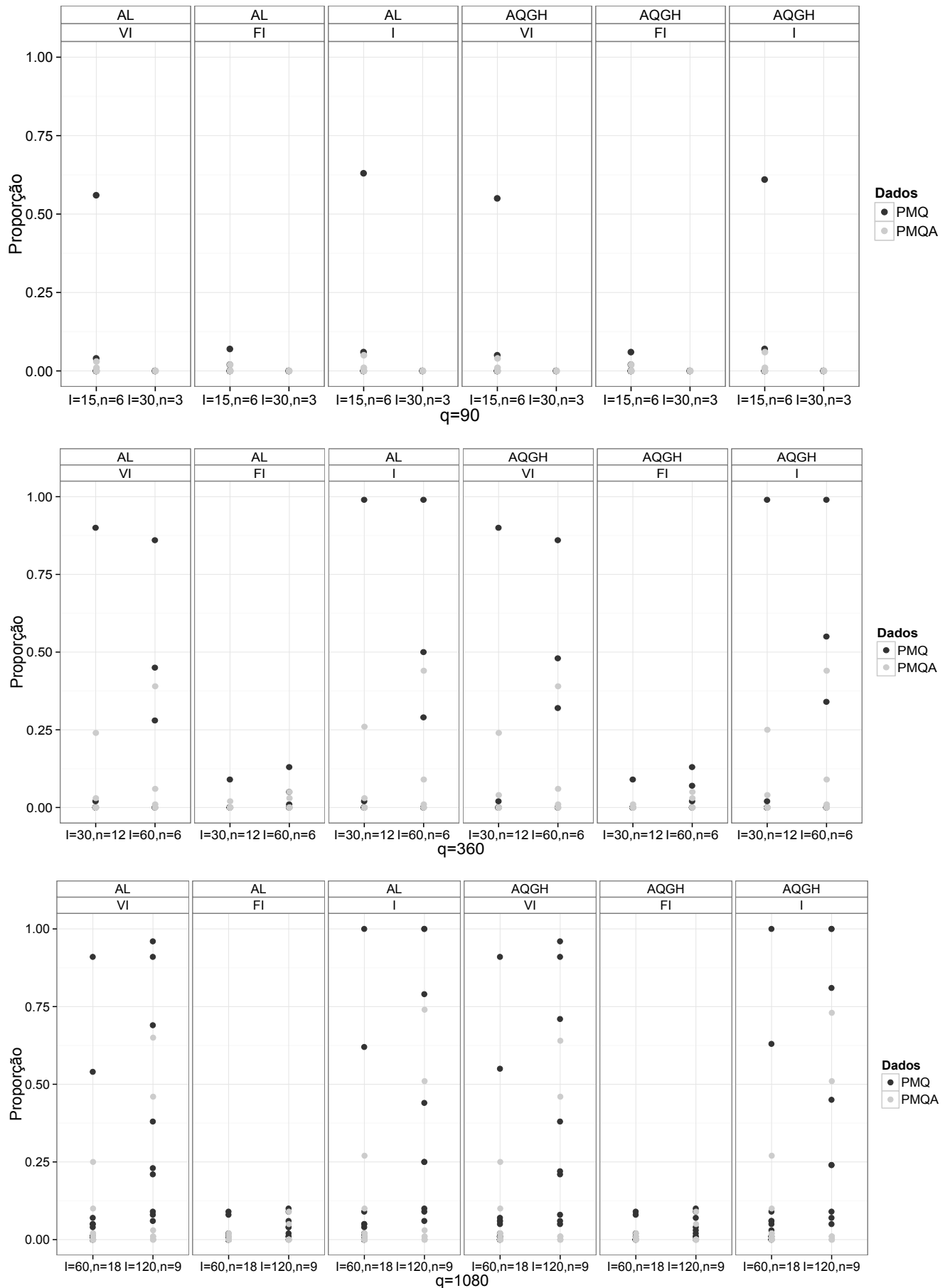


Figura 4.11: Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de l e n_i pelos métodos AL e AQGH, com $\omega^2=1,2$. (Estudo de simulação 2)(MRPEM)

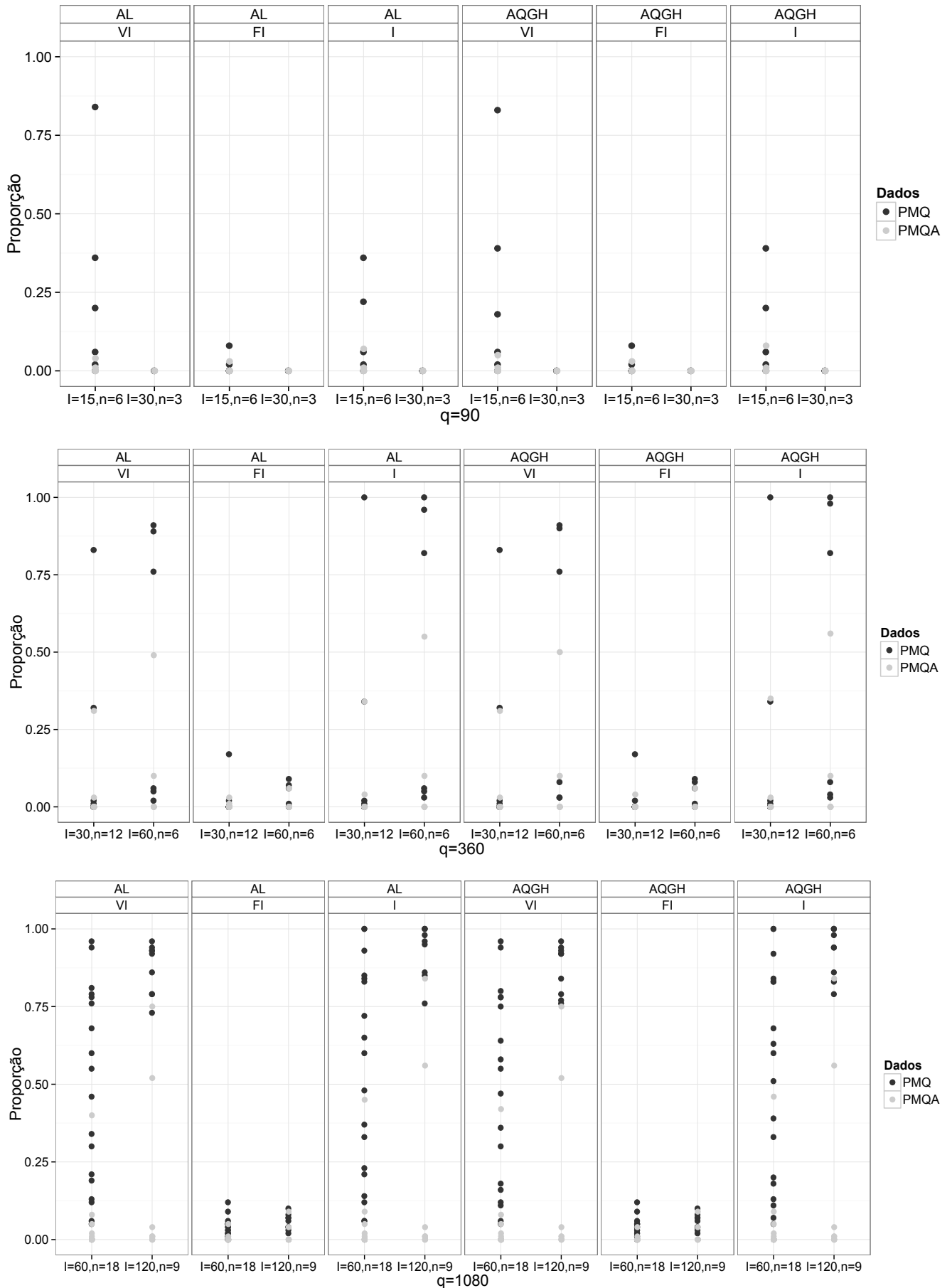


Figura 4.12: Proporções de VI, FI e I para os dados PMQ e PMQA no caso em que para um valor fixado de q consideram-se diferentes valores de l e n_i pelos métodos AL e AQQH, com $\omega^2=1,5$. (Estudo de simulação 2) (MRPEM)

4.4 Aplicações

Nesta seção ilustramos a metodologia desenvolvida mediante três aplicações a conjuntos de dados reais. A aplicação 1 refere-se a um conjunto de dados relacionados a um ensaio clínico em pacientes com epilepsia apresentado por [Thall e Vail \(1990\)](#), o qual é abordado frequentemente na literatura. A aplicação 2 refere-se a um conjunto de dados relacionados a um estudo em pacientes que padecem de enxaquecas apresentado por [McKnight e Eeden \(1993\)](#). Por último, na aplicação 3 analisamos um conjunto de dados correspondentes a um estudo de caso-controle envolvendo pacientes com diabetes mellitus tipo I apresentado por [Giampaoli *et al.* \(2014\)](#).

4.4.1 Aplicação 1

Um conjunto de dados balanceados relacionados a um ensaio clínico que teve por objetivo analisar o efeito de um tratamento em 59 pacientes com epilepsia foram apresentados por [Thall e Vail \(1990\)](#). Os pacientes foram aleatoriamente atribuídos aos grupos tratamento e placebo e o número de crises vivenciadas por cada um deles foram registrados durante quatro períodos de dois semanas cada um. Outras variáveis tais como o logaritmo do número de crises de um pre-ensaio clínico e o logaritmo da idade também foram consideradas.

Para este problema considera-se um MRPEM cuja variável aleatória Y_{ij} é dada pelo número de crises para o i -ésimo paciente no j -ésimo período com taxa média de ocorrências $\lambda_{ij} > 0$ e cujo intercepto aleatório b_i assume uma distribuição normal de media 0 e variância σ^2 . Assim, o MRPEM é dado por $Y_{ij} \sim Poisson(\lambda_{ij})$, em que

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + b_i, \quad (4.32)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 59$ ($q=236$), com x_{1ij} uma variável indicadora dos grupos tratamento (1) e placebo (0), x_{2ij} o logaritmo do número de crises de um pre-ensaio clínico, x_{3ij} a interação entre x_{1ij} e x_{2ij} , x_{4ij} o logaritmo da idade e x_{5ij} uma variável indicadora dos períodos (-3,-1,1 e 3), e $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ e β_5 os coeficientes de regressão. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \sigma)'$.

Para começar realizamos o ajuste do modelo obtendo os resultados das estimativas de máxima verossimilhança dos parâmetros e das predições BLUP do intercepto aleatório pelos métodos AL e AQRH (com cinco pontos). Na Tabela 4.1 observamos que com um nível de significância de 5% os coeficientes das covariáveis intercepto, grupo, logaritmo do número de crises de um pre-ensaio clínico

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	1,793	0,104	<0,001
	β_1	-0,334	0,147	0,023
	β_2	0,883	0,131	<0,001
	β_3	0,339	0,203	0,094
	β_4	0,481	0,346	0,164
	β_5	-0,030	0,010	0,004
	σ	0,501		
AQGH	β_0	1,793	0,105	<0,001
	β_1	-0,334	0,148	0,024
	β_2	0,883	0,0131	<0,001
	β_3	0,339	0,203	0,095
	β_4	0,481	0,346	0,165
	β_5	-0,030	0,010	0,004
	σ	0,502		

Tabela 4.1: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.32).

e períodos são significativos, e que o desvio padrão do intercepto aleatório é próximo a 0,500, tanto pelo método AL quanto pelo método AQGH. Na Tabela 4.2 observamos que as predições BLUP do intercepto aleatório praticamente não apresentam diferenças de um método a outro. Observamos também nas Figuras 4.13a e 4.13b que os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório não localizam-se totalmente no interior do envelope sugerindo indícios de sobredispersão, por ambos os métodos.

De acordo com o proposto na Seção 4.2, para realizar análise de diagnóstico de influência local primeiro amostramos $\{b_i^{(s)}; s = 1, \dots, 10.000, i = 1, \dots, 59\}$ da função de densidade dada por

$$p(b_i | \mathbf{Y}_o, \hat{\psi}) \propto \exp \left[-\frac{1}{2} b_i^2 \hat{\sigma}^{-2} + \sum_{j=1}^{n_i} y_{ij} (\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + \hat{\beta}_3 x_{3ij} + \hat{\beta}_4 x_{4ij} + \hat{\beta}_5 x_{5ij} + b_i) - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + \hat{\beta}_3 x_{3ij} + \hat{\beta}_4 x_{4ij} + \hat{\beta}_5 x_{5ij} + b_i) \right],$$

com as estimativas dos parâmetros obtidas tanto pelo método AL quanto pelo método AQGH. Os resultados mostram-se nos histogramas e funções de densidades estimadas das Figuras B.7 e B.8 do Apêndice B. Além disso, na Tabela B.4 observamos que as predições do intercepto aleatório baseadas na amostragem são similares de um método a outro e com as predições BLUP nos métodos respectivos (vide Tabela 4.2).

Uma vez amostradas as observações calculamos as matrizes $-\ddot{Q}_\psi(\hat{\psi})$, $\Delta_{\omega_{0pmq}}$ e $\Delta_{\omega_{0pmqa}}$ e, por conseguinte, as curvaturas normais individuais sob cada um dos esquemas. Sob o esquema de perturbação multiplicativa quadrática pelo método AL, a Figura 4.14a mostra que as observações

Paciente	Método	
	AL	AQGH
1	0,055	0,055
2	0,067	0,067
3	0,338	0,339
4	0,157	0,157
5	0,020	0,020
6	-0,202	-0,202
7	-0,121	-0,121
8	0,428	0,428
9	-0,182	-0,182
10	0,941	0,941
11	0,167	0,167
12	-0,013	-0,013
13	-0,070	-0,070
14	-0,079	-0,079
15	-0,218	-0,218
16	-0,856	-0,856
17	-0,756	-0,757
18	0,195	0,196
19	-0,247	-0,247
20	-0,110	-0,110
21	0,020	0,020
22	0,335	0,336
23	-0,312	-0,312
24	0,098	0,098
25	0,961	0,962
26	-0,457	-0,458
27	0,039	0,039
28	0,264	0,264
29	-0,312	-0,312
30	-0,154	-0,154
31	-0,321	-0,322
32	0,539	0,540
33	0,459	0,459
34	-0,305	0,306
35	1,019	1,020
36	0,549	0,550
37	0,275	0,276
38	-0,689	-0,689
39	-0,047	-0,047
40	0,005	0,005
41	-0,569	-0,570
42	0,134	0,135
43	0,393	0,394
44	0,016	0,016
45	0,102	0,102
46	0,380	0,381
47	0,110	0,110
48	-0,373	-0,374
49	0,687	0,687
50	-0,212	-0,212
51	-0,195	-0,195
52	-0,771	-0,772
53	0,443	0,444
54	-0,388	-0,389
55	0,202	0,202
56	1,102	1,102
57	-0,646	-0,647
58	-0,937	-0,939
59	0,095	0,095

