DIAGNÓSTICO EM MODELOS SIMÉTRICOS DE REGRESSÃO

LUIS HERNANDO VANEGAS PENAGOS

Orientador: Prof. Dr. Francisco José de Azevêdo Cysneiros
Área de Concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, dezembro de 2005

Vanegas Penagos, Luis Hernando
Diagnóstico em modelos simétricos de regressão / Luis Hernando Vanegas Penagos . – Recife : O Autor, 2005.
xii, 119 folhas. : il., fig., quadros.

Dissertação (mestrado) – Universidade Federal de Pernambuco. MEI. Estatística, 2005.

Inclui bibliografia e apêndices.

1.Estatística aplicada – Modelagem . 2. Modelos simétricos de regressão – Tipos de resíduos – Componentes desvio e quantal . 3. Influência global – Métodos de diagnóstico . I. Título.

519.233.5	**CDU (2.ed.)**	**UFPE**
519.536	**CDD (22.ed.)**	**BC2005-678**

Universidade Federal de Pernambuco
Mestrado em Estatística

20 de dezembro de 2005
(data)

Nós recomendamos que a dissertação de mestrado de autoria de

Luis Hernando Vanegas Penagos

intitulada

Diagnóstico em Modelos Simétricos de Regressão

seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Klaus Leite Pinto Vasconcellos
Coordenador da Pós-Graduação em Estatística

Prof. Klaus L. P. Vasconcellos
Coordenador da
Pós-graduação em
Estatística, UFPE

Banca Examinadora:

Francisco José de Azevedo Cysneiros
Francisco José de Azevedo Cysneiros orientador

Lucia Pereira Barroso (USP)

Klaus Leite Pinto Vasconcellos
Klaus Leite Pinto de Vasconcellos

Este documento será anexado à versão final da dissertação.

À minha esposa, Luz Marina,

À minha mãe, María Eugenia,

À minha irmã, Luz Dary.

Agradecimentos

Ao Programa de Mestrado em Estatística da Universidade Federal de Pernambuco, pela oportunidade e pelo apoio a mim concedidos, que me permitiram realizar o mestrado neste maravilhoso país, e em especial, aos seus coordenadores, os professores Francisco Cribari Neto e Klaus Vasconcellos.

Ao Professor Dr. Francisco José A. Cysneiros, meu orientador, pela oportunidade concedida, confiança, apoio, incentivo, disponibilidade, competência, paciência, e excelente orientação.

Ao corpo docente do Programa de Mestrado em Estatística da Universidade Federal de Pernambuco, pela sua contribuição na minha formação pessoal, acadêmica e profissional.

À minha esposa, Luz Marina, pelo apoio, paciência, carinho e amor por ela sempre oferecidos. Agradeço também, sua compreensão da minha ausência, mesmo quando eu estava presente fisicamente.

À minha mãe, María Eugenia, pelo apoio, amor, e em especial, pelo seu imensurável esforço para me educar e me fornecer princípios básicos.

À minha irmã, Luz Dary, pelo incentivo e compreensão.

Aos meus amigos, Gabriel, Angela, Alexandra, Angélica, Rafael, Javier e Alexander, pelo incentivo e amizade.

Aos meus colegas do mestrado, e em especial, Artur José Lemonte e Themis da Costa Abensur, pela amizade por eles oferecida.

À banca examinadora, pelos comentários e sugestões.

À Valeria Bittencourt, pela eficiência e disponibilidade.

Ao CNPq pelo apoio financeiro.

Resumo

A modelagem estatística sob a suposição de erros normalmente distribuídos pode ser altamente influenciada por observações extremas, o que motiva o estudo de técnicas de regressão mais robustas frente a este tipo de observações. Como uma alternativa, podem ser considerados modelos em que a distribuição assumida para o erro pertence à classe simétrica. Estes modelos são chamados de modelos simétricos de regressão. Neste trabalho, estudamos analiticamente e através de simulações de Monte Carlo as propriedades estatísticas de dois tipos de resíduos propostos: componente desvio e quantal para os modelos simétricos de regressão com componentes sistemáticas não-lineares. Além disso, desenvolvemos métodos de diagnóstico usando o enfoque de influência global. Neste sentido, propomos algumas medidas tais como distância de Cook, estatística W-K, aproximação a um passo, afastamento da verossimilhança e alguns métodos gráficos.

