



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

**UM NOVO RESÍDUO PARA CLASSES DE
MODELOS DE REGRESSÃO NA
FAMÍLIA EXPONENCIAL.**

LELIO ALEJANDRO ARIAS VIZCAINO

DISSERTAÇÃO DE MESTRADO

Recife
2016

LELIO ALEJANDRO ARIAS VIZCAINO

UM NOVO RESÍDUO PARA CLASSES DE
MODELOS DE REGRESSÃO NA
FAMÍLIA EXPONENCIAL.

ORIENTADOR: PROF. DR RAYDONAL OSPINA MARTÍNEZ

COORIENTADOR:PROFA. DRA PATRÍCIA LEONE ESPINHEIRA OSPINA

Área de Concentração: Estatística Aplicada

Dissertação apresentada ao programa de Pós-graduação
em Estatística do Departamento de Estatística da Univer-
sidade Federal de Pernambuco, como requisito parcial para
obtenção do grau de **Mestre em Estatística**.

Recife

2016

A696n Arias Vizcaino, Lelio Alejandro.
Um novo resíduo para classes de modelos de regressão na família exponencial / Lelio Alejandro Arias Vizcaino. – 2016.
73f.: fig., tab.

Orientador: Raydonal Ospina Martínez
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN. Estatística, Recife, 2016.
Inclui referências e apêndice.

1. Modelos de regressão. 2. Modelos lineares generalizados. 3. Análises de diagnóstico. I. Ospina Martínez, Raydonal. (Orientador). II. Título.

310 CDD (22. ed.)

UFPE-MEI 2016-162

LÉLIO ALEJANDRO ARIAS VIZCAINO

**UM NOVO RESÍDUO PARA CLASSES DE MODELOS DE REGRESSÃO NA
FAMÍLIA EXPONENCIAL**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 05 de dezembro de 2016.

BANCA EXAMINADORA

Prof.^a Patrícia Leone Espinheira Ospina
UFPE

Prof. Francisco Cribari Neto
UFPE

Prof.^a Tarciana Liberal Pereira
UFPB

À minha mãe, Idda Patricia, por ser a
melhor professora que já tive em minha
vida.

GRACIAS TOTALES.

Agradecimentos

À minha mãe, por seu gigantesco esforço, dedicação e apoio investido em mim nestes 29 anos.

A meu irmão Diego Espinel, pelo apoio, pelas oportunidades brindadas, por ser mais um filho para minha mãe.

Ao professor Raydonal, pela confiança depositada em mim para desenvolver esta dissertação, por sua amizade e pelo seu compromisso de me tornar um melhor aluno.

À professora Patricia por compartilhar a ideia central desta dissertação, por seus conselhos e as, muito necessárias, correções.

A todos os professores do Programa de Pós Graduação em Estatística da UFPE, por todo conhecimento transmitido.

À FACEPE, pelo apoio financeiro.

Aos membros da banca examinadora, pelas sugestões.

À minha melhor amiga Olivia, por compartilhar mais uma aventura acadêmica.

A meus colegas, Yuri e Vinícius por todas as conversas na varanda, e por todas aquelas que faltam.

À meus colegas Jucelino e Rodney, por toda sua ajuda. Meus caros, seu sucesso esta garantido.

Aos colegas da pós-graduação Sásbatien, Telma, Ramon, Jodavid, Walmir, Elisandra, Camila, Fernando, Renata, Cristina, por todos os momentos compartilhados.

Ao Brasil, por me acolher.

“Somewhere, something incredible is
waiting to be known.”

Carl Sagan.

Resumo

Dentre as principais metodologias estatísticas, a análise de regressão é uma das formas mais efetivas para modelar dados. Neste sentido, a análise de diagnóstico é imprescindível para determinar o que poder ter acontecido no processo gerador dos dados caso os pressupostos impostos a este não sejam plausíveis. Uma das ferramentas mais úteis em diagnóstico é a avaliação dos resíduos. Neste trabalho, propomos um novo resíduo para as classes de modelos de regressão linear e não linear baseados na família exponencial com dispersão variável (Smyth (1989)). A proposta permite incorporar de forma simultânea informações relativas aos submodelos da média e da dispersão sem fazer uso de matrizes de projeção para sua padronização. Resultados de simulação e de aplicações a dados reais mostram que o novo resíduo é altamente competitivo em relação ao resíduos amplamente usados e consolidados na literatura.

Palavras-chave: Análise de diagnóstico. Modelo de regressão da família exponencial. Modelagem conjunta. Matriz de projeção. Resíduos

Abstract

In statistical methodologies, regression analysis can be a very effective way to model data. In this sense, the diagnostic analysis is needed to try to determine what might happened in the data generating process if the conditions imposed to it are not true. One of the most useful techniques to detect the goodness of fit to the model is the evaluation of residuals. In this work, we propose a new residual to the class of linear and nonlinear regression models based on exponential family with variable dispersion (Smyth (1989)). The proposal incorporates simultaneously information from the sub-models of the mean and the dispersion without using projection matrices for its standardization. Simulation results and applications in real data show that the new residual is highly competitive with respect to residuals widely used and established in the literature.

Keywords: Goodness of fit. The exponential family regression model. Joint modelling. Projection matrix. Residual.

Lista de Figuras

4.1	Gráficos normais de probabilidade com envelopes simulados para Modelo 4 e Modelo 5.	47
4.2	Gráficos normais de probabilidade com envelopes simulados para Modelo 1, Modelo 2 e Modelo 3.	48
4.3	Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 1. .	51
4.4	Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 1.	52
4.5	Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 2. .	53
4.6	Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 2.	54
4.7	Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 3. .	55
4.8	Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 3.	56
4.9	Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 4. .	57
4.10	Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 4.	58
4.11	Gráficos normais de probabilidade com envelopes simulados para Modelo 1, Modelo 2 e Modelo 3.	60
4.12	Gráficos normais de probabilidade com envelopes simulados para Modelo 4 e Modelo 5.	61

-
- 5.1 Gráficos normais de probabilidade referentes à modelagem conjunta da média e da dispersão supondo resposta gama ajustada aos dados sobre *snack* com: (a) resíduo componente desvio para a média e (b) resíduo componente desvio para a dispersão. 64
- 5.2 Gráfico de probabilidade do resíduo combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta gama ajustada aos dados sobre *snack*. 64
- 5.3 Gráfico de probabilidade do residual combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta binomial negativa ajustada aos dados sobre *maças*. 66
- 5.4 Gráfico de probabilidade do residual combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta binomial negativa ajustada aos dados sobre *maças*. 67

Lista de Tabelas

2.1	Exemplos de distribuições pertencentes à família exponencial	19
2.2	Algumas quantidades para distribuições da família exponencial.	21
4.1	Médias, erros-padrão, assimetria e curtose empíricas dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim N(\mu_i; \phi_1^{-1})$ com $\mu_i = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = -1.4$, $\beta_2 = 0.8$, $\gamma_1 = 0.9$, $\gamma_2 = -1$, $x_i \sim U(0.1, 1.1)$, $z_i \sim U(0.4, 1.4)$, $i = 1, \dots, 30$	40
4.2	Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{gama}(\mu_i; \phi_1)$ com $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1.5$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 30$	42
4.3	Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{NI}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = -0.45$, $\beta_2 = 0.9$, $\gamma_1 = -0.2$, $\gamma_2 = -0.4$, $x_i \sim U(0.1, 1.1)$, $z_i \sim U(0.4, 1.4)$, $i = 1, \dots, 30$	43
4.4	Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{BN}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1.6$, $\beta_2 = 0.8$, $\gamma_1 = -0.3$, $\gamma_2 = 0.5$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 30$. . .	44
4.5	Médias, erros-padrão, assimetria e curtoses empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{PD}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = -0.4$, $\gamma_2 = -0.2$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 30$. .	45
5.1	Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta gama ajustada aos dados sobre <i>snack</i>	63

-
- 5.2 Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta binomial negativa ajustada aos dados sobre *maçãs*. 66
- 5.3 Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta binomial negativa ajustada aos dados sobre *maçãs*. 67

Sumário

1	Introdução	16
1.1	Apresentação dos capítulos	17
1.2	Suporte computacional	17
2	Resíduo combinado padronizado no modelo linear generalizado com dispersão variável	18
2.1	Modelagem simultânea da média e da dispersão na família exponencial . .	18
2.2	Resíduos ordinários	22
2.3	Resíduo tipo desvio	24
2.4	Resíduo combinado padronizado	25
3	Resíduo combinado padronizado para classes de modelos não lineares da família exponencial	26
3.1	Modelos não lineares da família exponencial	26
3.2	Modelos não lineares para dados de contagem	28
3.2.1	Modelo binomial negativo não Linear	28
3.2.2	Modelo Poisson duplo não linear	31
3.3	Resíduos ordinários	33
3.4	Resíduo combinado padronizado	34
3.4.1	Resíduo combinado padronizado generalizado	35
4	Avaliação numérica	39
4.1	Avaliação dos resíduos através de simulações de Monte Carlo	39
4.2	Envelopes simulados	45
4.3	Modelos sob especificação incorreta e novos limites para detecção de pontos aberrantes	47

<i>Sumário</i>	15
4.4 Observações aberrantes	59
5 Aplicação	62
5.1 Aplicação 1	62
5.2 Aplicação 2	65
6 Considerações finais	68
Referências	70
A Métodos iterativos scoring de Fisher e resíduos ordinários	71
B Forma canônica da família exponencial	73

Introdução

Os modelos de regressão têm o objetivo de avaliar o comportamento de uma variável de interesse (também chamada variável resposta) em função de algumas outras variáveis denominadas de covariáveis, sendo os modelos lineares com erros normais amplamente usados para descrever a maioria dos fenômenos aleatórios. Porém, quando o fenômeno sob estudo tem uma variável resposta para a qual não é apropriado supor normalidade, os modelos lineares generalizados são, sob algumas condições, uma alternativa muito útil para ajustar dados. A classe dos modelos lineares generalizados introduzida por Nelder e Wedderburn (1972) tem como objetivo utilizar um conjunto mais amplo de distribuições para a variável resposta univariada, tais distribuições pertencem à família exponencial.

O processo de diagnóstico tem como finalidade avaliar a adequacidade do modelo proposto aos dados. Neste sentido, a análise de resíduos é fundamental, por exemplo com base nos gráficos de resíduos é possível a detecção de pontos aberrantes, especificação inadequada da função de ligação, presença de dispersão variável (que poderiam mudar significativamente as estimativas dos parâmetros), e o próprio processo aleatório de geração dos dados (distribuição de probabilidades). Na literatura, tem-se uma ampla variedade de resíduos, sendo os mais conhecidos o resíduo studentizado e o resíduo tipo desvio (deviance). Estes resíduos são padronizados através do uso da matriz de projeção presente no processo iterativo de maximização da verosimilhança dos dados, pelo uso do método gradiente. Para maiores detalhes acerca desses resíduos podem ser encontrados em McCullagh e Nelder (1989) e Paula (2013).

Esta padronização requer o cálculo de matrizes de projeção $n \times n$, o que computacionalmente pode ser muito intenso quando modela-se dados com um número consideravelmente grande de observações. O objetivo desta dissertação é propor um resíduo alternativo utilizando um método de padronização que não depende da matriz de projeção, avaliá-lo e compará-lo com outros resíduos difundidos na literatura para classes de modelos lineares e não lineares da família exponencial com dispersão variável introduzidos por Smyth (1989), não lineares com resposta binomial negativa introduzidos por Svetliza (2002) e modelos não lineares da família Poisson dupla introduzida por Efron (1986).

Além de, não depender da matriz de projeção, o resíduo proposto leva em consideração as modelagens feitas sob a média e a dispersão da variável resposta de forma simultânea.

1.1 Apresentação dos capítulos

No Capítulo 2 são apresentados os modelos lineares generalizados com dispersão variável, o método de estimação Scoring de Fisher e a construção de alguns resíduos com base nestes métodos de estimação dos parâmetros. No Capítulo 3 são apresentados os modelos não lineares generalizados com dispersão variável, modelos não lineares binomial negativo com dispersão variável e Poisson duplo, apresentado o método de estimação Scoring de Fisher, a construção de resíduos com base neste método de estimação e por último, uma generalização que chamamos de resíduo combinado padronizado. No Capítulo 4 apresentamos resultados de simulações de Monte Carlo com o objetivo de determinar de forma empírica a distribuição dos resíduos ordinários, resíduo tipo desvio e combinado para modelos não lineares da família exponencial. Também apresentamos resultados de simulações com modelos sob especificação incorreta e simulações em que a variável resposta foi perturbada induzindo a presença de observações aberrantes. No Capítulo 5 apresentamos aplicações a dados reais, e por último, no Capítulo 6 discutiremos sob as conclusões deste trabalho.

1.2 Suporte computacional

Todas as simulações apresentadas nesta dissertação foram implementadas na linguagem de programação R em sua versão 3.3.1. R é um software usado para o desenvolvimento de cálculos e gráficos estatísticos. O R encontra-se disponível gratuitamente em <http://www.R-project.org/>.

Resíduo combinado padronizado no modelo linear generalizado com dispersão variável

Uma das classes de distribuições comumente usadas na estatística é a família exponencial. Por sua flexibilidade é possível incorporar a modelagem de dados assimétricos, discretos, contínuos na reta real ou restritos a um intervalo. Na presença de covariáveis, Nelder e Wedderburn (1972) introduzem a classe de modelos lineares generalizados em que a resposta pertence à família exponencial. Smyth (1989) definiu uma classe mais ampla, ao considerar uma relação funcional da dispersão com a variável resposta e não só a típica modelagem da média.

2.1 Modelagem simultânea da média e da dispersão na família exponencial

Sejam $y_1, y_2, \dots, y_n$ variáveis aleatórias independentes cada uma com densidade dada por

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

em que $b(\cdot), c(\cdot; \cdot)$ são funções conhecidas, θ_i e y_i pertencem a $\mathbb{R}$, e $\phi_i^{-1} > 0$ é chamado de parâmetro de dispersão. A média e a variância de y_i são:

$$E(y_i) = \mu_i = b'(\theta_i) \quad \text{e} \quad \text{Var}(y_i) = \phi_i^{-1} b''(\theta_i).$$

Além disso, $b''(\theta_i) = \partial \mu_i / \partial \theta_i$ é uma função que só depende de μ_i e é representada por $V(\mu_i)$, denominada função de variância. A tabela 2.1 apresenta um resumo das distribuições que pertencem à família exponencial, em que CV é o coeficiente de variação definido como $CV = \sqrt{\text{Var}(y)}/E(y)$ e $\Gamma(\cdot)$ é a função gama.

Para obter uma estrutura de regressão tem-se que μ_i e ϕ_i satisfazem as relações fun-

Tabela 2.1: Exemplos de distribuições pertencentes à família exponencial .

Distribuição	θ	$b(\theta)$	ϕ	$V(\mu)$	$c(y; \phi)$
Normal	μ	$\theta^2/2$	σ^{-2}	1	$[\log(\phi/2\pi) - \phi y^2]/2$
Gama	$-1/\mu$	$-\log(-\theta)$	$1/(CV)^2$	μ^2	$(\phi - 1)\log y + \phi \log \phi - \log \Gamma(\phi)$
Normal Inversa	$-1/2\mu^2$	$-\sqrt{-2\theta}$	ϕ	μ^3	$[\log\{\phi/(2\pi y^3)\} - \phi/y]/2$
Poisson	$\log \mu$	e^θ	1	μ	$-\log y!$
Binomial	$\log(\mu/(1 - \mu))$	$\log(1 + e^\theta)$	n	$\mu(1 - \mu)$	$\log \binom{\phi}{\phi y/n}$

cionais dadas por:

$$g(\mu_i) = \eta_{1i} = x_i^\top \beta = \sum_{j=1}^k x_{ij} \beta_j, \quad (2.2)$$

$$h(\phi_i) = \eta_{2i} = z_i^\top \gamma = \sum_{j=1}^q z_{ij} \gamma_j, \quad (2.3)$$

para $i = 1, \dots, n$, com $\beta \in \mathbb{R}^k$ e $\gamma \in \mathbb{R}^q$ vetores de parâmetros desconhecidos, $x_{t1}, \dots, x_{tk_1}$ e $z_{t1}, \dots, z_{tq_1}$ são observações de k_1 e q_1 covariáveis com $k_1 \leq k < n$ e $q_1 \leq q < n$, respectivamente; as funções $g(\cdot)$ e $h(\cdot)$ são chamadas de funções de ligação, obedecendo a monotonicidade crescente e a dupla diferenciabilidade. Por último, temos que η_{1i} e η_{2i} são os preditores lineares. Seja $\boldsymbol{\theta} = (\beta^\top, \gamma^\top)^\top$, o vetor de parâmetros conjunto do modelo. O logaritmo da função de verosimilhança é dado por:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\mu_i; \phi_i), \quad (2.4)$$

em que

$$\ell_i(\mu_i; \phi_i) = \phi_i [y_i \theta_i - b(\theta_i)] + c(y_i; \phi_i).$$

A função escore é dada por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = U_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (U_{\beta}^\top, U_{\gamma}^\top)^\top,$$

com $U_{\beta} = \partial \ell(\boldsymbol{\theta}) / \partial \beta$ e $U_{\gamma} = \partial \ell(\boldsymbol{\theta}) / \partial \gamma$. A estimação dos parâmetros do modelo baseado em (2.1) e (2.2) é feita através do método da máxima verosimilhança. O j -ésimo componente da função escore para β é:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad j = 1, \dots, k, \quad (2.5)$$

com

$$\frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \mu_i} = \phi_i \left[y_i \frac{\partial \theta_i}{\partial \mu_i} - \frac{\partial b(\theta_i)}{\partial \mu_i} \right] = \phi_i [y_i - b'(\theta_i)] \frac{\partial \theta_i}{\partial \mu_i}. \quad (2.6)$$

Note que a Equação (2.5) reduz-se a

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \phi_i x_{ij} \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i), \quad (2.7)$$

em que $V_i = \partial \mu_i / \partial \theta_i$, $x_{ij} = \partial \eta_{1i} / \partial \beta_j$ e $\omega_i = (\partial \mu_i / \partial \eta_{1i})^2 / V_i$. Pode-se escrever a função escore para β na forma matricial como

$$U_\beta = X^\top \Phi W^{1/2} V^{-1/2} (y - \mu),$$

em que X é a matriz de covariadas de ordem $n \times k$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, $V = \text{diag}\{V_1, \dots, V_n\}$, $y = (y_1, \dots, y_n)^\top$ e $\mu = (\mu_1, \dots, \mu_n)^\top$.

Note que, se o parâmetro da precisão ϕ_i é conhecido e constante a Equação (2.1) define a classe de modelos da família exponencial com parâmetro canônico θ_i . No entanto, se ϕ_i é desconhecido, a Equação (2.1) define a classe de modelos da família exponencial só quando $c(y_i; \phi_i)$ é da forma

$$c(y_i; \phi_i) = d(\phi_i) + \phi_i a(y) + u(y), \quad (2.8)$$

com $d(\cdot)$, $a(\cdot)$ e $u(\cdot)$ funções quaisquer (para maior detalhe ver Smyth (1989)).

De forma similar aos passos para obter U_β e supondo que a função $c(y_i; \phi_i)$ é como na Equação (2.8), tem-se que o j -ésima componente da função escore para γ é dado por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j = 1, \dots, q, \quad (2.9)$$

com

$$\frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \phi_i} = t_i + d'(\phi_i), \quad (2.10)$$

em que $t_i = y_i \theta_i - b(\theta_i) + a(y_i)$. Logo a Equação (2.9) reduz-se a

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^n (t_i + d'(\phi_i)) \frac{z_{ij}}{h'(\phi_i)}, \quad (2.11)$$

sendo que $z_{ij} = \partial \eta_{2i} / \partial \gamma_j$ e $h'(\phi_i) = \partial \eta_{2i} / \partial \phi_i$. Perceba que, se θ_i é fixado, a Equação (2.4) pode ser reescrita como

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \{\phi_i t_i + d(\phi_i) + u(y_i)\}. \quad (2.12)$$

Desta forma, a Equação (2.12) é o logaritmo da função de verosimilhança de um modelo na família exponencial em que as variáveis aleatórias $t_1, \dots, t_n$ são independentes, com parâmetros canônicos $\phi_1, \dots, \phi_n$ e dispersão constante igual a 1 (mais detalhes em Paula

(2013)). Então, pelas condições usuais de regularidade da família exponencial, tem-se

$$E(t_i) = \mu_i^* = -d'(\phi_i) \quad \text{e} \quad \text{Var}(t_i) = -d''(\phi_i).$$

Essas quantidades são descritas na tabela 2.2 para as distribuições normal, normal inversa (N. Inversa) e gama, sendo $\psi(\cdot)$ a função digama e $\psi'(\cdot)$ a função trigama.

Tabela 2.2: Algumas quantidades para distribuições da família exponencial.

	Normal	N. Inversa	Gama
t_i	$y_i \mu_i - (y_i^2 + \mu_i^2)/2$	$-\{y_i/(2\mu_i^2) + \mu_i^{-1} + (2y_i)^{-1}\}$	$\log(y_i/\mu_i) - y_i/\mu_i$
$d(\phi_i)$	$\frac{1}{2}\log(\phi_i)$	$\frac{1}{2}\log(\phi_i)$	$\phi_i \log(\phi_i) - \phi_i \Gamma(\phi_i)$
$d'(\phi_i)$	$(2\phi_i)^{-1}$	$(2\phi_i)^{-1}$	$(1 + \log(\phi_i)) - \psi(\phi_i)$
$d''(\phi_i)$	$-(2\phi_i^2)^{-1}$	$-(2\phi_i^2)^{-1}$	$\phi_i^{-1} - \psi'(\phi_i)$

E forma matricial a função escore para γ é

$$U_\gamma = Z^\top H_\gamma^{-1} (t - \mu^*),$$

em que Z é a matriz de covariadas de ordem $n \times q$, $H_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$, $t = (t_1, \dots, t_n)^\top$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$. Pelas propriedades da família exponencial (ver Lehmann e Casella (1998)), tem-se que θ_i e ϕ_i são ortogonais, logo, segue-se diretamente a ortogonalidade de β e γ . Assim, a matriz de informação de Fisher para θ é bloco diagonal dada por $K_{\theta\theta} = \text{diag}\{K_{\beta\beta}, K_{\gamma\gamma}\}$, em que $K_{\beta\beta}$ e $K_{\gamma\gamma}$ são as matrizes de informação de Fisher para β e γ respectivamente. Para obtermos ditas matrizes é preciso obter as segundas derivadas da função de verosimilhança com relação ao parâmetro θ . Desta forma, para o componente β de θ temos que

$$\frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \beta_\ell} = \sum_{i=1}^n \phi_i \frac{\partial}{\partial \beta_\ell} \left(\sqrt{\frac{\omega_i}{V_i}} \right) (y_i - \mu_i) x_{ij} - \sum_{i=1}^n \phi_i \sqrt{\frac{\omega_i}{V_i}} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 x_{ij} x_{i\ell}, \quad j, \ell = 1, \dots, k,$$

cujos valores esperados são

$$E \left(\frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \beta_\ell} \right) = - \sum_{i=1}^n \phi_i \omega_i x_{ij} x_{i\ell}.$$

Assim, pode-se escrever a informação de Fisher para β em forma matricial como

$$K_{\beta\beta} = E \left(- \frac{\partial^2 \ell(\theta)}{\partial \beta \partial \beta^\top} \right) = X^\top \Phi W X. \quad (2.13)$$

Da mesma forma, temos para o componente γ de $\boldsymbol{\theta}$ que

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} = \sum_{i=1}^n (t_i + d'(\phi_i)) h''(\phi_i) \frac{z_{ij} z_{i\ell}}{(h'(\phi_i))^3} - \sum_{i=1}^n d''(\phi_i) \frac{z_{ij} z_{i\ell}}{(h'(\phi_i))^2}, \quad j, \ell = 1, \dots, q,$$

e aplicando esperança tem-se

$$E \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} \right) = - \sum_{i=1}^n \frac{d''(\phi_i)}{(h'(\phi_i))^2} z_{ij} z_{i\ell}.$$

Em forma matricial, pode-se escrever a matriz de informação de Fisher para γ como

$$K_{\gamma\gamma} = E \left(- \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \gamma^\top} \right) = Z^\top P Z, \quad (2.14)$$

em que $P = V_\gamma H_\gamma^{-2}$ e $V_\gamma = \text{diag}\{-d''(\phi_1), \dots, -d''(\phi_n)\}$.