Tabela 4.2: *Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.32).*

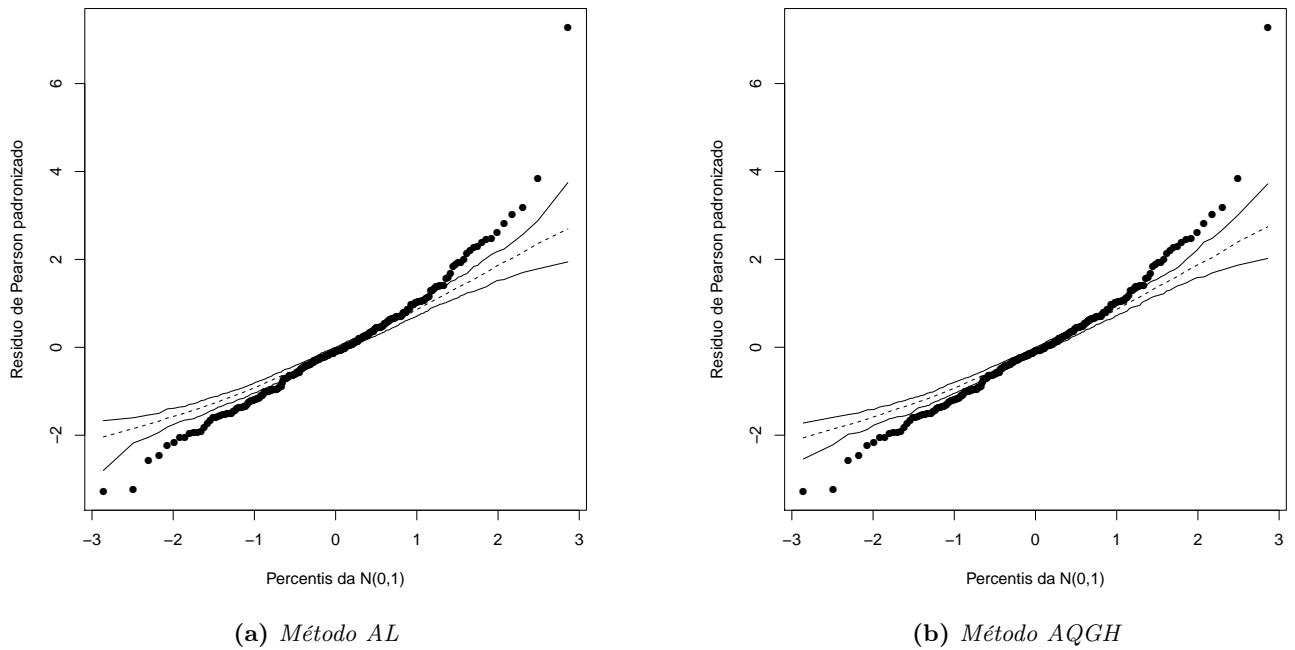
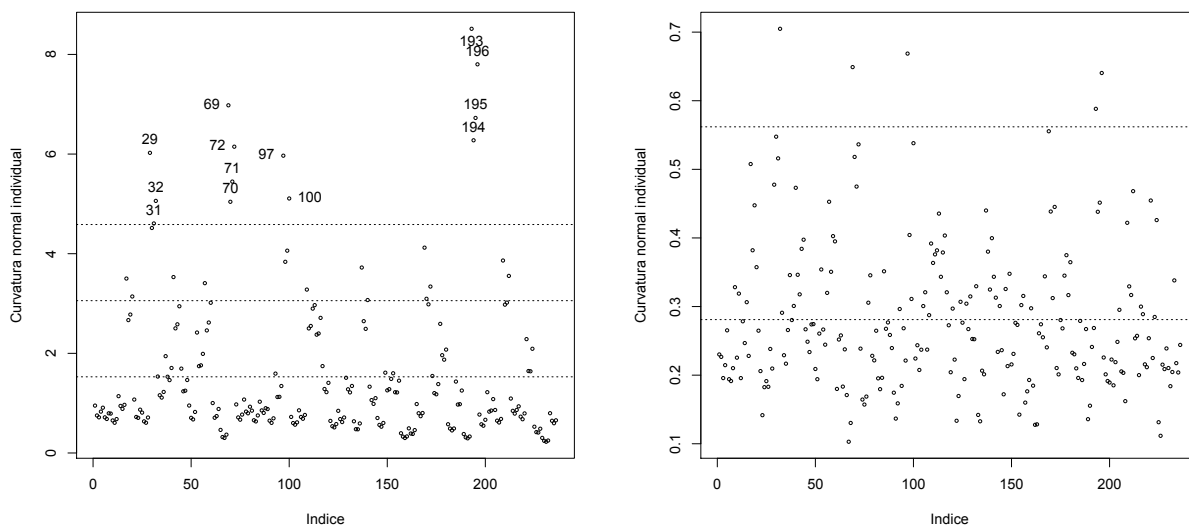


Figura 4.13: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as previsões BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.32).

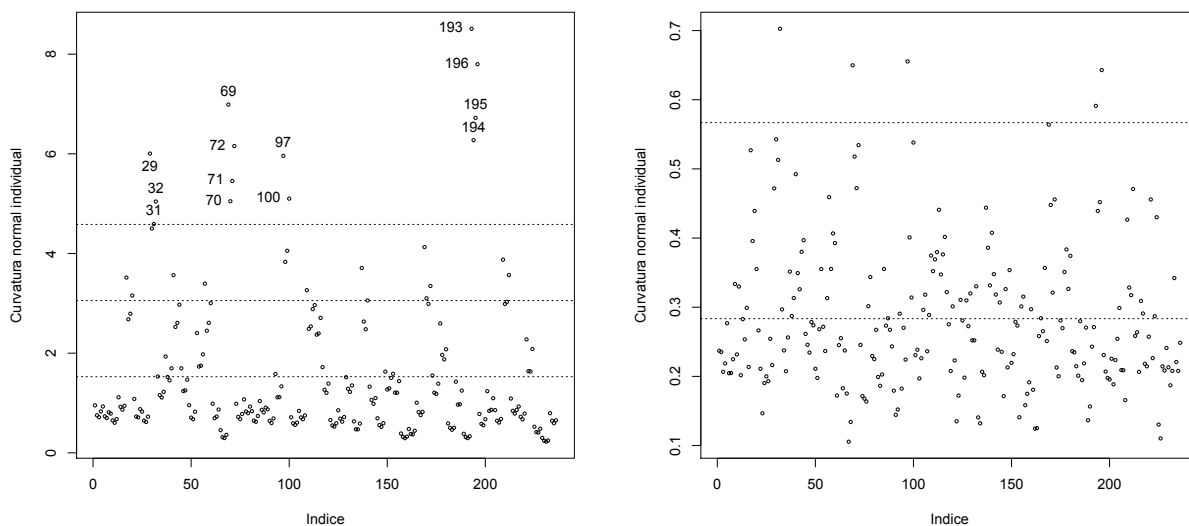
#29, #31 e #32 do paciente 8, #69, #70, #71 e #72 do paciente 18, #97 e #100 do paciente 25 e #193, #194, #195 e #196 do paciente 49 foram detectadas como pontos influentes. Entretanto, sob o esquema perturbação multiplicativa quadrática apropriada pelo método AL, a Figura 4.14b mostra que não houve detecção de pontos influentes. Resultados idênticos observamos pelo método AQGH (vide Figura 4.15a e 4.15b).

Para avaliar a magnitude do impacto exercido pelos conjuntos de observações influentes sobre as estimativas dos parâmetros, calculamos o desvio relativo percentual. Nas Tabelas 4.3 e 4.4 observamos as combinações nas quais obtiveram-se os maiores totais de desvios relativos percentuais pelos métodos AL e AQGH, respectivamente. Pelo método AL, observamos que para cada uma das combinações as estimativas de β_3 e β_5 foram as mais sensíveis ao esquema de perturbação multiplicativa quadrática. Especificamente, o maior total de desvios relativos percentuais obteve-se ao excluir a combinação dada pelos conjuntos #29, #31 e #32 do paciente 8, #69, #70, #71, #72 do paciente 18 e #193, #194, #195 e #196 do paciente 49, com valores relacionados a β_3 e β_5 aproximadamente do 33% e 66%, respectivamente. Por outro lado, as estimativas de β_0 , β_1 , β_2 , β_4 e σ ficaram bastante estáveis e, portanto, foram menos sensíveis ao esquema de perturbação



(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.14: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.32).



(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.15: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.32).

multiplicativa quadrática. Pelo método AQGH observamos os mesmos resultados (vide Tabela 4.4).

Por último, a Tabela 4.5 mostra os resultados do ajuste do modelo sem a combinação mencionada. Logo, comparando estes resultados com os do ajuste do modelo com todas as observações (vide Tabela 4.1), temos que pelo método AL os valores-p das estimativas de β_0 , β_1 e β_2 são os mesmos, mantendo a significância os coeficientes das covariáveis intercepto, grupo e logaritmo do número de crises de um pre-ensaio clínico. O valor-p da estimativa de β_3 aumenta e, portanto, o coeficiente da interação entre grupo e logaritmo do número de crises de um pre-ensaio clínico segue sendo não significativo. Por último, o valor-p da estimativa de β_5 aumenta perdendo a significância o coeficiente da covariável períodos. Os mesmos resultados observamos pelo método AQGH.

Em conclusão, os resultados do análise de diagnóstico de influência local foram importantes porque levaram a mudanças inferenciais no modelo.

Esquema de perturbação multiplicativa quadrática					
Método	Observações excluídas	Pacientes	Parâmetro	Estimativa	DRP_j
AL	Nenhuma	-	β_0	1,793	-
			β_1	-0,334	-
			β_2	0,883	-
			β_3	0,339	-
			β_4	0,481	-
			β_5	-0,030	-
			σ	0,501	-
			#193, #194, #195, #196	49	β_0
	β_1	-0,355			-6,085
	β_2	0,884			-0,041
	β_3	0,193			42,909
	β_4	0,468			2,756
	β_5	-0,022			26,884
	σ	0,491			2,116
	#29, #31, #32 e #193, #194, #195, #196	8 e 49			β_0
			β_1	-0,347	-3,852
			β_2	0,873	1,139
			β_3	0,198	41,483
			β_4	0,430	10,631
			β_5	-0,012	58,231
			σ	0,488	2,639
#29, #31, #32 e #69, #70, #71, #72 e #193, #194, #195, #196			8 e 18 e 49	β_0	1,778
	β_1	-0,337		-0,911	
	β_2	0,845		4,359	
	β_3	0,226		33,316	
	β_4	0,420		12,654	
	β_5	-0,010		66,397	
	σ	0,493		1,546	
	#29, #31, #32 e #69, #70, #71, #72 e #97, #100 e #193, #194, #195, #196	8 e 18 e 25 e 49		β_0	1,790
β_1			-0,354	-5,954	
β_2			0,867	1,861	
β_3			0,207	39,005	
β_4			0,425	11,644	
β_5			-0,014	52,619	
σ			0,513	2,346	

Tabela 4.3: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelo método AL relacionados ao modelo (4.32).

Esquema de perturbação multiplicativa quadrática					
Método	Observações excluídas	Pacientes	Parâmetro	Estimativa	DRP_j
AQGH	Nenhuma	-	β_0	1,793	-
			β_1	-0,334	-
			β_2	0,883	-
			β_3	0,339	-
			β_4	0,481	-
			β_5	-0,030	-
			σ	0,502	-
	#193, #194, #195, #196	49	β_0	1,795	-0,132
			β_1	-0,355	-6,097
			β_2	0,884	-0,040
			β_3	0,193	42,903
			β_4	0,467	2,752
			β_5	-0,022	26,885
			σ	0,492	2,110
	#29, #31, #32 e #193, #194, #195, #196	8 e 49	β_0	1,789	0,234
			β_1	-0,347	-3,867
			β_2	0,873	1,139
			β_3	0,198	41,474
			β_4	0,430	10,623
			β_5	-0,012	58,233
			σ	0,489	2,629
#29, #31, #32 e #69, #70, #71, #72 e #193, #194, #195, #196	8 e 18 e 49	β_0	1,778	0,854	
		β_1	-0,337	-0,923	
		β_2	0,845	4,360	
		β_3	0,226	33,304	
		β_4	0,420	12,649	
		β_5	-0,010	66,398	
		σ	0,495	1,532	
#29, #31, #32 e #69, #70, #71, #72 e #97, #100 e #193, #194, #195, #196	8 e 18 e 25 e 49	β_0	1,790	0,150	
		β_1	-0,354	-5,963	
		β_2	0,867	1,862	
		β_3	0,207	38,997	
		β_4	0,425	11,638	
		β_5	-0,014	52,618	
		σ	0,514	2,358	

Tabela 4.4: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelo método AQGH relacionados ao modelo (4.32).

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	1,778	0,105	<0,001
	β_1	-0,337	0,148	0,023
	β_2	0,845	0,141	<0,001
	β_3	0,226	0,224	0,312
	β_4	0,420	0,343	0,221
	β_5	-0,010	0,012	0,401
	σ	0,493		
AQGH	β_0	1,778	0,106	<0,001
	β_1	-0,337	0,148	0,023
	β_2	0,845	0,141	<0,001
	β_3	0,226	0,225	0,315
	β_4	0,420	0,344	0,222
	β_5	-0,010	0,012	0,401
	σ	0,495		

Tabela 4.5: Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.32).

4.4.2 Aplicação 2

Um conjunto de dados relacionados a um ensaio clínico de tipo cruzado (*cross-over trial*) em 27 pacientes que padecem de enxaquecas é apresentado por McKnight e Eeden (1993). O ensaio consistiu na aplicação dos tratamentos placebo (P) e aspartame (A), subministrados em cinco períodos. No final de cada período registrou-se o número de enxaquecas sofridos pelos pacientes. No primeiro período foi subministrado placebo e nos outros quatro placebo ou aspartame, isto é, PAPAP, PAPP, PPAPA e PPAAP. Os ordenamentos dos tratamentos foram atribuídos aleatoriamente aos pacientes. Para eliminar os efeitos dos tratamentos anteriores, os períodos foram separados por um dia. Todos os períodos foram de sete dias, mas para alguns pacientes estes períodos tiveram uma menor duração. Além disso, para alguns pacientes não houve registro para todos os períodos constituindo assim um conjunto de dados desbalanceados.

Considera-se um MRPEM cuja variável aleatória Y_{ij} é dada pelo número de enxaquecas para o i -ésimo paciente no j -ésimo período com taxa média de ocorrências $\lambda_{ij} > 0$ e cujo intercepto aleatório b_i assume uma distribuição normal de media 0 e variância σ^2 . Logo, o MRPEM é dado por $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$, em que

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (4.33)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 27$ ($q = 122$), com x_{1ij} a variável indicadora do tratamento: 1 para aspartame e 0 para placebo, x_{2ij} o logaritmo do número de dias de cada período e β_0 , β_1 e β_2 os coeficientes de regressão. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \sigma^2)'$.

Nas Tabelas 4.6 e 4.7 encontram-se os resultados das estimativas de máxima verossimilhança dos parâmetros e das predições BLUP do intercepto aleatório pelos métodos AL e AQGH, respectivamente. Na Tabela 4.6 observamos que com um nível de significância de 5% não existem

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-0,889	0,763	0,244
	β_1	0,575	0,386	0,136
	β_2	0,258	0,145	0,075
	σ	0,677		
AQGH	β_0	-0,893	0,763	0,242
	β_1	0,576	0,386	0,136
	β_2	0,258	0,145	0,075
	σ	0,682		

Tabela 4.6: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.33).

Paciente	Método	
	AL	AQGH
1	-0,003	-0,003
2	0,393	0,396
3	0,371	0,373
4	-1,087	-1,094
5	-0,241	-0,241
6	-0,117	-0,117
7	0,371	0,373
8	-0,311	-0,312
9	-0,025	-0,024
10	-1,087	-1,094
11	-0,174	-0,174
12	-0,241	-0,241
13	1,477	1,479
14	-0,003	-0,003
15	-0,003	-0,003
16	0,130	0,131
17	0,677	0,682
18	-0,527	-0,529
19	-0,527	-0,529
20	0,678	0,682
21	0,657	0,659
22	0,377	-0,378
23	0,778	0,780
24	-0,377	-0,378
25	0,657	0,659
26	-0,352	-0,355
27	0,425	0,427

Tabela 4.7: Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.33).

coeficientes significativos e que o desvio padrão do intercepto aleatório é próximo a 0,680, pelo método AL e AQGH. Na Tabela 4.7 observamos que as previsões BLUP do intercepto aleatório são similares entre os dois métodos. Verificamos também mediante as Figuras 4.16a e 4.16b que o ajuste do modelo parece ser razoável, por ambos os métodos.

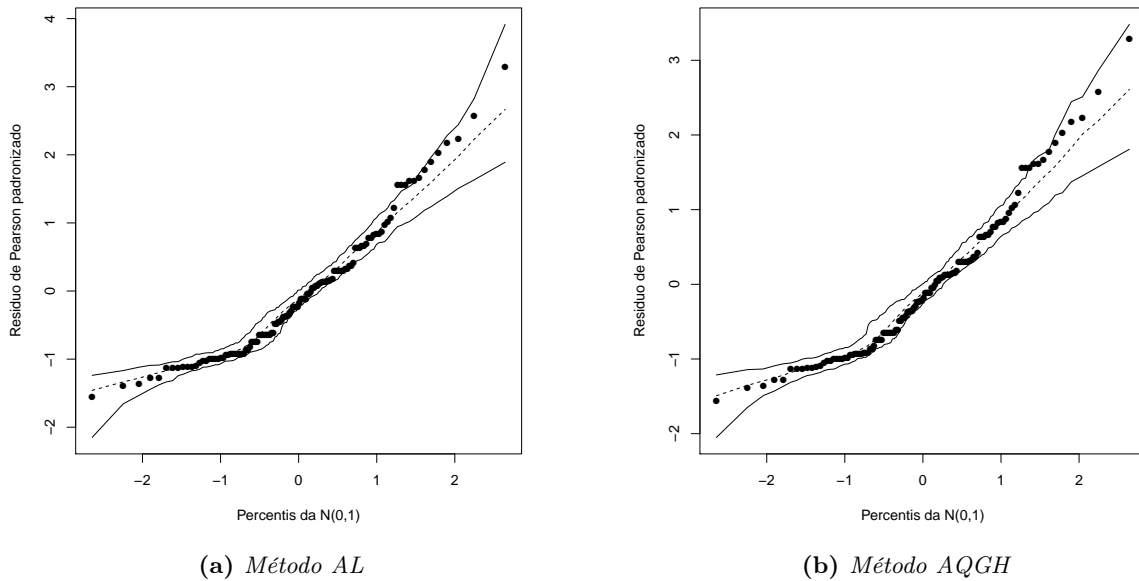


Figura 4.16: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as previsões BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.33).