Abstract

The statistical modelling using the assumption of normally distributed errors can be strongly influenced by extreme observations. This motivates the study on regression techniques that are robust in presence of this type of observations. As an alternative, models in which the error term belongs to the symmetrical class can be considered. This class of models is called symmetrical regression models. In these models, the tails of the distribution of the error term are heavier than the tails of the normal distribution. In this work, we studied analytically and numerically the properties for two types of proposed residuals for symmetrical nonlinear regression models: deviance and quantal. Besides, we developed diagnostic methods on symmetrical regression models with systematic nonlinear components using the global influence approach. Through this approach we obtain measures as the Cook distance, W-K statistic, one step approximation, likelihood displacement and some graphics methods.

Sumário

Lista de Quadros

Lista de Figuras

CAPÍTULO 1

Introdução

A modelagem estatística é uma das ferramentas usadas com maior freqüência em diversas áreas do conhecimento por pesquisadores, cujo interesse é responder às perguntas que motivaram sua investigação através da análise estatística de um conjunto de dados. Na literatura estatística existem muitas técnicas de modelagem, dentre as quais, a análise baseada em modelos de resposta normal é a mais popular quando a variável de interesse é contínua, devido a que esta distribuição tem propriedades estatísticas desejáveis e uma ampla teoria desenvolvida. Entretanto, é bem conhecido que a modelagem sobre a suposição de respostas normalmente distribuídas pode ser altamente influenciada por outliers (observações extremas ou aberrantes). Sendo assim, a aplicação destes modelos na presença de observações extremas pode levar a conclusões inadequadas.

Como uma alternativa à análise clássica para conjuntos de dados com observações extremas, podem ser considerados modelos em que a distribuição assumida para a resposta pertence à classe simétrica de distribuições (veja Fang, Kotz e Ng, 1990). Por exemplo, modelos em que as respostas são assumidas como variáveis aleatórias independentes seguindo distribuições com caudas mais pesadas do que a normal podem reduzir e controlar a influência das observações aberrantes. Estes modelos, chamados de modelos simétricos de regressão (veja Cysneiros, Paula e Galea, 2005), podem ser considerados como uma extensão da teoria clássica de regressão para resposta contínua, na qual se proporciona alta flexibilidade na especificação da distribuição para a variável resposta. Exemplos típicos desta classe são modelos em que é assumido que a resposta segue distri-

buições tais como normal, logística tipos I e II, t-Student, t-Student generalizada, exponencial potência, entre outras.

Dado que o modelo simétrico de regressão pode ser uma alternativa para a modelagem de conjuntos de dados com observações extremas, é importante avaliar a adequabilidade e/ou robustez deste tipo de modelo na presença de observações aberrantes. Para esta avaliação podem ser usados os métodos de diagnóstico, os quais permitem avaliar se a diferença entre o observado e o explicado pelo modelo é estatisticamente significativa. Além disso, as técnicas de diagnóstico permitem identificar observações com uma influência desproporcional nas estimativas dos parâmetros do modelo. Portanto, a validação do modelo ajustado através de métodos de diagnóstico é uma componente fundamental na aplicação dos modelos simétricos de regressão, o que incentiva o estudo e aprofundamento deste tópico.

Galea, Paula e Cysneiros (2005) desenvolvem métodos de diagnóstico em modelos simétricos de regressão, propondo um resíduo seguindo a linha de Cox e Snell (1968), avaliando influência local seguindo a linha de Cook (1986), e medindo alavancagem seguindo a linha de Wei, Hu e Fung (1998). Neste trabalho desenvolvemos métodos de diagnóstico em modelos simétricos de regressão usando o enfoque de influência global. Este enfoque foi usado, por exemplo, em Cook (1977) para os modelos normais lineares, em Pregibon (1981) para os modelos lineares generalizados, em Tang, Wei e Wang (2000) para os modelos não-lineares reprodutivos, e em Fung et *al* (2002) para os modelos mistos semi-paramétricos.