2.2 Resíduos ordinários

A estimação dos parâmetros nos modelos lineares generalizados geralmente é feita pelo método de máxima verossimilhança, em que $\hat{\beta}$ e $\hat{\gamma}$ são os estimadores de máxima verossimilhança, obtidos como soluções do sistema

$$\begin{cases} U_\beta &= 0, \\ U_\gamma &= 0. \end{cases}$$

Porém, em muitas casos, não é possível obter uma solução em forma fechada para o sistema, logo o sistema é resolvido empregando algum método numérico de otimização. Utilizando o método iterativo scoring de Fisher, o qual tem forma genérica dada por

$$\beta^{(m+1)} = \beta^{(m)} + \{K_{\beta\beta}^{(m)}\}^{-1} U_\beta^{(m)}, \quad (2.15)$$

$$\gamma^{(m+1)} = \gamma^{(m)} + \{K_{\gamma\gamma}^{(m)}\}^{-1} U_\gamma^{(m)}. \quad (2.16)$$

Logo, fazendo as substituições respectivas dessas quantidades (ver Apêndice A), as expressões (2.15) e (2.16) podem ser escritas na forma de um processo iterativo de mínimos quadrados ponderados como

$$\begin{aligned} \beta^{(m+1)} &= (X^\top \Phi^{(m)} W^{(m)} X)^{-1} X^\top \Phi^{(m)} W^{(m)} u_1^{(m)}, \\ \gamma^{(m+1)} &= (Z^\top P^{(m)} Z)^{-1} Z^\top P^{(m)} u_2^{(m)}, \end{aligned} \quad (2.17)$$

em que $u_1^{(m)} = \eta_1^{(m)} + W^{1/2(m)} V^{-1/2(m)} (y - \mu^{(m)})$ e $u_2^{(m)} = \eta_2^{(m)} + V_\gamma^{-1(m)} H_\gamma^{(m)} (t - \mu^{*(m)})$.

Após a convergência do processo iterativo em (2.17), tem-se que

$$\hat{\beta} = \left(X^\top \hat{\Phi} \hat{W} X \right)^{-1} X^\top \hat{\Phi} \hat{W} u_1, \quad (2.18)$$

$$\hat{\gamma} = \left(Z^\top \hat{P} Z \right)^{-1} Z^\top \hat{P} u_2, \quad (2.19)$$

com $u_1 = \hat{\eta}_1 + \hat{W}^{1/2} \hat{V}^{-1/2} (y - \hat{\mu})$ e $u_2 = \hat{\eta}_2 + \hat{V}_\gamma^{-1} \hat{H}_\gamma (t - \hat{\mu}^*)$.

As equações (2.18) e (2.19) podem ser vistas como estimadores de mínimos quadrados ordinários; logo para obter $\hat{\beta}$ considera-se a regressão da variável auxiliar $\hat{\Phi}^{1/2} \hat{W}^{1/2} u_1$ sob a covariável $\hat{\Phi}^{1/2} \hat{W}^{1/2} X$, e para obter $\hat{\gamma}$ considera-se a regressão da variável auxiliar $\hat{P}^{1/2} u_2$ sob a covariável $\hat{P}^{1/2} Z$. Consequentemente, os resíduos ordinários das regressões anteriores são

$$r^\beta = \hat{\Phi}^{1/2} \hat{W}^{1/2} (u_1 - \hat{\eta}_1),$$

$$r^\gamma = \hat{P}^{1/2} (u_2 - \hat{\eta}_2).$$

Por tanto, o resíduo para a i -ésima observação é dado por (mais detalhes ver Apêndice A):

$$r_i^\beta = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}_i^{-1} \hat{V}_i}}, \quad i = 1, \dots, n, \quad (2.20)$$

$$r_i^\gamma = \frac{t_i - \hat{\mu}_i^*}{\sqrt{-d''(\hat{\phi}_i)}} \quad i = 1, \dots, n. \quad (2.21)$$

Esses resíduos podem ser padronizados usando a matriz de projeção da solução de mínimos quadrados ordinários. Para cada componente temos

$$\hat{S}^\beta = \hat{\Phi}^{1/2} \hat{W}^{1/2} X \left(X^\top \hat{\Phi} \hat{W} X \right)^{-1} X^\top \hat{W}^{1/2} \hat{\Phi}^{1/2}, \quad (2.22)$$

$$\hat{S}^\gamma = \hat{P}^{1/2} Z \left(Z^\top \hat{P} Z \right)^{-1} Z^\top \hat{P}^{1/2}. \quad (2.23)$$

Desta forma, os resíduos ordinários padronizados podem ser determinados por

$$r_{pi}^\beta = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}_i^{-1} \hat{V}_i (1 - \hat{s}_{ii}^\beta)}}, \quad i = 1, \dots, n, \quad (2.24)$$

$$r_{pi}^\gamma = \frac{t_i - \hat{\mu}_i^*}{\sqrt{-d''(\hat{\phi}_i) (1 - \hat{s}_{ii}^\gamma)}}, \quad i = 1, \dots, n, \quad (2.25)$$

em que $\hat{s}_{ii}^\beta$ e $\hat{s}_{ii}^\gamma$ são elementos da diagonal principal de $\hat{S}^\beta$ e $\hat{S}^\gamma$ respectivamente. Logo, r_{pi}^β é conhecido como resíduo studentizado. Note que seu cálculo depende da obtenção da matriz de projeção $\hat{S}^\beta$. Aqui, r_{pi}^β será referido como resíduo ordinário padronizado.

2.3 Resíduo tipo desvio

Uma importante estatística discutida em problemas relacionados à família exponencial para avaliar a qualidade do ajuste do modelo é o deviance, traduzida como desvio pela primeira vez por Cordeiro (1986).

Define-se a função desvio para a média como a distância entre o logaritmo da função de verosimilhança do modelo saturado (com n parâmetros) e do modelo de interesse (com k parâmetros, sendo $k < n$) avaliado na estimativa de máxima verosimilhança $\hat{\beta}$, isto é

$$D^*(y, \hat{\mu}) = 2 \sum_{i=1}^n \phi_i \left\{ y_i \left[\tilde{\theta}_i - \hat{\theta}_i \right] - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\},$$

em que $\tilde{\theta}_i = \theta(y_i)$ e $\hat{\theta}_i = \theta(\hat{\mu}_i)$ representam, respectivamente, as estimativas de máxima verosimilhança de θ para o modelo saturado (n parâmetros) e para o modelo sob estudo ($k < n$ parâmetros).

Analogamente, pode-se definir a função desvio para a precisão como

$$D^*(y, \hat{\phi}_i, \hat{\mu}) = 2 \sum_{i=1}^n \left\{ t_i \left[\tilde{\phi}_i - \hat{\phi}_i \right] + \left[d(\tilde{\phi}_i) - d(\hat{\phi}_i) \right] \right\},$$

em que, $\tilde{\phi}_i$ é solução da Equação $d'(\tilde{\phi}_i) = -t_i$. Para maiores detalhes acerca da função desvio ver Paula (2013) e McCullagh e Nelder (1989).

Desta forma, um tipo de resíduo muito utilizado nos modelos lineares generalizados é o resíduo deviance (ou componente do desvio). Portanto, os resíduos deviance padronizados para a média e a dispersão, respectivamente, são dados por

$$d_{\mu}(\hat{\mu}_i, \hat{\phi}_i) = \frac{d_{\mu}^*(\hat{\mu}_i, \hat{\phi}_i)}{\sqrt{1 - \hat{s}_{ii}^{\beta}}}, \quad i = 1, \dots, n, \quad (2.26)$$

$$d_{\phi}(\hat{\mu}_i, \hat{\phi}_i) = \frac{d_{\phi}^*(\hat{\mu}_i, \hat{\phi}_i)}{\sqrt{1 - \hat{s}_{ii}^{\gamma}}}, \quad i = 1, \dots, n, \quad (2.27)$$

em que, $\hat{s}_{ii}^{\beta}$ e $\hat{s}_{ii}^{\gamma}$ são i -ésimos elementos diagonais das matrizes de projeção dadas em (2.22) e (2.23) respectivamente, e

$$d_{\mu}^*(\hat{\mu}_i, \hat{\phi}_i) = \text{sign}\{y_i - \hat{\mu}_i\} \left\{ 2y_i \left[\tilde{\theta}_i - \hat{\theta}_i \right] - 2 \left[b(\tilde{\theta}_i) - b(\hat{\theta}_i) \right] \right\}^{1/2},$$

$$d_{\phi}^*(\hat{\mu}_i, \hat{\phi}_i) = \text{sign}\{\hat{t}_i + d'(\hat{\phi}_i)\} \left\{ 2t_i \left[\tilde{\phi}_i - \hat{\phi}_i \right] + 2 \left[d(\tilde{\phi}_i) - d(\hat{\phi}_i) \right] \right\}^{1/2}.$$

2.4 Resíduo combinado padronizado

O resíduo combinado surge como uma solução para a seguinte pergunta: “Como podemos obter mais informação sob a qualidade do modelo ajustado aos dados levando em consideração as informações dos ajustes dos componentes da média (r_i^β) e da dispersão (r_i^γ) em um único resíduo?”. Note que, por sua construção, os resíduos dados em (2.24) e (2.25) contêm informação da modelagem da média e da dispersão de forma separada. A proposta desta dissertação é obter uma classe de resíduos que levem em consideração as informações do processo de estimação de todos os parâmetros envolvidos no ajuste de forma simultânea.

Podemos definir o resíduo combinado para a i -ésima observação como

$$r_i^{\beta\gamma} = (y_i - \hat{\mu}_i) + (t_i - \hat{\mu}_i^*). \quad (2.28)$$

Note que (2.28) é uma combinação linear dos numeradores dos resíduos dados em (2.24) e (2.25). Porém, como o resíduo combinado não está baseado em um único submodelo de regressão (o modelo para a média ou o modelo para a dispersão), é necessário especificar um método de padronização. A proposta que será apresentada neste trabalho consiste em não usar nenhuma das duas matrizes de projeção definidas nas equações (2.22) e (2.23), simplesmente obter de forma direta $\text{Var}(r_i^{\beta\gamma})$. Isto implica em um baixo custo computacional para seu cálculo, uma vez que não é necessária a obtenção de matrizes de projeção cuja dimensão é $n \times n$.

Seja $\zeta_i = \text{Var}(r_i^{\beta\gamma})$ com $i = 1, \dots, n$. Daí tem-se que

$$\begin{aligned} \zeta_i &= \text{Var}((y_i - \hat{\mu}_i) + (t_i - \hat{\mu}_i^*)) \\ &= \text{Var}(y_i + y_i\theta_i + a(y_i)) \\ &= (1 + \theta_i)^2 \text{Var}(y_i) + \text{Var}(a(y_i)) + 2(1 + \theta_i) \text{Cov}(y_i; a(y_i)). \end{aligned}$$

Pelas equações (B.2), (B.4) e (B.5) do Apêndice B temos que

$$\zeta_i = \phi_i^{-1} b''(\theta_i) - d''(\phi_i), \quad i = 1, \dots, n. \quad (2.29)$$

Assim, o resíduo combinado padronizado é dado por

$$r_{pi}^{\beta\gamma} = \frac{(y_i - \hat{\mu}_i) + (t_i - \hat{\mu}_i^*)}{\sqrt{\hat{\zeta}_i}}, \quad (2.30)$$

em que $\hat{\zeta}_i$ é dado pela Equação (2.29) avaliada em $\hat{\mu}_i$ e $\hat{\phi}_i$. Se o modelo tiver dispersão constante considera-se que $\phi_i = \dots = \phi_n = \phi$.

Resíduo combinado padronizado para classes de modelos não lineares da família exponencial

A ideia dos modelos não lineares da família exponencial é a de generalizar os modelos apresentados no Capítulo 2, ampliando o espectro funcional das componentes sistemáticas dadas em (2.2) e (2.3), considerando preditores não lineares.

3.1 Modelos não lineares da família exponencial

Sejam $y_1, y_2, \dots, y_n$ variáveis aleatórias independentes, em que cada y_i segue a distribuição dada na Equação (2.1), em que μ_i e $\phi_i > 0$ satisfazem as relações funcionais

$$g(\mu_i) = \eta_{1i} = f_1(x_i^\top; \beta), \quad (3.1)$$

$$h(\phi_i) = \eta_{2i} = f_2(z_i^\top; \gamma), \quad (3.2)$$

em que $\beta \in \mathbb{R}^k$ e $\gamma \in \mathbb{R}^q$ são vetores de parâmetros desconhecidos com $k + q < n$, η_{1i} e η_{2i} são os preditores não lineares, $x_i^\top = (x_{i1}, \dots, x_{ik_1})$ e $z_i^\top = (z_{i1}, \dots, z_{iq_1})$ são observações de k_1 e q_1 covariáveis fixas e conhecidas, com $i = 1, \dots, n$, $k_1 \leq k$ e $q_1 \leq q$; as funções $g(\cdot)$ e $h(\cdot)$ chamadas de ligações, obedecem monotonicidade crescente e dupla diferenciabilidade; por último, temos que f_j com $j \in \{1, 2\}$ são funções contínuas e diferenciáveis, de modo que as matrizes de derivadas $J_1 = \partial\eta_1/\partial\beta$ e $J_2 = \partial\eta_2/\partial\gamma$ têm posto completo k e q , respectivamente.

O logaritmo da função de verosimilhança é dado pela Equação (2.4). Assim, o j -ésimo componente da função escore para β é a seguinte

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^n \frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad j = 1, \dots, k, \\ &= \sum_{i=1}^n \phi_i (y_i - \mu_i) \sqrt{\frac{\omega_i}{V_i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \end{aligned} \quad (3.3)$$

em que $V_i = \partial\mu_i/\partial\theta_i$ e $\omega_i = (\partial\mu_i/\partial\eta_{1i})^2/V_i$. Pode-se escrever a função escore na forma matricial

$$U_\beta = J_1^\top \Phi W^{1/2} V^{-1/2} (y - \mu),$$

em que $J_1 = \partial\eta_1/\partial\beta$ é uma matriz de ordem $n \times k$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, $V = \text{diag}\{V_1, \dots, V_n\}$, $y = (y_1, \dots, y_n)^\top$ e $\mu = (\mu_1, \dots, \mu_n)^\top$.

Sob a suposição de que $c(y_i; \phi_i) = d(\phi_i) + \phi_i a(y) + u(y)$, tem-se que a função escore para γ é dada por

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} &= \sum_{i=1}^n \frac{\partial \ell_i(\mu_i; \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j = 1, \dots, q, \\ &= \sum_{i=1}^n (t_i + d'(\phi_i)) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \end{aligned} \quad (3.4)$$

em que $t_i = y_i \theta_i - b(\theta_i) + a(y_i)$. De forma semelhante ao Capítulo 2, a função escore para γ pode ser escrita como

$$U_\gamma = J_2^\top H_\gamma^{-1} (t - \mu^*),$$

em que $J_2 = \partial\eta_2/\partial\gamma$ é uma matriz de ordem $n \times q$, $H_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$, $h'(\phi_i) = \partial\eta_{2i}/\partial\phi_i$, $t = (t_1, \dots, t_n)^\top$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$.

Para obter a matriz de informação de Fisher, precisa-se das segundas derivadas da função de log-verosimilhança com relação aos parâmetros. Desta forma, para $j, \ell = 1, \dots, k$ tem-se que

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} = \sum_{i=1}^n \phi_i \frac{\partial}{\partial \beta_\ell} \left(\sqrt{\frac{\omega_i}{V_i}} \right) (y_i - \mu_i) \frac{\partial \eta_{1i}}{\partial \beta_j} - \sum_{i=1}^n \phi_i \sqrt{\frac{\omega_i}{V_i}} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_\ell},$$

para $j, \ell = 1, \dots, q$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} = \sum_{i=1}^n (t_i + d'(\phi_i)) \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial}{\partial \gamma_\ell} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) + \sum_{i=1}^n d''(\phi_i) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell},$$

e por último para $j = 1, \dots, k$ e $\ell = 1, \dots, q$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_\ell} = \sum_{i=1}^n (y_i - \mu_i) \sqrt{\frac{\omega_i}{V_i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell}.$$

Dado que $E(y_i - \mu_i) = 0$ e $E(t_i - \mu_{T_i}) = 0$ (sob condições de regularidade), as esperanças

destas derivadas são dadas por

$$E \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} \right) = - \sum_{i=1}^n \phi_i \omega_i \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_\ell}, \quad (3.5)$$

$$E \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} \right) = - \sum_{i=1}^n \frac{d''(\phi_i)}{(h'(\phi_i))^2} \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell}, \quad (3.6)$$

$$E \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_\ell} \right) = 0. \quad (3.7)$$

Então, das equações (3.5), (3.6) e (3.7) tem-se que a matriz de informação de Fisher para $\boldsymbol{\theta}$ é bloco diagonal dada por $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{diag}\{K_{\beta\beta}; K_{\gamma\gamma}\}$, em que $K_{\beta\beta}$ e $K_{\gamma\gamma}$ são as matrizes de informação de Fisher para β e γ respectivamente. Logo, pode-se escrever na forma matricial $K_{\beta\beta}$ e $K_{\gamma\gamma}$ da seguinte forma

$$K_{\beta\beta} = E \left(- \frac{\partial^2 \ell(\beta; \gamma)}{\partial \beta \partial \beta^\top} \right) = J_1^\top \Phi W J_1, \quad (3.8)$$

$$K_{\gamma\gamma} = E \left(- \frac{\partial^2 \ell(\beta; \gamma)}{\partial \gamma \partial \gamma^\top} \right) = J_2^\top P J_2, \quad (3.9)$$

em que, $P = V_\gamma H_\gamma^{-2}$ e $V_\gamma = \text{diag}\{-d''(\phi_1), \dots, -d''(\phi_n)\}$.

3.2 Modelos não lineares para dados de contagem

Para ajustar dados de contagem, temos como alternativa o modelo Poisson. Este modelo apresenta algumas dificuldades, como, por exemplo, quando os dados apresentam superdispersão (quando a variância é maior do que a média). Nestas situações o modelo Poisson não sempre é adequado, pois sua principal característica é que a média e a variância são iguais, logo pensar em uma modelagem conjunta da média e da dispersão é inviável. O mesmo acontece no modelo binomial. Neste caso o parâmetro de dispersão é n (ver tabela 2.1), mas, no contexto de regressão, n sempre é conhecido. Desta forma, é incoerente modelar a dispersão.

3.2.1 Modelo binomial negativo não Linear

Nos últimos anos a distribuição binomial negativa tem sido muito usada como uma alternativa em caso de superdispersão, uma vez que, ignorar este fenômeno, traz grandes consequências para os erros padrão obtidos através do modelo de regressão de Poisson, pois, ou estes estão errados ou subestimam a variabilidade das estimativas de seus parâmetros.

Seja y uma variável aleatória com distribuição binomial negativa com parâmetros $\mu > 0$

e $\phi > 0$, denotada $y \sim \text{BN}(\mu, \phi)$, com função de probabilidade dada por

$$f(y; \mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \mu} \right)^\phi \left(\frac{\mu}{\phi + \mu} \right)^y, \quad y = 0, 1, 2, \dots, \quad (3.10)$$

onde $\Gamma(\cdot)$ é a função gama e denotamos $y \sim \text{BN}(\mu, \phi)$. A média e a variância de y são dadas por

$$E(y) = \mu \quad \text{e} \quad \text{Var}(y) = \mu + \mu^2/\phi.$$

Note que, se o parâmetro ϕ for conhecido, a distribuição binomial negativa faz parte da família exponencial com $\theta = \log\{\mu/(\mu + \phi)\}$, $b(\theta) = -\phi \log\{\mu/(\mu + \phi)\}$, $c(y; \phi) = \Gamma(y + \mu)/\Gamma(y + 1)\Gamma(\phi)$ e função de variância $V(\mu) = \mu + \mu^2/\phi$.

Svetliza (2002) apresenta a classe de modelos não lineares com resposta binomial negativa. Suponha agora que $y_1, \dots, y_n$ são variáveis aleatórias independentes tais que $y_i \sim \text{BN}(\mu_i, \phi_i)$, ϕ_i é desconhecido. Suponha também uma estrutura de regressão na qual μ_i e ϕ_i satisfazem as relações funcionais dadas nas equações (3.1) e (3.2), respectivamente. Considerando $\boldsymbol{\theta} = (\beta^\top, \gamma^\top)^\top$ o vetor de parâmetros do modelo, o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} \ell(\boldsymbol{\theta}) = \sum_{i=1}^n \{ & y_i \log(\mu_i) + \phi_i \log(\phi_i) - (y_i + \phi_i) \log(\mu_i + \phi_i) + \log \Gamma(\phi_i + y_i) \\ & - \log \Gamma(\phi_i) - \log \Gamma(y_i + 1) \}. \end{aligned} \quad (3.11)$$

As derivadas de primeira e segunda ordem de $\ell(\boldsymbol{\theta})$ em relação aos parâmetros para $j, \ell = 1, \dots, k$ são dadas por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} - \frac{y_i + \phi_i}{\mu_i + \phi_i} \right\} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{V(\mu_i)} \left(\frac{\partial \eta_{1i}}{\partial \mu_i} \right)^{-1} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad (3.12)$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} = \sum_{i=1}^n \left\{ \frac{-V(\mu_i) \partial \mu_i / \partial \eta_{1i} - (y_i - \mu_i) \nu_i}{[V(\mu_i) \partial \mu_i / \partial \eta_{1i}]^2} \right\} \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_\ell}, \quad (3.13)$$

para $j = 1, \dots, k$ e $\ell = 1, \dots, q$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_\ell} = \sum_{i=1}^n \frac{y_i - \mu_i}{(\phi + \mu_i)^2} \left(\frac{\partial \eta_{1i}}{\partial \mu_i} \right)^{-1} \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell}, \quad (3.14)$$

e, por último, para $j, \ell = 1, \dots, q$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^n c_i \frac{\partial \eta_{2i}}{\partial \phi_i} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad (3.15)$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} = \sum_{i=1}^n c'_i \left(\frac{\partial \eta_{2i}}{\partial \phi_i} \right)^2 \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell} + \sum_{i=1}^n c_i \frac{\partial}{\partial \gamma_\ell} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad (3.16)$$

em que

$$\begin{aligned} V'(\mu_i) &= \mu_i + \mu_i^2 / \phi_i, \\ \nu_i &= V'(\mu_i) \partial \mu_i / \partial \eta_{1i} + V(\mu_i) \partial^2 \mu_i / \partial \eta_{1i}^2, \\ c_i &= \psi(y_i + \phi_i) - \psi(\phi_i) + \log(\phi_i / (\mu_i + \phi_i) - (y_i + \phi_i) / (\mu_i + \phi_i) + 1), \\ c'_i &= \psi'(y_i + \phi_i) - \psi'(\phi_i) + (y_i - \mu_i) / (\mu_i + \phi_i)^2 + \mu_i / \phi_i (\mu_i + \phi_i). \end{aligned}$$

Assim, temos que as funções escore para β e γ podem ser escritas em forma matricial, respectivamente, como

$$\begin{aligned} U_\beta &= J_1^\top W(y - \mu) \quad \text{e} \\ U_\gamma &= J_2^\top W_\gamma^{-1} c, \end{aligned}$$

com $J_1 = \partial \eta_1 / \partial \beta$, $J_2 = \partial \eta_2 / \partial \gamma$ matrizes de ordem $n \times k$ e $n \times q$ respectivamente, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, com $\omega_i = (\partial \mu_i / \partial \eta_{1i})^2 V(\mu_i)^{-1}$, $W_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$, com $h'(\phi_i) = \partial \eta_{2i} / \partial \phi_i$, $y = (y_1, \dots, y_n)^\top$, $\mu = (\mu_1, \dots, \mu_n)^\top$ e $c = (c_1, \dots, c_n)^\top$.