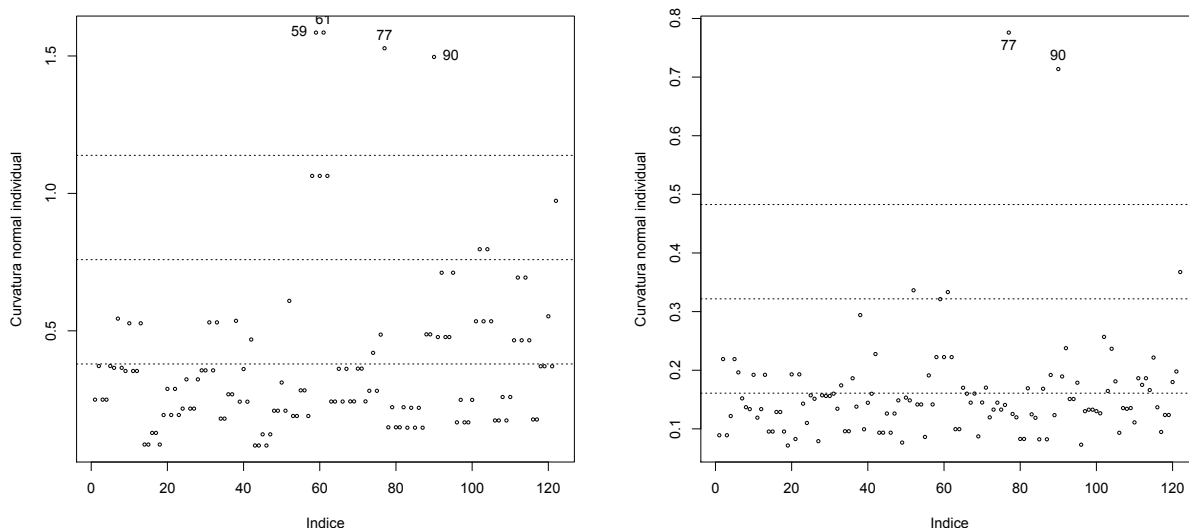
Seguindo com o procedimento amostramos $\{b_i^{(s)}; s = 1, \dots, 10.000, i = 1, \dots, 27\}$ da função de densidade

$$p(b_i | \mathbf{Y}_o, \hat{\boldsymbol{\psi}}) \propto \exp \left[-\frac{1}{2} b_i^2 \hat{\sigma}^{-2} + \sum_{j=1}^{n_i} y_{ij} (\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i) - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{\beta}_2 x_{2ij} + b_i) \right],$$

utilizando o algoritmo Metropolis-Hastings e as estimativas dos parâmetros obtidas tanto pelo método AL quanto pelo método AQGH. Os resultados mostram-se nas Figuras B.9 e B.10. Além disso, na Tabela B.5 observamos que as previsões do intercepto aleatório baseadas na amostragem são similares entre os métodos e com as previsões BLUP nos métodos respectivos (vide Tabela 4.7).

Uma vez amostradas as observações calculamos as matrizes $-\ddot{Q}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$, $\Delta_{\omega_{0pmq}}$ e $\Delta_{\omega_{0pmqa}}$ e, por conseguinte, as curvaturas normais individuais sob cada um dos esquemas. Sob o esquema de perturbação multiplicativa quadrática pelo método AL, a Figura 4.17a mostra que as observações #59 e #61 do paciente 13, #77 do paciente 17 e #90 do paciente 20 foram detectadas como pontos influentes. Sob o esquema de perturbação multiplicativa quadrática apropriada pelo método AL, a

Figura 4.17b mostra que as observações #77 do paciente 17 e #90 do paciente 20 foram detectadas como pontos influentes. Os mesmos resultados observamos pelo método AQGH (vide Figura 4.18a e 4.18b).



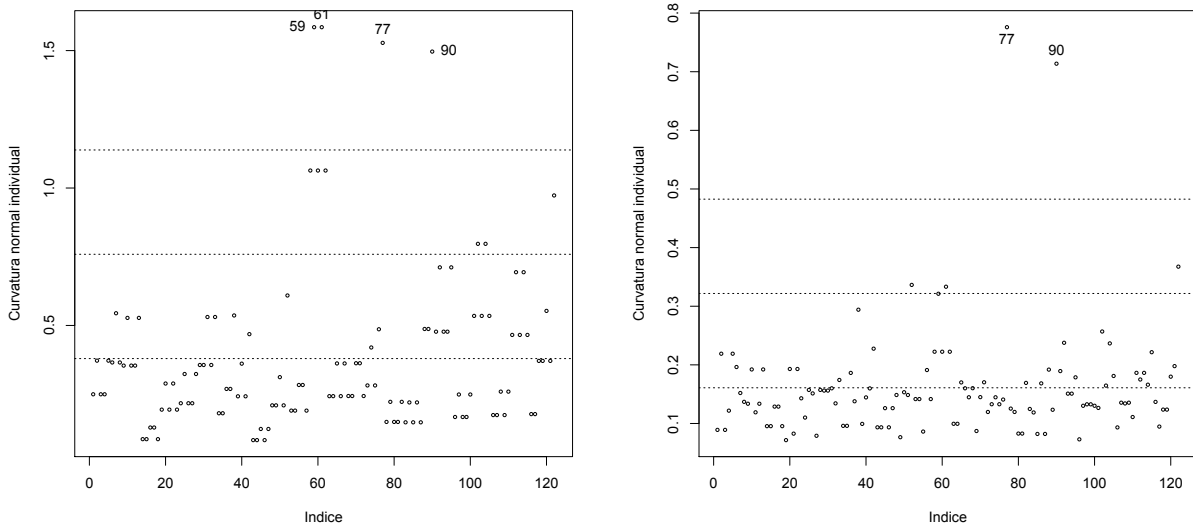
(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.17: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.33).

Para avaliar a magnitude do impacto exercido nas estimativas dos parâmetros pelas observações influentes e conjuntos de observações influentes, calculamos o desvio relativo percentual. As Tabelas 4.8 e 4.9 mostram os resultados para cada uma das combinações onde se teve os maiores totais de desvios relativos percentuais, tanto pelo método AL quanto pelo método AQGH, para os esquemas de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada, respectivamente.

Observamos que sob o esquema de perturbação multiplicativa quadrática pelo método AL, o maior total de desvios relativos percentuais obteve-se ao excluir a combinação dada pelas observações #59 e #61 do paciente 13, #77 do paciente 17 e #90 do paciente 20. Especificamente, as mudanças foram aproximadamente do 88% e 72% para as estimativas β_0 e β_1 , respectivamente. No entanto, as mudanças foram pequenas para as estimativas β_2 e σ , de modo que estas estimativas foram menos sensível ao esquema de perturbação multiplicativa quadrática. Os mesmos resultados observamos pelo método AQGH.

Sob o esquema de perturbação multiplicativa quadrática apropriada pelo método AL, o maior total de desvios relativos percentuais obteve-se ao excluir as observações #77 do paciente 17 e



(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.18: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.33).

#90 do paciente 20, com mudanças para as estimativas β_0 e β_1 aproximadamente do 92% e 74%, respectivamente. No entanto, as mudanças foram pequenas para as estimativas β_2 e σ , de modo que estas estimativas foram menos sensível ao esquema de perturbação multiplicativa quadrática apropriada. Os mesmos resultados observamos pelo método AQGH.

Para finalizar, as Tabelas 4.10 e 4.11 mostram os resultados do ajuste do modelo sem as combinações mencionadas para os esquemas de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada, respectivamente. Assim, comparando estes resultados com os do ajuste do modelo com todas as observações (vide Tabela 4.6) observamos que sob o esquema de perturbação multiplicativa quadrática, tanto pelo método AL quanto pelo método AQGH, os valores-p das estimativas de β_0 e β_1 aumentaram e o valor-p da estimativa de β_2 diminuiu ficando muito perto do nível de significância de 5%. Resultados similares observamos para o esquema de perturbação multiplicativa quadrática apropriada.

Em conclusão, os resultados da análise de diagnóstico de influência local sob os esquemas de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada, pelos métodos AL e AQGH, levaram a importantes variações nas estimativas dos parâmetros mas não a mudanças inferenciais.

Esquema de perturbação multiplicativa quadrática					
Método	Observações excluídas	Pacientes	Parâmetro	Estimativa	DRP_j
AL	Nenhuma	-	β_0	-0,889	-
			β_1	0,575	-
			β_2	0,258	-
			σ	0,677	-
	#90	20	β_0	-0,639	28,132
			β_1	0,444	22,722
			β_2	0,263	-1,935
			σ	0,681	-0,532
	#77 e #90	17 e 20	β_0	-0,073	91,758
			β_1	0,151	73,784
			β_2	0,265	-2,703
			σ	0,689	-1,786
	#59, #61 e #77 e #90	13 e 17 e 20	β_0	-0,103	88,454
			β_1	0,159	72,374
			β_2	0,299	-15,914
			σ	0,704	-3,998
AQGH	Nenhuma	-	β_0	-0,893	-
			β_1	0,576	-
			β_2	0,258	-
			σ	0,682	-
	#90	20	β_0	-0,642	28,114
			β_1	0,445	22,742
			β_2	0,263	-1,937
			σ	0,686	-0,536
	#77 e #90	17 e 20	β_0	-0,072	91,929
			β_1	0,150	74,028
			β_2	0,265	-2,707
			σ	0,694	-1,797
	#59, #61 e #77 e #90	13 e 17 e 20	β_0	-0,102	88,576
			β_1	0,158	72,592
			β_2	0,299	-16,128
			σ	0,710	-4,046

Tabela 4.8: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.33).

Esquema de perturbação multiplicativa quadrática apropriada					
Método	Observações excluídas	Pacientes	Parâmetro	Estimativa	DRP_j
AL	Nenhuma	-	β_0	-0,889	-
			β_1	0,575	-
			β_2	0,258	-
			σ	0,677	-
	#90	20	β_0	-0,639	28,132
			β_1	0,444	22,722
			β_2	0,263	-1,935
			σ	0,681	-0,532
	#77 e #90	17 e 20	β_0	-0,073	91,758
			β_1	0,151	73,784
			β_2	0,265	-2,703
			σ	0,689	-1,786
AQGH	Nenhuma	-	β_0	-0,893	-
			β_1	0,576	-
			β_2	0,258	-
			σ	0,682	-
	#90	20	β_0	-0,642	28,114
			β_1	0,445	22,742
			β_2	0,263	-1,937
			σ	0,686	-0,536
	#77 e #90	17 e 20	β_0	-0,072	91,929
			β_1	0,150	74,028
			β_2	0,265	-2,707
			σ	0,694	-1,797

Tabela 4.9: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.33).

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-0,103	1,273	0,936
	β_1	0,159	0,653	0,808
	β_2	0,299	0,157	0,058
	σ	0,704		
AQGH	β_0	-0,102	1,274	0,936
	β_1	0,158	0,654	0,809
	β_2	0,299	0,157	0,057
	σ	0,710		

Tabela 4.10: Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.33).

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	-0,073	1,267	0,954
	β_1	0,151	0,651	0,817
	β_2	0,265	0,145	0,068
	σ	0,689		
AQGH	β_0	-0,072	1,267	0,955
	β_1	0,150	0,651	0,818
	β_2	0,265	0,145	0,068
	σ	0,694		

Tabela 4.11: Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.33).

4.4.3 Aplicação 3

Nesta aplicação consideramos um estudo de tipo caso-controle correspondente ao “Ciclo vigília/sono em pacientes com diabetes mellitus tipo I” apresentado em [Giampaoli *et al.* \(2014\)](#). Entre as informações do estudo encontram-se as observações das seguintes variáveis: grupo (caso: indivíduos com diabetes mellitus tipo I e controle: indivíduos sem a doença), latência do sono em minutos e quantidade de despertares noturnos. As variáveis foram registradas no decorrer de dez dias para 27 indivíduos, mas houve registros com menor duração, constituindo assim um conjunto de dados desbalanceados.

Para este problema vamos supor um MRPEM cuja variável aleatória Y_{ij} é dada pela quantidade de despertares noturnos para o i -ésimo indivíduo no j -ésimo dia com taxa média de ocorrências $\lambda_{ij} > 0$, e cujo intercepto aleatório b_i assume uma distribuição normal de media 0 e variância σ^2 . Logo, o MRPEM é dado por $Y_{ij} \sim Poisson(\lambda_{ij})$, em que

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i, \quad (4.34)$$

para $j = 1, \dots, n_i$ e $i = 1, \dots, 27$ ($q = 264$), com x_{1ij} a variável indicadora do grupo: 1 para caso e 0 para controle, x_{2ij} a latência de sono em minutos e β_0, β_1 e β_2 os coeficientes de regressão. O vetor de parâmetros desconhecidos é dado por $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \sigma^2)'$.

Os resultados das estimativas de máxima verossimilhança dos parâmetros e das predições BLUP do intercepto aleatório pelos métodos AL e AQGH encontram-se nas Tabelas 4.12 e 4.13, respectivamente. Observamos que com um nível de significância de 5% os coeficientes de regressão não são significativos e que o desvio padrão do intercepto aleatório é próximo a 0,303. Observamos também que as predições BLUP do intercepto aleatório não apresentam diferenças entre os métodos AL e AQGH. Por outro lado, verificamos mediante as Figuras 4.19a e 4.19b que os resíduos de Pearson padronizados com as predições BLUP do intercepto aleatório cruzam o envelope gerado sugerindo sobredispersão, pelos métodos AL e AQGH.

Realizamos a análise de diagnóstico de influência local da mesma maneira que nas aplicações anteriores. Nas Figuras B.11 e B.12 mostram-se os resultados da amostragem. Na Tabela B.6 observamos que as predições do intercepto aleatório baseadas na amostragem são similares entre os métodos AL e AQGH e com as predições BLUP nos métodos respectivos (vide Tabela 4.13).

A Figura 4.20a mostra que sob o esquema de perturbação multiplicativa quadrática, pelo método AL, as observações #88 do indivíduo 10 e #185 e #188 do indivíduo 20 foram detectadas como pontos influentes. A Figura 4.20b mostra que sob o esquema de perturbação multiplicativa

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	1,483	0,199	<0,001
	β_1	-0,098	0,141	0,486
	β_2	0,002	0,002	0,409
	σ	0,302		
AQGH	β_0	1,483	0,199	<0,001
	β_1	-0,098	0,141	0,486
	β_2	0,002	0,002	0,409
	σ	0,303		

Tabela 4.12: Estimativas de máxima verossimilhança, erros padrão e valores-p pelos métodos AL e AQGH relacionados ao modelo (4.34).

Indivíduo	Método	
	AL	AQGH
1	0,332	0,332
2	-0,012	-0,012
3	0,176	0,176
4	-0,046	-0,046
5	0,116	0,116
6	0,301	0,301
7	0,212	0,212
8	0,401	0,401
9	0,059	0,059
10	0,051	0,051
11	-0,588	-0,559
12	-0,392	-0,392
13	0,123	0,124
14	-0,357	-0,357
15	-0,149	-0,149
16	-0,009	-0,009
17	-0,033	-0,033
18	-0,340	-0,340
19	-0,291	-0,291
20	0,272	0,272
21	-0,337	-0,337
22	0,499	0,499
23	0,238	0,238
24	0,056	0,056
25	0,031	0,032
26	0,147	0,147
27	-0,246	-0,246

Tabela 4.13: Predições BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.34).