Este trabalho está organizado em seis capítulos. No Capítulo 2 definimos os modelos simétricos de regressão, descrevendo suas componentes aleatória e sistemática, bem como suas principais propriedades e características no processo de estimação e de inferência. Também são apresentadas as expressões da função de escore e da matriz de informação esperada de Fisher. Além disso, estudamos o efeito da especificação da distribuição para a variável resposta no processo de estimação.

No Capítulo 3, propomos uma definição geral de resíduos para os modelos simétricos de regressão. A adequabilidade e as propriedades estatísticas mais importantes dos resíduos propostos são estudadas analiticamente usando expansões de Cox e Snell (1968), e via simulação de Monte Carlo aplicando os resíduos es-

tudados em modelos em que a componente sistemática é dada pelas curvas de Gompertz e Michaelis-Menten. A distribuição assintótica dos resíduos propostos é estudada usando as propriedades dos estimadores de máxima verossimilhança. Além disso, os resíduos considerados naquele capítulo são comparados através de simulação de Monte Carlo com o resíduo proposto em Galea, Paula e Cysneiros (2005). Vale salientar que os estudos de simulação desenvolvidos neste trabalho foram conduzidos usando as rotinas computacionais nos softwares `R` e `S-Plus` disponíveis no site `www.de.ufpe.br/~cysneiros/elliptical/elliptical.html` (veja Cysneiros, Paula e Galea, 2005).

No Capítulo 4, desenvolvemos algumas medidas para identificar observações influentes na classe dos modelos simétricos de regressão. Utilizando o enfoque baseado no modelo de *deleção de casos* (CDM) proposto por Cook (1977), desenvolvemos várias medidas tais como distância de Cook (Cook, 1977), estatística W-K (Pregibon, 1981), aproximação a um passo e o afastamento da verossimilhança (Cook e Weisberg, 1982). Além disso, utilizando simulação de Monte Carlo e o modelo de Michaelis-Menten, estudamos o desempenho do teste para identificar outliers baseado na metodologia denominada *modelo de mudança de médias* (MSOM) (veja Wei e Fung, 1999). Finalmente, naquele capítulo apresentamos alguns métodos gráficos para avaliar a influência das observações na significância estatística dos parâmetros do modelo. Estes métodos são uma extensão das técnicas estudadas por Cook e Weisberg (1982) e Wang (1985).

No Capítulo 5 apresentamos dois exemplos através dos quais ilustramos a aplicação e interpretação da metodologia estudada neste trabalho. Para o primeiro exemplo, consideramos uma aplicação analisada em Ratkowsky (1983), cujo interesse principal é estudar a relação entre o peso das lentes dos olhos de coelhos europeus, y (em mg) e a idade do animal, x (em dias), numa amostra de 71 observações. A motivação para o uso deste conjunto de dados é a suspeita da presença de outliers quando este é analisado sob a suposição de erros normais. Como um segundo exemplo, consideramos o conjunto de dados analisado em Draper e Smith (1998), que corresponde a um experimento desenvolvido com o objetivo de explicar a concentração de cloro disponível num produto, y em relação ao tempo desde que o mesmo foi fabricado, x (em semanas), numa amostra de 44 observações. A motivação para o uso deste conjunto de dados é a

presença de outliers quando o mesmo é analisado com o modelo de resposta normal. Para finalizar, no Capítulo 6 apresentamos as conclusões mais importantes deste trabalho.

CAPÍTULO 2

Modelos Simétricos de Regressão

Os modelos de regressão são formulados e aplicados com o objetivo de explicar e/ou predizer o comportamento de uma variável de interesse, denominada variável resposta, em relação a um conjunto de variáveis explicativas. Isto é feito supondo que para cada indivíduo na população de estudo, a variável resposta segue uma distribuição de probabilidade cujos parâmetros dependem dos valores das variáveis explicativas. Desta maneira, os modelos de regressão podem ser definidos com base em duas componentes principais, a aleatória e a sistemática. A componente aleatória descreve a distribuição assumida pelo modelo para a variável resposta, a qual é escolhida dependendo de características particulares da variável resposta na população de estudo. Uma destas características pode ser a escala na qual é medida esta variável; desta maneira, se a resposta é contínua no intervalo $(-\infty, \infty)$ é possível assumir uma distribuição normal, enquanto que, para respostas do tipo discreto, por exemplo, podem-se usar distribuições como poisson ou binomial, entre outras.