Segundo Lawless (1987), a matriz de informação de Fisher para $\boldsymbol{\theta} = (\beta^\top, \gamma^\top)^\top$ é bloco diagonal e é dada por $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{diag}\{K_{\beta\beta}, K_{\gamma\gamma}\}$ em que $K_{\beta\beta}$ e $K_{\gamma\gamma}$ são as matrizes de informação de Fisher para β e γ respectivamente. Assim, pode-se escrever $K_{\beta\beta}$ e $K_{\gamma\gamma}$ em forma matricial como

$$K_{\beta\beta} = E \left(- \frac{\partial^2 \ell(\beta; \gamma)}{\partial \beta \partial \beta^\top} \right) = J_1^\top W J_1, \quad (3.17)$$

$$K_{\gamma\gamma} = E \left(- \frac{\partial^2 \ell(\beta; \gamma)}{\partial \gamma \partial \gamma^\top} \right) = J_2^\top P J_2, \quad (3.18)$$

com $P = T W_\gamma^{-2}$ e $T = \text{diag} = \{E(c'_1), \dots, E(c'_n)\}$.

A função desvio para a média supondo ϕ_i conhecido, para $i = 1, \dots, n$, é dada por

$$D^*(y, \hat{\mu}) = 2 \sum_{i=1}^n \left\{ \phi_i \log \left[\frac{\hat{\mu}_i + \phi_i}{y_i + \phi_i} \right] + y_i \log \left[\frac{y_i (\hat{\mu}_i + \phi_i)}{\hat{\mu}_i (y_i + \phi_i)} \right] \right\}. \quad (3.19)$$

Na prática, ϕ_i é substituído pela estimativa de máxima verosimilhança $\hat{\phi}_i$. No caso em

que $y_i = 0$ para algum i , o i -ésimo componente do desvio é dado por

$$2\phi_i \log \left[\frac{y_i + \phi_i}{\widehat{\mu}_i + \phi_i} \right].$$

3.2.2 Modelo Poisson duplo não linear

A distribuição Poisson dupla foi introduzida por Efron (1986). Essa distribuição é obtida pela combinação exponencial de duas distribuições de Poisson de médias diferentes. A função de densidade de uma variável y com distribuição Poisson Dupla de média μ e parâmetro de dispersão ϕ^{-1} , denotada como $y \sim \text{PD}(\mu, \phi)$, é dada por

$$f(y; \mu, \phi) = k(\mu, \phi) \phi e^{-\phi\mu} \frac{(\mu\phi)^{\phi y}}{\Gamma(\phi y + 1)} \quad y = 0, 1, 2, \dots, \quad (3.20)$$

em que $k(\mu, \phi)$ é a constante de normalização cujo valor depende dos valores de μ e ϕ da seguinte forma

$$\frac{1}{k(\mu, \phi)} \cong 1 + \frac{1 - \phi}{12\phi\mu} \left(1 + \frac{1}{\phi\mu} \right).$$

Essa distribuição tem média aproximadamente igual a μ e variância aproximadamente igual a μ/ϕ . Note que, se $\phi = 1$, a Equação (3.20) se reduz à densidade Poisson padrão. Agora, suponha que $y_1, \dots, y_n$ são variáveis aleatórias independentes tais que $y_i \sim \text{PD}(\mu_i, \phi_i)$. Suponha também uma estrutura de regressão na qual μ_i e ϕ_i satisfazem as relações funcionais dadas nas equações (3.1) e (3.2), respectivamente. Considerando $\boldsymbol{\theta} = (\beta^\top, \gamma^\top)^\top$, como sendo o vetor de parâmetros do modelo, o logaritmo da função de verosimilhança sem levar em consideração a constante de integração $k(\mu_i, \phi_i)$ é dado por (para mais detalhe ver Efron (1986))

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \{ \log(\phi_i) - \phi_i \mu_i + \phi_i y_i \log(\mu_i / y_i) - \log \Gamma(\phi_i y_i + 1) \}. \quad (3.21)$$

Assim, as derivadas de primeira e segunda ordem da função de log-verosimilhança para $j, \ell = 1, \dots, k$ são dadas por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \phi_i (y_i - \mu_i) \left(\frac{\omega_i}{\mu_i} \right)^{-1/2} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad (3.22)$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} = \sum_{i=1}^n (y_i - \mu_i) \frac{\partial}{\partial \beta_\ell} \left(\mu_i^{-1} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j} \right) - \sum_{i=1}^n \frac{\phi_i}{\mu_i} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_\ell}, \quad (3.23)$$

considerando agora, $j = 1, \dots, k$ e $\ell = 1, \dots, q$,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_\ell} = \sum_{i=1}^n (y_i - \mu_i) \left(\frac{\omega_i}{\mu_i} \right)^{-1/2} \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell}, \quad (3.24)$$

e, por último, para $j, \ell = 1, \dots, q$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^n c_i \frac{\partial \eta_{2i}}{\partial \phi_i} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad (3.25)$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} = \sum_{i=1}^n c'_i \left(\frac{\partial \eta_{2i}}{\partial \phi_i} \right)^2 \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_\ell} + \sum_{i=1}^n c_i \frac{\partial}{\partial \gamma_\ell} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad (3.26)$$

em que $\omega_i = \mu_i^{-1} (\partial \mu_i / \partial \eta_{1i})^2$, $c_i = \{y_i [\log(\mu_i \phi_i) + 1] + \phi_i^{-1} - \mu_i - \psi(\phi_i y_i + 1)\}$ e $c'_i = \{\phi_i^{-2} + y_i \phi_i^{-1} - \psi'(\phi_i y_i + 1)\}$. Desta forma, tem-se que as funções escore para β e γ podem ser escritas em forma matricial, respectivamente, como

$$U_\beta = J_1^\top \Phi W^{1/2} V^{-1/2} (y - \mu) \quad \text{e} \\ U_\gamma = J_2^\top W_\gamma^{-1} c,$$

com $J_1 = \partial \eta_1 / \partial \beta$, $J_2 = \partial \eta_2 / \partial \gamma$ matrizes de ordem $n \times k$ e $n \times q$ respectivamente, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, $V = \text{diag}\{V(\mu_1), \dots, V(\mu_n)\}$ com $V(\mu_i) = \mu_i$, $W_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$ com $h'(\phi_i) = \partial \eta_{2i} / \partial \phi_i$, $y = (y_1, \dots, y_n)^\top$, $\mu = (\mu_1, \dots, \mu_n)^\top$ e $c = (c_1, \dots, c_n)^\top$.

Sob condições de regularidade, aplicando esperança nas equações (3.23), (3.24) e (3.26), obtemos que a matriz de informação de Fisher para $\boldsymbol{\theta}$ é bloco diagonal dada por $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{diag}\{K_{\beta\beta}, K_{\gamma\gamma}\}$, em que $K_{\beta\beta}$ e $K_{\gamma\gamma}$ são as matrizes de informação de Fisher para β e γ respectivamente. Logo, pode-se escrever, respectivamente, na forma matricial $K_{\beta\beta}$ e $K_{\gamma\gamma}$ da seguinte forma

$$K_{\beta\beta} = E \left(-\frac{\partial^2 \ell(\beta; \gamma)}{\partial \beta \partial \beta^\top} \right) = J_1^\top W J_1, \quad (3.27)$$

$$K_{\gamma\gamma} = E \left(-\frac{\partial^2 \ell(\beta; \gamma)}{\partial \gamma \partial \gamma^\top} \right) = J_2^\top P J_2, \quad (3.28)$$

com $P = T W_\gamma^{-2}$ e $T = \text{diag} = \{E(c'_1), \dots, E(c'_n)\}$.

A função desvio para a média supondo ϕ_i conhecido, para $i = 1, \dots, n$, é dada por

$$D^*(y, \hat{\mu}) = 2 \sum_{i=1}^n \phi_i \left\{ \hat{\mu}_i - y_i \left[1 + \log \left(\frac{\hat{\mu}_i}{y_i} \right) \right] \right\}. \quad (3.29)$$

No caso em que $y_i = 0$ para algum i , o i -ésimo componente do desvio é dado por $2\phi_i \hat{\mu}_i$.

3.3 Resíduos ordinários

Como desenvolvido no Capítulo 2, a construção dos resíduos ordinários depende do processo iterativo scoring de Fisher usado para obter as estimativas de máxima verosimilhança. Como apresentado no Apêndice A, se substituirmos a matriz X e Z por J_1 e J_2 , respectivamente, pode-se escrever o processo iterativo scoring de Fisher como um processo iterativo de mínimos quadrados ponderados da seguinte forma:

$$\beta^{(m+1)} = (J_1^\top \Phi^{(m)} W^{(m)} J_1)^{-1} J_1^\top \Phi^{(m)} W^{(m)} u_1^{(m)}, \quad (3.30)$$

$$\gamma^{(m+1)} = (J_2^\top P^{(m)} J_2)^{-1} J_2^\top P^{(m)} u_2^{(m)}, \quad (3.31)$$

em que $u_1^{(m)} = J_1 \beta^{(m)} + W^{1/2(m)} V^{-1/2(m)} (y - \mu^{(m)})$ e $u_2^{(m)} = J_2 \gamma^{(m)} + V_\gamma^{-1(m)} H_\gamma^{(m)} (t - \mu^{*(m)})$.

Após a convergência do processo iterativo, tem-se que

$$\hat{\beta} = (J_1^\top \hat{\Phi} \hat{W} J_1)^{-1} J_1^\top \hat{\Phi} \hat{W} u_1, \quad (3.32)$$

$$\hat{\gamma} = (J_2^\top \hat{P} J_2)^{-1} J_2^\top \hat{P} u_2, \quad (3.33)$$

com $u_1 = \hat{\eta}_1 + \hat{W}^{1/2} \hat{V}^{-1/2} (y - \hat{\mu})$ e $u_2 = \hat{\eta}_2 + \hat{V}_\gamma^{-1} \hat{H}_\gamma (t - \hat{\mu}^*)$.

As Equações (3.32) e (3.33) podem ser vistas como estimadores de mínimos quadrados ordinários de β e γ , respectivamente. Isto é, para obter $\hat{\beta}$ considere a regressão da variável auxiliar $\hat{\Phi}^{1/2} \hat{W}^{1/2} u_1$ sob a covariável $\hat{\Phi}^{1/2} \hat{W}^{1/2} J_1$, e para obter $\hat{\gamma}$ se considera a regressão da variável auxiliar $\hat{P}^{1/2} u_2$ sob a covariável $\hat{P}^{1/2} J_2$. Consequentemente, os resíduos ordinários das regressões anteriores são, respectivamente,

$$r^\beta = \hat{\Phi}^{1/2} \hat{W}^{1/2} (u_1 - \hat{\eta}_1),$$

$$r^\gamma = \hat{P}^{1/2} (u_2 - \hat{\eta}_2).$$

Logo, os resíduos para a i -ésima observação baseados nos submodelos da média (Equação (2.20)) e para o submodelo da dispersão (Equação (2.21)) são, respectivamente,

$$r_i^\beta = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}_i^{-1} \hat{V}_i}}, \quad i = 1, \dots, n \quad \text{e}$$

$$r_i^\gamma = \frac{t_i - \hat{\mu}_i^*}{\sqrt{-d''(\hat{\phi}_i)}}, \quad i = 1, \dots, n.$$

Esses resíduos podem ser padronizados usando as matrizes de projeção da solução de

mínimos quadrados ordinários dadas em (3.32) e (3.33), isto é,

$$\widehat{S}^\beta = \widehat{\Phi}^{1/2} \widehat{W}^{1/2} J_1 \left(J_1^\top \widehat{\Phi} \widehat{W} J_1 \right)^{-1} J_1^\top \widehat{W}^{1/2} \widehat{\Phi}^{1/2}, \quad (3.34)$$

$$\widehat{S}^\gamma = \widehat{P}^{1/2} J_2 \left(J_2^\top \widehat{P} J_2 \right)^{-1} J_2^\top \widehat{P}^{1/2}. \quad (3.35)$$

Logo, os resíduos padronizados são

$$r_{pi}^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\phi}_i^{-1} \widehat{V}_i(1 - \widehat{s}_{ii}^\beta)}}, \quad i = 1, \dots, n,$$

$$r_{pi}^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d''(\widehat{\phi}_i)(1 - \widehat{s}_{ii}^\gamma)}}, \quad i = 1, \dots, n,$$

em que $\widehat{s}_{ii}^\beta$ e $\widehat{s}_{ii}^\gamma$ são elementos da diagonal principal das matrizes dadas em (3.34) e (3.35), respectivamente.

3.4 Resíduo combinado padronizado

O resíduo combinado, construído a partir dos processos iterativos de maximização da verossimilhança dados nas Equações (3.30) e (3.31), é dado pela combinação

$$r_i^{\beta\gamma} = (y_i - \widehat{\mu}_i) + (t_i - \widehat{\mu}_i^*); \quad i = 1, \dots, n. \quad (3.36)$$

A variância de $r_i^{\beta\gamma}$ pode ser obtida através das Equações (B.2), (B.4) e (B.5) no Apêndice B. Seja $\text{Var}(r_i^{\beta\gamma}) = \zeta_i$. Então, tem-se que

$$\zeta_i = \phi_i^{-1} b''(\theta_i) - d''(\phi_i). \quad (3.37)$$

Assim, a proposta deste trabalho é padronizar o resíduo combinado $r_i^{\beta\gamma}$ da seguinte forma

$$r_{pi}^{\beta\gamma} = \frac{(y_i - \widehat{\mu}_i) + (t_i - \widehat{\mu}_i^*)}{\sqrt{\widehat{\zeta}_i}}, \quad i = 1, \dots, n, \quad (3.38)$$

em que $\widehat{\zeta}_i$ é dado por ζ_i avaliado em $\widehat{\mu}_i$ e $\widehat{\phi}_i$.

Outra proposta para padronizar o resíduo combinado é obtendo as matrizes de covariâncias de u_1 e u_2 , considerando

$$\text{Cov}(\widehat{\beta}) \approx (J_1^\top \Phi W J_1)^{-1},$$

$$\text{Cov}(\widehat{\gamma}) \approx (J_2^\top P J_2)^{-1}.$$

Assim, pode-se reescrever as equações (3.32) e (3.33) como

$$\begin{aligned}(J_1^\top \Phi W J_1) \hat{\beta} &\approx J_1^\top \Phi W u_1, \\ (J_2^\top P J_2) \hat{\gamma} &\approx J_2^\top P u_2.\end{aligned}$$

Considerando $\widehat{W} \approx W$, $\widehat{\Phi} \approx \Phi$ e $\widehat{P} \approx P$, temos que

$$\begin{aligned}\widehat{\text{Cov}}(u_1) &\approx \widehat{W}^{-1} \widehat{\Phi}^{-1}, \\ \widehat{\text{Cov}}(u_2) &\approx \widehat{P}^{-1}.\end{aligned}$$

Desta forma,

$$\begin{aligned}\widehat{\text{Cov}}(r^\beta) &\approx (I - \widehat{S}^\beta), \\ \widehat{\text{Cov}}(r^\gamma) &\approx (I - \widehat{S}^\gamma).\end{aligned}$$

Portanto, o resíduo combinado padronizado para a i -ésima observação baseado nas covariâncias de u_1 e u_2 é dado por

$$r_{pi}^{\beta\gamma} = \frac{r_i^\beta + r_i^\gamma}{\sqrt{(1 - \widehat{s}_{ii}^\beta) + (1 - \widehat{s}_{ii}^\gamma) + 2\widehat{m}_{ii}}}, \quad i = 1, \dots, n, \quad (3.39)$$

em que $\widehat{m}_{ii}$ é o i -ésimo elemento da diagonal principal da matriz $M = \text{Cov}(r^\beta, r^\gamma)$. Porém, neste trabalho, o interesse não é avaliar o resíduo dado em (3.39), já que este último envolve o calculo das matrizes de projeção dadas nas equações (3.34) e (3.35), o que pode trazer um alto custo computacional quando o número observações é consideravelmente grande.

3.4.1 Resíduo combinado padronizado generalizado

Para o desenvolvimento do resíduo combinado foi necessário impor condições sob a forma funcional da função $c(y_i; \phi_i)$, a fim de encontrar expressões de forma fechada para a variância. Nesta seção, relaxaremos a forma funcional da função dita, a fim de se obter uma versão generalizada do resíduo combinado. A eliminação dessa restrição só afeta a estimação dos parâmetros do modelo para dispersão (ϕ_i), como será mostrado posteriormente.

Sejam $y_1, \dots, y_n$ variáveis aleatórias independentes cada uma com densidade dada por (2.1), com μ_i e $\phi_i > 0$ satisfazendo as relações funcionais dadas pelas equações (3.1) e (3.2) respectivamente. O logaritmo da função de verosimilhança do modelo não linear é dado pela Equação (2.4). Desta forma, para β temos que a função score e a matriz de

informação de Fisher são

$$U_\beta = J_1^\top \Phi W^{1/2} V^{-1/2} (y - \mu),$$

$$K_{\beta\beta} = J_1^\top \Phi W J_1,$$

em que $J_1 = \partial\eta_1/\partial\beta$ é uma matriz de ordem $n \times k$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, $\omega_i = (\partial\mu_i/\partial\eta_{1i})^2/V_i$, $V = \text{diag}\{V_1, \dots, V_n\}$, $V_i = \partial\mu_i/\partial\theta_i$, $y = (y_1, \dots, y_n)^\top$ e $\mu = (\mu_1, \dots, \mu_n)^\top$.

Obtendo a função escore para γ , tem-se que

$$\begin{aligned} \frac{\partial\ell(\boldsymbol{\theta})}{\partial\gamma_j} &= \sum_{i=1}^n \frac{\partial\ell_i(\mu_i; \phi_i)}{\partial\phi_i} \frac{\partial\phi_i}{\partial\eta_{2i}} \frac{\partial\eta_{2i}}{\partial\gamma_j}, \quad j = 1, \dots, q \\ &= \sum_{i=1}^n t_i \frac{\partial\phi_i}{\partial\eta_{2i}} \frac{\partial\eta_{2i}}{\partial\gamma_j}, \end{aligned} \quad (3.40)$$

em que, $t_i = \theta_i y_i - b(\theta_i) + c'(y_i; \phi_i)$. Consequentemente, pode-se escrever a função escore para γ em forma matricial como

$$U_\gamma = J_2^\top H_\gamma^{-1} t,$$

em que, $J_2 = \partial\eta_2/\partial\gamma$ é uma matriz de ordem $n \times q$, $H_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$, $h'(\phi_i) = \partial\eta_{2i}/\partial\phi_i$ e $t = (t_1, \dots, t_n)^\top$.

É necessário para obter a matriz de informação de Fisher a partir da segunda derivada do logaritmo da função de verosimilhança, isto é,

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_\ell} = \sum_{i=1}^n t'_i \left(\frac{\partial\phi_i}{\partial\eta_{2i}} \right)^2 \frac{\partial\eta_{2i}}{\partial\gamma_j} \frac{\partial\eta_{2i}}{\partial\gamma_\ell} + \sum_{i=1}^n t_i \frac{\partial\eta_{2i}}{\partial\gamma_j} \frac{\partial}{\partial\gamma_\ell} \left(\frac{\partial\phi_i}{\partial\eta_{2i}} \right). \quad (3.41)$$

Note que, sob condições usuais de regularidade, temos que

$$E(U_\gamma) = \mathbf{0} \quad \Rightarrow \quad E\left(\sum_{i=1}^n t_i \frac{\partial\phi_i}{\partial\eta_{2i}} \frac{\partial\eta_{2i}}{\partial\gamma_j}\right) = 0 \quad \Rightarrow \quad E(t_i) = 0.$$

Assim, a esperança da Equação (3.41) é dada por

$$E\left(\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_\ell}\right) = \sum_{i=1}^n E(t'_i) \left(\frac{\partial\phi_i}{\partial\eta_{2i}} \right)^2 \frac{\partial\eta_{2i}}{\partial\gamma_j} \frac{\partial\eta_{2i}}{\partial\gamma_\ell}.$$

Desta forma, tem-se que a matriz de informação de Fisher para γ pode ser escrita em forma matricial como

$$K_{\gamma\gamma} = J_2^\top T H_\gamma^{-2} J_2, \quad (3.42)$$

com $T = \text{diag}\{-E(t'_1), \dots, -E(t'_n)\}$. Pela Equação (3.7), temos que a matriz de informação

de Fisher para θ é bloco diagonal dada por $K_{\theta\theta} = \text{diag}\{K_{\beta\beta}; K_{\gamma\gamma}\}$. O nosso interesse é na construção de um resíduo que dependa dos processos iterativos Scoring de Fisher, uma vez que o mesmo pode ser representado por um processo iterativo de mínimos quadrados ponderados. Assim, para β , o processo é dado na Equação (3.30) e, para γ , tem-se

$$\gamma^{(m+1)} = (J_2^\top P^{(m)} J_2)^{-1} J_2^\top P^{(m)} u^{*(m)},$$

em que $P^{(m)} = T^{(m)} H_\gamma^{-2(m)}$ e $u^{*(m)} = J_2 \gamma^{(m)} + T^{-1(m)} H_\gamma^{(m)} t$.

Após convergência do processo iterativo, tem-se que

$$\hat{\gamma} = (J_2^\top \hat{P} J_2)^{-1} J_2^\top \hat{P} u^*.$$

Logo, $\hat{\gamma}$ pode ser obtido como o estimador de mínimos quadrados ordinários da regressão da variável auxiliar $\hat{P}^{1/2} u^*$ sobre a covariável $\hat{P}^{1/2} J_2$. Assim, temos que o resíduo ordinário para γ é

$$\begin{aligned} r^\gamma &= \hat{P}^{1/2} (u^* - J_2 \hat{\gamma}), \\ &= \hat{P}^{1/2} \hat{T}^{-1} \hat{H}_\gamma t, \\ &= \hat{T}^{-1/2} t. \end{aligned}$$

Então, para a i -ésima observação tem-se que

$$r_i^\gamma = \frac{t_i}{\sqrt{-E(t_i')}}, \quad i = 1, \dots, n. \quad (3.43)$$

Consequentemente, o resíduo combinado é dado por

$$r_i^{\beta\gamma} = y_i - \hat{\mu}_i + t_i,$$

em que a variância de r_i^γ é dada por

$$\begin{aligned} \zeta_i &= \text{Var}(r_i^{\beta\gamma}) \\ &= \text{Var}(y_i + t_i) \\ &= \text{Var}[(1 + \theta_i)y_i + c'(y_i; \phi_i)] \\ &= (1 + \theta_i)^2 \text{Var}[y_i] + \text{Var}[c'(y_i; \phi_i)] \\ &\quad + 2(1 + \theta_i) \text{Cov}[y_i; c'(y_i; \phi_i)]. \end{aligned} \quad (3.44)$$

Finalmente, o resíduo combinado padronizado é dado por

$$r_{pi}^{\beta\gamma} = \frac{y_i - \hat{\mu} + \hat{t}_i}{\sqrt{\hat{\zeta}_i}}, \quad i = 1, \dots, n, \quad (3.45)$$

em que, $\widehat{\zeta}_i$ é ζ_i avaliado em $\widehat{\mu}_i, \widehat{\phi}_i, \widehat{\text{Var}} \left[c'(y_i; \widehat{\phi}_i) \right]$ e $\widehat{\text{Cov}} \left[y_i; c'(y_i; \widehat{\phi}_i) \right]$.

