Regressão beta

Patrícia Leone Espinheira Ospina

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

Área de Concentração: Estatística
Orientadora: Prof^a. Dr^a. Silvia Lopes de Paula Ferrari
Co-orientador: Prof. Dr. Francisco Cribari–Neto

Durante a elaboração deste trabalho a autora recebeu auxílio financeiro da FAPESP

São Paulo, 29 de março de 2007

Regressão beta

Este exemplar corresponde à redação final da
tese devidamente corrigida e defendida
por Patrícia Leone Espinheira Ospina
e aprovada pela Comissão Julgadora

São Paulo, 29 de março de 2007

Banca Examinadora:

Prof^a. Dr^a. Silvia Lopes de Paula Ferrari (orientadora)-IME/USP

Prof^a. Dr^a. Clarice Garcia Borges Demétrio-ESALQ/USP

Prof. Dr. Gilberto Alvarenga Paula-IME/USP

Prof. Dr. Klaus Leite Pinho Vasconcellos-UFPE

Prof^a. Dr^a. Reiko Aoki-ICMC/USP

A minha avó, Enedina (in memorian)

dedico com todo o meu amor.

Agradecimentos

As forças divinas que regem meu caminho me protegendo, amparando e orientando.

À professora Silvia Ferrari pela orientação constante e cuidadosa na elaboração desta tese, por todo aprendizado pessoal e profissional e pela convivência enriquecedora dos últimos quatro anos.

Ao professor Francisco Cribari Neto por ter me conduzido até aqui, ter confiado em meu potencial, pela dedicação na minha formação acadêmica e por todo o apoio dos últimos seis anos.

Ao meu marido Raydonal Ospina por tanto amor, **paciência**, dedicação, por ser essa pessoa tão linda e especial e por me fazer tão feliz.

A Sarah que chegará para iluminar minha vida.

Aos meus pais Raildes e Antônio, meu irmão Antônio e minha cunhada Aninha, e ao meu tio Roberto pela certeza de que sou amada incondicionalmente.

A toda minha família Leone e Espinheira pelo previlégio de ter uma família unida, solidária, com princípios de ética que regem meu comportamento e são decisivos em meu sucesso profissional e pessoal.

Aos professores do Instituto de Matemática e Estatística da Universidade de São Paulo, que ajudaram na minha formação acadêmica.

A Michelli e a Diana, amigas queridas e companheiras, por termos conseguido vencer juntas, desde o início, todos os desafios desse doutorado e da cidade de São Paulo.

A Tatiane e a Luz Mery, pela amizade sincera, tanto carinho e lealdade.

À minha família em São Paulo, Michelli e Horácio, Diana e Gustavo, Tatiane e Alessandro, Luz Mery, Rejelie e Gilberto, pela nossa união e dedicação uns com os outros.

A todos os amigos queridos que fiz em São Paulo e em Recife, Cristian e Sandra, Gladys e Jucirei, Lilian e família, Viviana Giampaoli e família, Claúdia e Manoel, Maurício, Diane, Arthur, Edjane, Elier, Clodis, Maria Kelly, Iracema, Gisela, Lívia, Lourdes, Márcio, Juvêncio e Jaqueline, Marcos e Daniela, Augusto e toda comunidade colombiana do IME-USP.

À cidade de São Paulo por todas as oportunidades pessoais e profissionais.

Aos participantes da banca examinadora, pelas sugestões.

À FAPESP, pelo apoio financeiro.

Resumo

Muitos estudos em diferentes áreas examinam como um conjunto de variáveis influencia algum tipo de percentagem, proporção ou frações. Modelos de regressão linear não são satisfatórios para modelar tais dados. Uma classe de modelos de regressão beta que em muitos aspectos é semelhante aos modelos lineares generalizados foi proposta por Ferrari e Cribari–Neto (2004). A resposta média é relacionada com um preditor linear por uma função de ligação e o preditor linear envolve covariáveis e parâmetros de regressão desconhecidos. O modelo também é indexado por um parâmetro de precisão. Smithson e Verkuilen (2006), entre outros, consideram o modelo de regressão beta em que esse parâmetro varia ao longo das observações. Nesta tese foram desenvolvidas técnicas de diagnóstico para os modelos de regressão beta com dispersão constante e com dispersão variável, sendo que o método de influência local (Cook, 1986) mostrou-se particularmente útil, sendo capaz inclusive de identificar dispersão variável nos dados. Adicionalmente, avaliamos através de estudos de simulação o desempenho dos estimadores de máxima verossimilhança do modelo de regressão beta com dispersão variável, as conseqüências de estimar o modelo supondo dispersão constante quando de fato ela é variável e o desempenho de testes assintóticos para testar a hipótese de dispersão constante. Finalmente, utilizando o método bootstrap (Davison e Hinkley, 1997) desenvolvemos um procedimento de obtenção de limites de predição para o modelo de regressão beta com dispersão constante. Ilustramos a teoria desenvolvida com várias aplicações a dados reais.

Abstract

Practitioners oftentimes wish to investigate how certain variables influence a continuous variable that assumes values on the standard unit interval $(0, 1)$, such as percentages, proportions, rates and fractions. Linear regression models are not suitable for modelling such data. A class of beta regression models which is in many aspects similar to that of generalised linear models was proposed by Ferrari and Cribari–Neto (2004). The mean response is related to a linear predictor, which involves covariates and unknown regression parameters, through a link function. The model is also indexed by a precision parameter. Smithson and Verkuilen (2006), among others, consider the beta regression model with variable dispersion, i.e., beta regression in which the precision parameter is not constant across observations. In this thesis we develop diagnostic methods for beta regression models with both constant and variable dispersion. The method of local influence (Cook, 1986) proved to be particularly useful, since it is able to identify variable dispersion in the data. We have also used Monte Carlo simulation to evaluate the finite sample performance of maximum likelihood estimators in beta regression models with variable dispersion; we have also evaluated the consequences of misspecifying the model by incorrectly assuming constant dispersion when dispersion is variable and the finite sample behavior of heteroskedasticity tests based on first order asymptotics is studied. Prediction bootstrap intervals (Davison and Hinkley, 1997) for the beta regression model with constant dispersion are also considered. Practical applications that employ real data are presented and discussed.

Índice

1. Introdução . . . 1
   1.1. Introdução . . . 1
   1.2. Organização da tese . . . 7
   1.3. Suporte computacional . . . 8
2. Análise de resíduos . . . 9
   2.1. Introdução . . . 9
   2.2. Resíduo ordinário . . . 10
   2.3. Resíduo ponderado padronizado 1 . . . 11
   2.4. Resíduo ponderado padronizado 2 . . . 12
   2.5. Resíduo componente desvio . . . 12
   2.6. Resultados numéricos . . . 13
   2.7. Aplicações . . . 21
      2.7.1. Aplicação I: dados de gastos com alimentação, Griffiths et al. (1993) . . . 21
      2.7.2. Aplicação II: dados de habilidade de leitura, Smithson e Verkuilen (2006) . . . 21
      2.7.3. Exemplo simulado . . . 26
3. Análise de influência . . . 32
   3.1. Introdução . . . 32
   3.2. Distância de Cook . . . 33
   3.3. Influência local . . . 35
   3.4. Esquemas de perturbação . . . 38
      3.4.1. Ponderação de casos . . . 39
      3.4.2. Perturbação da variável resposta . . . 40
      3.4.3. Perturbação individual de covariáveis . . . 41
      3.4.4. Perturbação no parâmetro de precisão . . . 41
   3.5. Comparação entre influência local e matrizes de alavanca . . . 42
   3.6. Aplicações . . . 43
      3.6.1. Aplicação I: dados de gastos com alimentação, Griffiths et al. (1993) . . . 44
      3.6.2. Aplicação II: dados de habilidade de leitura, Smithson e Verkuilen (2006) . . . 49
      3.6.3. Aplicação III: dados de ansiedade e estresse, Smithson e Verkuilen (2006) . . . 58
      3.6.4. Aplicação IV: dados de veredito, Smithson e Verkuilen (2006) . . . 62
      3.6.5. Aplicação V: dados de cloro disponível, Draper e Smith (1981) . . . 69

3.A. Influência local . . . 74
3.A.1. Introdução . . . 74
3.A.2. Influência local no modelo de regressão beta . . . 77
4. Modelo de regressão beta com dispersão variável . . . 82
4.1. Introdução . . . 82
4.2. Definição e estimação do modelo . . . 83
4.3. Estimação intervalar . . . 89
4.4. Testes de heteroscedasticidade . . . 91
4.5. Exemplos . . . 99
5. Diagnóstico para o modelo de regressão beta com dispersão variável . . . 105
5.1. Introdução . . . 105
5.2. Resíduo ponderado padronizado 2 . . . 105
5.3. Influência local . . . 106
5.3.1. Ponderação de casos . . . 107
5.3.2. Perturbação da variável resposta . . . 107
5.3.3. Perturbação individual de covariáveis . . . 107
5.4. Aplicações . . . 109
5.4.1. Aplicação I: dados de habilidade de leitura, Smithson e Verkuilen (2006) . . . 110
5.4.2. Aplicação II: dados de ansiedade e estresse, Smithson e Verkuilen (2006) . . . 116
5.A. Influência local (modelo de regressão beta com dispersão variável) . . . 120
5.A.1. Introdução . . . 120
5.A.2. Esquemas de perturbação . . . 120
6. Intervalos de predição bootstrap para o modelo de regressão beta . . . 128
6.1. Introdução . . . 128
6.2. O método bootstrap . . . 129
6.3. Intervalos de predição bootstrap . . . 130
6.4. Aplicação: Dados de indigência . . . 132
6.4.1. Escolha do modelo . . . 133
6.4.2. Método bootstrap . . . 136
7. Considerações finais . . . 141
7.1. Conclusões . . . 141
7.2. Trabalhos futuros . . . 142
A Dados . . . 144
A.1. Proporção da renda familiar total gasta com alimentação . . . 144

A.2. Escore em teste de habilidade de leitura 145
A.3. Ansiedade e estresse 146
A.4. Veredito 151
A.5. Cloro disponível 154
⋆ Referências 155

Capítulo 1

Introdução

1.1. Introdução

Muitos estudos em diferentes áreas examinam como um conjunto de variáveis se relaciona com algum tipo de percentagem ou proporção. Por exemplo, Kieschnick e McCullough (2003) estudam como a renda domiciliar, a percentagem de domicílios com crianças, entre outras variáveis, se relacionam com a proporção de casas que têm TV a cabo em cada área do mercado. De fato, problemas envolvendo variáveis respostas fracionárias surgem naturalmente em muitos campos de pesquisa. Alguns exemplos são composições de rochas, concentrações de agentes químicos e proporção da renda gasta com contribuições tributárias.

Entre as categorias em que dados de proporções podem ser colocados destacamos dois tipos, a saber: dados distribuídos continuamente no intervalo $(0, 1)$ e dados distribuídos no intervalo fechado $[0, 1]$ (com probabilidade positiva em ao menos um dos extremos). No entanto, enquanto os dados na primeira categoria podem ser modelados usando uma distribuição contínua, o mesmo não ocorre com os dados na segunda, porque eles seguem uma distribuição mista contínua–discreta. Nosso interesse é modelar variáveis que se encontram na primeira categoria, isto é, que se distribuem no intervalo aberto $(0, 1)$.

Não existe consenso quanto ao tipo de modelagem mais adequada para este tipo de dados. Kieschnick e McCullough (2003) identificam sete tipos de modelos de regressão que têm sido utilizados para modelar a distribuição de dados de proporções no intervalo $(0, 1)$. No entanto, segundo os autores, alguns destes modelos violam duas suposições essenciais. Primeira, por essa variável ser limitada, sua função de esperança condicional deve ser não linear. Segunda, as variâncias dos erros devem ser diferentes ao longo das observações. De fato, a variância condicional deve se aproximar de zero à medida que a média condicional se aproxima dos extremos do intervalo. Kieschnick e McCullough (2003) apresentam uma discussão a respeito dos seis primeiros modelos e propõem um sétimo modelo supondo distribuição beta para a variável resposta. Os principais pontos desta discussão serão apresentados nos próximos parágrafos.

Inicialmente, os sete modelos podem ser divididos em duas categorias. A primeira categoria é a que viola ao menos uma das duas suposições discutidas anteriormente. Nesta primeira categoria encontram-se o modelo normal linear, o modelo logito, o modelo normal

censurado e o modelo normal não linear. Na segunda categoria, aquela que não viola as suposições, estão o modelo de regressão baseado na distribuição beta, o modelo simplex e um modelo de quasi-verossimilhança.

Quanto ao modelo normal linear, os autores argumentam que, além da imposição inconsistente de que a função de esperança condicional é linear e de que a variância não é necessariamente função da média, também o fato de proporções em $(0, 1)$ não estarem definidas sobre todos os reais, que é o domínio da distribuição normal, sugere que a suposição de normalidade é inadequada.

Outra prática tipicamente utilizada é aplicar uma transformação na variável dependente e então ajustar uma função de resposta linear para a variável transformada. Segundo Kieschnick e McCullough (2003), na grande maioria dos trabalhos é utilizada a transformação logito. O modelo de regressão logito (Dyke e Patterson, 1952) tem uma longa história em economia e áreas afins. O principal inconveniente quanto à aplicação deste modelo para analisar proporções é que ele assume que a transformação logito estabiliza a variância, enquanto outros modelos distribucionais para este tipo de dados, como o beta e o simplex, consideram que tal transformação não estabiliza a variância.

No que se refere aos modelos normais censurados, ou modelo Tobit, os autores salientam que, para dados observados no intervalo $(0, 1)$, a regressão Tobit é equivalente à distribuição normal. Logo, esta modelagem está exposta às mesmas críticas referentes aos modelos normal linear e não–linear.

O modelo de regressão que supõe distribuição beta para a variável resposta, proposto por Kieschnick e McCullough (2003), utiliza a função de ligação logito para a esperança condicional, de forma que os dois parâmetros da distribuição são expressos em função do inverso da exponencial do preditor linear. Esta proposta, apesar de utilizar a distribuição beta para a variável resposta, limita-se ao uso da função de ligação logito para a esperança condicional. Outro modelo paramétrico utilizado para modelar proporções no intervalo $(0, 1)$, desenvolvido por Barndorff-Nielsen e Jorgensen (1991), é baseado na distribuição simplex.

Kieschnick e McCullough (2003) também discutem o uso da técnica de quasi-verossimilhança para a modelagem de proporções. Esta técnica especifica o primeiro e o segundo momentos da distribuição condicional, mas não especifica a distribuição completa. Cox (1996) investiga o ajuste de quatro especificações para os dois primeiros momentos da distribuição condicional de duas amostras de dados de proporções no intervalo $(0, 1)$. Mais precisamente, Cox (1996) examina o uso das funções logito e complemento log-log para a esperança, com especificações canônica e ortogonal para a função de variância. O autor conclui que a função de ligação logito associada à função de variância ortogonal é a combinação que se adequa melhor ao seu conjunto de dados. A combinação que Cox (1996) denomina ortogonal é:

$\mu(\theta) = 1/(1 + e^{-\theta})$, $V(\mu) = \mu^2(1 - \mu^2)$. Papke e Wooldridge (1996) usam uma técnica similar à de Cox, mas que se mostra mais robusta para estimação de erros padrões.

Sabemos que o objetivo central da modelagem de regressão é realizar inferências sobre os parâmetros do modelo. A principal implicação da escolha inadequada de um tipo de modelo de regressão é exatamente fazer inferências incorretas a respeito destes parâmetros. Um estudo comparativo entre os modelos de regressão discutidos acima, apresentado por Kieschnick e McCullough (2003), evidencia que as conclusões inferenciais associadas aos modelos normal linear, normal não linear, normal censurado e logito, aplicados a um conjunto de dados envolvendo uma variável dependente, proporção de casas dentro de uma área de mercado que assinam televisão a cabo, e mais cinco covariáveis (Federal Communications Commission, 1993 e 1994), são equivalentes entre si, mas incompatíveis com as conclusões obtidas pelos modelos que usam a distribuição beta, simplex e de quasi-verossimilhança (especificações segundo Papke e Wooldridge, 1996). De fato, segundo Kieschnick e McCullough (2003), o modelo de regressão que supõe a distribuição beta com a ligação logito para a esperança condicional da variável resposta, o modelo simplex e o modelo de quasi-verossimilhança não se diferenciam quanto aos resultados inferenciais. No entanto, de acordo com uma variação do critério AIC, a estatística $\mathrm{AIC_c}$, considerada um critério de seleção apropriado para modelos não normais e de quasi-verossimilhança (McQuarrie e Tsai, 1998), o modelo que usa a suposição da distribuição beta supera os outros dois modelos (simplex e de quasi-verossimilhança). A superioridade do modelo associado à distribuição beta é tanto mais evidente quanto menor é o tamanho da amostra. Os autores finalizam a discussão argumentando que o fato de um modelo paramétrico beta dominar o modelo de quasi-verossimilhança, em um tamanho de amostra finito, é consistente com o fato de que modelos de quasi-verossimilhança são melhores aproximações para modelos paramétricos à medida que o tamanho da amostra aumenta.

Diante do exposto até o momento, apesar da proposta do modelo de regressão que utiliza a distribuição beta para a variável resposta ser relativamente específica, pois impõe a função de ligação logito para a esperança condicional, este modelo mostrou-se superior aos outros seis modelos considerados. De fato, a distribuição beta é muito flexível para modelar proporções, uma vez que dependendo dos dois parâmetros que indexam a distribuição, a densidade assume formas bem variadas. Adicionalmente, pelo fato da distribuição beta ser um membro da família exponencial, seus estimadores de máxima verossimilhança dispõem de boas propriedades amostrais. Várias aplicações e propriedades da distribuição beta são discutidas em Bury (1999). Johnson, Kotz e Balakrishnan (1995) fornecem vários exemplos de diferentes áreas da ciência em que a distribuição beta foi considerada a mais apropriada para ajustar dados de proporção, quando comparada a outras alternativas consideradas.

Dentro do contexto de utilização da distribuição beta[1] para situações em que a variável resposta se distribui continuamente no intervalo $(0,1)$ e pode ser explicada por outras variáveis através de uma estrutura de regressão, encontra-se um modelo de regressão beta proposto por Ferrari e Cribari–Neto (2004). Os autores propõem uma parametrização diferente para a distribuição beta que permite a modelagem da média da resposta envolvendo também o parâmetro de precisão. A função de densidade beta nessa reparametrização tem a forma

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \tag{1.1}$$

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

$$\mathrm{E}(y) = \mu$$

e

$$\mathrm{var}(y) = \frac{V(\mu)}{1+\phi},$$

em que $V(\mu) = \mu(1-\mu)$.

Nota-se que uma vez fixada a média da variável resposta, μ, quanto maior for o valor que ϕ assume, menor será a variância de y. Neste sentido, ϕ pode ser interpretado como um parâmetro de precisão do modelo.

Sejam $y_1, \ldots, y_n$ variáveis aleatórias independentes, em que cada y_t, $t = 1, \ldots, n$, segue a densidade em (1.1) com média μ_t e precisão desconhecida ϕ. O modelo proposto por Ferrari e Cribari–Neto (2004) é obtido assumindo que a média de y_t pode ser escrita como

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i = \eta_t, \tag{1.2}$$

em que $\beta = (\beta_1, \ldots, \beta_k)^\top$ é um vetor de parâmetros de regressão desconhecido ($\beta \in I\!R^k$), $x_{t1}, \ldots, x_{tk}$ são observações de k covariáveis ($k < n$), assumidas fixas e conhecidas e $g(\cdot)$ é uma função estritamente monótona e duas vezes diferenciável. Assim, de acordo com este modelo, temos que $\mu_t = g^{-1}(\eta_t)$ e $\mathrm{var}(y_t) = V(g^{-1}(\eta_t))/(1+\phi)$. Ou seja, uma vez que a variância da resposta depende de μ, mesmo o parâmetro de precisão sendo constante para todas as observações, as variâncias não são constantes.

Baseado em (1.1) segue que o logaritmo da função de verossimilhança é

$$\ell(\beta,\phi) = \sum_{t=1}^{n} \ell_t(\mu_t,\phi),$$

[1] Mais detalhes sobre a distribuição beta podem ser obtidos ainda em Ospina et. al (2006) e Oliveira (2004, Capítulo 2).

em que

$$\begin{aligned}\ell_t(\mu_t,\phi) &= \log\Gamma(\phi) - \log\Gamma(\mu_t\phi) - \log\Gamma((1-\mu_t)\phi) + (\mu_t\phi - 1)\log y_t \\ &\quad + \{(1-\mu_t)\phi - 1\}\log(1-y_t).\end{aligned} \tag{1.3}$$

A função escore para β é dada por

$$U_\beta(\beta,\phi) = \phi X^\top T(y^* - \mu^*) \tag{1.4}$$

com os t-ésimos elementos de y^* e μ^* dados, respectivamente, por

$$y_t^* = \log\left\{\frac{y_t}{(1-y_t)}\right\} \quad \text{e} \quad \mu_t^* = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi) \tag{1.5}$$

em que $\psi(\cdot)$ é a função digama, i.e., $\psi(z) = \mathrm{d}\log\Gamma(z)/\mathrm{d}z$ para $z > 0$ e X é uma matriz $n \times k$ cuja t-ésima linha é $x_t^\top$ e

$$T = \mathrm{diag}\{g'(\mu_1)^{-1}, \ldots, g'(\mu_n)^{-1}\}. \tag{1.6}$$

Adicionalmente, como veremos no próximo capítulo, tem-se que $\mu_t^* = \mathrm{E}(y_t^*)$.

Tem-se ainda, a função escore para o parâmetro de precisão,

$$U_\phi(\beta,\phi) = \sum_{t=1}^{n} u_t,$$

em que

$$u_t = \{\mu_t(y_t^* - \mu_t^*) + \log(1-y_t) - \psi((1-\mu_t)\phi) + \psi(\phi)\}, \tag{1.7}$$

e a matriz de informação de Fisher

$$K = K(\beta,\phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}. \tag{1.8}$$

Em (1.8) tem-se que

$$K_{\beta\beta} = \phi X^\top W X, \tag{1.9}$$

em que $W = \mathrm{diag}\{w_1, \ldots, w_n\}$, com

$$w_t = \phi v_t \frac{1}{g'(\mu_t)^2}, \tag{1.10}$$

sendo que

$$v_t = \psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi). \tag{1.11}$$

Adicionalmente, $K_{\beta\phi} = K_{\phi\beta}^\top = X^\top T c$ em que $c = (c_1, \ldots, c_n)^\top$, com

$$c_t = \phi\left\{\psi'(\mu_t\phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\right\}, \tag{1.12}$$

sendo $\psi'(\cdot)$ a função trigama. Finalmente, $K_{\phi\phi} = \mathrm{tr}(D)$, $D = \mathrm{diag}\{d_1, \ldots, d_n\}$, com

$$d_t = \psi'(\mu_t\phi)\mu_t^2 + \psi'((1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi). \tag{1.13}$$

Ferrari e Cribari–Neto (2004) ressaltam que, diferentemente do que acontece em modelos lineares generalizados (McCullagh e Nelder, 1989), no modelo de regressão beta, os parâmetros β e ϕ não são ortogonais. Os autores argumentam ainda que sob condições de regularidade, quando o tamanho da amostra é grande,

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\phi} \end{pmatrix} \sim \mathcal{N}_{k+1}\left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right),$$

aproximadamente, em que $\widehat{\beta}$ e $\widehat{\phi}$ são os estimadores de máxima verossimilhança de β e ϕ, respectivamente. Adicionalmente, usando expressões padrões para a inversa de matrizes particionadas (Rao, 1973, p. 33), Ferrari e Cribari–Neto (2004) obtêm K^{-1} dada por

$$K^{-1} = K^{-1}(\beta, \phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}, \tag{1.14}$$

em que

$$K^{\beta\beta} = \frac{1}{\phi}(X^\top W X)^{-1}\left\{ \mathcal{I}_k + \frac{X^\top T c c^\top T^\top X (X^\top W X)^{-1}}{\xi\phi} \right\},$$

com $\xi = \mathrm{tr}(D) - \phi^{-1} c^\top T^\top X (X^\top W X)^{-1} X^\top T c$,

$$K^{\beta\phi} = (K^{\phi\beta})^\top = -\frac{1}{\xi\phi}(X^\top W X)^{-1} X^\top T c,$$

e $K^{\phi\phi} = \xi^{-1}$. Aqui, $\mathcal{I}_k$ é a matriz identidade de ordem k.

Os estimadores de máxima verossimilhança de β e ϕ, são obtidos numericamente maximizando-se o logaritmo da função de verossimilhança usando um algoritmo de otimização não linear, como por exemplo um algoritmo de Newton (Newton–Raphson, Fisher's scoring, BHHH, etc.) ou um algoritmo quasi-Newton (BFGS, etc.).

A modelagem proposta por Ferrari e Cribari–Neto (2004) não requer transformação da resposta, permite interpretação dos parâmetros em termos da variável original (sem ser transformada) e assimetria da distribuição da variável de interesse, como também diversas possíveis escolhas para a função de ligação $g(\cdot)$. Por exemplo, pode-se utilizar a especificação logística $g(\mu) = \log\{\mu/(1-\mu)\}$, a função probito $g(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma variável aleatória normal padrão, a ligação log-log $g(\mu) = -\log\{-\log(\mu)\}$, entre outras. Para uma comparação destas e outras funções de ligação, ver Atkinson (1985, Ch. 7) e McCullagh e Nelder (1989, §4.3.1). Os autores propõem ainda testes de hipóteses usando a normalidade assintótica do estimador de máxima verossimilhança e algumas medidas de validação do modelo (métodos de diagnóstico).

Outros pesquisadores têm desenvolvido trabalhos baseados no modelo de regressão beta proposto por Ferrari e Cribari–Neto (2004). Ospina et. al (2006) estudaram métodos melhorados de estimação pontual e intervalar para o modelo de regressão beta. Oliveira (2004) apresenta um estudo sobre o comportamento assintótico dos estimadores e de testes de hipótese no modelo de regressão beta. O autor apresenta ainda duas aplicações a dados reais, uma das quais evidencia a vantagem do uso do modelo de regressão beta em comparação ao modelo normal linear quando a variável resposta é medida de forma contínua no intervalo $(0, 1)$. Segundo o autor: "o ajuste da regressão normal linear pode fornecer valores preditos fora do intervalo unitário, sendo que o mesmo não ocorre no modelo de regressão beta".

A presente proposta de tese de doutorado objetiva desenvolver vários aspectos de inferência e diagnóstico para a classe de modelos de regressão beta. No que se refere a técnicas de diagnóstico, propomos novos resíduos, medidas de influência global e local. A motivação inicial do desenvolvimento de novos métodos de validação para o modelo de regressão beta proposto por Ferrari e Cribari–Neto (2004) foi a obtenção de medidas de diagnóstico capazes de captar a presença de dispersão variável nos dados. O modelo de regressão beta com dispersão variável é o segundo objetivo deste projeto. Este trabalho apresentará ainda intervalos de predição baseados no método bootstrap.

1.2. Apresentação dos capítulos

No Capítulo 2 trataremos da análise de resíduos para o modelo de regressão beta com dispersão constante. Neste sentido proporemos um resíduo ordinário e padronizações para este resíduo. Apresentaremos simulações de Monte Carlo realizadas com o objetivo de investigar as distribuições empíricas dos resíduos propostos por Ferrari e Cribari–Neto (2004) e dos novos resíduos. Ainda naquele capítulo serão apresentadas duas aplicações a dados reais e um exemplo simulado.

No Capítulo 3, dando continuidade à análise de diagnóstico no modelo de regressão beta com dispersão constante, desenvolveremos a análise de influência global e local para este modelo. Propomos uma nova aproximação para a distância de Cook e quanto à análise de influência local desenvolveremos quatro esquemas de perturbação, alterando as estruturas da ponderação de casos, da variável resposta, de covariáveis individualmente e da dispersão. As novas técnicas serão ilustradas através de cinco aplicações a dados reais.

No quarto capítulo será apresentada uma extensão do modelo proposto por Ferrari e Cribari–Neto (2004) para situações em que o parâmetro de precisão ϕ não é constante para todas as observações. Apesar desta abordagem não ser inédita, (Paolino, 2001; Smithson e Verkuilen, 2006; Cuervo e Gamerman, 2004), alguns aspectos matemáticos não foram ainda

bem explorados. Neste sentido, serão desenvolvidos testes para a hipótese nula $\mathcal{H}_0 : \phi_1 = \ldots = \phi_n = \phi$, sendo ϕ_t o parâmetro de precisão associado a y_t, $t = 1, \ldots, n$. Adicionalmente, avaliaremos através de estudos de simulação as propriedades dos estimadores de máxima verossimilhança do modelo beta com dispersão variável, as conseqüências de estimar o modelo supondo dispersão constante quando de fato ela é variável e ainda o desempenho dos testes assintóticos da razão de verossimilhanças, escore e Wald para testar a hipótese de dispersão constante. Investigaremos ainda técnicas de diagnósticos no sentido de identificar a dispersão variável dos dados.

No quinto capítulo trataremos do problema de diagnóstico para o modelo de regressão beta com dispersão variável. Apesar de já existirem aplicações do modelo de regressão beta com dispersão variável, nada foi feito quanto à validação deste tipo de modelo. Neste sentido, desenvolveremos inicialmente resíduos, tornando possível a construção de gráficos de diagnóstico como, por exemplo, gráficos normais de probabilidades com envelope simulado. Em seguida desenvolveremos a análise de influência local, considerando três esquemas de perturbação, a saber: a ponderação de casos, a perturbação da variável resposta e perturbações individuais de covariáveis. Apresentaremos ainda aplicações a dados reais.

No sexto capítulo desenvolveremos um procedimento de obtenção de limites de predição para o modelo de regressão beta com dispersão constante a partir do método bootstrap (Efron, 1979). O procedimento terá como base a proposta de Davison e Hinkley (1997) para modelos lineares generalizados. Apresentaremos uma aplicação a dados reais para avaliar o desempenho do procedimento proposto. Por fim, no último capítulo apresentaremos as considerações finais que incluem conclusões e trabalhos futuros.

1.3. Suporte computacional

A linguagem de programação matricial `Ox`, criada por Jurgen Doornik em 1994, constitui a plataforma computacional utilizada no desenvolvimento da tese. Esta linguagem permite a implementação de técnicas estatísticas com facilidade e atende a requisitos como precisão e eficiência, o que tem contribuído para sua ampla utilização no campo da computação numérica. Detalhes sobre esta linguagem de programação podem ser encontrados em Doornik (2001). As análises gráficas desta tese foram produzidas utilizando o ambiente de programação, análise de dados e construção de gráficos `R` em sua versão 2.3.1 que, se encontra disponível gratuitamente no endereço `http://www.r-project.org`.

Capítulo 2

Análise de resíduos

2.1. Introdução

Em muitos casos temos conhecimento limitado sobre a relação entre variáveis envolvidas em um problema de interesse. Se visualizamos os valores observados de tais variáveis como os resultados de um experimento devemos ter, então, uma ferramenta teórica, um modelo matemático, através do qual estas variáveis estejam relacionadas, para atuar como base do processo gerador de dados. No entanto, todos os modelos são inevitavelmente simplificações, aproximações da realidade. Assim, uma etapa imprescindível da análise de regressão é a validação do modelo, no sentido de avaliar a qualidade desta aproximação. E é este o contexto da análise de diagnóstico.

Em uma direção, o interesse recai em avaliar possíveis afastamentos das suposições admitidas para o modelo, entre as quais está a distribuição de probabilidades dos dados. Em outra direção, o interesse recai em investigar a robustez do modelo sob pequenas perturbações nas formulações iniciais, no sentido de avaliar a estabilidade dos resultados inferenciais. O modelo é considerado não robusto se pequenas perturbações na sua constituição original implicam em resultados significativamente distintos.

A análise de diagnóstico teve início com a análise de resíduos objetivando detectar pontos mal ajustados ou aberrantes e avaliar indícios de afastamentos das suposições sobre o modelo, entre estas, a adequação da distribuição proposta para a variável resposta. A análise de resíduos pode se basear nos resíduos ordinários e suas possíveis padronizações e nos resíduos construídos a partir dos componentes da função desvio (vide, por exemplo, McCullagh e Nelder, 1989, p. 37–39), freqüentemente utilizados em modelos lineares generalizados. Muitos autores têm apresentado padronizações para o resíduo componente do desvio e têm obtido boas aproximações pela distribuição normal, a saber: McCullagh (1987), Williams (1984,1987), Davison e Gigli (1989), Farhrmeir e Tutz (1994), Paula (1995), de Souza e Paula (2002), entre outros. As técnicas gráficas utilizando resíduos são freqüentemente adotadas para a análise de diagnóstico. O uso de envelopes simulados, por exemplo, conforme proposto por Atkinson (1981) inicialmente para o modelo de regressão com erros normais, permite avaliar a qualidade do ajuste do modelo postulado.

Realizamos a seguir uma discussão sobre análise de resíduos no modelo de regressão beta apresentado no Capítulo 1. Inicialmente apresentaremos dois resíduos propostos por Ferrari e Cribari–Neto (2004). Em seguida, propomos um novo resíduo para o modelo de regressão beta e duas padronizações para este resíduo.

Através de simulações, investigaremos as distribuições empíricas dos dois resíduos propostos por Ferrari e Cribari–Neto (2004) e das versões padronizadas do novo resíduo. Adicionalmente, a partir de aplicações a dados reais, compararemos as análises de diagnóstico obtidas a partir da utilização destes resíduos. Devemos ressaltar que para a definição dos resíduos estamos considerando o modelo definido em (1.2) com ϕ conhecido. No entanto, na prática usaremos a estimativa de máxima verossimilhança de ϕ obtida do processo iterativo conjunto para β e ϕ.

2.2 Resíduo ordinário

Podemos definir como resíduo uma medida que objetiva identificar discrepâncias entre o modelo ajustado e os dados. Assim, é compreensível que a maioria dos resíduos esteja baseada na diferença $y_t - \widehat{\mathrm{E}(y_t)}$. Neste sentido foi definido o primeiro resíduo ordinário para o modelo de regressão beta por Ferrari e Cribari–Neto (2004), dado por

$$r_t = \frac{y_t - \widehat{\mu}_t}{\sqrt{\widehat{\mathrm{var}}(y_t)}}, \tag{2.1}$$

em que $\widehat{\mu}_t = g^{-1}(x_t^\top \widehat{\beta})$ e $\widehat{\mathrm{var}}(y_t) = \widehat{\mu}_t(1-\widehat{\mu}_t)/(1+\widehat{\phi})$.

No entanto, respeitado o formato da distribuição de probabilidade da variável resposta, é mais interessante pensar no resíduo como uma função $r(y_t, \widehat{\mathrm{E}(y_t)})$, definição geral de resíduos proposta por Cox e Snell (1968). É sob este ponto de vista que propomos um novo resíduo para o modelo de regressão beta. Considerando ϕ conhecido, desenvolvemos o processo iterativo Scoring de Fisher para β dado por

$$\beta^{(m+1)} = \beta^{(m)} + (K_{\beta\beta}^{(m)})^{-1} U_\beta^{(m)}(\beta), \tag{2.2}$$

em que $m = 0, 1, 2, \ldots.$ denota os passos do processo iterativo que é repetido até que a distância entre $\beta^{(m+1)}$ e $\beta^{(m)}$ seja menor que um valor de tolerância especificado, ocorrendo assim a convergência do processo. Considerando a função escore e a matriz de informação de Fisher para β definidas em (1.4) e (1.9), temos que o processo iterativo em (2.2) é dado por

$$\beta^{(m+1)} = \beta^{(m)} + (X^\top W^{(m)} X)^{-1} X^\top T^{(m)} (y^* - \mu^{*(m)}),$$

em que os t-ésimos elementos de y^* e de μ^* estão definidos em (1.5), as matrizes T e W estão definidas em (1.6) e (1.10), respectivamente, e X é uma matriz $n \times k$ cuja t-ésima linha é $x_t^\top$. Vale a pena ressaltar que $\mu^* = \mathrm{E}(y^*)$.

Podemos representar este processo como um processo iterativo de mínimos quadrados reponderados dado por $\beta^{(m+1)} = (X^\top W^{(m)} X)^{-1} X^\top W^{(m)} z^{(m)}$, em que $z^{(m)} = \eta^{(m)} + W^{-1(m)} T^{(m)}(y^* - \mu^{*(m)})$, com $\eta = (\eta_1, \ldots, \eta_n)^\top = X\beta$. Na convergência do processo temos que

$$\widehat{\beta} = (X^\top \widehat{W} X)^{-1} X^\top \widehat{W} z \tag{2.3}$$

com

$$z = \widehat{\eta} + \widehat{W}^{-1} \widehat{T}(y^* - \widehat{\mu}^*), \tag{2.4}$$

que pode ser interpretado como a solução de mínimos quadrados da regressão linear de z contra X com pesos $\widehat{W}$ em que o resíduo ordinário é dado por $r^* = \widehat{W}^{1/2}(z - \widehat{\eta}) = \widehat{W}^{-1/2}\widehat{T}(y^* - \widehat{\mu}^*)$. Assim, usando as definições de T e W dadas em (1.6) e (1.10), respectivamente, propomos um novo resíduo para o modelo de regressão beta, que chamamos de resíduo ponderado, definido como

$$r_t^* = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{\phi v_t}}, \tag{2.5}$$

em que v_t é dado em (1.11).

2.3. Resíduo ponderado padronizado 1

A análise do resíduo ponderado (2.5) indica que os resíduos no modelo de regressão beta devem se basear na diferença $y_t^* - \widehat{\mu}_t^*$, ou seja, na diferença entre o logito da observação e a estimativa de máxima verossimilhança de seu valor esperado segundo o modelo postulado. Assim, parece razoável calcular a variância de y_t^* e verificar sua relação com o resíduo ponderado.

Um resultado importante é que a densidade beta (1.1) pertence à família exponencial biparamétrica canônica. De fato,

$$f(y_t; \mu_t, \phi) = \exp\{\tau_1 T_1 + \tau_2 T_2 - \mathcal{A}(\tau)\}(1/y_t(1 - y_t)),$$

em que $\tau = (\tau_1, \tau_2) = (\mu_t\phi, \phi)$, $(T_1, T_2) = (\log\{y_t/(1 - y_t)\}, \log(1 - y_t))$ e

$$\mathcal{A}(\tau) = \{-\log\Gamma(\phi) + \log\Gamma(\mu_t\phi) + \log\Gamma((1 - \mu_t)\phi)\}.$$

Assim, temos $\mathrm{E}(T_1) = \mathrm{E}(y_t^*) = \partial\mathcal{A}(\tau)/\partial\tau_1 = \psi(\mu_t\phi) - \psi((1 - \mu_t)\phi) = \mu_t^*$, resultado que já havíamos apresentado e

$$\mathrm{var}(T_1) = \mathrm{var}(y_t^*) = \partial^2\mathcal{A}(\tau)/\partial\tau_1^2 = \psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi) = v_t;$$

ver Lehmann e Casella (1998, p. 27). Logo, uma proposta razoável de resíduo padronizado é considerar a diferença $y_t^* - \widehat{\mu}_t^*$ dividida apenas pela raiz quadrada de $\mathrm{var}(y_t^*)$, o que conduz

ao resíduo que chamamos de resíduo ponderado padronizado 1, dado por

$$r_t^p = \phi^{1/2} r_t^* = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{v_t}}. \tag{2.6}$$

2.4. Resíduo ponderado padronizado 2

Uma outra alternativa de padronização para o resíduo ponderado se baseia na variância de z. Podemos reescrever (2.3) como $(X^\top \widehat{W} X)\widehat{\beta} = X^\top \widehat{W} z$. Uma vez que $\text{cov}(\widehat{\beta}) = \phi^{-1}(X^\top W X)^{-1}$, segue que $\text{cov}(z) \approx \phi^{-1}\widehat{W}^{-1}$. Então, usando (2.4) obtemos

$$\text{cov}(r^*) \approx \phi^{-1}[\mathcal{I}_n - H^*],$$

em que

$$H^* = \widehat{W}^{1/2} X (X^\top \widehat{W} X)^{-1} X^\top \widehat{W}^{1/2} \tag{2.7}$$

é a matriz de projeção da solução de mínimos quadrados da regressão linear de z contra X com pesos $\widehat{W}$ e $\mathcal{I}_n$ é a matriz identidade de ordem n. Desta forma, podemos definir o resíduo ponderado padronizado 2 como

$$r_t^{pp} = \frac{r_t^*}{\sqrt{\phi^{-1}(1 - h_{tt}^*)}} = \frac{r_t^p}{\sqrt{(1 - h_{tt}^*)}} = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{v_t(1 - h_{tt}^*)}}, \tag{2.8}$$

sendo h_{tt}^* o t-ésimo elemento da diagonal principal de H^*.

No processo de obtenção do resíduo ponderado foi feita uma analogia com o método de mínimos quadrados reponderados. No entanto, devemos ressaltar que no modelo de regressão beta, assim como em modelos lineares generalizados, os pesos "$\widehat{w}_t$" são determinados pelo ajuste, o que não pode ser confundido com o problema de mínimos quadrados ponderados, em que os pesos determinam o ajuste.

2.5. Resíduo componente do desvio

Os resíduos construídos a partir dos componentes da função desvio (McCullagh e Nelder 1989) se baseiam na distância para cada observação entre o máximo do logaritmo da função de verossimilhança do modelo saturado e o máximo do logaritmo da função de verossimilhança do modelo investigado.

Como, para valores grandes de ϕ, $\widetilde{\mu}_t \approx y_t$, em que $\widetilde{\mu}_t$ é a estimativa de μ do modelo saturado, Ferrari e Cribari–Neto (2004) propõem um resíduo componente desvio para o modelo de regressão beta dado por

$$r_t^d = \text{sign}(y_t - \widehat{\mu}_t) \left\{ 2(\ell_t(y_t, \widehat{\phi}) - \ell_t(\widehat{\mu}_t, \widehat{\phi})) \right\}^{1/2}, \tag{2.9}$$

em que $\widehat{\mu}_t = g^{-1}(x_t^\top \widehat{\beta})$ é a estimativa de μ para o modelo postulado.

Uma vez que a aproximação $\widetilde{\mu}_t \approx y_t$ é válida apenas para valores grandes de ϕ, é razoável que o resíduo componente do desvio definido acima não seja adequado a situações em que ϕ é pequeno. No entanto, como veremos nos exemplos reais, mesmo em situações em que a estimativa de ϕ é superior a 400, para algumas observações este resíduo não pode ser calculado. Constatamos que este problema ocorre devido ao valor negativo da diferença $\ell_t(y_t, \widehat{\phi}) - \ell_t(\widehat{\mu}_t, \widehat{\phi})$. Para investigar questões como esta relacionada ao resíduo componente do desvio e para avaliar as distribuições dos resíduos apresentados nas seções anteriores, realizamos simulações em cenários que considerem diferentes valores de ϕ, de médias e de tamanhos amostrais.

2.6. Resultados numéricos (simulações)

Realizamos uma simulação de Monte Carlo com 1000 réplicas, objetivando investigar as distribuições empíricas de alguns dos resíduos discutidos na Seção 2.3, a saber: o resíduo componente do desvio; expressão (2.9), o resíduo ordinário padronizado; expressão (2.1), e os resíduos ponderados padronizados 1 e 2; expressões (2.6) e (2.8), respectivamente.

Consideramos o modelo de regressão beta com

$$g(\mu_t) = \beta_1 + \beta_2 x_t, \ t = 1, \ldots, n,$$

sendo g a função de ligação logito. Consideramos três distribuições de probabilidades diferentes para gerar os valores da covariável x_t: $x_t \sim \mathcal{U}(0,1)$, $x_t \sim t_3$ e $x_t \sim \exp(2)$ (exponencial de média 2), devidamente padronizadas. Consideramos $\phi = \exp(\delta)$ e admitimos $\delta = 2.5$ e $\delta = 5$, o que conduz a $\phi \approx 12$ e $\phi \approx 148$, respectivamente. No caso em que $x_t \sim \mathcal{U}(0,1)$, consideramos dois cenários diferentes. No primeiro, utilizamos $\beta_1 = 4.0$ e $\beta_2 = -0.8$, o que conduz a valores da média da variável resposta próximos de um, mais especificamente, obtivemos $\mu \in (0.96, 0.98)$. No segundo cenário, os verdadeiros valores dos parâmetros são $\beta_1 = -2.5$ e $\beta_2 = -1.2$ e neste caso obtivemos $\mu \in (0.024, 0.075)$, ou seja, valores da média próximos de zero. Quanto à utilização da distribuição t–Student com três graus de liberdade (t_3), os verdadeiros valores dos parâmetros são $\beta_1 = 1.21$ e $\beta_2 = 1.25$ e aqui $\mu \in (0.05, 0.93)$. Finalmente, para a distribuição exponencial tomamos $\beta_1 = -1.2$ e $\beta_2 = -1.3$ obtendo $\mu \in (0.04, 0.53)$.

Objetivando confrontar os quantis empíricos dos resíduos com os quantis teóricos da distribuição normal padrão construímos, gráficos normais de probabilidades apresentados nas Figuras 2.1–2.6. Com relação ao resíduo componente do desvio, avaliamos o comportamento de sua distribuição para os três últimos valores de ϕ e considerando inicialmente dois tamanhos amostrais ($n = 20$ e $n = 60$), Figuras 2.1 e 2.2, respectivamente. Nota-se que para

$\phi = 20$, quando os valores de μ estão próximos de um ou próximos de zero, mais de 80% das observações não têm resíduos válidos, tanto para $n = 20$ quanto para $n = 60$ (Figuras 2.1a,d e 2.2a,d). Quando a precisão aumenta, $\phi = 148$ e $\phi = 403$, o percentual de resíduos não calculados cai para 45% e 30%, respectivamente (Figuras 2.1b,c,e,f e 2.2b,c,e,f). Nas situações em que as médias não se encontram concentradas próximas aos extremos do intervalo $(0, 1)$ e mesmo quando $\phi = 403$, o percentual de resíduos não calculados ainda é considerável, cerca de 25% (Figuras 2.1g–k e 2.2g–k). Em todos os casos investigados, a distribuição do resíduo componente do desvio aproxima-se mais da distribuição normal à medida que ϕ aumenta. No entanto, nos cenários em que $\mu \in (0.96, 0.98)$ e $\mu \in (0.025, 0.075)$, mesmo quando $\phi = 403$, a aderência à distribuição normal não é satisfatória.

Adicionalmente, ao que parece, o aumento do tamanho amostral não contribui para que a distribuição deste resíduo esteja mais próxima da distribuição normal. Para visualizar melhor este fato, na Figura 2.3 apresentamos gráficos normais de probabilidades quando fixamos $\phi = 148$ e variamos o tamanho amostral considerando $n = 40$, $n = 60$ e $n = 120$. Cada coluna desta figura representa um tamanho amostral e cada linha um cenário para a média da resposta. O que podemos notar, como suspeitávamos, é que em todos os cenários investigados para a média da resposta, a aderência da distribuição do resíduo componente do desvio à distribuição normal não parece melhorar à medida que n aumenta.

Para investigar a distribuição dos demais resíduos fixamos o tamanho amostral em $n = 20$ e trabalhamos com $\phi = 12$ (Figura 2.4), $\phi = 20$ (Figura 2.5) e $\phi = 148$ (Figura 2.6). A análise destas figuras revela que os resíduos ponderados padronizados 1 e 2 apresentam distribuição consideravelmente próxima da normal padrão, em especial para o valor $\phi = 148$, mesmo em situações extremas, quando os valores da média da variável resposta encontram-se próximos de um ou de zero. Observa-se, no entanto, que para ϕ pequeno, $\phi = 12$ e $\phi = 20$, a distribuição destes resíduos apresenta uma pequena assimetria, à direita quando $\mu \in (0.96, 0.98)$ e à esquerda quando $\mu \in (0.024, 0.075)$. (Figuras 2.4b,c,e,f e 2.5b,c,e,f).

No que se refere ao resíduo ordinário padronizado (Ferrari e Cribari–Neto 2004), nas situações em que $\mu \in (0.96, 0.98)$ e $\mu \in (0.024, 0.075)$ é necessário que a precisão dos dados aumente para que ocorra uma melhor aderência à distribuição normal (Figuras 2.4a,d, 2.5a,d e 2.6a,d). No entanto, assim como para os outros dois resíduos no cenário em que as médias não se concentram próximas aos extremos do intervalo $(0, 1)$, não há evidências para desconsiderar que a aproximação normal para a distribuição dos resíduos seja adequada.

Um resultado obtido, porém não apresentado, é que as conclusões acima são válidas para os tamanhos amostrais $n = 40$ e $n = 60$ e também considerando a função de ligação probito. Assim, ao que parece, os resíduos ponderados padronizados 1 e 2 apresentaram distribuições empíricas mais próximas da normal padrão que as dos demais resíduos.

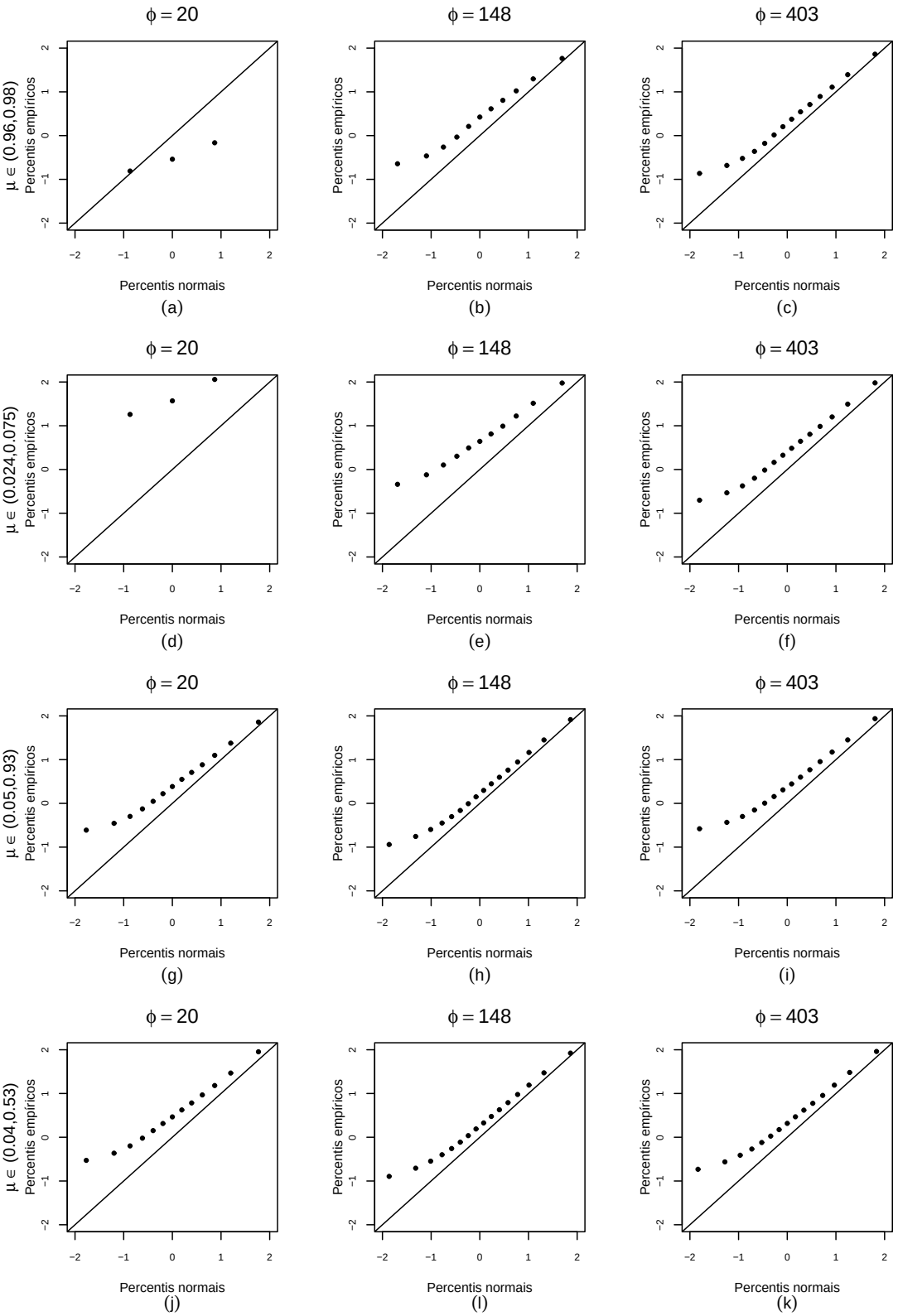

Figura 2.1. Gráficos normais de probabilidades do resíduo componente do desvio, $n = 20$.

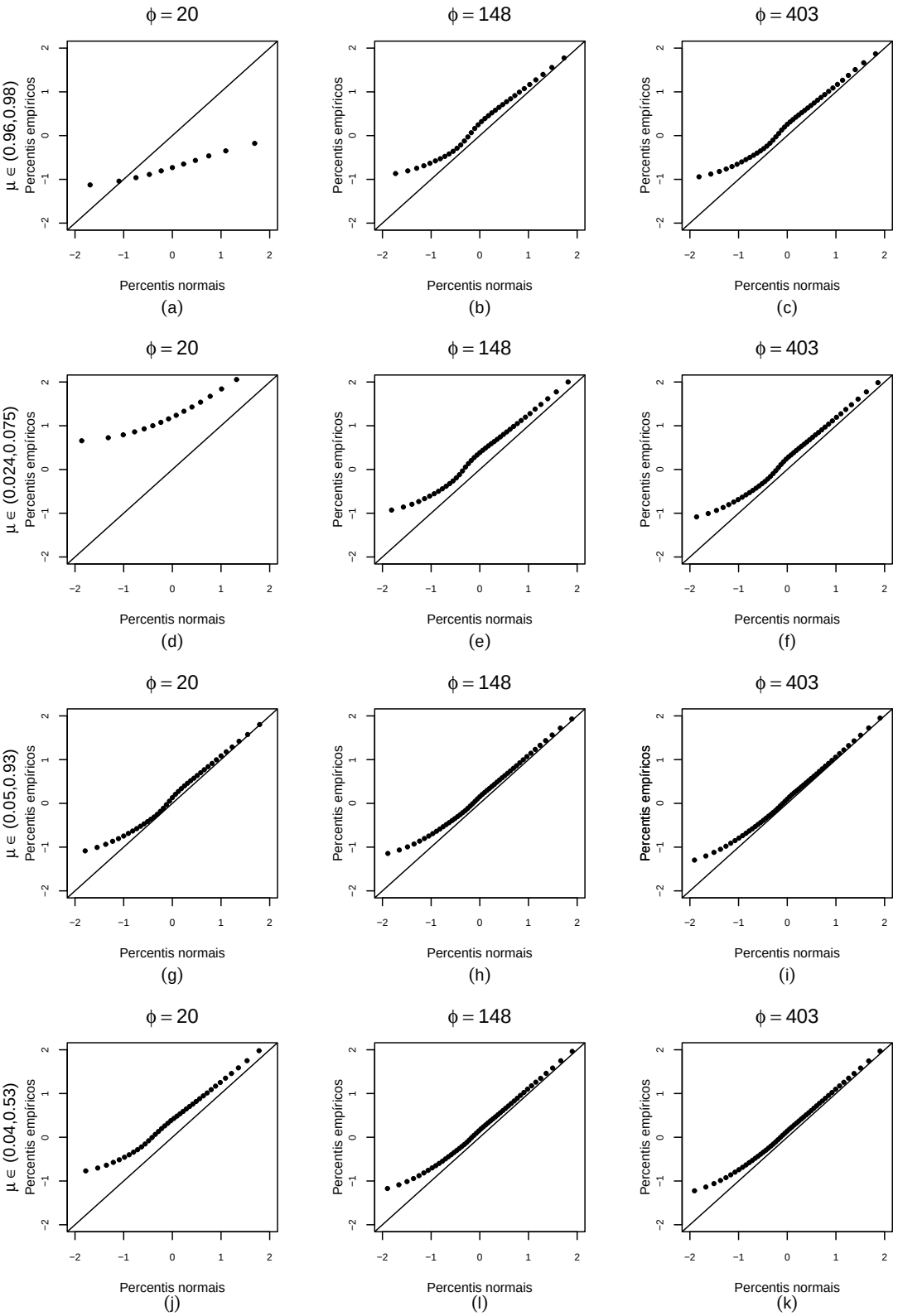

Figura 2.2. Gráficos normais de probabilidades do resíduo componente do desvio, $n = 60$.

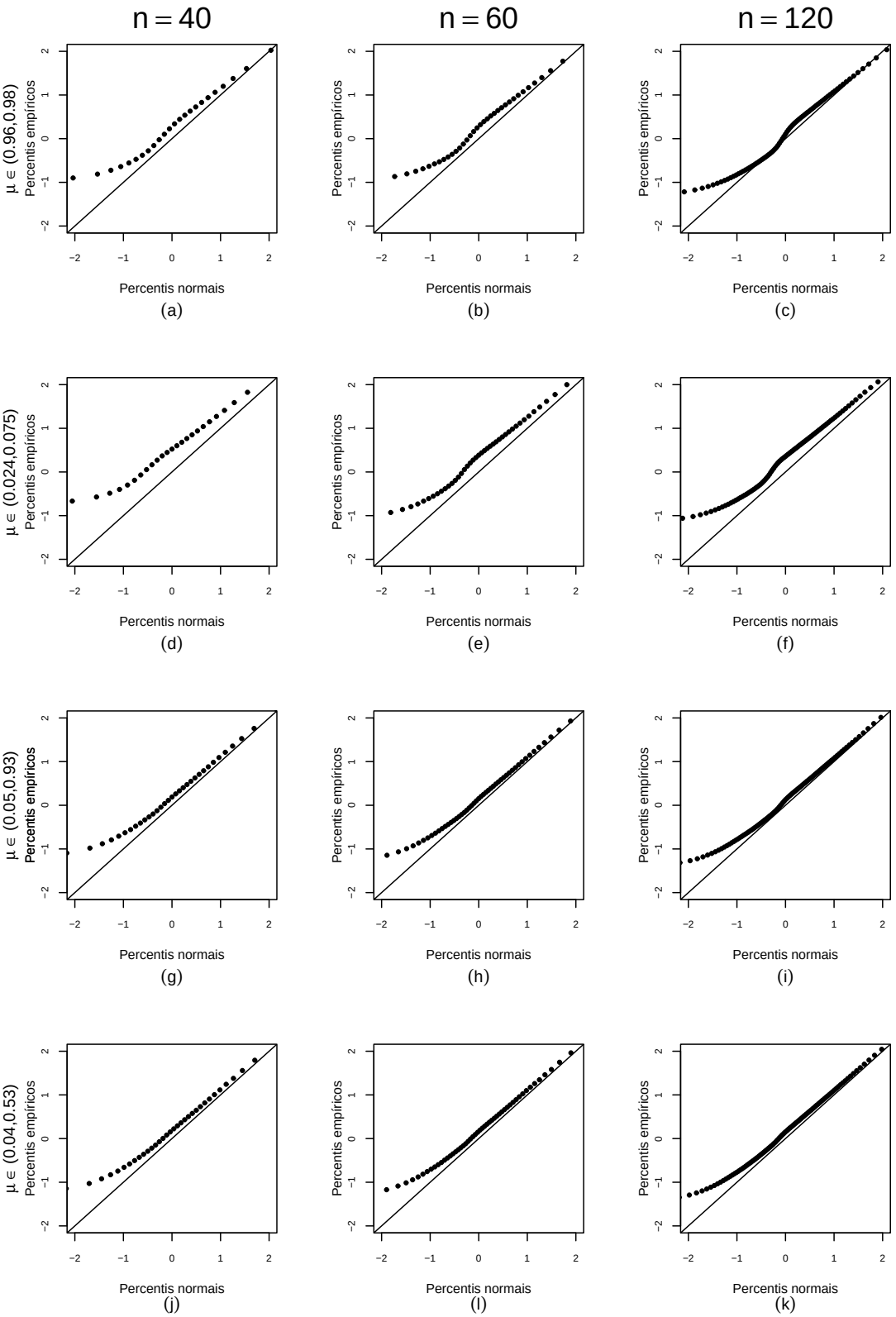

Figura 2.3. Gráficos normais de probabilidades do resíduo componente do desvio, variando tamanho da amostra, $\phi = 148$.