Por outro lado, a componente sistemática do modelo descreve a relação assumida entre a variável resposta e as variáveis explicativas. Geralmente, a média da distribuição da variável resposta é um indicador adequado do comportamento desta variável; portanto, muitos dos modelos de regressão são formulados supondo que a média da distribuição da resposta é uma função das variáveis explicativas e de um vetor de parâmetros desconhecidos. Desta maneira, a componente sistemática determina a relação funcional entre os parâmetros da distribuição da resposta e as variáveis explicativas. Dependendo do fenômeno em estudo,

a componente sistemática pode ser considerada como linear ou não-linear nos parâmetros, dando lugar aos modelos de regressão lineares e não-lineares, respectivamente.

Assim, os modelos simétricos de regressão são uma extensão da teoria clássica de regressão para resposta contínua, na qual se abre o leque de opções para a componente aleatória do modelo, permitindo que a distribuição para a variável resposta possa ser estendida à classe simétrica de distribuições. Além disso, estes modelos podem considerar componentes sistemáticas não-lineares, possibilitando o estudo de grande variedade de fenômenos, muitos dos quais não podem ser abordados usando modelos lineares. Entretanto, a motivação para o uso de modelos simétricos de regressão não está baseada somente na flexibilidade da especificação da componente aleatória, pois estes modelos proporcionam também uma modelagem apropriada para a análise de conjuntos de dados com outliers (observações extremas ou aberrantes), em que a aplicação de modelos de resposta normal pode ser inadequada. A aplicação de um modelo em que a distribuição assumida para a resposta tem caudas mais pesadas do que a normal pode reduzir e controlar a influência das observações extremas, permitindo uma inferência apropriada do conjunto de dados em estudo.

Os modelos simétricos de regressão têm recebido muitas contribuições nos últimos anos, podendo-se relacionar alguns trabalhos recentes. Por exemplo, Cordeiro et *al* (2000) derivam expressões para o viés de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros nos modelos simétricos de regressão. Ferrari e Uribe-Opazo (2001) desenvolvem correções de Bartlett para a estatística da razão de verossimilhanças em modelos com componentes sistemáticas lineares. Cordeiro (2004) estende os resultados de Ferrari e Uribe-Opazo (2001) aos modelos com componentes sistemáticas não-lineares. Galea, Paula e Uribe-Opazo (2003) avaliam influência local em modelos simétricos de regressão com componentes sistemáticas lineares. Cysneiros e Paula (2004) desenvolvem testes restritos em modelos lineares com erros t-multivariados. Cysneiros (2004) e Galea, Paula e Cysneiros (2005) propõem um resíduo padronizado e avaliam influência local em modelos simétricos de regressão com componentes sistemáticas não-lineares. Podemos citar algumas referências de texto no assunto tais como Fang, Kotz e Ng (1990), Fang e Anderson (1990), Fang e Zhang (1990)

e Cysneiros, Paula e Galea (2005).

Neste capítulo definimos os modelos simétricos de regressão, descrevendo suas componentes aleatória e sistemática, bem como suas propriedades e características mais importantes no processo de estimação e de inferência.