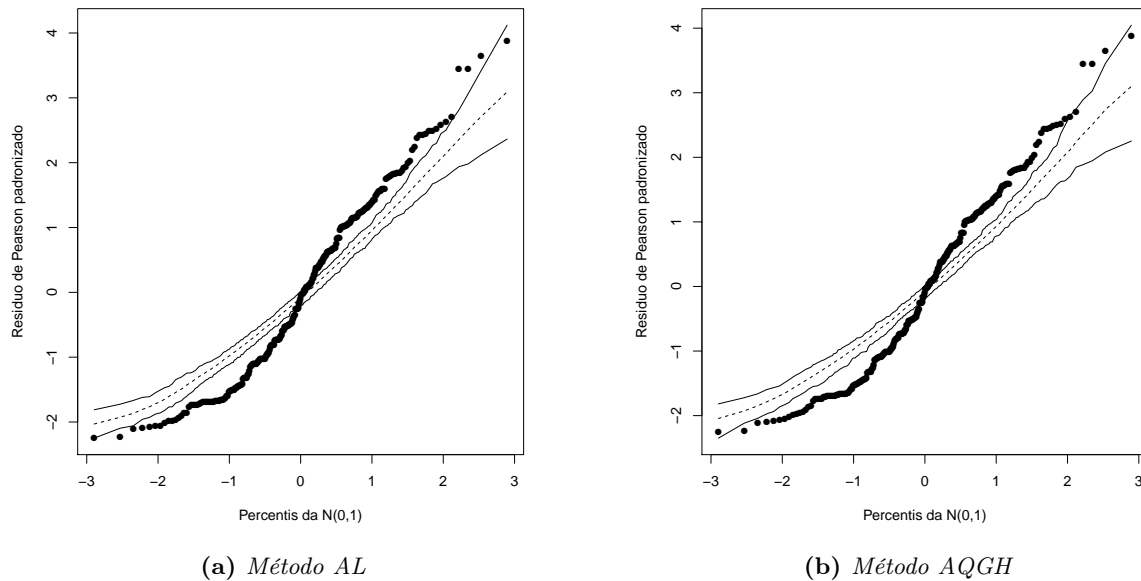
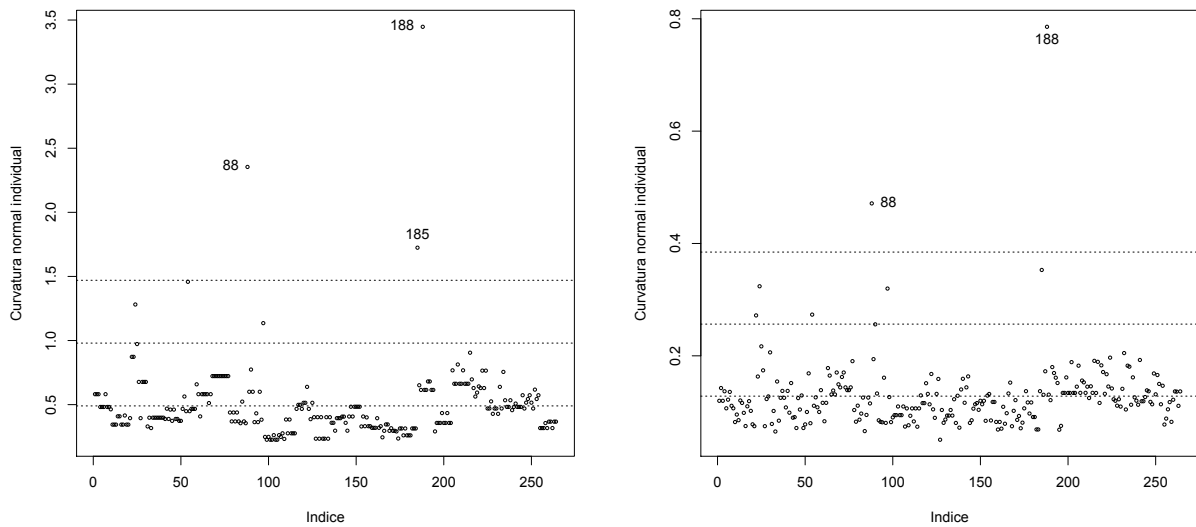


Figura 4.19: Gráficos de probabilidade normal com envelope simulado para os resíduos de Pearson padronizados com as previsões BLUP do intercepto aleatório pelos métodos AL e AQGH relacionados ao modelo (4.34).

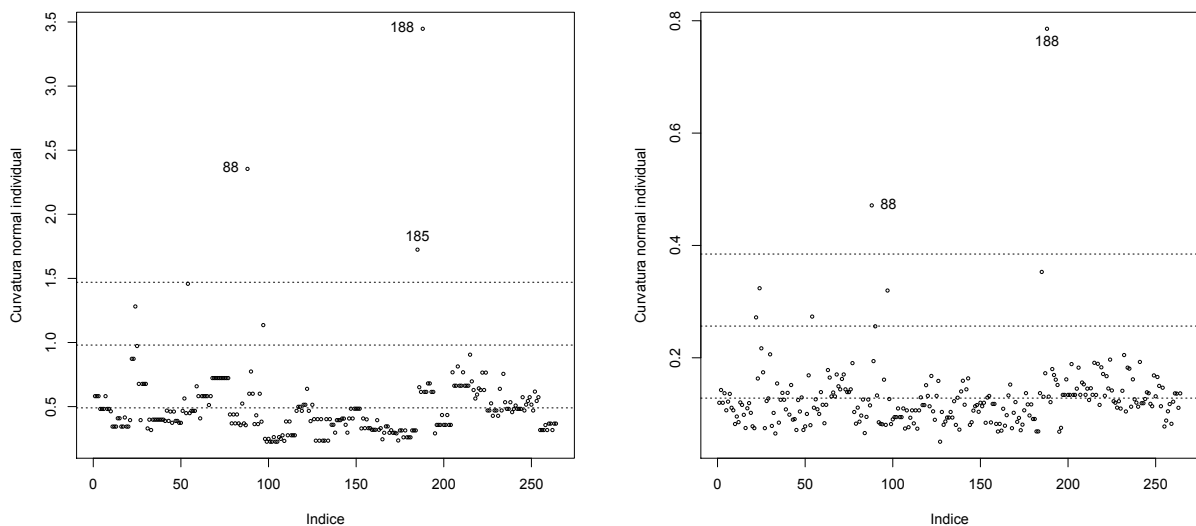
quadrática apropriada, pelo método AL, as observações #88 do indivíduo 10 e #188 do indivíduo 20 foram detectadas como pontos influentes. Os mesmos resultados observamos pelo método AQGH (vide Figura 4.21a e 4.21b).

Ao excluir as observações mencionadas, observamos nas Tabelas 4.14 e 4.15 que sob os dois esquemas por ambos os métodos, o maior impacto foi sofrido pela estimativa de β_2 . Entretanto, ao realizar os ajustes sem estas observações, observamos nas Tabelas 4.16 e 4.17 que não houve mudanças inferenciais a serem destacadas.



(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.20: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AL relacionados ao modelo (4.34).



(a) Esquema de perturbação multiplicativa quadrática (b) Esquema de perturbação multiplicativa quadrática apropriada

Figura 4.21: Gráficos das curvaturas normais individuais versus o índice das observações para o (a) esquema de perturbação multiplicativa quadrática e o (b) esquema de perturbação multiplicativa quadrática apropriada pelo método AQGH relacionados ao modelo (4.34).

Esquema de perturbação multiplicativa quadrática					
Método	Observações excluídas	Indivíduos	Parâmetro	Estimativa	DRP_j
AL	Nenhuma	-	β_0	1,483	-
			β_1	-0,098	-
			β_2	0,002	-
			σ	0,302	-
	#88	10	β_0	1,486	-0,210
			β_1	-0,009	10,456
			β_2	-0,000	103,431
			σ	0,302	0,124
	#88 e #185, #188	10 e 20	β_0	1,541	-3,943
			β_1	-0,098	0,131
			β_2	-0,004	345,203
			σ	0,297	1,796
AQGH	Nenhuma	-	β_0	1,483	-
			β_1	-0,098	-
			β_2	0,002	-
			σ	0,303	-
	#88	10	β_0	1,486	-0,211
			β_1	-0,009	10,450
			β_2	-0,000	103,428
			σ	0,303	0,125
	#88 e #185, #188	10 e 20	β_0	1,541	-3,942
			β_1	-0,098	0,128
			β_2	-0,004	345,141
			σ	0,297	1,792

Tabela 4.14: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.34).

Esquema de perturbação multiplicativa quadrática apropriada					
Método	Observações excluídas	Indivíduos	Parâmetro	Estimativa	DRP_j
AL	Nenhuma	-	β_0	1,483	-
			β_1	-0,098	-
			β_2	0,002	-
			σ	0,302	-
	#88	10	β_0	1,486	-0,210
			β_1	-0,009	10,456
			β_2	-0,000	103,431
			σ	0,303	0,124
	#88 e #188	10 e 20	β_0	1,487	-0,292
			β_1	-0,088	10,278
			β_2	-0,000	109,043
			σ	0,303	0,126
AQGH	Nenhuma	-	β_0	1,483	-
			β_1	-0,098	-
			β_2	0,002	-
			σ	0,303	-
	#88	10	β_0	1,486	-0,211
			β_1	-0,009	10,450
			β_2	-0,000	103,428
			σ	0,303	0,125
	#88 e #188	10 e 20	β_0	1,487	-0,291
			β_1	-0,088	10,274
			β_2	-0,000	108,972
			σ	0,303	0,128

Tabela 4.15: Estimativas de máxima verossimilhança e desvios relativos percentuais para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.34).

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	1,541	0,201	<0,001
	β_1	-0,098	0,139	0,480
	β_2	-0,004	0,004	0,328
	σ	0,297		
AQGH	β_0	1,541	0,202	<0,001
	β_1	-0,098	0,139	0,481
	β_2	-0,004	0,004	0,328
	σ	0,297		

Tabela 4.16: Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática pelos métodos AL e AQGH relacionados ao modelo (4.34).

Método	Parâmetro	Estimativa	Erro padrão	Valor-p
AL	β_0	1,487	0,202	<0,001
	β_1	-0,088	0,141	0,533
	β_2	-0,000	0,003	0,965
	σ	0,303		
AQGH	β_0	1,487	0,202	<0,001
	β_1	-0,088	0,141	0,533
	β_2	-0,000	0,003	0,965
	σ	0,303		

Tabela 4.17: Estimativas de máxima verossimilhança, erros padrão e valores-p sem as observações influentes para o esquema de perturbação multiplicativa quadrática apropriada pelos métodos AL e AQGH relacionados ao modelo (4.34).

4.5 Aspectos computacionais

4.5.1 Recursos computacionais

Para os estudos de simulação e as aplicações apresentadas foram utilizados os mesmos recursos computacionais descritos na Seção 3.5.1.

4.5.2 Problemas computacionais dos estudos de simulação

Os estudos de simulação 1 e 2 apresentaram os mesmos problemas computacionais que os estudos de simulação 1 e 2 do Capítulo 3. Deste modo, estes foram desenvolvidos considerando aquelas réplicas que não apresentavam falta de convergência (falsa ou singular) e que cumpriam com as condições $\hat{\beta} \in (-0.5, 1.5)$ e $\hat{\sigma} \in (0, 01; 3\sigma)$. Na Tabela 4.18 encontram-se os resultados do número de réplicas adicionais que tiveram que ser geradas para completar as 100 réplicas para cada tamanho amostral q e valor da perturbação ω_{ij}^2 . Observamos que o número de réplicas adicionais foi pequeno indicando poucos problemas de convergência e poucos valores da estimativa de σ^2 próximos a zero.

Número de réplicas adicionais						
Estudo de simulação	q	I	n_i	ω_{ij}^2		
				0,8	1,2	1,5
1	90	30	3	3	2	2
	360	90	6	0	0	0
	1080	120	9	0	0	0
2	90	15	6	2	2	2
		30	3	3	2	2
	360	30	12	0	0	1
		60	6	0	0	0
	1080	60	18	0	0	0
		120	9	1	0	0

Tabela 4.18: Número de réplicas adicionais para cada tamanho amostral q e valor da perturbação ω_{ij}^2 nos estudos de simulação 1 e 2. (MRPEM)

4.5.3 Tempos de execução dos estudos de simulação

Com respeito aos tempos de execução dos cenários dos estudos de simulação 1 e 2, a Tabela 4.19 mostra os resultados em horas para cada tamanho amostral q e valor da perturbação ω_{ij}^2 . Em ambos estudos pode-se observar que para cada valor da perturbação, à medida que o tamanho amostral aumenta o tempo de execução aumenta, conforme esperado. Além disso, para o estudo de simulação 2 cabe destacar que o tempo de execução foi maior no caso em que para um valor fixado de q tem-se um número de grupos I maior.

Tempos de execução (em horas)						
Estudo de simulação	q	I	n_i	ω_{ij}^2		
				0,8	1,2	1,5
1	90	30	3	15,28	23,07	22,45
	360	60	6	52,73	55,32	54,85
	1080	120	9	157,17	159	158,95
2	90	15	6	11,6	12,8	12,87
		30	3	15,28	23,07	22,45
	360	30	12	32,03	31,93	31,95
		60	6	52,73	55,32	54,85
	1080	60	18	89,45	89,52	89,52
		120	9	157,17	159	158,95

Tabela 4.19: Tempos de execução (em horas) dos estudos de simulação 1 e 2 para cada tamanho amostral q e valor da perturbação ω_{ij}^2 . (MRPEM)

Capítulo 5

Conclusões

Neste capítulo apresentamos as considerações finais e as sugestões para pesquisas futuras.

5.1 Considerações finais

Neste trabalho realizamos análise de diagnóstico de influência local em modelos lineares generalizados mistos com variável resposta discreta. Especificamente, desenvolvemos a técnica de análise de diagnóstico de influência local no modelo de regressão logística com efeitos mistos sob uma nova estratégia de perturbação indireta para a variável resposta binária, a qual é dada pelo esquema de perturbação multiplicativa da probabilidade de sucesso. Também desenvolvemos a técnica no modelo de regressão Poisson com efeitos mistos sob uma estratégia de perturbação para a variável resposta de contagem dada pelo esquema de perturbação multiplicativa quadrática da taxa média de ocorrências. Nos dois casos, desenvolvemos as análises sob as versões apropriadas dos esquemas de perturbação, de acordo com [Chen *et al.* \(2010\)](#).

Para ambos os modelos, avaliamos o desempenho das propostas mediante as proporções de tipos de pontos influentes, isto é, verdadeiros influentes, falsos influentes e influentes, com dois estudos de simulação de Monte Carlo, para diferentes cenários definidos pelo número de grupos e tamanho dos grupos. Para o modelo de regressão logística com efeitos mistos, em relação ao primeiro estudo de simulação, obtivemos que as proporções para os verdadeiros influentes foram maiores sob o esquema de perturbação multiplicativa apropriada do que sob o esquema de perturbação multiplicativa, para os falsos influentes foram maiores sob o esquema de perturbação multiplicativa do que sob o esquema de perturbação multiplicativa apropriada e para os influentes foram similares entre os esquemas. Com respeito ao segundo estudo de simulação, obtivemos que para um determinado tamanho amostral, em geral, as proporções não foram afetadas pela relação número de grupos e

tamanho dos grupos sob cada um dos esquemas. Para o modelo de regressão Poisson com efeitos mistos, em relação ao primeiro estudo de simulação, obtivemos que as proporções para os verdadeiros influentes e influentes foram maiores sob o esquema de perturbação multiplicativa quadrática do que sob o esquema de perturbação multiplicativa quadrática apropriada, entretanto, as proporções para os falsos influentes foram similares entre os esquemas. Com respeito ao segundo estudo de simulação, também obtivemos que para um determinado tamanho amostral, em geral, as proporções não foram afetadas pela relação número de grupos e tamanho dos grupos sob cada um dos esquemas. Por último, para ambos os modelos, nos dois estudos de simulação, em geral, as proporções foram maiores para os verdadeiros influentes que para os falsos influentes, conforme esperado.

Para ilustrar as propostas apresentamos três aplicações a conjuntos de dados reais para cada modelo. Para o modelo de regressão logística com efeitos mistos os resultados da análise de diagnóstico de influência local sob os esquemas de perturbação multiplicativa e multiplicativa apropriada da probabilidade de sucesso, levaram a importantes variações nas estimativas dos parâmetros e em algumas aplicações a mudanças inferenciais. Além disso, para todas as aplicações a exclusão das observações detectadas como influentes levaram ao melhoramento da capacidade preditiva do modelo. Para o modelo de regressão Poisson com efeitos mistos os resultados da análise de diagnóstico de influência local sob os esquemas de perturbação multiplicativa quadrática e multiplicativa quadrática apropriada da taxa média de ocorrências também revelaram ser importantes.