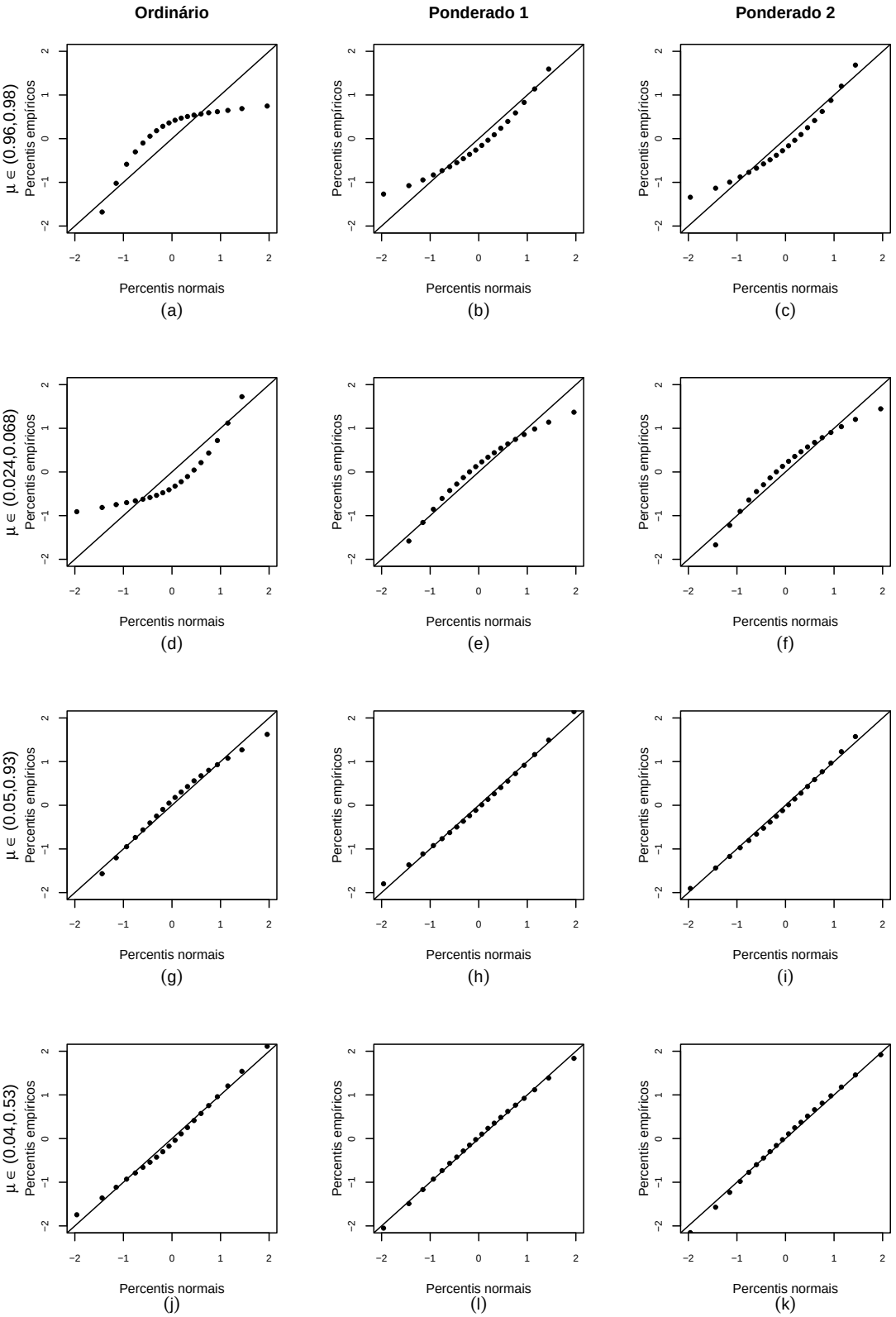

Figura 2.4. Gráficos normais de probabilidades dos resíduos ordinário padronizado, ponderado padronizado 1 e ponderado padronizado 2, $n = 20$, $\phi = 12$.

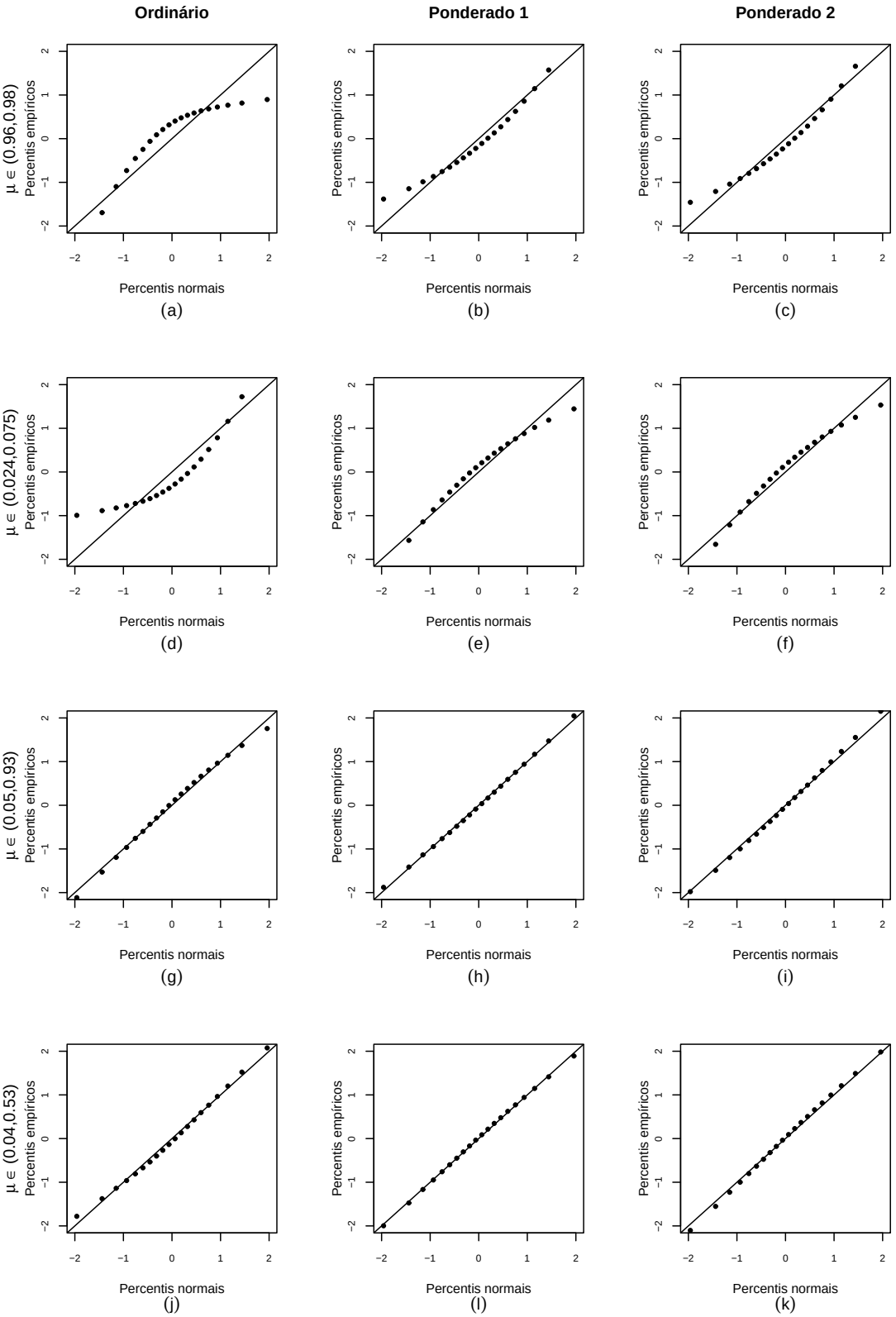

Figura 2.5. Gráficos normais de probabilidades dos resíduos ordinário padronizado, ponderado padronizado 1 e ponderado padronizado 2, $n = 20$, $\phi = 20$.

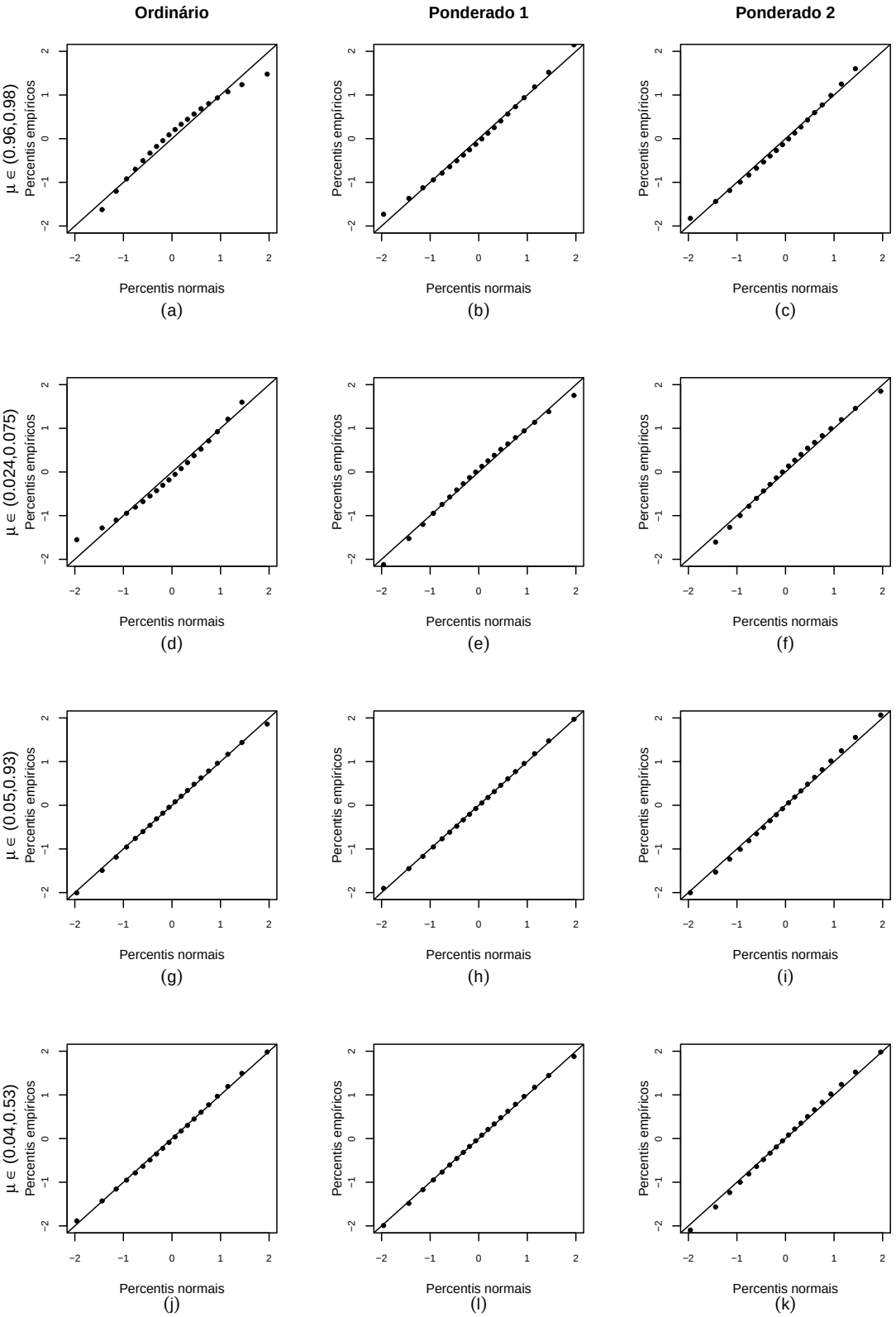

Figura 2.6. Gráficos normais de probabilidades dos resíduos ordinário padronizado, ponderado padronizado 1 e ponderado padronizado 2, $n = 20$, $\phi = 148$.

2.7. Aplicações

Nesta seção apresentaremos três aplicações, duas delas baseadas em dados reais e uma baseada em dados simulados. As aplicações têm como objetivo comparar o desempenho dos resíduos componente do desvio, ordinário padronizado, ponderado padronizado 1 e ponderado padronizado 2, no sentido de identificar possíveis diferenças quanto às conclusões obtidas a partir da análise de diagnóstico.

2.7.1. Aplicação I: dados de gastos com alimentação

Nesta primeira aplicação o interesse recai em modelar a proporção da renda familiar gasta com alimentação como função da renda total, x_2 e o número de pessoas no domicílio, x_3, em um total de 38 domicílios de uma grande cidade dos Estados Unidos (Griffiths et al. 1993, Tabela 15.4). Ferrari e Cribari–Neto (2004) analisaram esses dados utilizando a função de ligação logito. Nesta aplicação utilizamos as funções de ligação logito e probito. Na Tabela 2.1 estão apresentados os resultados da estimação considerando as duas funções de ligação. Nota-se a partir desta tabela que as estimativas do parâmetro de precisão para este conjunto de dados são praticamente iguais quando comparados os modelos probito e logito.

Tabela 2.1. Resultados inferenciais. Dados de gastos com alimentação

Parâmetros	Ligação Logito		Ligação Probito	
	Estimativas	p–valor	Estimativas	p–valor
β_1	-0.622	0.0054	-0.388	0.0042
β_2	-0.012	0.0000	-0.007	0.0000
β_3	0.118	0.0008	0.069	0.0010
ϕ	35.609		35.133	

A variável resposta neste caso encontra-se no intervalo $(0.108, 0.562)$ com mediana igual a 0.261. As estimativas de ϕ considerando as duas funções de ligação são próximas de 35; neste cenário as distribuições dos resíduos ordinário padronizado, e ponderados padronizados 1 e 2 são melhor aproximadas pela distribuição normal padrão que a distribuição do resíduo componente do desvio, o que pode ser comprovado através dos gráficos normais de probabilidades com envelopes (Figuras 2.7e–h e 2.8e–h). Adicionalmente, tanto para o modelo logito quanto para o modelo probito, para aproximadamente 8% das observações o resíduo componente do desvio não pôde ser calculado. Apesar deste fato, os gráficos dos resíduos contra os índices das observações apresentam comportamento semelhante e evidenciam que os resíduos se distribuem aleatoriamente em torno do zero e dentro dos limites $(-2, 2)$. Devemos ressaltar, no entanto, que o gráfico que utiliza o resíduo ponderado padronizado 2 destaca

mais observações aberrantes que os gráficos dos demais resíduos, a saber: as observações 5, 11 e 20. Com relação aos gráficos normais de probabilidades, nota-se uma leve tendência dos resíduos entre -1 e 1 estarem próximos ou acima da banda superior do envelope. Contudo, não há indícios a partir dos demais gráficos de resíduos para desconsiderar que o modelo de regressão beta seja adequado aos dados. Uma vez que no Capítulo 3 voltaremos a tratar deste conjunto de dados, complementando a análise de diagnóstico através da análise de influência, não excluiremos os pontos destacados como aberrantes para realizar uma análise confirmatória.

2.7.2. Aplicação II: dados de habilidade de leitura

A segunda aplicação se refere ao exemplo 3 apresentado por Smithson e Verkuilen (2006). Os dados foram obtidos de Pammer e Kevin (2004). Trata-se de um estudo sobre a habilidade em leitura de um grupo de 44 crianças realizado na escola de psicologia da Universidade Nacional da Austrália. Smithson e Verkuilen (2006) consideram a contribuição relativa da sensitividade visual (presença de dislexia) e do QI não–verbal na habilidade de leitura das 44 crianças. A variável resposta é o escore em um teste de habilidade em leitura (y) e as variáveis independentes são a ocorrência de dislexia (x_2) e o escore padronizado de QI não–verbal (x_3), chamadas daqui em diante de dislexia e QI, respectivamente. A covariável x_2 assume os valores 1 e -1 indicando se a criança é ou não disléxica, respectivamente. A variável resposta é obtida após transformações dos dados originais. Originalmente, os escores do teste de leitura, que denotaremos por y', encontram-se no intervalo $[a, b]$, em que b é o escore máximo possível para o teste de leitura e a é o escore mínimo possível. A primeira transformação realizada nos dados foi $y'' = (y' - a)/(b - a)$. Para garantir que a variável resposta pertence ao intervalo aberto $(0, 1)$ os autores fazem uma segunda transformação, $y = [y''(n - 1) + 0.5]/n$, em que n é o tamanho da amostra.

Inicialmente os autores consideram um modelo de regressão linear normal em que as covariáveis são o escore padronizado de QI , ausência ou presença de dislexia e a interação entre as duas covariáveis. Os autores suspeitam da possibilidade de variância não constante, o que é confirmado através do teste de Levene. Em seguida, Smithson e Verkuilen (2006) aplicam a transformação logito à resposta e consideram um modelo linear com as mesmas covariáveis anteriores. Neste caso, apenas o efeito de dislexia é significativo. Ao que parece, os escores de QI trazem pouca ou nenhuma contribuição para a explicação dos escores de leitura. Os autores ainda consideram um modelo de regressão beta com ligação logito para a média, com as mesmas covariáveis do modelo normal e chegam ao mesmo resultado anterior, apenas dislexia se mostra significativa. Smithson e Verkuilen (2006) argumentam que esta conclusão pode ser enganosa, já que o efeito das covariáveis na dispersão dos dados não está

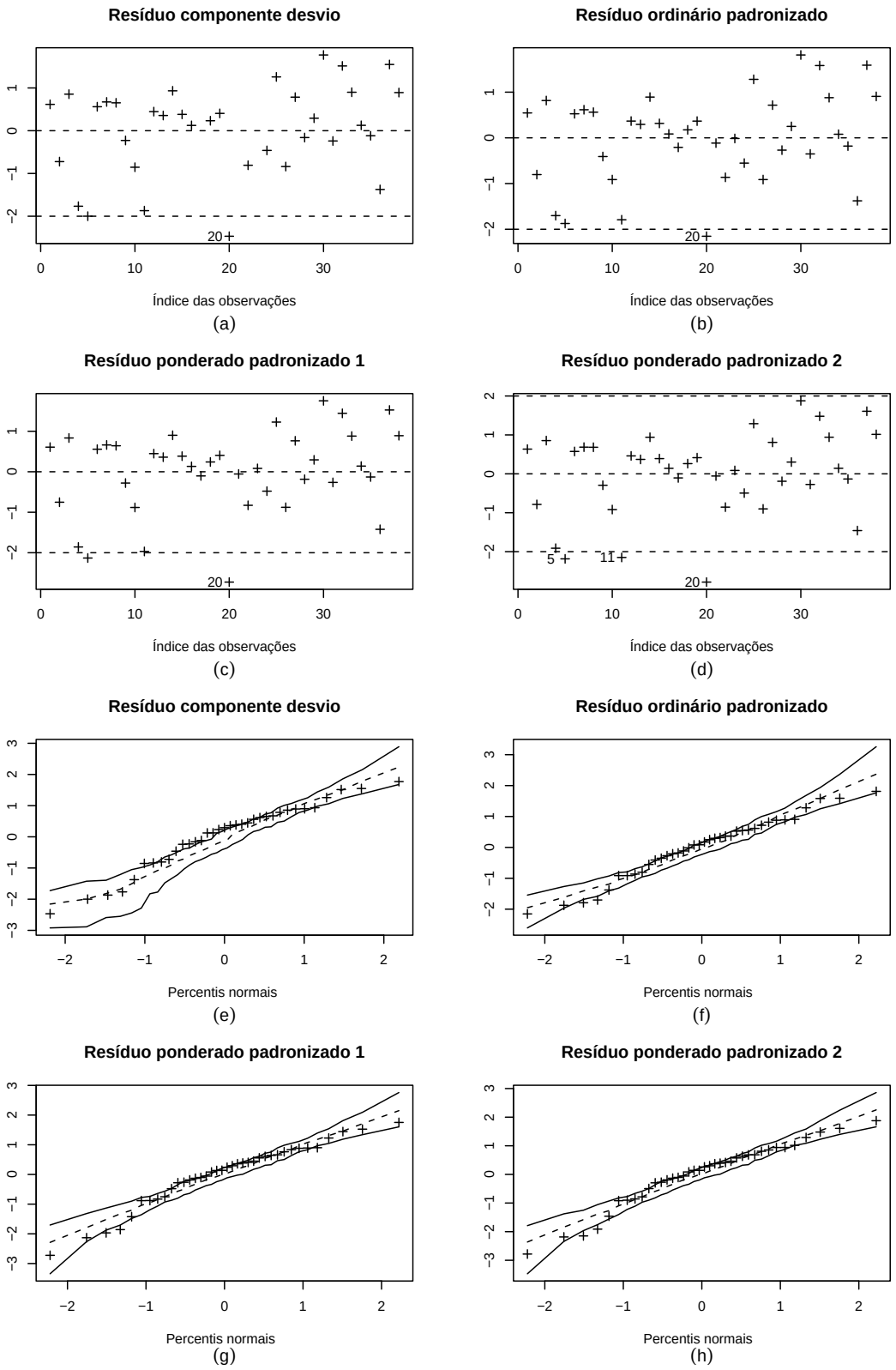

Figura 2.7. Gráficos de resíduos. Dados de gastos com alimentação. Ligação logito.

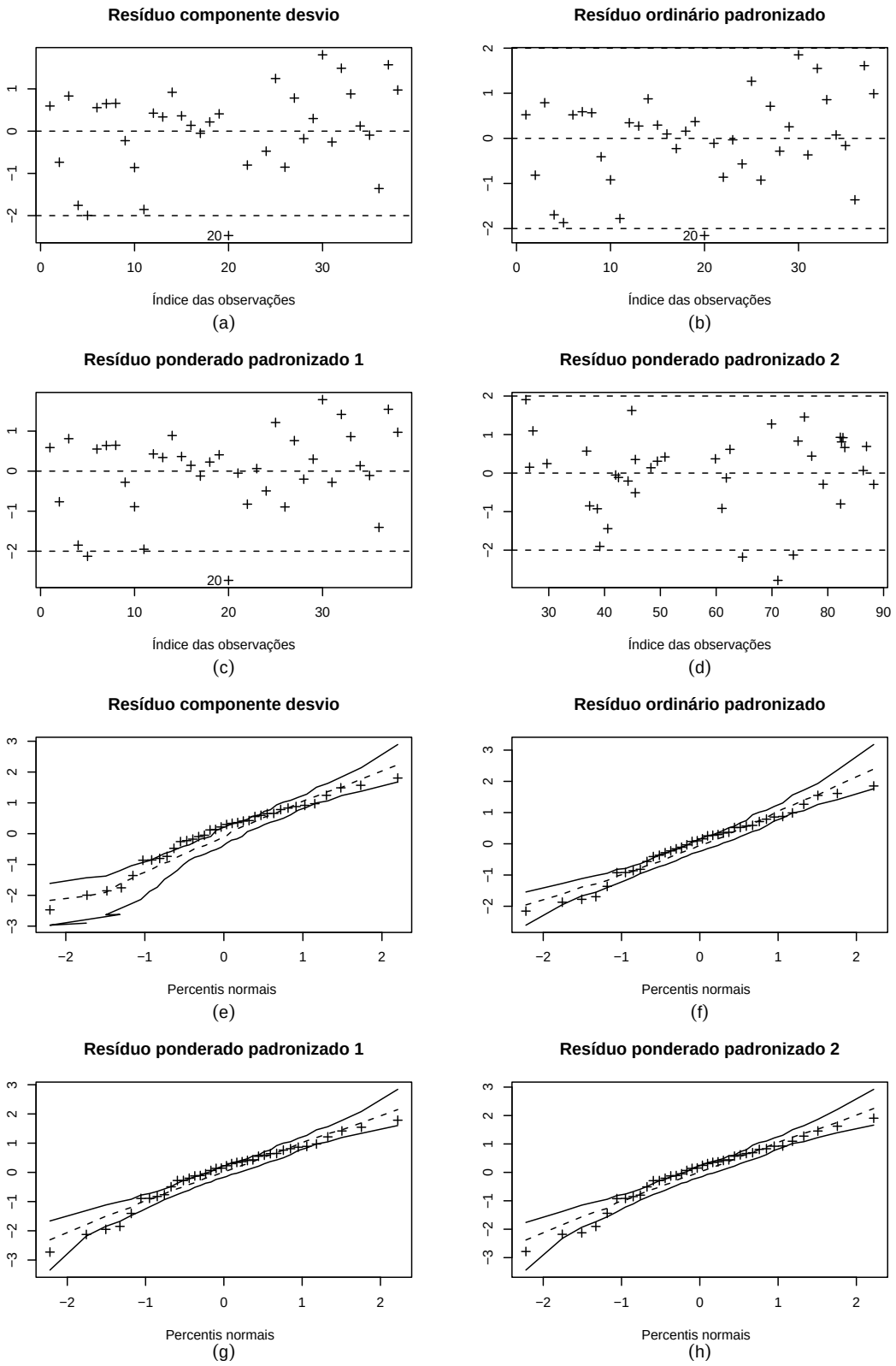

Figura 2.8. Gráficos de resíduos. Dados de gastos com alimentação. Ligação probito.

sendo considerada. Com base neste argumento, os autores propõem um modelo de regressão beta em que a média e a dispersão são modeladas simultaneamente.

Nesta segunda aplicação realizaremos uma análise de resíduos, usando apenas o resíduo ordinário padronizado e os resíduos ponderados padronizados 1 e 2, para um modelo de regressão beta com dispersão constante ajustado a esses dados. O modelo para a média considera as covariáveis dislexia, QI e a interação e a função de ligação logito. Um dos objetivos da análise a seguir é identificar se existem indícios da necessidade da modelagem da dispersão, ou se um modelo de regressão beta com dispersão constante considerando apenas como covariável a presença de dislexia está adequado aos dados. Os resultados inferenciais da segunda aplicação encontram-se na Tabela 2.2.

Tabela 2.2. Resultados inferenciais. Dados de habilidade de leitura

Parâmetros	Ligação Logito		Ligação Probito	
	Estimativas	p–valor	Estimativas	p–valor
β_1	1.333	0.0000	0.779	0.0000
β_2	-0.973	0.0000	-0.554	0.0000
β_3	0.160	0.2317	0.077	0.2855
β_4	-0.218	0.1042	-0.113	0.1188
ϕ	11.133		11.162	

De acordo com os resultados apresentados na Tabela 2.2 nota-se que tanto no modelo logito quanto no probito, além do intercepto e da precisão, apenas a covariável dislexia é considerada significativa (β_2). Apesar deste fato, para efeito da análise de diagnóstico consideraremos o modelo com as duas covariáveis e a interação. As estimativas de ϕ dos modelos logito e probito são praticamente idênticas e notadamente pequenas. Uma vez que a dispersão dos dados é alta comparada aos dados da aplicação I, aproximadamente três vezes maior que a dispersão dos dados de gastos com alimentação, nesta situação é muito interessante investigar o ajuste de um modelo de regressão beta. No entanto, deve-se ressaltar que a variável resposta não se encontra concentrada nos extremos do intervalo $(0, 1)$. De fato, $y \in (0.459, 0.990)$ e a mediana de y é 0.706. A mediana do escore de habilidade em leitura para as crianças não–disléxicas é 0.990 enquanto que para as disléxicas é 0.609. Dado que as impressões são praticamente idênticas para os modelos logito e probito, na investigação a seguir consideraremos apenas o modelo beta logito descrito acima.

O valor baixo da estimativa de ϕ, como era de se esperar, interferiu fortemente no resíduo componente do desvio, o que não havia ocorrido nas aplicações anteriores. Neste caso, para aproximadamente 40% das observações não foi possível calcular este resíduo e também tornou-se inviável a construção do correspondente gráfico normal de probabilidades.

Apesar de não se tratar de uma situação em que a resposta está concentrada em um dos extremos do intervalo $(0, 1)$, a alta dispersão dos dados conduz a uma situação em que as distribuições dos resíduos ponderados padronizados 1 e 2 são levemente assimétricas à esquerda. Nota-se que a assimetria à direita relacionada à distribuição do resíduo ordinário padronizado é consideravelmente acentuada (Figura 2.9d–f). Nesta situação é possível notar diferenças nos gráficos dos três resíduos contra os índices das observações. Através do gráfico do resíduo ordinário padronizado seriam destacadas as observações 2, 17, 19 e 24 como aberrantes (Figura 2.9a–c), enquanto os gráficos dos resíduos ponderados destacam apenas o caso 8. Salientamos que a análise confirmatória será realizada no próximo capítulo, quando complementaremos a análise de diagnóstico usando influência local.

É importante notar ainda, a partir da observação da Figura 2.9, que os gráficos normais de probabilidades com envelopes simulados dos resíduos ponderado padronizado 1 e ponderado padronizado 2 evidenciam uma leve falta de qualidade no ajuste do modelo, que pode estar relacionada à variância assumida. Essa suspeita é reforçada pelo gráfico dos resíduos contra os índices das observações, uma vez que é notável a diferença entre as dispersões da primeira metade e da segunda metade dos dados. Isto provavelmente ocorre porque os dados se dividem de acordo com a variável que indica a presença de dislexia. Até a observação 26 estão os casos de pessoas sem dislexia e as observações finais referem-se a não dislexia. Claramente o grupo de indivíduos sem dislexia apresenta maior dispersão que o grupo com dislexia, no que diz respeito aos escores de habilidade de leitura.

Objetivando investigar possíveis relações entre as covariáveis e a dispersão, a qual é assumida constante ao longo das observações, construímos gráficos dos resíduos contra os valores das covariáveis (Figura 2.10). A partir dessa figura nota-se que, de fato, o grupo de crianças com dislexia apresenta maior variabilidade quanto à habilidade de leitura que o outro grupo. Quanto a relação entre a covariável QI e a dispersão dos dados, nota-se que existe maior dispersão para os indivíduos cuja medida de QI está entre -1 e 1 que para aqueles em que a medida de QI se encontram nos extremos (Figura 2.10d–f). Ao que parece, tanto a covariável QI quanto a covariável dislexia são importantes para a explicação da dispersão dos dados.

2.7.3. Exemplo simulado

Como vimos nas aplicações anteriores, ao que parece, os gráficos que utilizam o resíduo ponderado padronizado 2 identificam como observações aberrantes casos que não são detectados pelos gráficos que utilizam os outros dois resíduos. Simulamos um exemplo com uma única amostra, para investigar melhor este fato. Geramos $n - 1$ observações de uma covariável $x_t \sim \mathcal{U}(0, 1)$ e adicionamos um n–ésimo valor à amostra, $x_n = 3.0$, induzindo

a ocorrência de uma observação alavanca. Considerando o modelo de regressão beta com $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, $t = 1, \ldots, n$, $\beta_1 = 3.5$, $\beta_2 = -3.4$ e $\phi = 403.43$ ($\phi = \exp(\delta)$ e $\delta = 6.0$), geramos $n = 30$ valores da variável resposta y, formando assim um conjunto de dados que chamaremos de dados corretos.

Alteramos o conjunto de dados corretos, simulando um possível erro de entrada dos dados, de forma que o valor da variável resposta relacionada a $x_{30} = 3.0$, y_{30}, fosse uma réplica de outro valor da resposta presente nos dados. Mais especificamente, escolhemos o máximo de y_t, $t = 1, \ldots, 29$, induzindo a ocorrência de um ponto alavanca e influente. Assim, criamos um conjunto de dados mantendo os valores da covariável x e com o trigésimo valor de y alterado. Chamamos este conjunto de dados de "dados com erro". Ajustamos o modelo de regressão beta logito a três conjuntos de dados: o conjunto de dados correto, o conjunto de "dados com erro" e o conjunto de dados excluindo a observação 30. Na Tabela 2.3 apresentamos os resultados inferenciais dos modelos considerando esses três conjuntos. A variação percentual apresentada na sexta linha desta tabela representa uma comparação entre as estimativas dos parâmetros do modelo ajustado aos dados sem a observação 30 e as do modelo ajustado aos "dados com erro". Da Tabela 2.3, nota-se que a ocorrência do "erro" reduz consideravelmente a estimativa de ϕ, aumentando a estimativa da dispersão dos dados, de forma a acomodar a observação errada. As estimativas dos coeficientes também são seriamente afetadas ao ponto da covariável não ser considerada significativa. Na Figura 2.11 apresentamos uma comparação entre as curvas estimadas através dos dados corretos e através dos "dados com erro". Nota-se como a observação que não pertence aos conjunto de dados reais prejudica o ajuste do modelo.

Realizamos uma análise de resíduos para o modelo ajustado aos dados com erro, os dados disponíveis (Figura 2.12). Notamos que são os gráficos relacionados ao resíduo ponderado padronizado 2 que destacam mais enfaticamente a observação 30, tanto o gráfico dos resíduos contra os índices das observações (Figura 2.12c) quanto o gráfico normal de probabilidades com envelopes simulados (Figura 2.12f). Nota-se que, em nenhum dos gráficos que utilizam o resíduo ordinário padronizado, o caso 30 é destacado (Figura 2.12a,d) e que o gráfico do resíduo ponderado padronizado 1 contra os índices das observações destaca discretamente essa observação (Figura 2.12b). Considerando os três resíduos, nota-se que os gráficos normais de probabilidades com envelopes simulados revelam má qualidade do ajuste do modelo aos "dados com erro", uma vez que os resíduos não se encontram distribuídos de forma aleatória dentro dos envelopes (Figura 2.12d–f). Voltando à Tabela 2.3, notamos que a exclusão do caso 30 do conjunto de "dados com erro" conduz a alterações importantes no modelo estimado, de forma que as estimativas dos parâmetros passam a ser consideravelmente próximas às obtidas usando os dados corretos e a covariável passa a ser significativa.

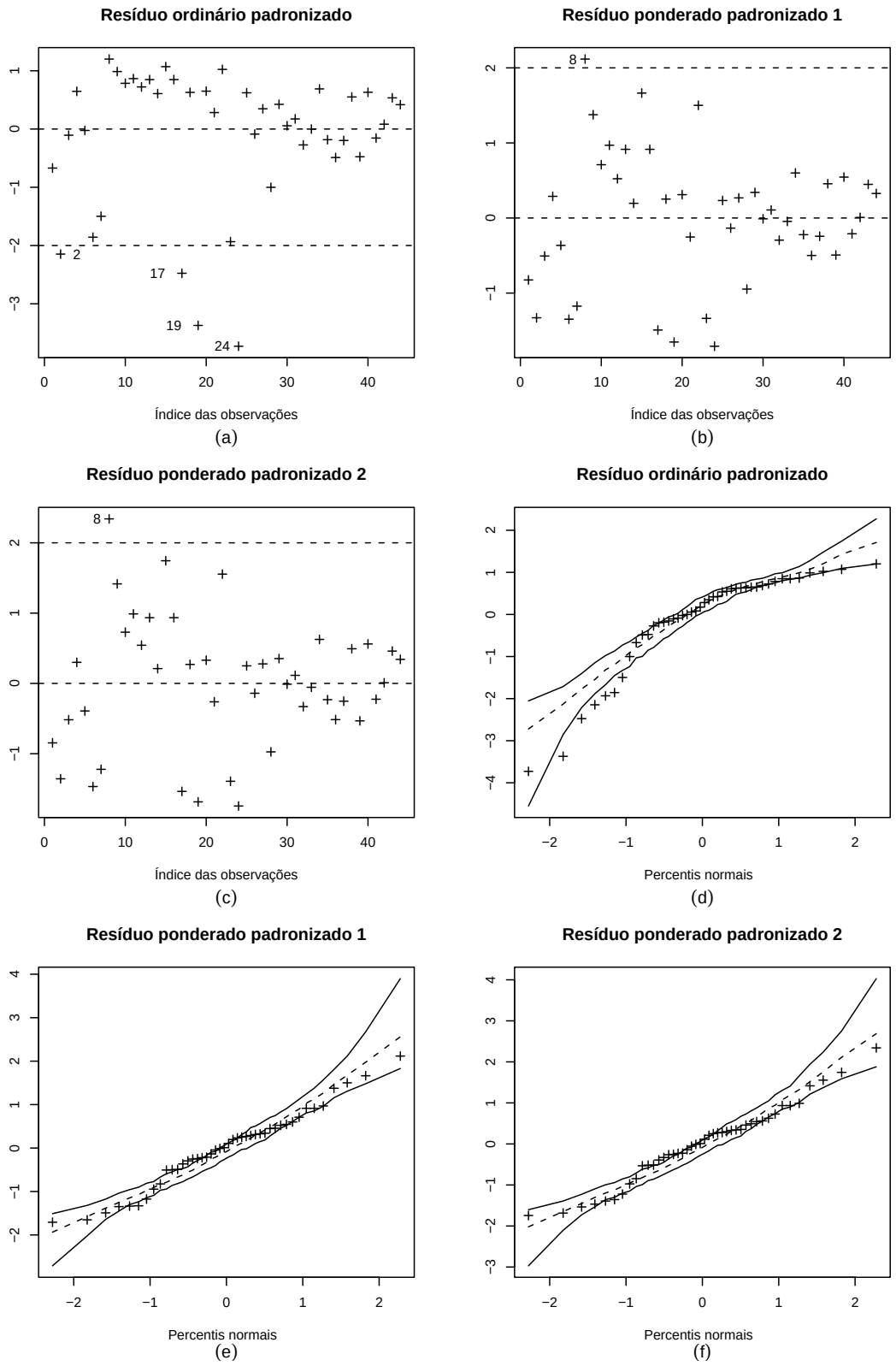

Figura 2.9. Gráficos de resíduos. Dados de habilidade de leitura.

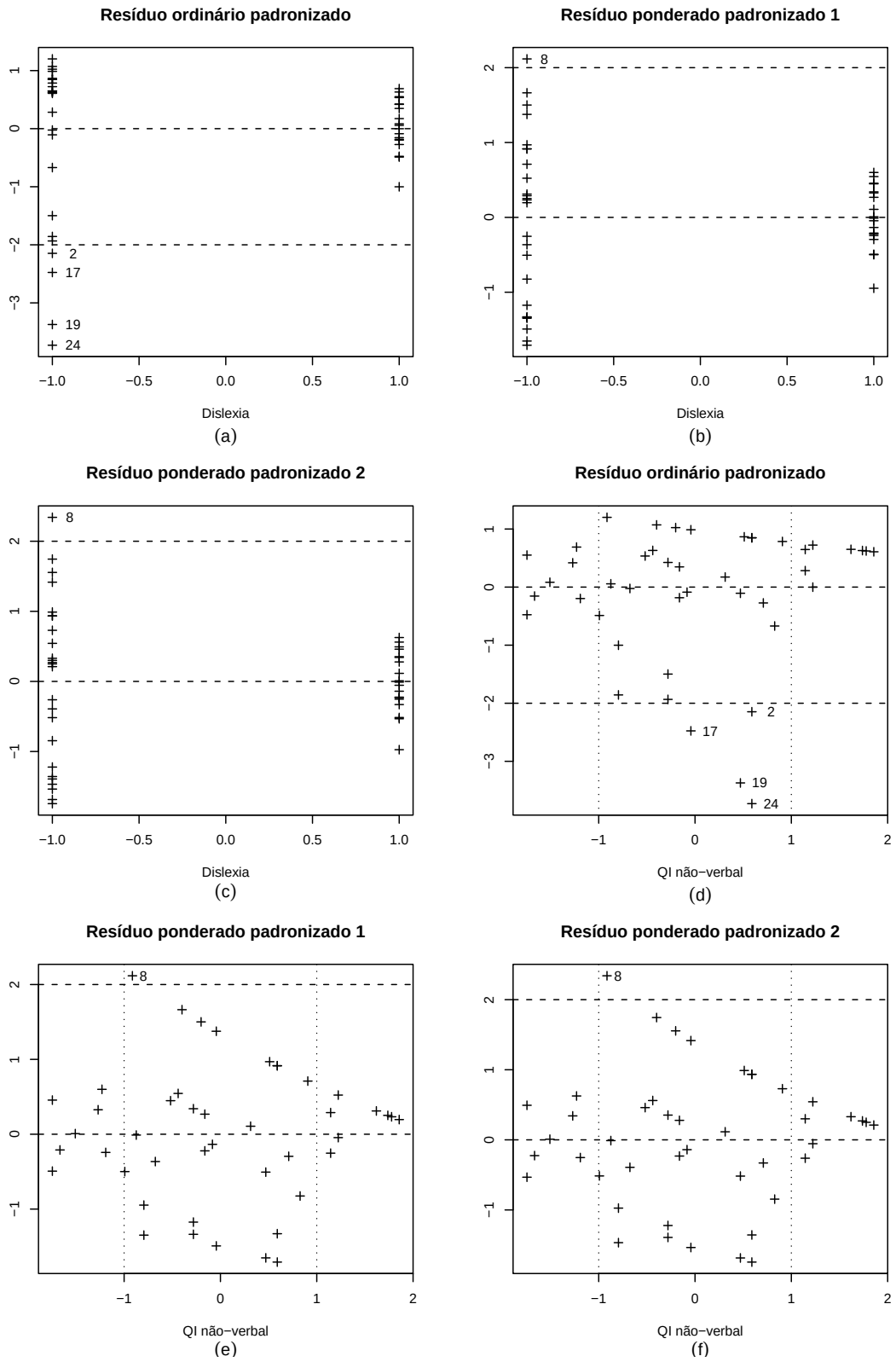

Figura 2.10. Gráficos de resíduos. Dados de habilidade de leitura.

Este exemplo mostra que o uso do resíduo ponderado padronizado 2 é mais indicado se o objetivo é a captação de casos aberrantes e que adicionalmente exercem uma influência desproporcional no ajuste do modelo. Quanto à avaliação da qualidade de ajuste do modelo, os três resíduos conduzem às mesmas conclusões.

Tabela 2.3. Resultados inferenciais do exemplo simulado.

Dados	Parâmetros	β_1	β_2	ϕ
Corretos	Estimativa	3.505	-3.407	330.166
	p–valor	0.0000	0.0000	
Com erro	Estimativa	1.633	-0.195	7.720
	p–valor	0.0000	0.4571	
Sem o caso 30	Estimativa	3.524	-3.437	330.284
	Variação%	115.80	1654.50	4177.93
	p–valor	0.0000	0.000	

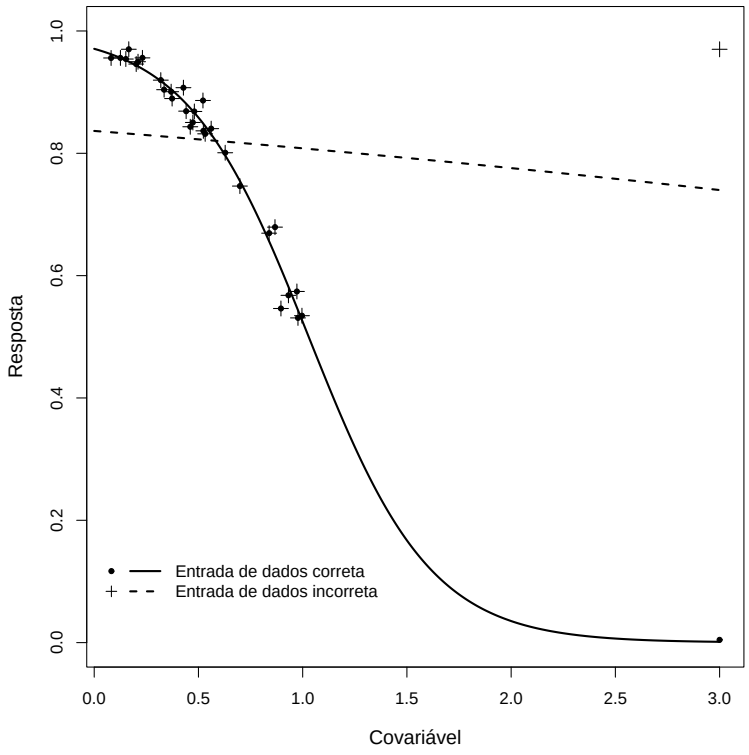

Figura 2.11. Curvas estimadas pela regressão beta. Comparação entre os modelos estimados pelos dados com erro e pelos dados correto.

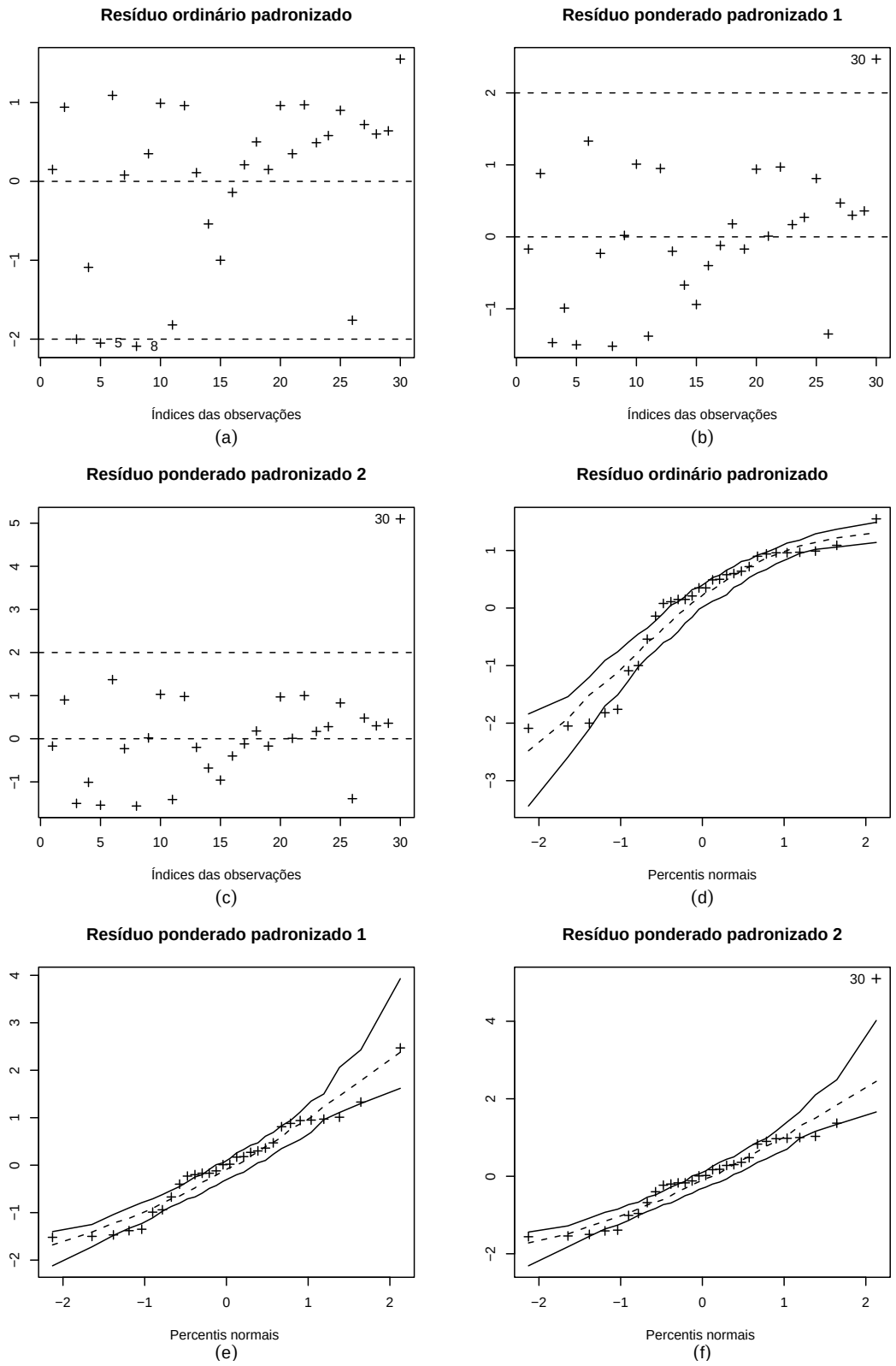

Figura 2.12. Gráficos de resíduos. Dados incorretos.

Capítulo 3

Análise de influência

3.1. Introdução

Um modelo ajustado é uma representação suavizada de aspectos essenciais dos dados. No entanto, aspectos importantes de um modelo podem ser dominados por uma única observação. Assim, na evolução dos métodos de diagnóstico uma etapa que se mostrou relevante foi a detecção de observações que exercem um efeito desproporcional no ajuste, podendo interferir inclusive em resultados inferenciais. Neste contexto, encontram-se a distância de Cook (1977), as matrizes de alavanca e as medidas de influência local.

A terminologia pontos de alavanca deve-se ao fato de tais pontos exercerem uma influência desproporcional no próprio valor ajustado. Em modelos lineares normais a medida de alavancagem está associada à matriz de projeção da solução de mínimos quadrados da regressão linear de y contra X, dada por $H = X(X'X)^{-1}X'$ (Hoaglin e Welsch, 1978). De uma forma geral, como pontuaram alguns autores, tais como Yoshizoe (1991), St. Laurent e Cook (1992), entre outros, uma medida de alavancagem deve refletir mais diretamente a influência de y_t no próprio valor ajustado. Sob este ponto de vista, Wei, Hu e Fung (1998) propõem uma matriz de alavanca generalizada que pode ser obtida para uma classe mais ampla que a dos estimadores de mínimos quadrados, tais como os estimadores de máxima verossimilhança, estimadores de método dos momentos e até estimadores obtidos sob o enfoque bayesiano.

A distância de Cook (1977) visa a medir o impacto de uma observação particular nas estimativas dos coeficientes da regressão a partir de sua exclusão do conjunto de dados. Apesar de ter sido desenvolvida originalmente para modelos normais lineares, aproximações para a distância de Cook têm sido utilizadas em diversas classes de modelos. Algumas referências são: Pregibon (1981), Cook e Weisberg (1982), Atkinson (1985), Cook, Peña e Weisberg (1988), Cordeiro e Paula (1992). A abordagem de deleção individual de casos, em que se baseia a distância de Cook, é um exemplo de uma análise de influência global.

Apesar de tipicamente a detecção de observações (casos) influentes se basear em deleção, esta é apenas uma das muitas maneiras de perturbar a formulação dos dados para acessar influência. Pequenas modificações dos valores de uma covariável, por exemplo, podem revelar estruturas relevantes nos dados que normalmente não seriam detectadas por deleção. Uma

análise de influência mais adequada deve considerar pequenas perturbações em diferentes elementos dos dados como, por exemplo, as covariáveis, o vetor de respostas ou a dispersão assumida. Este tipo de diagnóstico pode ser obtido utilizando o método de influência local desenvolvido por Cook (1986). A análise de influência local, que visa a avaliar conjuntamente o impacto das observações sob pequenas perturbações no modelo ou nos dados, representou um grande avanço para a análise de diagnóstico. Se modificações discretas na formulação inicial do modelo causam efeitos desproporcionais em determinados resultados, existem fortes indícios de falta de qualidade no ajuste ou violação das suposições iniciais, sugerindo a escolha de um modelo mais adequado aos dados. A proposta de influência local apresentada por Cook (1986) tem sido vastamente utilizada na modelagem de regressão. Por exemplo, Tsai e Wu (1992) investigam influência local em modelos auto-regressivos de primeira ordem e modelos heteroscedásticos, Paula (1996) em modelos próprios de dispersão, Galea, Paula e Bolfarine (1997), Liu (2000), Galea, Paula e Uribe–Opazo (2003) em modelos elípticos lineares.

Diante do exposto apresentamos para o modelo de regressão beta definido no Capítulo 1 uma medida de influência baseada na deleção de casos e medidas de influência local considerando diversos esquemas de perturbação. Adicionalmente realizamos uma comparação entre medidas de alavanca e influência local.

3.2. Distância de Cook

Para determinar a distância de Cook para o modelo de regressão beta apresentado em (1.2) consideraremos ϕ conhecido. No entanto, na prática usaremos a estimativa de máxima verossimilhança de ϕ obtida do processo iterativo conjunto para β e ϕ. Uma vez que supomos ϕ conhecido, temos que o logaritmo da função de verossimilhança para o modelo de regressão beta depende apenas do vetor de parâmetros desconhecido, β, ou seja,

$$\ell(\beta) = \sum_{t=1}^{n} \ell_t(\mu_t), \tag{3.1}$$

em que

$$\begin{aligned}\ell_t(\mu_t) &= \log\Gamma(\phi) - \log\Gamma(\mu_t\phi) - \log\Gamma((1-\mu_t)\phi) + (\mu_t\phi - 1)\log y_t \\ &\quad + \{(1-\mu_t)\phi - 1\}\log(1-y_t)\end{aligned}$$

e $\mu_t = g^{-1}(x_t^\top\beta)$.

No esquema de exclusão de observações uma medida de influência bastante geral motivada pela estatística da razão de verossimilhanças é dada por

$$LD_t = 2\left\{\ell_t(\widehat{\beta}) - \ell_t(\widehat{\beta}_{(-t)})\right\}.$$

Note que LD_t é calculada para cada observação, $t = 1, \ldots, n$, e que $\ell_t(\widehat{\beta})$ e $\ell_t(\widehat{\beta}_{(-t)})$ são, respectivamente, o logaritmo da função de verossimilhança avaliada na estimativa de máxima verossimilhança de β para os dados completos e o logaritmo da função de verossimilhança avaliada na estimativa de máxima verossimilhança de β quando a t-ésima observação é excluída do modelo. Essa medida foi apresentada por Cook e Weisberg (1982) como "likelihood displacement" (deslocamento pela verossimilhança).

Em geral, não é possível obter uma forma analítica para LD_t sendo usual utilizar a seguinte aproximação por série de Taylor em torno de $\widehat{\beta}$:

$$2\left\{\ell_t(\widehat{\beta}) - \ell_t(\beta)\right\} \approx (\beta - \widehat{\beta})^\top \left\{-\left(\frac{\partial^2 \ell(\beta)}{\partial\beta\partial\beta^\top}\right)\right\}(\beta - \widehat{\beta}).$$

Substituindo $-\left(\partial^2\ell(\beta)/\partial\beta\partial\beta^\top\right)$ por $K_{\beta\beta} = \mathrm{E}\left\{-\left(\partial^2\ell(\beta)/\partial\beta\partial\beta^\top\right)\right\}$ e β por $\widehat{\beta}_{(-t)}$ obtemos

$$LD_t \approx (\widehat{\beta} - \widehat{\beta}_{(-t)})^\top K_{\beta\beta}(\widehat{\beta} - \widehat{\beta}_{(-t)}). \tag{3.2}$$

O passo seguinte é obter $\widehat{\beta}_{(-t)}$, i.e., obter a estimativa de máxima verossimilhança de β quando a t-ésima observação é excluída do modelo. Segundo Pregibon (1981) isto corresponde a utilizar um método simples de perturbação que é uma generalização de um dos métodos usados por Welsh e Kuh (1977). De fato, a exclusão de pontos corresponde ao método de perturbação descrito a seguir.

Considere o logaritmo da função de verossimilhança em (3.1) sendo perturbado de forma que

$$\ell_\delta(\beta) = \sum_{t=1}^{n} \delta_t \ell_t(\mu_t), \tag{3.3}$$

em que $\delta_t = 0$ ou $\delta_t = 1$, $t = 1, \ldots, n$. Quando $\delta_t = 1$, $\forall t$, não há perturbação no modelo e $\delta_t = 0$ implica que a t-ésima observação foi excluída. O estimador de máxima verossimilhança de β será uma função de $(\delta_1, \ldots, \delta_n)$ e pode ser obtido maximizando-se (3.3). De fato, com base em (3.3) temos que a função escore e a matriz de informação de Fisher para β passam a ser, respectivamente, dadas por

$$U_\beta(\beta) = \phi X^\top \Lambda T(y^* - \mu^*)$$

e

$$K_{\beta\beta} = \phi X^\top W^{1/2} \Lambda W^{1/2} X,$$

em que $\Lambda = \mathrm{diag}\{\delta_1, \ldots, \delta_n\}$, os t-ésimos elementos de y^* e μ^* estão definidos em (1.5) e as matrizes T e W estão definidas em (1.6) e (1.9), respectivamente. Assim, utilizando o método Scoring de Fisher temos que o processo iterativo para β é dado por

$$\beta^{(m+1)} = \beta^{(m)} + \{X^\top [W^{(m)}]^{1/2} \Lambda [W^{(m)}]^{1/2} X\}^{-1} X^\top \Lambda T^{(m)}(y^* - \mu^{*(m)}).$$

Pregibon (1981) sugere começar o processo iterativo acima com o estimador de máxima verossimilhança de β e terminar após um passo. Essa aproximação de um passo gera $\widehat{\beta}_\delta$, que é dado por

$$\widehat{\beta}_\delta = \widehat{\beta} + \{X^\top \widehat{W}^{1/2} \Lambda \widehat{W}^{1/2} X\}^{-1} X^\top \Lambda \widehat{T}(y^* - \widehat{\mu}^*) \tag{3.4}$$

e que pode ser reescrito como

$$\widehat{\beta}_\delta = \{X^\top \widehat{W}^{1/2} \Lambda \widehat{W}^{1/2} X\}^{-1} X^\top \widehat{W}^{1/2} \Lambda \widehat{W}^{1/2} z,$$

em que

$$z = \widehat{\eta} + \widehat{W}^{-1} \widehat{T}(y^* - \widehat{\mu}^*),$$

com $\widehat{\eta} = (\widehat{\eta}_1, \ldots, \widehat{\eta}_n)^\top = X\widehat{\beta}$. Podemos interpretar $\widehat{\beta}_\delta$ como a solução de mínimos quadrados da regressão linear de $\widehat{W}^{1/2} z$ contra $\widehat{W}^{1/2} X$ com pesos Λ. A partir de (3.4) e considerando $\delta_t = 0$, $\delta_l = 1$, $\forall l \neq t$, o que implica na exclusão da t-ésima observação, segue que

$$\widehat{\beta} - \widehat{\beta}_{(-t)} \approx \frac{(X^\top \widehat{W} X)^{-1} x_t \, \widehat{w}_t^{1/2} r_t^*}{(1 - h_{tt}^*)},$$

em que r^* é o resíduo ponderado definido em (2.5) e h_{tt}^* é t–ésimo elemento diagonal principal da matriz H^* definida em (2.7).

Fazendo as devidas substituições em (3.2) obtém-se a distância de Cook aproximada para o modelo de regressão beta definido em (1.2), dada por

$$LD_t \approx \frac{h_{tt}^*}{v_t} \left(\frac{y_t^* - \widehat{\mu}_t^*}{1 - h_{tt}^*} \right)^2 = \frac{h_{tt}^*}{(1 - h_{tt}^*)} \frac{\phi r_t^{*2}}{(1 - h_{tt}^*)} = \frac{h_{tt}^*}{1 - h_{tt}^*} (r_t^{pp})^2, \tag{3.5}$$

em que r_t^{pp} é o resíduo ponderado padronizado 2 definido em 2.8.

Ressaltamos que Ferrari e Cribari–Neto (2004) propõem uma aproximação para a distância de Cook dada por $h_{tt}^* \, r_t^2 / [k(1 - h_{tt}^*)^2]$, em que r_t é o resíduo ordinário padronizado definido em (2.1) e k é o número de covariáveis do modelo da média. Em estudos realizados, mas não apresentados aqui, constatamos que a aproximação proposta por Ferrari e Cribari–Neto (2004), em geral, apresenta comportamento semelhante ao da distância de Cook definida em (3.5). No entanto, em cenários em que a estimativa de ϕ é pequena, digamos, menor que 20, as duas aproximações conduzem a conclusões diferentes, sendo que as conclusões relevantes foram obtidas a partir da distância de Cook definida em (3.5). Assim, adotamos a aproximação definida em (3.5) como a distância de Cook para o modelo de regressão beta.

3.3. Influência local

Para um certo conjunto de dados observados seja $\ell(\theta)$ o logaritmo da função de verossimilhança correspondente ao modelo postulado, em que θ é um vetor $s \times 1$ desconhecido.

Considere agora uma perturbação neste modelo postulado realizada através de um vetor δ, em geral $n \times 1$, o qual é restrito a algum sobconjunto aberto $\mathcal{D}$ de $I\!R^n$. Seja $\ell_\delta(\theta)$ o logaritmo da função de verossimilhança do modelo perturbado para um dado $\delta \in \mathcal{D}$. Geralmente $\ell_\delta(\theta)$ pode refletir qualquer esquema de perturbação bem definido e não se restringir apenas ao esquema de ponderação de casos.