2.1 Classe simétrica de distribuições

A família de distribuições elípticas, introduzida por Kelker (1970), consiste em uma generalização da distribuição normal multivariada. Esta família de distribuições tem sido considerada por muitos autores, como, por exemplo, Cambanis, Huang e Simons (1981), Berkane e Bentler (1986), Fang, Kotz e Ng (1990), Fang e Anderson (1990), Fang e Zhang (1990), Rao (1990), Landsman e Valdez (2002), entre outros. No caso univariado, a família de distribuições elípticas corresponde à classe simétrica de distribuições, na qual está baseada a componente aleatória dos modelos simétricos de regressão. A distribuição da variável aleatória y pertence à classe simétrica de distribuições se sua função de densidade pode ser escrita como

$$f(y;\mu,\phi) = \frac{1}{\sqrt{\phi}}\, g\left[\frac{(y-\mu)^2}{\phi}\right], \quad y \in I\!R, \tag{2.1}$$

para alguma função $g(\cdot)$ denominada função geradora de densidades, com $g(u) > 0$, para $u > 0$ e $\int_0^\infty u^{-1/2} g(u)du = 1$, em que $\mu \in I\!R$ e $\phi > 0$ são os parâmetros de locação e escala, respectivamente. Se esta condição é satisfeita, escrevemos $y \sim S(\mu,\phi)$ e dizemos que a distribuição da variável aleatória y pertence à classe simétrica. Como exemplos de distribuições pertencentes a esta classe podemos citar a normal, cauchy, t-Student, t-Student generalizada, logística tipos I e II, logística generalizada, Kotz, Kotz generalizada, normal contaminada, exponencial dupla, exponencial potência, entre outras. No Quadro 2.1 apresentamos algumas destas distribuições.

Se $y \sim S(\mu,\phi)$, sua função característica, $\varsigma_y(t) = E(e^{ity})$, pode ser escrita como $\varsigma_y(t) = e^{it\mu}\varphi(t^2\phi)$, $t \in I\!R$ para alguma função φ, com $\varphi(u) \in I\!R$ para $u > 0$. Quando existem, $E(y) = \mu$ e $Var(y) = \xi\phi$, em que $\xi > 0$ é uma constante dada por $\xi = -2\varphi'(0)$, com $\varphi'(0) = \partial\varphi(u)/\partial u|_{u=0}$ e que não depende dos parâmetros μ e ϕ. As distribuições pertencentes à classe simétrica apresentam algumas propriedades análogas às da distribuição normal, como por exemplo, se

Quadro 2.1 *Algumas distribuições pertencentes à classe simétrica.*

Distribuição	Função geradora de densidades $g(u)$		Curtose
Normal	$\frac{1}{\sqrt{2\pi}}\exp(-u/2)$		3
Cauchy	$\frac{1}{\pi(1+u)}$		não existe
t-Student	$\frac{1}{\mathrm{B}(1/2,\nu/2)}\nu^{\nu/2}\,(\nu+u)^{-\frac{\nu+1}{2}},$	$\nu>0$	$3+\frac{6}{\nu-4},\ \nu>4$
t-Student generalizada	$\frac{1}{\mathrm{B}(1/2,r/2)}s^{r/2}\,(s+u)^{-\frac{r+1}{2}},$	$s,r>0$	$3+\frac{6}{r-4},\ r>4$
Logística tipo I	$\frac{c}{\mathrm{e}^u(1+\mathrm{e}^{-u})^2}$		2,3851
Logística tipo II	$\frac{1}{\mathrm{e}^{u^{1/2}}(1+\mathrm{e}^{-u^{1/2}})^2}$		4,2

$\mathrm{B}(\cdot,\cdot)$ é a função beta; $c\approx 1,484300029$.