No caso em que a variável resposta do modelo têm distribuição binomial negativa, como é apresentado na Seção 3.2.1, a estimação de $\widehat{\text{Var}} \left[c'(y_i; \widehat{\phi}_i) \right]$ e $\widehat{\text{Cov}} \left[y_i; c'(y_i; \widehat{\phi}_i) \right]$ pode ser feita de forma numérica da seguinte forma

$$\begin{aligned} \widehat{\text{Var}} \left[c'(y_i; \widehat{\phi}_i) \right] &= \widehat{\text{E}} \left[c'(y_i; \widehat{\phi}_i)^2 \right] - \widehat{\text{E}} \left[c'(y_i; \widehat{\phi}_i) \right]^2 \\ \widehat{\text{Cov}} \left[y_i; c'(y_i; \widehat{\phi}_i) \right] &= \widehat{\text{E}} \left[y_i c'(y_i; \widehat{\phi}_i) \right] - \text{E}[y_i] \widehat{\text{E}} \left[c'(y_i; \widehat{\phi}_i) \right]. \end{aligned}$$

Obtendo os estimadores de forma numérica para as esperanças anteriores, tem-se

$$\begin{aligned} \widehat{\text{E}} \left[c'(y_i; \widehat{\phi}_i)^2 \right] &= \sum_{j=0}^{\infty} c'(j; \widehat{\phi}_i)^2 P(y_i = j), \\ \widehat{\text{E}} \left[c'(y_i; \widehat{\phi}_i) \right] &= \sum_{j=0}^{\infty} c'(j; \widehat{\phi}_i) P(y_i = j), \\ \widehat{\text{E}} \left[y_i c'(y_i; \widehat{\phi}_i) \right] &= \sum_{j=0}^{\infty} j c'(j; \widehat{\phi}_i) P(y_i = j). \end{aligned}$$

em que $P(y_i = j)$ é a probabilidade de y_i assumir o valor j .

Avaliação numérica

Neste capítulo serão apresentadas simulações que têm como objetivo investigar a distribuição empírica dos resíduos ordinário padronizado, componente de desvio e combinado. Os modelos considerados pertencem a família exponencial com dispersão variável. Ademais, foram realizadas simulações com modelos sob especificação incorreta e modelos com presença de observações aberrantes.

4.1 Avaliação dos resíduos através de simulações de Monte Carlo

Para descrever a distribuição empírica dos resíduos ordinário padronizado, componente de desvio padronizado ¹ e resíduo combinado padronizado, no modelo não linear generalizado com dispersão variável foram realizadas simulações de Monte Carlo com 5000 réplicas sob diferentes cenários. A avaliação da distribuição empírica desses resíduos é feita através de medidas que caracterizam uma distribuição, tais como: média, o erro padrão, assimetria e curtose; lembrando que para a distribuição normal padrão tem-se que os valores destas medidas são 0, 1, 0 e 0 respectivamente.

1. Neste cenário consideramos que a variável resposta do modelo tem distribuição normal de média μ_i e variância ϕ_i^{-1} , sendo esta distribuição a mais difundida da família exponencial. Consideramos amostras de tamanho $n = 30$. Desta forma, $y_1, \dots, y_{30}$ são variáveis aleatórias independentes, em que y_i tem distribuição normal de média μ_i e variância ϕ_i^{-1} , isto é denotado como $y_i \sim N(\mu_i; \phi_i^{-1})$, com $i = 1, \dots, 30$. Neste caso, foi considerado que a média da variável resposta e o parâmetro de dispersão satisfazem relações funcionais dadas por

$$\begin{aligned}\mu_i &= \beta_1 + x_i^{\beta_2}, \\ \log(\phi_i) &= \gamma_1 + z_i^{\gamma_2},\end{aligned}\tag{4.1}$$

¹Neste trabalho só consideramos o resíduo tipo desvio para a média dado na Equação (2.26).

para $i = 1, \dots, n$, em que as covariáveis são geradas da distribuição uniforme de forma independente, isto é, $x_i \sim U(0.1; 1.1)$ e $z_i \sim U(0.4; 1.4)$, e foram mantidas constantes em todos os cenários analisado. Além disso, consideramos $\beta_1 = -1.4$, $\beta_2 = 0.8, \gamma_1 = 0.9, \gamma_2 = -1$.

Tabela 4.1: Médias, erros-padrão, assimetria e curtose empíricas dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim N(\mu_i; \phi_1^{-1})$ com $\mu_i = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = -1.4$, $\beta_2 = 0.8, \gamma_1 = 0.9, \gamma_2 = -1$, $x_i \sim U(0.1, 1.1)$, $z_i \sim U(0.4, 1.4)$, $i = 1, \dots, 30$.

i	Média			Erro-padrão			Assimetria			Curtose		
	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0529	0.0529	0.0501	1.0362	1.0362	0.9573	0.0019	0.0019	-0.9437	-0.2490	-0.2490	0.8401
2	-0.1111	-0.1111	-0.0974	1.1502	1.1502	1.0590	-0.0058	-0.0058	-0.8852	-0.3649	-0.3649	0.5246
3	0.0635	0.0635	0.0592	1.0394	1.0394	0.9667	-0.0389	-0.0389	-0.9625	-0.2414	-0.2414	0.8919
4	0.0313	0.0313	0.0254	1.0583	1.0583	1.0049	-0.0578	-0.0578	-1.0292	-0.3378	-0.3378	1.0147
5	0.0013	0.0013	0.0005	1.0915	1.0915	1.0096	-0.0015	-0.0015	-0.7594	-0.3119	-0.3119	0.4026
6	0.0214	0.0214	0.0165	1.0540	1.0540	0.9989	-0.0304	-0.0304	-0.9679	-0.3951	-0.3951	0.7662
7	-0.0923	-0.0923	-0.0863	1.1760	1.1760	1.0462	0.0809	0.0809	-0.5716	-0.4337	-0.4337	-0.1301
8	-0.0854	-0.0854	-0.0751	1.0852	1.0852	1.0171	0.0140	0.0140	-0.8861	-0.2477	-0.2477	0.7573
9	-0.0742	-0.0742	-0.0497	0.9771	0.9771	0.8851	0.0639	0.0639	-0.5967	-0.3986	-0.3986	0.0725
10	-0.0161	-0.0161	-0.0163	1.0562	1.0562	1.0027	0.0037	0.0037	-0.9609	-0.0796	-0.0796	1.1221
11	0.0200	0.0200	0.0195	1.0335	1.0335	0.9795	0.0081	0.0081	-0.9239	-0.1361	-0.1361	1.0070
12	0.0029	0.0029	0.0076	1.0231	1.0231	0.9762	-0.0230	-0.0230	-0.9498	-0.4250	-0.4250	0.7264
13	-0.0619	-0.0619	-0.0539	1.0405	1.0405	1.0025	0.0003	0.0003	-0.9311	-0.1951	-0.1951	0.9274
14	0.0795	0.0795	0.0664	1.0798	1.0798	1.0051	-0.0481	-0.0481	-0.9750	-0.1878	-0.1878	0.9593
15	0.0031	0.0031	0.0062	1.0848	1.0848	0.9493	0.0045	0.0045	-0.5384	-0.6729	-0.6729	-0.3277
16	-0.0058	-0.0058	-0.0100	1.1008	1.1008	1.0191	0.0464	0.0464	-0.7126	-0.3134	-0.3134	0.2808
17	-0.0310	-0.0310	-0.0266	1.1289	1.1289	0.9966	-0.0003	-0.0003	-0.6177	-0.5437	-0.5437	-0.0570
18	0.0017	0.0017	0.0018	1.0455	1.0455	0.9959	-0.0378	-0.0378	-0.9825	-0.3026	-0.3026	0.9239
19	0.0218	0.0218	0.0200	1.0805	1.0805	0.9983	0.0028	0.0028	-0.7909	-0.2823	-0.2823	0.4899
20	-0.0164	-0.0164	-0.0070	1.0136	1.0136	0.9509	0.0280	0.0280	-0.9129	-0.3559	-0.3559	0.6219
21	-0.1167	-0.1167	-0.0963	1.2312	1.2312	1.0411	-0.0140	-0.0140	-0.9626	-0.1495	-0.1495	0.9737
22	0.0592	0.0592	0.0540	1.0718	1.0718	0.9827	-0.0275	-0.0275	-0.6926	-0.4931	-0.4931	0.1835
23	0.0689	0.0689	0.0649	1.0327	1.0327	0.9681	-0.0809	-0.0809	-1.0778	-0.3255	-0.3255	1.1394
24	0.0498	0.0498	0.0486	1.0264	1.0264	0.9447	-0.0106	-0.0106	-0.9274	-0.4568	-0.4568	0.6166
25	0.0531	0.0531	0.0473	1.0527	1.0527	0.9813	-0.0466	-0.0466	-1.0164	-0.2997	-0.2997	1.0107
26	0.0298	0.0298	0.0283	1.0374	1.0374	0.9528	0.0317	0.0317	-0.8929	-0.4148	-0.4148	0.5735
27	0.0651	0.0651	0.0570	1.1328	1.1328	0.9910	-0.0783	-0.0783	-0.5823	-0.7951	-0.7951	-0.3369
28	-0.0468	-0.0468	-0.0326	1.0019	1.0019	0.9511	0.0320	0.0320	-0.8650	-0.1997	-0.1997	0.7188
29	0.0811	0.0811	0.0726	1.0818	1.0818	0.9745	0.0105	0.0105	-0.7302	-0.3677	-0.3677	0.2215
30	-0.0857	-0.0857	-0.0780	1.1094	1.1094	1.0294	0.0496	0.0496	-0.8707	-0.3277	-0.3277	0.5302
M	-0.0012	-0.0012	0.0006	1.0711	1.0711	0.9879	-0.0041	-0.0041	-0.8505	-0.3435	-0.3435	0.5815

Na Tabela 4.1, observa-se que as médias dos resíduos avaliados são próximas de zero², a similaridade também ocorre com os erros padrão que são próximos a um. As medidas de assimetria e com maior claridade a curtose, mostram o afastamento de todos os resíduos da distribuição normal padrão.

Porém, não é necessariamente de interesse obter um resíduo cuja distribuição seja aproximadamente normal, já que é possível obter uma estimação da distribuição empírica deste via simulações (ATKINSON, 1985). As vantagens que o resíduo combinado padronizado tem sob os outros propostos na literatura é levar em conta a

²M: é a média aritmética por coluna.

modelagem conjunta dos submodelos da média e da dispersão, assim como a redução do custo computacional ao negligenciar o cálculo da matriz de projeção. Note que os valores da média, erro-padrão, assimetria e curtose para o resíduo ordinário e componente desvio na Tabela 4.1 são iguais, devido a que sua forma funcional é a mesma, quando a variável resposta do modelo tem distribuição normal e a função de ligação é a identidade.

2. Seguidamente elaboramos uma outra simulação considerando que a variável resposta do modelo tem distribuição gama com média μ_i e parâmetro de dispersão ϕ_i^{-1} , denotada como $y_i \sim G(\mu_i; \phi_i)$, sendo esta distribuição muito útil quando as realizações da variável resposta pertencem aos reais positivos ($\mathbb{R}^+$). Desta forma, $y_1, \dots, y_{30}$ são variáveis aleatórias independentes, em que y_i tem distribuição gama de média μ_i e parâmetro de dispersão ϕ_i , isto é denotado como $y_i \sim G(\mu_i; \phi_i)$, com $i = 1, \dots, 30$. Neste caso, foi considerado que a média da variável resposta e o parâmetro de dispersão satisfazem relações funcionais dadas por:

$$\begin{aligned}\mu_i^{-1} &= \beta_1 + x_i^{\beta_2}, \\ \log(\phi_i) &= \gamma_i + z_i^{\gamma_2}\end{aligned}\tag{4.2}$$

para $i = 1, \dots, n$, em que $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1.5$.

Pode-se observar na Tabela 4.2 que o comportamento da média dos resíduos está próxima do valor zero para os três casos, sendo a do resíduo deviance a mais distante. O comportamento do erro-padrão está próximo ao valor um para todos os resíduos, as maiores diferenças apresentam-se nos valores da assimetria e a curtose, já que a distribuição empírica do resíduo deviance é aproximadamente simétrica, mas com curtose negativa, o que indica afastamento da distribuição normal padrão.

3. Posteriormente, consideramos uma outra alternativa para dados reais positivos, considerando que a variável resposta do modelo tem distribuição normal inversa com média μ_i e parâmetro de dispersão ϕ_i^{-1} , denotada por $y_i \sim \text{NI}(\mu_i; \phi_i)$. Desta forma, $y_1, \dots, y_{30}$ são variáveis aleatórias independentes, em que y_i tem distribuição normal inversa de média μ_i e parâmetro de dispersão ϕ_i^{-1} , com $i = 1, \dots, 30$. Neste caso, foi considerado que a média da variável resposta e o parâmetro de dispersão satisfazem as relações funcionais dadas por

$$\begin{aligned}\log(\mu_i) &= \beta_1 + x_i^{\beta_2}, \\ \log(\phi_i) &= \gamma_i + z_i^{\gamma_2},\end{aligned}\tag{4.3}$$

para $i = 1, \dots, n$, em que $\beta_1 = -0.45$, $\beta_2 = 0.9$, $\gamma_1 = -0.2$, $\gamma_2 = -0.4$.

Na Tabela 4.3 pode-se observar mais uma vez que os resíduos não têm distribuição aproximadamente $N(0, 1)$. Basta ver o comportamento da média e da variância

Tabela 4.2: Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{gama}(\mu_i; \phi_1)$ com $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1.5$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 30$.

i	Média			Erro-padrão			Assimetria			Curtose		
	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$
1	-0.0681	-0.1901	-0.0633	1.0069	1.0473	1.0426	0.6518	0.0229	-1.6062	0.3642	-0.2360	2.9941
2	0.1376	0.0352	0.1295	1.2066	1.1792	0.9105	0.3711	-0.0892	-0.6799	-0.1334	-0.3099	0.2738
3	-0.0911	-0.2193	-0.0826	0.9845	1.0308	1.0341	0.6961	0.0606	-1.4265	0.3802	-0.2877	2.4229
4	-0.0654	-0.1620	-0.0551	1.0403	1.0691	1.0100	0.4759	0.0433	-1.2266	-0.1053	-0.4049	1.3695
5	0.0408	-0.1035	0.0218	1.0884	1.0788	0.8775	0.7183	0.0388	-1.8128	0.3847	-0.3716	3.5617
6	-0.0752	-0.1766	-0.0710	1.0448	1.0744	1.0416	0.5371	0.0464	-1.2865	0.1252	-0.3143	1.6389
7	0.0568	-0.0897	0.0466	1.1075	1.0969	0.9690	0.6480	-0.0024	-0.9617	0.1305	-0.4083	0.8442
8	0.0802	-0.0409	0.0756	1.1169	1.0936	0.9398	0.6013	-0.0229	-0.8558	0.3787	-0.2179	0.7340
9	0.1150	-0.0191	0.1128	1.0787	1.0251	0.6807	0.6784	0.0249	-2.0631	0.3947	-0.3186	5.3635
10	0.0056	-0.1215	0.0204	1.0252	1.0285	0.8790	0.6721	0.0293	-1.9264	0.3551	-0.2699	4.7902
11	-0.0916	-0.2250	-0.0857	0.9613	1.0251	1.0574	0.6735	-0.0072	-1.2530	0.3850	-0.2520	1.6858
12	-0.0650	-0.1516	-0.0471	1.0229	1.0411	0.9241	0.4629	0.0932	-0.5624	-0.2520	-0.5480	-0.1751
13	0.0638	-0.0612	0.0579	1.0820	1.0748	0.9571	0.5854	-0.0347	-0.9877	0.1551	-0.2563	1.1088
14	-0.0942	-0.2283	-0.0873	0.9743	1.0290	1.0520	0.7399	0.0479	-1.3609	0.5323	-0.2109	2.1490
15	0.0059	-0.1462	-0.0058	1.0640	1.0737	0.9096	0.7334	0.0395	-1.7826	0.3235	-0.4162	3.4976
16	-0.0241	-0.1645	-0.0194	1.0013	1.0359	0.9927	0.6848	-0.0055	-1.1316	0.3598	-0.3201	1.2526
17	0.0183	-0.1264	0.0210	1.0286	1.0448	0.9516	0.6639	-0.0099	-1.0117	0.1783	-0.3976	1.0217
18	0.0409	-0.0710	0.0315	1.0911	1.0818	0.9075	0.5661	0.0108	-1.9305	0.1144	-0.2480	5.1510
19	0.0227	-0.1132	0.0254	1.0315	1.0427	0.9057	0.6252	-0.0231	-1.8073	0.2184	-0.3521	3.6561
20	0.0523	-0.0366	0.0622	1.1244	1.1179	0.8100	0.3460	-0.0199	-1.4416	-0.3247	-0.4876	2.1293
21	0.2360	0.1138	0.2038	1.2859	1.2057	0.8969	0.4666	-0.0862	-0.7408	0.0220	-0.2888	0.4351
22	-0.0534	-0.2005	-0.0506	1.0163	1.0415	0.9697	0.7783	0.0948	-1.3372	0.4787	-0.3966	1.7910
23	-0.1021	-0.1970	-0.0892	1.0148	1.0521	1.0181	0.4988	0.0796	-1.0082	-0.0953	-0.4596	0.7226
24	-0.0968	-0.1919	-0.0987	1.0644	1.1025	1.0552	0.4335	0.0694	-0.8332	-0.3531	-0.6255	0.2011
25	-0.1021	-0.2108	-0.1049	1.0456	1.0892	1.0939	0.5640	0.0527	-1.3735	0.0698	-0.3166	2.0412
26	-0.0757	-0.1747	-0.0750	1.0639	1.1027	1.0617	0.4483	0.0358	-1.1779	-0.2370	-0.4699	1.1893
27	-0.0344	-0.1896	-0.0438	1.0327	1.0671	0.9761	0.7311	0.0517	-1.4981	0.2856	-0.4978	2.2192
28	0.1027	-0.0232	0.0702	1.1020	1.0902	0.8793	0.5139	-0.1023	-2.1642	0.0722	-0.2594	5.7557
29	-0.0879	-0.2298	-0.0737	0.9649	1.0198	1.0101	0.7244	0.0576	-1.2519	0.3863	-0.3727	1.4555
30	0.0895	-0.0173	0.0767	1.2022	1.1978	0.9942	0.3846	-0.0833	-0.7137	-0.1861	-0.3712	0.2669
M	-0.0020	-0.1244	-0.0033	1.0625	1.0753	0.9603	0.5892	0.0137	-1.3071	0.1469	-0.3562	2.0516

empírica do resíduo deviance, e para os demais tipos de resíduos a assimetria e a curtose. Sendo o resíduo deviance o que apresenta pior comportamento, já que a média e o erro padrão diferem consideravelmente dos valores da distribuição normal padrão.

- Para analisar dados de contagem, a distribuição binomial negativa tem sido muito usada como uma alternativa à de Poisson em caso de sobredispersão. Neste cenário consideramos que a variável resposta do modelo tem distribuição binomial negativa de média μ_i e parâmetro dispersão ϕ_i^{-1} , denotada $y_i \sim \text{BN}(\mu_i, \phi_i)$. Desta forma, $y_1, \dots, y_{30}$ são variáveis aleatórias independentes, em que y_i tem distribuição binomial negativa de média μ_i e parâmetro de dispersão ϕ_i^{-1} , com $i = 1, \dots, 30$. Neste caso, foi considerado que a média da variável resposta e o parâmetro de dispersão satisfazem relações funcionais dadas na Equação (4.3), em que $\beta_1 = 1.6$, $\beta_2 = 0.8, \gamma_1 = -0.3, \gamma_2 = 0.5$.

Tabela 4.3: Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{NI}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = -0.45$, $\beta_2 = 0.9$, $\gamma_1 = -0.2$, $\gamma_2 = -0.4$, $x_i \sim \text{U}(0.1, 1.1)$, $z_i \sim \text{U}(0.4, 1.4)$, $i = 1, \dots, 30$.

i	Média			Erro-padrão			Assimetria			Curtose		
	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.1093	-0.3234	0.1032	1.0611	1.4007	1.0224	1.6000	-0.0137	0.4466	3.3368	-0.3908	0.9456
2	-0.1667	-0.6437	-0.1358	0.8971	1.4081	0.9823	1.2245	-0.0270	-0.8427	1.8915	-0.3926	0.9439
3	0.0996	-0.3231	0.0841	1.0926	1.4206	1.0276	1.7995	0.0397	0.2460	4.7126	-0.2011	1.0853
4	0.1128	-0.3366	0.1004	1.0890	1.4225	1.0355	1.6730	0.0360	0.4411	3.6695	-0.3967	0.8769
5	-0.0101	-0.4639	-0.0013	0.9499	1.3508	0.9471	1.5946	-0.0067	0.7523	3.4482	-0.3329	1.3493
6	0.1084	-0.3394	0.1006	1.1031	1.4101	1.0332	1.7601	0.0940	0.5602	4.1436	-0.3957	0.9542
7	-0.1175	-0.5480	-0.1074	0.9410	1.4120	0.9854	1.2048	0.0426	-0.5765	1.7194	-0.4286	0.5012
8	-0.1185	-0.5685	-0.1025	0.9185	1.3916	0.9855	1.3471	0.0120	-0.8529	2.3364	-0.3026	1.4265
9	-0.1864	-0.7934	-0.1752	0.7176	1.2638	0.7471	1.5659	0.0023	0.8128	3.3358	-0.3195	1.3763
10	0.0238	-0.4481	0.0290	0.9855	1.3582	0.9735	1.7326	0.0008	0.8750	3.8052	-0.2624	1.6053
11	0.0674	-0.3200	0.0620	1.0323	1.4047	0.9971	1.5250	-0.0360	-0.2924	3.4876	-0.2037	0.8526
12	-0.0076	-0.4208	0.0030	0.9706	1.3722	0.9581	1.3744	0.0570	-0.4015	2.4670	-0.4719	0.3383
13	-0.0959	-0.5233	-0.0749	0.9079	1.3590	0.9713	1.3549	-0.0174	-0.7717	2.3474	-0.2642	1.4153
14	0.0506	-0.3670	0.0347	1.0512	1.4193	1.0163	1.6107	0.0125	-0.0660	3.5881	-0.2619	0.7162
15	0.0479	-0.3829	0.0484	1.0291	1.4034	0.9840	1.4119	0.1017	0.7719	2.2877	-0.5801	0.8494
16	-0.0155	-0.4031	-0.0071	0.9877	1.3955	0.9814	1.3742	0.0212	-0.4087	2.6827	-0.2890	0.6487
17	-0.0921	-0.5093	-0.0839	0.9608	1.4159	0.9756	1.2148	0.0758	-0.4736	1.7728	-0.4597	0.2306
18	0.0230	-0.4610	0.0284	0.9835	1.3532	0.9680	1.8352	0.0311	0.9112	4.7223	-0.3412	1.9152
19	0.0684	-0.3762	0.0675	1.0502	1.3978	1.0221	1.7285	-0.0117	0.8049	4.2068	-0.2745	1.7230
20	-0.0324	-0.5466	-0.0228	0.9283	1.3268	0.9174	1.8440	0.0693	1.0149	4.7498	-0.3704	2.0863
21	-0.1724	-0.6437	-0.1068	0.9212	1.4511	0.9582	1.1546	-0.0950	-0.9650	1.7367	-0.2621	1.3915
22	0.0628	-0.3381	0.0516	1.0718	1.4345	1.0157	1.4677	0.1032	0.2225	2.8493	-0.4636	0.3202
23	0.1068	-0.3255	0.0980	1.0955	1.4015	1.0130	1.7240	0.1198	0.3676	3.8983	-0.3703	0.9237
24	0.0939	-0.3418	0.0764	1.0765	1.4313	1.0281	1.5342	0.0510	0.1184	3.0591	-0.4712	0.4936
25	0.0974	-0.3486	0.0944	1.0784	1.3995	1.0182	1.7553	0.0799	0.5558	4.4933	-0.3461	1.1892
26	0.1018	-0.3511	0.0896	1.1014	1.4122	1.0335	1.7790	0.0991	0.5271	4.6299	-0.4332	1.0196
27	0.1016	-0.2805	0.0958	1.0744	1.4277	1.0023	1.2800	0.1152	0.4578	1.8635	-0.6681	0.1886
28	-0.1422	-0.7401	-0.1180	0.8133	1.3977	0.8243	1.8185	-0.6816	1.0622	5.1946	3.3745	2.6832
29	0.1012	-0.2774	0.0906	1.0600	1.4336	1.0148	1.3300	-0.0321	-0.0864	2.1824	-0.3847	0.3986
30	-0.1209	-0.5643	-0.0835	0.9070	1.3702	0.9568	1.3087	0.0295	-0.7676	2.1934	-0.3329	1.1623
M	0.0033	-0.4437	0.0080	0.9952	1.3948	0.9799	1.5309	0.0091	0.1481	3.2271	-0.2432	1.0537

A Tabela 4.4 apresenta o comportamentos no caso em que a variável resposta do modelo têm distribuição binomial negativa. Esses resultados são similares aos casos em que as distribuições são contínuas, tendo assimetrias distantes de zero para o resíduo ordinário e combinado padronizado, e curtoses diferentes de zero para o resíduo deviance.