Sejam $\widehat{\theta}$ e $\widehat{\theta}_\delta$ os estimadores de máxima verosimilhança de θ para o modelo postulado e para o modelo perturbado, respectivamente. O deslocamento pela verossimilhança ("likelihood displacement"), que neste caso mais geral é expresso por

$$LD_\delta = 2\left\{\ell(\widehat{\theta}) - \ell(\widehat{\theta}_\delta)\right\}, \tag{3.6}$$

pode ser utilizado como métrica para avaliar a influência sobre a estimativa de θ ao se variar δ através de $\mathcal{D}$. No entanto, avaliar o comportamento de LD_δ para todo $\delta \in \mathcal{D}$ pode ser inviável. Neste sentido, Cook (1986) propôs estudar o comportamento local de LD_δ ao redor do valor δ_0 de δ que representa a ausência de perturbação do modelo postulado, de tal forma que $LD_{\delta_0} = 0$. Uma vez que qualquer valor δ em uma vizinhança de δ_0 representa um incremento em LD_{δ_0}, este procedimento descreve a sensibilidade de $\ell(\widehat{\theta})$ com respeito a uma leve perturbação introduzida em $\ell(\theta)$. Para estudar esta sensibilidade Cook (1986) trabalha com curvaturas normais. O autor sugere avaliar a curvatura do gráfico $LD_{\delta_0+aI} \times a$, em que $a \in I\!R$ e I é uma direção de norma igual a um, ou seja, $\|I\| = 1$; esse gráfico é chamado de linha projetada. Cook (1986) sugere inspecionar a direção I_{max} que corresponde à linha projetada de maior curvatura C_{max}. Uma descrição mais detalhada do procedimento proposto por Cook (1986) está apresentada no Apêndice 3.A, Seção 3.A.1.

Segundo Cook (1986) o vetor I_{max} é a direção unitária de perturbação que implica na máxima mudança local em $LD_{\delta_0} = 0$. Isto corresponde a modificar os elementos mais influentes dos dados, os quais podem ser identificados pelos componentes de maior valor absoluto do vetor I_{max}. Como pode ser visto no Apêndice 3.A, Seção 3.A.1, Cook (1986) mostra que ao utilizarmos como métrica o deslocamento pela verossimilhança, a curvatura normal na direção de algum vetor I pode ser expressa por

$$C_I(\theta) = 2|I^\top \Delta^\top \ddot{\ell}^{-1} \Delta I|, \tag{3.7}$$

em que $\ddot{\ell} = \partial^2\ell(\widehat{\theta})/\partial\theta\partial\theta^\top$ e Δ é uma matriz $s \times n$ dada por $\Delta = \partial^2\ell_\delta(\theta)/\partial\theta\partial\delta^\top$, avaliada em $\theta = \widehat{\theta}$ e $\delta = \delta_0$. Desta forma C_{max} é o maior autovalor da matriz $-\boldsymbol{\Delta}^\top \ddot{\ell}^{-1} \boldsymbol{\Delta}$ e I_{max} é um correspondente autovetor de norma igual a um.

É possível também avaliar a influência local apenas para parte do vetor de parâmetros. Suponha que possamos particionar o vetor de parâmetros como $\theta = (\theta_1^\top, \theta_2^\top)^\top$, tal que

$$\ddot{\ell} = \begin{pmatrix} \ddot{\ell}_{\theta_1\theta_1} & \ddot{\ell}_{\theta_1\theta_2} \\ \ddot{\ell}_{\theta_2\theta_1} & \ddot{\ell}_{\theta_2\theta_2} \end{pmatrix},$$

em que $\ddot{\ell}_{\theta_1\theta_1} = \partial^2\ell(\theta)/\partial\theta_1\partial\theta_1^\top$, $\ddot{\ell}_{\theta_1\theta_2} = \partial^2\ell(\theta)/\partial\theta_1\partial\theta_2^\top$, $\ddot{\ell}_{\theta_2\theta_1} = \partial^2\ell(\theta)/\partial\theta_2\partial\theta_1^\top$ e $\ddot{\ell}_{\theta_2\theta_2} = \partial^2\ell(\theta)/\partial\theta_2\partial\theta_2^\top$. Segundo Cook (1986) se o interesse é calcular a influência local apenas para θ_1 o afastamento pela verossimilhança é dado por

$$LD_{\delta;\theta_1} = 2\left\{\ell(\widehat{\theta}_1, \widehat{\theta}_2) - \ell(\widehat{\theta}_{1\delta}, \mathcal{J}(\widehat{\theta}_{1\delta}))\right\},$$

em que $\mathcal{J}(\theta_{1\delta})$ é uma função que maximiza $\ell(\theta_1, \theta_2)$ para cada θ_1 fixo, $\widehat{\theta}_{1\delta}$ é obtido da partição $(\widehat{\theta}_{1\delta}, \widehat{\theta}_{2\delta})^\top$ e $\ell(\theta_{1\delta}, \mathcal{J}(\theta_{1\delta}))$ é o logaritmo da função de verossimilhança perfilada para θ_1. Neste caso, a curvatura normal na direção do vetor I é dada por

$$C_{I;\theta_1} = |I^\top \Delta^\top (\ddot{\ell}^{-1} - \ddot{\ell}_{22})\Delta I|,$$

em que

$$\ddot{\ell}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\ell}_{\theta_2\theta_2}^{-1} \end{pmatrix},$$

e $I_{max;\theta_1}$ neste caso é autovetor de norma igual a um correspondente ao maior autovalor da matriz $-\mathbf{\Delta}^\top(\ddot{\ell}^{\mathbf{-1}} - \ddot{\ell}_{\mathbf{22}})\mathbf{\Delta}$. O mesmo procedimento pode ser feito para avaliar a influência local apenas para θ_2.

Uma vez que o vetor I_{max} mostra como perturbar o modelo postulado (o modelo para $\delta = \delta_0$) no sentido de obter a maior mudança local no deslocamento pela verossimilhança, Cook (1986) propõe inspecionar I_{max} independentemente do tamanho de C_{max}, porque isto pode revelar observações que são **simultaneamente** influentes. No entanto, podem existir casos individualmente influentes e que não são revelados através da análise dos componentes de I_{max}. Neste sentido, Lesaffre e Verbeke (1998) sugerem avaliar a curvatura na direção do t-ésimo indivíduo, isto é, o vetor em que o t-ésimo componente assume o valor um e os demais o valor zero. Neste caso, a curvatura normal é dada por

$$C_t = 2|\Delta_t{}^\top \ddot{\ell}^{-1}\Delta_t|,$$

em que Δ_t é a t-ésima coluna da matriz Δ. Uma vez que C_t reflete a situação em que foi atribuído o maior valor (o valor total) possível à t-ésima coordenada de I, tal que $\|I\| = 1$, os autores denotam tal curvatura por influência local total do t-ésimo indivíduo. Também é possivel calcular a influência local total do t-ésimo indivíduo na estimação de parte do vetor θ. Por exemplo, se o interesse recai em θ_1 temos que

$$C_{t;\theta_1} = 2|\Delta_t{}^\top(\ddot{\ell}^{-1} - \ddot{\ell}_{22})\Delta_t|.$$

Lesaffre e Verbeke (1998) argumentam que os componentes de I_{max} e as medidas de influência C_t, $t = 1, \ldots, n$, contêm informações diferentes. Os autores mostram que podemos expressar C_t como

$$C_t = 2\sum_{p=1}^{s} \zeta_p \upsilon_{pt}^2, \tag{3.8}$$

em que ζ_p, $p = 1, \ldots, s$, são os autovalores diferentes de zero da matriz $-\mathbf{\Delta}^\top(\ddot{\ell})^{-\mathbf{1}}\mathbf{\Delta}$, tais que $\zeta_1 \geq \zeta_2 \geq \zeta_3 \geq \ldots \geq \zeta_s > \zeta_{s+1} = \ldots = \zeta_n = 0$ e $\upsilon_p = (\upsilon_{p1}, \ldots, \upsilon_{pn})^\top$, $p = 1, \ldots, s$, correspondentes autovetores ortogonais de norma igual a um, tal que uma base ortonormal da imagem de $-\mathbf{\Delta}^\top(\ddot{\ell})^{-\mathbf{1}}\mathbf{\Delta}$ é dada pela coluna da matriz

$$\begin{bmatrix} \upsilon_{11}, \ldots, \upsilon_{s1} \\ \vdots \qquad \vdots \\ \upsilon_{1n}, \ldots, \upsilon_{sn} \end{bmatrix}.$$

Assim, $C_{max}/2 = \zeta_1$ e $I_{max} = (\upsilon_{11}, \upsilon_{12}, \ldots, \upsilon_{1t}, \ldots, \upsilon_{1n})^\top$. A partir de (3.8) nota-se que podem existir casos individuais com uma curvatura total C_t grande sem ter o t-ésimo componente expressivo na direção I_{max} de máxima curvatura. Isto ocorrerá com indivíduos com um valor expressivo em qualquer autovetor υ_p, $p \neq 1$, correspondente a um autovalor relativamente grande, digamos ζ_2 ou ζ_3.

Assim, parece razoável considerar em uma análise de influência o estudo dos componentes de I_{max} e adicionalmente a medida de influência total referente ao t-ésimo indivíduo, C_t. Gráficos dos componentes de I_{max} contra os índices das observações podem sugerir quais observações são conjuntamente influentes, enquanto os gráficos de C_t contra os índices das observações destacam casos individualmente influentes. Lesaffre e Verbeke (1998) apresentam um exemplo em que a medida C_t identificou uma observação como influente, que se mostrou relevante no ajuste do modelo, que não foi identificada através do uso dos componentes de I_{max}. Por outro lado, os componentes de I_{max} podem ser utilizados para fornecer uma ferramenta de diagnóstico extra. Como veremos nas próximas seções, segundo Cook (1986), gráficos de I_{max} contra os índices das observações ou contra covariáveis podem revelar tendências importantes do conjunto de dados. Ressaltamos ainda que Lesaffre e Verbeke (1998) sugerem utilizar como ponto de corte nos gráficos de C_t, duas vezes o valor médio desta medida, ou seja, observações com C_t maior que $2\sum_{t=1}^{n} C_t/n$ podem ser classificadas como influentes.

3.4. Esquemas de perturbação

Inicialmente obteremos a expressão do inverso da matriz de informação observada e em seguida apresentamos quatro esquemas de perturbação para o modelo de regressão beta. Considere o modelo definido em (1.2) e sejam $s = k + 1$ e $\theta = (\beta^\top, \phi)^\top$. Temos que

$$\ddot{\ell} = \ddot{\ell}(\beta, \phi) = \begin{pmatrix} \ddot{\ell}_{\beta\beta} & \ddot{\ell}_{\beta\phi} \\ \ddot{\ell}_{\phi\beta} & \ddot{\ell}_{\phi\phi} \end{pmatrix}, \tag{3.9}$$

em que $\ddot{\ell}_{\beta\beta} = \phi X^\top Q X$, $\ddot{\ell}_{\beta\phi} = \ddot{\ell}_{\phi\beta}^\top = X^\top T f$, e $\ddot{\ell}_{\phi\phi} = \mathrm{tr}(D)$, sendo que $Q = \mathrm{diag}\{q_1, \ldots, q_n\}$, com

$$q_t = \left\{ \phi[\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)] + (y_t^* - \mu_t^*)\frac{g''(\mu_t)}{g'(\mu_t)} \right\} \frac{1}{\{g'(\mu_t)\}^2} \tag{3.10}$$

e $f = (f_1, \ldots, f_n)^\top$, com

$$f_t = \{c_t - (y_t^* - \mu_t^*)\}. \tag{3.11}$$

Os componentes de y^* e μ^* estão definidos em (1.5), as matrizes T e D e os componentes c_t estão definidos em (1.6), (1.13) e (1.12), respectivamente. A partir de (3.9) e usando a fórmula do inverso de matriz particionada (Rao, 1973, p. 33) obtemos

$$\ddot{\ell}^{-1} = \begin{pmatrix} \ddot{\ell}^{\beta\beta} & \ddot{\ell}^{\beta\phi} \\ \ddot{\ell}^{\phi\beta} & \ddot{\ell}^{\phi\phi} \end{pmatrix},$$

em que

$$\ddot{\ell}^{\beta\beta} = \frac{1}{\phi}(X^\top QX)^{-1}\left\{\mathcal{I}_k + \frac{X^\top Tff^\top T^\top X(X^\top QX)^{-1}}{\varrho\phi}\right\},$$

com

$$\varrho = \mathrm{tr}(D) - \phi^{-1}f^\top T^\top X(X^\top QX)^{-1}X^\top Tf, \tag{3.12}$$

$$\ddot{\ell}^{\beta\phi} = (\ddot{\ell}^{\phi\beta})^\top = -\frac{1}{\varrho\phi}(X^\top QX)^{-1}X^\top Tf,$$

$\ddot{\ell}^{\phi\phi} = \varrho^{-1}$ e $\mathcal{I}_k$ é a matriz identidade de ordem k. Ressaltamos que nas Seções 3.4.1–3.4.4 as quantidades assinaladas com "$\widehat{\ }$" são avaliadas em $(\widehat{\beta}^\top, \widehat{\phi})^\top$.

3.4.1. Ponderação de casos

Considere o esquema de perturbação

$$\ell_\delta(\beta, \phi) = \sum_{t=1}^{n} \delta_t \ell_t(\mu_t, \phi),$$

com $0 \leq \delta_t \leq 1$. Para o método de influência local discutido anteriormente temos que $\delta_0 = (1, 1, \ldots, 1)^\top$ e $\Delta_t = \partial\ell_t(\widehat{\theta})/\partial\theta$, $t = 1, \ldots, n$. Assim, da Seção 3.A.2.1 do Apêndice 3.A segue que

$$\Delta = \begin{pmatrix} \widehat{\phi}X^\top\widehat{T}\mathcal{E} \\ \widehat{u}^\top \end{pmatrix}, \tag{3.13}$$

em que

$$\mathcal{E} = \mathrm{diag}\{(y_1^* - \widehat{\mu}_1^*), \ldots, (y_n^* - \widehat{\mu}_n^*)\} \tag{3.14}$$

e $\widehat{u}^\top = (\widehat{u}_1, \ldots, \widehat{u}_n)$ sendo que T e u_t estão definidos em (1.6) e (1.7), respectivamente.

A ponderação de casos tem sido o esquema de perturbação mais utilizado para análise de influência e pode ser interpretado como uma perturbação na variância do t–ésimo caso, em especial para modelos normais lineares (Thomas e Cook, 1989). Cook (1986), baseado neste tipo de perturbação, utiliza dois exemplos de regressão linear simples para ilustrar o comportamento de I_{max} na presença de heteroscedasticidade. O autor conclui que os gráficos

de I_{max} versus valores de covariáveis evidenciam a heteroscedasticidade latente dos dados auxiliando a identificar que covariáveis devem ser usadas para a modelagem da variância.

3.4.2. Perturbação da variável resposta

Considere um esquema aditivo de perturbação da resposta em que $y = (y_1, \ldots, y_n)^\top$ é alterado através da adição de um vetor δ de pequenas perturbações. Em situações em que cada y_t apresenta uma variância diferente é comum utilizar um fator de escala para padronizar os componentes de δ, por exemplo, a estimativa do desvio padrão de y_t, de forma que

$$y_t(\delta) = y_t + \delta_t s(y_t), \tag{3.15}$$

em que $s(y_t) = \sqrt{\{\widehat{\mu}_t(1-\widehat{\mu}_t)\}/(1+\widehat{\phi})\}}$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e da Seção 3.A.2.2 do Apêndice 3.A segue que

$$\Delta = \begin{pmatrix} \widehat{\phi} X^\top \widehat{T} M S_y \\ \widehat{b}^\top S_y \end{pmatrix}, \tag{3.16}$$

em que $M = \text{diag}\{m_1, \ldots, m_n\}$ com

$$m_t = \frac{1}{\{y_t(1-y_t)\}}, \tag{3.17}$$

$S_y = \text{diag}\{s(y_1), \ldots, s(y_n)\}$ e $\widehat{b} = (\widehat{b}_1, \ldots, \widehat{b}_n)^\top$ com

$$b_t = -\frac{(y_t - \mu_t)}{\{y_t(1-y_t)\}}. \tag{3.18}$$

Pertubações aditivas na variável resposta estão fortemente relacionadas com o conceito de alavanca que apresentaremos na Seção 3.5. De fato, ao impor uma mudança aditiva na resposta a medida de influência local resultante (quando $\delta_t \downarrow 0$), I_{max} ou C_t pode ser utilizada para identificar observações que exercem forte influência no próprio valor ajustado. No caso normal linear, por exemplo, Schawarzmann (1991) mostra que os vetores I_{max} e $r = y - \widehat{y}$ são proporcionais e conduzem às mesmas conclusões de diagnóstico. Emerson, Hoaglin e Kempthorne (1984) utilizam este tipo de perturbação em um estudo sobre alavanca em análise de tabelas de contigência 2×2. Na Seção 3.5 apresentaremos uma comparação formal entre medidas de influência local baseadas na perturbação aditiva da resposta e alavanca para o modelo de regressão beta.

3.4.3. Perturbação individual de covariáveis

Thomas e Cook (1990) sugerem modificar a p–ésima coluna da matriz X, x_p, $p = 2, \ldots, k$, adicionando um vetor δ de pequenas perturbações ponderado por um fator de escala s_{x_p}, desvio padrão da coluna modificada, de forma que

$$x_{tp}(\delta) = x_{tp} + \delta_t s_{x_p}. \tag{3.19}$$

Neste caso, por exemplo, se $p \neq 2$ ou $p \neq k$

$$\eta_t(\delta) = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_p(x_{tp} + \delta_t s_{x_p}) + \ldots + \beta_k x_{tk} \tag{3.20}$$

e $\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e da Seção 3.A.2.3 do Apêndice 3.A, segue que

$$\Delta = \begin{pmatrix} -\widehat{\phi} s_{x_p}(\widehat{\beta}_p X^\top \widehat{Q} - P\widehat{T}\mathcal{E}) \\ -\widehat{\beta}_p s_{x_p} \widehat{f}^\top \widehat{T} \end{pmatrix}, \tag{3.21}$$

em que P é uma matriz $k \times n$ de zeros exceto a p-ésima linha, que é composta por uns e os componentes das matrizez diagonais Q e $\mathcal{E}$ e do vetor f estão definidos em (3.10), (3.14) e (3.11), respectivamente.

A utilização do fator de escala $s(x_p)$ permite comparar os valores de C_{max} obtidos ao perturbar separadamente cada covariável e assim estabelecer, por exemplo, quais covariáveis causam maior mudança local no deslocamento pela verossimilhança sob a adição de pequenas modificações em seus valores originais. Ou seja, através deste esquema de perturbação é possível acessar a influência individual de cada covariável no processo de estimação do modelo. No entanto, este tipo de perturbação faz sentido apenas se a covariável é medida de forma contínua.

3.4.4. Perturbação no parâmetro de precisão

Modificaremos a precisão do modelo definido em (1.2) de forma que no modelo perturbado a precisão não seja constante ao longo das observações, ou seja,

$$\phi_t(\delta) = \phi/\delta_t, \tag{3.22}$$

$t = 1, \ldots, n$. Neste caso, $\delta_0 = (1, \ldots, 1)^\top$ e da Seção 3.A.2.4 do Apêndice 3.A, segue que

$$\Delta = \begin{pmatrix} \widehat{\phi} X^\top \widehat{T} \widehat{F} \\ \widehat{\phi} \widehat{d}^\top - \widehat{u}^\top \end{pmatrix}, \tag{3.23}$$

sendo que os componentes da matriz diagonal F e dos vetores u e d estão definidos em (3.11), (1.7) e (1.13), respectivamente.

A curvatura máxima obtida a partir deste esquema de perturbação pode ser comparada com as demais curvaturas dos outros esquemas, ponderação de casos, variável resposta e covariáveis, no sentido de avaliar quão sensível os dados se mostram em relação a uma perturbação heteroscedástica. Adicionalmente, a análise de I_{max} e C_t pode revelar quais casos contribuem mais para esta possível sensibilidade.

3.5. Comparação entre influência local e matrizes de alavanca

Para estender o conceito de alavanca, originalmente definido no modelo de regressão normal linear, para modelos mais gerais, Wei, Hu e Fung (1998) buscaram captar o sentido essencial deste termo em Estatística. Baseados no ponto de vista de que uma medida de alavancagem deve refletir diretamente a influência de y_t no próprio valor ajustado, os autores propõem a matriz de alavanca generalizada ("generalized leverage") para estimadores de θ, dada por

$$\mathrm{GL}_{(\theta)} = \frac{\partial \widehat{y}}{\partial y^\top},$$

em que $\widehat{y}$ é um estimador de y. A proposta de Wei, Hu e Fung (1998) é a mesma da matriz de alavanca jacobiana ("Jacobian leverage") desenvolvida por St. Laurent e Cook (1992) para modelos normais não lineares.

Os autores mostram que a matriz de alavanca generalizada para $\widehat{\theta}$, estimador de máxima verossimilhança de θ, tal que $\mathrm{E}(y) = \mu = \mu(\theta)$, é dada por

$$\mathrm{GL}_\theta = D_\theta \left(-\frac{\partial^2 \ell}{\partial\theta\partial\theta^\top} \right)^{-1} \frac{\partial^2 \ell}{\partial\theta\partial y^\top}$$

avaliada em $\widehat{\theta}$,em que $D_\theta = \partial\mu/\partial\theta^\top$.

Utilizando este resultado Ferrari e Cribari–Neto (2004) obtêm no modelo de regressão beta a matriz de alavanca generalizada para $\widehat{\beta}$ e $\widehat{\phi}$ dada por

$$\mathrm{GL}_{\beta,\phi} = \mathrm{GL}_\beta + \frac{1}{\varrho\phi} TX(X^\top QX)^{-1}X^\top Tf \left(f^\top TX(X^\top QX)^{-1}X^\top TM - b^\top \right),$$

em que

$$\mathrm{GL}_\beta = TX(X^\top QX)^{-1}X^\top TM,$$

em que ϱ está definido em (3.12) e os componentes das matrizes T, Q e M e dos vetores f e b estão definidos em (1.6), (3.10), (3.17), (3.11) e (3.18), respectivamente. Em geral, nas análises gráficas são utilizados os elementos da diagonal principal da matriz descrita acima. Denotaremos esses elementos apenas por alavanca. A partir das expressões acima é possível

estabelecer uma relação entre influência local e a alavanca generalizada. Sob perturbação aditiva da resposta (ver Secção 3.4.2) temos que

$$-2\Delta^\top \ddot{\ell}^{-1}\Delta\Big|_{(\theta=\widehat{\theta},\delta=\delta_0)} = 2S_y\{\phi M\mathrm{GL}_{\beta,\phi} + \frac{1}{\varrho}b[b^\top - f^\top \mathrm{GL}_\beta]\}S_y\Big|_{\theta=\widehat{\theta}},$$

o que demonstra uma forte conexão entre as medidas de influência local baseadas no esquema de perturbação da variável resposta e os elementos das matrizes $\mathrm{GL}_{\beta,\phi}$ e GL_β. Assim, parece interessante realizar uma comparação entre os diagnósticos obtidos através da matriz de alavanca e da influência local.

3.6. Aplicações

As aplicações a seguir têm como objetivo utilizar a análise de influência para complementar as análises de resíduos realizadas para os modelos de regressão beta ajustados aos dados de gastos com alimentação (Griffiths et al. 1993) e aos dados de habilidade de leitura (Smithson e Verkuilen 2006) apresentadas na Seção 2.7. Em seguida, realizaremos uma análise de diagnóstico, incluindo análise de influência, para modelos de regressão beta ajustados a três novos conjuntos de dados. Ressaltamos que para todos os conjuntos de dados ajustamos o modelo de regressão beta (1.2) considerando as funções de ligação logito e probito e então realizamos a análise de diagnóstico. Não constatamos diferenças importantes nos ajustes quanto ao uso das duas funções de ligação. Assim, apresentaremos aqui as análises de diagnóstico considerando apenas os modelos de regressão beta que utilizam a função de ligação logito.

Em quatro das aplicações construímos os gráficos de influência local considerando os quatro esquemas de perturbação apresentados nas seções anteriores. Na quarta aplicação não é considerado o esquema de perturbação da covariável, uma vez que as covariáveis são categorizadas. Na maioria das aplicações seguimos a mesma rotina. Inicialmente, apresentamos uma figura com os gráficos dos valores de I_{max} contra os índices das observações. Nessa figura a primeira coluna se refere à influência local avaliada para todo o vetor de parâmetros $\theta = (\beta^\top, \phi)^\top$, $I_{max;\theta}$. A segunda coluna se refere à influência local apenas para o vetor β, $I_{max;\beta}$, e a terceira coluna, à influência local apenas para ϕ, $I_{max;\phi}$. Quanto à composição das linhas, consideramos na primeira linha o esquema de ponderação de casos, na segunda linha o esquema de perturbação da variável resposta, na terceira linha o esquema de perturbação da covariável (caso seja possível) e na última linha o esquema de perturbação do parâmetro de precisão ϕ. Em seguida, apresentamos uma figura com os gráficos de influência local total C_t contra os índices das observações, sendo que a primeira coluna se refere a $C_{t;\theta}$, influência local total para todo o vetor de parâmetros, a segunda coluna se refere à influência local

total apenas para o vetor β, $C_{t;\beta}$, e a terceira coluna à influência local total apenas para ϕ, $C_{t;\phi}$. As disposições das linhas dizem respeito aos tipos de esquemas de perturbação. O objetivo dos gráficos de I_{max} é avaliar quais observações são conjuntamente influentes, enquanto o objetivo dos gráficos de C_t é identificar observações individualmente influentes, que eventualmente não tenham sido destacadas pelos gráficos de I_{max}.

Adicionalmente, seguimos a proposta de Cook (1986) de utilizar os gráficos dos valores de I_{max}, considerando o esquema de ponderação de casos, contra os valores das covariáveis ou contra os índices das observações, para captar possíveis tendências nos dados. Temos interesse especial em identificar se a dispersão dos dados não é constante ao longo das observações e quais covariáveis devem ser utilizadas na modelagem do parâmetro ϕ.

3.6.1. Aplicação I: dados de gastos com alimentação

Esta aplicação complementa a aplicação I do capítulo anterior, em que a variável resposta, proporção da renda familiar gasta com alimentação, é modelada como função da renda total, x_2, e do número de pessoas no domicílio, x_3, através do modelo de regressão beta. Na Tabela 2.1 (Seção 2.7.2) apresentamos os resultados inferenciais referentes a esse modelo. Ressaltamos que a estimativa de ϕ é próxima de 35 e a variável resposta encontra-se no intervalo $(0.108, 0.562)$ com mediana igual a 0.261. No Capítulo 2 havíamos argumentado que os gráficos normais de probabilidades apresentavam uma leve tendência dos resíduos entre -1 e 1 estarem acima ou próximos da banda superior do envelope.

Dando continuidade à análise de diagnóstico construímos os gráficos da distância de Cook contra os valores ajustados e de alavanca generalizada contra os valores preditos (Figura 3.1a,b) e os gráficos de influência local. Na Figura 3.2 estão apresentados os gráficos dos valores de I_{max} contra os índices das observações e na Figura 3.3 apresentamos os gráficos de influência local total C_t contra os índices das observações. No esquema de perturbação individual de uma covariável utilizamos a covariável renda total. Verificamos a partir da Figura 3.1 que a distância de Cook destaca os casos 11 e 30, enquanto o gráfico de alavanca destaca os casos 11 e 38. Já a análise de influência local, a partir de uma visão geral de todos os esquemas de perturbação, destaca como conjuntamente influentes as observações 5, 11 e 20 (Figura 3.2).

Considerando o esquema de ponderação de casos e o esquema de perturbação do parâmetro de precisão, são destacados como conjuntamente influentes para a estimativa de β os casos 5, 11, 20, 25, 30, 37 e 38. Quanto à análise da influência individual de cada observação, Figura 3.3, notamos que os casos 11 e 38 são destacados como individualmente influentes para a estimativa de β, em especial pelo esquema de perturbação da variável resposta (Figura 3.3e), indicando que estes casos são pontos de alavanca, destacados pelo gráfico de alavanca

generalizada, que se confirmam como influentes. Já a observação 20 é consideravelmente influente para a estimativa ϕ a partir de todos os esquemas de perturbação investigados. A Figura 3.3b revela a influência individual do caso 30 na estimação dos coeficientes da regressão. Ainda são destacadas como individualmente influentes as observações 4, 5 e 37.

Note que o esquema de perturbação da covariável renda total foi aquele que apresentou os menores valores de curvatura na direção I_t, $t = 1, \ldots, n$ (Figura 3.3g–i). Além disto, os gráficos de I_{max} contra os índices das observações baseados nesse esquema de perturbação não destacam enfaticamente qualquer observação (Figura 3.2g–i). Destacamos ainda a semelhança entre os gráficos de influência local total referentes aos esquemas de ponderação de casos e pertubação da precisão (Figura 3.3a–c; 3.3j–k).

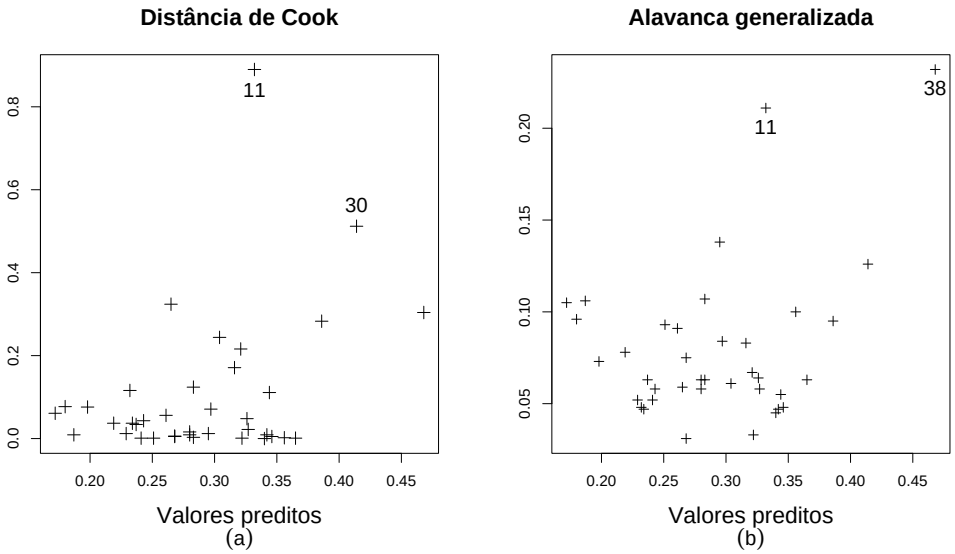

Figura 3.1. Gráficos de distância de Cook e alavanca generalizada. Dados de gastos com alimentação.

Excluímos individualmente os casos 4, 5, 11, 20, 30, 37 e 38, conjuntamente os casos 5, 11 e 20 e conjuntamente os casos 5, 11, 20, 25, 30, 37 e 38. Ressaltamos que as exclusões individuais dos casos 4 e 37 não se mostraram relevantes. As variações percentuais nas estimativas dos parâmetros devido às exclusões mais expressivas encontram-se na Tabela 3.1. Ao observar essa tabela comprovam-se as assertivas anteriores. O caso 20 influencia mais fortemente a estimativa de ϕ, enquanto o caso 11 influencia mais fortemente as estimativas do intercepto e do coeficiente da covariável que representa o número de pessoas no domicílio. É notável a

influência conjunta das observações 5, 11 e 20 no processo de estimação. Adicionalmente, a exclusão conjunta dos casos 5, 11, 20, 25, 30, 37 e 38 afeta consideravelmente a estimativa de ϕ, mas o impacto nas estimativas dos coeficientes do modelo, quando comparado ao impacto referente à exclusão dos casos 5, 11 e 20, é mais expressivo, de fato, apenas para o coeficiente da covariável renda total. Nenhuma das exclusões alterou a significância dos coeficientes do modelo, nem os sinais das estimativas.

Tabela 3.1. Variação percentual das estimativas dos parâmetros retirando observações influentes. Dados de gastos com alimentação.

Casos	β_1	β_2	β_3	ϕ
5	6.76	3.33	8.70	9.72
11	22.56	7.86	24.16	9.86
20	6.89	7.97	2.35	17.39
30	13.17	13.96	8.57	7.46
37	0.48	4.63	12.18	4.68
38	2.42	8.63	13.44	0.51
5, 11, 20	42.74	21.82	42.20	70.99
5, 11, 20, 25, 30, 37, 38	45.50	47.61	15.49	137.10

Voltando à tendência relatada no início da aplicação, seguindo sugestão de Cook (1986), consideramos gráficos dos valores de I_{max} obtidos através do esquema de ponderação de casos contra os valores das covariáveis renda total e número de pessoas no domicílio, Figura 3.4a,b, respectivamente. A análise dessa figura evidencia que a dispersão de I_{max} é diferente para cada valor da covariável número de pessoas no domicílio. De fato, domicílios com 1, 2 e 6 pessoas apresentam componentes de I_{max} bem concentrados. Contudo, em domicílios com 3 pessoas já ocorre um caso que se destaca dos demais e quando o número de pessoas é igual a 4, 5 e 7 a dispersão de I_{max} aumenta consideravelmente. Ao que parece, a dispersão é influenciada pela covariável número de pessoas no domicílio. Já o gráfico de I_{max} contra os valores da covariável renda total não evidencia uma relação entre a dispersão de I_{max} e esta covariável. Adicionalmente, a Figura 3.4a,b confirma a influência conjunta dos casos 5, 11 e 20. Excluímos os casos 5, 11 e 20 e refizemos os dois gráficos acima (Figura 3.5a,b). É possível notar, no modelo estimado sem essas observações, uma relação entre a dispersão de I_{max} e a covariável renda total de forma que à medida que aumenta o valor da covariável a dispersão diminui. No Capítulo 5 ajustaremos a esse conjunto de dados um modelo em que a média e a dispersão são modeladas simultaneamente, através das covariáveis número de pessoas no domicílio e renda total.

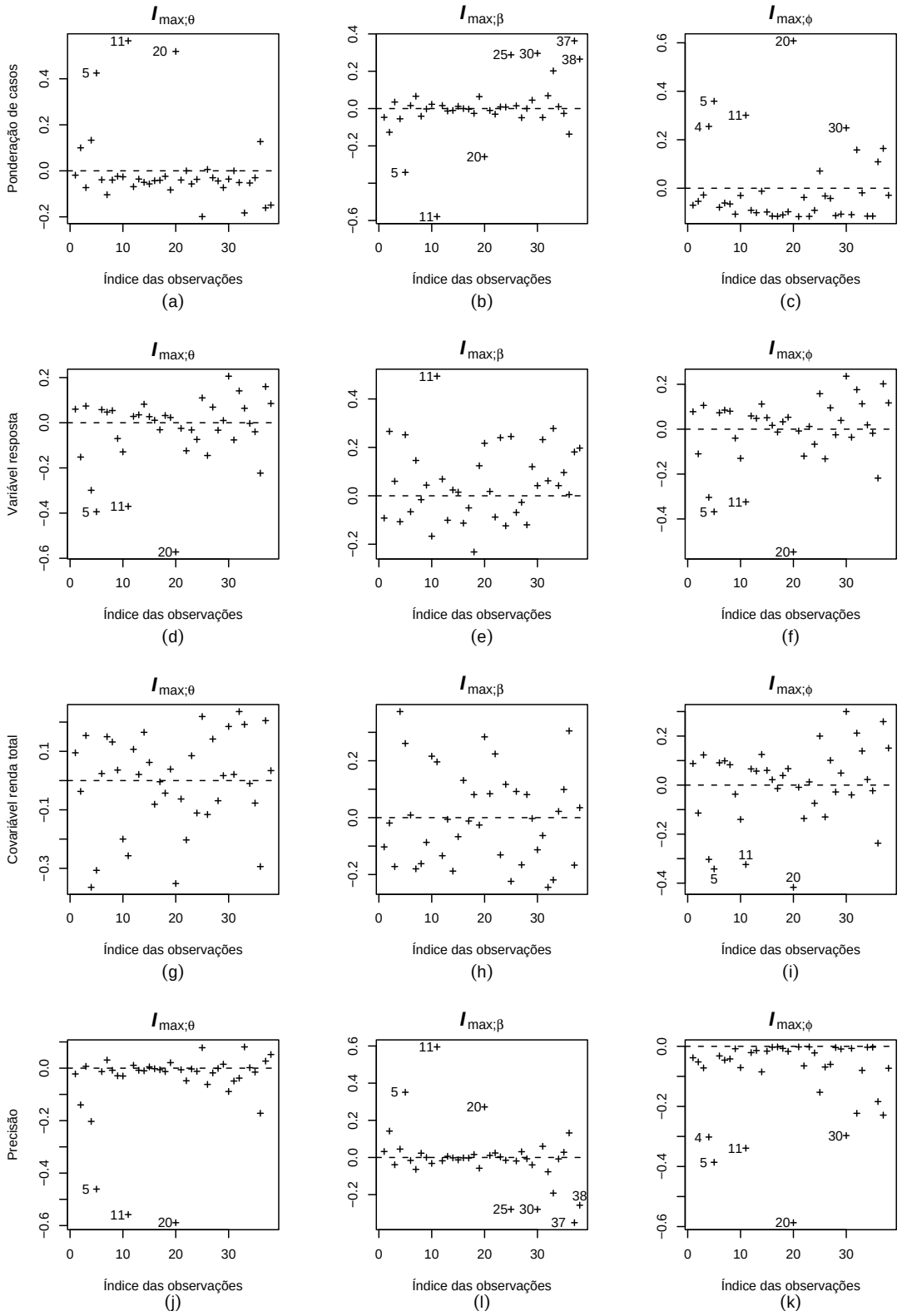

Figura 3.2. Gráficos de influência local. Dados de gastos com alimentação.

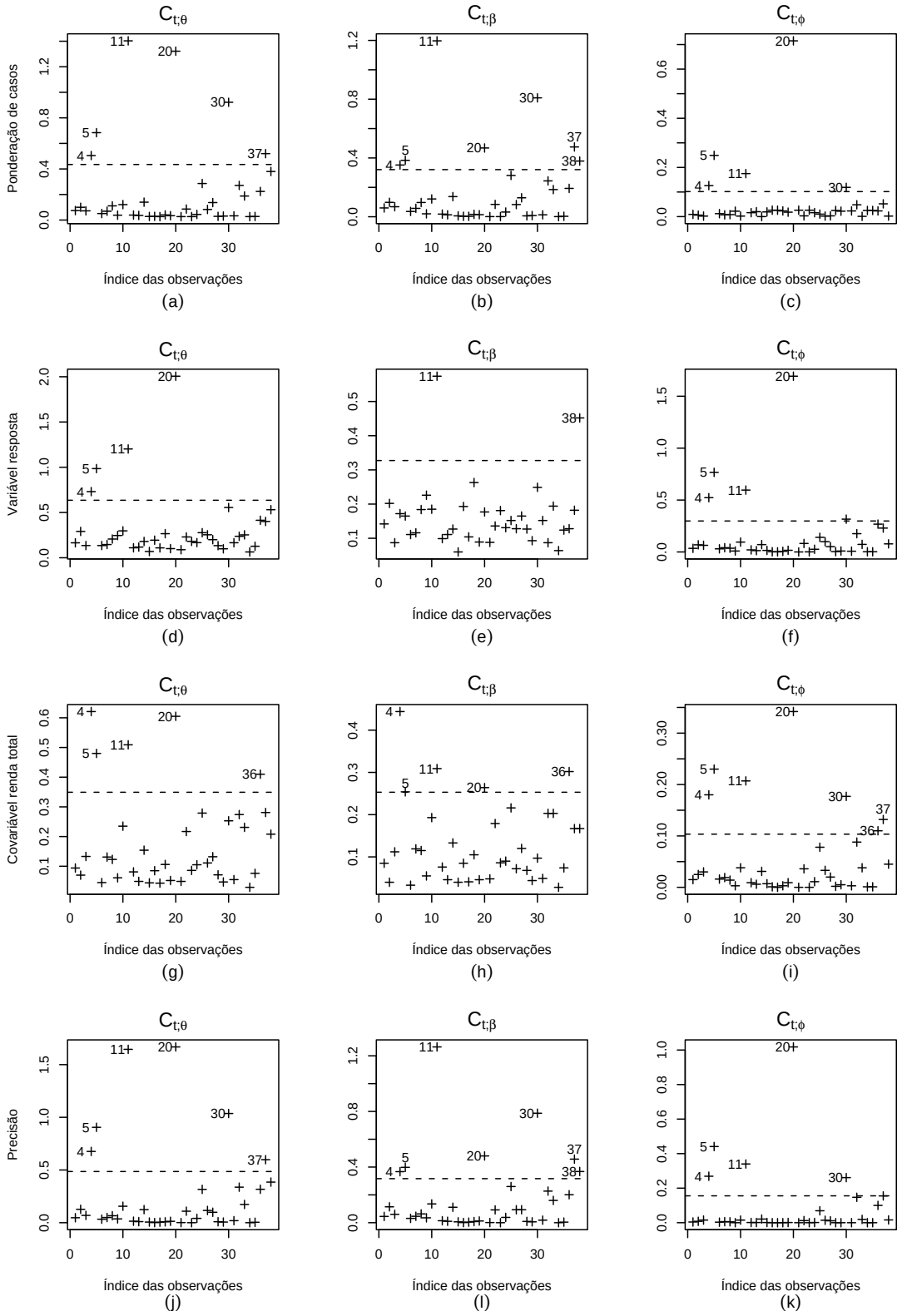

Figura 3.3. Gráficos de influência local total. Dados de gastos com alimentação.

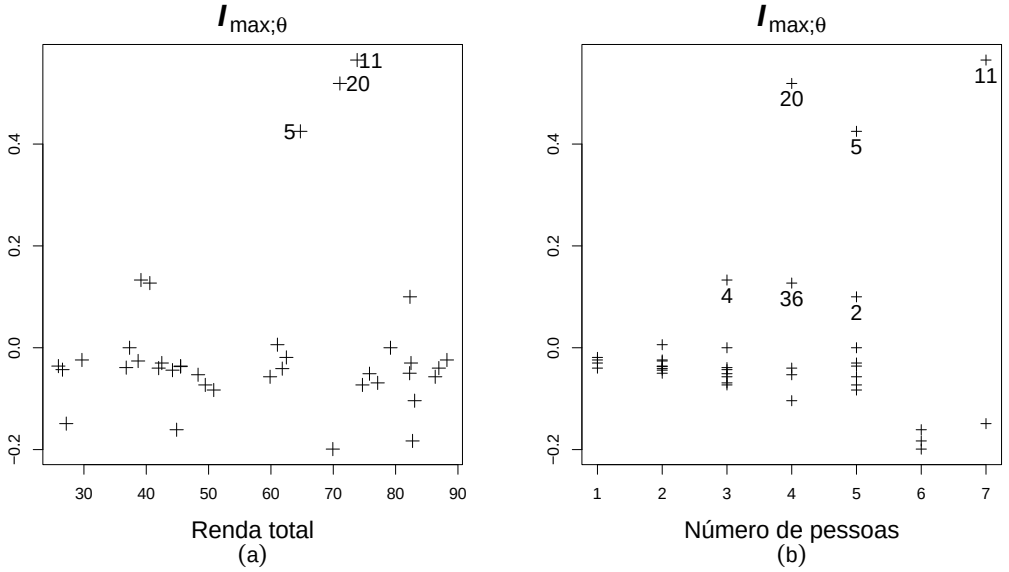

Figura 3.4. Gráficos de influência. Dados de gastos com alimentação.

Investigamos as observações 5, 11 e 20 e chegamos a conclusões interessantes. As três observações apresentam uma característica comum, a proporção de renda gasta com alimentação é pequena levando em conta a quantidade de pessoas no domicílio. As observações 5, 11 e 20 correspondem a domicílios com 5, 4 e 7 pessoas, respectivamente. A observação 11, por exemplo, representa um domicílio com sete pessoas, com uma renda familiar 26% maior que a média da renda familiar de todos os 38 domicílios e que gasta apenas 20% com alimentação. Esperava-se que quanto maior o número de pessoas no domicílio, maior a proporção da renda gasta com alimentação, e isto não ocorre com os casos 5, 11 e 20. Notamos que ao excluírmos esses casos a estimativa do coeficiente do número de pessoas no domícilio passa de 0.1185 para 0.1685. É interessante relembrar que no Capítulo 2 apenas o gráfico do resíduo ponderado padronizado 2 versus os índices da observações destacou conjuntamente os casos 5, 11 e 20.

3.6.2. Aplicação II: dados de habilidade de leitura

Esta aplicação complementará a análise de resíduos realizada no modelo de regressão beta logito ajustado aos dados de habilidade de leitura (ver Seção 2.7.3.). Os dados são provinientes do exemplo 3 apresentado em Smithson e Verkuilen (2006). A variável resposta é o escore transformado em um teste de habilidade de leitura aplicado a um grupo de 44 crianças.

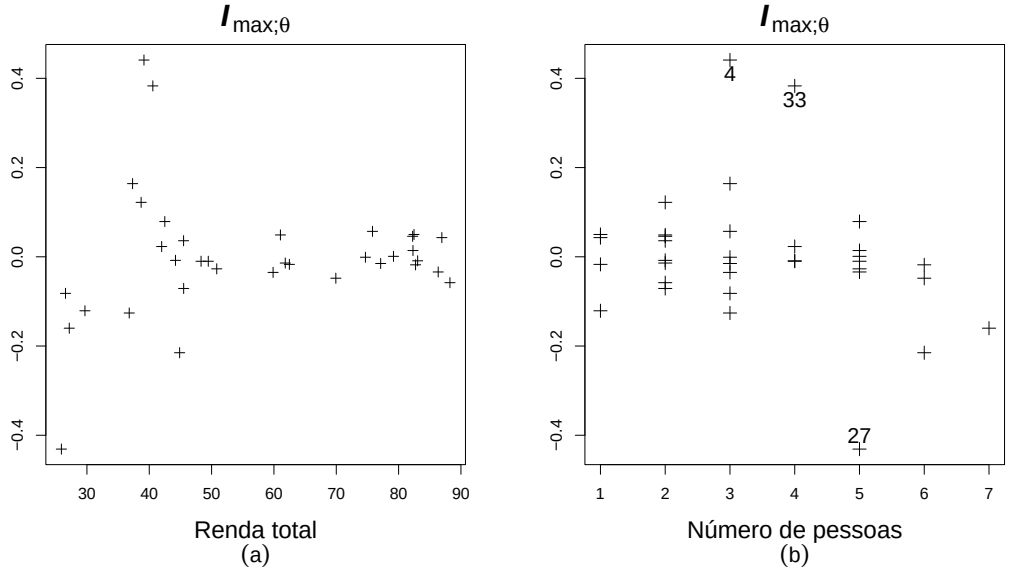

Figura 3.5. Gráficos de influência. Dados de gastos com alimentação excluindo os casos 5, 11 e 20.

As covariáveis utilizadas para modelar a média da resposta são uma variável indicadora de ausência ou presença de dislexia (-1: ausência e 1: presença) (x_2), o escore padronizado de QI não–verbal (x_3), chamadas de dislexia e QI, respectivamente e a interação (produto) entre as duas covariáveis (x_4). Os resultados inferenciais do modelo beta logito ajustado a esses dados estão apresentados na Tabela 2.2. É importante ressaltar que a estimativa de ϕ é consideravelmente pequena, menor que 12. No entanto, $y \in (0.459, 0.990)$ e a mediana de y é 0.706. Temos ainda que, a mediana do escore de habilidade em leitura para as crianças não–disléxicas é 0.990, enquanto que para as disléxicas é 0.609. Vimos que apenas a covariável dislexia era considerada significativa para a explicação da média. Smithson e Verkuilen (2006) argumentam que esta conclusão pode ser enganosa, já que o efeito das covariáveis na dispersão dos dados não está sendo considerado.

Apesar do fato de apenas a covariável dislexia ter sido considerada significativa, para efeito da análise de resíduos realizada na Seção 2.7.3 consideramos o modelo com as duas covariáveis e a interação. A análise revelou que a possibilidade da modelagem simultânea da média e da dispersão sugerida por Smithson e Verkuilen (2006) é bastante razoável, uma vez que os gráficos dos resíduos evidenciam que os indivíduos com dislexia apresentam maior dispersão que os indivíduos sem, no que diz respeito aos escores de habilidade de leitura.

Ainda foi detectada uma relação entre a covariável QI e a dispersão dos dados. A análise de influência que realizaremos a seguir considera o mesmo modelo em que foi realizada a análise de resíduos do Capítulo 2. Começamos a análise a partir dos gráficos da distância de Cook e de alavanca generalizada (Figura 3.6a,b). Nota-se que a distância de Cook e a alavanca generalizada destacam claramente a observação 8.

Partindo para a análise de influência local, construímos os gráficos de I_{max} e C_t contra os índices das observações, segundo a rotina descrita no início da Seção 3.7 (Figuras 3.7 e 3.8). Aqui, a covariável perturbada foi QI. Inicialmente analisaremos a influência simultânea das observações (Figura 3.7). A partir dos esquemas de ponderação de casos e de perturbação da precisão, que conduzem a gráficos quase idênticos (Figura 3.7a–c; 3.7j–k), nota-se a influência de dois sub–conjuntos de dados na estimativa de β, em sentidos opostos, $\{8, 9, 15, 22\}$ e $\{6, 17, 19, 23, 24\}$. Nota-se também a influência conjunta dos casos 17, 19 e 24 na estimativa de ϕ (Figura 3.7a–c; 3.7j–k). Já o esquema de perturbação da resposta (Figura 3.7d–f) destaca os casos 8, 9, 15 e 22 como conjuntamente influentes para as estimativas de β e ϕ. Segundo a perturbação da covariável QI (Figura 3.7g–i), os casos 8, 9, 15, 22 seriam destacados.

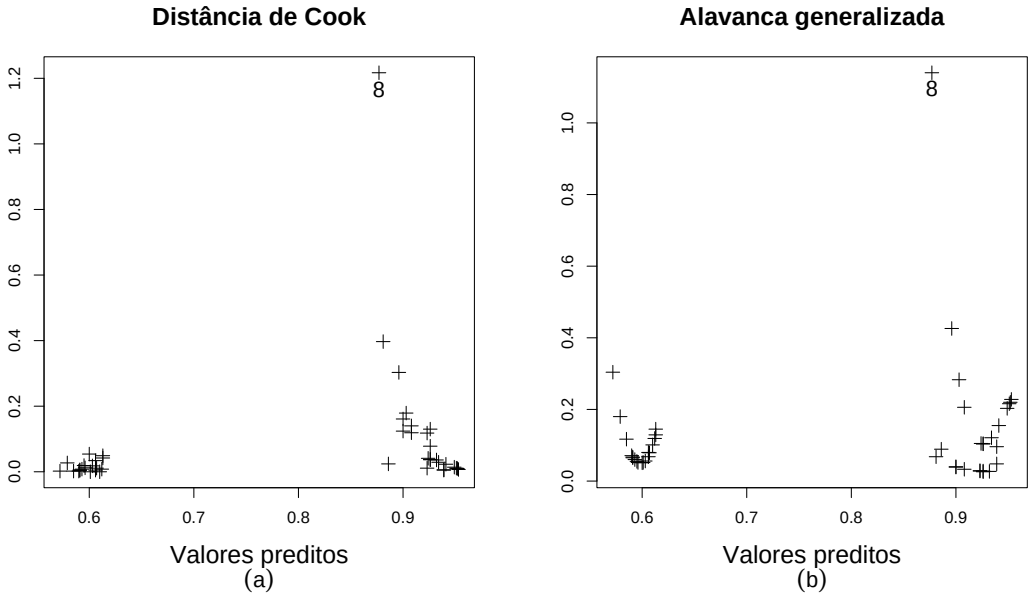

Figura 3.6. Gráficos de distância de Cook e alavanca generalizada. Dados de habilidade de leitura.

Quanto à influência individual das observações, o caso 8 influencia a estimativa de β considerando todos os esquemas de perturbação (Figura 3.8), enquanto que influencia a estimativa de ϕ apenas através da variável resposta e da covariável QI (Figura 3.8f,i). Note que o caso 8 é um ponto de alavanca e também influente. O caso 15 também é considerado influente a partir dos esquemas de perturbação da covariável e da resposta, tanto para a estimativa de β quanto para a de ϕ (Figura 3.8d–i). Devem ser analisados individualmente também os casos 6, 17, 19, 22 e 24. De acordo com a discussão acima, excluímos individualmente os casos 6, 8, 9, 15, 17, 19, 22, e 24, conjuntamente os casos 8, 9, 15 e 22, os casos 6, 17, 19, 23 e 24 e os casos 17,19 e 24. As variações percentuais nas estimativas dos parâmetros relativas às exclusões mais relevantes e os p-valores para os testes de significância associados aos parâmetros são dados na Tabela 3.2.

Nota-se que o caso 8 influencia consideravelmente as estimativas de β_3 e de β_4 (interação entre QI e dislexia). De fato, a covariável QI passa a ser quase significante ao nível de 5% e a interação torna-se significativa a esse nível. Ressaltamos que as exclusões individuais das observações 6, 9, 15, 17, 19, 22 e 24 resultam em mudanças inferenciais bem menos expressivas que a relacionada à exclusão do caso 8, uma vez que a interação quase é significante, mas a covariável QI continua sendo considerada sem importância para a explicação do escore em habilidade de leitura. Por isso não apresentamos esses resultados aqui.

É notável o impacto da exclusão conjunta dos casos 8, 9, 15 e 22 nas estimativas dos coeficientes do modelo. Quando a observação 8 é excluída dos dados a estimativa do coeficiente da variável QI passa de 0.1607 para 0.266, reforçando a relação positiva entre a resposta, escore em habilidade de leitura, e a covariável QI. Ou seja, espera-se que quanto maior o escore de QI não–verbal, maior o escore em habilidade de leitura. O caso 8 é influente exatamente porque contradiz essa relação, uma vez que ele representa uma criança não disléxica, com um escore em habilidade de leitura igual 0.99 (o maior valor possível) mas com um escore de QI não–verbal muito baixo (-0.914). Note que QI varia de -1.745 a 1.856 e apenas 18% das crianças têm escores de QI não–verbal inferior a -0.914. Os casos 9, 15 e 22 apresentam a mesma característica, mas de forma menos acentuada. Assim, quando excluímos os casos 8, 9, 15 e 22, a estimativa do coeficiente de QI passa de 0.1607 para 0.4899, realçando a importância desta covariável para a explicação da habilidade de leitura.

Por outro lado, a exclusão dos casos 6, 17, 19, 23 e 24 conduz à direção oposta, isto porque esses casos representam crianças disléxicas, com um escore em habilidade de leitura no intervalo $(0.64, 0.73)$, ou seja, próximos de 0.7 mas com escores de QI não–verbal consideravelmente distintos, entre -0.795 e 0.590. Para esse grupo de observações não fica clara a relação entre a resposta e QI . O gráfico de influência na Figura 3.7a mostra a direção oposta dos componentes de $I_{max;\theta}$ dos dois conjuntos de dados, $\{8, 9, 15, 22\}$ e $\{6, 17, 19, 23, 24\}$.

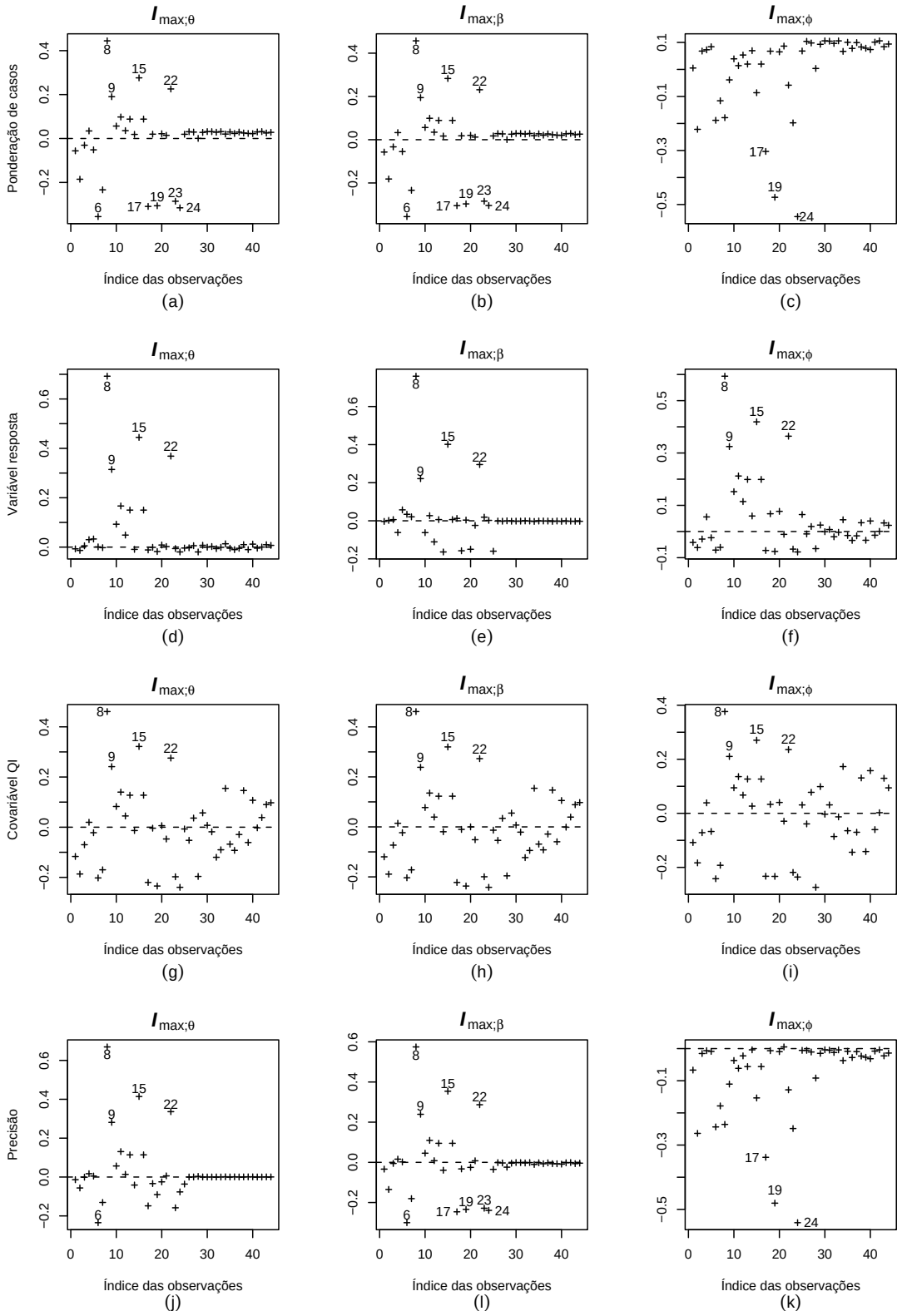

Figura 3.7. Gráficos de influência local. Dados de habilidade de leitura.

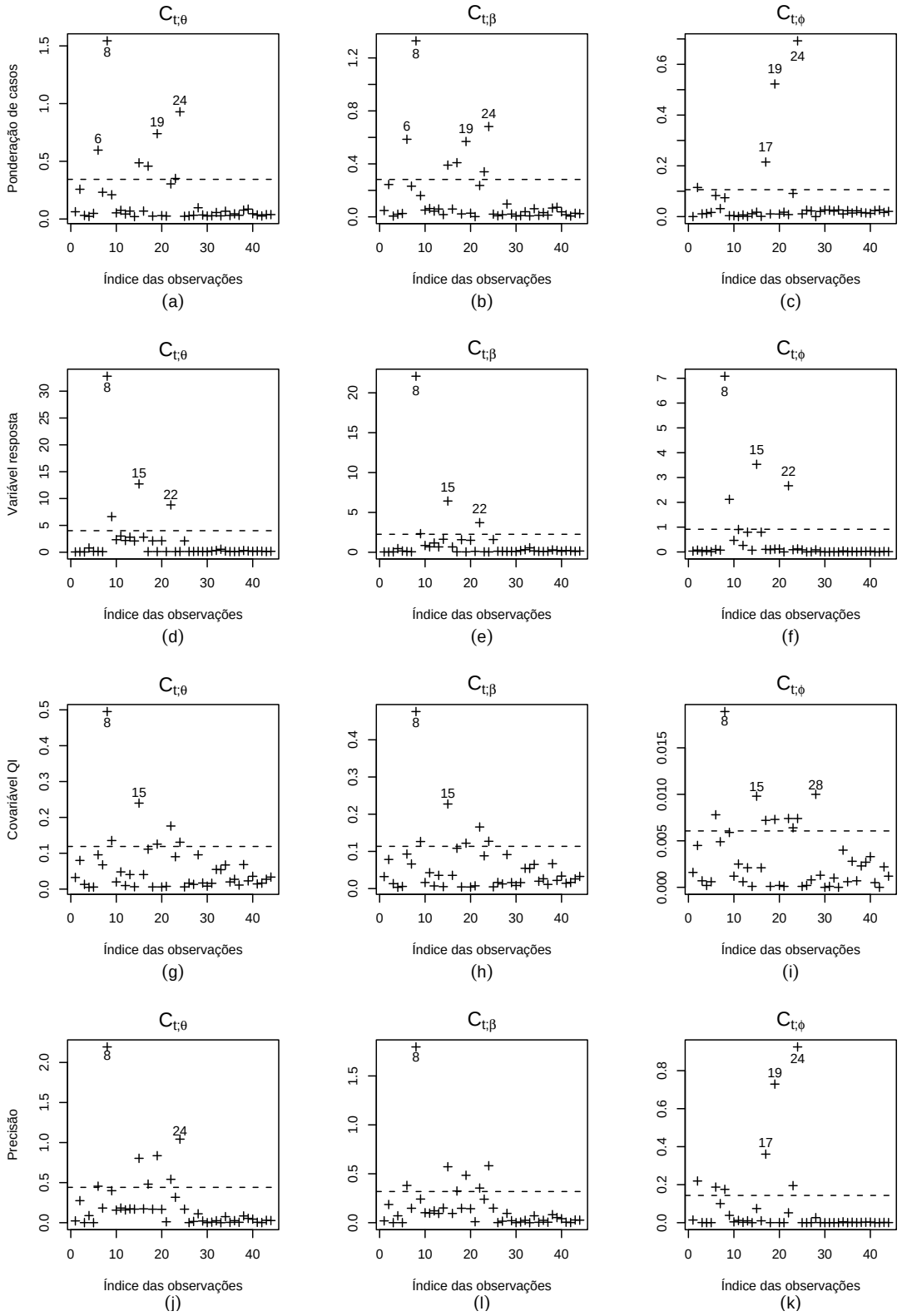

Figura 3.8. Gráficos de influência local total. Dados de habilidade de leitura.