Distribuição	Função geradora de densidades $g(u)$	Curtose
Logística generalizada	$\frac{\alpha}{\mathrm{B}(m,m)}\frac{\mathrm{e}^{-m\alpha\sqrt{u}}}{(1+\mathrm{e}^{-\alpha\sqrt{u}})^{2m}}$, $m>0,\ \alpha>0$	$3+\frac{\psi^{(3)}(m)}{2\psi^{(1)}(m)^2}$
Exponencial potência	$\frac{\exp[-\frac{1}{2}u^{1/(k+1)}]}{\Gamma(1+\frac{1+k}{2})2^{1+(1+k)/2}}$, $-1<k\leq 1$	$\frac{\Gamma(\frac{5}{2}(1+k))\Gamma(\frac{1+k}{2})}{\Gamma^2(\frac{3}{2}(k+1))}$
Kotz	$\frac{r^{(2N-1)/2}}{\Gamma(\frac{2N-1}{2})}u^{N-1}\mathrm{e}^{-ru}$, $r>0,\ N\geq 1$	$\frac{2N+1}{2N-1}$
Kotz generalizada	$\frac{sr^{(2N-1)/2s}}{\Gamma(\frac{2N-1}{2s})}u^{N-1}\mathrm{e}^{-ru^s}$, $r,s>0,\ N\geq 1$	$\frac{\Gamma(\frac{2N-1}{2s})\Gamma(\frac{2N+3}{2s})}{\Gamma^2(\frac{2N+1}{2s})}$
Normal contaminada	$(1-\varepsilon)\frac{\exp(-u/2)}{\sqrt{2\pi}}+\varepsilon\frac{\exp(-u/(2\sigma^2))}{\sqrt{2\pi}\sigma}$, $\sigma>0,\ 0\leq\varepsilon\leq 1$	$\frac{3[1+\varepsilon(\sigma^4-1)]}{[1+\varepsilon(\sigma^2-1)]^2}$

$\mathrm{B}(\cdot,\cdot)$ é a função beta, $\Gamma(\cdot)$ a função gama e $\psi^{(1)}(\cdot), \psi^{(3)}(\cdot)$ a primeira e terceira derivadas da função digama.

$y \sim S(\mu, \phi)$ então $a + by \sim S(a + b\mu, b^2\phi)$, em que a, $b \in \mathbb{R}$ com $b \neq 0$, isto é, a distribuição de uma combinação linear de uma variável aleatória cuja distribuição pertence à classe simétrica também pertence a esta classe. Este resultado permite que, a partir de uma variável aleatória simétrica padrão, possa ser construída uma família de variáveis aleatórias cujas distribuições também pertencem à classe simétrica. Ou seja, se z segue distribuição $S(0,1)$ com função geradora de densidades $g(\cdot)$, então para cada $\mu \in \mathbb{R}$ e $\phi > 0$, a distribuição da variável aleatória $y = \mu + \sqrt{\phi}\, z$, pertence à classe simétrica de distribuições com função geradora de densidades $g(\cdot)$ e primeiros momentos (quando existem) dados por $E(y) = \mu$, $Var(y) = \xi\phi$.

Apesar de que a classe simétrica de distribuições proporciona grande flexibilidade para a componente aleatória dos modelos simétricos de regressão, as distribuições pertencentes a esta classe com caudas mais pesadas do que a normal são de maior interesse, pois os modelos baseados nessas distribuições são menos afetados pelas observações extremas do que o modelo de resposta normal. Entre estas podemos citar, por exemplo, as distribuições t-Student, t-Student generalizada ($s \geq r - 2 > 0$), logística tipo II, exponencial potência ($k > 0$), entre outras. Como ilustração, apresentamos nas Figuras 2.1 a 2.4 os gráficos das densidades de algumas destas distribuições em suas versões padronizadas. Em todos os casos, a densidade considerada é comparada com a densidade da distribuição normal padrão, a qual é representada pela linha pontilhada. Todavia, podemos relacionar algumas propriedades destas distribuições; por exemplo, a distribuição t-Student tende para a distribuição normal à medida que ν, o número de graus de liberdade, tende para infinito. Este comportamento também é observado para a distribuição exponencial potência quando k tende para zero. A distribuição t-Student generalizada de parâmetros r e s é equivalente à distribuição t-Student com ν graus de liberdade quando $r = s = \nu$.