- Por último, apresentamos uma outra simulação em que a variável resposta do modelo tem distribuição Poisson dupla como alternativa ao modelo Poisson quando os dados apresentam o fenômeno de sobredispersão. Logo, consideramos que a variável resposta do modelo tem distribuição Poisson duplo de média μ_i e parâmetro dispersão ϕ_i^{-1} , denotada por $y_i \sim \text{PD}(\mu_i, \phi_i)$. Desta forma, $y_1, \dots, y_{30}$ são variáveis aleatórias independentes, em que y_i tem distribuição Poisson duplo de média μ_i e parâmetro de dispersão ϕ_i^{-1} , com $i = 1, \dots, 30$. Neste caso, foi considerado que a média da variável resposta e o parâmetro de dispersão satisfazem relações funcionais

Tabela 4.4: Médias, erros-padrão, assimetria e curtose empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{BN}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1.6$, $\beta_2 = 0.8$, $\gamma_1 = -0.3$, $\gamma_2 = 0.5$, $x_i \sim \text{U}(0.1; 1.1)$, $z_i \sim \text{U}(0.4; 1.4)$, $i = 1, \dots, 30$.

i	Média			Erro-padrão			Assimetria			Curtose		
	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0800	-0.1576	0.0784	1.0543	1.0440	1.0303	1.1830	-0.0145	1.0567	1.6297	-0.2099	1.2593
2	-0.1285	-0.3632	-0.1148	0.9790	1.0834	0.9020	0.8964	-0.1011	0.7846	0.6302	-0.3266	0.3963
3	0.0521	-0.1940	0.0507	1.0299	1.0495	1.0129	1.1202	-0.0737	0.9810	1.4170	-0.2828	1.0352
4	0.0664	-0.1639	0.0637	1.0782	1.0744	1.0498	1.0030	0.0108	0.9080	0.8986	-0.4082	0.6631
5	-0.0020	-0.2557	-0.0025	1.0275	1.0544	0.9994	1.1614	0.0250	1.0509	1.3966	-0.3572	1.0897
6	0.0681	-0.1624	0.0659	1.0745	1.0688	1.0474	1.0042	0.0051	0.9082	0.8893	-0.3812	0.6530
7	-0.0718	-0.3220	-0.0684	0.9586	1.0343	0.9302	1.0829	-0.0360	0.9278	1.2114	-0.3675	0.7840
8	-0.1105	-0.3527	-0.1037	0.9465	1.0401	0.9102	1.0490	-0.0747	0.8936	1.0519	-0.3181	0.6783
9	-0.1257	-0.3551	-0.1152	0.8751	0.9842	0.8423	0.9571	-0.1475	0.8891	0.9414	-0.3212	0.8004
10	0.0243	-0.2173	0.0238	1.0112	1.0400	0.9847	1.0850	-0.0388	0.9744	1.1698	-0.3251	0.8798
11	-0.0044	-0.2560	-0.0046	1.0181	1.0488	1.0032	1.2489	0.0140	1.0970	1.7721	-0.2085	1.3381
12	-0.0398	-0.2538	-0.0381	0.9745	1.0296	0.9511	0.8840	-0.0380	0.7806	0.5482	-0.4332	0.3003
13	-0.0670	-0.3078	-0.0641	0.9572	1.0285	0.9328	1.1419	-0.0571	0.9733	1.4541	-0.2718	0.9966
14	0.0434	-0.2018	0.0431	1.0196	1.0434	1.0044	1.1190	-0.0618	0.9772	1.3927	-0.2649	1.0077
15	0.0335	-0.2275	0.0310	1.0548	1.0662	1.0233	1.1033	0.0519	1.0104	1.1712	-0.5602	0.9200
16	-0.0280	-0.2797	-0.0269	0.9978	1.0400	0.9800	1.2202	0.0170	1.0566	1.8418	-0.3072	1.3151
17	-0.0786	-0.3355	-0.0763	0.9710	1.0383	0.9467	1.1916	0.0270	1.0297	1.5911	-0.3855	1.1109
18	0.0472	-0.1847	0.0453	1.0313	1.0588	1.0002	0.9745	-0.1008	0.8750	0.8118	-0.2650	0.5907
19	0.0579	-0.1931	0.0562	1.0403	1.0593	1.0149	1.0319	-0.0463	0.9250	0.8743	-0.3993	0.6241
20	-0.0049	-0.2136	-0.0073	1.0074	1.0394	0.9587	0.8862	0.0134	0.8140	0.5088	-0.5098	0.3754
21	-0.1951	-0.4552	-0.1689	0.9736	1.1166	0.8616	0.9830	-0.1202	0.8286	1.0077	-0.3096	0.6065
22	0.0570	-0.1967	0.0560	1.0409	1.0551	1.0235	1.0913	-0.0007	0.9649	1.1634	-0.4341	0.8314
23	0.0643	-0.1635	0.0620	1.0729	1.0719	1.0470	0.9652	-0.0001	0.8688	0.8063	-0.4161	0.5635
24	0.0578	-0.1561	0.0571	1.0357	1.0446	1.0123	0.9089	-0.0156	0.8136	0.6062	-0.4052	0.3901
25	0.0559	-0.1771	0.0540	1.0592	1.0703	1.0342	1.0613	-0.0404	0.9528	1.2137	-0.2377	0.9341
26	0.0624	-0.1613	0.0604	1.0590	1.0676	1.0318	0.9458	-0.0259	0.8515	0.7540	-0.3824	0.5307
27	0.0724	-0.1926	0.0694	1.0702	1.0731	1.0468	1.0881	0.0216	0.9785	1.0999	-0.5273	0.8026
28	-0.0610	-0.2829	-0.0571	0.9281	1.0002	0.8965	0.9891	-0.1037	0.8868	0.8720	-0.2565	0.6463
29	0.0468	-0.2085	0.0457	1.0385	1.0554	1.0217	1.1541	-0.0216	1.0146	1.3559	-0.2962	0.9869
30	-0.0928	-0.3137	-0.0850	0.9592	1.0384	0.9066	1.0092	-0.0420	0.8788	1.0390	-0.2298	0.7176
M	-0.0040	-0.2435	-0.0023	1.0115	1.0506	0.9802	1.0513	-0.0325	0.9317	1.1040	-0.3466	0.7942

dadas na Equação (4.3), em que $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = -0.4$, $\gamma_2 = -0.2$.

A Tabela 4.5 mostra o comportamento no caso em que variável tem distribuição Poisson dupla, mostrando para todos os tipos de resíduos analisados, um afastamento muito forte da distribuição normal padrão. Uma crítica ao modelo em que a variável de interesse tem distribuição Poisson duplo é feita por Hilbe (2011) (para outras discussões ver Cameron e Trivedi (2013)), em que discute o fato de que a média e a variância da distribuição são uma aproximação e, também, ao estimador proposto em Efron (1986) para a constante de normalização onde este apresenta algumas dificuldades. Desta forma, supor esta modelagem traz algumas consequências, como a não convergência nos métodos de otimização usados para estimar os parâmetros do modelo. Assim, recomenda-se usar outros modelos para analisar dados de contagem com sobredispersão.

Tabela 4.5: Médias, erros-padrão, assimetria e curtoses empírica dos resíduos ordinário padronizado r_p^β , deviance $d_\mu(\hat{\mu}_i; \hat{\phi}_i)$ e combinado $r_p^{\beta\gamma}$. Modelo: $y_i \sim \text{PD}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = -0.4$, $\gamma_2 = -0.2$, $x_i \sim \text{U}(0.1; 1.1)$, $z_i \sim \text{U}(0.4; 1.4)$, $i = 1, \dots, 30$.

i	Média			Erro-padrão			Assimetria			Curtose		
	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^β	$d_\mu(\hat{\mu}_i; \hat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0233	-0.1658	1.4118	3.3167	3.5686	2.6543	0.5352	-0.1450	0.8083	0.8123	0.4426	1.0319
2	-0.1037	-0.3751	0.8548	3.8579	4.3109	2.3268	0.8648	-0.0387	1.2593	1.7674	0.6244	2.1948
3	0.0615	-0.1496	1.3403	3.5216	3.7657	2.7926	0.6165	-0.0839	0.8689	0.8094	0.3001	1.0851
4	0.0463	-0.1527	1.4415	3.9185	4.1153	3.0228	0.9254	-0.0206	1.2355	3.3578	1.8486	3.0421
5	-0.0016	-0.2054	1.6672	4.3049	4.6379	3.3868	0.5709	-0.1810	0.9803	1.5157	1.0457	1.8564
6	0.0635	-0.1361	1.4354	3.7256	3.9687	2.9319	0.6426	-0.1530	1.0031	1.4686	0.9935	1.6238
7	-0.0933	-0.3979	0.9768	4.9688	5.2090	3.0480	2.1098	0.3302	1.7778	21.2966	3.3476	8.8820
8	-0.0647	-0.3235	0.8992	3.4578	3.7909	2.3544	0.8915	0.0818	1.1611	1.6709	0.1812	2.1290
9	-0.3374	-0.5158	1.6586	4.1449	4.7605	2.7663	0.1114	-0.8286	1.0090	3.9539	4.7466	2.9466
10	-0.0323	-0.2191	1.5739	3.6045	3.8846	2.9438	0.5688	-0.1215	0.8459	1.0055	0.6263	1.2320
11	0.0069	-0.2295	1.1594	3.6421	3.9128	2.7782	0.8287	0.0206	1.0989	1.6480	0.4310	2.0542
12	-0.0201	-0.2809	1.0686	4.7487	5.1345	2.8904	1.3168	-0.1040	1.5475	7.6618	4.5142	4.3919
13	-0.0649	-0.3008	0.9345	3.1557	3.4906	2.2213	0.7113	-0.0139	0.9614	0.7537	0.0196	1.0298
14	0.0895	-0.1376	1.2945	3.7493	3.9578	2.9200	0.7406	-0.0080	0.9939	1.3571	0.3503	1.6857
15	0.0194	-0.2279	1.7977	6.6654	7.0479	4.5443	1.2710	-0.0740	1.5437	10.9736	5.2700	5.3047
16	-0.0592	-0.3287	1.0744	4.2024	4.5966	2.9267	0.9173	0.0138	1.2347	2.3228	0.6937	2.4515
17	0.0102	-0.2943	1.1006	5.2006	5.5565	3.3215	1.1066	0.0598	1.4255	3.5479	1.2652	3.3694
18	0.0069	-0.1714	1.5834	3.4867	3.7450	2.8046	0.4870	-0.1907	0.7964	0.8338	0.6903	0.9775
19	-0.0149	-0.2235	1.5777	4.2101	4.4967	3.3186	0.7519	-0.0853	1.1014	2.1757	1.1462	2.4559
20	-0.0655	-0.2604	1.6440	5.5428	5.7653	3.4208	2.0548	0.0715	1.7183	23.1532	12.7579	7.2832
21	-0.1503	-0.4306	0.7868	3.7096	4.2006	2.1232	0.8248	0.0206	1.1638	1.3389	0.1311	1.9946
22	0.0535	-0.2025	1.4024	5.0639	5.3102	3.6818	1.1248	0.0734	1.3769	4.7469	1.8224	3.6726
23	0.0967	-0.1154	1.4131	4.2417	4.3887	3.1568	1.0820	0.0249	1.3385	4.8203	2.4179	3.5112
24	0.1074	-0.1237	1.3676	5.3277	5.1619	3.4436	2.7325	0.6202	1.9467	27.4171	9.0764	9.4828
25	0.0627	-0.1290	1.4403	3.5540	3.7567	2.8391	0.7231	-0.0904	1.0249	2.0506	1.0417	2.0577
26	0.0864	-0.1182	1.4713	4.0637	4.2597	3.1574	0.8106	-0.0747	1.1975	2.3497	1.4990	2.5252
27	0.1782	-0.1030	1.7318	7.1841	7.4618	4.8472	1.1150	-0.0332	1.4564	5.2160	3.0318	3.6466
28	-0.1599	-0.3145	1.5901	3.0414	3.3301	2.4662	0.4244	-0.1989	0.7015	0.7759	0.6846	0.7803
29	0.0484	-0.2111	1.3164	4.6417	4.9355	3.4256	0.9671	0.0090	1.2822	2.9733	1.1306	3.0575
30	-0.0717	-0.3348	0.8960	3.8000	4.1193	2.4258	1.0736	0.0992	1.3362	3.3738	0.8143	3.2316
M	-0.0093	-0.2393	1.3303	4.2684	4.5547	3.0314	0.9634	-0.0340	1.2065	4.9049	2.0982	3.0329

4.2 Envelopes simulados

Como foi mostrado na Seção 4.1, o resíduo proposto não tem distribuição normal padrão, desta forma, em particular para detectar pontos aberrantes não é conveniente utilizar os limites usuais $(-2, 2)$ nos gráficos dos resíduos contra os índices das observações, ou contra valores de covariáveis e valores preditos. A nossa proposta para solucionar esta questão é utilizar quantis empíricos dos resíduos gerados a partir dos envelopes simulados, em que, mesmo sob violação de algumas das suposições feitos sob os dados e sobre o modelo, podem representar bem o verdadeiro comportamento probabilístico dos resíduos.

Os gráficos de probabilidade com envelope simulado são uma ferramenta muito útil para avaliar a qualidade do ajuste do modelo proposto, como também para identificar possíveis observações aberrantes, isto é, resíduos com valores absolutos que excedem os limites do envelope simulado são considerados aberrantes (ATKINSON, 1985).

Tais gráficos são construídos da seguinte forma:

1. ajuste o modelo e gere uma amostra simulada de n observações independente usando o modelo ajustado como o verdadeiro processo gerador dos dados,
2. ajuste o modelo para a nova amostra gerada no passo 1, e calcule os valores absolutos dos resíduos de interesse e ordene de forma crescente,
3. repita os passos 1 e 2 no mínimo m vezes,
4. considere os n conjuntos dos m resíduos ordenados, para cada conjunto calcule as respectivas medidas, o percentil 2.5, a média e o percentil 97.5,
5. faça o gráfico desses valores e do resíduo de interesse ordenado da amostra original contra os quantis da distribuição normal padrão.

O ajuste do modelo é considerado bom se nos gráficos de probabilidade com envelopes simulados, os verdadeiros resíduos estão dentro das bandas de confiança simuladas, isto é, a distribuição do resíduo tem que ser próxima a distribuição empírica estimada.

As Figuras 4.1 e 4.2 apresentam os gráficos dos envelopes simulados. Estes foram feitos considerando a especificação correta do modelo e um tamanho de amostra de 80. As linhas nestes gráficos representam o percentil 2.5, a média e o percentil 97.5 de 100 pontos simulados para cada observação.

Foram considerados diferentes cenários para a variável resposta, e modelo sob a média e a dispersão descritos a seguir:

- (i) Modelo 1: $y_i \sim N(\mu_i; \phi_i^{-1})$ com $\mu_i = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = -1.4$, $\beta_2 = 0.8, \gamma_1 = 0.9, \gamma_2 = -1$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 80$.
- (ii) Modelo 2: $y_i \sim \text{gama}(\mu_i; \phi_1)$ com $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8, \gamma_1 = 1, \gamma_2 = 1.5$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 80$.
- (iii) Modelo 3: $y_i \sim \text{NI}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = -0.45$, $\beta_2 = 0.9, \gamma_1 = -0.2, \gamma_2 = -0.4$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 80$.
- (iv) Modelo 4: $y_i \sim \text{BN}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1.6$, $\beta_2 = 0.8, \gamma_1 = -0.3, \gamma_2 = 0.5$, $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 80$.
- (v) Modelo 5: $y_i \sim \text{PD}(\mu_i; \phi_1)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ e $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8, \gamma_1 = -0.4, \gamma_2 = -0.2$. $x_i \sim U(0.1; 1.1)$, $z_i \sim U(0.4; 1.4)$, $i = 1, \dots, 80$.

Pode-se observar que as distribuições empíricas dos resíduos ordinário e tipo desvio estão próximas das distribuições estimadas, indicando um bom ajuste. Isso já era esperado, pois assumimos a especificação correta dos modelos.

Desta forma, podemos concluir que o resíduo combinado padronizado apresenta comportamento satisfatório e competitivo em todos os cenários analisados, uma vez que todos os quantis observados estão dentro dos envelopes simulados, o que sugere um bom ajuste.

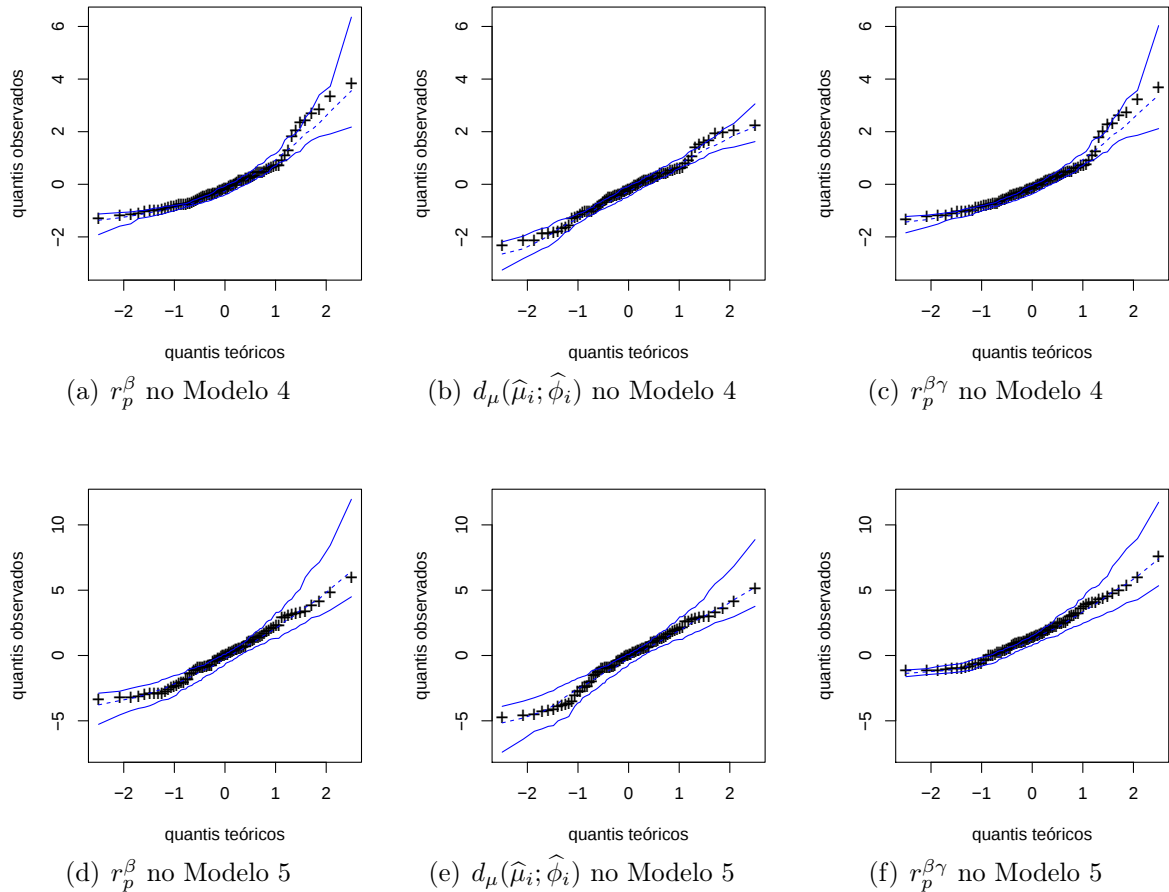


Figura 4.1: Gráficos normais de probabilidade com envelopes simulados para Modelo 4 e Modelo 5.

4.3 Modelos sob especificação incorreta e novos limites para detecção de pontos aberrantes

Um dos erros mais frequentes na modelagem de um conjunto de dados é ajustar o modelo assumindo que os dados têm dispersão constante, mas, em vários casos, o processo gerador dos dados fornece observações com dispersão variável. Uma forma de detectar este tipo de erro de especificação no ajuste, é avaliando a distribuição empírica dos resíduos do modelo usando gráficos normais de probabilidade. Em geral, a distribuição empírica dos resíduos é avaliada quando os modelos estão corretamente especificados. No entanto, se

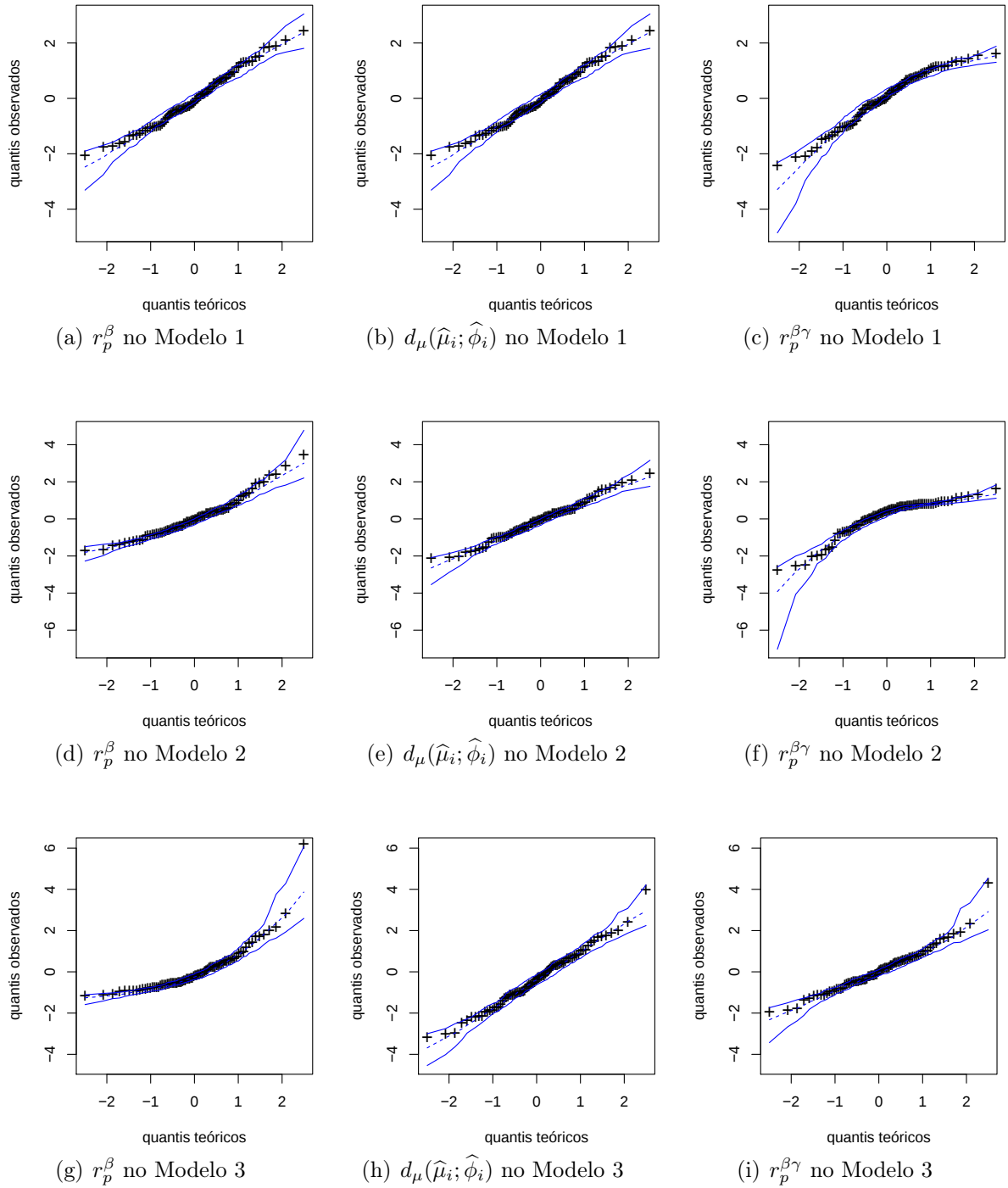


Figura 4.2: Gráficos normais de probabilidade com envelopes simulados para Modelo 1, Modelo 2 e Modelo 3.

existir alguma violação das suposições impostas sob o modelo postulado, as distribuições desses resíduos podem mudar de forma drástica.