Assim, quando excluímos os casos 6, 17, 19, 23 e 24 enfatizamos a influência dos casos 8, 9, 15 e 22 e reforçamos a idéia de que QI não é significante. Notamos ainda que os casos 17, 19 e 24 influenciam bastante na estimativa de ϕ. No entanto, este fato não interfere nas estimativas dos parâmetros relacionados com a média. Ressaltamos que no Capítulo 2 os gráficos dos resíduos ponderados padronizados 1 e 2 destacaram como aberrante o caso 8, que efetivamente mostrou-se decisivo para o ajuste do modelo, enquanto o gráfico do resíduo ordinário padronizado destacou os casos 2, 17, 19 e 24.

Outra informação importante que obtemos a partir da análise do gráfico dos valores de $I_{max;\theta}$ contra os índices das observações, considerando o esquema de ponderação de casos, é a diferença muito evidente entre as dispersões da primeira (grupo sem dilexia) e da segunda metade dos dados (grupo com dilexia)(Figura 3.7a). O gráfico de $I_{max;\theta}$ (considerando o esquema de ponderação de casos) contra os valores da covariável dislexia também mostra essa diferença (Figura 3.9a). Já o gráfico dos valores de $I_{max;\theta}$ contra os valores da covariável (Figura 3.9b) torna mais evidente a tendência apresentada pelo gráfico do resíduo ponderado padronizado 2 contra os valores da covariável QI (Figura 2.10f). Notamos que os valores de $I_{max;\theta}$ que correspondem aos valores de QI entre $(-1, 1)$ estão muito mais dispersos que aqueles correspondentes aos valores extremos.

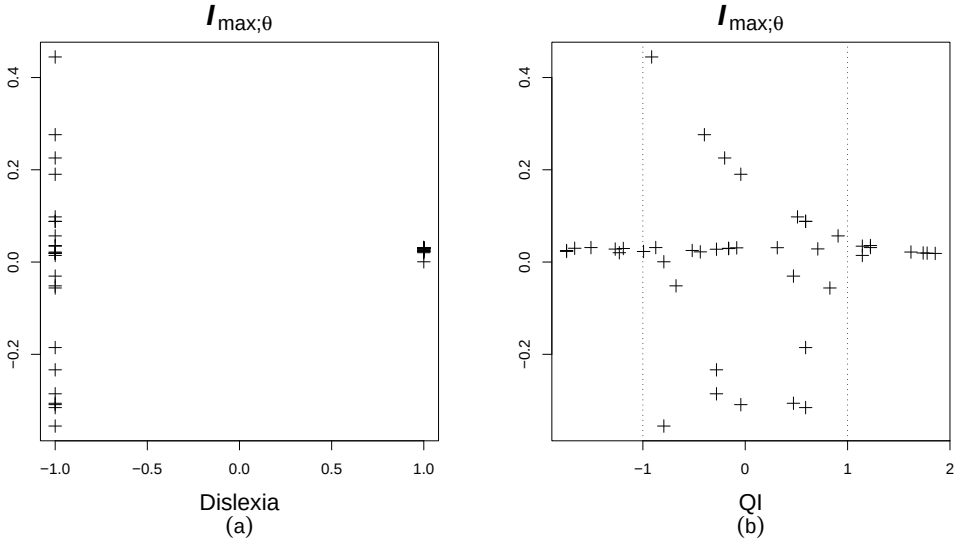

Figura 3.9. Gráficos de influência. Dados de habilidade de leitura.

Tabela 3.2. Variação percentual das estimativas dos parâmetros e p–valores retirando observações influentes. Dados de habilidade de leitura.

Parâmetros	β_1	β_2	β_3	β_4	ϕ
8	5.97	8.43	65.63	48.46	7.71
p–valor	0.0000	0.0000	**0.0516**	**0.0178**	
8, 9, 15 e 22	21.98	31.30	204.75	151.52	51.77
p–valor	0.0000	0.0000	**0.0001**	**0.0000**	
17, 19 e 24	14.22	18.20	8.38	7.14	58.07
p–valor	0.0000	0.0000	0.1546	0.0559	
6, 17, 19, 23 e 24	29.53	38.46	52.14	36.84	128.04
p–valor	0.0000	0.0000	0.5315	0.2617	

Para finalizar realizamos uma análise de diagnóstico para o modelo excluindo o caso 8 (Figura 3.10). Consideramos o gráfico do resíduo ponderado padronizado 2 contra os índices das observações (a), contra os valores de QI (b) e o gráfico normal de probabilidades (c). Em seguida, a partir do esquema de ponderação de casos construímos os gráficos de $I_{max;\theta}$ contra os índices das observações (d) e versus os valores das covariáveis dislexia (e) e QI (f). A principal conclusão obtida a partir da Figura 3.10 é que a diferença entre as dispersões dos dados subdivididos segundo as duas categorias da covariável dislexia se mantém. É notável a diferença entre as dispersões da primeira e da segunda metade dos dados, através do gráfico dos resíduos contra os índices das observações e mais claramente através dos gráficos de $I_{max;\theta}$ contra os índices das observações. As relações entre a dispersão dos dados e as covariáveis QI e dislexia são mantidas. Assim, o modelo parece não muito bem ajustado aos dados, o que é também evidenciado pelo gráfico normal de probabilidades (Figura 3.10c).

Diante do que discutimos até agora parece razoável avaliar um modelo de regressão beta em que a média e a dispersão sejam modeladas simultaneamente considerando como covariáveis QI e dislexia. De fato, Smithson e Verkuilen (2006) consideram tal modelo e chegam à conclusão que as covariáveis QI e dislexia são significantes tanto para o modelo da média quanto para o modelo da dispersão, enquanto a interação é significante apenas para o modelo da média. No Capítulo 5 faremos a análise de diagnóstico desse modelo e então veremos se a modelagem da dispersão acomoda os pontos que influenciaram tanto o ajuste analisado nesta seção.

A partir desta aplicação foi possível ainda reforçar as conclusões apresentadas no Capítulo 2, ou seja, que o uso do resíduo ponderado padronizado 2 em gráficos de diagnóstico, entre os resíduos investigados, é o mais indicado se o objetivo é captar pontos mal ajustados e que são relevantes para o ajuste do modelo.

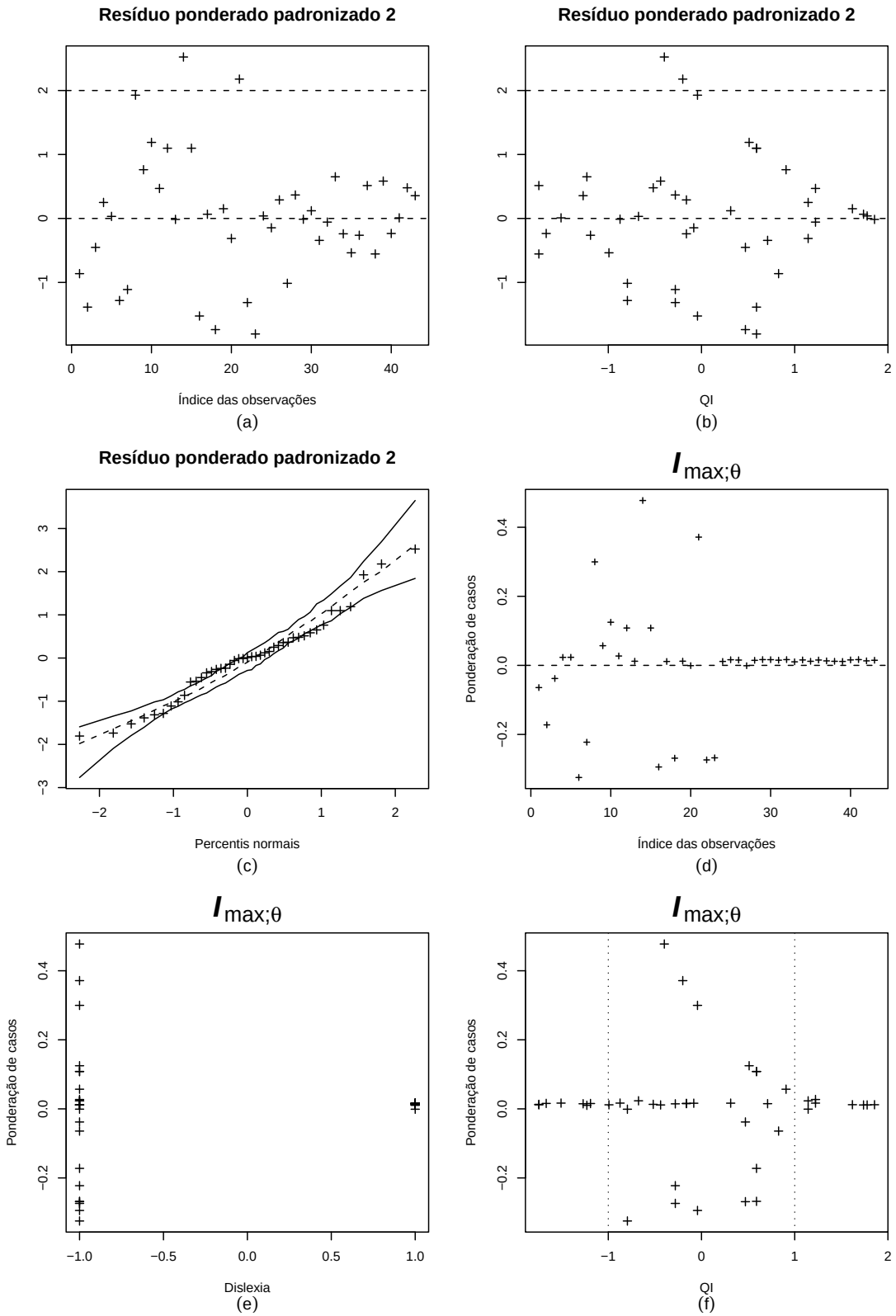

Figura 3.10. Gráficos de diagnóstico. Dados de habilidade de leitura excluindo o caso 8.

3.6.3. Aplicação III: dados de ansiedade e estresse

A terceira aplicação é referente ao exemplo 2 analisado por Smithson e Verkuilen (2006). Os dados são de um estudo sobre ansiedade e estresse realizado em um grupo de 166 mulheres "normais", ou seja, fora de um quadro clínico patológico (Townsville, Queensland, Austrália). Smithson e Verkuilen (2006) argumentam que é provável que o indivíduo sob algum tipo de ansiedade esteja também sob estresse. No entanto, o contrário não é verdade. Muitas pessoas podem experimentar estresse sem estarem ansiosas. Segundo os autores, o estresse pode ser pensado como uma condição necessária mas não suficiente para que a ansiedade ocorra e, nesta situação, um modelo em que a variável resposta é o nível de ansiedade e a variável independente é o nível de estresse deve considerar dispersão variável. Os autores propõem diretamente um modelo de regressão beta em que estresse é preditor tanto no modelo da média quanto no modelo da dispersão. Nossa proposta é considerar um modelo de regressão beta em que apenas a média da variável resposta, nível de ansiedade, seja modelada e então realizar uma análise de diagnóstico que justifique a modelagem da dispersão. Assim, ajustamos a estes dados um modelo de regressão beta com função de ligação logito e além do intercepto consideramos a covariável estresse (x_2). As estimativas dos parâmetros deste modelo bem como os p–valores dos testes de significância estão apresentados na Tabela 3.3.

Tabela 3.3. Resultados inferenciais. Dados de ansiedade e estresse

Parâmetros	β_1	β_2	ϕ
Estimativas	-3.479	3.749	11.674
p–valor	0.0000	0.0000	

Nota-se que a estimativa de ϕ é muito pequena. Outra questão que torna o cenário ainda mais crítico é o fato dos dados estarem concentrados próximos ao extremo inferior do intervalo $(0, 1)$, $y \in [0.01, 0.69]$ com mediana de $y = 0.04$. Considerando o resíduo ponderado padronizado 2, na Figura 3.11 apresentamos o gráfico normal de probabilidades com envelopes simulados e, na Figura 3.12, os gráficos dos resíduos contra contra os valores da covariável estresse (a), contra os índices das observações (b) e contra os valores preditos (c). Na Figura 3.13, apresentamos o gráfico da distância de Cook contra os valores preditos e o gráfico de alavanca generalizada contra os valores preditos. Os gráficos de I_{max} e os gráficos de C_t contra os índices das observações são similares e destacam o caso 89 consideravelmente. Assim, apresentamos na Figura 3.14 apenas os gráficos de C_t. Neste caso, a covariável perturbada é estresse. Seguindo a sugestão de Cook (1986) construímos gráficos de $I_{max;\theta}$ contra os valores da covariável estresse (Figura 3.15).

Como havíamos constatado em simulação, neste cenário a distribuição do resíduo ponderado padronizado 2 apresenta uma assimetria à esquerda (Figura 3.11). A Figura 3.11 revela ainda uma certa falta de qualidade do ajuste. O gráfico do resíduo ponderado padronizado 2 contra a covariável estresse revela uma relação crescente entre a dispersão dos dados e os valores da covariável (Figura 3.12a). Esta tendência é mais evidente no gráfico de $I_{max;\theta}$ contra os valores da covariável estresse (Figura 3.15a). Considerando o gráfico do resíduo ponderado padronizado 2 contra os índices das observações, o gráfico de alavanca generalizada, o gráfico da distância de Cook e os gráficos de influência local total (Figura 3.14), temos que o caso 89 é considerado ponto de alavanca e influente. Ao investigarmos essa observação notamos que ela representa uma mulher com um nível de estresse muito alto, 0.65 (note que o nível de estresse varia de 0.01 a 0.85), com o menor nível de ansiedade encontrado nos dados, igual a 0.01.

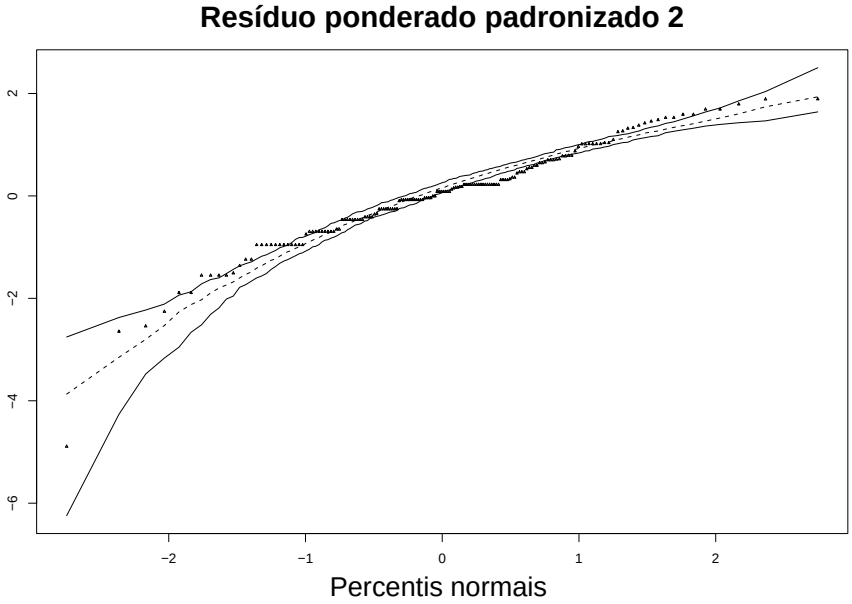

Figura 3.11. Gráfico normal de probabilidades com envelopes simulados. Dados de ansiedade e estresse.

No entanto, excluindo esta observação as alterações nas estimativas dos parâmetros não são relevantes e a relação entre a covariável estresse e a dispersão dos dados é ainda mais evidente (Figura 3.15b). Desta forma, mais uma vez, um modelo de regressão beta em que a dispersão é modelada parece ser uma boa alternativa, podendo inclusive absorver melhor a influência do caso 89. De fato, Smithson e Verkuilen (2006) ajustam tal modelo e concluem que a covariável estresse é significativa tanto para a média quanto para a dispersão. No Capítulo 5 realizaremos a análise de diagnóstico para esse modelo.

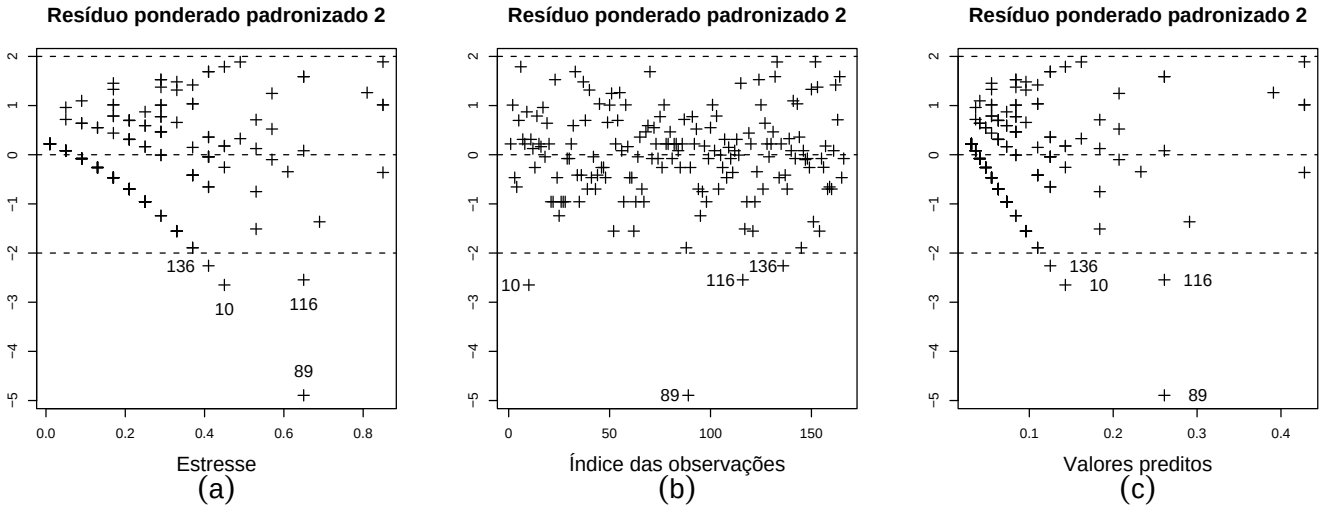

Figura 3.12. Gráficos de resíduos. Dados de ansiedade e estresse.

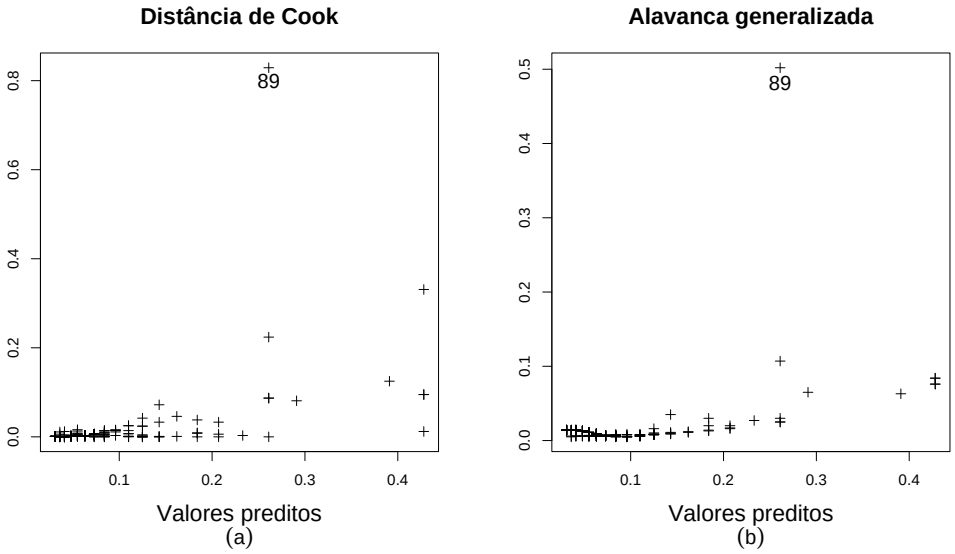

Figura 3.13. Gráficos de distância de Cook e alavanca generalizada. Dados de ansiedade e estresse.

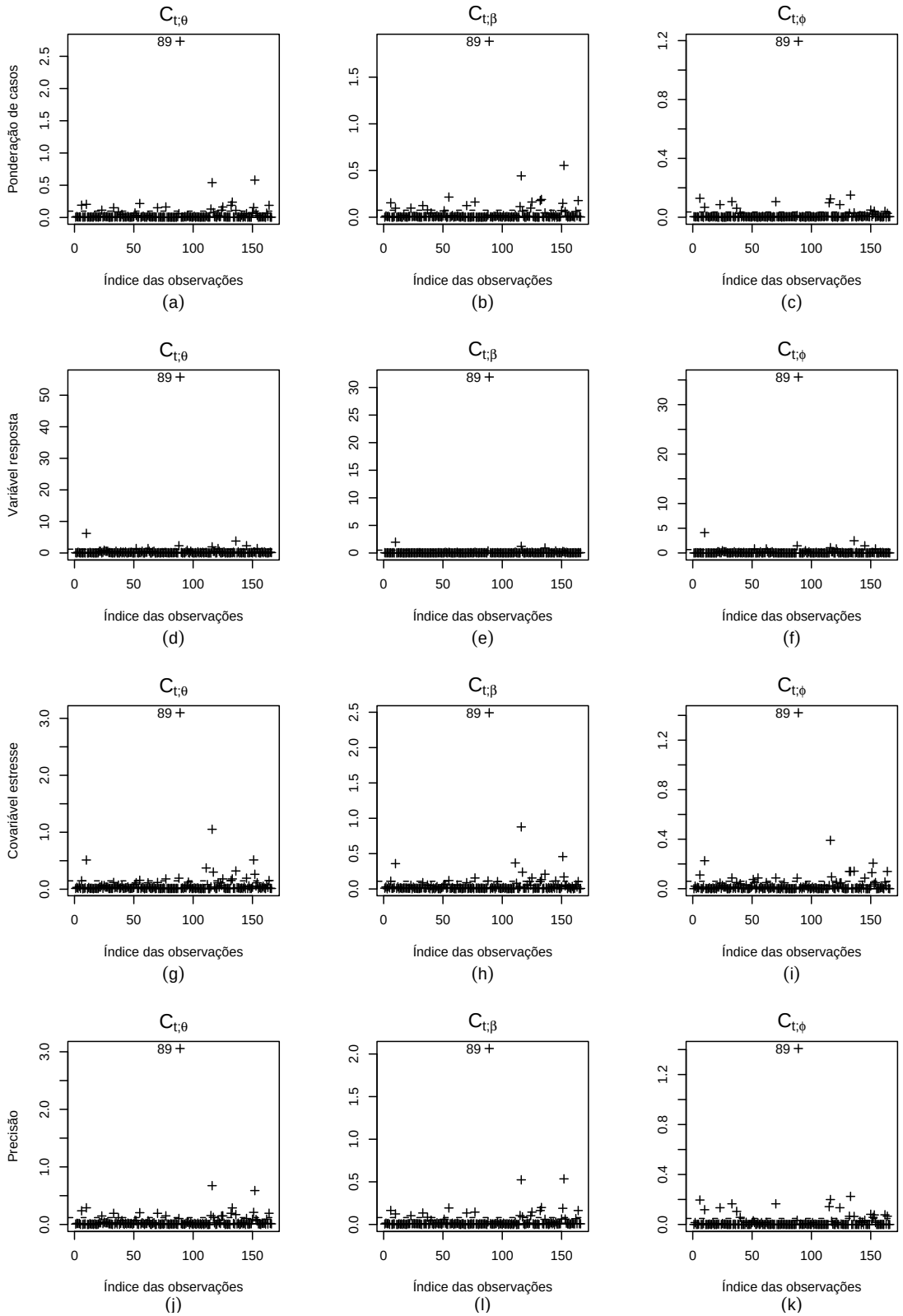

Figura 3.14. Gráficos de influência local total. Dados de ansiedade e estresse.

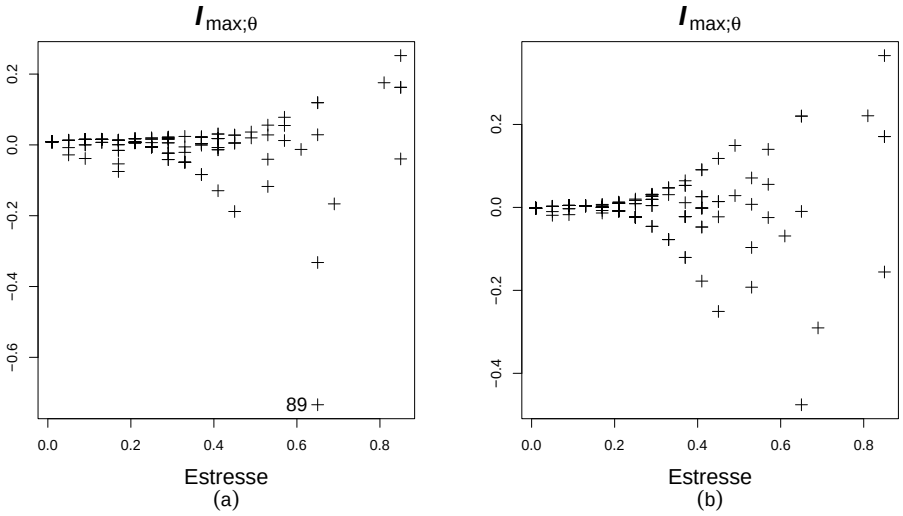

Figura 3.15. Gráficos de influência considerando o esquema de ponderação de casos. Dados completos (a). Dados sem a observação 89 (b). Dados de ansiedade e estresse.

3.6.4. Aplicação IV: dados de confiança em vereditos

A quarta aplicação trata do exemplo 1 de Smithson e Verkuilen (2006). Os dados se referem a um estudo em que 104 estudantes do primeiro ano de psicologia da Universidade Nacional da Austrália simularam vereditos para uma suposta infração. O interesse recai em modelar o grau de confiança. Uma possível covariável é o tipo de veredito disponível, que no caso é dois tipos. O primeiro é o convencional: culpa, absolvição, duas opções. O segundo é: culpa ou absolvição ou não há provas suficientes para uma conclusão, três opções. Outra possível covariável é a evidência de testemunhos contraditórios: conflito no testemunho. O grau de confiança do jurado em seu veredito é uma taxa percentual $[0, 100]$. Assim, para trabalhar com uma variável resposta no intervalo $(0, 1)$ os autores consideram a seguinte transformação:

$$y = \frac{\text{grau} \times 99}{100} + 0.005.$$

Inicialmente, Smithson e Verkuilen (2006) ajustam um modelo normal 2×2 sendo y a variável dependente. Um dos fatores é tipo de veredito e o outro fator é evidência de testemunho contraditório. O fator tipo de veredito é igual a -1 caso os jurados tenham duas opções de veredito e igual a 1, caso os jurados tenham três opções. Já o fator 2 é -1 para ocorrência de conflito no testemunho e 1 para não conflito. A partir de agora

chamaremos a covariável indicadora de presença ou ausência de conflito, apenas de conflito. Os autores argumentam que mesmo utilizando a transformação logito nos dados, os fatores 1 e 2 parecem não interferir no grau de confiança dos jurados. Em seguida, os autores ajustam um modelo de regressão beta modelando apenas a média através de tipo de veredito, conflito e a interação entre as duas. O resultado foi o mesmo do modelo normal 2×2. Já quando os autores consideram um modelo de regressão beta com dispersão variável, as covariáveis tipo de veredito, conflito e a interação são significativos para o modelo da dispersão, enquanto no modelo da média apenas a interação mostrou-se significativa.

Nesta aplicação ajustaremos um modelo de regressão beta utilizando a função de ligação logito para a média da variável resposta, grau de confiança do jurado em seu veredito. Consideraremos um intercepto e como covariáveis tipo de veredito (x_2), conflito (x_3) e o produto, interação, entre tipo de veredito e conflito (x_4) (Tabela 3.4). Realizaremos uma análise de diagnóstico, incluindo análise de resíduos e análise de influência, para averiguar a necessidade da modelagem da dispersão defendida por Smithson e Verkuilen (2006) e os motivos da não significância dos coeficientes do modelo ajustado. Antes de ajustarmos o modelo realizaremos uma breve análise descritiva para entender melhor os dados. Na Figura 3.16 estão os boxplots do grau de confiança no veredito considerando as covariáveis tipo de veredito e conflito.

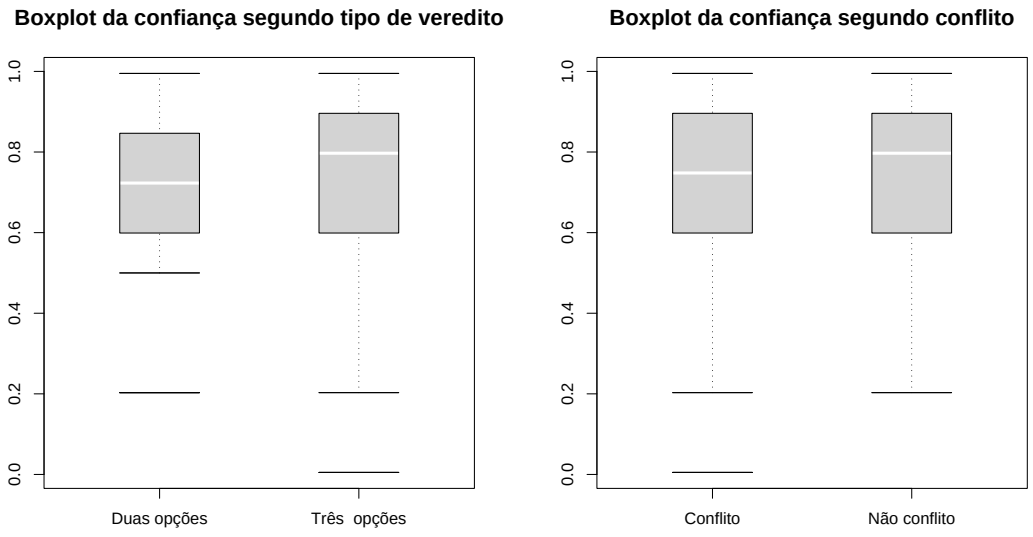

Figura 3.16. Boxplots. Dados de veredito.

Essa figura revela que a mediana do grau de confiança no veredito é um pouco maior para o grupo com três opções de veredito que para o grupo com duas opções; de fato, as medianas são iguais a 0.72 e 0.80, respectivamente. Quanto à covariável conflito, para o grupo sem ocorrência de conflito a mediana do grau de confiança é 0.74 e para o grupo com ocorrência de conflito a mediana é 0.69. Segundo Smithson e Verkuilen (2006) esperava-se um grau de confiança no veredito mediano mais expressivo para os grupos com três opções de veredito e sem ocorrência de conflito no testemunho. Nota-se ainda a partir da Figura 3.16 que o grupo com três opções de veredito apresenta maior dispersão que o grupo com duas opções e que o grupo com conflito no testemunho apresenta maior dispersão que o grupo sem.

Tabela 3.4. Resultados inferenciais. Dados de veredito

Parâmetros	β_1	β_2	β_3	β_4	ϕ
Estimativas	0.851	0.075	0.101	0.177	2.698
p–valor	0.0000	0.4565	0.3164	0.0788	

A partir da Tabela 3.4 nota-se que a estimativa de ϕ é extremamente pequena. Neste caso, os dados não estão concentrados próximo de zero ou um. De fato, $y \in (0.005, 0.995)$ com mediana de $y = 0.75$, sendo que as proporções de valores de y iguais a 0.005 e iguais a 0.995 são aproximadamente 1.3% é 6.9%, respectivamente. Nesta situação, a distribuição do resíduo ponderado padronizado 2 está melhor aproximada pela normal padrão que no cenário da aplicação anterior. Isto porque, apesar da estimativa de ϕ ser a menor, y está melhor distribuído no intervalo $(0, 1)$.

Nas Figuras 3.17–3.19 apresentamos gráficos de diagnósticos incluindo análise de resíduos e de influência. Considerando o resíduo ponderado padronizado 2, na Figura 3.17 apresentamos os gráficos dos resíduos contra os valores das covariáveis tipo de veredito (a) e conflito (b), o gráfico dos resíduos contra os índices das observações (c), o gráfico normal de probabilidades com envelopes simulados (d) e os gráficos da distância de Cook (e) e alavanca generalizada contra os valores preditos (f).

Os gráficos do resíduo ponderado padronizado 2 contra as covariáveis tipo de veredito e conflito não revelam uma grande diferença entre as dispersões das categorias (Figura 3.17a,b). O gráfico dos resíduos contra os índices das observações revela a existência de muitas observações aberrantes, que são as mesmas destacadas pelo gráfico da distância de Cook, os casos 21, 23, 37, 39, 44, 48, 49 e 54. O gráfico normal de probabilidades destaca as observações 39 e 44 e o gráfico de alavanca destaca adicionalmente o caso 89 (Figura 3.17). Ainda, é possível notar a partir do gráfico normal de probabilidades com envelopes simulados uma falta de qualidade do ajuste do modelo.

Nas Figuras 3.18 e 3.19 encontram-se, respectivamente, os gráficos de influência local e de influência local total. Ressaltamos que para este conjunto de dados o esquema de perturbação da covariável é inviável, pois as duas covariáveis são categorizadas. Analisando de forma geral os gráficos de influência local e os gráficos de diagnósticos, ao que parece os casos 39 e 44 são forte e conjuntamente influentes. Também se destacam como conjuntamente influentes as observações 21, 23, 37, 48, 49 e 54, considerando a grande maioria dos esquemas de perturbação. Considerando apenas o esquema de perturbação da variável resposta, em adição aos casos anteriores, estão os casos 89, 92, 94 e 103 (Figura 3.18d).

De fato, o gráfico na Figura 3.18d encerra uma informação importante, uma vez que revela as direções opostas de influência de dois conjuntos de dados $\{39, 44\}$ e $\{21, 23, 37, 48, 49, 54, 89, 92, 94, 103\}$. Segundo a influência individual das observações medida pela influência local total (Figura 3.19), os casos 39 e 44 devem também ser analisados separadamente. Realizamos várias combinações de exclusões para as observações acima; apresentamos apenas os resultados referentes às exclusões mais relevantes (Tabela 3.5).

Nota-se a partir desta tabela que de fato os casos 39 e 44 são conjuntamente influentes, em especial para estimativa de β_2 (tipo de veredito), a ponto de suas exclusões implicarem na quase significância deste coeficiente (p–valor $= 0.058$). Maior impacto ocorre ao excluírmos conjuntamente as observações 21, 23, 37, 39, 44, 48, 49 e 54; tanto tipo de veredito, quanto a interação passam a ser significativas ao nível de 5% e conflito no testemunho é significativo ao nível de 10%. No entanto, se em seguida ainda excluímos os casos 89, 92, 94 e 103 os coeficientes voltam a ser não–significativos, como no modelo original (esse resultado não está na Tabela 3.5). É interessante notar ainda os efeitos opostos das exclusões dos conjuntos $\{39, 44\}$ e $\{21, 23, 37, 48, 49, 54, 89, 92, 94, 103\}$. Uma questão importante, que pode auxilar a compreender melhor estes dados, é que no caso das observações 39 e 44, o valor da resposta é $y = 0.005$, o que implica que o grau de confiança original é igual a 0. Já para os casos 21, 23, 37, 48, 49 , 54, 89, 92, 94 e 103, $y = 0.995$, ou seja grau de confiança original igual a 1. Mas são as observações 21, 23, 37, 39, 44, 48, 49 e 54 as que mais interferem na estimação do modelo. Além de serem originalmente zeros ou uns, estas observações apresentam comportamentos que contradizem fortemente a relação positiva entre as categorias três opções de veredito e não conflito com a resposta e a relação negativa entre as categorias duas opções de veredito e conflito com a resposta. Os casos 39 e 44 representam indivíduos com a opção de três vereditos que presenciaram testemunhos com conflito e apresentam o menor grau de confiança no veredito 0.005.

Assim, ao que parece, a presença do conflito interferiu fortemente de forma negativa na confiança no veredito, anulando inclusive a relação positiva do tipo de veredito com três opções. Mas isto não é o que ocorre com os casos 21 e 23, que representam pessoas com duas

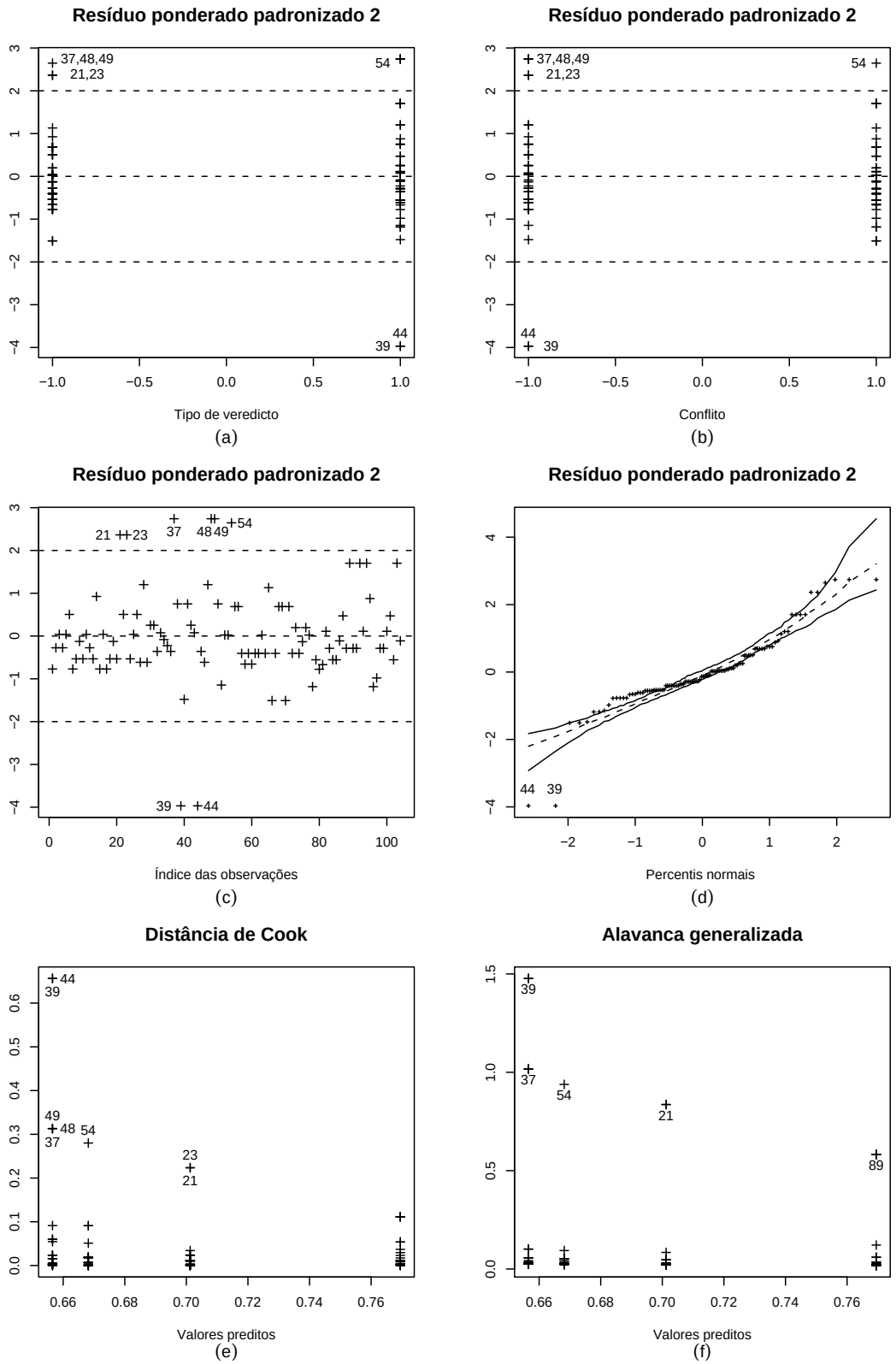

Figura 3.17. Gráficos de diagnóstico. Dados de veredito.

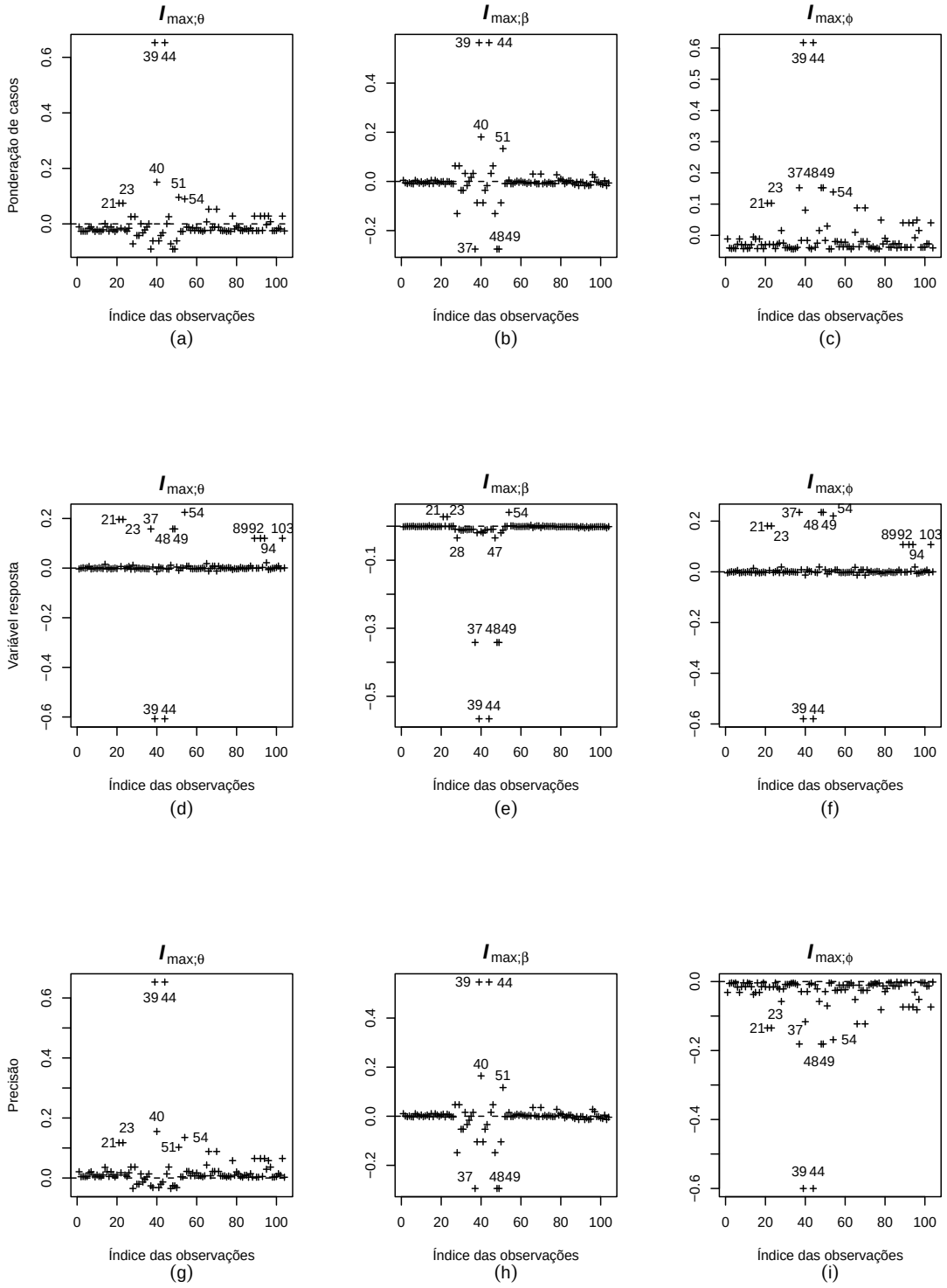

Figura 3.18. Gráficos de influência local total. Dados de veredito.

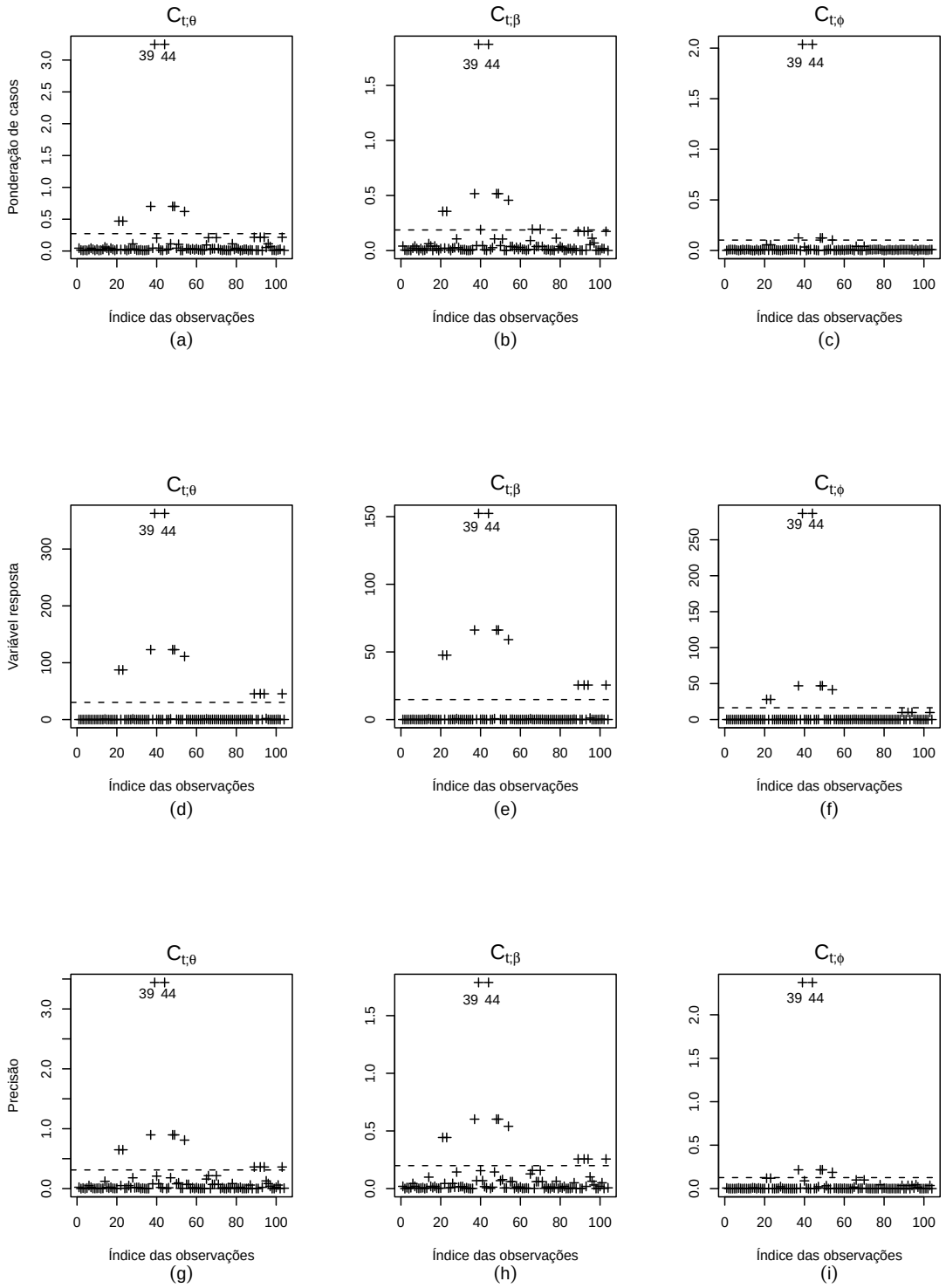

Figura 3.19. Gráficos de influência local total. Dados de veredito.

opções de veredito, com conflito no testemunho e com o maior grau de confiança possível: 0.995. Outra contradição são os casos 37, 48 e 49, que apresentam o mesmo perfil dos casos 39 e 44 quanto a tipo de veredito e conflito, mas com o grau de confiança 0.995. Finalmente, o caso 54 representa uma pessoa que teve apenas duas opções de veredito, mas apresenta um grau de confiança 0.995.

Tabela 3.5. Variação percentual das estimativas dos parâmetros e p–valores retirando observações influentes. Dados de veredito.

Parâmetros	39 e 44		21, 23, 37, 39, 44, 48, 49 e 54		21, 23, 37, 48, 49, 54, 89, 92, 94, 103	
	%	p–valor	%	p–valor	%	p–valor
β_1	22.98	0.0000	9.32	0.0000	-22.13	0.0000
β_2	136.22	0.0581	145.11	0.0336	-136.62	0.7726
β_3	-78.76	0.8180	57.24	0.0663	56.45	0.0973
β_4	-39.34	0.2484	8.60	0.0258	6.59	0.0471
ϕ	42.70		94.17		40.04	

De acordo com a análise de diagnóstico realizada, não existem fortes indícios de que um modelo de regressão beta considerando a modelagem simultânea da média e da dispersão, em que as covariáveis para os dois modelos são tipo de veredito, existência de conflito, e a interação seja adequado para estes dados. De fato, como vimos no início da aplicação, neste caso apenas a interação entre tipo de veredito e conflito seria significativa para o modelo da média. Uma possibilidade mais razoável seria considerar uma regressão que modelasse mais adequadamente a ocorrência de zeros e uns, como por exemplo um modelo de regressão beta inflacionado com zeros e uns.

3.6.5. Aplicação V: dados de cloro disponível

Nesta última aplicação utilizaremos os dados da fração de cloro disponível após semanas de fabricação (Draper e Smith 1981, Tabela 10.2). Um certo produto deve ter uma fração de cloro disponível igual a 0.50 no momento de sua fabricação. É sabido que a fração de cloro no produto decai com o tempo. Em oito semanas após a produção, antes que o produto seja consumido, em teoria, ocorre um declínio para o nível de 0.49. Ajustamos a esses dados um modelo de regressão beta com dispersão constante em que a variável resposta é a fração de cloro disponível e a covariável é o número de semanas após a fabricação do produto (Tabela 3.6) e realizamos a análise de diagnóstico para esse modelo (Figuras 3.20–3.22).

Notamos que a distribuição postulada para os dados parece adequada (Figura 3.20d). No entanto, parece haver falta de algum componente não–linear na parte sistemática do

Tabela 3.6. Resultados inferenciais. Dados de cloro disponível.

Parâmetros	β_1	β_2	ϕ
Estimativas	-0.043	-0.012	1139.0
p–valor	0.0685	0.0000	

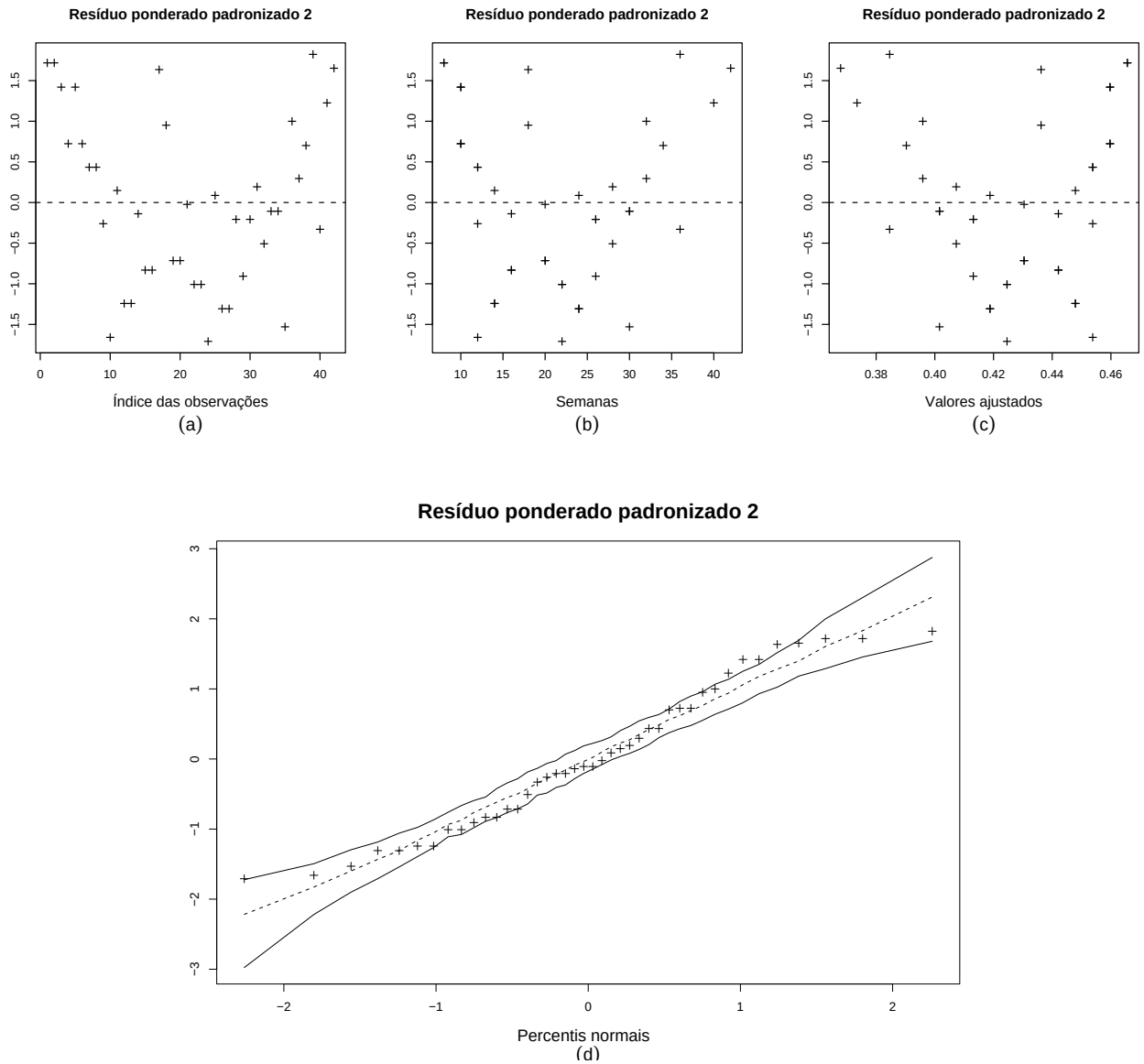

Figura 3.20. Gráficos de resíduos. Dados de cloro disponível.

modelo (Figura 3.20a–c). Os valores de I_{max} contra os índices das observações parecem também formar tendências não–lineares (Figura 3.21a–c). A partir das gráficos de I_{max} e C_t (Figuras 3.21 e 3.22) destacamos as observações 1, 2, 10, 17, 24, 35, 39, 41 e 42. A exclusão conjunta ou parcial dessas observações causa impacto expressivo na estimativa de ϕ e implica na diminuição da dispersão dos dados. No entanto, essas observações são pouco influentes para a estimativa de β_2 (Tabela 3.7).

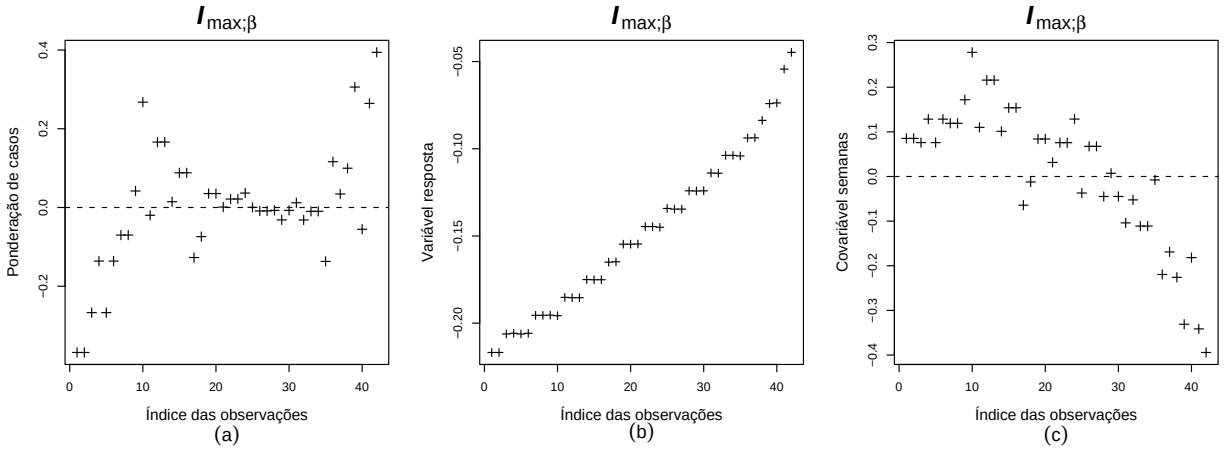

Figura 3.21. Gráficos de influência local. Dados de cloro disponível.

Ajustamos outros modelos de regressão beta admitindo funções não–lineares para a covariável semanas. As análises de diagnóstico desses modelos apontaram para falta de qualidade do ajuste, ou ainda, a falta de algum componente não–linear na parte sistemática. Para exemplificar, apresentamos na Figura 3.23 os gráficos dos resíduos contra os índices das observações obtidos a partir dos ajustes de dois modelos em que as covariáveis são $\exp(-x_2)$ (Figura 3.23a) e x_2^2 (Figura 3.23b).

Tabela 3.7. Variação percentual das estimativas dos parâmetros e p–valores retirando observações influentes. Dados de cloro disponível.

Parâmetros	β_1	β_2	ϕ
39, 41 e 42	79.5	17.3	22.1
p–valor	0.7088	0.0000	
1, 2, 10, 17, 24, 35, 39, 41 e 42	27.3	8.2	79.5
p–valor	0.1616	0.0000	

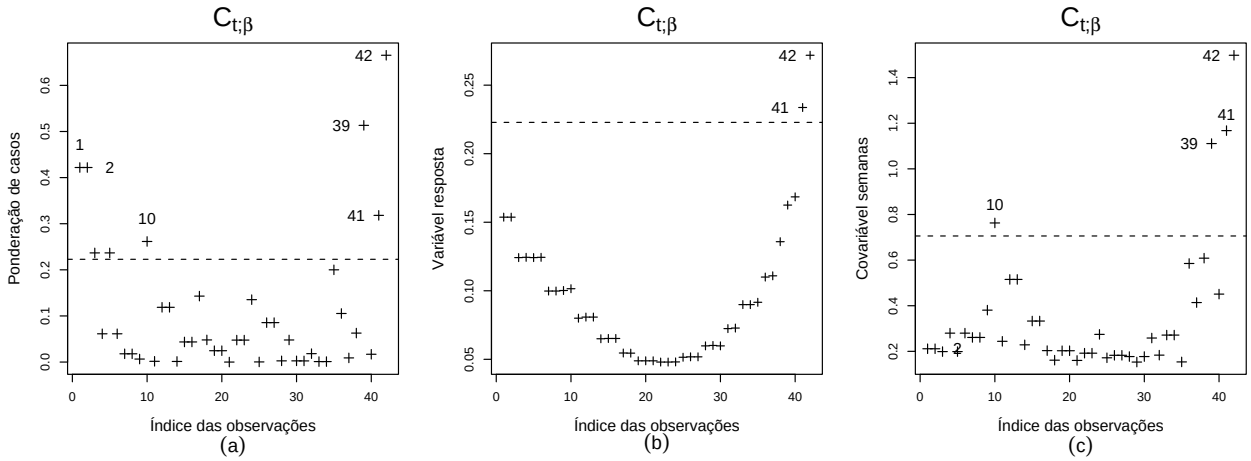

Figura 3.22. Gráficos de influência local. Dados de cloro disponível.