Por fim, para ambos os modelos, tanto nos estudos de simulação como nas aplicações, as análises de diagnóstico de influência local com as estimativas de máxima verossimilhança obtidas pelo método de integração de aproximação de Laplace como pelo método de adaptativa quadratura de Gauss-Hermite, em geral, apresentaram resultados similares.

Em conclusão, com este trabalho mostramos a aplicabilidade da técnica de análise de diagnóstico de influência local em os modelos lineares generalizados mistos com variável resposta discreta, apoiando com evidências empíricas a capacidade desta mesma para a detecção de observações influentes sob as estratégias de perturbação propostas.

5.2 Sugestões para pesquisas futuras

Como perspectivas de pesquisas futuras pretende-se desenvolver:

- Para o modelo de regressão logística de efeitos mistos:
 - o método de estimação de parâmetros SAEM (Meza *et al.*, 2009);

- análise de diagnóstico de influência local de primeira ordem para a probabilidade de sucesso (Cadigan e Farrell, 2002);
 - análise de diagnóstico de influência local de primeiro ordem para a razão de chances (Larsen *et al.*, 2000);
 - análise de diagnóstico de influência local para a predição da probabilidade de sucesso (Thomas e Cook, 1990).
- Para o modelo de regressão Poisson de efeitos mistos:
 - o método de estimação de parâmetros SAEM;
 - análise de diagnóstico de influência local de primeira ordem para a taxa média de ocorrências;
 - análise de diagnóstico de influência local para a predição da taxa média de ocorrências.
 - Estender as propostas anteriores considerando outra distribuição para os efeitos aleatórios. (Fabio *et al.*, 2012; Pinheiro *et al.*, 2001)

Apêndice A

Derivadas

Neste apêndice apresentamos as derivadas primeiras e segundas relacionadas às matrizes $-\ddot{Q}_\psi(\hat{\psi})$ e Δ_{ω_0} dos Capítulos 3 e 4.

A.1 Derivadas relacionadas ao Capítulo 3

Seja o logaritmo da função de verossimilhança de ψ dado os dados completos dado por

$$L(\psi|\mathbf{Y}_c) = \sum_{i=1}^I \left\{ \sum_{j=1}^{n_i} \left[y_{ij} \log \left(\frac{p_{ij}}{1-p_{ij}} \right) - \log \left(\frac{1}{1-p_{ij}} \right) \right] - \frac{1}{2} \mathbf{b}'_i \Sigma^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\Sigma)) \right\},$$

com

$$p_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \quad e \quad \eta_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i.$$

Assim, as derivadas primeiras e segundas relacionadas à matriz $-\ddot{Q}_\psi(\hat{\psi})$ são dadas por

$$\frac{\partial L(\psi|\mathbf{Y}_c)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^I \sum_{j=1}^{n_i} \left[y_{ij} \mathbf{x}_{ij} - \frac{\exp(\eta_{ij}) \mathbf{x}_{ij}}{1 + \exp(\eta_{ij})} \right]_{(p_1 \times 1)},$$

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^I \sum_{j=1}^{n_i} \frac{\exp(\eta_{ij}) \mathbf{x}_{ij} \mathbf{x}'_{ij}}{[1 + \exp(\eta_{ij})]^2} \Big|_{(p_1 \times p_1)},$$

$$-\frac{\partial^2 L(\psi|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = \mathbf{0}_{(p_1 \times p_4)},$$

$$\frac{\partial L(\psi|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma}} = -\frac{I}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}'_i \Sigma^{-1} \Big|_{(p_2 \times p_2)},$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} = \mathbf{0}_{(p_4 \times p_1)}$$

e

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\frac{I}{2}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \right) \otimes \boldsymbol{\Sigma}_{(p_4 \times p_4)}^{-1}.$$

Seja o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos sob o esquema de perturbação multiplicativa dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c) = \sum_{i=1}^I \left(\sum_{j=1}^{n_i} \left\{ y_{ij} \log \left(\frac{p_{ij}(\omega_{ij})}{1 - p_{ij}(\omega_{ij})} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij})} \right) \right\} - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right), \quad (\text{A.1})$$

com

$$p_{ij}(\omega_{ij}) = p_{ij} \omega_{ij}, \quad \omega_{ij} \in (0, 1].$$

Então, as derivadas primeiras e segundas relacionadas à matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pm}}$ são dadas por

$$\frac{\partial L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c)}{\partial \omega_{ij}} = \frac{y_{ij}(1 + \exp(\eta_{ij}))}{\omega_{ij}(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})} - \frac{\exp(\eta_{ij})}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = \frac{(y_{ij} - 1) \exp(\eta_{ij}) \mathbf{x}_{ij}}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})^2},$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0},$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Para a definição do esquema de perturbação multiplicativa apropriada considere a expressão dada em (A.1). Logo, a função de densidade associada é dada por

$$p(\mathbf{Y}_c|\boldsymbol{\psi}, \boldsymbol{\omega}) = \prod_{i=1}^I \left(\prod_{j=1}^{n_i} \left\{ y_{ij} \log \left(\frac{p_{ij}(\omega_{ij})}{1 - p_{ij}(\omega_{ij})} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij})} \right) \right\} - \frac{1}{2} \mathbf{b}_i' \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right).$$

Assim, as derivadas primeiras e segundas relacionadas à matriz $G(\boldsymbol{\omega})$ são dadas por

$$\frac{\partial \log p(\mathbf{Y}_c|\boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_{ij}} = \frac{y_{ij}(1 + \exp(\eta_{ij}))}{\omega_{ij}(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})} - \frac{\exp(\eta_{ij})}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})},$$

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = \frac{y_{ij}(1 + \exp(\eta_{ij}))(1 + \exp(\eta_{ij}) - 2 \exp(\eta_{ij})\omega_{ij})}{\omega_{ij}^2(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})^2} + \frac{\exp(2\eta_{ij})}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij})^2}, \quad (\text{A.2})$$

para $i = j$, e

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0,$$

para $i \neq j$, com $i, j = 1, \dots, q$.

Como $G(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \neq a\mathbf{I}_q$ escolhe-se um novo vetor de perturbação dado por

$$\omega_{ij}^* = 1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\omega_{ij} - 1),$$

para $i, j = 1, \dots, q$, do modo que o esquema de perturbação multiplicativa apropriada fica dado por

$$p_{ij}(\omega_{ij}^*) = p_{ij}\omega_{ij}^*, \quad \omega_{ij}^* \in (0, 1],$$

com o qual verifica-se $G(\boldsymbol{\omega}_0) = \mathbf{I}_q$, com $a = 1$.

Seja o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos sob o esquema de perturbação multiplicativa apropriada dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left(\sum_{j=1}^{n_i} \left\{ y_{ij} \log \left(\frac{p_{ij}(\omega_{ij}^*)}{1 - p_{ij}(\omega_{ij}^*)} \right) - \log \left(\frac{1}{1 - p_{ij}(\omega_{ij}^*)} \right) \right\} - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right).$$

Então, as derivadas primeiras e segundas relacionadas à matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_{0pma}}$ são dadas por

$$\frac{\partial L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \omega_{ij}} = \frac{g_{ij}(\boldsymbol{\omega}_0)^{-1/2} y_{ij} (1 + \exp(\eta_{ij}))}{\omega_{ij}^* (1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij}^*)} - \frac{g_{ij}(\boldsymbol{\omega}_0)^{-1/2} \exp(\eta_{ij})}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij}^*)},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_{ij}} = \frac{g_{ij}(\boldsymbol{\omega}_0)^{-1/2} (y_{ij} - 1) \exp(\eta_{ij}) \mathbf{x}_{ij}}{(1 + \exp(\eta_{ij}) - \exp(\eta_{ij})\omega_{ij}^*)^2},$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}} = \mathbf{0},$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

A.2 Derivadas relacionadas ao Capítulo 4

Seja o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos dado por

$$L(\boldsymbol{\psi}|\mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} (y_{ij} \log(\lambda_{ij}) - \lambda_{ij} - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right], \quad (\text{A.3})$$

com

$$\lambda_{ij} = \exp(\eta_{ij}) \quad e \quad \eta_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i.$$

Assim, as derivadas primeiras e segundas relacionadas a $-\ddot{Q}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}})$ são dadas por

$$\frac{\partial L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^I \sum_{j=1}^{n_i} [y_{ij} - \exp(\eta_{ij})] \mathbf{x}_{ij(p_1 \times 1)},$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^I \sum_{j=1}^{n_i} \exp(\eta_{ij}) \mathbf{x}_{ij} \mathbf{x}'_{ij(p_1 \times p_1)},$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = \mathbf{0}_{(p_1 \times p_4)},$$

$$\frac{\partial L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma}} = -\frac{I}{2} \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}'_i \boldsymbol{\Sigma}^{-1}_{(p_2 \times p_2)},$$

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} = \mathbf{0}_{(p_4 \times p_1)}$$

e

$$-\frac{\partial^2 L(\boldsymbol{\psi}|\mathbf{Y}_c)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\frac{I}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^I \mathbf{b}_i \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \right) \otimes \boldsymbol{\Sigma}^{-1}_{(p_4 \times p_4)}.$$

Seja o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos sob o esquema de perturbação multiplicativa quadrática dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} (y_{ij} \log(\lambda_{ij}(\omega_{ij})) - \lambda_{ij}(\omega_{ij}) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right], \quad (\text{A.4})$$

com

$$\lambda_{ij}(\omega_{ij}) = \lambda_{ij} \omega_{ij}^2, \quad \omega_{ij}^2 > 0.$$

Então, as derivadas primeiras e segundas relacionadas a $\Delta_{\omega_{0_{pmq}}}$ são dadas por

$$\frac{\partial L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \omega_{ij}} = 2 \left[\frac{y_{ij}}{\omega_{ij}} - \exp(\eta_{ij}) \right] \omega_{ij},$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \beta \partial \omega_{ij}} = -2 \exp(\eta_{ij}) \mathbf{x}_{ij} \omega_{ij},$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma \partial \omega_{ij}} = \mathbf{0},$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Para a definição do esquema de perturbação multiplicativa quadrática apropriada considere a expressão dada em (A.4). Logo, a função de densidade associada é dada por

$$p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega}) = \prod_{i=1}^I \left[\prod_{j=1}^{n_i} \exp(y_{ij} \log(\lambda_{ij}(\omega_{ij})) - \lambda_{ij}(\omega_{ij}) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right].$$

Então, as derivadas primeiras e segundas relacionadas a $G(\boldsymbol{\omega})$ são dadas por

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_{ij}} = 2 \left[\frac{y_{ij}}{\omega_{ij}} - \exp(\eta_{ij}) \omega_{ij} \right],$$

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 2 \left[\frac{y_{ij}}{\omega_{ij}^2} + \exp(\eta_{ij}) \right],$$

para $i = j$, e

$$\frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_i} \frac{\partial \log p(\mathbf{Y}_c | \boldsymbol{\psi}, \boldsymbol{\omega})}{\partial \omega_j} = 0,$$

para $i \neq j$, com $i, j = 1, \dots, q$.

Como $G(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \neq a\mathbf{I}_q$ escolhe-se um novo vetor de perturbação dado por

$$\omega_{ij}^* = 1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2} (\omega_{ij} - 1),$$

para $i, j = 1, \dots, q$, de tal modo que o esquema de perturbação multiplicativa quadrática apropriada

é dado por

$$\lambda_{ij}(\omega_{ij}^*) = \lambda_{ij}\omega_{ij}^{*2}, \quad \omega_{ij}^{*2} > 0,$$

com o qual obtemos $G(\omega_0) = \mathbf{I}_q$, com $a = 1$.

Seja o logaritmo da função de verossimilhança de $\boldsymbol{\psi}$ dado os dados completos sob o esquema de perturbação multiplicativa apropriada dado por

$$L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^I \left[\sum_{j=1}^{n_i} (y_{ij} \log(\lambda_{ij}(\omega_{ij}^*)) - \lambda_{ij}(\omega_{ij}^*) - \log(y_{ij}!)) - \frac{1}{2} \mathbf{b}'_i \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log(\det(\boldsymbol{\Sigma})) \right].$$

Assim, as derivadas primeiras e segundas relacionadas a $\boldsymbol{\Delta}_{\omega_{0_{pmqa}}}$ são dadas por

$$\frac{\partial L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \omega_{ij}} = 2g_{ij}(\omega_0)^{-1/2} \left[\frac{y_{ij}}{\omega_{ij}^*} - \exp(\eta_{ij}) \omega_{ij}^* \right],$$

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \beta \partial \omega_{ij}} = -2g_{ij}(\omega_0)^{-1/2} \omega_{ij}^* \exp(\eta_{ij}) \mathbf{x}_{ij},$$

$$i, j = 1, \dots, q_{(p_1 \times q)},$$

e

$$\frac{\partial^2 L(\boldsymbol{\psi}, \boldsymbol{\omega} | \mathbf{Y}_c)}{\partial \gamma \partial \omega_{ij}} = \mathbf{0},$$

$$i, j = 1, \dots, q_{(p_4 \times q)}.$$

Apêndice B

Gráficos e tabelas

Neste apêndice apresentamos os gráficos e tabelas relacionados às aplicações dos Capítulos 3 e 4.

B.1 Gráficos e tabelas relacionados às aplicações do Capítulo 3

B.1.1 Aplicação 1

Terrário	Método	
	AL	AQGH
1	-0,651	-0,724
2	-0,219	-0,270
3	-0,268	-0,283
4	-0,086	-0,055
5	0,839	0,923
6	0,502	0,600
7	0,195	0,233
8	-0,051	-0,024
9	-1,078	-1,189
10	0,297	0,354
11	0,278	0,310
12	-0,770	-0,861
13	0,699	0,796
14	0,332	0,360

Tabela B.1: *Predições do intercepto aleatório para os terrários baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.31).*

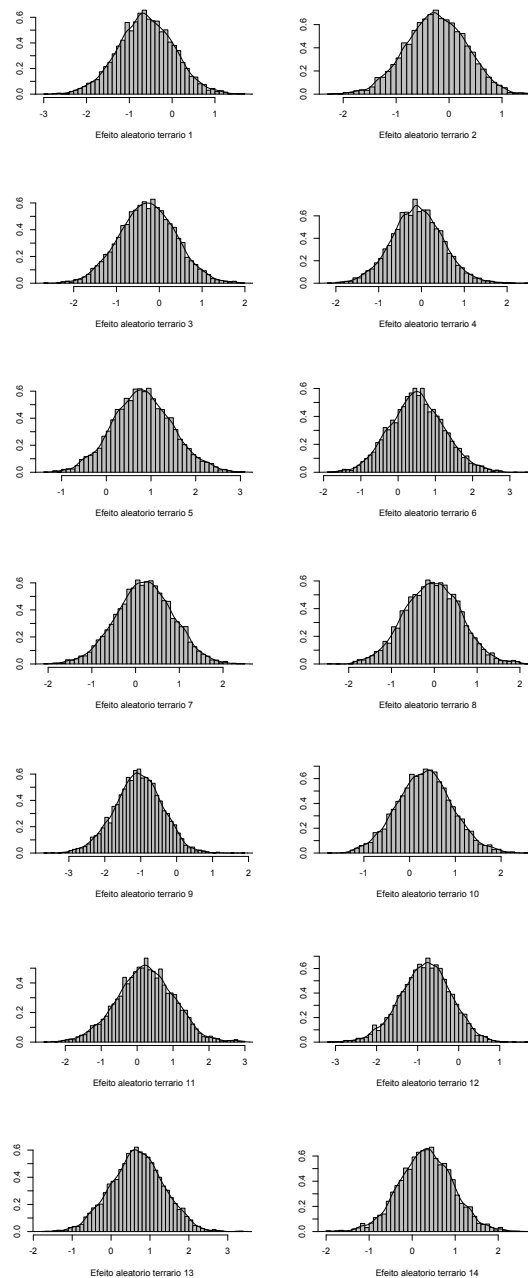


Figura B.1: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os terrários com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (3.31).