Com base nas seções anteriores notamos que as distribuições dos resíduos podem não ser bem aproximadas pela distribuição normal padrão. Mesmo quando há fortes indícios que os resíduos investigados são aproximadamente normais, há de se ter cautela. Por exemplo, se para detectar observações aberrantes usamos os limites convencionais da distribuição normal padrão $(-2, 2)$, podemos estar cometendo um erro, pois como apresentamos na Seção 4.1, a distribuição empírica dos resíduos tem valores da assimetria e da curtose afastados dos valores da distribuição normal padrão. Desta forma, propomos que os limites sejam obtidos a partir dos envelopes simulados. Logo, para a construção dos gráficos de resíduos versus índices das observações estabelecemos como bandas de corte $(\omega_{0.025}; \omega_{0.975})$, que são os quantis 2.5% e 97.5% da distribuição empírica dos resíduos gerados para a obtenção do envelope. Esta etapa possui um custo computacional baixo, visto que os resíduos já foram simulados.

Com o objetivo de avaliar o comportamento dos resíduos no enfoque de diagnóstico, apresentamos exemplos simulados em que a modelagem da dispersão é desconsiderada (assumimos um modelo com dispersão constante). Para o submodelo verdadeiro da dispersão usado em todos os cenários desta simulação temos que a relação funcional do parâmetro de precisão é dada por $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, em que a covariável $z_i = 1$ se $i = 1, \dots, 25$ e $z_i = 1.5$ se $i = 26, \dots, 50$. Além disso, definimos o grau de heterogeneidade da dispersão como

$$\lambda = \frac{\max\{\phi_1, \dots, \phi_n\}}{\min\{\phi_1, \dots, \phi_n\}}.$$

Note que λ depende dos valores verdadeiros dos parâmetros γ_1 e γ_2 . Para os modelos em que a resposta tenha uma distribuição contínua (distribuição normal, gama ou normal inversa, ver tabela 2.1), consideramos que $\gamma_1 = 1$ e $\gamma_2 = 2$; implicando $\lambda \approx 20$, se $\gamma_2 = 2.46$; implicando $\lambda \approx 40$ e por último, $\gamma_2 = 3.07$; implicando $\lambda \approx 100$. No caso em que o modelo tenha resposta com distribuição Binomial negativa considera-se $\gamma_1 = 1$ e $\gamma_2 = 3.07$; implicando $\lambda \approx 100$, se $\gamma_2 = 3.54$; implicando $\lambda \approx 200$, se $\gamma_2 = 4$; implicando $\lambda \approx 400$. No entanto, para o processo de estimação dos parâmetros desconsidera-se a dispersão variável, isto é, só é feita a modelagem da média da variável resposta.

Desta forma, tem-se:

- (i) Modelo 1: $y_i \sim N(\mu_i; \phi_i^{-1})$ com $\mu_i = \beta_1 + x_i^{\beta_2}$, $\beta_1 = -1.4$, $\beta_2 = 0.8$, $x_i \sim U(0.1; 1.1)$, $i = 1, \dots, 50$.
- (ii) Modelo 2: $y_i \sim \text{gama}(\mu_i; \phi_i)$ com $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ e $\beta_1 = 1$, $\beta_2 = 0.8$, $x_i \sim U(0.1; 1.1)$, $i = 1, \dots, 50$.
- (iii) Modelo 3: $y_i \sim \text{NI}(\mu_i; \phi_i)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ com $\beta_1 = -0.45$, $\beta_2 = 0.9$, $x_i \sim U(0.1; 1.1)$, $i = 1, \dots, 50$.

- (iv) Modelo 4: $y_i \sim \text{BN}(\mu_i; \phi_i)$ com $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ com $\beta_1 = 1.6$, $\beta_2 = 0.8$, $x_i \sim U(0.1; 1.1)$, $i = 1, \dots, 50$.

Os gráficos em 4.3 apresentam o resultados quando a variável resposta tem distribuição normal. Pode-se observar que todos os tipos de resíduos detectam a especificação incorreta do modelo, sendo isto mais evidente à medida que o grau de heterogeneidade λ aumenta. Na Figura 4.4 notamos que nas primeiras observações a amplitude dos resíduos é maior, e torna-se menor à medida em que os índices das observações aumentam. Note que o resíduo combinado não detecta as observações 7 e 8 como aberrantes, contrastando com os outros resíduos que detectam esta observação, no entanto, esta observação está muito próxima do limite estabelecido.

Nas Figuras 4.5, 4.7 e 4.9, mostra-se o comportamento dos resíduos quando a variável resposta do modelo é gama, normal inversa e binomial negativa, respectivamente. Pode-se observar que todos os resíduos detectam a especificação incorreta do modelo, pois os resíduos observados caem fora dos envelope simulados e sendo mais evidente este comportamento quando o grau de heterogeneidade λ aumenta. Porém, no modelo com resposta binomial negativa, a detecção da especificação incorreta é mais sutil. Note à medida que λ aumenta, os gráficos de probabilidade têm poucas diferenças. Isto pode ser explicado pela robustez no processo de estimação dos parâmetros à medida que variam os valores verdadeiros do parâmetro de dispersão.

A avaliação dos resíduos na detecção de observações atípicas é feita nas Figuras 4.6, 4.8 e 4.10, e apresentam a distribuição dos resíduos contra os índices das observações quando a variável de interesse é gama, normal inversa e binomial negativa respectivamente. Observa-se um padrão de comportamento em todos eles, mostrando primeiro uma especificação incorreta do modelo, já que é considerado um ajuste bom quando os resíduos são distribuídos de forma aleatória. Também é importante ressaltar que o resíduo combinado destaca com maior força as observações consideradas aberrantes, isto, devido ao fato de que esses pontos estão mais afastados das bandas de corte estabelecidas. Além disso, observações que não são consideradas atípicas usando os resíduos ordinários e resíduo deviance por estarem próximas as bandas de corte, são detectadas pelo resíduo proposto. Exemplo disso, é a observação 20 na Figura 4.8, que é detectada em todos os cenários analisados para λ .

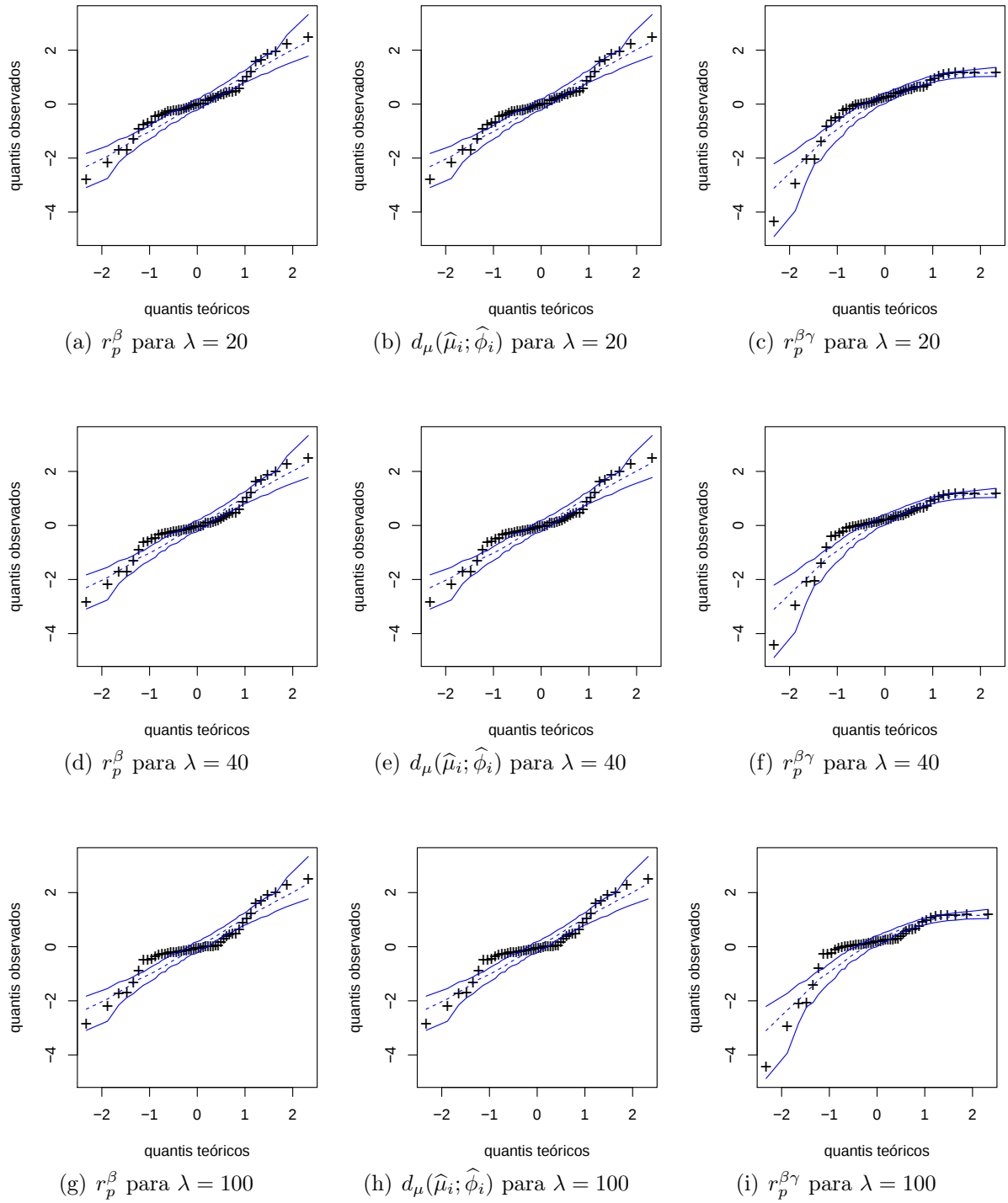


Figura 4.3: Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 1.

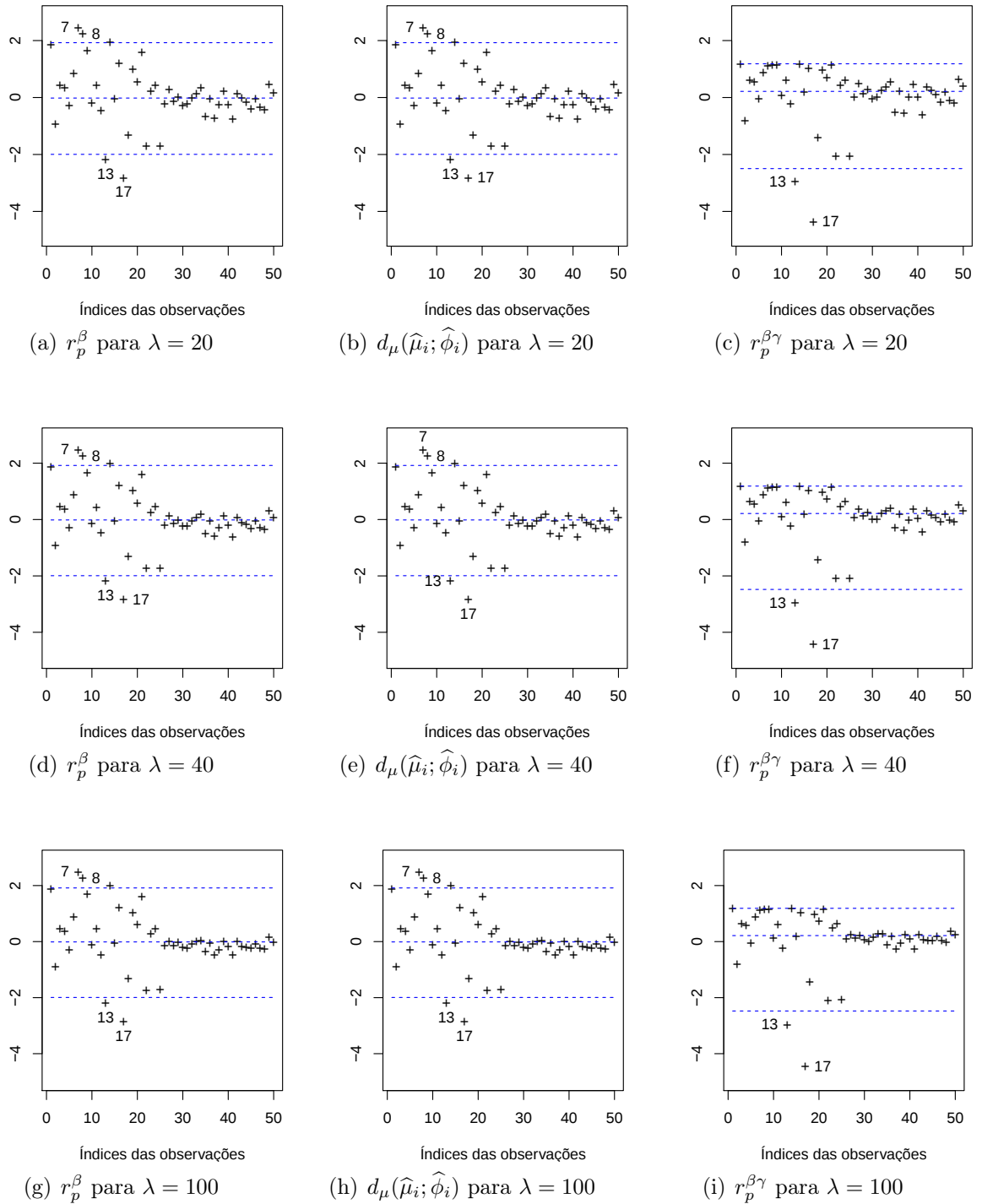


Figura 4.4: Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 1.

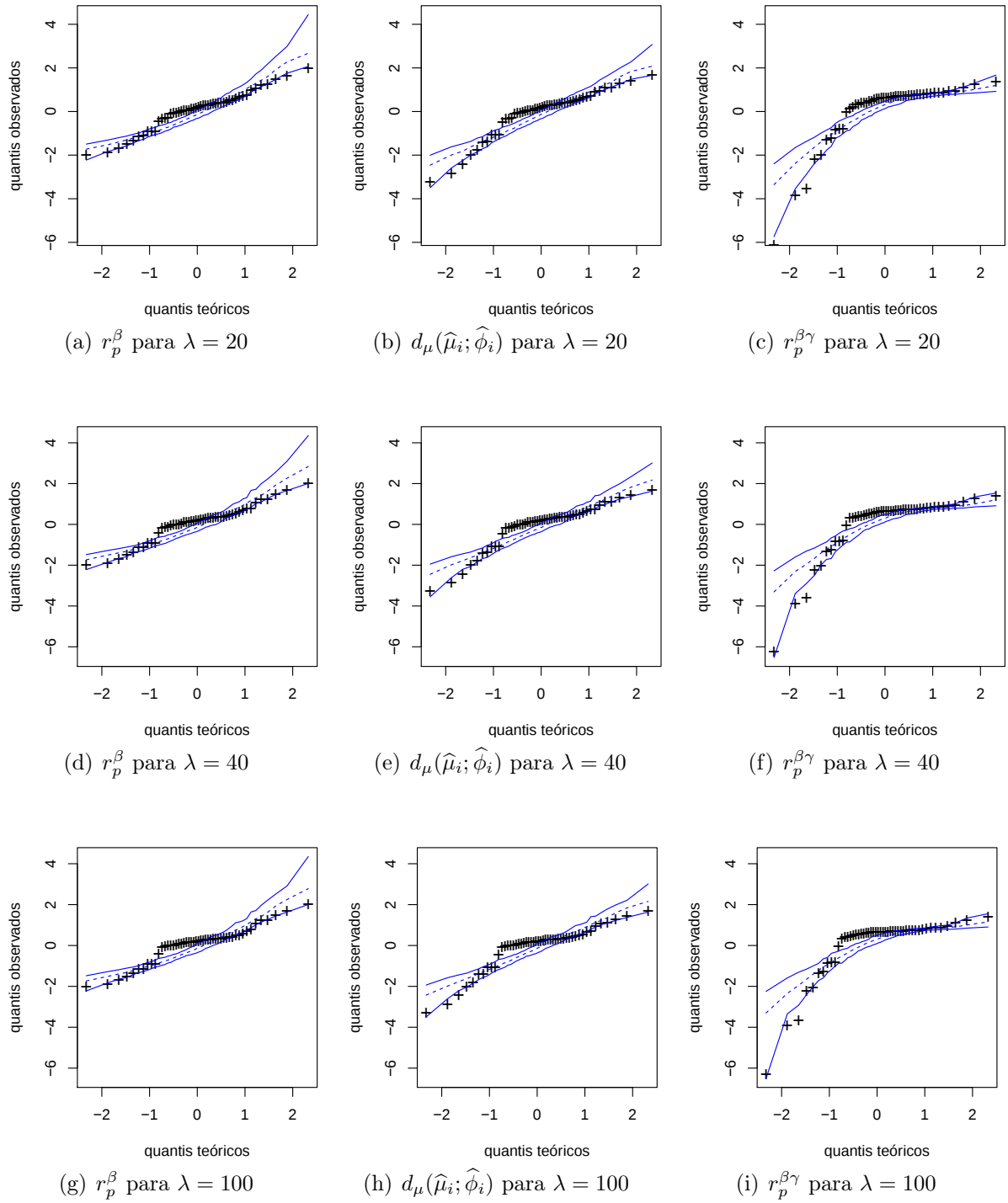


Figura 4.5: Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 2.

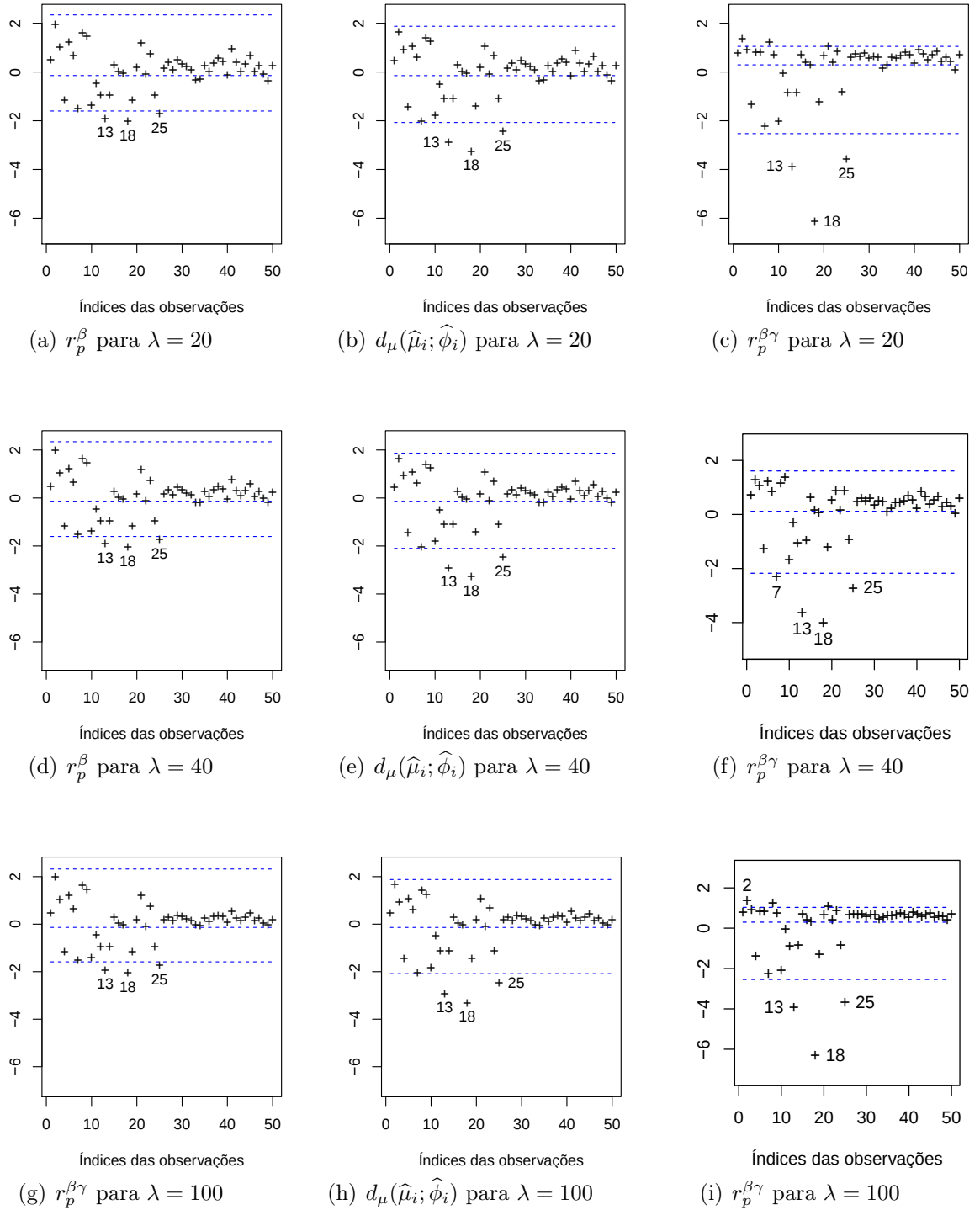


Figura 4.6: Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 2.

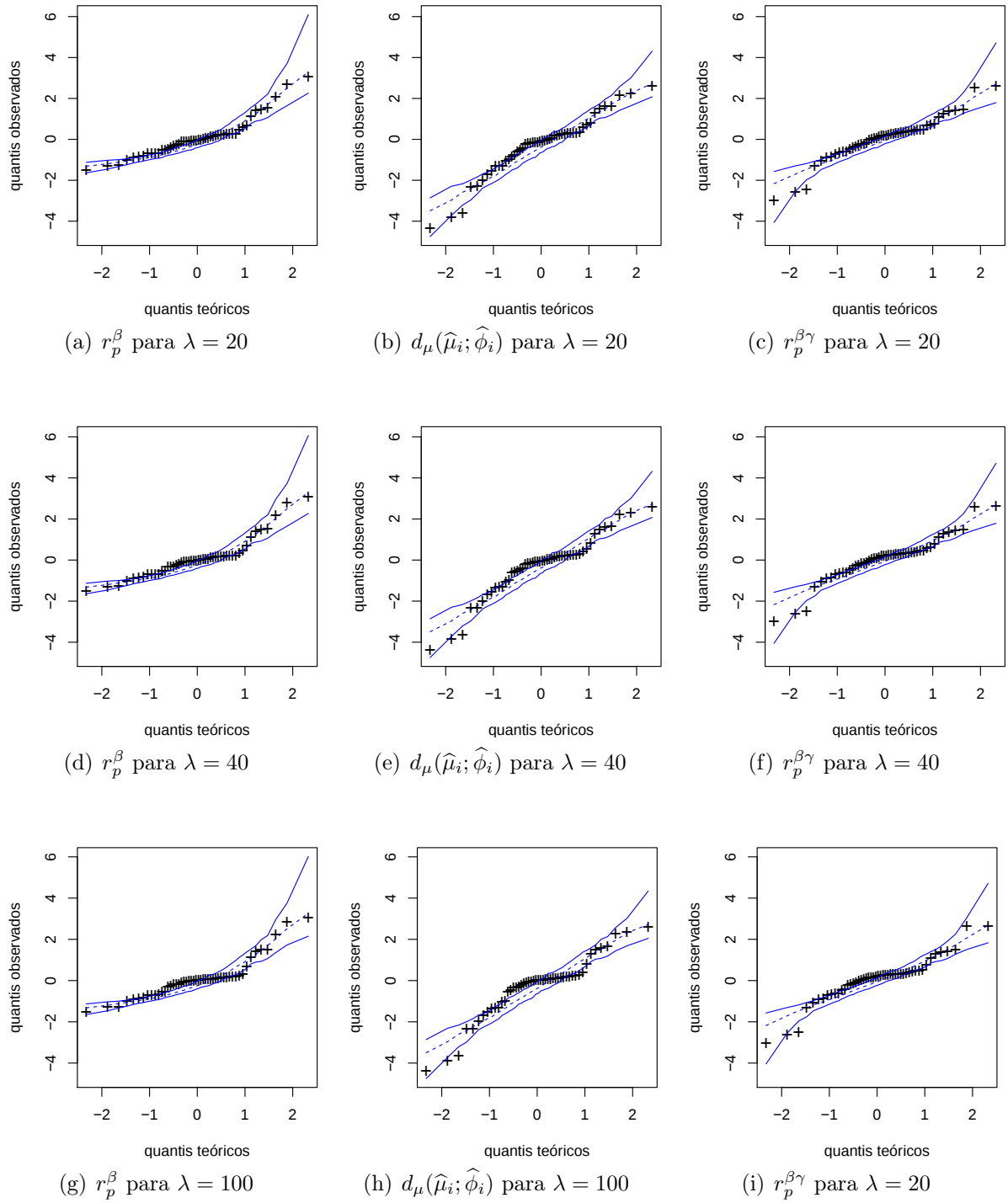


Figura 4.7: Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 3.