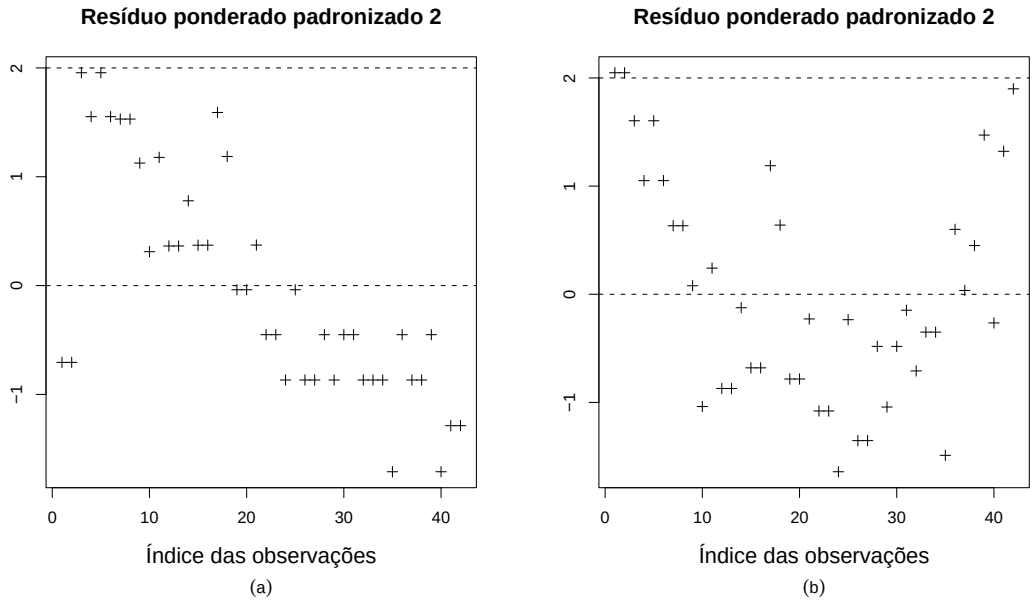

Figura 3.23. Gráficos de resíduos. Dados de cloro disponível. Modelo $\exp(-x_2)$ (a) e modelo x_2^2 (b).

Não podemos, baseados apenas nas análises gráficas, identificar precisamente que componentes da parte sistemática do modelo, o vetor de parâmetros β ou as covariáveis, ou ambos, devem ser descritos por funções não–lineares. No entanto, a análise de diagnóstico foi útil no sentido de identificar um problema de formulação do preditor (η) do modelo que propomos para os dados. De fato, neste caso temos uma informação adicional sobre a natureza dos dados. A teoria relacionada ao problema indica que a fração de cloro disponível decai segundo uma função não linear do número de semanas após a fabricação do produto e de parâmetros desconhecidos (Draper e Smith, 1981 p. 276).

A partir das cinco aplicações anteriores chegamos a algumas impressões interessantes. Primeira, das cinco aplicações, três mostram indícios da necessidade da modelagem da dispersão. Nos exemplos de Smithson e Verkuilen (2006) ocorreram situações em que a modelagem da dispersão implicou na significância de coeficientes para o modelo da média, compatível com resultados obtidos a partir da exclusão das observações influentes. Estes fatos sugerem que o modelo beta com dispersão variável, nestes casos, absorveria a influência de tais observações e explicaria melhor o comportamento da média. Segunda, a influência local pode ser muito útil no sentido de identificar a dispersão não constante dos dados, inclusive indicando quais variáveis devem modelar a dispersão. Terceira, o resíduo ponderado padronizado 2 é capaz de identificar observações aberrantes que através da análise de influência são confirmadas como decisivas no ajuste do modelo. Quarta, não existem ganhos aparentes em realizar o esquema de perturbação da precisão, uma vez que o esquema de ponderação de casos para o modelo de regressão beta parece refletir a perturbação na dispersão do t–ésimo caso, assim como ocorre no modelo normal linear. Quinta, os gráficos de influência baseados na perturbação da variável resposta são mais eficazes que o gráfico de alavancas. O esquema de perturbação da resposta destaca pontos que são alavanca e efetivamente influentes, enquanto que os gráficos de alavanca destacam pontos alavanca que podem ou não ser confirmados como influentes.

Apêndice 3.A

Influência local

3.A.1. Introdução

Como discutido na Capítulo 3, Seção 3.3, Cook (1986) propôs estudar o comportamento local de LD_δ, definido em (3.6), em uma vizinhança de δ_0, sendo que $LD_{\delta_0} = 0$. O procedimento proposto por Cook (1986) consiste em avaliar como a superfície geométrica $\alpha(\delta) = (\delta, LD_\delta)^\top$ desvia-se de seu plano tangente em δ_0, à medida que δ se afasta levemente de δ_0, ou seja, quando pequenas perturbações são introduzidas no modelo. Tal descrição pode ser obtida pelo estudo das curvaturas das secções normais da superfície $\alpha(\delta)$ em δ_0; que são chamadas de curvaturas normais. Para definir o que é uma secção normal da superfície $\alpha(\delta)$ em δ_0 consideremos o plano tangente a $\alpha(\delta)$ em δ_0, $\mathcal{T}_0$, e o vetor que é ortogonal a $\mathcal{T}_0$, que chamaremos de $\pi(\mathcal{T}_0)$. Na Figura A.1 apresentamos a curva $\alpha(\delta) = (\delta, LD_\delta)^\top$, o plano tangente a esta curva em δ_0 e o vetor $\pi(\mathcal{T}_0)$. Para esta ilustração estamos admitindo que $\delta = (\delta_1, \delta_2)$. Assim, $\delta_0 = (\delta_{01}, \delta_{02})$.

Os planos que contêm o vetor normal $\pi(\mathcal{T}_0)$ são chamados de planos normais a $\alpha(\delta)$ em δ_0. As secções normais em δ_0 são definidas como as intersecções desses planos normais com a superfície $\alpha(\delta)$. Consideremos agora um deslocamento ao redor de δ_0, ou seja, introduziremos uma pequena perturbação no modelo. Escolhemos uma direção I, tal que $\|I\| = 1$ e tomamos $\delta = \delta_0 + I$. Desta forma, teremos o ponto $LD_{\delta_0+I} \neq 0$. Por esse ponto passa o plano que contêm os vetores I e $\pi(\mathcal{T}_0)$ e a intersecção desse plano com $\alpha(\delta)$ forma uma secção normal (Figura 3.A.2). De fato, cada direção I escolhida, tal que $\|I\| = 1$, associada ao vetor $\pi(\mathcal{T}_0)$, especifica uma secção normal e define um valor de curvatura normal. Cada seção normal é equivalente ao gráfico $LD_{\delta_0+aI} \times a$, em que $a \in I\!R$. Esse gráfico é chamado de linha projetada.

Na Figura 3.A.3 ilustramos o comportamento de duas linhas projetadas $LD_{\delta_0+aI_1} \times a$ e $LD_{\delta_0+aI_2} \times a$, relacionadas, respectivamente, às direções I_1 e I_2. Uma vez que $LD_{\delta_0} = 0$, segue que LD_{δ_0+aI} tem um mínimo local em $a = 0$. Assim, é possível notar que a direção I_2 é a que provoca os maiores deslocamentos pela verossimilhança, à medida que os valores de a se afastam de zero. A idéia de Cook (1986) é analisar a direção I, tal que a linha projetada resultante apresente a maior das curvaturas normais (C_{max}). Segundo Cook (1986), este

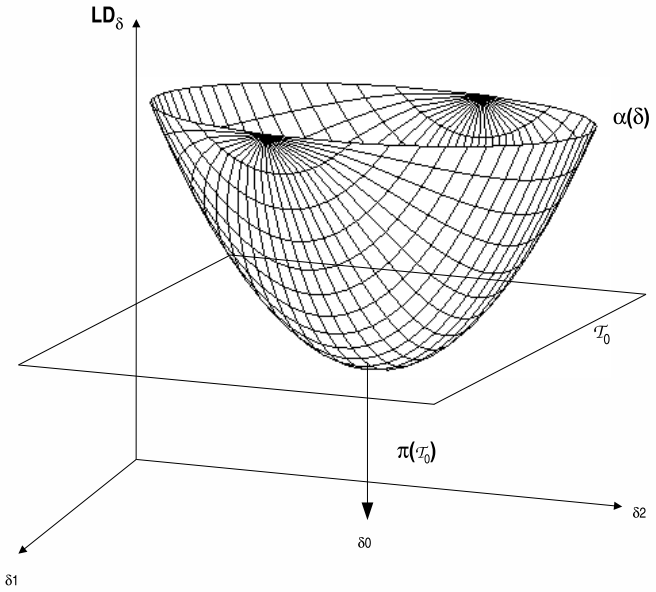

Figura 3.A.1. Curva $\alpha(\delta)$.

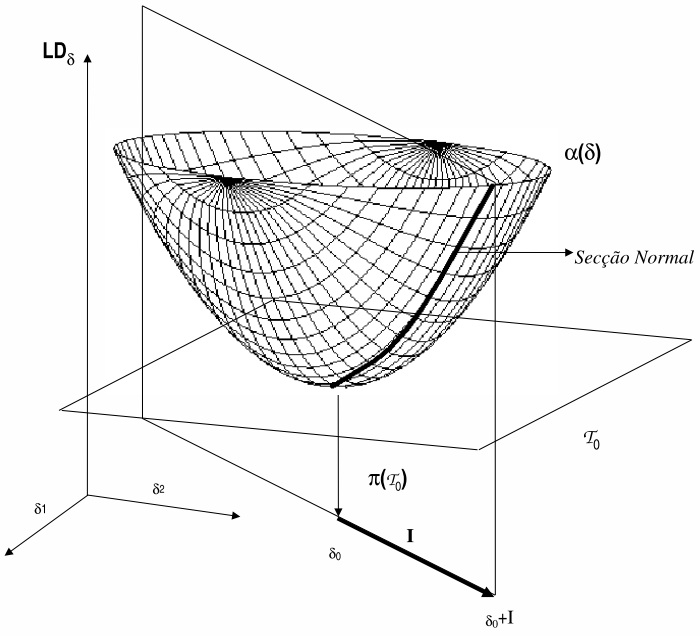

Figura 3.A.2. Seção normal.

vetor, denotado por I_{max}, indica que valores atribuir aos componentes de δ para obter a maior mudança local no deslocamento pela verossimilhança.

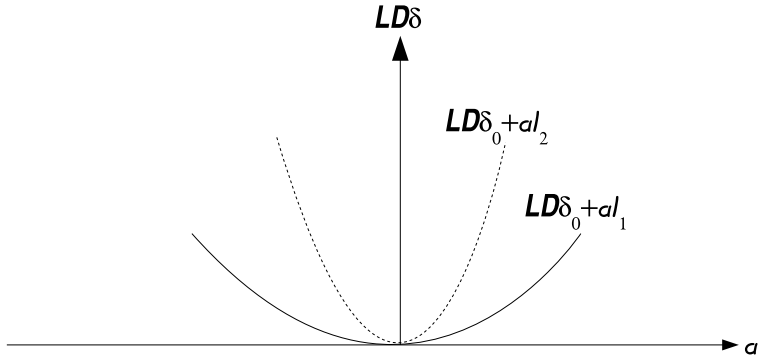

Figura 3.A.3. Linhas projetadas $LD_{\delta_0+aI_1} \times a$ e $LD_{\delta_0+aI_2} \times a$.

Cook (1986) mostra que a curvatura normal na direção I em que estamos interessados fica resumida ao módulo da segunda derivada de LD_{δ_0+aI} com respeito a a, que avaliada em $a = 0$ é dada por

$$C_I = |I^\top \ddot{LD}_\delta I|,$$

em que

$$\ddot{LD}_\delta = \frac{\partial^2 LD_\delta}{\partial\delta\partial\delta^\top},$$

avaliada em $\delta = \delta_0$ e $\theta = \widehat{\theta}$. Note que $\ddot{LD}_\delta(\cdot)$ é uma matriz simétrica e positiva definida. Pela definição do deslocamento pela verossimilhança , segue que

$$\ddot{LD}_\delta = -2\Delta^\top \ddot{\ell}^{-1}\Delta,$$

em que

$$\ddot{\ell} = \frac{\partial^2 \ell(\widehat{\theta})}{\partial\theta\partial\theta^\top} \quad \text{e} \quad \Delta = \frac{\partial^2 \ell_\delta(\widehat{\theta})}{\partial\theta\partial\delta^\top},$$

avaliada em $\delta = \delta_0$. Ressaltamos que estamos considerando $\delta = (\delta_1, \ldots, \delta_n)^\top$ e uma vez que θ um vetor $s \times 1$, temos que Δ é uma matriz $s \times n$. Finalmente, chegamos à expressão da curvatura normal em (3.7) definida por Cook (1986) [2], que também pode ser expressa como

$$C_I = 2I^\top(-\ddot{\mathcal{F}})I,$$

em que $\ddot{\mathcal{F}} = \Delta^\top \ddot{\ell}^{-1}\Delta$; ver Wei (1998, p. 126). Note que $-2\ddot{\mathcal{F}} = \ddot{LD}_\delta$ e, portanto, $-\ddot{\mathcal{F}}$ é uma matriz simétrica e positiva definida. Desta forma,

$$\max\{I^\top(-\ddot{\mathcal{F}})I : \|I\| = 1\} = \max_t \zeta_t;$$

[2] Para detalhes de todo o processo veja de Souza (1999).

em que ζ_t, $t = 1, \ldots, n$, são os autovalores da matriz $-\Delta^\top \ddot{\ell}^{-1} \Delta$. Isso quer dizer que existem vetores $\upsilon_t = (\upsilon_{t1}, \ldots, \upsilon_{tn})^\top$, $\|\upsilon_t\| = 1$, $t = 1, \ldots, n$, tais que

$$-\ddot{\mathcal{F}}\, \upsilon_t = \zeta_t \upsilon_t;$$

ver Bickel e Doksum (2001, p. 520). Os vetores υ_t, $t = 1, \ldots, n$, são os autovetores normalizados correspondentes aos autovalores da matriz $-\ddot{\mathcal{F}}$. Com base nestas definições, temos que $C_{max}/2 = \max_t \zeta_t = \zeta_{max}$ e I_{max} é o autovetor normalizado associado a $\max_t \zeta_t$, $t = 1, \ldots, n$, ou seja, $I_{max} = \upsilon_{\zeta_{max}}$, tal que, $\|\upsilon_{\zeta_{max}}\| = 1$. Na prática Cook (1986) considera $C_{max} = \max_t \zeta_t$.

3.A.2. Influência local no modelo de regressão beta

No modelo de regressão beta definido em (1.2) temos que $\theta = (\beta_1, \ldots, \beta_k, \phi)^\top$, então Δ é uma matriz $(k+1) \times n$ dada por

$$\Delta = \begin{pmatrix} \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_1 \partial \beta_1} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_n \partial \beta_1} \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_1 \partial \beta_2} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_n \partial \beta_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_1 \partial \beta_k} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_n \partial \beta_k} \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_1 \partial \phi} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial \delta_n \partial \phi} \end{pmatrix},$$

avaliada em $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ e δ_0. Podemos particionar Δ tal que

$$\Delta = \begin{pmatrix} \Delta_\beta \\ \Delta_\phi \end{pmatrix},$$

em que $\Delta_\beta = \Delta_\beta(\theta, \delta)$ é uma matriz $k \times n$ e $\Delta_\phi^\top = \Delta_\phi^\top(\theta, \delta)$ é um vetor $n \times 1$, dados, respectivamente, por

$$\Delta_\beta = \left. \frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \beta \partial \delta^\top} \right|_{(\theta = \widehat{\theta}, \delta = \delta_0)} \quad \text{e} \quad \Delta_\phi = \left. \frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \phi \partial \delta^\top} \right|_{(\theta = \widehat{\theta}, \delta = \delta_0)},$$

ou seja, ambas quantidades avaliadas em $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ e δ_0.

3.A.2.1. Ponderação de casos

No esquema de ponderação de casos temos $\ell_\delta(\beta, \phi) = \sum_{t=1}^n \delta_t \ell_t(\mu_t, \phi)$ e, assim, a t–ésima coluna de Δ, $t = 1 \ldots, n$, é dada por

$$\Delta_t = \frac{\partial \ell_t(\theta)}{\partial \theta} = \left. \left(\frac{\partial \ell_t(\theta)}{\partial \beta_1}, \ldots, \frac{\partial \ell_t(\theta)}{\partial \beta_k}, \frac{\partial \ell_t(\theta)}{\partial \phi} \right)^\top \right|_{\theta = \widehat{\theta}}.$$

Com base em (1.3), para $i = 1, \ldots, k$, obtemos

$$\frac{\partial \ell_t(\theta)}{\partial \beta_i} = \phi \frac{1}{g'(\mu_t)}(y_t^* - \mu_t^*)x_{ti},$$

e

$$\frac{\partial \ell_t(\theta)}{\partial \phi} = \{\mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1 - \mu_t)\phi) + \psi(\phi)\} = u_t,$$

em que, os t-ésimos elementos de y^* e μ^* são dados em (1.5) e $g(\mu_t)$ está definida em (1.2). Então, temos que a i–ésima linha de Δ, $i = 1, \ldots, k$, é dada por

$$\Delta_i = \left(\phi \frac{1}{g'(\mu_1)}(y_1^* - \mu_1^*)x_{1i}, \phi \frac{1}{g'(\mu_2)}(y_2^* - \mu_2^*)x_{2i}, \ldots, \phi \frac{1}{g'(\mu_n)}(y_n^* - \mu_n^*)x_{ni} \right) \Big|_{\theta = \widehat{\theta}}$$

e a $(k + 1)$–ésima linha é o vetor $u^\top = (u_1, \ldots, u_n)|_{\theta = \widehat{\theta}}$. Desta forma, podemos escrever $\Delta_\beta = \widehat{\phi} X^\top \widehat{T} \mathcal{E}$ e $\Delta_\phi = \widehat{u}^\top$, em que $\widehat{T}$, $\mathcal{E}$ e $\widehat{u}$ estão definidos na Seção 3.4.1. Assim, a matriz Δ fica expressa na forma apresentada em (3.13).

3.A.2.2. Perturbação da resposta

Com base na definição da variável resposta sob perturbação aditiva apresentada em (3.15) e considerando (1.3) temos que o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned} \ell_\delta(\mu, \phi) = \sum_{t=1}^{n} \{ & \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log(y_t + \delta_t s(y_t)) \\ & + [(1 - \mu_t)\phi - 1] \log(1 - y_t - \delta_t s(y_t))\}. \end{aligned}$$

Obtemos

$$\frac{\partial \ell_\delta(\beta, \phi)}{\partial \delta_t} = \frac{s(y_t)}{y_t + \delta_t s(y_t)}(\mu_t \phi - 1) - \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)}[(1 - \mu_t)\phi - 1].$$

Assim, para $i = 1, \ldots, k$

$$\begin{aligned} \frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \beta_i} &= \phi \frac{1}{g'(\mu_t)} x_{ti} \left(\frac{s(y_t)}{y_t + \delta_t s(y_t)} + \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \right) \\ &= \phi \frac{1}{g'(\mu_t)} x_{ti} \frac{s(y_t)}{(y_t + \delta_t s(y_t))(1 - y_t - \delta_t s(y_t))} \end{aligned}, \tag{3.A1}$$

e

$$\begin{aligned} \frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \phi} &= \mu_t \left(\frac{s(y_t)}{y_t + \delta_t s(y_t)} + \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \right) - \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \\ &= \frac{-(y_t - \mu_t)s(y_t) - \delta_t s(y_t)^2}{(y_t + \delta_t s(y_t))(1 - y_t - \delta_t s(y_t))}. \end{aligned} \tag{3.A2}$$

Avaliando as expressões (3.A1) e (3.A2) em $\delta_0 = (0, 0, \ldots, 0)^\top$ e $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ podemos escrever $\Delta_\beta = \phi X^\top \widehat{T} M S_y$ e $\Delta_\phi = \widehat{b}^\top S_y$, em que M e $\widehat{b}$ estão definidos na Seção 3.4.2. Assim, a matriz Δ fica expressa na forma apresentada em (3.16).

3.A.2.3. Perturbação individual de covariáveis

Com base na perturbação aditiva da p–ésima covariável, x_p, $p = 1, \ldots, k$, definida em (3.19) e considerando (1.3), temos que, neste caso, o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}\ell_\delta(\mu, \phi) = \sum_{t=1}^{n} \{&\log\Gamma(\phi) - \log\Gamma(\mu_t(\delta)\phi) - \log\Gamma((1-\mu_t(\delta))\phi) + (\mu_t(\delta)\phi - 1)\log(y_t) \\ &+ \{(1-\mu_t(\delta))\phi - 1\}\log(1-y_t)\},\end{aligned}$$

em que $\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$, com $\eta_t(\delta)$ exemplificado em 3.19. Assim, temos que

$$\frac{\partial \ell(\beta, \phi)}{\partial \delta_t} = \phi[\psi((1-\mu_t(\delta))\phi) - \psi(\mu_t(\delta)\phi) + y_t^*]\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p}, \tag{3.A3}$$

em que y^* é o logito da resposta e s_{x_p} é o desvio padrão de x_p. De (3.A3) segue que para $i = 1, \ldots, k$, com $i \neq p$,

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \beta_i} = &- \phi s_{x_p}\left\{\phi\left[\psi'((1-\mu_t(\delta))\phi) + \psi'(\mu_t(\delta)\phi)\right]\frac{1}{\{g'(\mu_t(\delta))\}^2}\beta_p x_{ti}\right\} \\ &- \phi s_{x_p}\left[(y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\frac{1}{\{g'(\mu_t(\delta))\}^2}\beta_p x_{ti}\right],\end{aligned} \tag{3.A4}$$

em que, para este esquema de perturbação, $\mu_t^*(\delta) = \psi(\mu_t(\delta)\phi) - \psi((1-\mu_t(\delta))\phi)$. Podemos reescrever (3.A4) como se segue:

$$\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \beta_i} = -\phi s_{x_p}\left\{\phi v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right\}\frac{1}{\{g'(\mu_t(\delta))\}^2}\beta_p x_{ti},$$

em que $v_t(\delta) = \psi'((1-\mu_t(\delta))\phi) + \psi'(\mu_t(\delta)\phi)$. Desta forma, obtemos, para $i \neq p$

$$\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \beta_i} = -\phi s_{x_p}\beta_p x_{ti} q_t(\delta), \tag{3.A5}$$

em que

$$q_t(\delta) = \left\{\phi v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right\}\frac{1}{\{g'(\mu_t(\delta))\}^2}.$$

Com base em (3.A5) e considerando $i = p$, segue que

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial \delta_t \partial \beta_i} = &- \phi s_{x_p}\left\{\phi v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right\}\frac{1}{\{g'(\mu_t(\delta))\}^2}\beta_p(x_{tp} + \delta_t s_{x_p}) \\ &+ \phi s_{x_p}(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\end{aligned}$$

Reescrevendo essa expressão temos que

$$\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\beta_p} = -\phi s_{x_p}\left[\beta_p(x_{tp}+\delta_t s_{x_p})q_t(\delta) - (y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\right]. \tag{3.A6}$$

Ainda de (3.A3), segue que

$$\begin{aligned}
&\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\phi} = [\psi((1-\mu_t(\delta))\phi) - \psi(\mu_t(\delta)\phi) + y_t^*]\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p} \\
&+ \phi[\psi'((1-\mu_t(\delta))\phi)(1-\mu_t(\delta)) - \psi'(\mu_t(\delta)\phi)\mu_t(\delta)]\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p} = \\
&\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p}\left\{(y_t^* - \mu_t^*(\delta)) + \phi[\psi'((1-\mu_t(\delta))\phi)(1-\mu_t(\delta)) - \psi'(\mu_t(\delta)\phi)\mu_t(\delta)]\right\},
\end{aligned}$$

que pode ser reescrita como

$$\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\phi} = -\beta_p s_{x_p} f_t(\delta)\frac{1}{g'(\mu_t(\delta))}, \tag{3.A7}$$

em que, considerando o esquema de perturbação da covariável x_p,

$$f_t(\delta) = \phi[\psi'(\mu_t(\delta)\phi)\mu_t(\delta) - \psi'((1-\mu_t(\delta))\phi)(1-\mu_t(\delta))] - (y_t^* - \mu_t^*(\delta)) = c_t(\delta) - (y_t^* - \mu_t^*(\delta)),$$

com $c_t(\delta) = \phi[\psi'(\mu_t(\delta)\phi)\mu_t(\delta) - \psi'((1-\mu_t(\delta))\phi)(1-\mu_t(\delta))]$.

Note que as quantidades que estão em função de $\mu(\delta)$, de fato, dependem de $\eta(\delta)$ definido em (3.20). Avaliando as expressões (3.A4), (3.A6) e (3.A7) em $\delta_0 = (0,0,\ldots,0)^\top$ e $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ e considerando que $\eta(\delta_0) = \eta$, obtemos $\Delta_\beta = -\widehat{\phi} s_{x_p}[\widehat{\beta}_p X^\top \widehat{Q} - P\widehat{T}\mathcal{E}]$ e $\Delta_\phi = -\widehat{\beta}_p s_{x_p}\widehat{f}^\top \widehat{T}$, em que $\widehat{Q}$, P e $\widehat{f}^\top$ estão definidos na Seção 3.4.3. Assim, a matriz Δ fica expressa na forma apresentada em (3.21).

3.A.2.4. Perturbação no parâmetro de precisão

Com base na definição do parâmetro de precisão sob a perturbação multiplicativa definida em (3.22) e considerando (1.3) temos que o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}
\ell_\delta(\mu,\phi) = &\sum_{t=1}^n \left\{\log\Gamma\left(\frac{\phi}{\delta_t}\right) - \log\Gamma\left(\frac{\mu_t\phi}{\delta_t}\right) - \log\Gamma\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) + \left(\frac{\mu_t\phi}{\delta_t} - 1\right)\log(y_t)\right\} \\
&+ \sum_{t=1}^n \left\{\left[\frac{(1-\mu_t)\phi}{\delta_t} - 1\right]\log(1-y_t)\right\}.
\end{aligned}$$

Assim, temos que

$$\frac{\partial \ell(\beta,\phi)}{\partial\delta_t} = -\frac{\phi}{\delta_t^2}\left[\psi\left(\frac{\phi}{\delta_t}\right) - \mu_t\psi\left(\frac{\mu_t\phi}{\delta_t}\right) - (1-\mu_t)\psi\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) + \mu_t y_t^* + \log(1-y_t)\right] \tag{3.A8}$$

Segue de (3.A8) que

$$\begin{aligned}
&\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\beta_i} = -\frac{\phi}{\delta_t^2}\frac{1}{g'(\mu_t)}x_{ti}\times \\
&\left\{y_t^* + \psi\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) - \psi\left(\frac{\mu_t\phi}{\delta_t}\right) - \frac{\mu_t\phi}{\delta_t}\psi'\left(\frac{\mu_t\phi}{\delta_t}\right) + \frac{(1-\mu_t)\phi}{\delta_t}\psi'\left(\frac{(1-\mu_t)\phi}{\delta_t}\right)\right\} \quad (3.A9)\\
&= -\frac{\phi}{\delta_t^2}\frac{1}{g'(\mu_t)}x_{ti}\left\{(y_t^* - \mu_t^*(\delta)) - \frac{\phi}{\delta_t}\left[\psi'\left(\frac{\mu_t\phi}{\delta_t}\right)\mu_t - \psi'\left(\frac{(1-\mu_t)\phi}{\delta_t}\right)(1-\mu_t)\right]\right\},
\end{aligned}$$

em que, neste caso, $\mu_t^*(\delta) = \psi\left(\mu_t\phi/\delta_t\right) - \psi\left((1-\mu_t)\phi/\delta_t\right)$. Podemos reescrever (A9) como

$$\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\beta_i} = \frac{\phi}{\delta_t^2}\frac{1}{g'(\mu_t)}x_{ti}f_t(\delta) \quad (3.A10)$$

em que, para este esquema de perturbação,

$$f_t(\delta) = \frac{\phi}{\delta_t}\left[\psi'\left(\frac{\mu_t\phi}{\delta_t}\right)\mu_t - \psi'\left(\frac{(1-\mu_t)\phi}{\delta_t}\right)(1-\mu_t)\right] - (y_t^* - \mu_t^*(\delta)).$$

Com base em (3.A8) temos ainda que

$$\begin{aligned}
\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\phi} =& -\frac{1}{\delta_t^2}\left[\psi\left(\frac{\phi}{\delta_t}\right) - \mu_t\psi\left(\frac{\mu_t\phi}{\delta_t}\right) - (1-\mu_t)\psi\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) + \mu_t y_t^* + \log(1-y_t)\right] \\
&- \frac{\phi}{\delta_t^2}\left[\frac{1}{\delta_t}\psi'\left(\frac{\phi}{\delta_t}\right) - \frac{\mu_t^2}{\delta_t}\psi'\left(\frac{\mu_t\phi}{\delta_t}\right) - \frac{(1-\mu_t)^2}{\delta_t}\psi'\left(\frac{(1-\mu_t)\phi}{\delta_t}\right)\right],
\end{aligned}$$

que pode ser escrito como

$$\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\phi} = -\frac{1}{\delta_t^2}\left[\mu_t(y_t^* - \mu_t^*(\delta)) - \psi\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) + \psi\left(\frac{\phi}{\delta_t}\right) + \log(1-y_t)\right] + \frac{\phi d_t(\delta)}{\delta_t^2},$$

em que, neste caso,

$$d_t(\delta) = \left[\frac{\mu_t^2}{\delta_t}\psi'\left(\frac{\mu_t\phi}{\delta_t}\right) + \frac{(1-\mu_t)^2}{\delta_t}\psi'\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) - \frac{1}{\delta_t}\psi'\left(\frac{\phi}{\delta_t}\right)\right].$$

Finalmente, temos que

$$\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\phi} = \frac{1}{\delta_t^2}[\phi d_t(\delta) - u_t(\delta)], \quad (3.A11)$$

com

$$u_t(\delta) = \mu_t(y_t^* - \mu_t^*(\delta)) + \log(1-y_t) - \psi\left(\frac{(1-\mu_t)\phi}{\delta_t}\right) + \psi\left(\frac{\phi}{\delta_t}\right).$$

Avaliando as expressões (3.A10) e (3.A11) em $\delta_0 = (1,1,\ldots,1)^\top$ e $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ podemos definir $\Delta_\beta = -\widehat{\phi}X^\top\widehat{T}\widehat{F}$ e $\Delta_\phi = \widehat{\phi}\widehat{d}^\top - \widehat{u}^\top$, em que $\widehat{F}$ e $\widehat{d}$ estão definidos na Seção 3.4.4. Assim, a matriz Δ fica expressa na forma apresentada em (3.23).

Capítulo 4

Modelo de regressão beta com dispersão variável

4.1. Introdução

O modelo de regressão beta proposto por Ferrari e Cribari–Neto (2004) assume que o parâmetro de precisão é uma constante na função de variância. Isto implica dizer que a dispersão (inverso da precisão) é constante para todas as observações. No entanto, as perdas de eficiência em usar modelos com dispersão constante, quando na verdade a dispersão é variável, pode ser substancial. De fato, a estimação eficiente dos parâmetros em uma regressão depende da modelagem correta da dispersão. Muitos autores têm considerado a modelagem da dispersão para dados normais. No contexto de modelos lineares generalizados, Smyth e Verbyla (1999) definem os modelos lineares generalizados duplos que permitem que a média e a variância sejam modeladas simultaneamente.

Neste contexto, apresentamos um modelo de regressão beta em que o parâmetro de dispersão varia com as observações, havendo assim uma estrutura heteroscedástica. Note-se que mesmo que ϕ seja constante ao longo das observações, as variâncias de $y_1, \ldots, y_n$ não serão constantes, pois dependerão das médias desconhecidas, estas variando de acordo com uma estrutura de regressão. Assim, o conceito de heteroscedasticidade no presente contexto difere daquele empregado em modelos lineares normais de regressão, em que, sob homoscedasticidade as variâncias condicionais são constantes. De fato, segundo Houaiss (2001), homoscedasticidade é a propriedade de apresentar a mesma variância ou dispersão. Em modelos normais, a medida de dispersão tipicamente utilizada é a variância, logo as duas medidas não se confundem. No entanto, nos modelos da família exponencial, homoscedasticidade significa que o parâmetro de dispersão é o mesmo para todas as observações.

Ressaltamos que alguns autores já apresentaram modelos de regressão baseados na distribuição beta em que a média e a dispersão são modeladas simultaneamente. Paolino (2001) e Smithson e Verkuilen (2006) utilizam máxima verossimilhança para estimar os parâmetros dos modelos da média e da dispersão. Smithson e Verkuilen (2006) apresentam, adicionalmente, testes de hipóteses tanto para o modelo da média quanto para o modelo da precisão. Cuervo e Gamerman (2004) propõem um método bayesiano para estimar parâmetros de distribuições que pertencem a famílias exponenciais biparamétricas e, naquele contexto, os autores modelam a média e dispersão da distribuição beta. Neste capítulo apresentamos

uma extensão do modelo proposto por Ferrari e Cribari–Neto (2004) para situações em que o parâmetro de precisão ϕ não é constante ao longo das observações. A precisão será modelada em termos de covariáveis e de parâmetros desconhecidos da mesma forma que a média condicional. A estimação conjunta de todos os parâmetros do modelo será realizada utilizando máxima verossimilhança. Apresentaremos, ainda, expressões para a função escore, para a matriz de informação de Fisher e para sua inversa. Adicionalmente, serão apresentados testes para a hipótese nula de homocedasticidade.

Apesar deste tema não ser inédito, apresentaremos aqui alguns aspectos ainda não explorados na literatura. Considerando como critério de avaliação as taxas de coberturas empíricas de intervalos de confiança para a média da resposta, avaliamos através de estudos de simulação, ao mesmo tempo, as propriedades dos estimadores de máxima verossimilhança do modelo beta com dispersão variável e as conseqüências de estimar o modelo supondo dispersão constante quando de fato ela é variável. Avaliamos ainda o desempenho quanto ao tamanho e quanto ao poder dos testes assintóticos da razão de verossimilhanças, escore e Wald para testar a hipótese de dispersão constante. Finalmente, sugerimos técnicas de diagnósticos no sentido de identificar a dispersão variável dos dados.

4.2. Definição e estimação do modelo

Nesta seção apresentamos uma extensão do modelo de regressão beta (Ferrari e Cribari–Neto, 2004) em que o parâmetro de precisão varia ao longo das observações. Mais precisamente, em adição à modelagem da média da variável resposta, definimos uma estrutura de regressão para o parâmetro de precisão.

A partir de agora assumimos que $y_1, \ldots, y_n$ são variáveis aleatórias independentes, em que cada y_t, $t = 1, \ldots, n$, segue a densidade em (1.1) com média μ_t definida em (1.2) e precisão ϕ_t; admitimos ainda que

$$h(\phi_t) = \sum_{j=1}^{q} z_{tj}\gamma_j = \vartheta_t, \tag{4.1}$$

em que $\gamma = (\gamma_1, \ldots, \gamma_q)^\top$ é um vetor de parâmetros desconhecido ($\gamma \in I\!R^q$), $z_{t1}, \ldots, z_{tq}$ são observações de q covariáveis ($q < n$), assumidas fixas e conhecidas e $h(\cdot)$ é uma função estritamente monótona e duas vezes diferenciável.

O logaritmo da função de verossimilhança é

$$\ell(\beta, \gamma) = \sum_{t=1}^{n} \ell_t(\mu_t, \phi_t),$$

com

$$\begin{aligned}\ell_t(\mu_t,\phi_t) &= \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)\log y_t \\ &\quad + \{(1-\mu_t)\phi_t - 1\}\log(1-y_t).\end{aligned} \tag{4.2}$$

Logo, para $i = 1, \ldots, k$, a função escore para β_i é dada por

$$\frac{\partial\ell(\beta,\gamma)}{\partial\beta_i} = \sum_{t=1}^{n} \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\frac{\partial\eta_t}{\partial\beta_i}, \tag{4.3}$$

com $\mathrm{d}\mu_t/\mathrm{d}\eta_t = 1/g'(\mu_t)$ e

$$\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t} = \phi_t\left[\log\frac{y_t}{1-y_t} - \{\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)\}\right]. \tag{4.4}$$

Definimos

$$\mu_t^* = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t) \tag{4.5}$$

e, conseqüentemente,

$$\frac{\partial\ell(\beta,\gamma)}{\partial\beta_i} = \sum_{t=1}^{n} \phi_t(y_t^* - \mu_t^*)\frac{1}{g'(\mu_t)}x_{ti}, \tag{4.6}$$

em que y_t^* é o logito de y_t, dado em (1.5). Podemos escrever o vetor $U_\beta(\beta,\gamma)$, função escore de $\beta = (\beta_1, \ldots, \beta_k)^\top$, através da seguinte expressão matricial:

$$U_\beta(\beta,\gamma) = X^\top\Phi T(y^* - \mu^*),$$

em que

$$\Phi = \mathrm{diag}\{\phi_1, \ldots, \phi_n\} \tag{4.7}$$

e $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$.

No contexto em que ϕ varia ao longo das observações, a densidade beta (1.1) escrita na forma da família exponencial biparamétrica canônica é dada por

$$f(y_t;\mu_t,\phi_t) = \exp\{\tau_1 T_1 + \tau_2 T_2 - \mathcal{A}(\tau)\}(1/y_t(1-y_t)),$$

em que $\tau = (\tau_1,\tau_2) = (\mu_t\phi_t, \phi_t)$, $(T_1, T_2) = (\log\{y_t/(1-y_t)\}, \log(1-y_t))$ e

$$\mathcal{A}(\tau) = -\log\Gamma(\phi_t) + \log\Gamma(\mu_t\phi_t) + \log\Gamma((1-\mu_t)\phi_t).$$

Assim, segue que

$$\mathrm{E}(T_1) = \mathrm{E}(y_t^*) = \partial\mathcal{A}(\tau)/\partial\tau_1 = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t) = \mu_t^* \tag{4.8}$$

e

$$\mathrm{E}(T_2) = \mathrm{E}(\log(1-y_t)) = \partial\mathcal{A}(\tau)/\partial\tau_2 = -\psi(\phi_t) + \psi((1-\mu_t)\phi_t). \tag{4.9}$$

Note que (4.8) é equivalente a $\mathrm{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\mu_t)=0$, ou ainda, $\mathrm{E}(U_\beta(\beta,\gamma))=0$.

Consideremos agora as derivadas do logaritmo da função de verossimilhança em relação aos parâmetros que modelam a precisão, γ_j, $j=1,\ldots,q$. Temos que

$$\frac{\partial\ell(\beta,\gamma)}{\partial\gamma_j}=\sum_{t=1}^{n}\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\frac{\partial\vartheta_t}{\partial\gamma_j}, \tag{4.10}$$

sendo que $\mathrm{d}\phi_t/\mathrm{d}\vartheta_t=1/h'(\phi_t)$. Também,

$$\begin{aligned}\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}=&\mu_t\left[\log\frac{y_t}{1-y_t}-(\psi(\mu_t\phi_t)-\psi((1-\mu_t)\phi_t))\right]+\\&\log(1-y_t)-\psi((1-\mu_t)\phi_t)+\psi(\phi_t).\end{aligned} \tag{4.11}$$

Utilizando as definições de y_t^* em (1.5) e de μ_t^* em (4.5) chegamos a

$$\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}=\mu_t(y_t^*-\mu_t^*)+\log(1-y_t)-\psi((1-\mu_t)\phi_t)+\psi(\phi_t).$$

Assim, a função escore para cada um dos parâmetros γ_j é dada por

$$\frac{\partial\ell(\beta,\gamma)}{\partial\gamma_j}=\sum_{t=1}^{n}[\mu_t(y_t^*-\mu_t^*)+\log(1-y_t)-\psi((1-\mu_t)\phi_t)+\psi(\phi_t)]\frac{1}{h'(\phi_t)}z_{tj},$$

que pode ser expressa como

$$\frac{\partial\ell(\beta,\gamma)}{\partial\gamma_j}=\sum_{t=1}^{n}a_t\frac{1}{h'(\phi_t)}z_{tj},$$

em que

$$a_t=\mu_t(y_t^*-\mu_t^*)+\log(1-y_t)-\psi((1-\mu_t)\phi_t)+\psi(\phi_t). \tag{4.12}$$

A função escore para $\gamma=(\gamma_1,\ldots,\gamma_q)^\top$ pode ser expressa em forma matricial como

$$U_\gamma(\beta,\gamma)=Z^\top Ha,$$

em que Z é uma matriz $n\times q$ cuja t-ésima linha é $z_t^\top$,

$$H=\mathrm{diag}\{1/h'(\phi_1),\ldots,1/h'(\phi_n)\} \tag{4.13}$$

e $a=(a_1,\ldots,a_n)^\top$. Note que (4.9) é equivalente a $\mathrm{E}(a)=0$, ou ainda, $\mathrm{E}(U_\gamma(\beta,\gamma))=0$.

Com o objetivo de construir a matriz de informação de Fisher conjunta dos vetores β e γ passaremos aos cálculos das segundas derivadas do logaritmo da função de verossimilhança. A partir de (4.3) temos que para $i,p=1,\ldots,k$

$$\begin{aligned}\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\beta_p}&=\sum_{t=1}^{n}\frac{\partial}{\partial\mu_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\right)\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\frac{\partial\eta_t}{\partial\beta_j}x_{ti}\\&=\sum_{t=1}^{n}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial}{\partial\mu_t}\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\right)\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}x_{ti}x_{tp}.\end{aligned}$$

Uma vez que $\mathrm{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\mu_t)=0$,

$$\mathrm{E}\left(\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\beta_p}\right)=\sum_{t=1}^{n}\mathrm{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\right)\left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\right)^2 x_{ti}x_{tp}.$$

De (4.4) segue que

$$\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}=-\phi_t^2\{\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)\},$$

definindo

$$w_t=\phi_t\left\{\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)\right\}\frac{1}{\{g'(\mu_t)\}^2}, \tag{4.14}$$

temos que

$$\mathrm{E}\left(\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\beta_p}\right)=-\sum_{t=1}^{n}\phi_t w_t x_{ti}x_{tp}, \tag{4.15}$$

A expressão em (4.15) pode ser escrita em notação matricial como

$$\mathrm{E}\left(\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta\partial\beta^\top}\right)=-X^\top\Phi W X,$$

em que $W=\mathrm{diag}\{w_1,\ldots,w_n\}$.

Adicionalmente, segue de (4.6) que as segundas derivadas de $\ell(\beta,\gamma)$ com respeito a β_i e γ_j, para $i=1,\ldots,k$ e $j=1,\ldots,q$, são dadas por

$$\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\gamma_j}=\sum_{t=1}^{n}\frac{\partial}{\partial\phi_t}\left(\phi_t(y_t^*-\mu_t^*)\frac{1}{g'(\mu_t)}x_{ti}\right)\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\frac{\partial\vartheta_t}{\partial\gamma_j}.$$

Como

$$\begin{aligned}\frac{\partial}{\partial\phi_t}\left(\phi_t(y_t^*-\mu_t^*)\frac{1}{g'(\mu_t)}x_{ti}\right)&=\left\{(y_t^*-\mu_t^*)-\phi_t\left[\psi'(\mu_t\phi_t)\mu_t-\psi'((1-\mu_t)\phi_t)(1-\mu_t)\right]\right\}\\&\quad\times\frac{1}{g'(\mu_t)}x_{ti},\end{aligned}$$

chegamos a

$$\begin{aligned}\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\gamma_j}&=\sum_{t=1}^{n}\left\{(y_t^*-\mu_t^*)-\phi_t\left[\psi'(\mu_t\phi_t)\mu_t-\psi'((1-\mu_t)\phi_t)(1-\mu_t)\right]\right\}\\&\quad\times\frac{1}{g'(\mu_t)}\frac{1}{h'(\phi_t)}x_{ti}z_{tj}.\end{aligned}$$

Uma vez que $\mathrm{E}(y_t^*-\mu_t^*)=0$ temos

$$\mathrm{E}\left(\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_i\partial\gamma_j}\right)=-\sum_{t=1}^{n}c_t\frac{1}{g'(\mu_t)}\frac{1}{h'(\phi_t)}x_{ti}z_{tj},$$

que em notação matricial pode ser escrito como

$$\mathrm{E}\left(\frac{\partial^2 \ell(\beta,\gamma)}{\partial\beta\partial\gamma}\right) = -X^\top CTHZ,$$

em que $C = \mathrm{diag}\{c_1, \ldots, c_n\}$, com

$$c_t = \phi_t \left[\psi'(\mu_t\phi_t)\mu_t - \psi'((1-\mu_t)\phi_t)(1-\mu_t)\right]. \tag{4.16}$$

Agora, de (4.10) podemos obter as segundas derivadas de $\ell(\beta,\gamma)$ com respeito a γ_j e γ_l, para $j, l = 1, \ldots, q$, dadas por

$$\frac{\partial^2 \ell(\beta,\gamma)}{\partial\gamma_j\partial\gamma_l} = \sum_{t=1}^{n} \frac{\partial}{\partial\phi_t}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\right)\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\frac{\partial\vartheta_t}{\partial\gamma_l} z_{tj} =$$

$$\sum_{t=1}^{n}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t} + \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial}{\partial\phi_t}\left(\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\right)\right)\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t} z_{tl} z_{tj}.$$

Mas $\mathrm{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\phi_t) = 0$ e, portanto,

$$\mathrm{E}\left(\frac{\partial^2 \ell(\beta,\gamma)}{\partial\gamma_j\partial\gamma_l}\right) = \sum_{t=1}^{n} \mathrm{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\right)\left(\frac{\mathrm{d}\phi_t}{\mathrm{d}\vartheta_t}\right)^2 z_{tl} z_{tj}.$$

Adicionalmente, a partir de (4.11) podemos obter

$$\mathrm{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\right) = -\left[\psi'(\mu_t\phi_t)\mu_t^2 + \psi'((1-\mu_t)\phi_t)(1-\mu_t)^2 - \psi'(\phi_t)\right].$$

Assim,

$$\begin{aligned}\mathrm{E}\left(\frac{\partial^2 \ell(\beta,\gamma)}{\partial\gamma_j\partial\gamma_l}\right) = &-\sum_{t=1}^{n}\left[\psi'(\mu_t\phi_t)\mu_t^2 + \psi'((1-\mu_t)\phi_t)(1-\mu_t)^2 - \psi'(\phi_t)\right] \\ &\times \frac{1}{\{h'(\phi_t)\}^2} z_{tl} z_{tj},\end{aligned}$$

que em notação matricial passa a ser

$$\mathrm{E}\left(\frac{\partial^2 \ell(\beta,\gamma)}{\partial\gamma\partial\gamma^\top}\right) = -Z^\top D^* Z,$$

em que $D^* = \mathrm{diag}\{d_1^*, \ldots, d_n^*\}$, com

$$d_t^* = \left[\psi'(\mu_t\phi_t)\mu_t^2 + \psi'((1-\mu_t)\phi_t)(1-\mu_t)^2 - \psi'(\phi_t)\right]\frac{1}{\{h'(\phi_t)\}^2}. \tag{4.17}$$

Portanto, a matriz de informação de Fisher é dada por

$$K^* = K^*(\beta,\gamma) = \begin{pmatrix} K^*_{\beta\beta} & K^*_{\beta\gamma} \\ K^*_{\gamma\beta} & K^*_{\gamma\gamma}\end{pmatrix},$$

em que $K^*_{\beta\beta} = X^\top \Phi W X$, $K^*_{\beta\gamma} = {K^*_{\gamma\beta}}^\top = X^\top CTHZ$ e $K^*_{\gamma\gamma} = Z^\top D^* Z$.

No caso em que $\phi_1 = \ldots = \phi_n = \phi$ (homoscedasticidade) tomamos $h(\phi_t) = \phi$ para $t = 1, \ldots, n$. Portanto, segue que $h'(\phi_t) = 1$, $Z = (1, 1, 1, \ldots, 1)^\top$ e $\Phi = \phi \mathcal{I}_n$. Assim, as expressões em (4.14), (4.16) e (4.17) passam a ser, respectivamente,

$$w_t = \phi \left[\psi'(\mu_t \phi) + \psi'((1-\mu)\phi)\right] \frac{1}{\{g'(\mu_t)\}^2},$$

$$c_t = \phi \left[\psi'(\mu_t \phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\right],$$

e

$$d_t^* = \left[\psi'(\mu_t \phi)\mu_t^2 + \psi'((1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi)\right].$$

Adicionalmente, considerando as definições de T, c, d_t e D apresentadas no Capítulo 1 chegamos a três resultados. Primeiro, $X^\top \Phi W X = \phi X^\top W X$. Segundo, $CTHZ = Tc$, e, portanto, $X^\top CTHZ = X^\top Tc$. Terceiro, $d_t^* = d_t$ e $Z^\top D^* Z = \sum_{t=1}^n d_t^* = \sum_{t=1}^n d_t = \mathrm{tr}(D)$. Assim, podemos verificar que a matriz de informação de Fisher do modelo heteroscedástico se reduz, como esperado, à matriz de informação de Fisher do caso homoscedástico, quando $\phi_1 = \ldots = \phi_n = \phi$.

Sob condições de regularidade, temos que para tamanhos de amostras grandes, a distribuição aproximada conjunta de $\widehat{\beta}$ e $\widehat{\gamma}$ é normal $(k+q)$ multivariada, de forma que

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} \sim \mathcal{N}_{k+q}\left(\begin{pmatrix} \beta \\ \gamma \end{pmatrix}, K^{*-1}\right). \tag{4.18}$$

em que

$$K^{*-1} = K^{*-1}(\beta, \gamma) = \begin{pmatrix} K_*^{\beta\beta} & K_*^{\beta\gamma} \\ K_*^{\gamma\beta} & K_*^{\gamma\gamma} \end{pmatrix},$$

com

$$K_*^{\beta\beta} = \left(X^\top \Phi W X - X^\top CTHZ(Z^\top D^* Z)^{-1} Z^\top HTC^\top X\right)^{-1},$$

$$K_*^{\beta\gamma} = (K_*^{\gamma\beta})^\top = -K_*^{\beta\beta} X^\top CTHZ(Z^\top D^* Z)^{-1}$$

e

$$K_*^{\gamma\gamma} = (Z^\top D^* Z)^{-1}\left\{\mathcal{I}_q + (Z^\top HTC^\top X) K_*^{\beta\beta} X^\top CTHZ(Z^\top D^* Z)^{-1}\right\}.$$

Aqui, $\mathcal{I}_q$ é matriz identidade de ordem q.

Devemos ressaltar que os estimadores de máxima verossimilhaça de β e γ são obtidos através de métodos iterativos, tal como o algoritmo quasi-Newton. Os algoritmos de otimização requerem a especificação de um valor de $\theta = (\beta^\top, \gamma^\top)^\top$ para inicializar o processo iterativo. No modelo homoscedástico, em que $\theta = (\beta^\top, \phi)^\top$, Ferrari e Cribari–Neto (2004)

sugerem utilizar como valor inicial de β no processo iterativo a estimativa de mínimos quadrados ordinários desse vetor obtida de uma regressão linear em que a variável resposta é transformada através da função de ligação $g(\cdot)$. Desta forma, as observações da variável resposta transformada são dadas por $\varpi = (g(y_1), \ldots, g(y_n))^\top$ e o estimador de mínimos quadrados ordinários é $(X^\top X)^{-1}X^\top\varpi$. Adicionalmente, uma vez que $\mathrm{var}(y_t) = \mu_t(1-\mu_t)/(1+\phi)$, isto é, $\phi = \mu_t(1-\mu_t)/\mathrm{var}(y_t) - 1$, o valor inicial para ϕ sugerido pelos autores é

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\check{\mu}_t(1-\check{\mu}_t)}{\check{\sigma}_t^2} - 1,$$

em que $\check{\mu}_t$ é obtido aplicando-se $g^{-1}(\cdot)$ ao t-ésimo valor ajustado da regressão linear de $g(y_1)$, $\ldots, g(y_n)$ em X, i.e.,

$$\check{\mu}_t = g^{-1}(x_t^\top(X^\top X)^{-1}X^\top\varpi)$$

e

$$\check{\sigma}_t^2 = \check{e}^\top\check{e}/[(n-k)\{g'(\check{\mu}_t)\}^2]. \tag{4.19}$$

Aqui, $\check{e} = \varpi - X(X^\top X)^{-1}X^\top\varpi$ é o vetor de resíduos de mínimos quadrados ordinários de uma regressão linear com a variável resposta transformada. Note que a expressão em (4.19) está baseada no fato de que

$$\mathrm{var}(g(y_t)) \approx \mathrm{var}\{g(\mu_t) + (y_t - \mu_t)g'(\mu_t)\} = \mathrm{var}(y_t)\{g'(\mu_t)\}^2,$$

ou seja, $\mathrm{var}(y_t) \approx \mathrm{var}\{g(y_t)\}\{g'(\mu_t)\}^{-2}$.

No modelo beta heteroscedástico a estimativa inicial para β pode ser a mesma. No que se refere a $\phi_1, \ldots, \phi_n$, sugerimos os seguintes valores iniciais:

$$\check{\phi}_t = \frac{\check{\mu}_t(1-\check{\mu}_t)}{\check{\sigma}_t^2} - 1.$$

Dado que $h(\phi_t) = \sum_{j=1}^{q} z_{tj}\gamma_j$, a estimativa inicial de γ pode ser $(Z^\top Z)^{-1}Z^\top\kappa$, em que $\kappa = (h(\check{\phi}_1), \ldots, h(\check{\phi}_n))^\top$.

4.3. Estimação intervalar

Através de simulações de Monte Carlo avaliamos em amostras de tamanho finito o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta, considerando que a média e a dispersão são modeladas simultaneamente. Todo o procedimento de cálculo foi programado na linguagem de programação matricial `Ox` (Doornik, 2001). O número de réplicas de Monte Carlo foi $R = 5000$.

Decidimos avaliar o comportamento dos estimadores de máxima verossimilhança quanto à estimação intervalar da média da resposta. Geramos dados com dispersão variável ao

longo das observações e consideramos dois contextos de estimação. Primeiro estimamos o modelo supondo dispersão constante; chamamos este processo de estimação incorreta. Em seguida, estimamos o modelo supondo dispersão variável; chamamos este processo de estimação correta. O objetivo é ao mesmo tempo avaliar a qualidade da estimação intervalar para a média da resposta no modelo beta com dispersão variável e quais as conseqüências do processo incorreto de estimação.

Com base em (4.18) e utilizando o método delta podemos construir intervalos de confiança para a média da variável resposta. Por exemplo, o intervalo de confiança assintótico de tamanho 95% para a resposta média dado um vetor de covariáveis x, é dado por

$$\left(g^{-1}\Big(\widehat{\eta} - 1.96 \text{ e.p.}(\widehat{\eta})\Big), \quad g^{-1}\Big(\widehat{\eta} + 1.96 \text{ e.p.}(\widehat{\eta})\Big)\right),$$

em que $\widehat{\eta} = x^\top \widehat{\beta}$ é o valor do erro padrão e.p.$(\widehat{\eta}) = \sqrt{(x^\top \widehat{\mathcal{K}}^{\beta\beta} x)}$. Se o processo de estimação supõe um modelo de regressão beta com dispersão constante $\mathcal{K}^{\beta\beta} = \widehat{K}^{\beta\beta} = K^{\beta\beta}(\widehat{\theta})$, sendo $K^{\beta\beta}(\widehat{\theta})$ a inversa da matriz de informação de Fisher de β definida em (1.14) avaliada em $\widehat{\theta}$. Se o processo de estimação se baseia no modelo de regressão beta com dispersão variável temos que $\mathcal{K}^{\beta\beta} = K_*^{\beta\beta}(\widehat{\theta})$, sendo $K_*^{\beta\beta}(\widehat{\theta})$ a inversa da matriz de informação de Fisher de β definida em (4.18) avaliada em $\widehat{\theta}$.

Consideraremos

$$\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t \tag{4.20}$$

e

$$\log(\phi_t) = \gamma_1 + \gamma_2 z_t, \tag{4.21}$$

$t = 1, \ldots, n$. Para esta investigação usamos $z_t = x_t$ e $x_1 \ldots, x_n$ são obtidos como uma amostra aleatória da distribuição $\mathcal{U}(0,1)$. Definimos uma medida para o *grau de heterogeneidade da dispersão* dos dados como

$$\lambda = \max(\phi_t)/\min(\phi_t), \tag{4.22}$$

$t = 1, \ldots, n$. Assim, $\lambda = 1$ indica que a dispersão é constante ao longo das obervações. Consideramos dois cenários diferentes. No primeiro cenário utilizamos $\beta_1 = 4.1$ e $\beta_2 = -0.8$, o que conduz a valores da média da variável resposta próximos de um, i.e., $\mu \in (0.96, 0.98)$. No segundo cenário, os verdadeiros valores dos parâmetros são $\beta_1 = -2.5$ e $\beta_2 = -1.2$, que conduzem a $\mu \in (0.024, 0.075)$, valores da média próximos de zero. Nos dois cenários admitimos $\gamma_1 = 5.0$ e variamos o valor de γ_2 para obter diferentes valores de λ. Desta forma, fixamos $\gamma_2 = -2.5$, $\gamma_2 = -3.5$, $\gamma_2 = -4.5$ e $\gamma_2 = -4.9$ o que conduz a valores $\lambda \approx 10$, $\lambda \approx 27$, $\lambda \approx 66$ e $\lambda \approx 109$, respectivamente. Consideramos $n = 20, 40$ e 60, sendo que são geradas

20 observações da variável x_t, que são replicadas duas e três vezes, respectivamente, a fim de se obter os tamanhos amostrais $n = 40, 60$. Este procedimento de replicação de valores das covariáveis assegura que o grau de heterogeneidade da dispersão dos dados mantém-se constante à medida que o tamanho da amostra cresce.

Para a análise de resultados de estimação intervalar calculamos a taxa de cobertura empírica de intervalos com nível de confiança nominal de 95% para a média da resposta, ou seja, intervalos de confiança assintóticos baseados no quantil 97.5% da distribuição normal padrão. Construímos gráficos considerando as diferenças (discrepâncias) entre o nível nominal e as taxas de cobertura estimadas, segundo a estimação correta e incorreta. A qualidade da estimação pode ser medida pela proximidade das discrepâncias do zero.

Na Figura 4.1 estão as discrepâncias entre as taxas de cobertura nominal e empírica, considerando a estimação correta e incorreta, no cenário em que a média da resposta está próxima de um. A Figura 4.2 considera o cenário em que a média da resposta está próxima de zero. As duas figuras apresentam a mesma organização; em cada linha é considerado um valor de λ, grau de heterogeneidade da dispersão dos dados, e em cada coluna em valor de n.

A análise das Figuras 4.1 e 4.2 revela que quando o modelo é estimado corretamente, ou seja, considerando que os dados apresentam uma dispersão que varia ao longo das observações, as taxas de cobertura empírica e nominal são muito próximas e assim as discrepâncias ficam próximas de zero, em especial quando o tamanho da amostra aumenta. No entanto, quando o modelo é estimado supondo dispersão constante quando de fato a dispersão é variável, as diferenças entre as taxas de cobertura nominal e empírica são consideráveis, em especial quando o valor de λ aumenta. Adicionalmente, fixando o valor de λ nota-se que a medida que o tamanho da amostra cresce as discrepâncias aumentam, e se afastam de zero. Para um grau de heterogeneidade $\lambda \approx 66$ as discrepâncias são próximas de -80% e quando $\lambda \approx 109$ este valor é próximo do máximo possível -95%.

A partir destas impressões podemos concluir que os estimadores de máxima verossimilhança para o modelo de regressão beta com dispersão variável apresentam um bom desempenho quanto à estimação intervalar da média da resposta e que estimar o modelo supondo dispersão constante quando de fato a dispersão é variável interfere drasticamente na precisão dos intervalos de confiança para a média. Este argumento justifica a modelagem da dispersão sempre que houver algum indício de dispersão não constante.