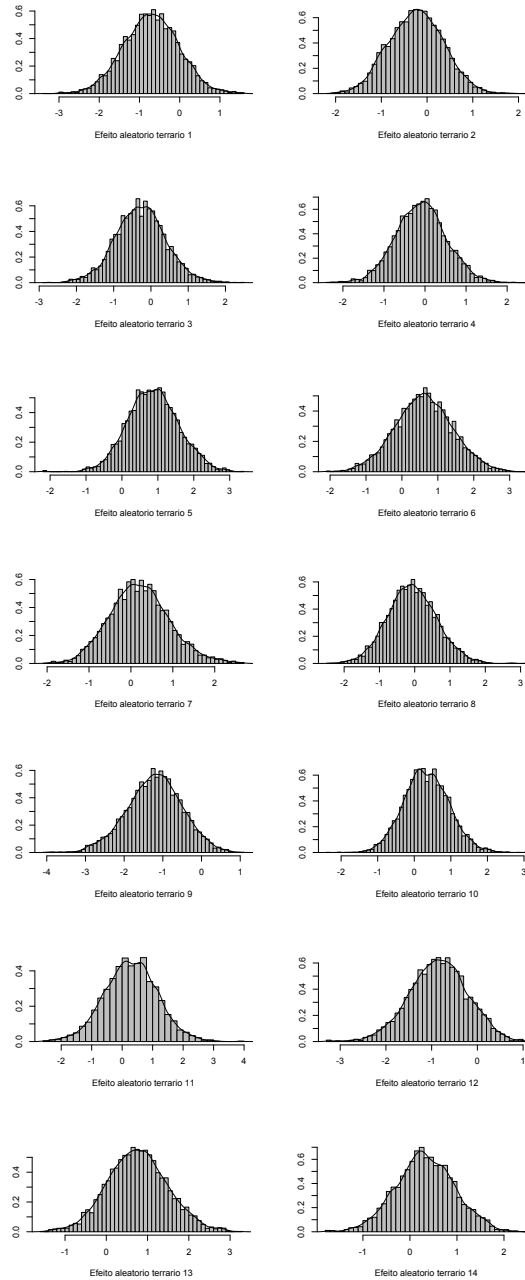


Figura B.2: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os terrários com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (3.31).

B.1.2 Aplicação 2

Bandeja	Método	
	AL	AQGH
1	0,860	0,848
2	1,974	1,976
3	1,956	2,013
4	2,231	2,261
5	1,451	1,463
6	0,426	0,426
7	4,539	4,603
8	4,712	4,779
9	2,760	2,785
10	3,973	4,001
11	0,848	0,879
12	0,695	0,694
13	0,861	0,865
14	0,699	0,682
15	1,987	2,002
16	-0,002	-0,004
17	0,252	0,261
18	1,898	1,904
19	1,297	1,287
20	2,735	2,780
21	0,266	0,245

Tabela B.2: *Predições do intercepto aleatório para as bandejas baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.32).*

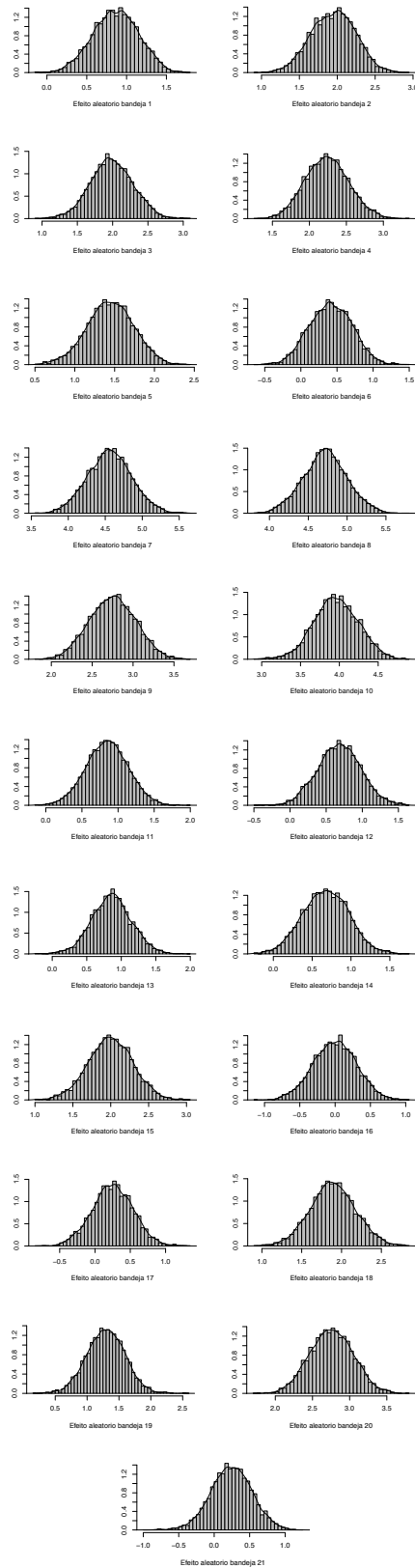


Figura B.3: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para as bandejas com $S = 10.000$ e $M_0 = 1.000$ pelo método AL relacionados ao modelo (3.32).

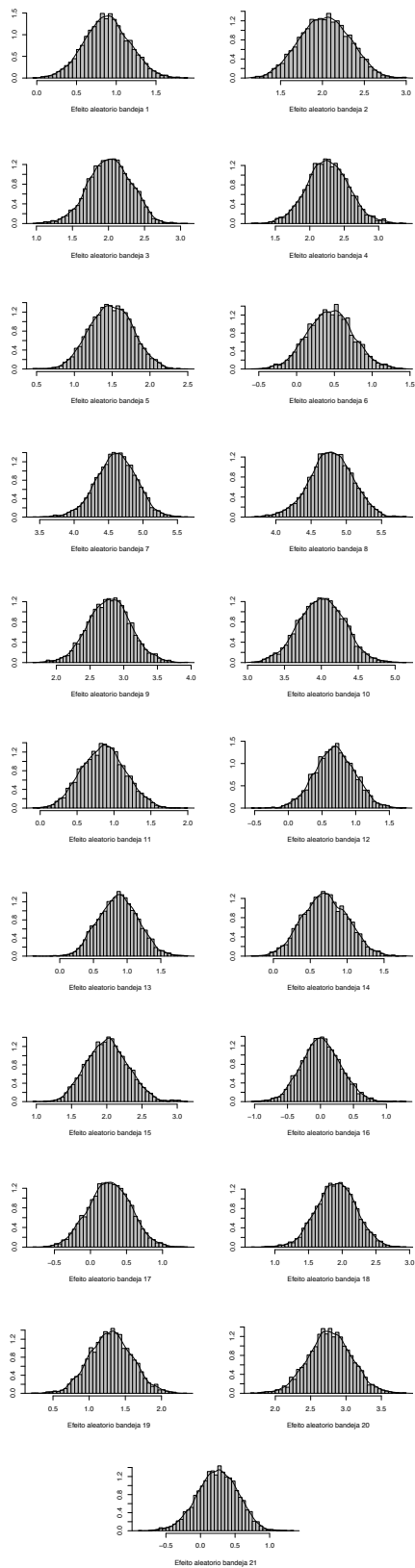


Figura B.4: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para as bandejas com $S = 10.000$ e $M_0 = 1.000$ pelo método AQGH relacionados ao modelo (3.32).

B.1.3 Aplicação 3

Grupo	Método	
	AL	AQGH
1	0,377	0,372
2	0,513	0,517
3	0,371	0,361
4	0,726	0,753

Tabela B.3: Predições do intercepto aleatório para os grupos baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (3.33).

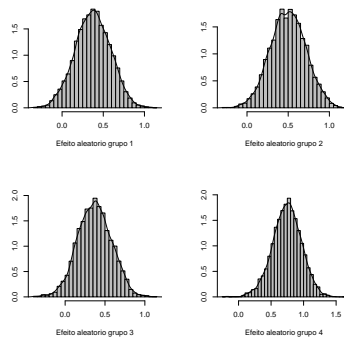


Figura B.5: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os grupos com $S = 10.000$ e $M_0 = 1.000$ pelo método AL relacionados ao modelo (3.33).

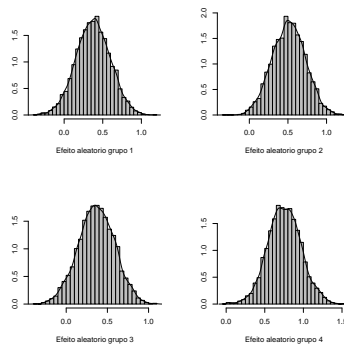


Figura B.6: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os grupos com $S = 10.000$ e $M_0 = 1.000$ pelo método AQGH relacionados ao modelo (3.33).

B.2 Gráficos e tabelas relacionados às aplicações do Capítulo 4

B.2.1 Aplicação 1

Paciente	Método	
	AL	AQGH
1	0,021	0,035
2	0,049	0,038
3	0,306	0,330
4	0,137	0,129
5	0,011	0,012
6	-0,228	-0,220
7	-0,149	-0,142
8	0,426	0,425
9	-0,199	-0,188
10	0,930	0,928
11	0,158	0,162
12	-0,019	-0,028
13	-0,095	-0,080
14	-0,090	-0,090
15	-0,227	-0,233
16	-0,884	-0,874
17	-0,787	-0,758
18	0,191	0,195
19	-0,261	-0,264
20	-0,137	-0,131
21	-0,006	0,003
22	0,315	0,314
23	-0,356	-0,331
24	0,084	0,078
25	0,960	0,960
26	-0,489	-0,493
27	0,019	-0,001
28	0,252	0,251
29	-0,321	-0,323
30	-0,163	-0,157
31	-0,364	-0,358
32	0,522	0,514
33	0,441	0,449
34	-0,335	-0,336
35	1,011	1,009
36	0,527	0,539
37	0,250	0,260
38	-0,697	-0,701
39	-0,061	-0,073
40	-0,014	-0,024
41	-0,602	-0,612
42	0,109	0,105
43	0,388	0,382
44	0,004	0,006
45	0,089	0,092
46	0,347	0,331
47	0,091	0,095
48	-0,401	-0,402
49	0,000	0,000
50	-0,227	-0,228
51	-0,214	-0,210
52	-0,791	-0,809
53	0,430	0,435
54	-0,410	-0,409
55	0,184	0,175
56	1,096	1,094
57	-0,670	-0,676
58	-0,951	-0,958
59	0,056	0,071

Tabela B.4: *Predições do intercepto aleatório para os pacientes baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.32).*

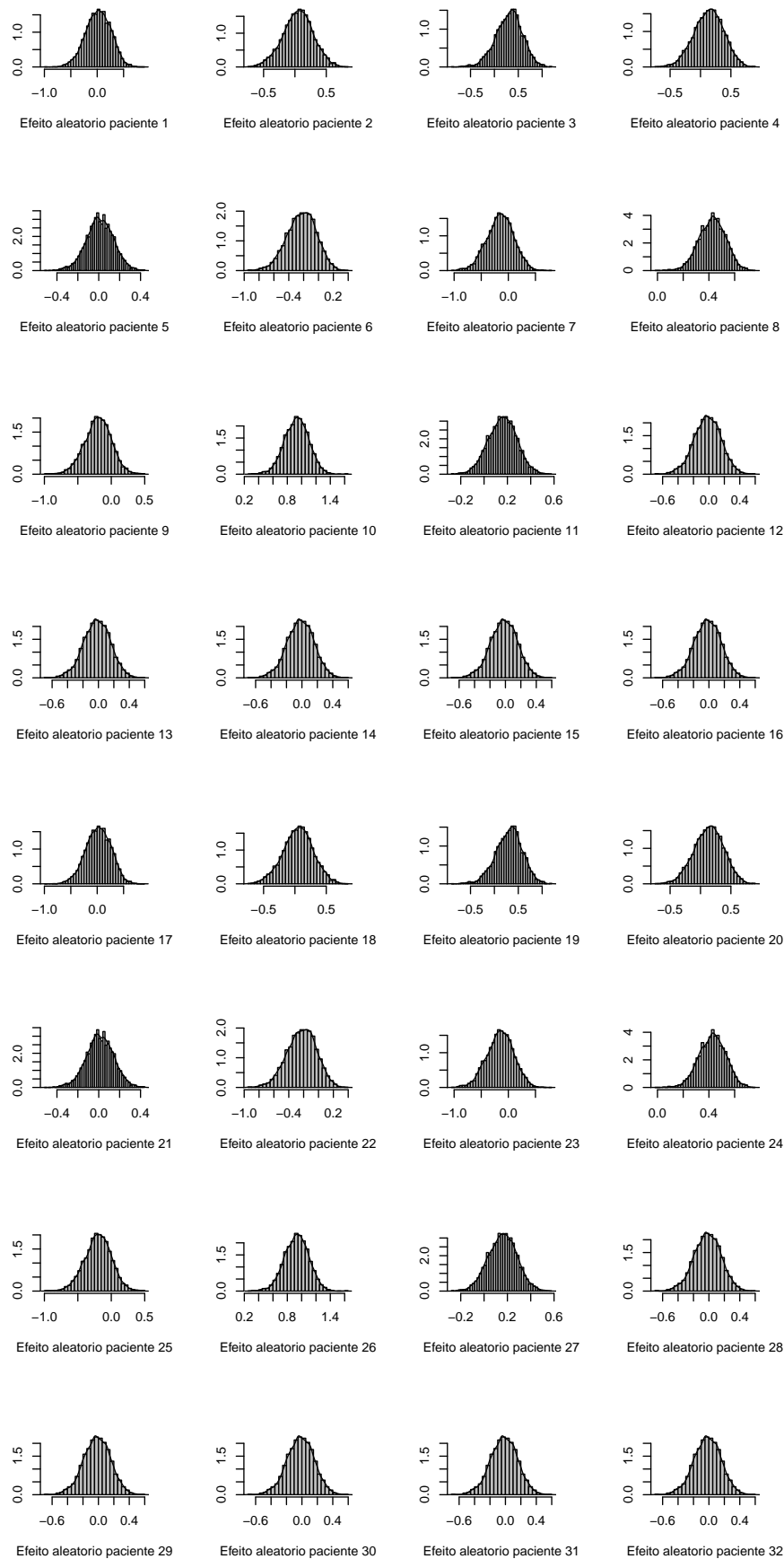


Figura B.7: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.32).

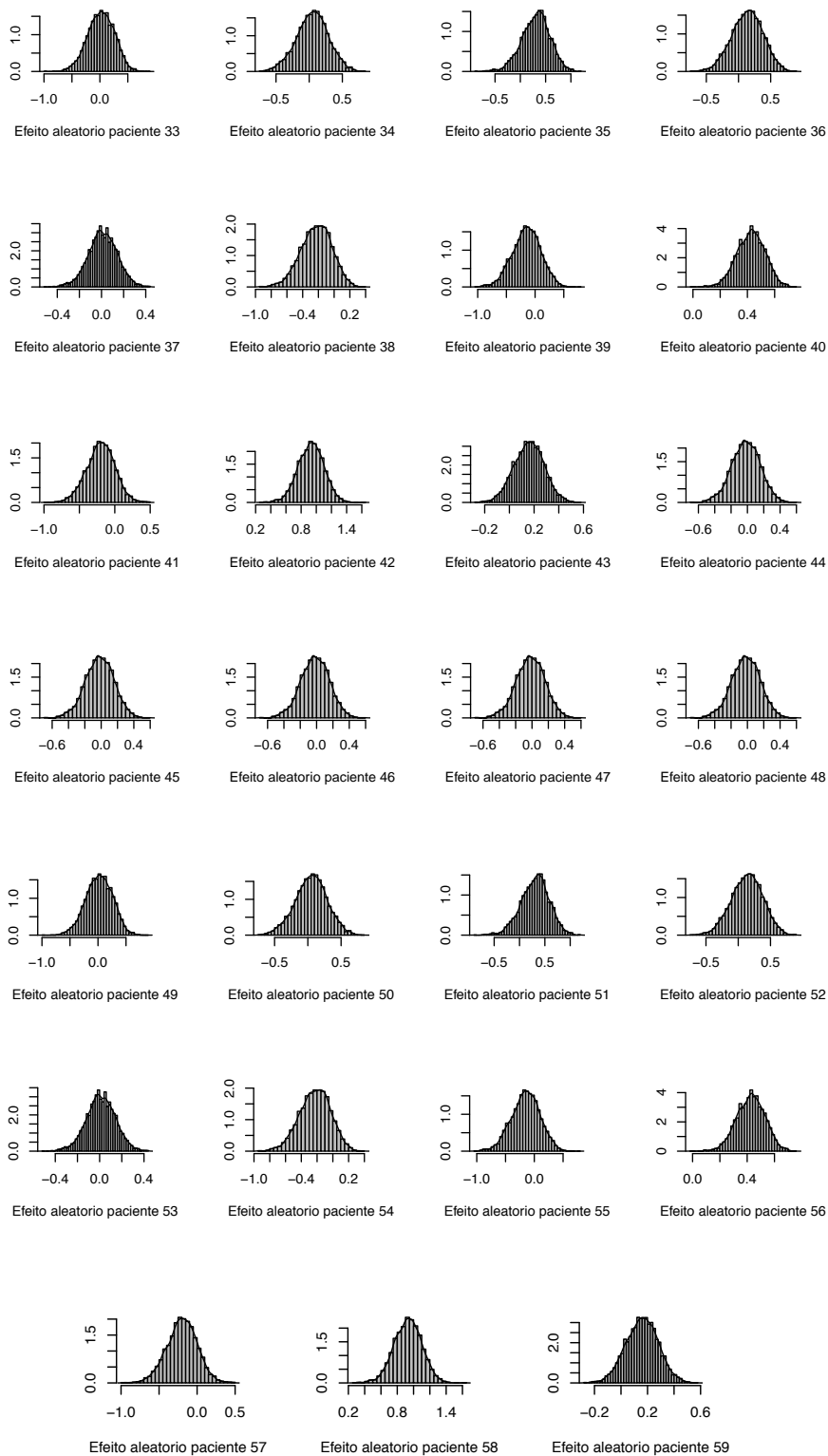


Figura B.7: (Continuação)