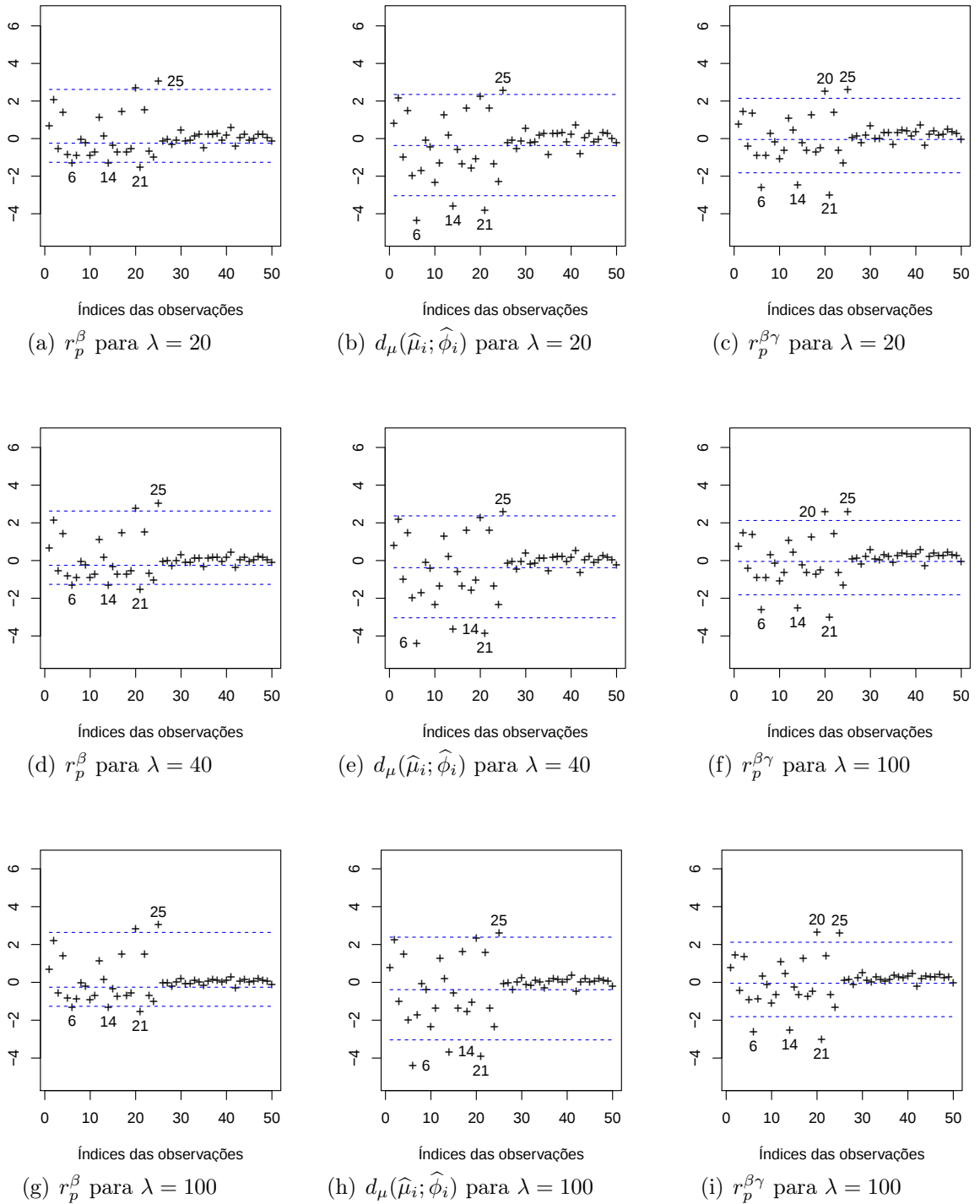


Figura 4.8: Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 3.

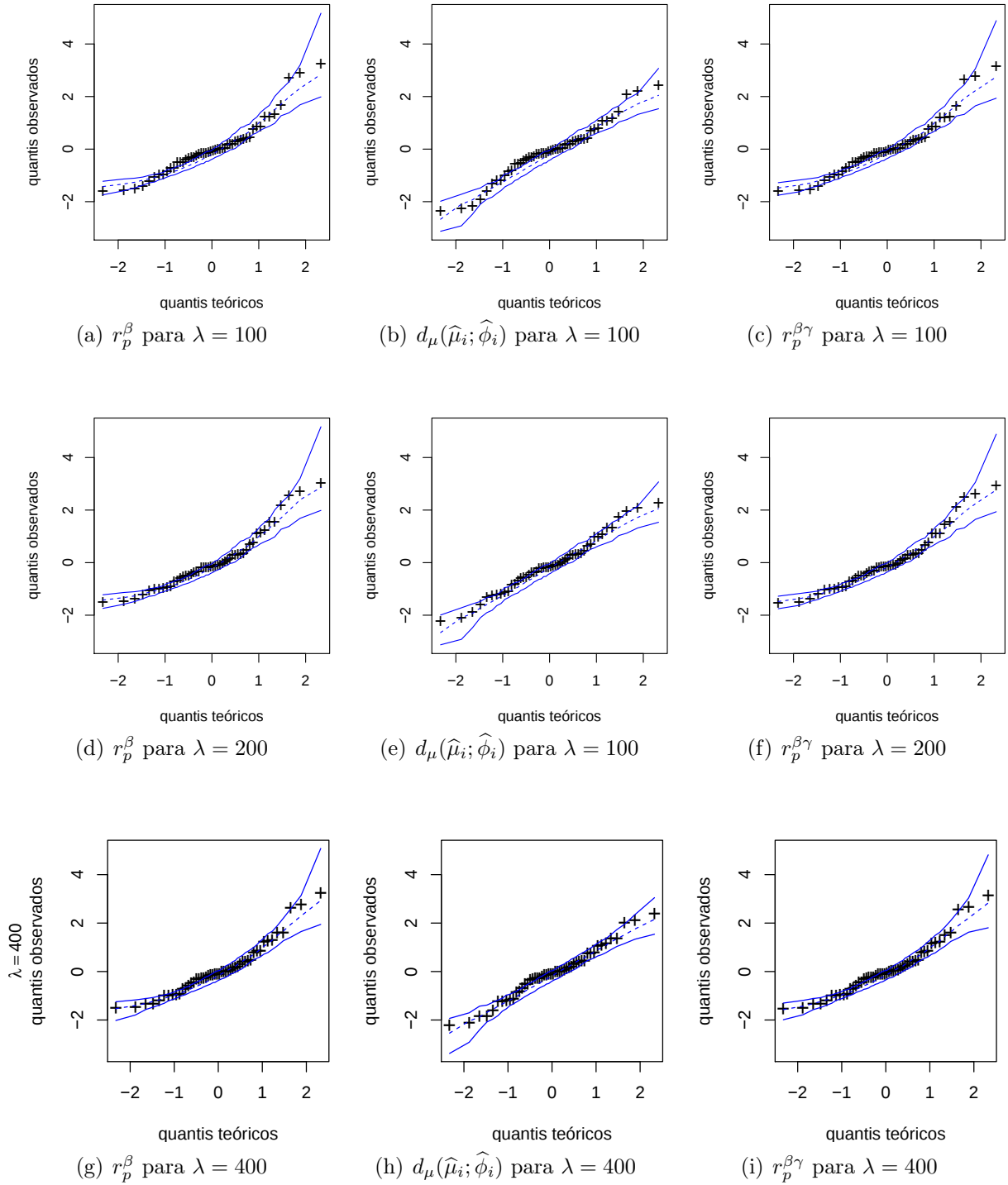


Figura 4.9: Gráficos normais de probabilidade dos resíduos ordinário padronizado, resíduo deviance e combinado sob especificação incorreta para o Modelo 4.

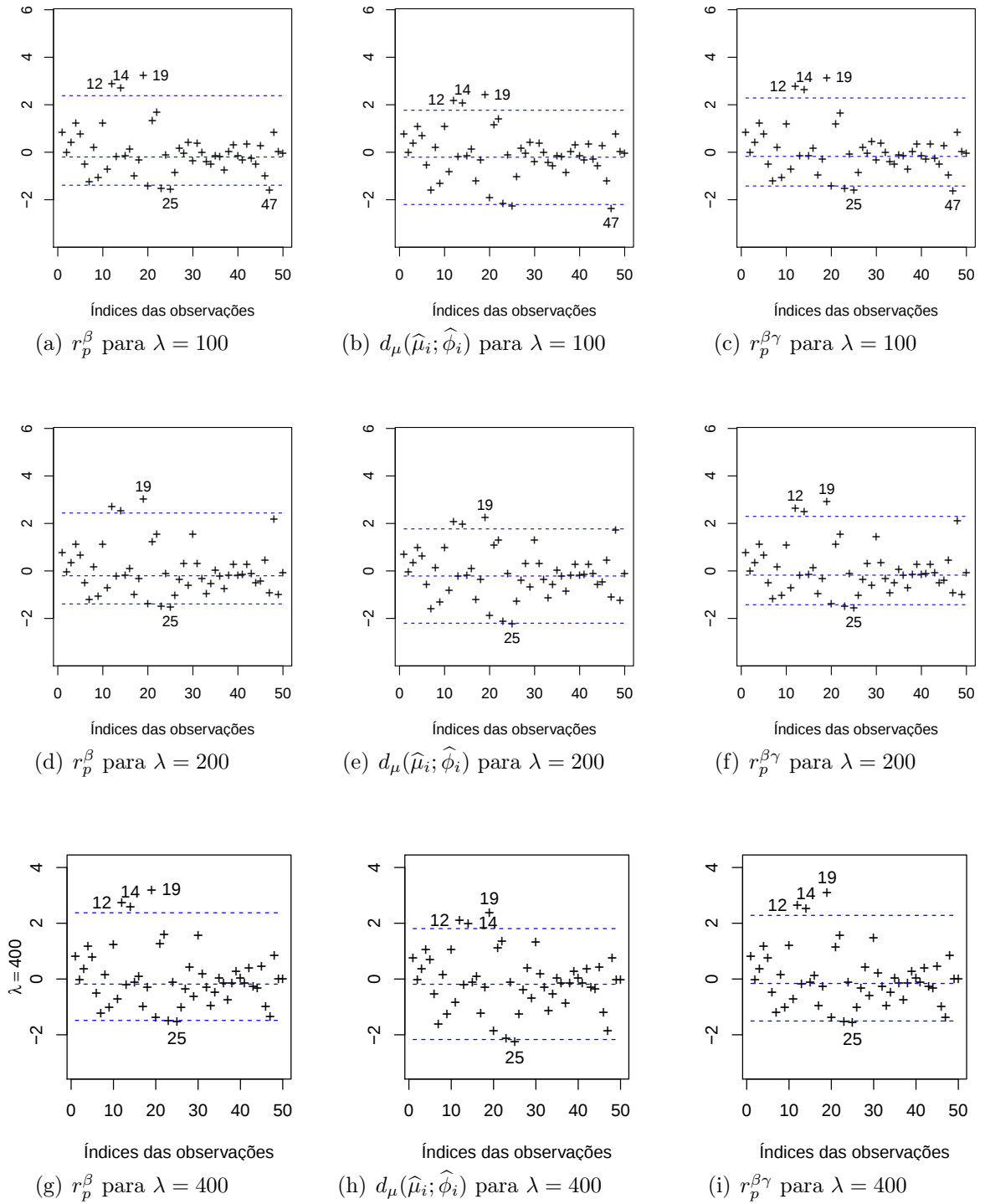


Figura 4.10: Gráficos de resíduos contra índices das observações sob especificação incorreta para o Modelo 4.

É muito interessante notar que os novos limites para a detecção de pontos aberrantes são, em muitas situações, bastante diferentes dos usuais $(-2, 2)$. Por exemplo, considerado o resíduo deviance, para o modelo em que a variável resposta tem distribuição normal inversa, sob especificação incorreta da dispersão, nota-se que os limites empíricos podem ser $(-3, 2)$, (Figura 4.8), já para o modelo em que variável resposta tem distribuição binomial negativa o limite superior é menor do que 2 (Figura 4.10).

4.4 Observações aberrantes

Para avaliar a capacidade dos resíduos de mensurar corretamente a discrepância entre as i -ésimas respostas observadas e as estimadas dado que essa resposta é atípica ou até um possível erro, multiplicamos uma observação por um fator de perturbação positivo. Para isso consideramos amostras de tamanho 50. Os modelos analisados foram apresentados na Seção 4.2 e assumimos especificação correta dos modelos. Os valores das covariadas são gerados através da distribuição uniforme de forma independente com: $x_i \sim U(0.1; 1, 1)$ e $z_i \sim U(0.4; 1, 4)$.

A observação 31 é a escolhida para ser perturbada, logo, espera-se que os gráficos de resíduo versus os índices das observações destaquem este ponto. A Figura 4.11 mostra o comportamento do resíduos na detecção de observações aberrantes quando a variável resposta do modelo tem distribuição continua (normal, gama e normal inversa), e pode-se observar um desempenho favorável do resíduo combinado, no sentido que o mesmo consegue atribuir uma alta discrepância entre a resposta perturbada e a resposta estimada em todos os cenários. O contrário ocorre com o resíduo deviance, que não consegue destacar o ponto 31 quando a variável resposta do modelo tem distribuição normal inversa.

A Figura 4.12 apresenta o desempenho para os três tipos de resíduos sob estudo, para os casos em que as variáveis respostas tem distribuição binomial negativa e Poisson dupla respectivamente. Pode-se observar que o resíduo combinado consegue destacar a observação 31, apresentando o pior comportamento quando a variável resposta é Poisson duplo, devido a que o ponto esta mais próximo das bandas de confiança. Porém, o comportamento quando a variável resposta é binomial negativa é satisfatório e contrasta com o resíduo deviance em que o destaque da observação aberrante é menor.

Deve-se ressaltar que os gráficos do resíduo ordinário padronizado também destacam corretamente a observação 31 e com ênfase similar aos gráficos do resíduo combinado padronizado. No entanto, o cálculo do resíduo ordinário padronizado depende da obtenção das matrizes de projeção de ordem $n \times n$, o que não é requerido quando se utiliza o resíduo combinado. Neste sentido, o uso do resíduo proposto apresenta a vantagem de diagnosticar corretamente problemas nos dados e no modelo, sem o alto custo de cálculo de matrizes $n \times n$; em especial quando n é significativamente grande.

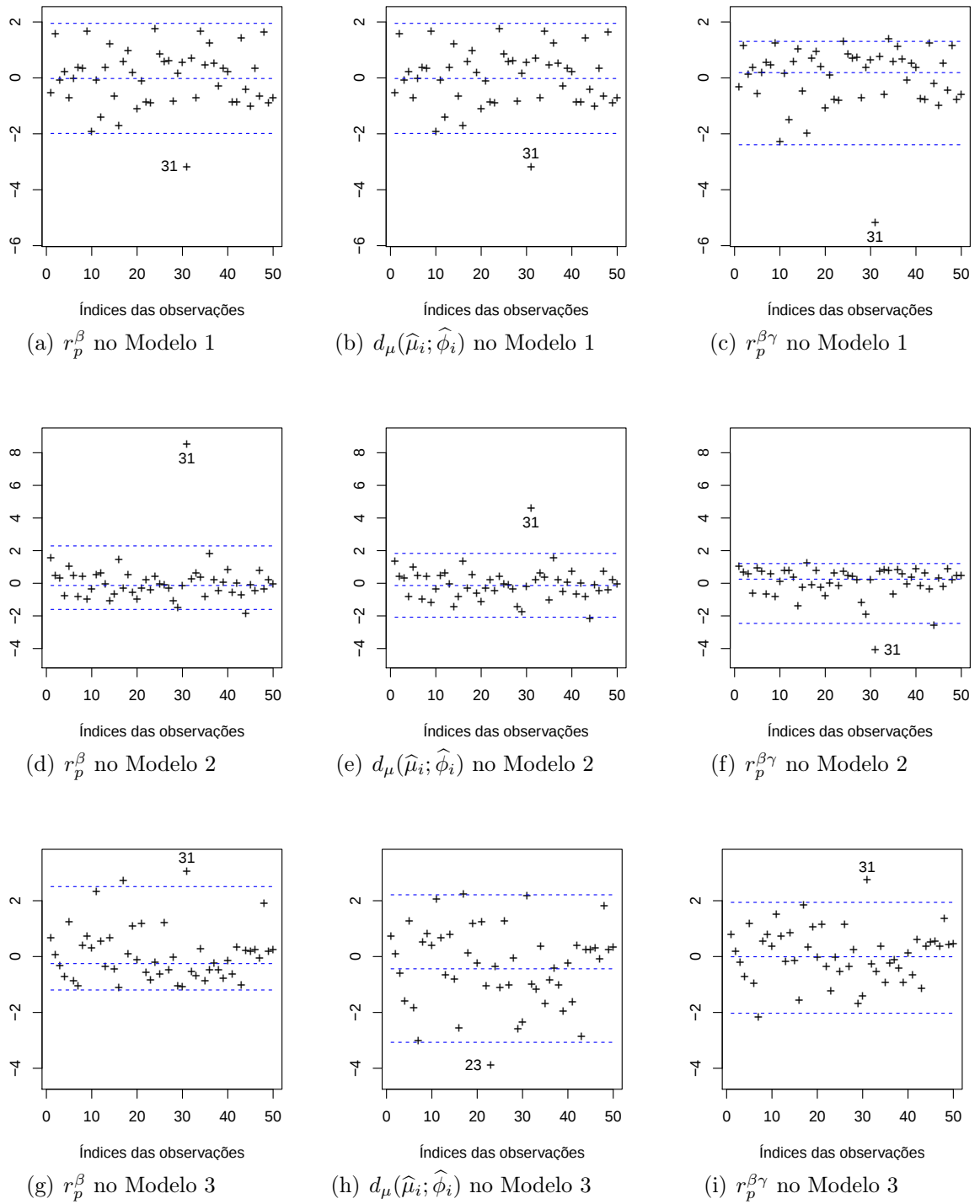


Figura 4.11: Gráficos normais de probabilidade com envelopes simulados para Modelo 1, Modelo 2 e Modelo 3.

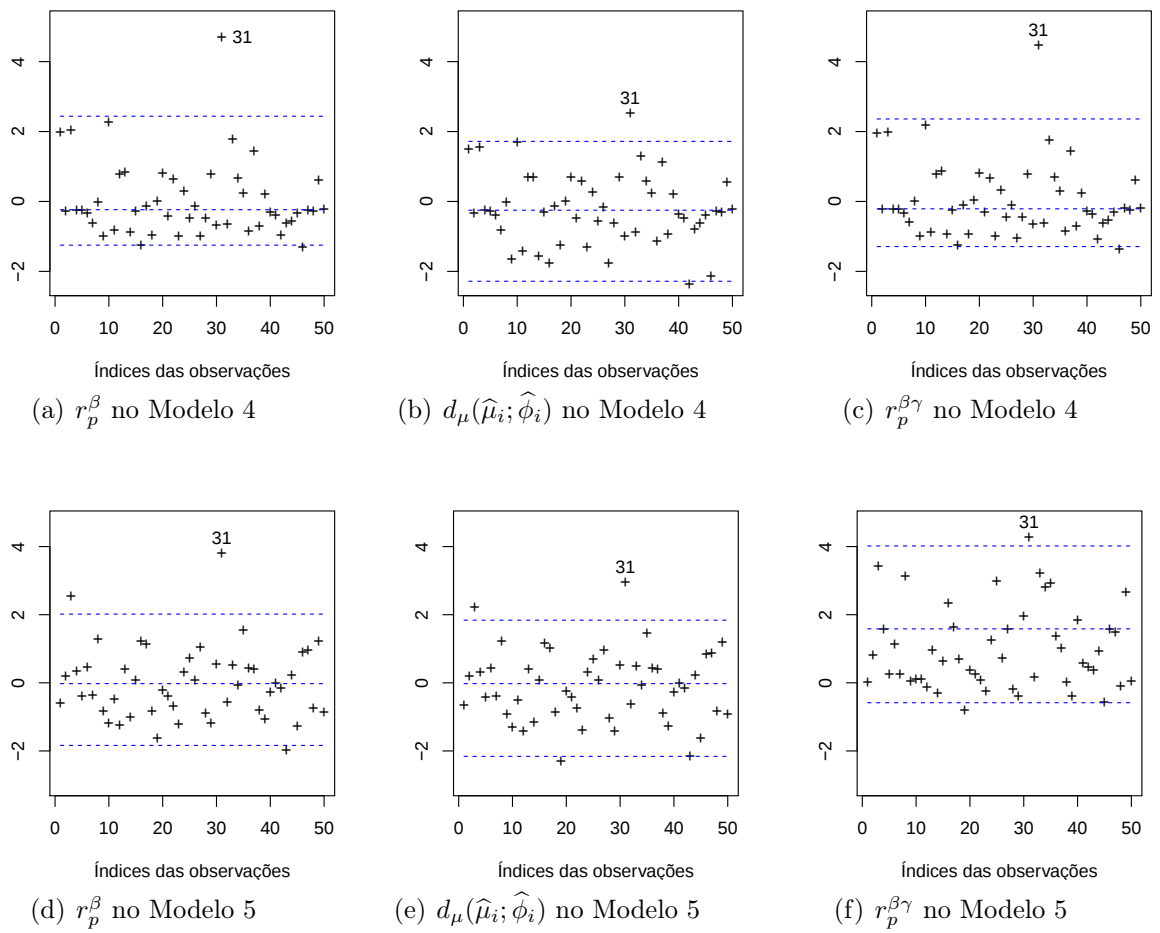


Figura 4.12: Gráficos normais de probabilidade com envelopes simulados para Modelo 4 e Modelo 5.

Aplicação

Neste capítulo serão apresentadas duas aplicações a dados reais. Foram realizadas análises de bondade de ajuste usando os gráficos de probabilidade com envelopes simulados e tendo como referência os quantis teóricos da distribuição normal padrão. Na primeira aplicação ajustamos um modelo linear generalizado duplo com resposta gama e, por último, ajustamos um modelo com resposta binomial negativa.

5.1 Aplicação 1

Os dados desta aplicação estão disponíveis em Paula (2013) e são parte de um experimento desenvolvido no Departamento de Nutrição da Faculdade de Saúde Pública da Universidade de São Paulo, em que um novo tipo de *snack* com 5 características diferentes foi comparado ao longo de 20 semanas. Para esse novo produto foi substituída a gordura vegetal hidrogenada, que é responsável pela fixação do aroma no *snack*, por óleo de canola de forma total ou parcial. As características são as seguintes:

- Característica A: 22% gordura, 0% de óleo de canola.
- Característica B: 0% gordura, 22% de óleo de canola.
- Característica C: 17% gordura, 5% de óleo de canola.
- Característica D: 11% gordura, 11% de óleo de canola.
- Característica E: 5% gordura, 17% de óleo de canola.