4.4. Testes de heteroscedasticidade

Uma vez que estamos considerando a possibilidade de um modelo de regressão beta em que a precisão depende de covariáveis e de parâmetros desconhecidos, parece razoável desenvolver

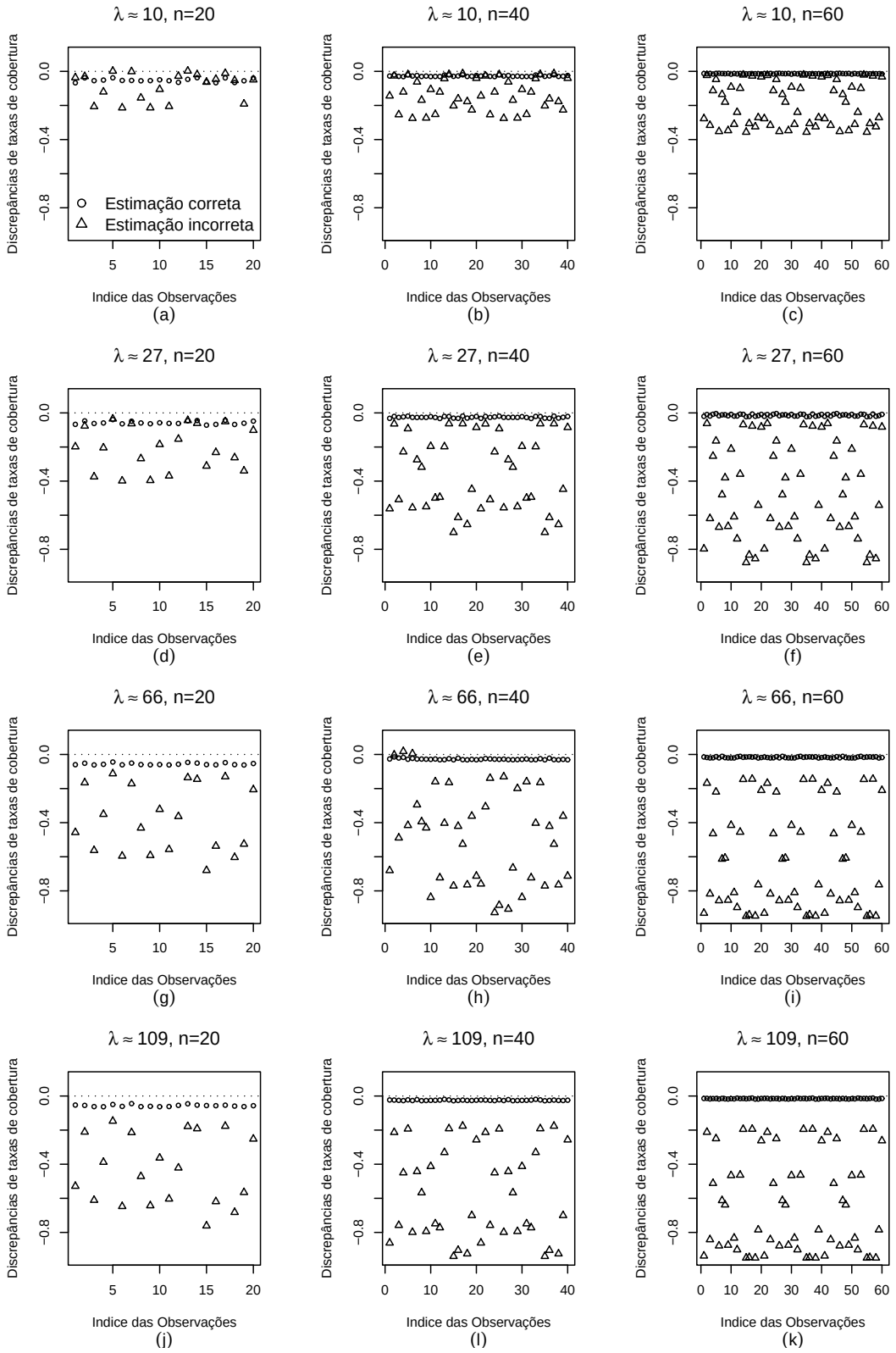

Figura 4.1. Discrepâncias de taxas de cobertura. $\mu \in (0.96, 0.98)$

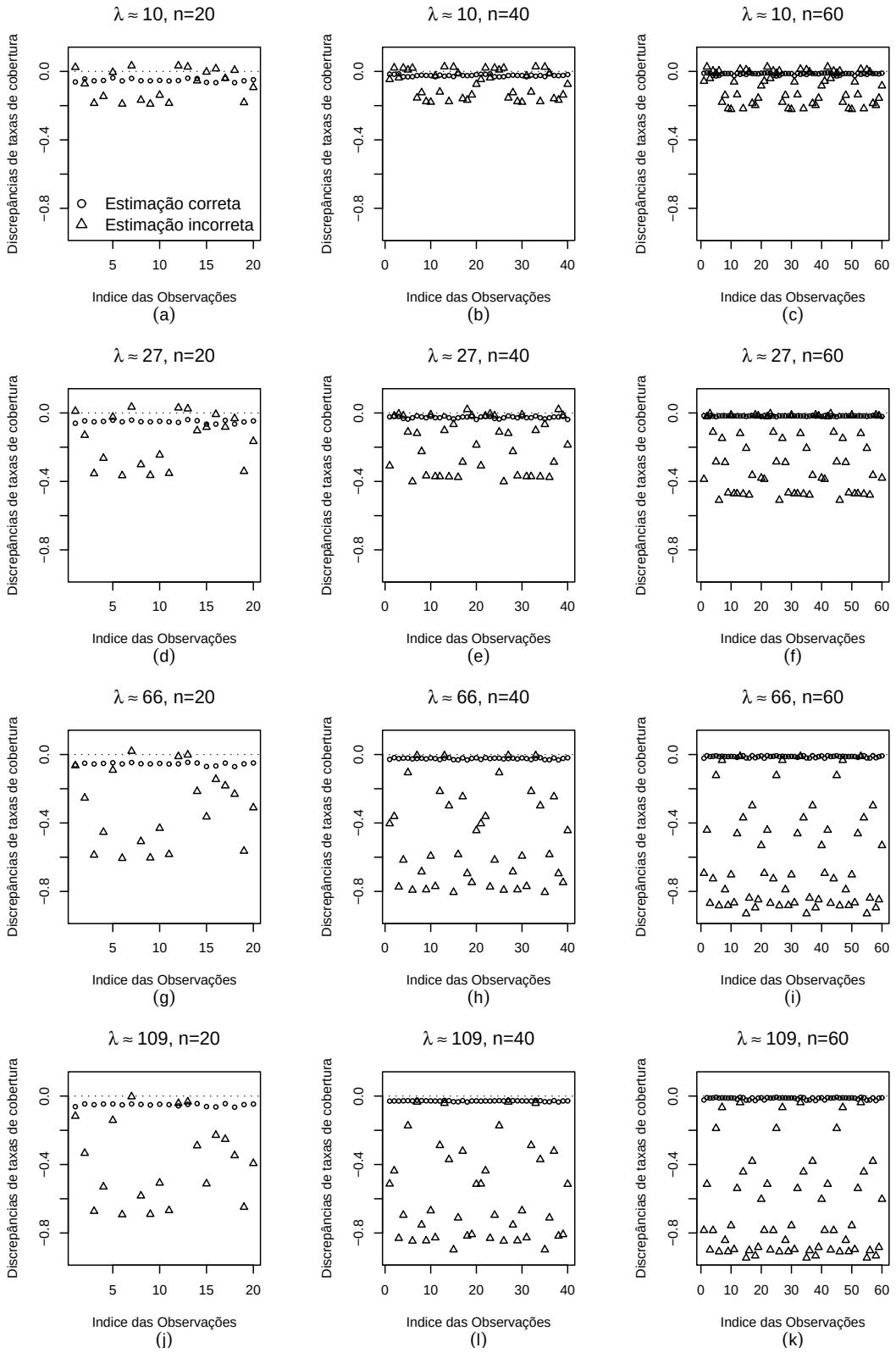

Figura 4.2. Discrepâncias de taxas de cobertura. $\mu \in (0.03, 0.08)$

testes para verificar quando a hipótese de dispersão constante é violada. Os testes que apresentamos se baseiam na consistência e normalidade assintótica do estimador de máxima verossimilhança.

4.4.1. Testes da razão de verossimilhanças, escore e Wald

Considere a hipótese nula de homoscedasticidade, $\mathcal{H}_0 : \phi_1 = \ldots = \phi_n = \phi$. Testar esta hipótese equivale a testar

$$\mathcal{H}_0 : \gamma_{(q-1)} = 0, \tag{4.23}$$

em que $\gamma_{(q-1)} = (\gamma_2, \ldots, \gamma_q)^\top$, no modelo definido em (4.1) com $z_{t1} = 1$ para $t = 1, \ldots, n$.

Neste contexto a estatística da razão de verossimilhanças (RV) é

$$\omega_1 = 2\left\{\ell(\widehat{\beta}, \widehat{\gamma}) - \ell(\widetilde{\beta}, \widetilde{\gamma})\right\},$$

em que $\ell(\beta, \gamma)$ é o logaritmo da função de verossimilhança e $(\widetilde{\beta}^\top, \widetilde{\gamma}^\top)^\top$ é o estimador de máxima verossimilhança restrito de $(\beta^\top, \gamma^\top)^\top$, obtido pela imposição da hipótese nula.

Considere agora $U_{(q-1)\gamma}$ um vetor coluna $q-1$ dimensional contendo os $q-1$ elementos finais da função escore de γ e $K^{\gamma\gamma}_{*(q-1)(q-1)}$ a matriz $(q-1) \times (q-1)$ formada das $q-1$ últimas linhas e as $q-1$ últimas colunas da matriz K^{*-1}. Então, a estatística escore para testar a hipótese em (4.23) pode ser escrita como

$$\omega_2 = \widetilde{U}^\top_{(q-1)\gamma} \widetilde{K}^{\gamma\gamma}_{*(q-1)(q-1))} \widetilde{U}_{(q-1)\gamma},$$

em que o til indica que as quantidades são avaliadas no estimador de máxima verossimilhança restrito.

A estatística de Wald para testar a hipótese em (4.23) é dada por

$$\omega_3 = \widehat{\gamma}^\top_{(q-1)} \left(\widehat{K}^{\gamma\gamma}_{*(q-1)(q-1)}\right)^{-1} \widehat{\gamma}_{(q-1)},$$

em que $\left(\widehat{K}^{\gamma\gamma}_{*(q-1)(q-1)}\right)^{-1}$ é igual $\left(K^{\gamma\gamma}_{*(q-1)(q-1)}\right)^{-1}$ avaliado no estimador de máxima verossimilhança irrestrito e $\widehat{\gamma}_{(q-1)}$ é o estimador de máxima verossimilhança de $\gamma_{(q-1)}$.

Sob condições usuais de regularidade e sob $\mathcal{H}_0$, $\omega_1, \omega_2, \omega_3$ convergem em distribuição para $\chi^2_{(q-1)}$. Assim, os testes acima podem ser realizados usando valores críticos aproximados obtidos de quantis da distribuição $\chi^2_{(q-1)}$.

4.4.2. Resultados numéricos

Para avaliar o desempenho dos testes assintóticos foram calculadas as taxas de rejeição sob a hipótese nula e o poder dos testes. Consideramos o modelo de regressão beta descrito em

(4.20) e (4.21), com $z_t = x_t$, sendo que, em adição aos cenários em que a média da resposta está próxima de zero ou próxima de um, estamos considerando dois novos cenários. Consideramos duas outras distribuições de probabilidade para gerar os valores da covariável x, a saber: a distribuição lognormal e uma distribuição assimétrica, a Weibull com parâmetros de escala e forma iguais a 1.0 e 0.5, respectivamente. Os valores da covariável x obtidos aleatoriamente da distribuição lognormal objetivam avaliar o impacto da existência de pontos de alavanca no comportamento dos testes assintóticos. Para a distribuição lognormal os verdadeiros valores dos parâmetros são $\beta_1 = 2.0$ e $\beta_2 = -0.6$ e aqui $\mu \in (0.002, 0.87)$. Para a distribuição Weibull tomamos $\beta_1 = -0.8$ e $\beta_2 = -1.2$ obtendo $\mu \in (0.007, 0.43)$. Consideramos, ainda, $n = 20, 40$ e 60 e os seguintes níveis nominais: $\alpha = 0.10, 0.05, 0.01$.

Calculamos as taxas de rejeição sob a hipótese nula ($\gamma_2 = 0.0$), variando o valor de ϕ. Uma vez que, neste caso, $\phi = \exp(\gamma_1)$, tomamos $\gamma_1 = 2.5$, $\gamma_1 = 3.0$, $\gamma_1 = 4.0$ e $\gamma_1 = 5.0$, o que conduz a $\phi \approx 12$, $\phi \approx 20$, $\phi \approx 55$ e $\phi \approx 148$, respectivamente. Os resultados desta investigação encontram-se nas Tabelas 4.1–4.4. Em seguida, calculamos o poder dos testes ($\gamma_2 \neq 0.0$). As simulações de poder foram realizadas utilizando valores críticos empíricos (exatos) obtidos das simulações de tamanho, a fim de forçar todos os testes a possuírem o mesmo tamanho. Neste caso, $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. Fixamos $\gamma_1 = 5.0$ e calculamos o poder variando os valores de γ_2. Desta forma obtemos diferentes valores para λ. Consideramos $\gamma_2 = -2.5$, $\gamma_2 = -3.5$, $\gamma_2 = -4.5$ e $\gamma_2 = -4.9$ o que conduz a valores $\lambda \approx 10$, $\lambda \approx 27$, $\lambda \approx 66$ e $\lambda \approx 109$, respectivamente. Para as simulações de poder consideramos apenas os cenários em que a média da resposta está concentrada em um dos extremos do intervalo unitário. Os resultados desta investigação encontram-se nas Tabelas 4.5 e 4.6.

A análise das Tabelas 4.1–4.4 mostra que, em geral, à medida que o número de observações cresce as distorções de tamanho dos testes são reduzidas, em especial no que se refere ao teste de Wald. De fato, o teste de Wald é consideravelmente liberal quando comparado aos outros dois testes. No cenário em que $\mu \in (0.96, 0.98)$ os testes da razão de verossimilhanaças e escore apresentam desempenhos semelhantes, sendo que o teste RV é tipicamente liberal enquanto que o teste escore é tipicamente conservador. No cenário em que $\mu \in (0.03, 0.08)$ o teste escore continua rejeitando a hipótese nula, quando esta é verdadeira, com freqüência menor que o nível nominal, mas apresenta um desempenho levemente superior ao teste RV. Apesar do comportamento tipicamente conservador, o teste escore é o teste que apresenta taxas de rejeição mais próximas do tamanho assintótico desejado quando o tamanho amostral é reduzido ($n = 20$). Nos cenários das Tabelas 4.1 e 4.2 não existem índicios de que o valor de ϕ interfira no desempenhos dos testes.

Com relação à Tabela 4.3, que representa o cenário em que as covariáveis seguem distribuição lognormal, é interessante notar que os desempenhos dos testes, em geral, são infe-

Tabela 4.1. Taxas de rejeição sob $\mathcal{H}_0 : \gamma_2 = 0$ dos testes: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim \mathcal{U}(0,1),\ t = 1, \ldots, n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.96, 0.98)$.

	ϕ	12			20			55			148		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	13.10	7.41	1.74	13.16	7.06	1.84	13.00	7.27	1.91	12.93	7.46	1.81
	Escore	8.01	3.48	0.38	8.46	3.94	0.54	8.55	4.06	0.53	8.59	3.85	0.55
	Wald	19.78	13.44	5.91	18.77	12.21	5.35	18.11	11.99	5.05	18.47	12.03	4.67
40	RV	11.51	6.03	1.21	11.79	6.00	1.47	11.44	5.91	1.52	11.38	5.88	1.40
	Escore	8.84	4.21	0.67	9.63	4.47	0.82	9.43	4.64	0.84	9.20	4.23	0.90
	Wald	14.55	8.72	2.82	14.04	8.34	2.77	13.95	8.30	2.54	13.89	7.80	2.47
60	RV	10.76	5.53	1.21	11.04	5.65	1.12	10.53	5.36	1.35	10.89	5.51	1.16
	Escore	8.99	4.35	0.71	9.61	4.56	0.77	9.28	4.47	0.92	9.57	4.48	0.88
	Wald	13.01	7.51	2.11	13.01	7.18	1.97	12.02	6.46	2.02	12.31	6.70	1.85

Tabela 4.2. Taxas de rejeição sob $\mathcal{H}_0 : \gamma_2 = 0$ dos testes: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim \mathcal{U}(0,1),\ t = 1, \ldots, n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.03, 0.08)$.

	ϕ	12			20			55			148		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	13.18	7.39	1.90	13.55	7.02	1.94	13.25	7.16	1.85	13.02	7.09	1.76
	Escore	8.96	4.16	0.64	8.91	4.09	0.52	8.92	4.03	0.55	8.83	3.70	0.54
	Wald	17.03	11.08	4.71	17.82	11.24	4.40	17.45	11.22	4.39	17.48	10.88	4.05
40	RV	10.88	5.46	1.16	11.59	5.86	1.44	11.05	5.69	1.41	10.96	5.84	1.12
	Escore	8.98	4.12	0.65	9.62	4.59	0.76	9.05	4.41	0.78	9.23	4.42	0.69
	Wald	12.96	7.29	2.13	13.88	7.70	2.45	12.90	7.35	2.32	12.88	7.36	2.03
60	RV	10.85	5.75	1.18	10.79	6.12	1.23	11.18	5.50	1.25	10.38	5.37	1.07
	Escore	9.55	4.65	0.87	9.53	4.78	0.75	9.72	4.67	0.94	9.13	4.59	0.79
	Wald	12.20	6.89	1.90	12.25	7.16	1.98	12.39	6.68	1.71	11.65	6.55	1.50

Tabela 4.3. Taxas de rejeição sob $\mathcal{H}_0 : \gamma_2 = 0$ dos testes: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim lognormal,\ t = 1, \ldots, n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.002, 0.869)$.

	ϕ	12			20			55			148		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	14.21	8.14	2.13	13.31	7.23	1.44	11.54	5.66	1.10	8.82	4.55	0.70
	Escore	6.76	3.00	0.53	5.95	2.56	0.63	5.67	3.25	1.06	7.31	4.86	2.01
	Wald	22.72	16.73	9.93	21.46	15.54	9.03	19.35	13.86	7.90	15.71	10.93	6.75
40	RV	12.72	6.99	1.69	12.64	7.07	2.02	14.60	8.46	2.52	15.03	8.90	2.77
	Escore	9.12	4.26	0.82	8.91	3.86	0.76	7.56	3.40	0.96	6.13	3.55	1.28
	Wald	16.59	10.71	4.49	16.78	10.83	4.81	18.83	12.97	6.19	19.71	13.93	7.87
60	RV	11.55	6.24	1.56	11.57	6.30	1.57	12.93	7.14	1.76	12.43	6.75	1.78
	Escore	9.38	4.52	0.84	9.46	4.42	0.86	8.66	3.73	0.99	6.59	3.29	1.21
	Wald	14.17	8.89	3.23	14.19	8.79	3.23	15.60	9.56	4.04	15.40	9.86	4.28

Tabela 4.4. Taxas de rejeição sob $\mathcal{H}_0 : \gamma_2 = 0$ dos testes: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim \text{Weibull}(1.0, 0.5),\ t = 1, \ldots, n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.007, 0.434)$.

	ϕ	12			20			55			148		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	18.67	11.14	3.72	19.03	12.19	4.34	19.34	12.71	4.74	19.48	12.65	4.48
	Escore	5.76	2.57	0.99	5.61	2.60	0.92	4.60	2.76	0.99	4.73	3.05	1.33
	Wald	27.71	21.49	2.59	27.96	21.65	13.77	28.55	22.28	4.35	29.36	22.75	14.76
40	RV	13.73	7.66	1.79	13.77	7.84	2.10	14.13	8.18	2.37	13.78	7.81	2.13
	Escore	7.92	3.81	1.22	7.71	3.55	0.91	7.51	3.45	1.00	5.94	3.13	1.17
	Wald	18.03	12.40	5.88	18.41	12.39	5.64	18.53	12.35	5.84	18.48	12.41	5.86
60	RV	11.96	6.42	1.52	12.50	6.39	1.51	12.08	6.50	1.69	12.77	7.01	1.64
	Escore	8.55	3.93	1.02	8.77	4.00	1.01	8.09	3.76	0.95	7.54	3.58	1.23
	Wald	15.02	9.47	4.02	15.44	9.52	3.70	14.82	9.04	3.65	15.42	10.10	4.18

Tabela 4.5. Poder dos testes: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim \mathcal{U}(0,1),\ t = 1,\ldots,n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.96, 0.98)$.

	λ	10			27			66			109		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	58.26	45.56	21.78	75.83	65.73	39.44	78.61	69.42	46.51	76.90	67.51	45.74
	Escore	56.54	44.77	25.07	70.48	59.59	40.08	70.92	61.48	44.24	69.42	59.72	44.12
	Wald	54.47	39.92	13.78	74.51	61.33	28.40	79.15	68.18	38.39	78.20	67.61	38.31
40	RV	87.76	81.03	58.58	96.12	93.12	81.86	96.95	94.64	85.83	96.28	93.94	83.88
	Escore	85.59	77.83	55.86	93.89	89.46	75.09	93.91	90.27	78.23	93.24	88.97	76.00
	Wald	87.88	80.43	53.30	96.69	94.01	79.88	97.68	95.90	86.31	97.28	95.35	85.32
60	RV	96.52	93.92	82.68	99.44	98.84	95.89	99.62	99.25	96.97	99.41	98.81	96.21
	Escore	95.46	92.20	80.26	98.80	97.73	93.18	99.01	98.03	93.81	98.63	97.30	92.84
	Wald	96.97	94.13	82.19	99.65	99.14	96.45	99.80	99.53	98.08	99.64	99.33	97.41

Tabela 4.6. Poder dos testes quasi–t: da razão de verossimilhanças, escore e Wald, no modelo de regressão beta $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$, em que $x_t \sim \mathcal{U}(0,1),\ t = 1,\ldots,n$ e $\phi = \exp(\gamma_1 + \gamma_2 x_t)$. $\mu \in (0.03, 0.08)$.

	λ	10			27			66			109		
n	Estatística	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
20	RV	65.32	52.84	29.50	83.42	74.60	51.74	91.32	84.72	66.04	92.40	87.60	71.36
	Escore	62.28	49.72	28.98	77.60	66.94	45.34	85.34	76.48	56.26	87.20	78.56	59.34
	Wald	60.66	47.46	18.38	81.76	71.94	38.64	91.04	84.14	57.98	92.92	87.52	64.56
40	RV	89.90	83.50	60.56	98.56	97.00	87.98	99.60	99.28	96.06	99.70	99.46	97.44
	Escore	86.74	77.66	52.06	96.76	93.66	77.48	98.98	97.50	89.10	99.22	97.84	91.42
	Wald	90.48	83.64	59.18	99.02	97.74	89.22	99.78	99.48	97.42	99.86	99.70	98.42
60	RV	98.62	97.10	88.18	99.94	99.80	98.92	100.00	99.98	99.84	100.00	100.00	99.94
	Escore	98.02	95.64	84.10	99.82	99.42	96.80	99.94	99.78	99.02	100.00	99.94	99.18
	Wald	98.92	97.42	88.40	100.00	99.86	99.22	100.00	99.98	99.90	100.00	100.00	99.98

riores aos dos dois cenários anteriores. O teste de Wald apresenta taxas de rejeição consideravelmente mais distantes dos níveis nominais. Por exemplo, para $n = 20$, $\phi = 12$ e a um nível nominal de 10% a taxa de rejeição empírica deste teste é de 22%. Em algumas situações, também o teste escore e RV apresentam distorções de tamanho mais expressivas. O teste escore parece ser o mais prejudicado neste cenário, uma vez que a tendência de ser conservador é acentuada mesmo quando o tamanho amostral aumenta. Para $n = 60$ e quando o nível nominal é 10% e $\phi = 148$ a taxa de rejeição empírica associada a esse teste é igual a 6.59%. O que ocorre é que a covariável gerada a partir da distribuição lognormal apresenta observações fortemente de alavanca. Por exemplo, no cenário em que $n = 60$ e $\phi = 148$, três observações apresentam valor da covariável igual a 14.021 enquanto que para as demais observações $x_t \in (0.17, 3.69)$. Os valores de y associados aos valores extremos de x são próximos de 0.0016 enquanto que para as demais observações $y_t \in (0.45, 0.87)$. De fato, essas observações de alavanca apresentam valor de medida de influência local total 100 vezes maior que as medidas das demais observações. Ao que parece os testes são sensíveis a pontos influentes. No que diz respeito aos resultados dos testes associados ao 'modelo Weibull' as distorções de tamanho são fortemente acentuadas. Os teste RV e Wald são consideravelmente liberais e o escore consideravelmente conservador. De uma forma geral, no cenário em que a covariável segue uma distribuição fortemente assimétrica, o teste escore mostra-se mais confiável, em especial quando o tamanho da amostra é reduzido.

Quanto ao poder dos testes, Tabelas 4.5–4.6, o teste escore apresenta valores inferiores porém próximos aos do teste da razão de verossimilhanças, especialmente com o aumento do número de observações. Assim, há fortes indícios de que o teste escore é o mais confiável entre os três testes investigados, uma vez que este apresenta resultados satisfatórios tanto em relação ao poder quanto no que diz respeito aos tamanhos. Adicionalmente, uma vez que exige somente o ajuste do modelo homoscedástico, o teste escore é uma opção muito atraente.

4.5. Exemplos simulados

Objetivando investigar gráficos de diagnóstico adequados para detectar dispersão variável, simulamos dois exemplos de uma amostra em que os dados são gerados com dispersão variável e o ajuste do modelo é realizado supondo dispersão constante. Nos dois casos, consideramos

$$\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t \ \text{ e } \ \log(\phi_t) = \gamma_1 + \gamma_2 z_t, \quad t = 1, \ldots, 40.$$

No Exemplo 1 admitimos que $z_t = x_t$, $t = 1, \ldots, 40$, i.e, cada observação apresenta uma precisão diferente. Os valores da covariável x_t são realizações independentes de uma variável aleatória distribuída uniformemente no intervalo $(0, 1)$, $\beta_1 = 0.8, \beta_2 = -2.5$, $\gamma_1 =$

4.5 e $\gamma_2 = -4.1$. Neste cenário, $\mu \in (0.16, 0.69)$, $\max_t \phi_t \approx 88$, $\min_t \phi_t \approx 2$ e $\lambda \approx 58$. Considerando os conjuntos de dados simulados, inicialmente ajustaremos um modelo de regressão beta considerando corretamente o modelo para a média da resposta, mas admitindo dispersão constante ao longo das observações. Na Tabela 4.7 apresentamos as estimativas dos coeficientes e p–valores para esse modelo. Em seguida utilizaremos as medidas de diagnóstico apresentadas nos capítulos anteriores para avaliar a qualidade do ajuste. Finalmente, se existirem índicios da necessidade da modelagem da dispersão, ajustaremos aos dados um modelo de regressão beta modelando a média e a dispersão simultaneamente.

Na Figura 4.3 apresentamos gráficos de diagnóstico para o modelo de regressão beta com dispersão constante referentes ao primeiro exemplo. Nessa figura, considerando o resíduo ponderado padronizado 2, construímos os gráficos de resíduos contra os índices das observações (a), contra os valores preditos (b), contra os valores da covariável (c) e o gráfico normal de probabilidades com envelopes simulados (d). Em seguida, considerando a medida de influência local I_{max}, obtida a partir do esquema de ponderação de casos, construímos os gráficos de $I_{max,\theta}$ contra os índices das observações (e), contra os valores preditos (f) e contra os valores da covariável (g). Finalmente, consideramos o gráfico da distância de Cook contra os índices das observações (h).

Tabela 4.7. Resultados inferenciais do Exemplo 1. Modelo supondo dispersão constante

Parâmetros	Estimativas	p–valor
β_1	1.1515	0.00000
β_2	-3.6199	0.00000
ϕ	9.3361	

A partir da Tabela 4.7 notamos que a estimativa de β_2 está consideravelmente distante do valor verdadeiro $\beta_2 = -2.5$ e que a estimativa de ϕ é muito pequena. Analisando os gráficos na Figura 4.3 notamos que os gráficos dos resíduos contra os valores preditos revelam uma leve tendência decrescente, ou seja, à medida que os valores preditos aumentam a dispersão dos resíduos diminui (Figura 4.3b). Já o gráfico dos resíduos contra os valores da covariável apresenta a tendência oposta; à medida que o valor da covariável aumenta a dispersão dos resíduos também cresce (Figura 4.3c). Essas duas tendências são mais evidentes nos correspondentes gráficos dos valores de $I_{max,\theta}$ contra os valores preditos e contra os valores das covariáveis, Figura 4.3f e Figura 4.3g, respectivamente.

No processo gerador dos dados o coeficiente que define a heterogeneidade da precisão (γ_2) é negativo e a covariável x_t assume apenas valores positivos. Assim, à medida que os valores

da covariável aumentam os valores da precisão diminuem, ou seja, aumenta a dispersão dos dados. Adicionalmente, uma vez que β_2 é negativo, temos que também quando os valores da covariável aumentam os valores da resposta diminuem. Assim, se estabelece uma relação positiva entre os valores da resposta e a precisão dos dados, ou, por outro lado, uma relação negativa entre a resposta e a dispersão. Esta última relação pode ser percebida através do gráfico dos resíduos contra os valores preditos e mais nitidamente através do gráfico dos valores de $I_{max,\theta}$ contra os valores preditos.

Logo, as tendências evidenciadas são coerentes e apontam para a natureza dos dados, que é ter dispersão variável ao longo das observações e esta dispersão podendo estar relacionada com a covariável x_t. Temos também o gráfico normal de probabilidades com envelopes simulados, que mostra a falta de qualidade do ajuste do modelo supondo dispersão constante (Figura 4.3d). Nota-se ainda nos gráficos dos resíduos (Figura 4.3a) e de $I_{max,\theta}$ (Figura 4.3e) contra os índices das observações que a segunda metade dos dados parece estar menos dispersa que a primeira. Investigando os dados constatamos que a segunda metade dos dados, em geral, apresenta menores valores da covariável. As exceções são os casos 30, 37 e 39. Quanto às observações influentes, destacam-se, considerando todos os gráficos inclusive a distância de Cook, os casos 3 e 8. Os dois casos apresentam a mesma característica: o valor da resposta muito pequeno, 0.001 e 0.002, respectivamente. A próxima observação em ordem crescente do valor da resposta é o caso 15 em que $y_{15} = 0.01$.

Com base nos indícios apresentados nos gráficos de diagnósticos, consideramos ajustar aos dados simulados um modelo de regressão beta em que a média e a dispersão são modeladas simultaneamente. Consideramos x_t a covariável utilizada nos dois modelos, a ligação logito para a média e a função logaritmo para a dispersão (Tabela 4.8). Realizamos os testes da razão de verossimilanças, escore e Wald para testar a hipótese $\mathcal{H}_0 : \gamma_2 = 0$ e obtemos nos três casos p–valores menores que 0.00000.

Tabela 4.8. Resultados inferenciais do Exemplo 1. Modelo supondo dispersão variável.

Parâmetros	β_1	β_2	γ_1	γ_2	λ
Estimativas	0.983	-2.977	4.896	-4.398	78.2
p–valor	0.0000	0.0000	0.0000	0.0000	

A Tabela 4.8 revela que as estimativas dos coeficientes dos modelos da média e da dispersão estão próximos dos valores verdadeiros. No entanto, nota-se que λ é consideravelmente superestimado. Para finalizar este exemplo seria necessário realizar uma análise de diagnóstico do modelo ajustado apresentado na Tabela 4.8.

No Exemplo 2 definimos $z_t = 0$, para $t = 1, \ldots, 20$ e $z_t = 1$, para $t = 21, \ldots, 40$, ou seja, estamos trabalhando dois sub–conjuntos de dados, cada um com uma precisão própria e estimaremos o modelo supondo dispersão constante. Consideramos $x_t \sim \mathcal{U}(0, 1)$, $\beta_1 = 0.8, \beta_2 = -2.5$ e $\gamma_1 = 5.7$. Variamos os valores de γ_2 no sentido de obter diferentes valores de λ. Admitimos $\gamma_2 = -2.3$, $\gamma_2 = -3.3$, $\gamma_2 = -4.2$ e $\gamma_2 = -4.7$ o que conduz a valores $\lambda \approx 10$, $\lambda \approx 27$, $\lambda \approx 66$ e $\lambda \approx 109$, respectivamente. Pretendemos comparar o comportamento de alguns gráficos de diagnósticos à medida que o valor de λ aumenta. Construímos gráficos de resíduos contra os índices das observações, gráficos normais de probabilidades com envelopes simulados e gráficos de $I_{max;\theta}$ contra os índices das observações. Na Figura 4.4 a–c consideramos esses gráficos para $\lambda \approx 10$, na Figura 4.4 d–f para $\lambda \approx 27$, na Figura 4.4 g–i para $\lambda \approx 66$ e na Figura 4.4 j–k para $\lambda \approx 109$.

Nota-se a partir da Figura 4.4 que, à medida que o grau de heterogeneidade da dispersão aumenta, o gráfico normal de probabilidades com envelopes simulados evidencia mais nitidamente a falta de qualidade do ajuste supondo dispersão constante. Adicionalmente, também fica mais evidente a partir dos gráficos dos resíduos e de I_{max} contra os índices das observações, a diferença entre as dispersões e da primeira da segunda metade dos dados.

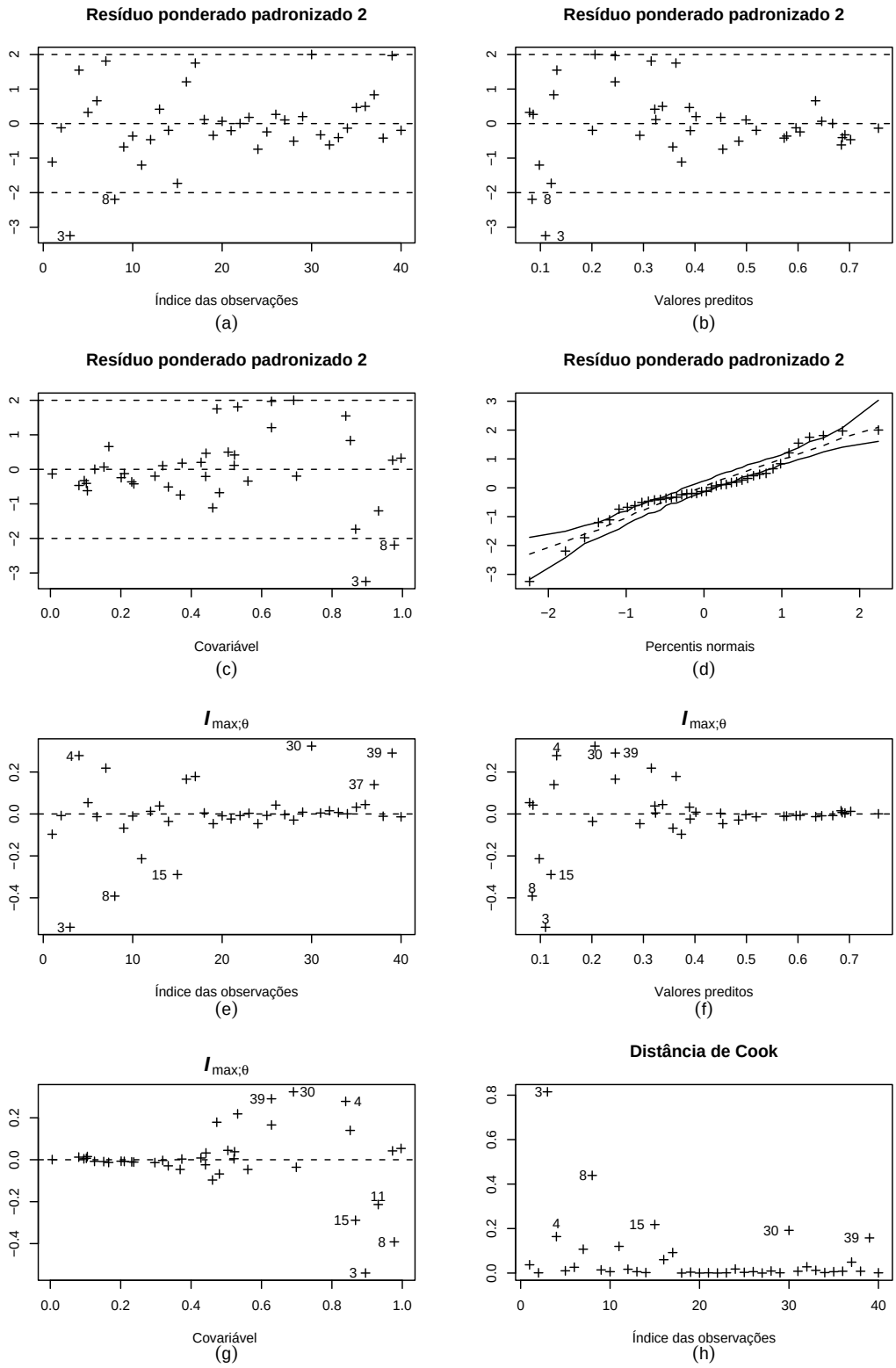

Figura 4.3. Gráficos de resíduos, influência local e distância de Cook. Exemplo 1. $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$ e $\log(\phi_t) = \gamma_1 + \gamma_2 x_t$, $x_t \sim \mathcal{U}(0,1)$, $t = 1, \ldots, 40$. O modelo é estimado supondo que $\phi_t = \phi$ para $t = 1, \ldots, 40$.

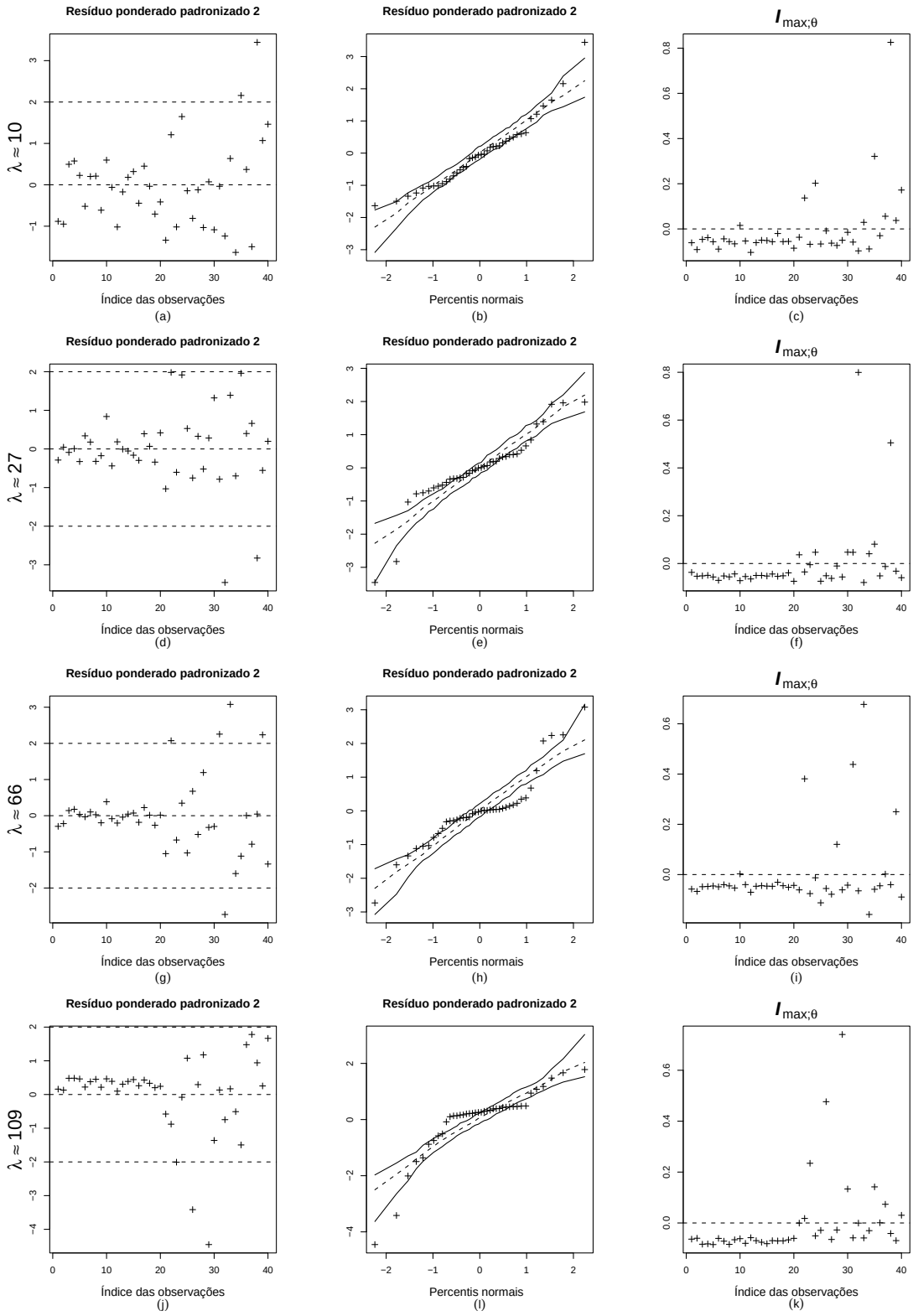

Figura 4.4. Gráficos de resíduos e influência local. Exemplo 2. $\log[\mu_t/(1-\mu_t)] = \beta_1 + \beta_2 x_t$ e $\log(\phi_t) = \gamma_1 + \gamma_2 z_t$, $x_t \sim \mathcal{U}(0,1)$, $t = 1, \ldots, 40$ e $z_t = 0$, para $t = 1, \ldots, 20$ e $z_t = 1$, para $t = 21, \ldots, 40$. O modelo é estimado supondo que $\phi_t = \phi$ para $t = 1, \ldots, n$.

Capítulo 5

Diagnóstico para o modelo de regressão beta com dispersão variável

5.1. Introdução

No Capítulo 4 apresentamos o modelo de regressão beta com dispersão variável. Adicionalmente, apresentamos algumas técnicas no sentido de identificar a necessidade da modelagem da dispersão. Uma vez ajustado um modelo em que a média e a dispersão são modeladas simultaneamente faz-se necessário avaliar a qualidade desse ajuste. Neste capítulo, apresentamos algumas técnicas de diagnóstico para o modelo de regressão beta com dispersão variável. Inicialmente, avaliaremos o uso do resíduo ponderado padronizado 2 adaptado à situação em que a dispersão é modelada. Utilizaremos esse resíduo para a construção de gráficos, em especial o gráfico normal de probabilidades com envelopes simulados. Em seguida desenvolveremos medidas de influência local. Ressaltamos que apresentaremos apenas as expressões de Δ considerando os diferentes esquemas de perturbação, ver Seções 5.3.1–5.3.3 e $\ddot{\ell}$ para o modelo de regressão beta com dispersão variável, ver (5.2). No entanto iremos considerar todas as medidas de influência discutidas no Capítulo 3, ou seja, $I_{max;\theta}$, $I_{max;\beta}$, $I_{max;\gamma}$, $C_{t;\theta}$, $C_{t;\beta}$ e $C_{t;\gamma}$. Apresentaremos ainda aplicações a dados reais.

5.3. Resíduo ponderado padronizado 2

No Capítulo 2, apresentamos o resíduo ponderado padronizado 2 (ver (2.8)) e constatamos que este resíduo apresenta desempenho superior aos demais resíduos investigados, especialmente no sentido de identificar observações efetivamente influentes para as estimativas das médias. Nossa proposta para o modelo de regressão beta com dispersão variável é considerar uma adaptação do resíduo ponderado padronizado 2 para a situação em que o parâmetro de precisão é modelado. Desta forma, podemos definir o novo resíduo ponderado padronizado 2 como

$$r_t^{pp} = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{v_t(1-h_{tt}^*)}}, \tag{5.1}$$

sendo que μ_t^* está definida em (4.5), $v_t = \psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)$ e h_{tt}^* é o t-ésimo elemento da diagonal principal de H^* que, neste caso, é dada por

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

em que a matriz Φ está definida em (4.7).

5.3. Influência local

Nesta seção, aplicaremos o método de influência local discutido no Capítulo 3 ao modelo de regressão beta com dispersão variável. Inicialmente, obteremos a expressão do inverso da matriz de informação observada e, em seguida, apresentaremos três esquemas de perturbação. Considere o modelo definido em (4.1) e sejam $s = k + q$ e $\theta = (\beta^\top, \gamma^\top)^\top$. Temos que

$$\ddot{\ell} = \ddot{\ell}(\beta, \gamma) = \begin{pmatrix} \ddot{\ell}_{\beta\beta} & \ddot{\ell}_{\beta\gamma} \\ \ddot{\ell}_{\gamma\beta} & \ddot{\ell}_{\gamma\gamma} \end{pmatrix}, \tag{5.2}$$

em que$\ddot{\ell}_{\beta\beta} = X^\top \Phi Q X$, $\ddot{\ell}_{\beta\gamma} = \ddot{\ell}_{\gamma\beta}^\top = X^\top F T H Z$ e $\ddot{\ell}_{\gamma\gamma} = Z^\top \mathcal{V} Z$, sendo que Φ está definido em 4.7, $Q = \mathrm{diag}\{q_1, \ldots, q_n\}$, com

$$q_t = \left\{ \phi_t[\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)] + (y_t^* - \mu_t^*)\frac{g''(\mu_t)}{g'(\mu_t)} \right\} \frac{1}{\{g'(\mu_t)\}^2}, \tag{5.3}$$

$F = \mathrm{diag}\{f_1, \ldots, f_n\}$ com

$$f_t = c_t - (y_t^* - \mu_t^*) \tag{5.4}$$

e $\mathcal{V} = \mathrm{diag}\{\nu_1, \ldots, \nu_n\}$ com

$$\nu_t = d_t^* + a_t \frac{h''(\mu_t)}{\{h'(\mu_t)\}^3}. \tag{5.5}$$

Temos que y_t^*, μ_t^*, a_t, c_t e d_t^* estão definidos em (1.5), (4.5), (4.12), (4.16) e (4.17), respectivamente, e as matrizes T e H estão definidas em (1.6) e (4.13), respectivamente. A partir de (5.2) e usando a fórmula do inverso de matriz particionada (Rao, 1973, p. 33) obtemos

$$\ddot{\ell}^{-1} = \begin{pmatrix} \ddot{\ell}^{\beta\beta} & \ddot{\ell}^{\beta\gamma} \\ \ddot{\ell}^{\gamma\beta} & \ddot{\ell}^{\gamma\gamma} \end{pmatrix},$$

em que

$$\ddot{\ell}^{\beta\beta} = \left(X^\top \Phi Q X - X^\top F T H Z (Z^\top \mathcal{V} Z)^{-1} Z^\top H T F^\top X \right)^{-1},$$

$$\ddot{\ell}^{\beta\gamma} = (\ddot{\ell}^{\gamma\beta})^\top = -\ddot{\ell}^{\beta\beta} X^\top F T H Z (Z^\top \mathcal{V} Z)^{-1}$$

$$\ddot{\ell}^{\gamma\gamma} = (Z^\top \mathcal{V} Z)^{-1} \left\{ \mathcal{I}_q + (Z^\top H T F^\top X) \ddot{\ell}^{\beta\beta} X^\top F T H Z (Z^\top \mathcal{V} Z)^{-1} \right\}.$$

Aqui, $\mathcal{I}_q$ é matriz identidade de dimensão $q \times q$.

Ressaltamos que nas Seções 5.3.1–5.3.3 as quantidades assinaladas com "$\widehat{\ }$" são avaliadas em $(\widehat{\beta}^\top, \widehat{\gamma}^\top)^\top$. Informamos ainda que os detalhes da obtenção dos resultados apresentados nessas seções podem ser encontrados no Apêndice 5.A.

5.3.1. Ponderação de casos

Considere o esquema de perturbação

$$\ell_\delta(\beta,\gamma) = \sum_{t=1}^{n} \delta_t \ell_t(\mu_t, \phi_t),$$

com $0 \leq \delta_t \leq 1$. Para o método de influência local discutido no Capítulo 3 temos que $\delta_0 = (1, 1, \ldots, 1)^\top$ e $\Delta_t = \partial \ell_t(\widehat{\theta})/\partial\theta$, $t = 1, \ldots, n$. Assim, da Seção 5.A.2.1 do Apêndice 5.A segue que

$$\Delta = \begin{pmatrix} X^\top \widehat{\Phi} \widehat{T} \mathcal{E} \\ Z^\top \widehat{H} \mathcal{A} \end{pmatrix}, \tag{5.6}$$

em que

$$\mathcal{E} = \text{diag}\{(y_1^* - \widehat{\mu}_1^*), \ldots, (y_n^* - \widehat{\mu}_n^*)\} \tag{5.7}$$

e

$$\mathcal{A} = \text{diag}\{\widehat{a}_1, \ldots, \widehat{a}_n\}, \tag{5.8}$$

sendo que T e a_t estão definidos em (1.6) e (4.12), respectivamente.

5.3.2. Perturbação da variável resposta

Consideremos o esquema aditivo de perturbação da resposta em que

$$y_t(\delta) = y_t + \delta_t s(y_t), \tag{5.9}$$

com $s(y_t) = \sqrt{\{\widehat{\mu}_t(1 - \widehat{\mu}_t)\}/(1 + \widehat{\phi}_t)\}}$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e da Seção 5.A.2.2 do Apêndice 5.A segue que

$$\Delta = \begin{pmatrix} X^\top \widehat{\Phi} \widehat{T} M S_y \\ Z^\top \widehat{H} B S_y \end{pmatrix}, \tag{5.10}$$

em que M está definida em (3.17), $S_y = \text{diag}\{s(y_1), \ldots, s(y_n)\}$ e $B = \text{diag}\{\widehat{b}_1, \ldots, \widehat{b}_n\}$ com b_t definido em (3.18).

5.3.3. Perturbação individual de covariáveis

Modificaremos a p–ésima coluna da matriz X, x_p, $p = 2, \ldots, k$, de forma que

$$x_{tp}(\delta) = x_{tp} + \delta_t s_{x_p}. \tag{5.11}$$

Neste caso em que o parâmetro de precisão e a média da variável resposta são modelados simultaneamente, consideraremos três cenários para o esquema de perturbação individual

de uma covariável. Primeiro, $X \neq Z$. Segundo, a matriz de regressores para o modelo da média X e a matriz de regressores para o modelo da precisão Z são iguais, i.e., $X = Z$. Terceiro, $X \neq Z$, no entanto, para algum par (p, p') $z_{tp'} = x_{tp}$ ou $z_{tp'} = \mathcal{G}(x_{tp})$, sendo $\mathcal{G}$ uma função diferenciável diferente da função identidade. Aqui não consideraremos esquemas de perturbação de covariáveis em que estejam envolvidas apenas na modelagem da dispersão.

5.3.3.1. A matriz Z totalmente diferente da matriz X

Neste caso, por exemplo, se $p \neq 2$ ou $p \neq k$,

$$\eta_t(\delta) = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_p(x_{tp} + \delta_t s_{x_p}) + \ldots + \beta_k x_{tk} \tag{5.12}$$

e $\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$. Uma vez que os regressores que determinam a média não interferem na precisão, $\phi_t(\delta) = \phi_t$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e da Seção 5.A.3.1 do Apêndice 5.A segue que

$$\Delta = \begin{pmatrix} -s_{x_p}[\widehat{\beta}_p X^\top \widehat{\Phi}\widehat{Q} - P\widehat{\Phi}\widehat{T}\mathcal{E}] \\ -\widehat{\beta}_p s_{x_p} Z^\top \widehat{F}\widehat{T}\widehat{H} \end{pmatrix}, \tag{5.13}$$

em que P é uma matriz $k \times n$ de zeros exceto a p-ésima linha, que é composta por uns e os componentes das matrizes diagonais Q e F estão definidos em (5.3) e (5.4), respectivamente e $\mathcal{E}$ está definida em (5.7).

5.3.3.2. A matriz Z igual à matriz X

Neste caso, $\eta_t(\delta)$ pode ser exemplificado segundo (5.12) e

$$\vartheta_t(\delta) = \gamma_1 + \gamma_2 x_{t2} + \ldots + \gamma_p(x_{tp} + \delta_t s_{x_p}) + \ldots + \gamma_k x_{tk}, \tag{5.14}$$

$\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$ e $\phi_t(\delta)$ é tal que $h(\phi_t(\delta)) = \vartheta_t(\delta)$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e da Seção 5.A.3.2 do Apêndice 5.A, segue que

$$\Delta = \begin{pmatrix} -s_{x_p}\{X^\top[\widehat{\beta}_p \widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\widehat{\gamma}_p] - P\widehat{\Phi}\widehat{T}\mathcal{E}\} \\ -s_{x_p}\{X^\top[\widehat{\gamma}_p \mathcal{V} + \widehat{F}\widehat{T}\widehat{H}\widehat{\beta}_p] - P\widehat{H}\mathcal{A}\} \end{pmatrix}, \tag{5.15}$$

em que P é uma matriz $k \times n$ de zeros exceto a p-ésima linha que é composta por uns e $\mathcal{V} = \text{diag}\{\widehat{\nu}_1, \ldots, \widehat{\nu}_n\}$, com ν_t definido em (5.5) e $\mathcal{A}$ está definida em (5.8).

5.3.3.3. A p'–ésima coluna da matriz Z igual à p–ésima coluna da matriz X

Aqui estamos considerando a situação em que algumas covariáveis (não todas) que determinam a média também estão envolvidas na modelagem da dispersão e pretendemos perturbar tais covariáveis. Ou seja, consideramos que $z_{tp'} = x_{tp}$, para algum par (p, p'), com

$p = 2, \ldots, k$ e $p' = 2, \ldots, q$. Neste caso $\eta_t(\delta)$ pode ser exemplificado segundo (5.12) e, por exemplo, se $p' \neq 2$ ou $p' \neq q$,

$$\vartheta_t(\delta) = \gamma_1 + \gamma_2 z_{t2} + \ldots + \gamma_{p'}(x_{tp} + \delta_t s_{x_p}) + \ldots + \gamma_q z_{tq}, \tag{5.16}$$

$\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$ e $\phi_t(\delta)$ é tal que $h(\phi_t(\delta)) = \vartheta_t(\delta)$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e

$$\Delta = \begin{pmatrix} -s_{x_p}\{X^\top[\widehat{\beta}_p \widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\widehat{\gamma}_{p'}] - P\widehat{\Phi}\widehat{T}\mathcal{E}\} \\ -s_{x_p}\{Z^\top[\widehat{\gamma}_{p'}\mathcal{V} + \widehat{F}\widehat{T}\widehat{H}\widehat{\beta}_p] - \mathcal{P}\widehat{H}\mathcal{A}\} \end{pmatrix}, \tag{5.17}$$

em que $\mathcal{P}$ é uma matriz $q \times n$ de zeros exceto a p'-ésima linha, que é composta por uns.

5.3.3.4. A p'–ésima coluna da matriz Z é função da p–ésima coluna da matriz X

Aqui estamos considerando a situação em que $z_{tp'} = \mathcal{G}(x_{tp})$, com $p = 2, \ldots, k$ e $p' = 2, \ldots, q$, em que $\mathcal{G}$ é uma função contínua e diferenciável; se $\mathcal{G}$ é a função identidade, temos a situação tratada na Seção 5.3.3.3. Aqui $\eta_t(\delta)$ pode ser exemplificado segundo (5.12) e, por exemplo, se $p' \neq 2$ ou $p' \neq q$,

$$\vartheta_t(\delta) = \gamma_1 + \gamma_2 z_{t2} + \ldots + \gamma_{p'}\mathcal{G}(x_{tp} + \delta_t s_{x_p}) + \ldots + \gamma_q z_{tq}, \tag{5.18}$$

$\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$ e $\phi_t(\delta)$ é tal que $h(\phi_t(\delta)) = \vartheta_t(\delta)$. Para este tipo de perturbação $\delta_0 = (0, 0, \ldots, 0)^\top$ e

$$\Delta = \begin{pmatrix} -s_{x_p}\{X^\top[\widehat{\beta}_p \widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\dot{G}\widehat{\gamma}_{p'}] - P\widehat{\Phi}\widehat{T}\mathcal{E}\} \\ -s_{x_p}\{Z^\top[\widehat{\gamma}_{p'}\dot{G}\mathcal{V} + \widehat{F}\widehat{T}\widehat{H}\widehat{\beta}_p] - \mathcal{P}\widehat{H}\mathcal{A}\dot{G}\} \end{pmatrix}, \tag{5.19}$$

em que $\mathcal{P}$ é uma matriz $q \times n$ de zeros exceto a p'-ésima linha, que é composta por uns e $\dot{G} = \mathrm{diag}\{\mathcal{G}'(x_{1p}), \ldots, \mathcal{G}'(x_{np})\}$, com $\mathcal{G}'(x_{tp}) = \mathrm{d}\mathcal{G}(x_{tp})/\mathrm{d}x_{tp}$.

5.4. Aplicações

No Capítulo 3 constatamos que, em três dos cinco conjuntos de dados que analisamos a modelagem da dispersão era uma possibilidade a se considerar. Neste sentido, ajustamos o modelo de regressão beta com dispersão variável a esses três conjuntos, a saber: dados de gastos com alimentação, dados de habilidade de leitura e dados de ansiedade e estresse. Apresentaremos aqui os ajustes e análises de diagnóstico referentes aos dois últimos conjuntos de dados, uma vez que esses dois exemplos foram apresentados em Smithson e Verkuilen (2006) utilizando a modelagem da dispersão. Quanto aos dados de gastos com alimentação, ressaltamos que ao considerarmos o modelo em que a média e a dispersão

são modeladas simultaneamente, através das covariáveis número de pessoas no domicílio e renda total, todas as covariáveis são consideradas significativas tanto no modelo da média quanto no modelo da dispersão. Esse resultado foi obtido com base nos testes assintóticos da razão de verossimilhanças, escore e Wald. Adicionalmente, notamos que a modelagem da dispersão ameniza consideravelmente a influência dos casos 5, 11 e 20. Assim, apesar de não apresentarmos aqui os resultados citados acima, acreditamos que o modelo de regressão beta em que a média e a dispersão são modeladas simultaneamente, através das covariáveis número de pessoas no domicílio e renda total, é o mais indicado para os dados de gastos com alimentação.

5.4.1. Aplicação I: dados de habilidade de leitura

Esta aplicação refere-se aos dados sobre habilidade em leitura analisados nas Seções 2.7.3 e 3.6.2. Nessas seções, verificamos que tanto a análise de resíduos quanto a análise de influência para o modelo de regressão beta ajustado a esses dados apontavam para a necessidade da modelagem da dispersão. Temos como variável resposta o escore transformado em um teste de habilidade de leitura aplicado a um grupo de 44 crianças e como covariáveis uma variável indicadora de ausência ou presença de dislexia (-1: ausência e 1: presença) (x_2), o escore padronizado de QI não–verbal (x_3), chamadas de dislexia e QI, respectivamente, e a interação (produto) entre as duas covariáveis (x_4). Os resultados inferenciais do modelo beta logito ajustado a esses dados, apresentados na Tabela 2.2, evidenciam que apenas a covariável dislexia é considerada significativa para a explicação da média.

Seguindo sugestão de Smithson e Verkuilen (2006), e a partir da análise de diagnóstico realizada nos Capítulos 2 e 3 decidimos ajustar um modelo de regressão beta em que a média e a dispersão são modeladas simultaneamente através das mesmas covariáveis: dislexia, QI e a interação (Tabela 5.1). Nota-se que a interação não é significativa para o modelo da dispersão. Mesmo assim realizamos uma breve análise de diagnóstico para esse modelo e apresentamos aqui apenas o gráfico de I_{max} contra os índices das observações, considerando o esquema da ponderação de casos, o qual evidencia que a diferença entre as dispersões da primeira e da segunda metade dos dados é mantida (Figura 5.1). Assim, decidimos avaliar o modelo de regressão beta com dispersão variável em que as covariáveis para a média são x_2, x_3 e x_4 e, para a dispersão, apenas x_2 e x_3, sem a interação (Tabela 5.2). Nota-se que a modelagem do parâmetro de precisão conduz aos mesmos resultados que aqueles apresentados devido às exclusões do conjunto $\{8, 9, 15, 22\}$ no modelo de regressão beta com dispersão constante (Tabela 3.2). Ou seja, tanto a covariável QI quanto a interação são significativas. A modelagem de ϕ parece absorver a influência desses casos. Ressaltamos ainda que $\widehat{\lambda} = 454.71$ representa a estimativa do *grau de heterogeneidade da dispersão* definido em (4.22).