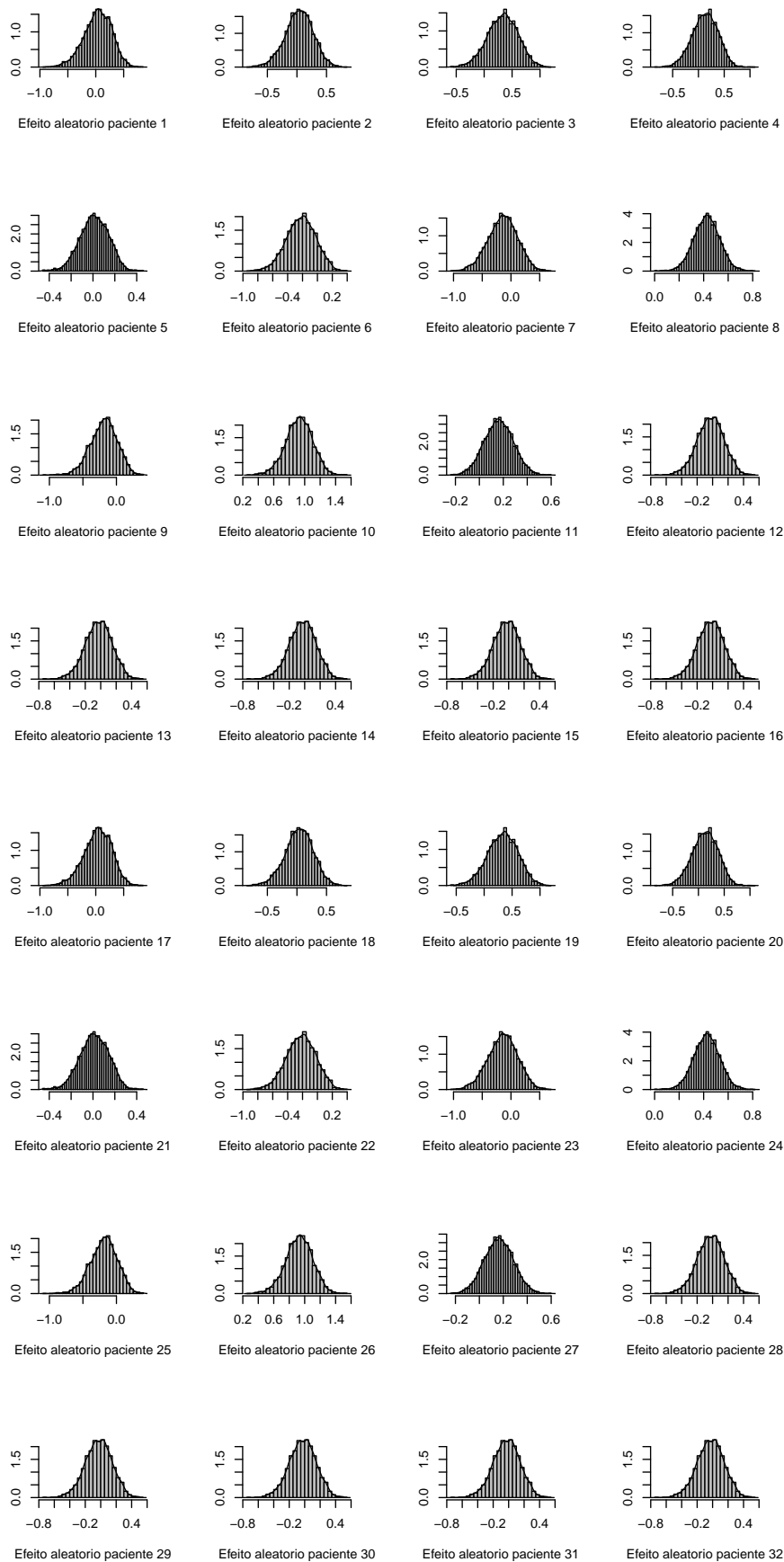


Figura B.8: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.32).

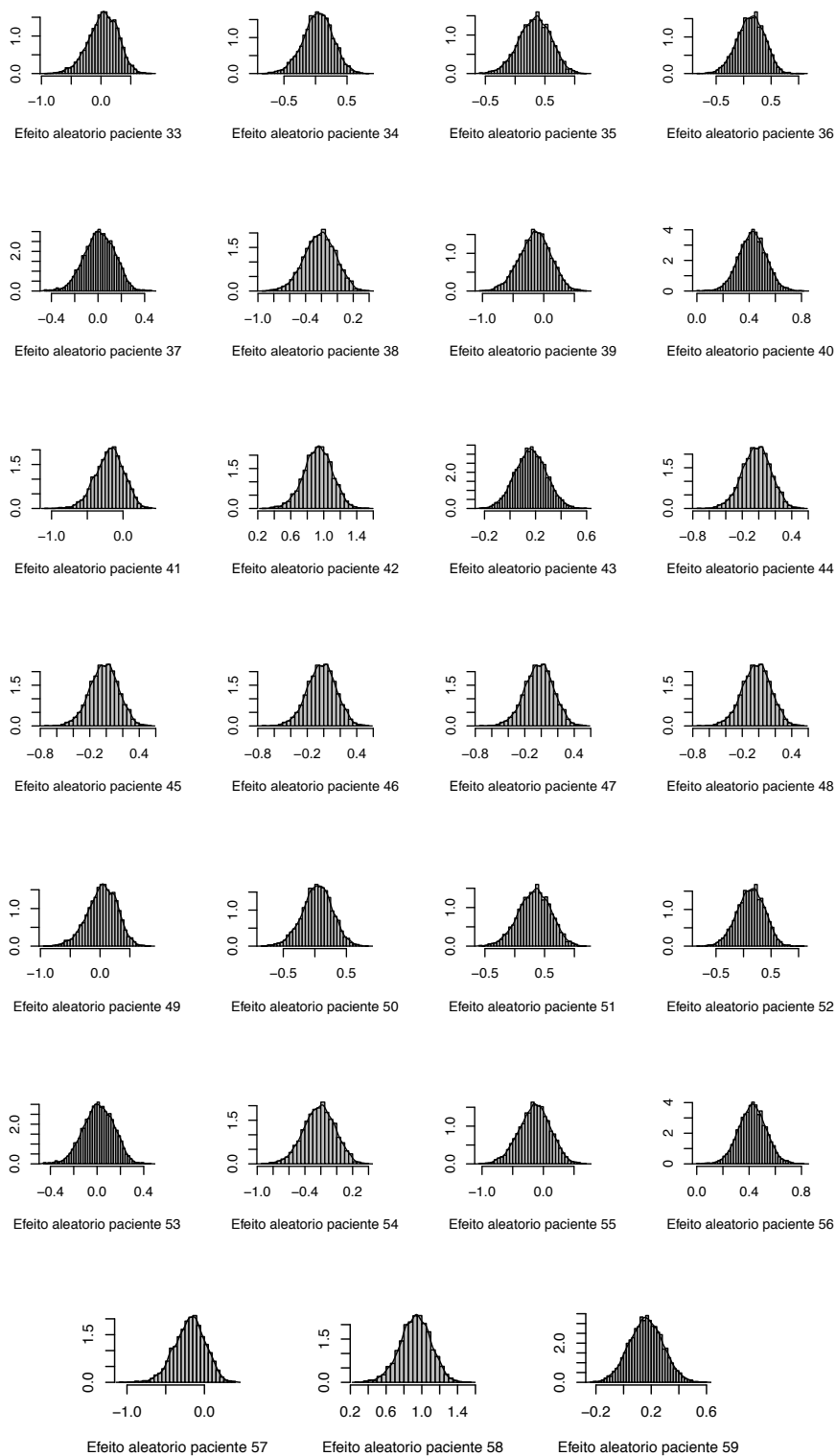


Figura B.8: (Continuação)

B.2.2 Aplicação 2

Paciente	Método	
	AL	AQGH
1	-0,031	-0,024
2	0,345	0,345
3	0,331	0,342
4	-1,170	-1,132
5	-0,277	-0,287
6	-0,141	-0,168
7	0,325	0,346
8	-0,347	-0,363
9	-0,090	-0,067
10	-1,136	-1,187
11	-0,229	-0,213
12	-0,295	-0,305
13	1,455	1,464
14	-0,039	-0,050
15	-0,023	-0,053
16	0,098	0,083
17	0,622	0,612
18	-0,570	-0,569
19	-0,576	-0,574
20	0,640	0,643
21	0,637	0,647
22	-0,432	-0,445
23	0,751	0,764
24	-0,421	-0,410
25	0,627	0,621
26	-0,393	-0,439
27	0,396	0,384

Tabela B.5: *Predições do intercepto aleatório para os pacientes baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.33).*

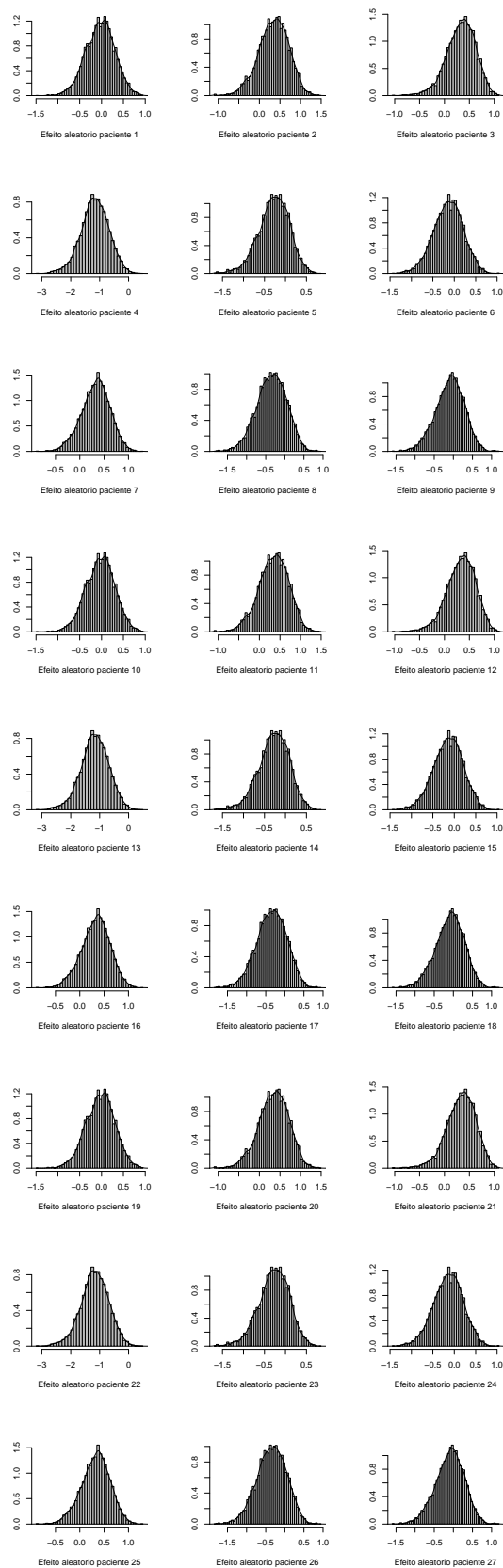


Figura B.9: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.33).

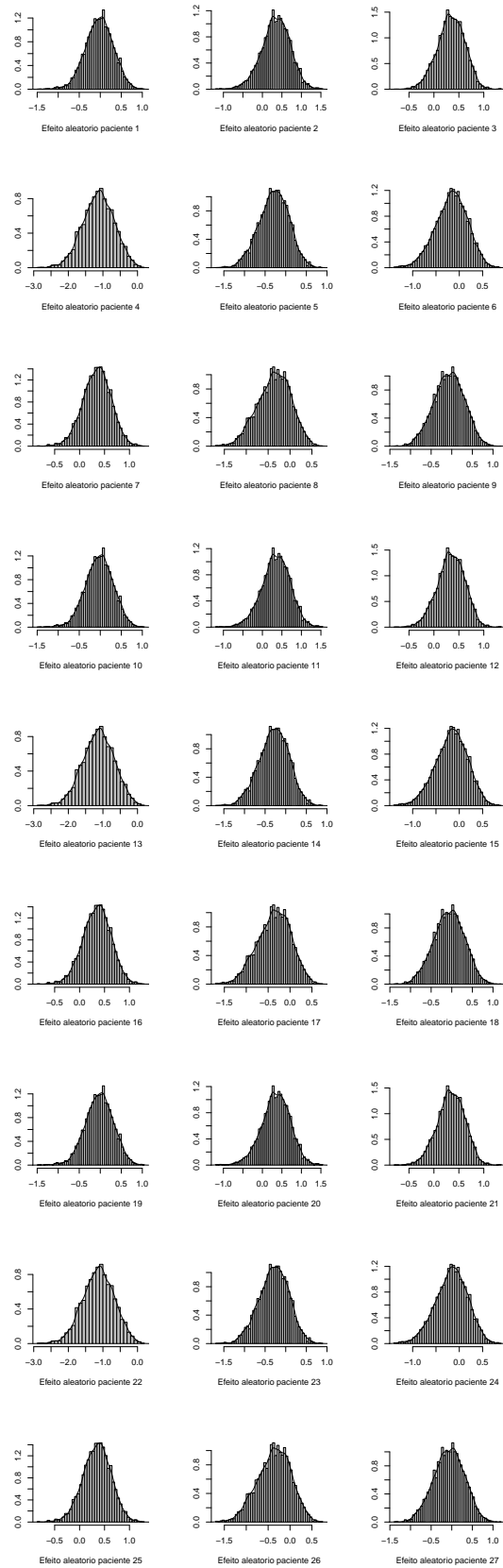


Figura B.10: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os pacientes com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.33).

B.2.3 Aplicação 3

Indivíduo	Método	
	AL	AQGH
1	0,323	0,330
2	-0,022	-0,009
3	0,166	0,169
4	-0,060	-0,051
5	0,111	0,112
6	0,299	0,298
7	0,209	0,201
8	0,395	0,392
9	0,041	0,062
10	0,042	0,047
11	-0,600	-0,594
12	-0,398	-0,406
13	0,115	0,109
14	-0,377	-0,356
15	-0,158	-0,159
16	-0,018	-0,013
17	-0,044	-0,046
18	-0,343	-0,351
19	-0,304	-0,290
20	0,266	0,266
21	-0,342	-0,339
22	0,493	0,490
23	0,230	0,232
24	0,050	0,050
25	0,029	0,022
26	0,137	0,143
27	-0,254	-0,254

Tabela B.6: *Predições do intercepto aleatório para os indivíduos baseadas na amostragem pelos métodos AL e AQGH relacionados ao modelo (4.34).*

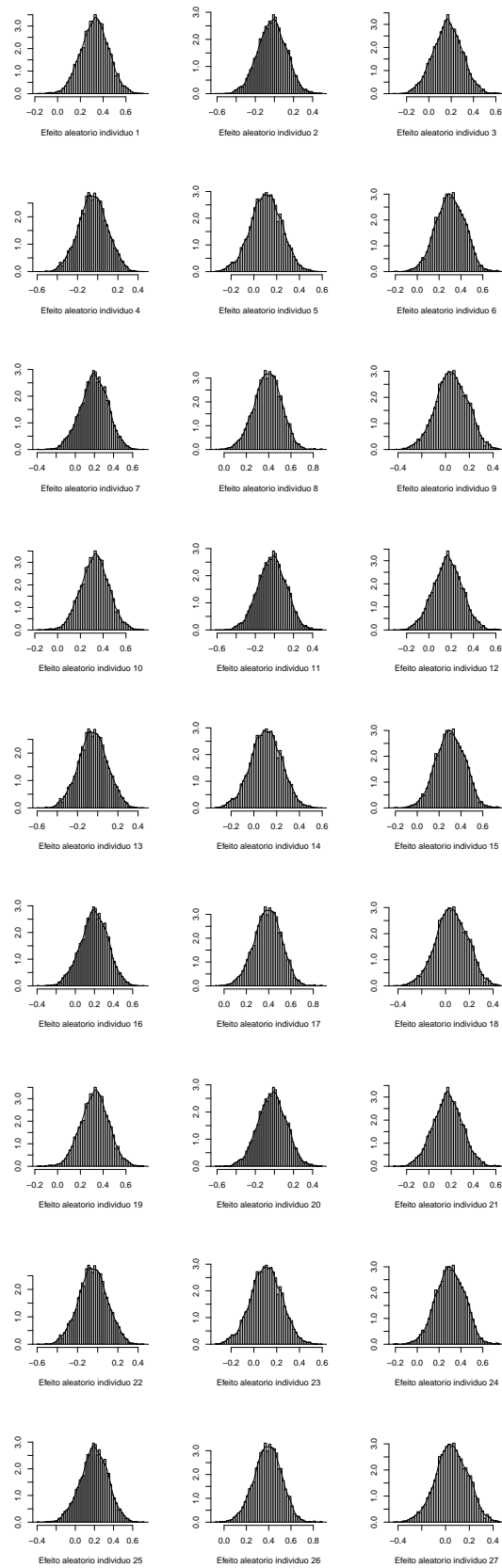


Figura B.11: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os indivíduos com $S=10.000$ e $M_0=1.000$ pelo método AL relacionados ao modelo (4.34).