O experimento foi realizado de maneira que 15 embalagens são coletados nas semanas pares, para cada um dos produtos A, B, C, D e E. Logo, tem-se que a variável resposta é a força necessária para o cisalhamento do produto, e as variáveis explicativas são o grupo, que pode ser A, B, C, D ou E, e a semana da coleta. Desta forma denotamos y_{ijk} força de cisalhamento concernente à k -ésima réplica do i -ésimo grupo na j -ésima semana, para $k = 1, \dots, 15$, $j = 2, 4, 6, \dots, 20$ e $i = 1(A), 2(B), 3(C), 4(D)$ e $5(E)$. Vamos supor

uma modelagem conjunta da média e a dispersão com resposta gama em que a parte sistemática é dada por

$$\begin{aligned}\mu_{ij} &= \beta_0 + \beta_i + \beta_6 \text{semana}_j + \beta_7 \text{semana}_j^2, \\ \log(\phi_{ij}) &= \gamma_0 + \gamma_i + \gamma_6 \text{semana}_j + \gamma_7 \text{semana}_j^2.\end{aligned}$$

Note que $\beta_1 = 0$ e $\gamma_1 = 0$. Deste modo, β_0 e γ_0 são os efeitos sob a média e a dispersão produzido pela característica A, dependendo da semana, e os efeitos das características B, C, D e E são dados por $\beta_0 + \beta_i$ e $\gamma_0 + \gamma_i$ para a média e precisão, respectivamente.

Tabela 5.1: Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta gama ajustada aos dados sobre *snack*.

Efeito	Média			Dispersão		
	Estimativa	E/E.padrão	p-valor	Estimativa	E/E.padrão	p-valor
Constante	36.990	3.208	0.00	1.560	0.214	0.00
Grupo B	-10.783	1.684	0.00	0.477	0.161	0.00
Grupo C	-3.478	1.763	0.00	0.050	0.160	0.07
Grupo D	-14.829	1.614	0.00	0.815	0.161	0.00
Grupo E	-15.198	1.622	0.00	0.817	0.161	0.00
semana	5.198	0.525	0.00	0.155	0.039	0.00
semana ²	-0.189	0.212	0.00	-0.005	0.001	0.00

A partir das estimativas dos parâmetros da média dadas na Tabela 5.1, pode-se observar que o grupo A tem a maior medida para a força de cisalhamento contrastando com os grupos D e E que tem as menores medidas. Note também que o grupo A tem a maior variabilidade no sentido de coeficiente de variação. Pelos gráficos normais de probabilidade apresentados na Figura 5.1, pode-se concluir que não existem indícios de inadequação do ajuste proposto. O resíduo combinado padronizado proposto neste trabalho, contém informações da modelagem da média e da dispersão de forma conjunta, o que facilita as análises de diagnostico. A Figura 5.2 apresenta o gráfico de probabilidade com envelopes simulados. Desta forma, pode-se observar simultaneamente que não existem indícios de especificação incorreta do modelo.

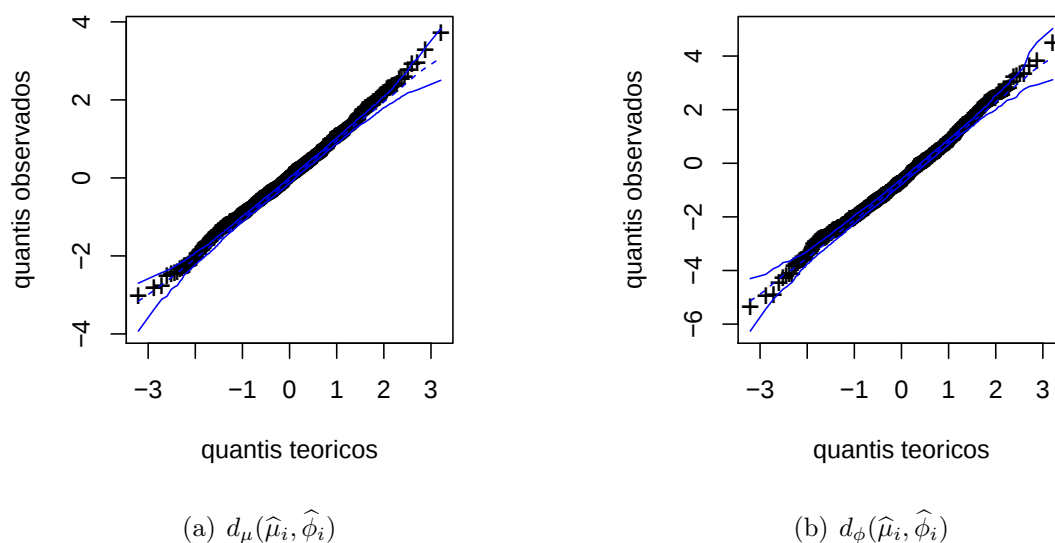


Figura 5.1: Gráficos normais de probabilidade referentes à modelagem conjunta da média e da dispersão supondo resposta gama ajustada aos dados sobre *snack* com: (a) resíduo componente desvio para a média e (b) resíduo componente desvio para a dispersão.

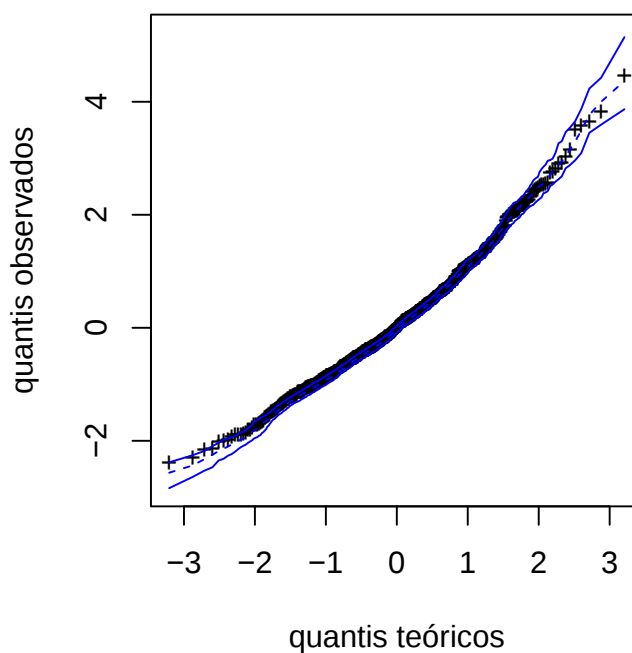


Figura 5.2: Gráfico de probabilidade do resíduo combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta gama ajustada aos dados sobre *snack*.

5.2 Aplicação 2

Para observar o comportamento do resíduo combinado padronizado em que a variável resposta do modelo é discreta, usamos um conjunto de dados que pode ser encontrado em Ridout et al. (2001). Os dados foram obtidos de um cultivo de maçãs e fazem referência ao número de raízes produzidas por 270 brotos micropropagados no cultivo. Os brotos foram produzidos sob condições de 8 ou 16 horas de exposição a luz (denominado fotoperíodo), e 5 tipos diferentes de concentração de citocinina BAP. Logo define-se as variáveis do modelo como:

- y_i : número de raízes do i -ésimo broto ;
- x_{i1} : fotoperíodo (0 = 8 horas, 1 = 16 horas);
- Grupo 1: concentração de citocinina BAP de 2.2;
- Grupo 2: concentração de citocinina BAP de 4.4;
- Grupo 3: concentração de citocinina BAP de 8.8;
- Grupo 4: concentração de citocinina BAP de 17.6;
- Grupo 5: concentração de citocinina BAP de 18.6;

Vamos supor uma modelagem conjunta da média e da dispersão do modelo com resposta binomial negativa e componente sistemática dada por

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_i \quad \text{e}$$

$$\log(\phi_{ij}) = \gamma_0 + \gamma_1 x_{i1} + \gamma_i.$$

Note que $\beta_2 = 0$ e $\gamma_2 = 0$. Desta forma, β_0 e γ_0 são os efeitos do grupo 1 sob a média e a dispersão respectivamente, dependendo do fotoperíodo. Assim, para os demais grupos tem-se que o efeito é dado por $\beta_0 + \beta_i$ e $\gamma_0 + \gamma_i$, para da média e da dispersão, respectivamente.

A partir das estimativas dos parâmetros da média dadas na Tabela 5.2, pode-se observar que o grupo 1 têm o maior número de raízes nos brotos, contrastando com os grupos 4 e 5 que tem as menores contagens de raízes dos brotos. Note também que o grupo 1 tem a maior variabilidade em relação ao coeficiente de variação. Pelo gráfico de probabilidade para o resíduo combinado padronizado apresentado na Figura 5.3, pode-se concluir que existe forte indício de inadequação do ajuste proposto.

Tabela 5.2: Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta binomial negativa ajustada aos dados sobre *maças*.

Efeito	Média			Dispersão		
	Estimativa	E/E.padrão	p-valor	Estimativa	E/E.padrão	p-valor
Constante	1.78	0.091	0.00	2.34	0.527	0.00
fotoperíodo	-0.92	0.151	0.00	-3.79	0.463	0.00
Grupo 2	0.26	0.119	0.03	0.46	0.485	0.34
Grupo 3	0.23	0.110	0.03	0.92	0.491	0.06
Grupo 4	0.17	0.112	0.12	0.69	0.484	0.15
Grupo 5	-0.87	1.013	0.16	10.08	105.6	0.92

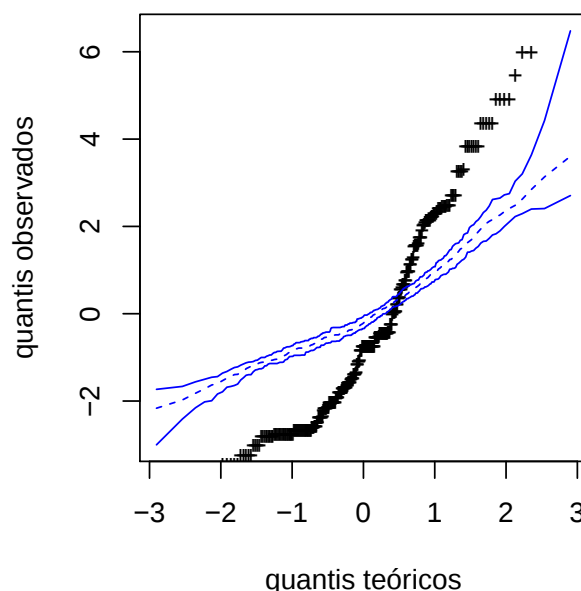


Figura 5.3: Gráfico de probabilidade do residual combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta binomial negativa ajustada aos dados sobre *maças*.

Consequentemente, a nova proposta para a parte sistemática do modelo é dada por

$$\begin{aligned}\log(\mu_i) &= \beta_0 + \beta_1 x_{i1} \quad \text{e} \\ \log(\phi_{ij}) &= \gamma_0 + \gamma_1 x_{i1},\end{aligned}\tag{5.1}$$

e só levamos em consideração o número de horas de exposição a luz (fotoperíodo). Na Tabela 5.3 são apresentadas as estimativas de máxima verossimilhança referentes ao modelagem conjunto da média e da dispersão dado na Equação (5.1).

Tabela 5.3: Estimativas de máxima verossimilhança para os parâmetros referentes à modelagem conjunta da média e da dispersão com resposta binomial negativa ajustada aos dados sobre *maçãs*.

Efeito	Média			Dispersão		
	Estimativa	E/E.padrão	p-valor	Estimativa	E/E.padrão	p-valor
Constante	1.96	0.419	0.00	2.77	0.038	0.00
fotoperíodo	-0.91	0.458	0.00	-3.68	0.152	0.00

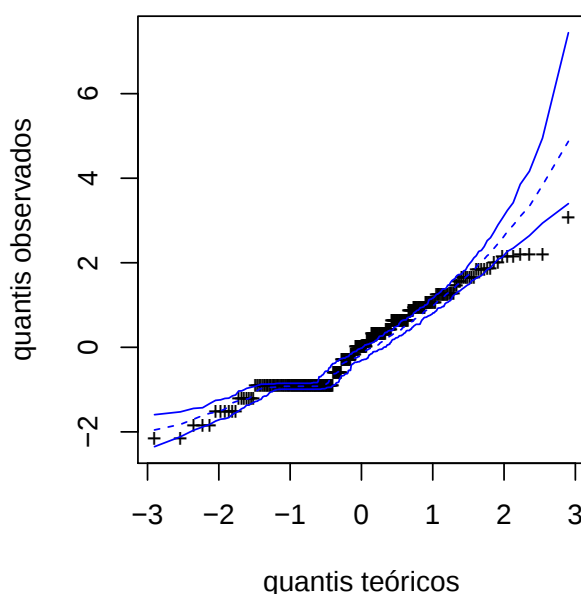


Figura 5.4: Gráfico de probabilidade do residual combinado padronizado referente à modelagem conjunta da média e da dispersão supondo resposta binomial negativa ajustada aos dados sobre *maçãs*.

Por outro lado, o gráfico de probabilidade apresentado na Figura 5.4 mostra que alguns dos pontos nas caldas da distribuição estão fora dos envelopes, contrastando com a Figura 5.3, em que a maioria dos pontos estão fora dos envelopes simulados. Desta forma, pode-se concluir que existem indícios de bom ajuste.

Considerações finais

Neste trabalho foi introduzido um novo resíduo para classes de modelos lineares e não lineares da família exponencial com o objetivo de avaliar sua qualidade de ajuste em diferentes cenários. Foram feitas simulações de Monte Carlo a fim de avaliar a distribuição empírica do resíduo. Estas simulações mostraram claramente que a distribuição de referência não é normal padrão.

O resíduo combinado padronizado tem várias vantagens em relação a seus competidores na literatura, uma delas sendo sua construção, visto que, além de considerar o processo iterativo scoring de Fisher para obter as estimativas de β , da mesma forma leva em consideração o processo iterativo Scoring de Fisher para estimar γ . Isto é, o resíduo combinado padronizado é uma medida que resume as duas modelagens feitas sobre a média e a dispersão, facilitando e unificando a análise de diagnóstico, uma vez que, analisamos só um grupo de resíduos que contém de forma simultânea estas informações, e não dois grupos como é comumente feito para essas análises. Esta avaliação de apenas um grupo de resíduos é apropriada nos casos em que se quer avaliar conjuntamente a bondade de ajuste da média e a dispersão.

Uma outra vantagem do resíduo proposto é que não é necessário obter as matrizes de projeção dos modelos da média e da dispersão para calcular a padronização dos resíduos. Isto poderia nos trazer um menor custo computacional para realizar as análises de diagnóstico. Também é importante destacar que o resíduo proposto apresenta comportamentos similares ao resíduo tipo desvio, que é usado com maior frequência em aplicações práticas.

O estudo de simulações na Seção 4.4 mostra claramente o quanto o resíduo combinado padronizado consegue identificar as observações atípicas quando a variável resposta é perturbada. Do mesmo modo, quando o modelo não está corretamente especificado, o resíduo combinado tem desempenho satisfatório, mostrando no gráfico de probabilidade ou no gráfico de resíduos versus índices das observações evidências de especificação incorreta.

Outra vantagem do resíduo combinado padronizado, é que só depende das estimativas dos parâmetros do modelo, essas podem ser obtidas por algum método de maximização não linear, por exemplo SANN ¹ ou algoritmos genéticos. Desta forma, o resíduo combinado

¹para maiores detalhes ver (BÉLISLE, 1992)

só depende das estimativas e não do método de maximização. Assim, temos uma boa alternativa para a análises de diagnóstico nas situações em que não é possível obter a matriz de projeção.

Referências

- ATKINSON, A. C. *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Oxford: Statistical Science Series, 1985.
- BÉLISLE, C. J. Convergence theorems for a class of simulated annealing algorithms on rd. *Journal of Applied Probability*, p. 885–895, 1992.
- CAMERON, A. C.; TRIVEDI, P. K. *Regression analysis of count data*. Cambridge: Cambridge university press, 2013. v. 53.
- CORDEIRO, G. M. Modelos lineares generalizados. *VII Simpósio Nacional de Probabilidade e Estatística.*, 1986.
- EFRON, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, v. 81, p. 709–721, 1986.
- HILBE, J. M. *Negative Binomial Regression*. 2.. ed. Cambridge: Cambridge University Press, 2011.
- LAWLESS, J. F. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, v. 15, n. 3, p. 209–225, 1987.
- LEHMANN, E.; CASELLA, G. *Theory of Point Estimation*. 2.. ed. New York: Springer, 1998.
- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. 2.. ed. London: Chapman and Hall, 1989. v. 37.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society Series A (General)*, Blackwell Publishing, v. 135, 1972.
- PAULA, G. A. *Modelos de regressão: com apoio computacional*. São Paulo: IME-USP, 2013.
- RIDOUT, M.; HINDE, J.; DEMÉTRIO, C. G. A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, Wiley Online Library, v. 57, n. 1, p. 219–223, 2001.
- SMYTH, G. K. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, p. 47–60, 1989.
- SVETLIZA, C. F. *Modelos Não Lineares com Resposta Binomial Negativa*. Tese (Doutorado) — IME-USP São Paulo, 2002.

Métodos iterativos scoring de Fisher e resíduos ordinários

O método iterativo scoring de Fisher para β é dado por

$$\begin{aligned}
 \beta^{(m+1)} &= \beta^{(m)} + \{K_{\beta\beta}^{(m)}\}^{-1} U_{\beta}^{(m)} \\
 &= \beta^{(m)} + (X^{\top} \Phi^{(m)} W^{(m)} X)^{-1} X^{\top} \Phi^{(m)} W^{1/2(m)} V^{-1/2(m)} (y - \mu^{(m)}) \\
 &= (X^{\top} \Phi^{(m)} W^{(m)} X)^{-1} X^{\top} \Phi^{(m)} W^{(m)} X \beta^{(m)} + \\
 &\quad (X^{\top} \Phi^{(m)} W^{(m)} X)^{-1} X^{\top} \Phi^{(m)} W^{1/2(m)} V^{-1/2(m)} (y - \mu^{(m)}) \\
 &= (X^{\top} \Phi^{(m)} W^{(m)} X)^{-1} X^{\top} \Phi^{(m)} W^{(m)} \left(\eta_1^{(m)} + W^{-1/2(m)} V^{-1/2(m)} (y - \mu^{(m)}) \right) \\
 &= (X^{\top} \Phi^{(m)} W^{(m)} X)^{-1} X^{\top} \Phi^{(m)} W^{(m)} u_1,
 \end{aligned}$$

em que $u_1 = \left(\eta_1^{(m)} + W^{-1/2(m)} V^{-1/2(m)} (y - \mu^{(m)}) \right)$.

Da mesma forma, o método iterativo scoring de Fisher para γ é dado por

$$\begin{aligned}
 \gamma^{(m+1)} &= \gamma^{(m)} + \{K_{\gamma\gamma}^{(m)}\}^{-1} U_{\gamma}^{(m)} \\
 &= \gamma^{(m)} + (Z^{\top} P^{(m)} Z)^{-1} Z^{\top} H_{\gamma}^{-1(m)} (t - \mu_T^{(m)}) \\
 &= (Z^{\top} P^{(m)} Z)^{-1} Z^{\top} P^{(m)} Z \gamma^{(m)} + (Z^{\top} P^{(m)} Z)^{-1} Z^{\top} H_{\gamma}^{-1(m)} (t - \mu_T^{(m)}) \\
 &= (Z^{\top} P^{(m)} Z)^{-1} Z^{\top} P^{(m)} (\eta_2 + P^{-1(m)} H_{\gamma}^{-1(m)} (t - \mu_T^{(m)})) \\
 &= (Z^{\top} P^{(m)} Z)^{-1} Z^{\top} P^{(m)} u_2,
 \end{aligned}$$

em que $u_2 = \eta_2 + P^{-1(m)} H_{\gamma}^{-1(m)} (t - \mu_T^{(m)})$.

O resíduo ordinário para β é dado por

$$\begin{aligned}
 r^\beta &= \widehat{\Phi}^{1/2} \widehat{W}^{1/2} (u_1 - \widehat{\eta}_1) \\
 &= \widehat{\Phi}^{1/2} \widehat{W}^{1/2} (\widehat{\eta}_1 + \widehat{W}^{-1/2} \widehat{V}^{-1/2} (y - \widehat{\mu}) - \widehat{\eta}_1) \\
 &= \widehat{\Phi}^{1/2} \widehat{W}^{1/2} \widehat{W}^{-1/2} \widehat{V}^{-1/2} (y - \widehat{\mu}) \\
 &= \widehat{\Phi}^{1/2} \widehat{V}^{-1/2} (y - \widehat{\mu}).
 \end{aligned}$$

Logo, o resíduo para a i -ésima observação é

$$r_i^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\phi_i^{-1} V_i}}, \quad i = 1, \dots, n.$$

O resíduo ordinário para γ é dado por

$$\begin{aligned}
 r^\gamma &= \widehat{P}^{1/2} (u_2 - \widehat{\eta}_2) \\
 &= \widehat{P}^{1/2} (\widehat{\eta}_2 + \widehat{P}^{-1} \widehat{H}_\gamma^{-1} (t - \widehat{\mu}^*) - \widehat{\eta}_2) \\
 &= \widehat{P}^{1/2} \widehat{P}^{-1} \widehat{H}_\gamma^{-1} (t - \widehat{\mu}^*) \\
 &= \widehat{P}^{-1/2} \widehat{H}_\gamma^{-1} (t - \widehat{\mu}^*)
 \end{aligned}$$

Como $P = V_\gamma H_\gamma^{-2}$ tem-se que

$$r^\gamma = V_\gamma^{-1/2} (t - \widehat{\mu}^*).$$

Logo, o resíduo para a i -ésima observação é

$$r_i^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d'''(\phi_i)}}, \quad i = 1, \dots, n.$$

Forma canônica da família exponencial

Seja $c(y_i; \phi_i) = d(\phi_i) + \phi_i a(y_i) + u(y_i)$. Logo a Equação (2.1) pode ser reescrita como

$$f(y_i, \theta_i, \phi_i) = \exp\{\phi_i \theta_i y_i + \phi_i (y_i) - (\phi_i b(\theta_i) - d(\phi_i)) + u(y_i)\}.$$

Desta forma, tem-se que $\tau_1(y_i) = y_i$, $\tau_1(y_i) = a(y_i)$, $c_1(\theta_i; \phi_i) = \theta_i \phi_i$, $c_2(\theta_i; \phi_i) = \phi_i$ e $\delta(\theta_i; \phi_i) = \phi_i b(\theta_i) - d(\phi_i)$. Logo, os parâmetros canônicos são dados por

$$\nu_1 = \theta_i \phi_i \Rightarrow \theta_i = \nu_1 / \nu_2 \quad \text{e} \quad \nu_2 = \phi_i.$$

Então, temos que $\delta_0(\nu_1; \nu_2) = \nu_2 b(\nu_1 / \nu_2) - d(\nu_2)$. Assim,

$$E(\tau_1) = \frac{\partial \delta_0(\nu_1; \nu_2)}{\partial \nu_1} = b' \left(\frac{\nu_1}{\nu_2} \right) = b''(\theta_i), \quad (\text{B.1})$$

$$\text{Var}(\tau_1) = \frac{\partial^2 \delta_0(\nu_1; \nu_2)}{\partial \nu_1^2} = \frac{1}{\nu_2} b'' \left(\frac{\nu_1}{\nu_2} \right) = \phi_i^{-1} b''(\theta_i), \quad (\text{B.2})$$

$$\begin{aligned} E(\tau_2) &= \frac{\partial \delta_0(\nu_1; \nu_2)}{\partial \nu_2} = b \left(\frac{\nu_1}{\nu_2} \right) - \frac{\nu_1}{\nu_2} b' \left(\frac{\nu_1}{\nu_2} \right) - d'(\nu_2), \\ &= b(\theta_i) - \theta_i b'(\theta_i) - d'(\phi_i), \end{aligned} \quad (\text{B.3})$$

$$\text{Var}(\tau_2) = \frac{\partial^2 \delta_0(\nu_1; \nu_2)}{\partial \nu_2^2} = \frac{\nu_1^2}{\nu_2^3} b'' \left(\frac{\nu_1}{\nu_2} \right) - d''(\nu_2) = \frac{\theta_i^2}{\phi_i} b''(\theta_i) - d''(\phi_i), \quad (\text{B.4})$$

$$\text{Cov}(\tau_1; \tau_2) = \frac{\partial^2 \delta_0(\nu_1; \nu_2)}{\partial \nu_1 \partial \nu_2} = -\frac{\nu_1}{\nu_2} b'' \left(\frac{\nu_1}{\nu_2} \right) = -\frac{\theta_i}{\phi_i} b''(\theta_i). \quad (\text{B.5})$$

Mais detalhes em Lehmann e Casella (1998).