Tabela 5.1. Resultados inferenciais. Dados de habilidade em leitura (considerando a interação no modelo da precisão).

Parâmetros	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4
Estimativa	1.019	-0.638	0.690	-0.776	3.040	1.768	1.437	-0.611
p–valor	0.0000	0.0001	0.0064	0.0025	0.0000	0.0000	0.0045	0.1970

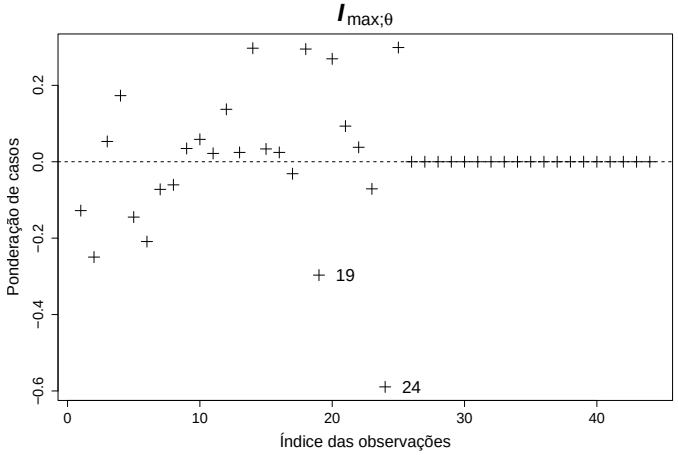

Figura 5.1. Gráfico de influência. Dados de habilidade de leitura (considerando a interação no modelo da precisão).

Tabela 5.2. Resultados inferenciais. Dados de habilidade em leitura

Parâmetros	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	λ
Estimativa	1.123	-0.742	0.486	-0.581	3.304	1.747	1.229	454.71
p–valor	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	

Para avaliar o modelo estimado apresentado na Tabela 5.2 realizamos inicialmente a análise de resíduos. Utilizamos o resíduo ponderado padronizado 2 adaptado ao modelo de regressão beta com dispersão variável, apresentado em (5.1). Na Figura 5.2 estão os gráficos dos resíduos contra os índices das observações (a) e o gráfico normal de probabilidades com envelopes simulados (b). A partir da Figura 5.2 chegamos a algumas conclusões interessantes. Primeira, a observação 8 não é mais destacada como aberrante e sim a observação 28. Segundo, não parece haver alguma tendência dos resíduos contra os índices das observações (Figura 5.2a). Terceiro, não existem indícios de falta de qualidade de ajuste (Figura 5.2b) e, se comparamos esse gráfico com o respectivo gráfico associado ao modelo com dispersão constante (Figura 2.9f), percebemos que a tendência dos resíduos entre -1 e 1 estarem levemente fora da banda superior dos envelopes foi corrigida.

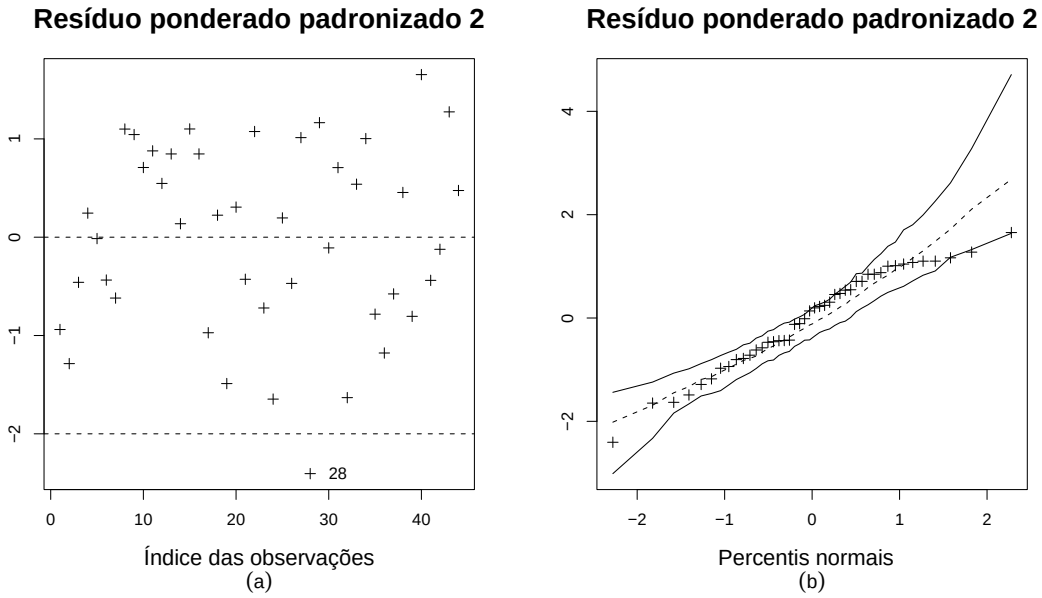

Figura 5.2. Gráficos de resíduos. Dados de habilidade de leitura

Em seguida realizamos a análise de influência. Na Figura 5.3 consideramos os gráficos de I_{max} contra os índices das observações. A organização das colunas é a mesma do Capítulo 3. A organização das linhas apenas não considera o esquema de perturbação da dispersão. Na Figura 5.4 estão os gráficos de C_t contra os índices das observações. Ressaltamos que a covariável perturbada é QI e que o esquema de perturbação é o descrito na Seção 5.3.3.3, a segunda coluna da matriz Z é igual à segunda coluna da matriz X, mas as matrizes Z e X não são iguais. Notamos a partir da Figura 5.3a–f que o problema da diferença entre as dispersões da primeira e da segunda metade dos dados é resolvida. Do ponto de vista da influência conjunta é interessante notar que cada esquema de perturbação destaca conjuntos diferentes: a ponderação de casos destaca o conjunto $\{2, 19, 24\}$, a perturbação da resposta destaca $\{8, 9, 15, 22\}$ e a perturbação da covariável QI, $\{28, 32, 33\}$ (Figura 5.3).

Já a Figura 5.4 traz uma informação importante: a observação 28 destacada pelos gráficos de resíduos, mas discretamente destacada pela medida I_{max} apenas através da perturbação da covariável, é considerada individualmente influente pela medida C_t, tanto através da ponderação de casos quanto através do esquema de perturbação da resposta. É notável ainda a influência individual dos casos 8 (Figura 5.4d–f) e 32 (Figura 5.4g–i). Consideramos algumas combinações de exclusões de casos; as mais interessantes estão na Tabela 5.3. A partir da Tabela 5.3 nota-se que o conjunto $\{8, 9, 15, 22\}$ continua sendo influente para as estimativas dos coeficientes dos dois sub–modelos, o da média e o da dispersão. No entanto,

essa influência é consideravelmente menor que aquela exercida no modelo com dispersão constante (Tabela 3.2). Já havíamos analisado esse conjunto na Seção 3.6.2 e verificamos que são observações que contradizem as relações estimadas entre as covariáveis e a resposta. Também é consideravelmente amenizada a influência individual do caso 8.

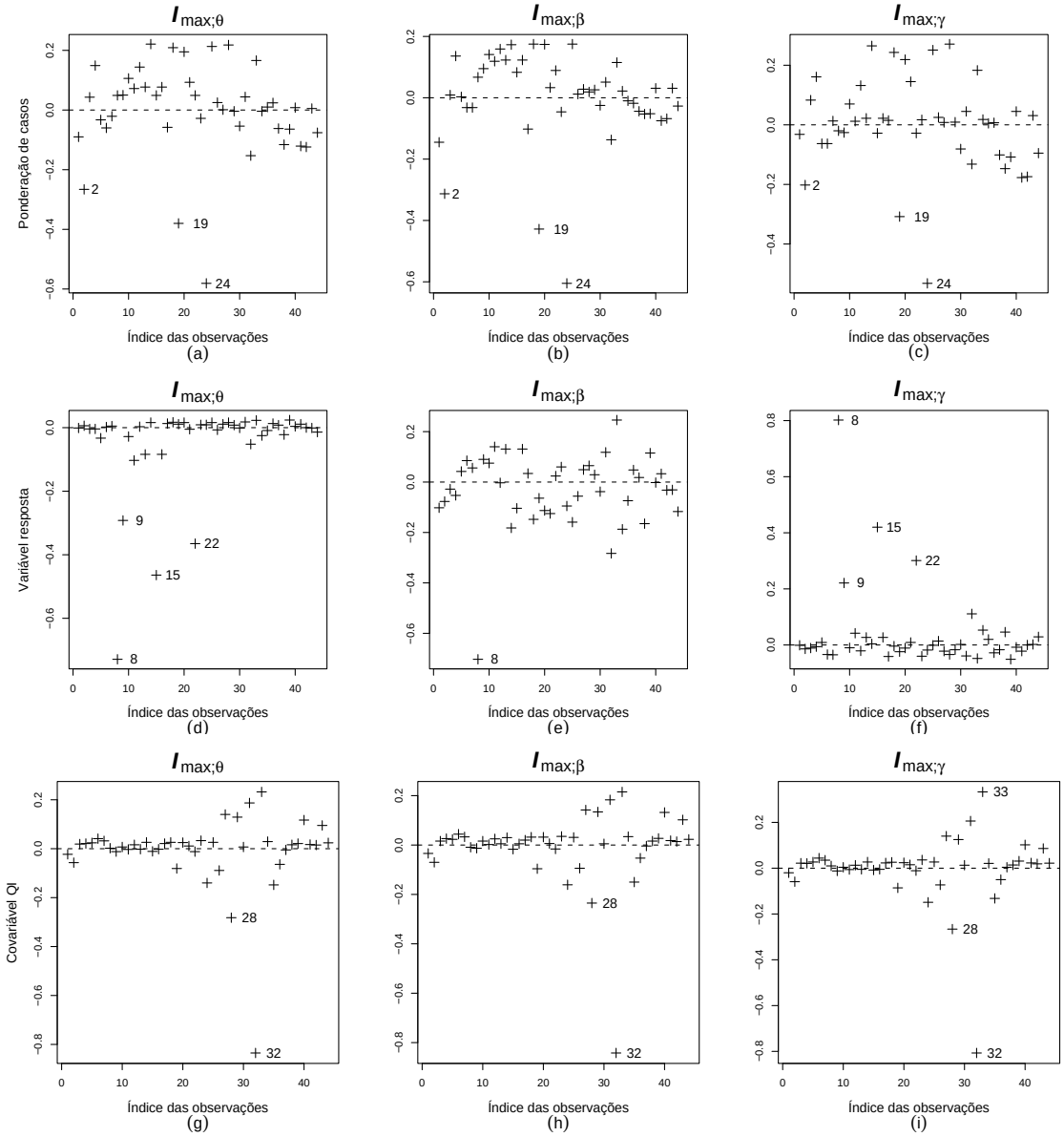

Figura 5.3. Gráficos de influência local. Dados de habilidade de leitura.

Os casos 2, 19 e 24 representam crianças não disléxicas, com valores de QI não verbal relativamente altos, mas com escores de habilidade em leitura muito pequenos quando comparados com os escores das demais crianças de mesmo perfil. O caso 24 é o mais atípico,

apresenta o menor valor da resposta entre as crianças não disléxicas, 0.646 e um valor de QI não verbal igual a 0.59; 42% das crianças não disléxicas apresentam QI não verbal inferior a este valor. A incompatibilidade maior é com o fato de se tratar de crianças sem dislexia, tanto que a exclusão do conjunto causa uma variação percentual maior em β_2, 38.1%. É interessante notar que esses três casos contribuem fortemente para que o grupo sem dislexia ainda apresente uma dispersão levemente superior ao grupo com dislexia (Figura 5.3). Esse fato é comprovado a partir da Tabela 5.3, uma vez que a estimativa de γ_2 decai em 17.7% quando as três observações não estão presentes nos dados.

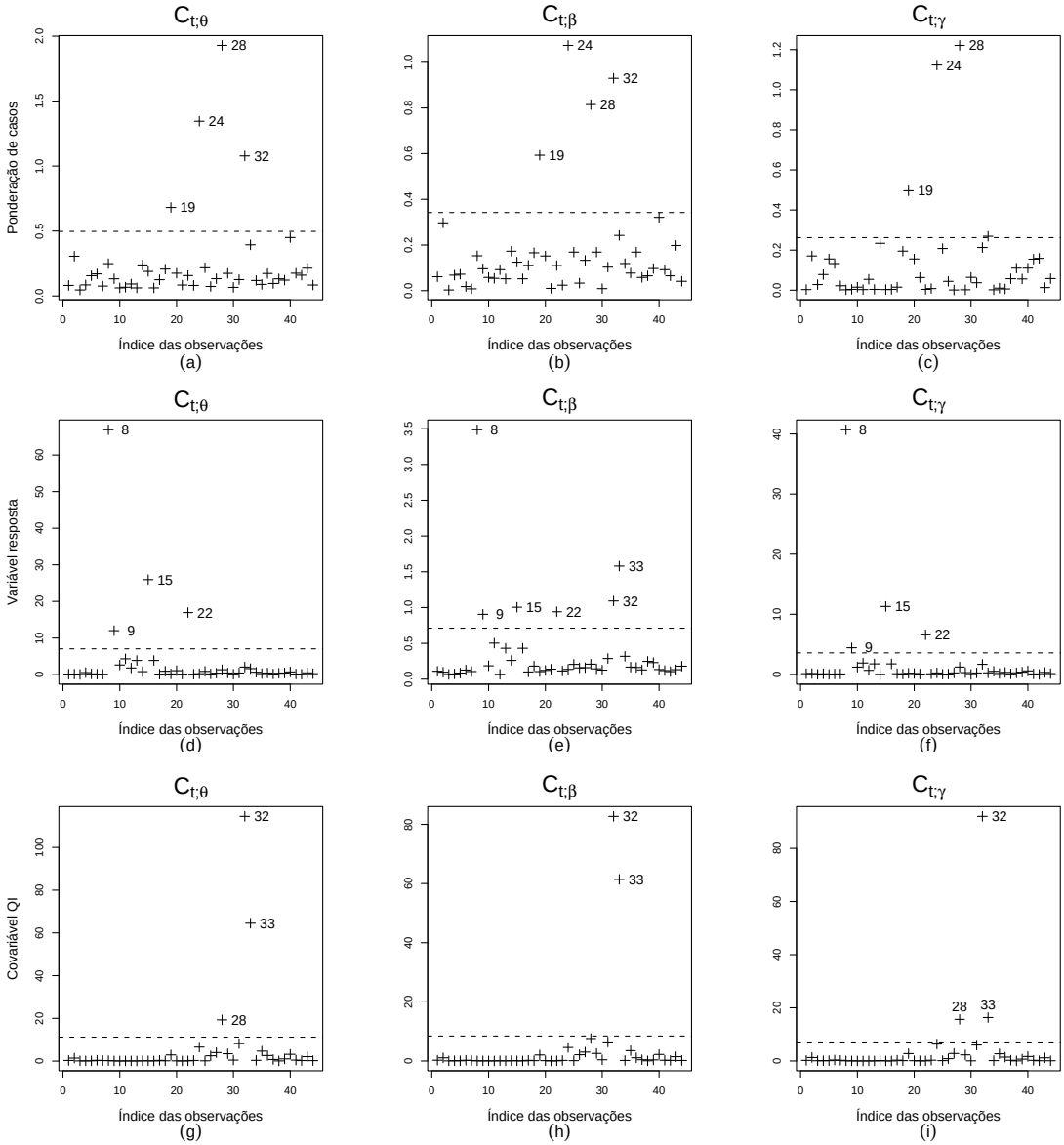

Figura 5.4. Gráficos de influência local. Dados de habilidade de leitura.

Tabela 5.3. Variação percentual das estimativas dos parâmetros e p–valores retirando observações influentes. Dados de habilidade de leitura.

Parâmetros	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	λ
8, 9, 15 e 22	-17.7	-26.8	32.3	26.4	2.0	-9.8	-15.0	-57.79
p–valor	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0002	
8	-3.7	-5.6	7.9	6.6	0.1	-0.5	-0.8	-16.61
p–valor	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	
2, 19 e 24	25.2	38.1	2.8	3.0	14.5	-17.7	29.3	16.40
p–valor	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
32	-2.9	-8.5	21.2	16.6	3.6	13.9	30.9	265.63
p–valor	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
28	2.3	0.2	-9.4	-3.6	4.6	6.7	-9.8	-2.36
p–valor	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	
33	0.5	2.5	-8.3	-3.2	-1.0	-3.8	-8.6	-60.82
p–valor	0.0000	0.0000	0.0014	0.0000	0.0000	0.0000	0.0001	

O modelo com dispersão variável permite visualizar três novas observações como influentes: os casos 28, 32 e 33. A influência do caso 32 é notadamente maior (Tabela 5.3). Os três casos pertencem ao conjunto de crianças disléxicas, ou seja, apresentam o valor da resposta mais concentrado $y \in (0.459, 0.702)$. Os casos 28 e 32 representam indivíduos com o menor e o terceiro menor valor da resposta, 0.459 e 0.540, respectivamente, entre todas as crianças. No entanto, apresentam valores de QI não tão pequenos, ou seja, que contradizem a relação estimada entre essa covariável e a resposta. O caso 28 tem o valor de QI igual a -0.795, superior ao QI de 47% das crianças com dislexia. O caso 32 tem QI igual a 0.709, superior a 90% dos valores de QI das crianças disléxicas. Já o caso 33 apresenta um valor da resposta maior, 0.572, o que o torna menos influente, apesar do seu alto valor QI não–verbal, 1.223.

O que torna o caso 32 mais influente é seu valor pequeno da resposta, incompatível com o valor de QI, tanto que sua exclusão causa o maior impacto entre as estimativas dos coeficientes do modelo da média em β_3, 21.2% e nas estimativas dos coeficientes do modelo da precisão em γ_3, 30.9%. As variações percentuais nas estimativas de γ_2 e γ_3 evidenciam a influência que o caso 32 exerce nas estimativas das precisões das observações. A exclusão desta observação implica em um aumento considerável nas precisões dos casos que representam as crianças com dislexia e uma redução nas precisões dos casos que representam as crianças sem dislexia, de forma que a estimativa de λ aumenta em cerca de 266% quando essa observação é excluída dos dados. Verificamos que as exclusões dos casos influentes não alteram os p–valores referentes aos coeficientes dos modelos da média e da precisão.

A partir desta primeira aplicação constatamos que a modelagem da precisão resultou em ganhos importantes para o modelo de regressão beta que consideramos para explicar o escore de um teste de habilidade de leitura através das covariáveis dislexia e escore de QI não–verbal. Além de evidenciar a importância da covariável QI para o modelo da média, a modelagem da precisão amenizou a influência de observações atípicas e revelou a existências de novas observações que deviam ser examinadas mais cuidadosamente. Ressaltamos a importância do esquema de perturbação da covariável que destacou enfaticamente o caso 32.

5.4.2. Aplicação II: dados de ansiedade e estresse

Nesta aplicação realizaremos uma análise de diagnóstico para o modelo de regressão beta com dispersão variável ajustado aos dados de ansiedade e estresse analisados inicialmente na Seção 3.7.3. Estamos seguindo a sugestão de Smithson e Verkuilen (2006). Os autores argumentam que um modelo em que a variável resposta é o nível de ansiedade e a variável independente é o nível de estresse deve considerar dispersão variável. De fato, as análises de resíduos e de influência realizadas para o modelo de regressão beta com dispersão constante na Seção 3.7.3 apontaram para a possibilidade da dispersão dos dados ser dependente da covariável estresse. Smithson e Verkuilen (2006) propõem um modelo de regressão beta em que estresse é preditor tanto no modelo da média quanto no modelo da dispersão. Ajustamos tal modelo e apresentamos as estimativas dos parâmetros e respectivos p–valores na Tabela 5.4. Note que a estimativa do parâmetro que determina a dispersão variável dos dados (γ_2) é considerada significativa. Os p–valores apresentados na Tabela 5.4 referem-se a testes de Wald. No entanto, realizamos também os testes escore e da razão de verossimilhanças para testar a hipótese de $\gamma_2 = 0$ e os p–valores foram menores que 0.0000. Apesar da modelagem da dispersão ser considerada importante, note que a estimativa do *grau de heterogeneidade da dispersão* dos dados é consideravelmente inferior à da aplicação anterior ($\widehat{\lambda} = 36.21$).

Tabela 5.4. Resultados inferenciais. Dados de ansiedade e estresse

Parâmetros	β_1	β_2	γ_1	γ_2	λ
Estimativas	-4.024	4.941	3.961	-4.273	36.21
p–valor	0.0000	0.0000	0.0000	0.0000	

Em seguida realizamos a análise de resíduos e de influência para esse modelo. Na Figura 5.5 apresentamos os gráficos dos resíduos contra os valores da covariável estresse (a), contra os índices das observações (b), contra os valores preditos (c), gráfico normal de probabilidades com envelopes simulados (d) e os gráficos de influência (e–g). Essa figura revela que o caso 89 continua sendo aberrante (Figura 5.5a–c). Os gráficos dos resíduos versus os valores da

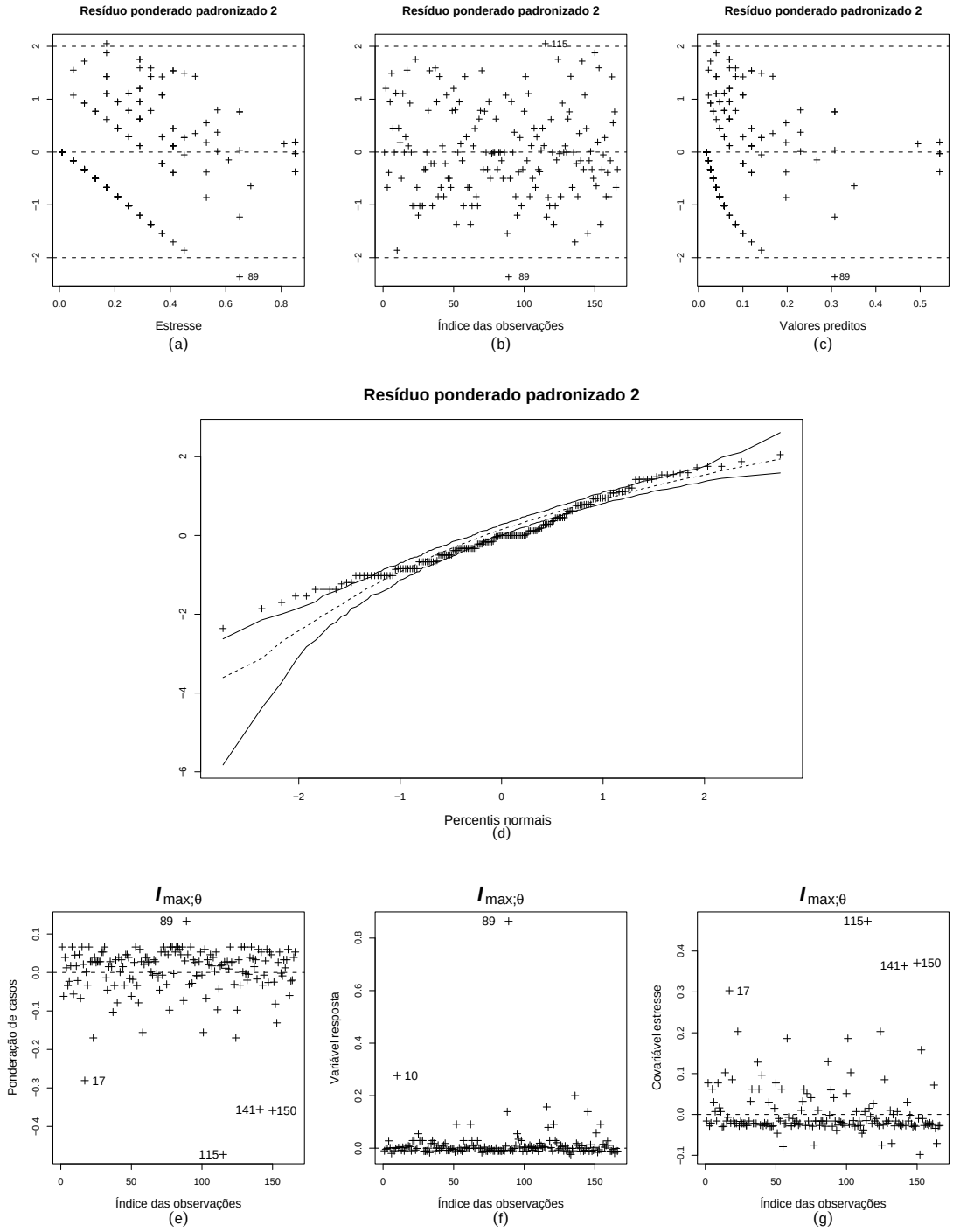

Figura 5.5. Gráficos de diagnóstico. Dados de ansiedade e estresse.

covariável estresse e contra os valores preditos revelam falta de aleatoriedade. Adicionalmente, nota-se a partir do gráfico normal de probabilidades com envelopes simulados uma certa falta de qualidade de ajuste. Quando realizamos a análise de influência (Figura 5.5e–g), notamos que a ponderação de casos aponta para dois conjuntos em direção oposta, os conjuntos $\{89\}$ e $\{17, 115, 141, 150\}$ (Figura 5.5e). A observação 89 é enfaticamente destacada pelo esquema de perturbação da variável resposta seguida pela observação 10 (Figura 5.5f). Já o esquema de perturbação da covariável destaca também o conjunto $\{17, 115, 141, 150\}$ (Figura 5.5g). Na Tabela 5.5 apresentamos as mudanças nas estimativas dos parâmetros e respectivos p–valores a partir das exclusões dos conjuntos de observações destacados como influentes.

Tabela 5.5. Variação percentual das estimativas dos parâmetros e p–valores retirando observações influentes. Dados de ansiedade e estresse.

Parâmetros	β_1	β_2	γ_1	γ_2	λ
89	0.0	1.1	-1.9	-8.0	-24.9
p–valor	0.0000	0.0000	0.0000	0.0000	
17, 115, 141 e 150	8.3	14.7	18.7	40.7	330.9
p–valor	0.0000	0.0000	0.0000	0.0000	

A partir da Tabela 5.5 ficam evidentes as influências opostas que os conjuntos $\{89\}$ e $\{17, 115, 141, 150\}$ exercem no processo de estimação do modelo. Enquanto a observação 89 influencia positivamente a modelagem da dispersão, uma vez que sua exclusão diminui o grau de heterogeneidade da dispersão, o segundo conjunto influencia negativamente essa decisão. De fato, a exclusão do conjunto $\{17, 115, 141, 150\}$ aumenta consideravelmente a dispersão dos dados (a estimativa de γ_2 passa de -4.273 para -6.0123). No entanto, apesar do caso 89 contribuir para que a dispersão dos dados seja considerada variável, sua exclusão não muda a significância da estimativa de γ_2, ou seja, a modelagem da dispersão continua sendo válida.

Como vimos no Capítulo 3, o caso 89 de fato representa uma observação atípica, uma vez que representa uma mulher com um nível de estresse muito alto, 0.65 (note que o nível de estresse varia de 0.01 a 0.85), com o menor nível de ansiedade encontrado nos dados, igual a 0.01. Já as outras observações apresentam característica oposta. Os valores da covariável estresse são pequenos quando comparados aos valores das outras observações com o mesmo valor da resposta. Para exemplicar tomemos o caso 115: o valor da resposta é 0.25 e o da covariável é 0.17. As demais observações com o valor da resposta igual a 0.25 apresentam valores da covariável entre 0.29 e 0.65.

Como visto acima, a modelagem da dispersão amenizou a influência do caso 89, e permitiu a identificação de novas observações. No entanto, não podemos considerar que o ajuste do modelo aos dados seja adequado (Figura 5.5a–d). Assim, ao que parece, também o modelo de regressão beta com dispersão variável não é adequado para a situação em que a resposta é o nível de ansiedade e o preditor é o nível de estresse. Esse modelo foi sugerido por Smithson e Verkuilen (2006), no entanto, os autores não realizaram uma análise de diagnóstico para verificar a qualidade do ajuste.

No modelo de regressão beta com dispersão variável, os diversos métodos de perturbação considerados revelaram precisamente os casos influentes, apontando para a necessidade de uma investigação mais cuidadosa com relação a algumas observações que passariam despercebidas. Em um dos exemplos, o método de perturbação individual de covariáveis, que neste caso influenciavam tanto a média quanto a precisão, revelou casos não destacados pela ponderação de casos ou pelo esquema da variável resposta, casos esses decisivos para a compreensão do problema estudado. Neste sentido, ressaltamos a importância dos mecanismos de diagnóstico para os modelos de regressão beta.

Apêndice 5.A

Influência local (modelo de regressão beta com dispersão variável)

5.A.1. Introdução

No modelo de regressão beta definido em (4.1) temos que $\theta = (\beta_1, \ldots, \beta_k, \gamma_1, \ldots, \gamma_q)^\top$, então Δ é uma matriz $(k+q) \times n$ dada por

$$\Delta = \begin{pmatrix} \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\beta_1} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\beta_1} \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\beta_2} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\beta_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\beta_k} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\beta_k} \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\gamma_1} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\gamma_1} \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\gamma_2} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\gamma_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_1\partial\gamma_q} & \cdots & \frac{\partial^2 \ell_\delta(\theta)}{\partial\delta_n\partial\gamma_q} \end{pmatrix},$$

avaliada em $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\gamma}^\top)^\top$ e δ_0. Podemos particionar Δ tal que

$$\Delta = \begin{pmatrix} \Delta_\beta \\ \Delta_\gamma \end{pmatrix},$$

em que $\Delta_\beta = \Delta_\beta(\theta, \delta)$ é uma matriz $k \times n$ e $\Delta_\gamma = \Delta_\gamma(\theta, \delta)$ é uma matriz $q \times n$, dadas, respectivamente, por

$$\Delta_\beta = \left.\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial\beta\partial\delta^\top}\right|_{(\theta=\widehat{\theta}, \delta=\delta_0)} \quad \text{e} \quad \Delta_\gamma = \left.\frac{\partial^2 \ell_\delta(\beta, \phi)}{\partial\gamma\partial\delta^\top}\right|_{(\theta=\widehat{\theta}, \delta=\delta_0)},$$

ou seja, ambas quantidades avaliadas em $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\gamma}^\top)^\top$ e δ_0.

5.A.2. Esquemas de perturbação

5.A.2.1. Ponderação de casos

No esquema de ponderação de casos temos $\ell_\delta(\beta, \gamma) = \sum_{t=1}^n \delta_t \ell_t(\mu_t, \phi_t)$ e, assim, a t–ésima coluna de Δ, $t = 1 \ldots, n$, é dada por

$$\Delta_t = \frac{\partial \ell_t(\theta)}{\partial\theta} = \left.\left(\frac{\partial \ell_t(\theta)}{\partial\beta_1}, \ldots, \frac{\partial \ell_t(\theta)}{\partial\beta_k}, \frac{\partial \ell_t(\theta)}{\partial\gamma_1}, \ldots, \frac{\partial \ell_t(\theta)}{\partial\gamma_q}\right)^\top\right|_{\theta=\widehat{\theta}}.$$

Com base em (4.2), para $i = 1, \ldots, k$ e $j = 1, \ldots, q$ obtemos

$$\frac{\partial \ell_t(\theta)}{\partial \beta_i} = \phi_t \frac{1}{g'(\mu_t)}(y_t^* - \mu_t^*)x_{ti},$$

e

$$\frac{\partial \ell_t(\theta)}{\partial \gamma_j} = a_t \frac{1}{h'(\mu_t)} z_{tj},$$

em que, y_t^*, μ_t^* e a_t são dados, respectivamente, em (1.5), (4.5) e (4.12), $g(\mu_t)$ está definida em (1.2) e $h(\phi_t)$ em (4.1). Então, temos que a i–ésima linha de Δ_β, $i = 1, \ldots, k$, é dada por

$$\Delta_i = \left(\phi_1 \frac{1}{g'(\mu_1)}(y_1^* - \mu_1^*)x_{1i}, \phi_2 \frac{1}{g'(\mu_2)}(y_2^* - \mu_2^*)x_{2i}, \ldots, \phi_n \frac{1}{g'(\mu_n)}(y_n^* - \mu_n^*)x_{ni} \right) \Bigg|_{\theta = \widehat{\theta}}$$

e a j–ésima linha de Δ_γ, $j = 1, \ldots, q$, é dada por

$$\Delta_j = \left(a_1 \frac{1}{h'(\phi_1)} z_{1j}, a_2 \frac{1}{h'(\phi_2)} z_{2j}, \ldots, a_n \frac{1}{h'(\phi_n)} z_{nj} \right) \Bigg|_{\theta = \widehat{\theta}}.$$

Desta forma, podemos escrever $\Delta_\beta = X^\top \widehat{\Phi} \widehat{T} \mathcal{E}$ e $\Delta_\gamma = Z^\top \widehat{H} \mathcal{A}$, em que $\mathcal{E}$ e $\mathcal{A}$ estão definidos na Seção 5.3.1. Assim, a matriz Δ fica expressa na forma apresentada em (5.6).

5.A.2.2. Perturbação da resposta

Com base na definição da variável resposta sob perturbação aditiva apresentada em (5.9) e considerando (4.2), temos que o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned} \ell_\delta(\mu, \phi) = \sum_{t=1}^{n} \{ & \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) - \log \Gamma((1 - \mu_t)\phi_t) + (\mu_t \phi_t - 1) \log(y_t + \delta_t s(y_t)) \\ & + [(1 - \mu_t)\phi_t - 1] \log(1 - y_t - \delta_t s(y_t)) \}. \end{aligned}$$

Obtemos

$$\frac{\partial \ell_\delta(\beta, \gamma)}{\partial \delta_t} = \frac{s(y_t)}{y_t + \delta_t s(y_t)}(\mu_t \phi_t - 1) - \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)}[(1 - \mu_t)\phi_t - 1].$$

Assim, para $i = 1, \ldots, k$

$$\begin{aligned} \frac{\partial^2 \ell_\delta(\beta, \gamma)}{\partial \delta_t \partial \beta_i} &= \phi_t \frac{1}{g'(\mu_t)} x_{ti} \left(\frac{s(y_t)}{y_t + \delta_t s(y_t)} + \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \right) \\ &= \phi_t \frac{1}{g'(\mu_t)} x_{ti} \frac{s(y_t)}{(y_t + \delta_t s(y_t))(1 - y_t - \delta_t s(y_t))} \end{aligned}, \tag{5.A1}$$

e para $j = 1, \ldots, q$

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\phi)}{\partial\delta_t\partial\gamma_j} &= \mu_t \frac{1}{h'(\mu_t)} z_{tj} \left(\frac{s(y_t)}{y_t + \delta_t s(y_t)} + \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \right) - \frac{1}{h'(\mu_t)} z_{tj} \frac{s(y_t)}{1 - y_t - \delta_t s(y_t)} \\ &= \frac{1}{h'(\mu_t)} z_{tj} \frac{-(y_t - \mu_t)s(y_t) - \delta_t s(y_t)^2}{(y_t + \delta_t s(y_t))(1 - y_t - \delta_t s(y_t))}.\end{aligned} \tag{5.A2}$$

Avaliando as expressões (5.A1) e (5.A2) em $\delta_0 = (0, 0, \ldots, 0)^\top$ e $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\gamma}^\top)^\top$, podemos escrever $\Delta_\beta = X^\top \widehat{\Phi}\widehat{T}MS_y$ e $\Delta_\gamma = Z^\top \widehat{H}BS_y$, em que M e B estão definidos na Seção 5.4.2. Assim, a matriz Δ fica expressa na forma apresentada em (5.10).

5.A.3. Perturbação individual de covariáveis

A seguir obteremos as matrizes Δ com base na pertubação aditiva da p–ésima covariável x_p, $p = 1, \ldots, k$, definida em (5.11), considerando (4.1) e os cenários descritos na Seção 5.3.3.

5.A.3.1. A matriz Z totalmente diferente da matriz X

Neste caso, o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}\ell_\delta(\mu,\phi) = \sum_{t=1}^{n} \{&\log\Gamma(\phi_t) - \log\Gamma(\mu_t(\delta)\phi_t) - \log\Gamma((1-\mu_t(\delta))\phi_t) + (\mu_t(\delta)\phi_t - 1)\log(y_t) \\ &+ \{(1-\mu_t(\delta))\phi_t - 1\}\log(1-y_t)\},\end{aligned}$$

em que $\mu_t(\delta)$ é tal que $g(\mu_t(\delta)) = \eta_t(\delta)$, com $\eta_t(\delta)$ exemplificado segundo 5.10. Assim, temos que

$$\frac{\partial \ell(\beta,\gamma)}{\partial\delta_t} = \phi_t(y_t^* - \mu_t^*(\delta)) \frac{1}{g'(\mu_t(\delta))} \beta_p s_{x_p}, \tag{5.A3}$$

em que y^* é o logito da resposta, $\mu_t^*(\delta) = \psi((1-\mu_t(\delta))\phi_t) - \psi(\mu_t(\delta)\phi_t)$ e s_{x_p} é o desvio padrão de x_p.

De (5.A3) segue que para $i = 1, \ldots, k$, com $i \neq p$,

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\beta_i} &= -\phi_t s_{x_p} \left\{ \phi_t \left[\psi'((1-\mu_t(\delta))\phi_t) + \psi'(\mu_t(\delta)\phi_t) \right] + (y_t^* - \mu_t^*(\delta)) \frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))} \right\} \\ &\quad \times \frac{1}{\{g'(\mu_t(\delta))\}^2} \beta_p x_{ti},\end{aligned}$$

em que $\mu_t^*(\delta) = \psi(\mu_t(\delta)\phi_t) - \psi((1-\mu_t(\delta))\phi_t)$. Definindo $v_t(\delta) = \psi'((1-\mu_t(\delta))\phi_t) + \psi'(\mu_t(\delta)\phi_t)$ temos então que

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\beta_i} = -\phi_t s_{x_p} \left[\phi_t v_t(\delta) + (y_t^* - \mu_t^*(\delta)) \frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))} \right] \frac{1}{\{g'(\mu_t(\delta))\}^2} \beta_p x_{ti},$$

ou ainda,

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\beta_i} = -\phi_t s_{x_p}\beta_p x_{ti} q_t(\delta), \tag{5.A4}$$

em que

$$q_t(\delta) = \left[\phi_t v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right]\frac{1}{\{g'(\mu_t(\delta))\}^2}.$$

Quando $i = p$ temos que

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\beta_p} &= -\phi_t s_{x_p}\left\{\phi_t[\psi'((1-\mu_t(\delta))\phi_t) + \psi'(\mu_t(\delta)\phi_t)] + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right\}\\ &\times \frac{1}{\{g'(\mu_t(\delta))\}^2}\beta_p(x_{tp} + \delta_t s_{x_p}) - \phi_t s_{x_p}\left[-(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\right]\end{aligned}$$

Reescrevendo essa expressão chegamos a

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\beta_p} = -\phi_t s_{x_p}\left[\beta_p(x_{tp} + \delta_t s_{x_p})q_t(\delta) - (y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\right]. \tag{5.A5}$$

Ainda de (5.A3), segue que, para $j = 1, \ldots, q$

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_j} = \Bigg\{&[\psi((1-\mu_t(\delta))\phi_t) - \psi(\mu_t(\delta)\phi_t) + y_t^*]\\ &+ \phi_t[\psi'((1-\mu_t(\delta))\phi_t)(1-\mu_t(\delta)) - \psi'(\mu_t(\delta)\phi_t)\mu_t(\delta)]\Bigg\}\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t)}z_{tj}\beta_p s_{x_p},\end{aligned}$$

que pode ser reescrita como

$$\frac{\partial^2 \ell_\delta(\beta,\gamma_j)}{\partial\delta_t\partial\gamma} = -\beta_p s_{x_p} f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t)}z_{tj}, \tag{5.A6}$$

em que, considerando o esquema de perturbação da covariável x_p, $f_t(\delta) = c_t(\delta) - (y_t^* - \mu_t^*(\delta))$, com $c_t(\delta) = \phi_t\left\{\psi'(\mu_t(\delta)\phi_t)\mu_t(\delta) - \psi'((1-\mu_t(\delta))\phi_t)(1-\mu_t(\delta))\right\}$.

Avaliando as expressões (5.A4), (5.A5) e (5.A6) em $\delta_0 = (0, 0, \ldots, 0)^\top$ e $\widehat{\theta}$ obtemos $\Delta_\beta = -s_{x_p}[\widehat{\beta}_p X^\top \widehat{\Phi}\widehat{Q} - P\widehat{\Phi}\widehat{T}\mathcal{E}]$ e $\Delta_\gamma = -\widehat{\beta}_p s_{x_p} Z^\top \widehat{F}\widehat{T}\widehat{H}$, em que $\widehat{Q}$, P e $\widehat{F}$ estão definidos na Seção 5.3.3.1. Assim, a matriz Δ fica expressa na forma apresentada em (5.13).

5.A.3.2. A matriz Z igual a matriz X

Neste caso, o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}\ell_\delta(\mu,\phi) = \sum_{t=1}^{n}\{&\log\Gamma(\phi_t(\delta)) - \log\Gamma(\mu_t(\delta)\phi_t(\delta)) - \log\Gamma((1-\mu_t(\delta))\phi_t)\\ &+ (\mu_t(\delta)\phi_t(\delta) - 1)\log(y_t) + \{(1-\mu_t(\delta))\phi_t(\delta) - 1\}\log(1-y_t)\},\end{aligned}$$

em que $\mu_t(\delta)$ e $\phi_t(\delta)$ dependem de $\eta_t(\delta)$ e $\vartheta_t(\delta)$, que podem ser exemplificados segundo 5.12 e 5.14, respectivamente. Assim, temos que

$$\frac{\partial \ell(\beta,\gamma)}{\partial \delta_t} = \mu_t(\delta)(y_t^* - \mu_t^*(\delta)) + \log(1-y_t) - \psi((1-\mu_t(\delta))\phi_t(\delta)) + \psi(\phi_t(\delta))\frac{1}{h'(\mu_t(\delta))}\gamma_p s_{x_p}$$
$$+ \phi_t(\delta)(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p},$$

em que, para este esquema de perturbação, $\mu_t^*(\delta) = \psi(\mu_t(\delta)\phi_t(\delta)) - \psi((1-\mu_t(\delta))\phi_t(\delta))$. Definindo $a_t(\delta) = \mu_t(\delta)(y_t^* - \mu_t(\delta)^*) + \log(1-y_t) - \psi((1-\mu_t(\delta))\phi_t(\delta)) + \psi(\phi_t(\delta))$ obtemos

$$\frac{\partial \ell(\beta,\gamma)}{\partial \delta_t} = a_t(\delta)\frac{1}{h'(\phi_t(\delta))}\gamma_p s_{x_p} + \phi_t(\delta)(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p}. \tag{5.A7}$$

De (5.A7) segue que para $i = 1, \ldots, k$, com $i \neq p$,

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial \delta_t \partial \beta_i} = \left\{ (y_t^* - \mu_t^*(\delta)) + \phi_t(\delta)\left[\psi'((1-\mu_t(\delta))\phi_t)(1-\mu_t(\delta)) - \psi'(\mu_t(\delta)\phi_t(\delta))\mu_t(\delta)\right] \right\}$$
$$\times \frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}x_{tj}\gamma_p s_{x_p} - \phi_t(\delta)\left[\phi_t(\delta)v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right]\frac{\beta_p x_{ti} s_{x_p}}{\{g'(\mu_t(\delta))\}^2},$$

em que $v_t(\delta) = \psi'((1-\mu_t(\delta))\phi_t(\delta)) + \psi'(\mu_t(\delta)\phi_t(\delta))$. Desta forma, obtemos, para $i \neq p$

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial \delta_t \partial \beta_i} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}x_{ti}\gamma_p s_{x_p} - \phi_t(\delta)q_t(\delta)x_{ti}\beta_p s_{x_p}, \tag{5.A8}$$

em que

$$q_t(\delta) = \left[\phi_t(\delta)v_t(\delta) + (y_t^* - \mu_t^*(\delta))\frac{g''(\mu_t(\delta))}{g'(\mu_t(\delta))}\right]\frac{1}{\{g'(\mu_t(\delta))\}^2} \quad \text{e} \quad f_t(\delta) = c_t(\delta) - (y_t^* - \mu_t^*(\delta)),$$

com $c_t(\delta) = \phi_t(\delta)\left[\psi'(\mu_t(\delta)\phi_t(\delta))\mu_t(\delta) - \psi'((1-\mu_t(\delta))\phi_t(\delta))(1-\mu_t(\delta))\right]$. Quando $i = p$,

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial \delta_t \partial \beta_p} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}(x_{tp} + \delta_t s_{x_p})\gamma_p s_{x_p} - \phi_t(\delta)q_t(\delta)(x_{tp} + \delta_t s_{x_p})\beta_p s_{x_p}$$
$$+ \phi_t(\delta)(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}s_{x_p}, \tag{5.A9}$$

Neste caso, para calcular as segundas derivadas com relação ao vetor γ também teremos que considerar $j = p$ e $j \neq p$. Assim, com base em (5.A7) para $j = 1, \ldots, k$, com $j \neq p$, temos que

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial \delta_t \partial \gamma_j} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}x_{tj}\beta_p s_{x_p}$$
$$+ \left[\psi'(\mu_t(\delta)\phi_t(\delta))\mu_t^2(\delta) + \psi'((1-\mu_t(\delta))\phi_t(\delta))(1-\mu_t(\delta))^2 - \psi'(\phi_t(\delta))\right]\frac{x_{tj}\gamma_p s_{x_p}}{\{h'(\phi_t(\delta))\}^2}$$
$$- a_t(\delta)\frac{h''(\phi_t(\delta))}{\{h'(\phi_t(\delta))\}^3}x_{tj}\gamma_p s_{x_p}.$$

Definindo

$$d_t^*(\delta) = \left[\psi'(\mu_t(\delta)\phi_t(\delta))\mu_t^2(\delta) + \psi'((1-\mu_t(\delta))\phi_t(\delta))(1-\mu_t(\delta))^2 - \psi'(\phi_t(\delta))\right] \frac{1}{\{h'(\phi_t(\delta))\}^2}$$

obtemos

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_j} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}x_{tj}\beta_p s_{x_p} - \left[d_t^*(\delta) + a_t(\delta)\frac{h''(\phi_t(\delta))}{\{h'(\phi_t(\delta))\}^3}\right] x_{tj}\gamma_p s_{x_p}\,.$$

Adicionalmente, definindo

$$\nu_t(\delta) = d_t^*(\delta) + a_t(\delta)\frac{h''(\mu_t(\delta))}{\{h'(\mu_t(\delta))\}^3}$$

chegamos a

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_j} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}x_{tj}\beta_p s_{x_p} - \nu_t(\delta)x_{tj}\gamma_p s_{x_p}\,. \tag{5.A10}$$

Finalmente, quando $j = p$ temos que

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_p} &= -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}(x_{tp}+\delta_t s_{x_p})\beta_p s_{x_p} - \nu_t(\delta)(x_{tp}+\delta_t s_{x_p})\gamma_p s_{x_p}\\ &+ a_t(\delta)\frac{1}{h'(\phi_t(\delta))}s_{x_p}.\end{aligned} \tag{5.A11}$$

Avaliando as expressões (5.A8)–(5.A11) em $\delta_0 = (0,0,\ldots,0)^\top$ e $\widehat{\theta}$ obtemos $\Delta_\beta = -s_{x_p}\{X^\top[\widehat{\beta}_p\,\widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\widehat{\gamma}_p] - P\,\widehat{\Phi}\widehat{T}\mathcal{E}\}$ e $\Delta_\gamma = -s_{x_p}\{X^\top[\widehat{\gamma}_p\mathcal{V} + \widehat{F}\widehat{T}\widehat{H}\widehat{\beta}_p] - P\widehat{H}\mathcal{A}\}$ em que $\mathcal{V}$ está definida na Seção 5.3.3.2. Assim, a matriz Δ fica expressa na forma apresentada em (5.15).

5.A.3.3. A p'–ésima coluna da matriz Z igual a p–ésima coluna da matriz X

Neste caso, o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}\ell_\delta(\mu,\phi) = \sum_{t=1}^{n}\{&\log\Gamma(\phi_t(\delta)) - \log\Gamma(\mu_t(\delta)\phi_t(\delta)) - \log\Gamma((1-\mu_t(\delta))\phi_t)\\ &+ (\mu_t(\delta)\phi_t(\delta)-1)\log(y_t) + \{(1-\mu_t(\delta))\phi_t(\delta)-1\}\log(1-y_t)\},\end{aligned}$$

em que $\mu_t(\delta)$ e $\phi_t(\delta)$ dependem de $\eta_t(\delta)$ e $\vartheta_t(\delta)$, que podem ser exemplificados segundo 5.12 e 5.16, respectivamente. Assim, temos que

$$\frac{\partial \ell(\beta,\gamma)}{\partial\delta_t} = a_t(\delta)\frac{1}{h'(\phi_t(\delta))}\gamma_{p'}s_{x_p} + \phi_t(\delta)(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p}. \tag{5.A12}$$

Note que a única diferença entre as expressões (5.A7) e (5.A12) é que na primeira aparece γ_p e na segunda, $\gamma_{p'}$. Ocorre o mesmo com a expressão de Δ_β em (5.15) e a referente ao esquema de perturbação atual, em que $\Delta_\beta = -s_{x_p}\{X^\top[\widehat{\beta}_p\widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\widehat{\gamma}_{p'}] - P\widehat{\Phi}\widehat{T}\mathcal{E}\}$.

Quanto as segundas derivadas com relação ao vetor γ temos que para $j = 1, \ldots, k$, com $j \neq p$,

$$\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_j} = -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}z_{tj}\beta_p s_{x_p} - \nu_t(\delta)z_{tj}\gamma_{p'}s_{x_p}\,. \tag{5.A13}$$

e quando $j = p$ temos que

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_p} &= -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}(x_{tp}+\delta_t s_{x_p})\beta_p s_{x_p} - \nu_t(\delta)(x_{tp}+\delta_t s_{x_p})\gamma_{p'}s_{x_p}\\ &+ a_t(\delta)\frac{1}{h'(\phi_t(\delta))}s_{x_p}.\end{aligned}$$

No entanto, considerando que $x_{tp} = z_{tp'}$, a expressão anterior pode ser escrita como

$$\begin{aligned}\frac{\partial^2 \ell_\delta(\beta,\gamma)}{\partial\delta_t\partial\gamma_p} &= -f_t(\delta)\frac{1}{g'(\mu_t(\delta))}\frac{1}{h'(\phi_t(\delta))}(z_{tp'}+\delta_t s_{x_p})\beta_p s_{x_p} - \nu_t(\delta)(z_{tp'}+\delta_t s_{x_p})\gamma_{p'}s_{x_p}\\ &+ a_t(\delta)\frac{1}{h'(\phi_t(\delta))}s_{x_p}.\end{aligned} \tag{5.A14}$$

Avaliando as expressões (5.A13) e (5.A14) em $\delta_0 = (0, 0, \ldots, 0)^\top$ e $\widehat{\theta}$ obtemos $\Delta_\gamma = -s_{x_p}\{Z^\top[\widehat{\gamma}_{p'}\mathcal{V} + \widehat{F}\widehat{T}\widehat{H}\widehat{\beta}_p] - \mathcal{P}\widehat{H}\mathcal{A}\}$ em que $\mathcal{P}$ está definida na Seção 5.3.3.3. Assim, a matriz Δ fica expressa na forma apresentada em (5.17).

5.A.3.3. A p'–ésima coluna da matriz Z é função da p–ésima coluna da matriz X

Neste caso, o logaritmo da função de verossimilhança do modelo perturbado é

$$\begin{aligned}\ell_\delta(\mu,\phi) = \sum_{t=1}^{n}\{&\log\Gamma(\phi_t(\delta)) - \log\Gamma(\mu_t(\delta)\phi_t(\delta)) - \log\Gamma((1-\mu_t(\delta))\phi_t)\\ &+ (\mu_t(\delta)\phi_t(\delta)-1)\log(y_t) + \{(1-\mu_t(\delta))\phi_t(\delta)-1\}\log(1-y_t)\},\end{aligned}$$

em que $\mu_t(\delta)$ e $\phi_t(\delta)$ dependem de $\eta_t(\delta)$ e $\vartheta_t(\delta)$, que podem ser exemplificados segundo 5.12 e 5.18, respectivamente. Assim, temos que

$$\frac{\partial \ell(\beta,\gamma)}{\partial\delta_t} = a_t(\delta)\frac{1}{h'(\phi_t(\delta))}\gamma_{p'}\mathcal{G}'(x_{tp}+\delta_t s_{x_p})s_{x_p} + \phi_t(\delta)(y_t^* - \mu_t^*(\delta))\frac{1}{g'(\mu_t(\delta))}\beta_p s_{x_p}. \tag{5.A15}$$

Com base em 5.A15 chegamos $\Delta_\beta = -s_{x_p}\{X^\top[\widehat{\beta}_p\widehat{\Phi}\widehat{Q} + \widehat{F}\widehat{T}\widehat{H}\dot{G}\widehat{\gamma}_{p'}] - P\widehat{\Phi}\widehat{T}\mathcal{E}\}$, em que $\dot{G}$ está definido na seção 5.3.3.4.

Quanto às segundas derivadas com relação ao vetor γ, temos que para $j = 1, \ldots, k$, com $j \neq p$,

$$\frac{\partial^2 \ell_\delta(\beta, \gamma)}{\partial \delta_t \partial \gamma_j} = -f_t(\delta) \frac{1}{g'(\mu_t(\delta))} \frac{1}{h'(\phi_t(\delta))} z_{tj} \beta_p s_{x_p} - \nu_t(\delta) z_{tj} \gamma_{p'} \mathcal{G}'(x_{tp} + \delta_t s_{x_p}) s_{x_p} \,. \tag{5.A16}$$

e quando $j = p$ temos que

$$\begin{aligned} \frac{\partial^2 \ell_\delta(\beta, \gamma)}{\partial \delta_t \partial \gamma_p} &= -f_t(\delta) \frac{1}{g'(\mu_t(\delta))} \frac{1}{h'(\phi_t(\delta))} \mathcal{G}(x_{tp} + \delta_t s_{x_p}) \beta_p s_{x_p} \\ &- \nu_t(\delta) \mathcal{G}(x_{tp} + \delta_t s_{x_p}) \gamma_{p'} \mathcal{G}'(x_{tp} + \delta_t s_{x_p}) s_{x_p} + a_t(\delta) \frac{1}{h'(\phi_t(\delta))} \mathcal{G}'(x_{tp} + \delta_t s_{x_p}) s_{x_p}. \end{aligned} \tag{5.A17}$$

Avaliando as expressões 5.A16 e 5.A17 em $\widehat{\theta}$ e $\delta_0 = (0, 0, \ldots, 0)^\top$ e considerando que $\mathcal{G}(x_{tp} + \delta_{0t} s_{x_p}) = \mathcal{G}(x_{tp}) = z_{tp'}$ obtemos $\Delta_\gamma = -s_{x_p}\{Z^\top[\widehat{\gamma}_{p'} \dot{G} \mathcal{V} + \widehat{F} \widehat{T} \widehat{H} \widehat{\beta}_p] - \mathcal{P} \widehat{H} \mathcal{A} \dot{G}\}$ em que $\mathcal{P}$ está definida na Seção 5.3.3.3. Assim, a matriz Δ fica expressa na forma apresentada em (5.19).

Capítulo 6

Intervalos de predição bootstrap para o modelo de regressão beta

6.1. Introdução

A inferência estatística tradicional foi desenvolvida fundamentalmente em termos da teoria das probabilidades. No entanto, o caráter matemático da teoria das probabilidades pode constituir um sério entrave para resultados práticos desejados pela estatística. Por outro lado, o processo empírico executado um grande número de vezes constitui uma alternativa para ultrapassar as suposições matemáticas complicadas ou ainda não muito bem desenvolvidas. Os métodos estatísticos computacionais surgiram exatamente a partir da necessidade de simular a experimentação diversas vezes, a fim de garantir a confiança nos resultados fornecidos pela teoria matemática. 'Bootstrap' é um método computacional de inferência estatística introduzido por Efron (1979) capaz de responder a muitas questões reais sem a necessidade de exaustivos, complicados e muitas vezes inviáveis cálculos analíticos. Efron e Tibshirani (1993, p. 2) escrevem: "The basic ideas of statistics haven't changed, but their implementation has. The modern computer lets us apply these ideas flexibly, quickly, easily, and with a minimum of mathematical assumptions."

Bootstrap é um método de estimação de variâncias, intervalos de confiança, p–valores e outros aspectos de interesse da inferência estatística através de reamostragem direta dos dados, os quais são tratados como se fossem a própria população. Tipicamente, este método fornece aproximações para a distribuição da estatística de interesse que podem ser consideravelmente mais precisas que as aproximações assintóticas de primeira ordem. De fato, testes de hipótese e intervalos de confiança baseados na teoria assintótica convencional, em alguns casos, podem conduzir a resultados enganosos quando o tamanho de amostra não é muito grande. Existem casos extremos de testes que mesmo para 50 observações ao nível nominal de 5% chegam a rejeitar a hipótese nula mais de 80% das vezes, sendo esta hipótese verdadeira (Davidson e MacKinnon, 2002). Adicionalmente, em algumas situações a teoria assintótica é intratável e, nestes casos, a inferência por bootstrap é justificada e fornece resultados equivalentes àqueles que seriam obtidos através das aproximações de primeira ordem. Assim, bootstrap é uma excelente alternativa para os casos em que a teoria assintótica tradicional não é viável e onde, apesar de viável, não é muito precisa para os tamanhos de amostras disponíveis. A aplicação de bootstrap para modelos de regressão foi estudada em

detalhes por Wu (1986) e mais recentemente muitos autores têm investigado o uso deste método em econometria, entre eles Horowitz (1997) Berkowitz e Kilian (2000), Davidson e MacKinnon (2002) e Monfardini (2003).

Neste capítulo utilizaremos o método bootstrap para obter limites de predição para novos valores da variável resposta com base no modelo de regressão beta com dispersão constante definido em (1.2). Um exemplo de aplicação deste método no contexto de regressão beta é um problema relacionado ao consumo de gás natural. Dada a potência nominal dos aparelhos instalados nos apartamentos ou edificações, a companhia de gás tem que predizer a "probabilidade" do uso concomitante de vários aparelhos a gás numa instalação e, então definir a potência máxima que deve ser fornecida. Apresentaremos aqui uma proposta de obtenção de limites de predição para o modelo de regressão beta baseada no mecanismo do bootstrap e utilizaremos dados reais para avaliar o desempenho do método proposto.

6.2. O método bootstrap

Considere uma amostra aleatória $x = (x_1, \ldots, x_n)$ de uma variável aleatória populacional X que apresenta comportamento probabilístico descrito completamente, por exemplo, por sua função de distribuição acumulada, denotada por $\mathbb{F}$.

Sejam $\theta = t(\mathbb{F})$ uma função de $\mathbb{F}$ denominada parâmetro e $\widehat{\theta} = S(x)$ um estimador de θ. Se $\mathbb{F}$ é desconhecida, as propriedades distribucionais de $\widehat{\theta}$ podem ser obtidas utilizando teoria assintótica de primeira ordem. No entanto, muitas vezes encontrar a distribuição assintótica é tarefa difícil e até mesmo inviável. A aplicação de bootstrap, neste caso, consiste em obter, a partir da amostra original x, um grande número de amostras $x^* = (x_1^*, \ldots, x_n^*)$, calcular as respectivas réplicas bootstrap de $\widehat{\theta}$, $\widehat{\theta}^* = S(x^*)$, e, com base na distribuição empírica de $\widehat{\theta}^*$, estimar a função de distribuição de $\widehat{\theta}$.

Existem duas versões do método bootstrap que se diferenciam unicamente quanto à forma de obtenção das amostras. Na primeira versão, a amostra bootstrap x^* é obtida de uma estimativa não-paramétrica de $\mathbb{F}$, a função de distribuição empírica da amostra original x, dada por

$$\widehat{\mathbb{F}}(\iota) = \frac{\#\{x_i \leq \iota\}}{n}, \ \iota \in \mathbb{R},$$

que atribui probabilidade $1/n$ a cada valor x_i, $i = 1, \ldots, n$. Afirmar que x^* é uma amostra aleatória de $\widehat{\mathbb{F}}$ equivale a dizer que a amostra x^* é formada por n observações extraídas aleatoriamente com reposição da amostra x, que é tratada como se fosse a população. Este é o esquema de boostrap não-paramétrico, que é uma aplicação direta de um princípio denominado "plug in". Este princípio utiliza $\widehat{\mathbb{F}}$ no lugar da distribuição desconhecida populacional para estimar características de interesse da mesma. Se temos que $\theta = t(\mathbb{F})$, a

respectiva estimativa "plug in" é $\widehat{\theta} = t(\widehat{I\!F})$. A estimativa bootstrap não-paramétrica de qualquer característica de $I\!F$ é uma estimativa "plug in", uma vez que todo processo de estimação parte das amostras x^* extraídas de $\widehat{I\!F}$. A segunda versão é viável quando $I\!F$ pertence a uma família paramétrica dimensionalmente finita e conhecida, $I\!F_\tau$. A obtenção de um estimador consistente para τ viabializa uma estimativa paramétrica de $I\!F$, representada por $I\!F_{\widehat{\tau}}$. É importante ressaltar que a existência de um modelo paramétrico não exclui a análise não-paramétrica, que pode ser muito útil para se avaliar a robustez das conclusões obtidas a partir do método paramétrico.