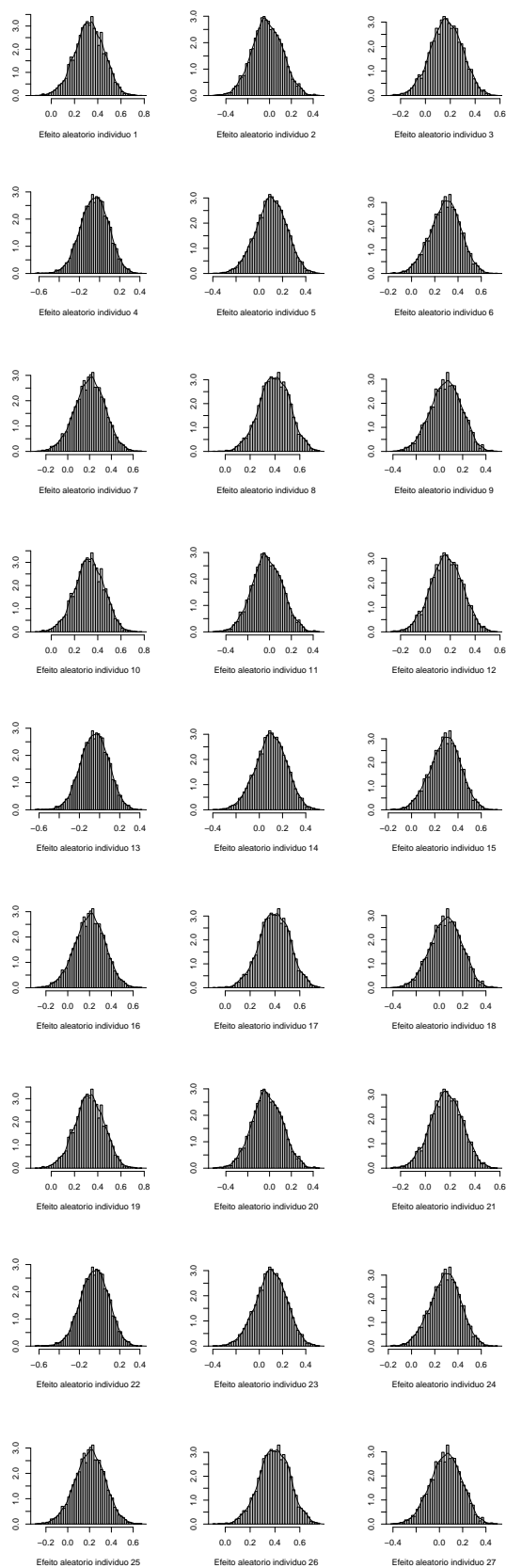


Figura B.12: Histogramas e funções de densidades estimadas da amostragem do intercepto aleatório para os indivíduos com $S=10.000$ e $M_0=1.000$ pelo método AQGH relacionados ao modelo (4.34).

Referências bibliográficas

- Abramowitz e Stegun (1972)** M. Abramowitz e I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. New York: Dover. Citado na pág. [10](#)
- Atkinson (1981)** A. C. Atkinson. Two graphical display for outlying and influential observations in regression. *Biometrika*, 68(1):13–20. Citado na pág. [11](#)
- Atkinson (1982)** A. C. Atkinson. Regression diagnostics, transformations and constructed variables. *Journal of the Royal Statistical Society, B*, 44(1):1–36. Citado na pág. [11](#)
- Belsley et al. (1980)** D. A. Belsley, E. Kuh e R. E. Welsch. *Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity*. New York: John Wiley and Sons. Citado na pág. [11](#)
- Box e Cox (1964)** G. E. P. Box e D. R. Cox. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, B*, 26(2):211–252. Citado na pág. [5](#)
- Cadigan e Farrell (2002)** N. G. Cadigan e P. J. Farrell. Generalized local influence with applications to fish stock cohort analysis. *Journal of the Royal Statistical Society, C*, 51(4):469–483. Citado na pág. [137](#)
- Chatterjee e Hadi (1986)** S. Chatterjee e A. L. Hadi. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 1(3):379–416. Citado na pág. [11](#)
- Chen et al. (2010)** F. Chen, H. T. Zhu, X. Y. Song e S. Y. Lee. Perturbation selection and local influence analysis for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 19(4):826–842. Citado na pág. [2](#), [3](#), [23](#), [24](#), [35](#), [36](#), [135](#)
- Cook (1977)** R. D. Cook. Detection of influential observations in linear regression. *Technometrics*, 19(1):15–18. Citado na pág. [12](#)
- Cook (1986)** R. D. Cook. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, B*, 48(2):133–169. Citado na pág. [1](#), [2](#), [3](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- Cook e Weisberg (1982)** R. D. Cook e S. Weisberg. *Residuals and Influence in Regression*. London - New York: Chapman and Hall. Citado na pág. [11](#)
- Cordeiro (2004)** G. M. Cordeiro. On Pearson’s residuals in generalized linear models. *Statistics and Probability Letters*, 66(3):213–219. Citado na pág. [11](#)
- Crowder (1978)** M. J. Crowder. Beta-binomial ANOVA for proportions. *Journal of the Royal Statistical Society, C*, 27(1):34–37. Citado na pág. [3](#), [55](#), [64](#)
- Davis e Rabinowitz (1975)** P. J. Davis e P. Rabinowitz. *Methods of Numerical Integration*. New York: Academic Press. Citado na pág. [10](#)
- Demidenko (2013)** E. Demidenko. *Mixed Models: Theory and Applications with R*. Hoboken: John Wiley and Sons. Citado na pág. [7](#), [9](#)

- Diggle et al. (1996)** P. J. Diggle, K. Y. Liang e S. L. Zeger. *Analysis of Longitudinal Data*. New York: Oxford University Press. Citado na pág. 2
- Fabio et al. (2012)** L. C. Fabio, G. A. Paula e M. de Castro. A Poisson mixed model with nonnormal random effect distribution. *Computational Statistics and Data Analysis*, 56(6):1499–1510. Citado na pág. 137
- Fawcett (2006)** T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8): 861–874. Citado na pág. 27
- Ferreccio et al. (2011)** C. Ferreccio, V. Van De Wyngard, F. Olcay, M. A. Domínguez, K. Puschel, A. H. Corvalán, S. Franceschi e P. JF. Snijders. High-risk HPV infection after five years in a population-based cohort of Chilean women. *Infectious Agents and Cancer*, 6(21). Citado na pág. 3, 55, 74
- Giampaoli et al. (2014)** V. Giampaoli, E. Y. Mitsuuchi e A. B. A. Sato. Relatório de análise estatística sobre o projeto: Ciclo vigília/sono em pacientes de diabetes mellitus tipo I. Relatório técnico, Centro de Estatística Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 3, 108, 125
- Gibbons et al. (2010)** R. Gibbons, D. Hedeker e S. Dutoit. Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology*, 6:79–107. Citado na pág. 2
- Goldstein et al. (2002)** H. Goldstein, W. Browne e J. Rasbash. Multilevel modelling of medical data. *Statistics in Medicine*, 21(21):3291–3315. Citado na pág. 2
- Hoaglin e Welsch (1978)** D. C. Hoaglin e R. E. Welsch. The hat matrix in regression and ANOVA. *The American Statistician*, 32(1):17–22. Citado na pág. 12
- Hosmer et al. (2013)** D. W. Hosmer, S. Lemeshow e R. X. Sturdivant. *Applied Logistic Regression*. Hoboken: John Wiley and Sons. Citado na pág. 27
- Ibacache-Pulgar et al. (2012)** G. Ibacache-Pulgar, G. A. Paula e M. Galea. Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, 12(2):165–193. Citado na pág. 1
- Jiang (2007)** J. Jiang. *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer. Citado na pág. 7
- Laird e Ware (1982)** N. M. Laird e J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974. Citado na pág. 2
- Langford e Lewis (1998)** I.H. Langford e T. Lewis. Outliers in multilevel data. *Journal of the Royal Statistical Society, A*, 161(2):121–160. Citado na pág. 11
- Larsen et al. (2000)** K. Larsen, J. H. Petersen, E. Budtz-Jorgensen e L. Endahl. Interpreting parameters in the logistic regression model with random effects. *Biometrics*, 56(3):909–914. Citado na pág. 137
- Lee et al. (2006)** Y. Lee, J. A. Nelder e Y. Pawitan. *Generalized Linear Models with Random Effects*. Boca Raton: Chapman and Hall/CRC. Citado na pág. 7
- Leiva et al. (2007)** V. Leiva, M. Barros, G. Paula e M. Galea. Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. *Computational Statistics and Data Analysis*, 51(12):5694–5707. Citado na pág. 12
- McCullagh e Nelder (1989)** P. McCullagh e J. A. Nelder. *Generalized Linear Models*. London: Chapman and Hall. Citado na pág. 5, 7

- McCulloch e Searle (2001)** C. E. McCulloch e S. R. Searle. *Generalized, Linear and Mixed Models*. New York: John Wiley and Sons. Citado na pág. 7, 56
- McKnight e Eeden (1993)** B. McKnight e S. K. Van Den Eeden. A conditional analysis for two-treatment multiple-period crossover designs with binomial or poisson outcomes and subjects who drop out. *Statistics in Medicine*, 12(9):825–834. Citado na pág. 3, 108, 117
- Meza et al. (2009)** C. Meza, F. Jaffrézic e J. L. Foulley. Estimation in the probit normal model for binary outcomes using the SAEM algorithm. *Computational Statistics and Data Analysis*, 53(4):1350–1360. Citado na pág. 136
- Naylor e Smith (1982)** J. C. Naylor e A. F. M. Smith. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society, C*, 31(3):214–225. Citado na pág. 10
- Nelder e Wedderburn (1972)** J. A. Nelder e R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society, A*, 135(3):370–384. Citado na pág. 5
- Osorio et al. (2009)** F. Osorio, G. A. Paula e M. Galea. On estimation and influence diagnostics for the Grubbs' model under heavy-tailed distributions. *Computational Statistics and Data Analysis*, 53(4):1249–1263. Citado na pág. 1
- Ouwens et al. (2001)** M. J. N. M. Ouwens, F. E. S. Tan e M. P. F. Berger. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, 57(4):1166–1172. Citado na pág. 12
- Paula (2013)** G. A. Paula. *Modelos de Regressão com Apoio Computacional*. São Paulo: IME-USP. Citado na pág. 5, 7
- Paula (1995)** G. A. Paula. Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation*, 51(2-4):315–331. Citado na pág. 11
- Paula (1999)** G. A. Paula. Leverage in inequality-constrained regression models. *The Statistician*, 48(4):529–538. Citado na pág. 12
- Pierce e Schafer (1986)** D. A. Pierce e D. W. Schafer. Residuals in generalized linear models. *Journal of the American Statistical Association*, 81(396):977–986. Citado na pág. 11
- Pinheiro e Bates (1995)** J. C. Pinheiro e D. M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35. Citado na pág. 10
- Pinheiro et al. (2001)** J. C. Pinheiro, Ch. Liu e Y. N. Wu. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276. Citado na pág. 137
- Pregibon (1981)** D. Pregibon. Logistic regression diagnostics. *The Annals of Statistics*, 9(4):705–724. Citado na pág. 11
- Press et al. (1992)** W. H. Press, S. A. Teukolsky, W. T. Vetterling e B. P. Flannery. *Numerical Recipes in FORTRAN*. New York: Cambridge University Press. Citado na pág. 10
- R Core Team (2014)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. Citado na pág. 3, 83
- Robert e Casella (2004)** C. P. Robert e G. Casella. *Monte Carlo Statistical Methods*. New York: Springer. Citado na pág. 22, 24

- Russo et al. (2009)** C. Russo, G. Paula e R. Aoki. Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis*, 53(12):4143–4156. Citado na pág. 12
- Sen et al. (2010)** P. Sen, J. Singer e A. C. Pedroso de Lima. *From Finite Sample to Asymptotic Methods in Statistics*. New York: Cambridge University Press. Citado na pág. 6, 23
- Stiratelli et al. (1984)** R. Stiratelli, N. Laird e J. Ware. Random effects models for serial observations with binary responses. *Biometrics*, 40(4):961–971. Citado na pág. 2
- Stoker (1956)** J. J. Stoker. *Differential Geometry*. New York: John Wiley and Sons. Citado na pág. 12
- Thall e Vail (1990)** P. F. Thall e S. C. Vail. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46(3):657–671. Citado na pág. 3, 108
- Thomas e Cook (1989)** W. Thomas e R. D. Cook. Assessing influence on regression coefficients in generalized linear models. *Biometrika*, 76(4):741–749. Citado na pág. 12
- Thomas e Cook (1990)** W. Thomas e R. D. Cook. Assessing influence on predictions from generalized linear models. *Technometrics*, 32(1):59–65. Citado na pág. 12, 137
- Verbeke e Molenberghs (2000)** G. Verbeke e G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag. Citado na pág. 7
- Villegas et al. (2013)** C. Villegas, G. A. Paula, F. J. Cysneiros e M. Galea. Influence diagnostics in generalized symmetric linear models. *Computational Statistics and Data Analysis*, 59:161–170. Citado na pág. 1
- Werneck (2012)** R. Werneck. Efeito da disponibilidade de sítios de nidificação sobre o sistema de acasalamento e o cuidado paternal em um opilião neotropical (Arachnida: Opiliones). Dissertação de Mestrado, Instituto de Biociências da Universidade de São Paulo, Departamento de Ecologia. Citado na pág. 3, 55
- Williams (1987)** D. A. Williams. Generalized linear model diagnostic using the deviance and single case deletion. *Journal of the Royal Statistical Society, C*, 36(2):181–191. Citado na pág. 11, 12
- Xiang et al. (2002)** L. Xiang, S. K. Tse e A. H. Leeb. Influence diagnostics for generalized linear mixed models: applications to clustered data. *Computational Statistics and Data Analysis*, 40(4):759–774. Citado na pág. 12
- Xiang et al. (2003)** L. Xiang, A. H. Lee e S. K. Tse. Assessing local cluster influence in generalized linear mixed models. *Journal of Applied Statistics*, 30(4):349–359. Citado na pág. 12
- Xu et al. (2006)** L. Xu, S. Y. Lee e W. Y. Poon. Deletion measures for generalized linear mixed effects models. *Computational Statistics and Data Analysis*, 51(2):1131–1146. Citado na pág. 12, 22, 35
- Zhu e Lee (2001)** H. T. Zhu e S. Y. Lee. Local influence for incomplete-data models. *Journal of the Royal Statistical Society, B*, 63(1):111–126. Citado na pág. 15
- Zhu e Lee (2003)** H. T. Zhu e S. Y. Lee. Local influence for generalized linear mixed models. *The Canadian Journal of Statistics*, 31(3):293–309. Citado na pág. 2, 3, 11, 12, 15, 16, 17, 18, 21, 22, 23, 86, 92