Figura 2.1 *Função de densidade da distribuição logística tipo II.*

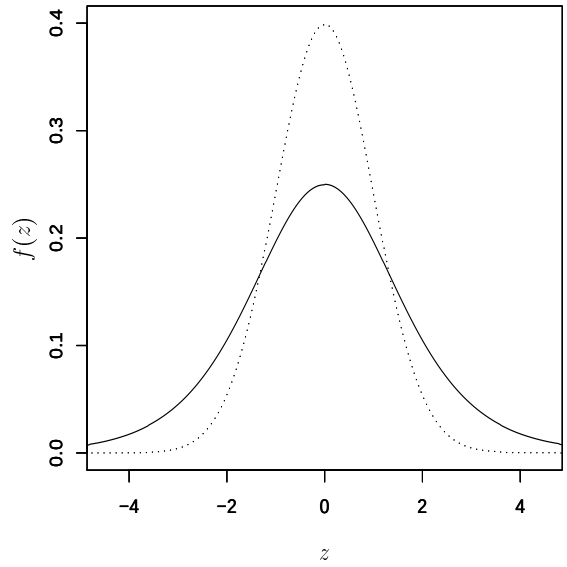


Figura 2.2 *Função de densidade da distribuição t-Student generalizada com parâmetros* $r = 3$, $s = 5$ *(esquerda) e* $s = 7$ *(direita).*

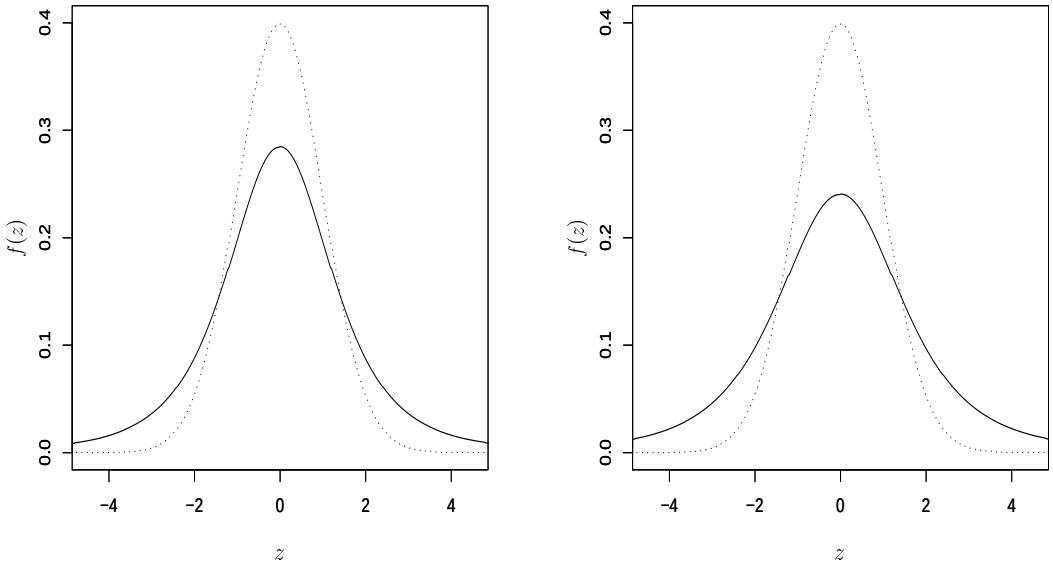


Figura 2.3 *Função de densidade da distribuição t-Student com* $\nu = 4$ *(esquerda) e* $\nu = 10$ *(direita) graus de liberdade.*