A utilização do método bootstrap se justifica quando a teoria assintótica é intratável ou quando, apesar de viável, as aproximações assintóticas de primeira ordem são imprecisas para os tamanhos amostrais disponíveis. Quando, por exemplo, a teoria assintótica fornece uma aproximação imprecisa para a distribuição de uma estatística de teste, as diferenças entre o nível empírico do teste (realizado com base em valores críticos assintóticos) e o nível desejado (nominal) podem ser substanciais. A aplicação de bootstrap, neste caso, é de grande utilidade, uma vez que o método pode reduzir consideravelmente, ou até eliminar, as distorções de tamanho de testes estatísticos em amostras finitas. No entanto, nestas circunstâncias, em quea distribuição assintótica pode ser obtida, não se pode esperar que o método bootstrap forneça aproximações mais precisas, a não ser que a distribuição da estatística em questão seja assintoticamente pivotal, i.e., independente de parâmetros desconhecidos. Segundo Horowitz (1994), procedimentos de bootstrap simples fornecem aproximações melhoradas para a distribuição de estatísticas assintoticamente pivotais, mas não para a distribuição de estatísticas que não apresentam esta propriedade. Beran (1988) mostra que se a distribuição assintótica da estatística, sob a hipótese nula, é pivotal, então, sob algumas condições de regularidade, os tamanhos de testes bootstrap apresentam erros de ordem menor, i.e., que convergem mais rapidamente para zero, que os erros dos testes baseados na teoria assintótica.

6.3. Intervalos de predição bootstrap

Usualmente o modelo de regressão ajustado a um certo conjunto de dados é utilizado para predizer novos valores da resposta. Para o modelo de regressão beta definido em (1.2) temos que $x_t^\top = (x_{t1}, \ldots, x_{tk})$ são observações de k covariáveis ($k < n$) e y_t é o valor observado da t–ésima resposta $t = 1, \ldots, n$. Denotamos por $(x_{+1}, \ldots, x_{+k})$ uma nova observação dessas k covariáveis, pertencente ou não ao conjunto de dados original e por y_+ a respectiva resposta não observada. Esse novo valor da resposta pode ser predito por $\widehat{\mu}_+ = g^{-1}(\sum_{i=1}^k x_{+i}\widehat{\beta}_i)$, em que $\widehat{\beta}_i$ é o estimador de máxima verossimilhança de β_i obtido com base no conjunto de dados original. Além da previsão pontual é interessante fornecer limites, usualmente chamados de

limites de predição, baseados em níveis de confiança; esses limites formarão um intervalo de predição, uma estimativa intervalar para a predição. Para acessar a precisão do ponto predito e construir o intervalo de predição utilizamos a distribuição aproximada do erro de predição.

Considere $\mathcal{R}(y, \mu)$ alguma função monótona em y que possui variância constante ao longo das observações. Suponhamos que a média μ_+ e a distribuição de $\mathcal{R}(y, \mu)$ são conhecidas e que $\wp_\alpha$ é o α-ésimo quantil dessa distribuição. Então, os limites de predição, para um intervalo com nível nominal $1 - \alpha$, são os valores $y_{+,\alpha/2}$ e $y_{+,1-\alpha/2}$ que satisfazem, respectivamente, $\mathcal{R}(y, \mu_+) = \wp_{\alpha/2}$ e $\mathcal{R}(y, \mu_+) = \wp_{(1-\alpha/2)}$. Se μ é estimado por $\widehat{\mu}$ independentemente de y_+ e se $\mathcal{R}(y_+, \widehat{\mu})$ tem quantis conhecidos, o mesmo método se aplica. Se não conhecemos a distribuição de $\mathcal{R}(y_+, \widehat{\mu})$, a mesma poderia ser aproximada tanto por métodos assintóticos quanto por métodos de reamostragem. O método bootstrap se enquadra no segundo contexto. Este aqui será utilizado para obter uma aproximação para a distribuição de $\mathcal{R}(y_+, \widehat{\mu})$ e, assim, fornecer os quantis empíricos utilizados para a obtenção dos limites de predição. A função $\mathcal{R}(y, \mu)$ corresponde a alguma definição dos resíduos do modelo e o algoritmo bootstrap utilizará para o processo de reamostragem uma versão padronizada da função $\mathcal{R}(y, \widehat{\mu})$, cuja distribuição tenha variância aproximadamente constante.

Para o modelo de regressão beta consideraremos

$$\mathcal{R}(y, \widehat{\mu}) = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{v_t}},$$

definição do resíduo ponderado padronizado 1, em que os t–ésimos elemento dos vetores y^*, μ^* e v estão definidos em (1.5) e (1.11). No processo de reamostragem utilizaremos

$$r^{pp} = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{v_t(1 - h_{tt}^*)}},$$

o resíduo ponderado padronizado 2 que pode ser considerado uma padronização do resíduo ponderado padronizado 1.

Desenvolvemos o método bootstrap para a obtenção de limites de predição empíricos para o modelo de regressão beta com base no método proposto por Davison e Hinkley (1997, p. 340) para modelos lineares generalizados. Considerando B réplicas booststrap e n_2 predições, sumarizamos a seguir o método bootstrap para a obtenção de limites de predição empíricos que propomos para o modelo de regressão beta:

Para $b = 1, \ldots B$:

1. Para cada $t = 1, \ldots, n$, selecione aleatoriamente $r_{t,b}^{pp}$ a partir de $r_1^{pp}, \ldots, r_n^{pp}$.

2. Forme uma amostra bootstrap (y_b, X), tal que

$$y_{t,b} = \frac{\exp(\widehat{\mu}_t^* + r_{t,b}^{pp}\sqrt{\widehat{v}_t})}{1 + \exp(\widehat{\mu}_t^* + r_{t,b}^{pp}\sqrt{\widehat{v}_t})}$$

é obtido a partir da solução da equação $\mathcal{R}(y, \widehat{\mu}) = r_{t,b}^{pp}$.

3. Usando (y_b, X) encontre $\widehat{\beta}_b$ e $\widehat{\phi}_b$, estimativas bootstrap de β e ϕ. Em seguida considerando a matriz $n_2 \times k$ de novas observações X_+, $\widehat{\beta}_b$ e $\widehat{\phi}_b$ obtenha $\widehat{\mu}_{+,b}$, $\widehat{\mu}_{+,b}^*$ $\widehat{v}_{+,b}$, neste caso vetores de dimensão n_2.
4. Para cada nova observação $a_+ = 1, \ldots, n_2$:
   a. Selecione aleatoriamente $r_{a_+,b}^{pp}$ a partir de $r_1^{pp}, \ldots, r_n^{pp}$.
   b. Obtenha

$$y_{a_+,b} = \frac{\exp(\widehat{\mu}_{a_+,b}^* + r_{a_+,b}^{pp}\sqrt{\widehat{v}_{a_+,b}})}{1 + \exp(\widehat{\mu}_{a_+,b}^* + r_{a_+,b}^{pp}\sqrt{\widehat{v}_{a_+,b}})}.$$

   c. Calcule o erro de predição simulado

$$\mathcal{R}_{a_+,b}(y_{a_+,b}, \widehat{\mu}_{a_+,b}^*) = \frac{y_{a_+,b}^* - \widehat{\mu}_{a_+,b}^*}{\sqrt{\widehat{v}_{a_+,b}}},$$

   em que $y_{a_+,b}^* = \log\left\{y_{a_+,b}/(1 - y_{a_+,b})\right\}$.

Para cada nova observação ordene os B valores $\mathcal{R}_{a_+}$, tal que, $\mathcal{R}_{a_+(1)} \leq \ldots \leq \mathcal{R}_{a_+(B)}$, obtenha os quantis bootstrap $\wp_{a_+(\alpha/2)}^* = \mathcal{R}_{a_+(B(\alpha/2))}$ e $\wp_{a_+(1-\alpha/2)}^* = \mathcal{R}_{a_+(B(1-\alpha/2))}$ e encontre os limites de predição

$$y_{a_+,I} = \frac{\exp(\widehat{\mu}_{a_+}^* + \wp_{a_+(\alpha/2)}^*\sqrt{\widehat{v}_{a_+}})}{1 + \exp(\widehat{\mu}_{a_+}^* + \wp_{a_+(\alpha/2)}^*\sqrt{\widehat{v}_{a_+}})} \text{ e } y_{a_+,S} = \frac{\exp(\widehat{\mu}_{a_+}^* + \wp_{a_+(1-\alpha/2)}^*\sqrt{\widehat{v}_{a_+}})}{1 + \exp(\widehat{\mu}_{a_+}^* + \wp_{a_+(1-\alpha/2)}^*\sqrt{\widehat{v}_{a_+}})},$$

em que $\widehat{\mu}_{a_+}^*$ e $\widehat{v}_{a_+}$ são as quantidades μ^* e v avaliadas em $\widehat{\mu}_{a_+} = g^{-1}(x_{a_+}^\top\widehat{\beta})$ e $\widehat{\phi}$, com a $x_{a_+}^\top$ a_+-ésima linha da matriz X_+ de novas observações, $a_+ = 1, \ldots, n_2$. Os valores $y_{a_+,I}$ e $y_{a_+,S}$, são obtidos, respectivamente, como solução das equações $\mathcal{R}(y_{a_+}, \widehat{\mu}_{a_+}) = \wp_{a_+(\alpha/2)}^*$ e $\mathcal{R}(y_{a_+}, \widehat{\mu}_{a_+}) = \wp_{a_+(1-\alpha/2)}^*$. A forma de obtenção dos quantis que descrevemos acima caracteriza o método denominado percentil.

6.4. Aplicação. Dados de indigência

Nesta aplicação utilizaremos um conjunto de dados reais para avaliar o método bootstrap proposto na seção anterior, como descreveremos mais adiante. Inicialmente apresentaremos os modelos que ajustamos a esse conjunto de dados e que validamos através da análise de diagnóstico desenvolvida nos capítulos anteriores.

6.4.1. Escolha do modelo

Os dados em questão foram obtidos a partir do ATLAS de Desenvolvimento Humano do Brasil, produto do Programa das Nações Unidas para o Desenvolvimento que tem como objetivo central o combate à pobreza. O ATLAS permite o acesso a 125 indicadores sociais e econômicos para os 5507 municípios brasileiros e as 27 unidades da Federação baseados nas informações dos Censos Demográficos de 1991 e de 2000 do IBGE (Instituto Brasileiro de Geográfia e Estatística). O conjunto de dados escolhido é formado pelos 645 municípios do estado economicamente mais importante do Brasil, o Estado de São Paulo, e o ano de referência é 2000. Nosso interesse é avaliar em que extensão indicadores relacionados a educação, em especial, a educação fundamental (crianças até quatorze anos), e condições relacionadas à infância, como, por exemplo, trabalho infantil e maternidade antes de quatorze anos, explicam o percentual de indigência do município.

A variável resposta é o percentual de indigentes, proporção de indivíduos com renda domiciliar per capta inferior a R$37.75 (1/4 do salário mínimo em agosto de 2000). Como covariáveis escolhemos **analfabetismo** (x_2): proporção de pessoas com vinte e cinco anos ou mais que não sabem ler nem escrever um bilhete simples, **analfabetismo 7–14** (x_3): proporção de pessoas entre sete e quatorze anos que não sabem ler nem escrever um bilhete simples, **mães jovens** (x_4): percentual de crianças do sexo feminino entre dez e quatorze anos de idade que já tiveram filhos, **trabalho infantil** (x_5): percentual de crianças entre dez e quatorze anos de idade que trabalharam em todos ou em parte dos últimos doze meses de 2000 (considera-se o trabalho remunerado ou não), **crianças fora da escola 7–14** (x_6): percentual de crianças entre sete e quatorze anos de idade que não freqüentam escola e **mais de doze anos de estudo** (x_7): percentual de pessoas entre dezoito e vinte e dois anos de idade com doze ou mais anos de estudo. Todas as covariáveis assumem valores entre zero e um.

Inicialmente consideramos o modelo normal linear, que não se mostrou adequado aos dados uma vez que a distribuição condicional da variável resposta é fortemente assimétrica. Como conseqüência, foram obtidos valores ajustados negativos. Em seguida consideramos o modelo de regressão beta com ligação logito. Ressaltamos que a covariável mães jovens não se revelou significativa para a explicação do comportamento da resposta (p–valor igual a 0.1499), enquanto a covariável analfabetismo 7–14 se mostrou significativa apenas ao nível de 10% (p–valor igual a 0.0710). Realizamos a análise de influência considerando o modelo completo. Na Figura 6.1 encontram-se os gráficos de I_{max} em relação a β contra os índices das observações; a Figura 6.1a corresponde a ponderação de casos, a Figura 6.1b corresponde a perturbação da covariável analfabetismo 7–14 e a Figura 6.1c, mães jovens.

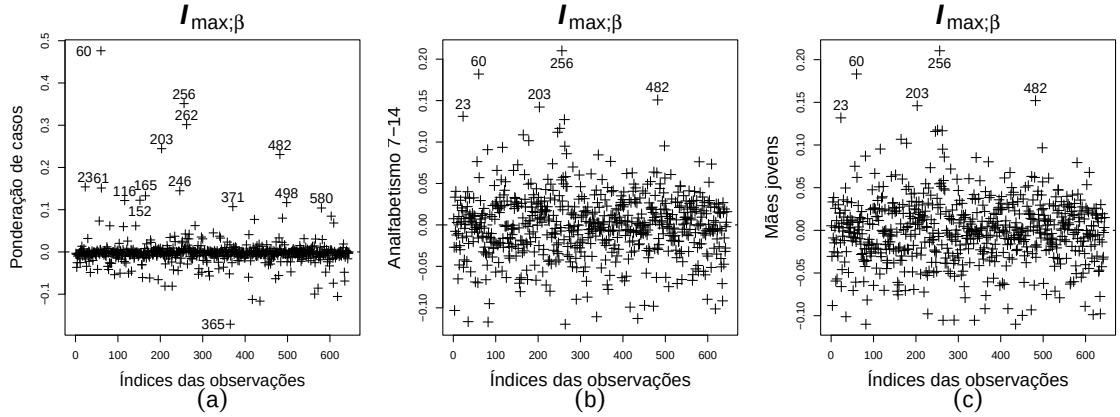

Figura 6.1 Gráficos de influência. Dados de indigência.

Nota-se a influência dos conjuntos de observações $\{23, 60, 203, 256, 482\}$ (Figura 6.1b,c) e $\{23, 60, 61, 116, 152, 165, 203, 246, 256, 262, 371, 482, 498\}$ (Figura 6.1a). A exclusão separada dos dois conjuntos conduz ao mesmo resultado, as covariáveis mães jovens e analfabetismo 7–14 passam a ser significativas ao nível de 5%. Investigando as observações influentes constatamos que o conjunto $\{23, 60, 203, 256, 482\}$ corresponde aos maiores percentuais de indigência de São Paulo, superiores a 28%, enquanto que a média do estado é igual a 7.1%, e todas as observações pertencem à microregião denominada Capão Bonito. Mais especificamente, entre os dez municípios que formam a microregião de Capão Bonito nove fazem parte do conjunto maior destacado como influente e que apresentam percentuais de indigência superiores a 20%.

Assim, decidimos investigar a introdução de uma covariável indicadora da região de Capão Bonito (x_8) ao modelo inicial (Tabela 6.1). A partir da Tabela 6.1 notamos que a inclusão da covariável indicadora de Capão Bonito revela que as covariáveis analfabetismo 7–14 e mães jovens podem ser importantes para a explicação da resposta. Realizamos uma análise de influência para o novo modelo (Figuras 6.2 e 6.3). Notamos, a partir das Figuras 6.2 e 6.3, que tanto o esquema de ponderação de casos quanto a perturbação das covariáveis destacam um número menor de observações como influentes. Adicionalmente, as influências dos casos 60 e 256 são consideravelmente amenizadas e agora destacam-se mais expressivamente os casos 61 e 484. Ainda são destacadas outras observações como influentes, entre elas os casos 81, 82, 152, 165, 178, 251, 264, 267 e 488. Apresentamos na Tabela 6.2 as exclusões mais relevantes. Nota-se que as exclusões individuais dos casos 61, 256 e 484 conduzem ao mesmo resultado: a covariável mães jovens é significativa ao nível de 5%.

Tabela 6.1. Resultados inferenciais para o modelo estendido. Dados de indigência.

Parâmetros	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ϕ
Estimativas	-3.406	5.474	2.868	7.904	-3.521	6.416	-2.753	0.994	78.280
p–valor	0.0000	0.0000	0.0167	0.0515	0.0001	0.0000	0.0000	0.0000	

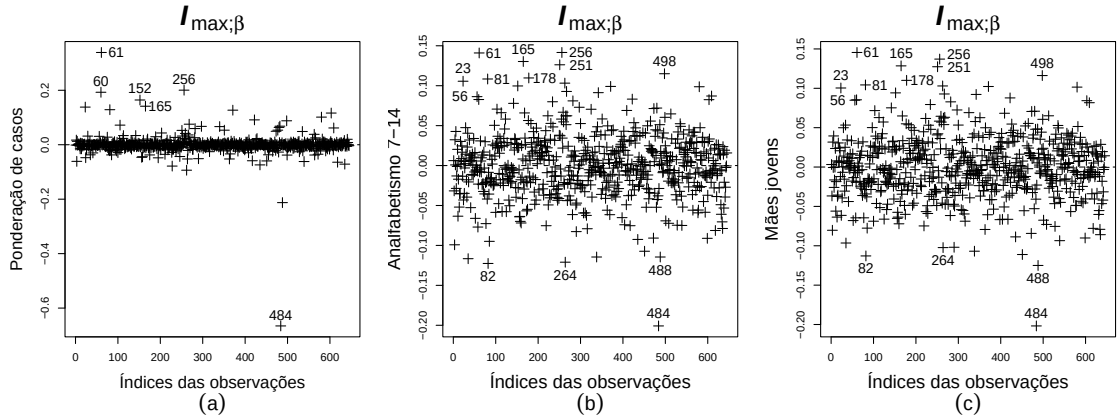

Figura 6.2. Gráficos de influência local para o modelo estendido. Dados de indigência.

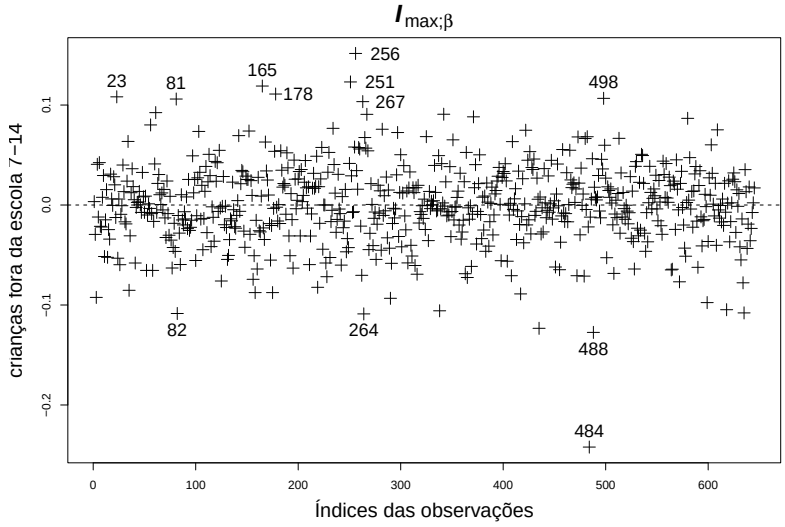

Figura 6.3. Gráficos de influência local para o modelo estendido. Dados de indigência.

Investigando os casos influentes, notamos que a observação 61 representa o quarto maior percentual de indigência de São Paulo, mas não pertence à região de Capão Bonito. Já os casos 256 e 484, apesar de pertencerem a Capão Bonito, são observações em que os valores

Tabela 6.2. Variação percentual das estimativas dos parâmetros para o modelo estendido e p–valores retirando observações influentes. Dados de indigência.

Parâmetros	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
484	1.4	-0.8	12.2	17.9	-2.2	9.2	-5.0	6.0
p–valor	0.0000	0.0000	0.0071	0.0212	0.0001	0.0000	0.0000	0.0000
256	0.5	-0.5	4.5	0.9	-1.7	3.3	-2.3	-6.9
p–valor	0.0000	0.0000	0.0123	0.0486	0.0001	0.0000	0.0000	0.0000
61	-1.6	-4.2	-3.3	2.2	-7.1	-11.6	5.2	5.0
p–valor	0.0000	0.0000	0.0206	0.0464	0.0003	0.0000	0.0000	0.0000

das covariáveis contradizem as relações estimadas entre as mesmas e a resposta. O caso 256 corresponde ao município de São Paulo com o maior percentual de indigentes, 42.2%, mas com valores das covariáveis, em geral, muito próximos da média, o que torna esta observação atípica. Já o caso 484 representa uma contradição no sentido oposto: apresenta o maior percentual de crianças fora da escola, 13.5% (a média do estado é 3.2%), o terceiro maior percentual de analfabetismo entre sete e quatorze anos, 10.2% (a média do estado é 4.5%), um dos maiores percentuais de mães jovens 1.3%, (a média do estado é 0.3%), um dos menores percentuais de pessoas com mais de doze anos de estudo 1.3% (média do estado 7.3%), mas apresenta apenas o oitavo maior percentual de indigência do estado, igual a 27.8%.

A análise de influência revelou a necessidade da inclusão de uma covariável indicadora da região de Capão Bonito. A introdução da covariável indicadora amenizou consideravelmente a influência das observações permitindo visualizar a importância das covariáveis analfabetismo 7–14 e mães jovens para a explicação da resposta. No modelo com a variável indicadora, a não–significância ao nível de 5% da covariável mães jovens (Tabela 6.1) está sendo mascarada, em especial, pela observação 484. Desta forma, acreditamos que o modelo cujas estimativas encontram-se apresentadas na Tabela 6.1 é útil no sentido de explicar como a probreza está relacionada à educação e à infância no Estado de São Paulo. Deve-se notar, por exemplo, a influência que as covariáveis mães jovens, crianças fora da escola 7–14 e analfabetismo exercem no percentual de indigentes. Um fato interessante e pouco desejável é o peso do trabalho infantil sobre a pobreza, uma vez que o modelo ajustado indica que quanto maior o percentual de crianças trabalhando menor o percentual médio de indigentes.

6.4.2. Método bootstrap

Considerando os dados de indigência e com base no modelo ajustado apresentado na Tabela 6.1, avaliaremos o desempenho do método bootstrap de obtenção de limites de predição. Para

isso dividimos os dados em dois sub–conjuntos, o primeiro utilizado para estimar o modelo e o segundo para obter intervalos de predição. Com o objetivo de cobrir todas as observações tomamos 13 partições dos dados. Considerando a seqüência das observações formamos 12 conjuntos com 595 observações para estimar o modelo e com 50 observações para construir os intervalos de predição e um conjunto com as 600 observações iniciais para estimar o modelo e as 45 observações finais a serem preditas. Esse procedimento permite avaliar como se comportam os intervalos de predição sobre todo o conjunto de dados e identificar em quais circunstâncias e por que o desempenho do intervalo é melhor ou pior.

Para avaliar o desempenho dos intervalos de predição calculamos a "cobertura empírica" para cada partição. Uma vez que conhecemos os valores da resposta do sub–conjunto utilizado para obter os intervalos de predição, definimos como cobertura empírica a razão entre o número de vezes que o verdadeiro valor da resposta é coberto por seu respectivo intervalo de predição e o número total de observações desse sub–conjunto. Consideramos $B = 1000$ réplicas bootstrap e um nível nominal igual a 95%. Considerando as 13 partições, a cobertura empírica variou de 86% a 100%. Os piores desempenhos se referem às seguintes partições: $51 \vdash\dashv 100$, $251 \vdash\dashv 300$, $451 \vdash\dashv 500$. As coberturas empíricas dessas partições foram, respectivamente, 88%, 86% e 90%. Ocorre que nesses três sub–conjuntos estão algumas das observações detectadas como mais influentes para o modelo ajustado aos dados completos (Tabela 6.1). Em especial, a pior cobertura está associada à presença de quatro observações consideravelmente influentes: as observações 256, 251, 264 e 267 (Figuras 6.2b e 6.3). Desconsiderando esses três sub–conjuntos, a menor cobertura empírica verificada foi 94%.

Na Figura 6.4 apresentamos as curvas de predição, formadas pelos limites superior e inferior de predição e o verdadeiro valor da resposta para cada observação predita, para os três sub–conjuntos: $51 \vdash\dashv 100$ (a), $251 \vdash\dashv 300$ (b) e $451 \vdash\dashv 500$ (c). Essa figura ilustra as assertivas anteriores, uma vez que a curva de predição referente ao conjunto que contém a observação 256 é a que apresenta o pior desempenho, em geral relacionado à subestimação do limite superior do intervalo de predição. Com relação à partição $451 \vdash\dashv 500$ (Figura 6.4b), nota-se que para as observações 484 e 488 o limite inferior do intervalo superou o valor verdadeiro e para a observação 498 o limite superior foi subestimado. Na Figura 6.4a alguns limites superiores de predição são subestimados; chamamos atenção especial para os casos 61 e 81.

Na Figura 6.5 apresentamos as curvas de predição para algumas partições em que as coberturas empíricas ou estiveram próximas do nível nominal ou chegaram a 100%. Escolhemos as partições $151 \vdash\dashv 200$ (a), $501 \vdash\dashv 550$ (b), $551 \vdash\dashv 600$ (c), cujas coberturas empíricas foram, respectivamente, 94%, 100% e 98%. Apesar da cobertura referente à partição $151 \vdash\dashv 200$ estar próxima do nível nominal, é notável a subestimação dos lim-

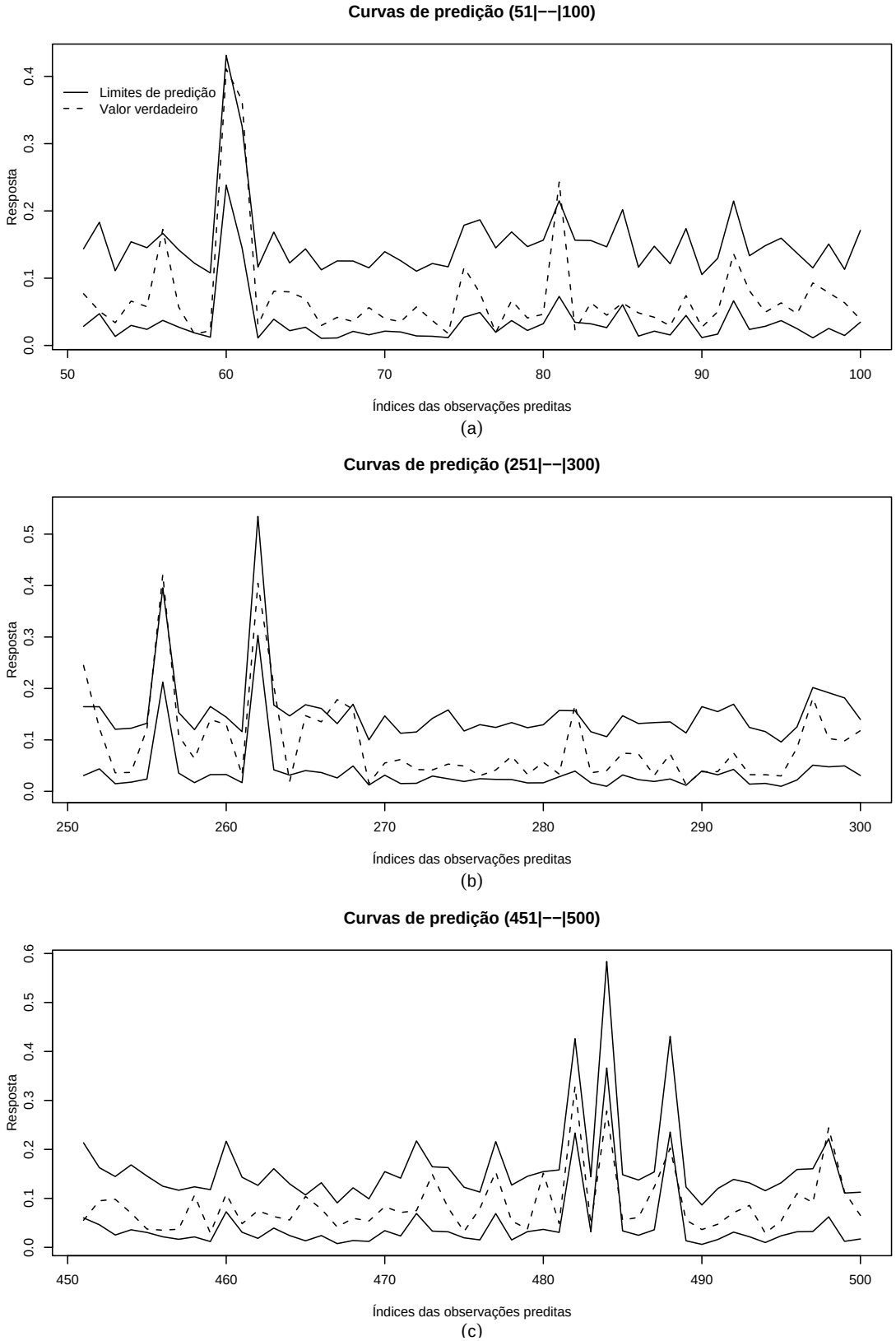

Figura 6.4. Curvas de predição. Dados de indigência.

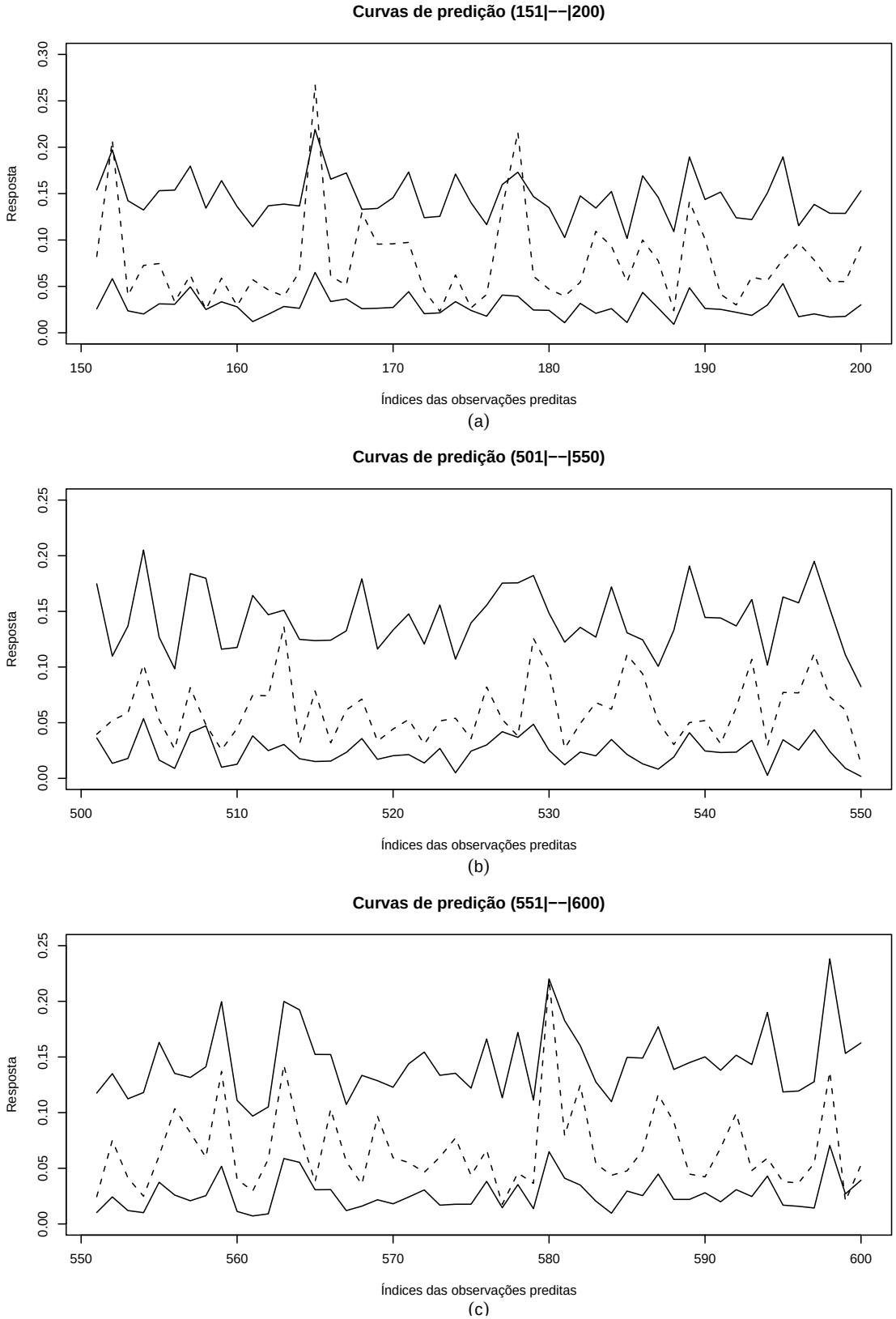

Figura 6.5. Curvas de predição. Dados de indigência.

ites superiores de predição das observações 152, 165 e 178 (Figura 6.5a), casos influentes no modelo com todas as observações. Por outro lado, quando as coberturas estão próximas de 100% as amplitudes dos intervalos de predição são grandes (Figura 6.5b,c).

Neste capítulo desenvolvemos um procedimento de obtenção de limites de predição para o modelo de regressão beta utilizando o método bootstrap proposto por Davison e Hinkley (1997) para modelos lineares generalizados. Notamos que o intervalo de predição bootstrap tem, em geral, bom desempenho, mas tende a falhar quando usado para prever casos que seriam considerados influentes se estivessem presentes no conjunto de dados observados. Devemos ressaltar que o trabalho desenvolvido neste capítulo é apenas o início de uma pesquisa sobre a obtenção de limites de predição para a classe de modelos de regressão beta. Inicialmente, vamos investigar uma correção do erro de predição através de medidas de influência para melhorar o desempenho do método apresentado na Seção 6.3. Pretendemos realizar simulações de Monte Carlo com o objetivo de avaliar melhor o desempenho dos métodos desenvolvidos. Pretendemos ainda aplicar outro método de obtenção dos limites de predição, por exemplo, o método percentil t ou o método BC_a ("bias-corrected accelerated"), tanto ao bootstrap tradicional quanto ao bootstrap corrigido pela medida de influência. O método BC_a, que no contexto de intervalos de predição foi proposto por Mojirsheibani and Tibshirani (1996), tipicamente corrige as distorções de tamanho associadas ao método percentil.

Capítulo 7

Considerações finais

7.1. Conclusões

Ao longo dos capítulos anteriores desenvolvemos diversos aspectos de inferência e diagnóstico para a classe de modelos de regressão beta. Resumimos nos itens a seguir as principais conclusões desse trabalho:

1. No Capítulo 2, propomos dois novos resíduos para o modelo de regressão beta (Ferrari e Cribari–Neto, 2004) com base no processo iterativo Scoring de Fisher. Avaliamos numericamente os desempenhos desses novos resíduos e também dos resíduos componente do desvio e ordinário padronizado propostos por Ferrari e Cribari–Neto (2004). Os resultados evidenciaram a inviabilidade do resíduo componente do desvio e favoreceram os novos resíduos, em especial o resíduo ponderado padronizado 2. Além de sua distribuição empírica estar mais próxima da distribuição normal padrão que as dos demais resíduos, o uso do resíduo ponderado padronizado 2 é mais indicado se o objetivo é a captação de casos decisivos no ajuste do modelo. Quanto à avaliação da qualidade de ajuste do modelo, os resíduos ordinário padronizado, ponderado padronizado 1 e ponderado padronizado 2 conduzem às mesmas conclusões.
2. No Capítulo 3, desenvolvemos para a classe de modelos de regressão beta (Ferrari e Cribari–Neto, 2004) uma medida de influência baseada na deleção de casos (distância de Cook) e medidas de influência baseadas no método de influência local desenvolvido por Cook (1986). A partir dos resultados obtidos recomendamos medidas de influência local considerando três esquemas de perturbação, a saber: a ponderação de casos, perturbação da resposta e perturbação individual de covariáveis. A aproximação para a distância de Cook e a alavanca generalizada também podem ser usadas; no entanto, as medidas de influência local, em geral, são suficientes para identificar observações influentes. Adicionalmente, ressaltamos a importância do uso das medidas de influência local em gráficos de diagnóstico no sentido de identificar possíveis tendências dos dados como, por exemplo, dispersão variável ao longo das observações.
3. No Capítulo 4, através de estudos de simulação comprovamos que os estimadores de máxima verossimilhança para o modelo de regressão beta com dispersão variável (Paolino, 2001 e Smithson e Verkuilen, 2006) apresentam um bom desempenho quanto à

estimação intervalar da resposta média e que estimar o modelo supondo dispersão constante quando, de fato, a dispersão é variável interfere drasticamente na precisão dos intervalos de confiança para a resposta média. Avaliamos ainda o desempenho dos testes assintóticos da razão de verossimilhanças, escore e Wald para testar a hipótese de dispersão constante. Nossos resultados favorecem o teste escore, uma vez que este apresenta resultados satisfatórios tanto no que diz respeito ao poder quanto em relação aos tamanhos. Finalmente, investigando o comportamento de alguns gráficos de diagnóstico verificamos que, à medida que o grau de heterogeneidade da dispersão aumenta, o gráfico normal de probabilidades com envelopes simulados evidencia mais nitidamente a falta de qualidade do ajuste supondo dispersão constante. Adicionalmente, a diferença das dispersões ao longo das observações fica mais evidente nos gráficos que utilizam a medida de influência I_{max} que nos gráficos dos resíduos.

4. No Capítulo 5, desenvolvemos métodos de diagnóstico para o modelo de regressão beta com dispersão variável. Adaptamos o resíduo ponderado padronizado 2 para a situação em que a dispersão varia ao longo das observações. Adicionalmente, desenvolvemos medidas de influência local considerando três esquemas de perturbação: a ponderação de casos, perturbação da resposta e perturbação individual de covariáveis, este último considerando os diversos cenários em relação às covariáveis envolvidas na modelagem da média e da dispersão. Através de exemplos com dados reais, comprovamos a sensibilidade da metodologia e a importância dos diferentes esquemas de perturbação no sentido de captar observações importantes quanto ao ajuste do modelo aos dados.
5. No Capítulo 6, propomos um método de obtenção de limites de predição para o modelo de regressão beta com dispersão constante a partir do método bootstrap. Constatamos com base em uma aplicação a dados reais que o método, em geral, tem bom desempenho, mas é prejudicado quando os casos preditos seriam considerados influentes se m presentes no conjunto de dados observados.

7.2. Trabalhos futuros

Serão foco de nossas pesquisas futuras:

1. Estender a classe de modelos de regressão beta proposta por Ferrari e Cribari–Neto (2004) para permitir que a média da variável resposta seja definida por um preditor *não-linear* envolvendo regressores e parâmetros de regressão desconhecidos. A estimação conjunta de todos os parâmetros do modelo será feita por máxima verossimilhança, devendo ser obtidas expressões de forma fechada para a função escore, para a matriz de informação de Fisher e sua inversa. Serão estudados também mecanismos de diagnóstico e identificação de observações influentes. Inferência via testes de hipóteses assintóticos será também

estudada.

2. Desenvolver a classe de testes não encaixados no contexto dos modelos de regressão beta, em especial o teste J para modelos lineares e o teste P para o contexto não–linear propostos por Davidson e MacKinnon (1981) e testes não encaixados baseados no método bootstrap (Davidson e MacKinnon, 2002 e Monfardini, 2003). Essa classe de testes permite a avaliação de formulações mais amplas para o modelo que está sendo investigado, uma vez que consideram modelos claramente distintos para as hipóteses nula e alternativa, ou seja, que não são casos particulares um do outro. Pretendemos avaliar o desempenho dos diferentes testes através de simulações de Monte Carlo e aplicar a teoria desenvolvida a dados reais.
3. Desenvolver um esquema bootstrap para obtenção de limites de predição considerando uma correção do erro de predição através de medidas de influência. Investigar os intervalos de predição do tipo BC_a e percentil t.

Apêndice A

Dados

A.1. Proporção da renda familiar total gasta com alimentação

Tabela A.1. Dados de gastos com alimentação

Obs	Renda Total	Número de pessoas	Gastos com alimentação
1	62.476	1	15.998
2	82.304	5	16.652
3	74.679	3	21.741
4	39.151	3	7.431
5	64.724	5	10.481
6	36.786	3	13.548
7	83.052	4	23.256
8	86.935	1	17.976
9	88.233	2	14.161
10	38.695	2	8.825
11	73.831	7	14.184
12	77.122	3	19.604
13	45.519	2	13.728
14	82.251	2	21.141
15	59.862	3	17.446
16	26.563	3	9.629
17	61.818	2	14.005
18	29.682	1	9.160
19	50.825	5	18.831
20	71.062	4	7.641
21	41.990	4	13.882
22	37.324	3	9.670
23	86.352	5	21.604
24	45.506	2	10.866
25	69.929	6	28.980
26	61.041	2	10.882
27	82.469	1	18.561
28	44.208	2	11.629
29	49.467	5	18.067
30	25.905	5	14.539
31	79.178	5	19.192
32	75.811	3	25.918

Tabela A.1. Dados de gastos com alimentação

Obs	Renda Total	Número de pessoas	Gastos com alimentação
33	82.718	6	28.833
34	48.311	4	15.869
35	42.494	5	14.910
36	40.573	4	9.550
37	44.872	6	23.066
38	27.167	7	14.751

A.2. Escore em teste de habilidade de leitura

Tabela A.2. Dados de habilidade de leitura

Obs	Escore do teste	Dislexia	QI
1	0.88386	-1	0.827
2	0.76524	-1	0.590
3	0.91508	-1	0.471
4	0.98376	-1	1.144
5	0.88386	-1	-0.676
6	0.70905	-1	-0.795
7	0.77148	-1	-0.281
8	0.99000	-1	-0.914
9	0.99000	-1	-0.043
10	0.99000	-1	0.907
11	0.99000	-1	0.511
12	0.99000	-1	1.223
13	0.99000	-1	0.590
14	0.99000	-1	1.856
15	0.99000	-1	-0.399
16	0.99000	-1	0.590
17	0.70281	-1	-0.043
18	0.99000	-1	1.738
19	0.66535	-1	0.471
20	0.99000	-1	1.619
21	0.95878	-1	1.144
22	0.99000	-1	-0.201
23	0.73402	-1	-0.281
24	0.64662	-1	0.590
25	0.99000	-1	1.777
26	0.57794	1	-0.083

Tabela A.2. Dados de habilidade de leitura

Obs	Escore do teste	Dislexia	QI
27	0.64038	1	-0.162
28	0.45932	1	-0.795
29	0.65286	1	-0.281
30	0.60916	1	-0.874
31	0.60916	1	0.313
32	0.54048	1	0.709
33	0.57170	1	1.223
34	0.70281	1	-1.230
35	0.56546	1	-0.162
36	0.53424	1	-0.993
37	0.57794	1	-1.191
38	0.69032	1	-1.745
39	0.54673	1	-1.745
40	0.68408	1	-0.439
41	0.59043	1	-1.666
42	0.62165	1	-1.507
43	0.67159	1	-0.518
44	0.66535	1	-1.270

A.3. Ansiedade e estresse

Tabela A.3. Dados de ansiedade e estresse

Obs	estresse	Ansiedade
1	0.01	0.01
2	0.29	0.17
3	0.17	0.01
4	0.41	0.05
5	0.21	0.09
6	0.45	0.41
7	0.21	0.05
8	0.01	0.01
9	0.25	0.13
10	0.45	0.01
11	0.21	0.05
12	0.53	0.17
13	0.13	0.01

Tabela A.3. Dados de ansiedade e estresse

Obs	estresse	Ansiedade
14	0.17	0.09
15	0.01	0.01
16	0.25	0.05
17	0.05	0.09
18	0.41	0.09
19	0.09	0.05
20	0.01	0.01
21	0.25	0.01
22	0.25	0.01
23	0.29	0.29
24	0.17	0.01
25	0.29	0.01
26	0.25	0.01
27	0.25	0.01
28	0.25	0.01
29	0.09	0.01
30	0.09	0.01
31	0.01	0.01
32	0.25	0.09
33	0.41	0.37
34	0.37	0.05
35	0.25	0.01
36	0.37	0.05
37	0.33	0.29
38	0.21	0.09
39	0.21	0.01
40	0.33	0.25
41	0.17	0.01
42	0.41	0.09
43	0.21	0.01
44	0.37	0.05
45	0.37	0.21
46	0.13	0.01
47	0.13	0.01
48	0.17	0.01
49	0.33	0.13
50	0.29	0.17
51	0.57	0.37
52	0.33	0.01

Tabela A.3. Dados de ansiedade e estresse

Obs	estresse	Ansiedade
53	0.01	0.01
54	0.21	0.09
55	0.81	0.57
56	0.05	0.01
57	0.25	0.01
58	0.17	0.13
59	0.25	0.05
60	0.17	0.01
61	0.17	0.01
62	0.33	0.01
63	0.25	0.01
64	0.41	0.09
65	0.41	0.13
66	0.21	0.01
67	0.25	0.01
68	0.29	0.09
69	0.25	0.09
70	0.41	0.37
71	0.09	0.01
72	0.13	0.05
73	0.01	0.01
74	0.09	0.01
75	0.29	0.13
76	0.13	0.01
77	0.85	0.57
78	0.01	0.01
79	0.01	0.01
80	0.29	0.09
81	0.09	0.01
82	0.01	0.01
83	0.01	0.01
84	0.05	0.01
85	0.13	0.01
86	0.01	0.01
87	0.05	0.05
88	0.37	0.01
89	0.65	0.01
90	0.13	0.01
91	0.29	0.13

Tabela A.3. Dados de ansiedade e estresse

Obs	estresse	Ansiedade
92	0.01	0.01
93	0.57	0.25
94	0.21	0.01
95	0.29	0.01
96	0.53	0.09
97	0.45	0.13
98	0.25	0.01
99	0.09	0.01
100	0.13	0.05
101	0.17	0.13
102	0.05	0.01
103	0.17	0.09
104	0.21	0.01
105	0.29	0.05
106	0.13	0.01
107	0.21	0.05
108	0.17	0.01
109	0.37	0.09
110	0.09	0.01
111	0.85	0.37
112	0.65	0.25
113	0.21	0.05
114	0.29	0.05
115	0.17	0.25
116	0.65	0.05
117	0.53	0.05
118	0.25	0.01
119	0.17	0.05
120	0.01	0.01
121	0.33	0.01
122	0.25	0.01
123	0.61	0.17
124	0.29	0.29
125	0.85	0.57
126	0.21	0.01
127	0.09	0.05
128	0.01	0.01
129	0.41	0.09
130	0.01	0.01

Tabela A.3. Dados de ansiedade e estresse

Obs	estresse	Ansiedade
131	0.29	0.09
132	0.65	0.49
133	0.49	0.45
134	0.17	0.01
135	0.01	0.01
136	0.41	0.01
137	0.37	0.05
138	0.21	0.01
139	0.49	0.17
140	0.05	0.01
141	0.09	0.13
142	0.09	0.01
143	0.37	0.21
144	0.41	0.13
145	0.37	0.01
146	0.05	0.01
147	0.57	0.17
148	0.09	0.01
149	0.13	0.01
150	0.17	0.21
151	0.69	0.13
152	0.85	0.69
153	0.29	0.25
154	0.33	0.01
155	0.09	0.01
156	0.45	0.09
157	0.45	0.13
158	0.21	0.01
159	0.41	0.05
160	0.21	0.01
161	0.05	0.01
162	0.37	0.29
163	0.53	0.25
164	0.65	0.49
165	0.17	0.01
166	0.09	0.01

A.4. Veredito

Tabela A.4. Dados de veredito

Obs	Veredito	Conflito	Confiança
1	-1.0	-1.0	0.500
2	-1.0	-1.0	0.698
3	-1.0	-1.0	0.797
4	-1.0	-1.0	0.698
5	-1.0	-1.0	0.797
6	-1.0	-1.0	0.896
7	-1.0	-1.0	0.500
8	-1.0	-1.0	0.599
9	-1.0	-1.0	0.748
10	-1.0	-1.0	0.599
11	-1.0	-1.0	0.797
12	-1.0	-1.0	0.698
13	-1.0	-1.0	0.599
14	-1.0	-1.0	0.946
15	-1.0	-1.0	0.500
16	-1.0	-1.0	0.797
17	-1.0	-1.0	0.500
18	-1.0	-1.0	0.599
19	-1.0	-1.0	0.748
20	-1.0	-1.0	0.599
21	-1.0	-1.0	0.995
22	-1.0	-1.0	0.896
23	-1.0	-1.0	0.995
24	-1.0	-1.0	0.599
25	-1.0	-1.0	0.797
26	-1.0	-1.0	0.896
27	1.0	-1.0	0.500
28	1.0	-1.0	0.946
29	1.0	-1.0	0.500
30	1.0	-1.0	0.797
31	1.0	-1.0	0.797
32	1.0	-1.0	0.599
33	1.0	-1.0	0.748
34	1.0	-1.0	0.698
35	1.0	-1.0	0.649
36	1.0	-1.0	0.599

Tabela A.4. Dados de veredito

Obs	Veredito	Conflito	Confiança
37	1.0	-1.0	0.995
38	1.0	-1.0	0.896
39	1.0	-1.0	0.005
40	1.0	-1.0	0.203
41	1.0	-1.0	0.896
42	1.0	-1.0	0.797
43	1.0	-1.0	0.748
44	1.0	-1.0	0.005
45	1.0	-1.0	0.599
46	1.0	-1.0	0.500
47	1.0	-1.0	0.946
48	1.0	-1.0	0.995
49	1.0	-1.0	0.995
50	1.0	-1.0	0.896
51	1.0	-1.0	0.302
52	-1.0	1.0	0.748
53	-1.0	1.0	0.748
54	-1.0	1.0	0.995
55	-1.0	1.0	0.896
56	-1.0	1.0	0.896
57	-1.0	1.0	0.599
58	-1.0	1.0	0.500
59	-1.0	1.0	0.599
60	-1.0	1.0	0.500
61	-1.0	1.0	0.599
62	-1.0	1.0	0.599
63	-1.0	1.0	0.748
64	-1.0	1.0	0.599
65	-1.0	1.0	0.946
66	-1.0	1.0	0.203
67	-1.0	1.0	0.599
68	-1.0	1.0	0.896
69	-1.0	1.0	0.896
70	-1.0	1.0	0.203
71	-1.0	1.0	0.896
72	-1.0	1.0	0.599
73	-1.0	1.0	0.797
74	-1.0	1.0	0.599
75	-1.0	1.0	0.698

Tabela A.4. Dados de veredito

Obs	Veredito	Conflito	Confiança
76	-1.0	1.0	0.797
77	-1.0	1.0	0.748
78	1.0	1.0	0.401
79	1.0	1.0	0.698
80	1.0	1.0	0.599
81	1.0	1.0	0.649
82	1.0	1.0	0.896
83	1.0	1.0	0.797
84	1.0	1.0	0.698
85	1.0	1.0	0.698
86	1.0	1.0	0.847
87	1.0	1.0	0.946
88	1.0	1.0	0.797
89	1.0	1.0	0.995
90	1.0	1.0	0.797
91	1.0	1.0	0.797
92	1.0	1.0	0.995
93	1.0	1.0	0.896
94	1.0	1.0	0.995
95	1.0	1.0	0.975
96	1.0	1.0	0.401
97	1.0	1.0	0.500
98	1.0	1.0	0.797
99	1.0	1.0	0.797
100	1.0	1.0	0.896
101	1.0	1.0	0.946
102	1.0	1.0	0.698
103	1.0	1.0	0.995
104	1.0	1.0	0.847

A.5. Cloro disponível

Tabela A.5. Dados de cloro disponível.

Obs	Semanas	Fração de cloro	Obs	Semanas	Fração de cloro
1	8	0.49	22	22	0.41
2	8	0.49	23	22	0.41
3	10	0.48	24	22	0.40
4	10	0.47	25	24	0.42
5	10	0.48	26	24	0.40
6	10	0.47	27	24	0.40
7	12	0.46	28	26	0.41
8	12	0.46	29	26	0.40
9	12	0.45	30	26	0.41
10	12	0.43	31	28	0.41
11	14	0.45	32	28	0.40
12	14	0.43	33	30	0.40
13	14	0.43	34	30	0.40
14	16	0.44	35	30	0.38
15	16	0.43	36	32	0.41
16	16	0.43	37	32	0.40
17	18	0.46	38	34	0.40
18	18	0.45	39	36	0.41
19	20	0.42	40	36	0.38
20	20	0.42	41	40	0.39
21	20	0.43	42	42	0.39

Referências

Atkinson, A.C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika*, 68, 13–20.

Atkinson, A.C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.

Barndorff–Nielsen O.E., Jorgensen B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis*, 39, 106–116.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83, 687–697.

Berkowitz, J., Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, 19, 1–48.

Bickel, J.P., Doksum, A.K. (2001). *Mathematical Statistics. Basic Ideas and Selected Topics*, vol. 1, 2nd ed. New Jersey: Prentice–Hall.

Bury, K. (1999). *Statistical Distributions in Engineering*. New York: Cambridge University Press.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

Cook, R.D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*, 48, 133–169.

Cook, R.D., Penã, D., Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics, Theory and Methods*, 17, 623–640.

Cook, R.D., Weisberg, S. (1982). *Residuals and influence in Regressions*. London: Chapman and Hall.

Cordeiro, G.M., Paula, G.A. (1992). Estimation, large–sample parametric tests and diagnostics for non–exponential family nonlinear models. *Communications in Statistics, Simulation and Computation*, 21, 149–172.

Cox C. (1996). Nonlinear quasi-likelihood models: applications to continuos proportions. *Computational Statistics and Data Analysis*, 21, 449–461.

Cox, D., Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, **30**, 248–275.

Cribari–Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis*, 45, 215–233.

Cuervo, E.C., Gamerman, D. (2004). Bayesian modelling of joint regressions for the mean and covariance matriz. *Biometrical Journal*, 46, 430–440.

Davidson, R., MacKinnon, J.G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Journal of Econometrics*. 109, 167–193.

Davidson, R., MacKinnon, J.G. (2002). Bootstrap J tests of nonnested linear regression models. *Journal of Econometrics*, 109, 167–93.

Davison, A.C., Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika*, 76, 211–221.

Davison, A.C., Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. New York: Cambridge University Press.

Doornik, J.A. (2001). *Ox: an Object-oriented matriz Programming Language*, 4th ed. London: Timberlake Consultants and Oxford: `http://www.doornik.com`.

Draper, N.R., Smith, H. (1981). *Applied Regression Analysis*. 2nd ed. New York: Wiley.

de Souza, F.A.M. (1999). *Influência Local e Análise de Resíduos em Modelos de Regressão Von Mises*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil.

de Souza, F.A.M., Paula, G.A. (2002). Deviance residuals for an angular response. *Australian and New Zealand Journal of Statistics*, 44, 345–356.

Dyke, G.V., Patterson, H.D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics*, 8, 1–12.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

Efron, B., Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Emerson, J.D., Hoaglin, D.C., Kempthorne,P.J. (1984). Leverage in least squares additive–plus–multiplicative fits for two–way tables. *Journal of the American Statistical Association*, 79, 329–335.

Farhrmeir, L., Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. New York: Springer.

Federal Communications Commission (1993). FCC 93–177, Report and Order and Further Notice of Proposed Rule Making, M.M. Docker 92–266 (3 May 1993), 6134.

Federal Communications Commission (1994). FCC 93–177, Second Order on Reconsideration, Fourth Report and Order and Fifth Notice of Proposed Rulemaking, M.M. Docket 92–266 (30 March 1994), 4277.

Ferrari, S.L.P., Cribari–Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799–815.

Galea, M., Paula, G.A., Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, 46, 71–79.

Galea, M., Paula, G.A., Uribe–Opazo, M. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers*, 44, 23–45.

Griffiths, W.E., Hill, R.C., Judge, G.G. (1993). *Learning and Practicing Econometrics*. New York: Wiley.

Hoaglin, D.C., Welsch, R.E. (1978). The hat matriz in regression and ANOVA. *The American Statistician*, 32, 17–22.

Horowitz, J.L. (1994). Bootstrap-based critical values for the information matriz test. *Journal of Econometrics*,61, 395–411.

Houaiss (2001). *Dicionário Eletrônico Houaiss 1.0 (BR)*. Disponível em `www.dicionariohouaiss.com.br/`.

Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*, vol. 2, 2nd ed. New York: Wiley.

Kieschnick R., McCullough B.D. (2003). Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Statistical Modelling*, 3, 193–213.

Lehmann, E.L., Casella, E. (1998). *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag.

Lesaffre, E., Verbeke,G. (1998). Local influence in linear mixed models. *Biometrics*, 54, 570–582.

Liu, S.Z. (2000). On local influence for elliptical linear models. *Statistical Papers*, 41, 211–224.

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.

McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

McQuarrie, A., Tsai, C. (1998). *Regression and Time Series Model Selection*. New Jersey: World Scientific Publishing Company.

Monfardini, C. (2003). An illustration of Cox's non–nested testing procedure for logit and probit models. *Computational Statistics and Data Analysis*. 42, 425–444.

Mojirsheibani, M., Tibshirani, R. (1996). Some results on bootstrap prediction intervals. *The Canadian Journal of Statistics*. 24, 549-568.

Oliveira, M. S. (2004). *Um Modelo de Regressão Beta: Teoria e Aplicações*. Dissertação de mestrado, Departamento de Estatística, Universidade de São Paulo, Brasil.

Ospina, R., Cribari–Neto, F., Vasconcellos, K.L.P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics and Data Analysis*, a aparecer.

Papke L., Wooldridge J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11, 619–632.

Paula, G.A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation*, 51, 315–352.

Paula, G.A. (1996). Influence diagnostics in proper dispersion models. *Australian Journal of Statistics*, 38, 307–316.

Paolino, P. (2001). Maximum likelihood estimation of models with beta–distributed dependent variables. *Political Analysis*, 9, 325–346.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, 9, 705–724.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley.

Schawarzmann, B. (1991). A connection between local–influence analysis and residual diagnostics.*Technometrics*, 33, 103–104.

Smithson, M., Verkuilen, J. (2006). A better lemon–squeezer? Maximum likelihood regression with beta–distribuited dependent variables. *Psychological Methods*, 11, 54–71.

Smyth, G.K., Verbyla, A.P. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics*, 10, 695–709.

St. Laurent, R.T., Cook, R.D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*. 87, 985–990.

Thomas, W., Cook, R.D. (1989). Assessing influence on regression coefficients in generalized linear models. *Biometrika*, 76, 741–749.

Thomas, W., Cook, R.D. (1990). Assessing influence on predictions from generalized linear models. *Biometrika*, 76, 741–749.

Tsai, C.H., Wu, X. (1992). Assessing local influence in linear regression models with first–order autoregressive or heteroscedastic error struture. *Statistics and Probability Letters*, 14, 247–252.

Williams, D.A. (1984). Residuals in generalized linear models. *In: Proceedings of the 12th International Conference*, Tokyo, p.59–68.

Williams, D.A. (1987). Generalized linear models diagnostic using the deviance and single case deletion. *Applied Statistics*, 36, 181–191.

Wei, B.-C., Hu, Y.-Q. and Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25, 25–37.

Wei, B.-C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer.

Welsch, R.E., Kuh, E. (1977). Linear regression diagnostics. Technical Report 923–77. Sloan School of Management, Massachusetts Institute of Technology.

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261–1295.

Yoshizoe, Y. (1991). Leverage points in nonlinear regression models. *Journal of Japan Statistics Society*, 21, 1–11. 25, 25–37.