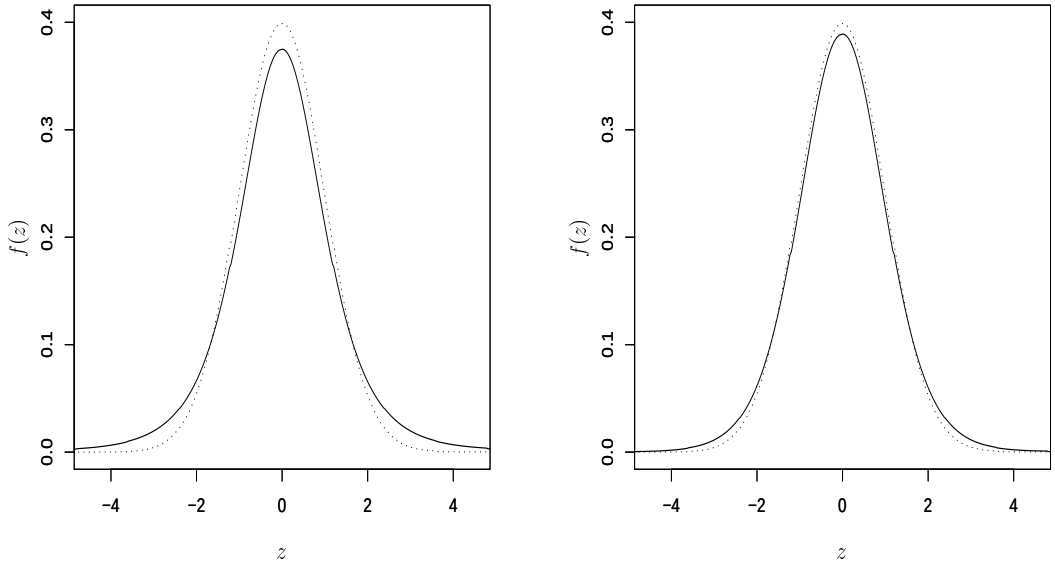
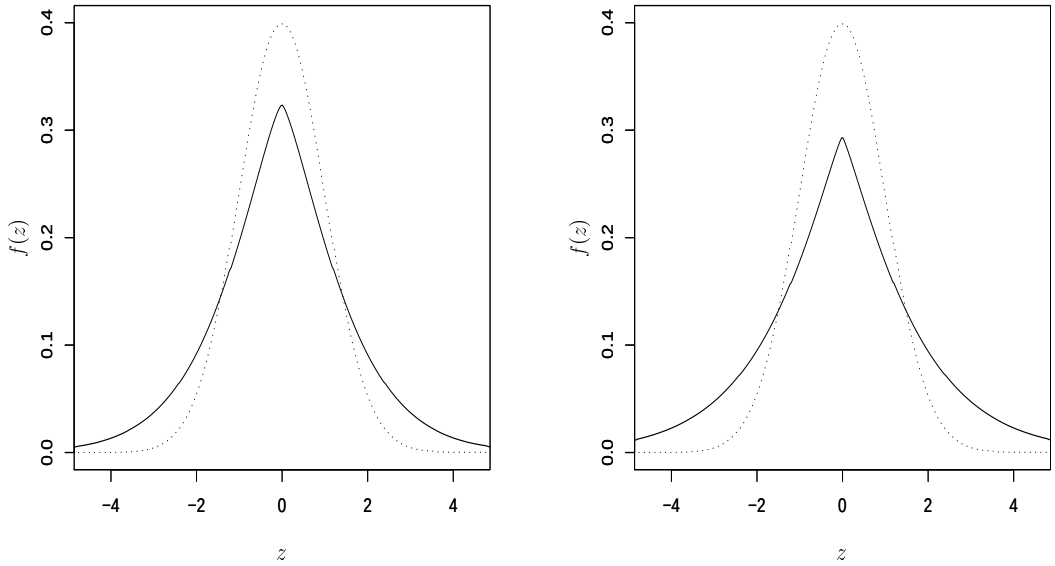


Figura 2.4 *Função de densidade da distribuição exponencial potência com parâmetro* $k = 0,5$ *(esquerda) e* $k = 0,7$ *(direita).*



2.2 Modelo Simétrico de Regressão

Sejam $y_1, \ldots, y_n$ os valores de uma variável de interesse medida em n indivíduos ou unidades experimentais, os quais são assumidos como variáveis aleatórias independentes cada uma com distribuição pertencente à classe simétrica definida em (2.1), com μ_i e $\phi > 0$ os parâmetros de locação e escala da distribuição de y_i. Com o objetivo de explicar e/ou predizer o comportamento das y_i, supomos que existe um conjunto de l variáveis, denominadas variáveis explicativas, cuja relação com a variável de interesse é chamada componente sistemática e pode ser descrita por

$$\mu_i = \mu(\boldsymbol{\beta}, \mathbf{x}_i), \qquad i = 1, \ldots, n, \tag{2.2}$$

em que $\mathbf{x}_i = (x_{i1}, \ldots, x_{il})$ são os valores das variáveis explicativas para o i-ésimo indivíduo e $\mu_i(\boldsymbol{\beta}) = \mu(\boldsymbol{\beta}, \mathbf{x}_i)$ é uma função contínua e duplamente diferenciável em relação ao vetor $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ de parâmetros desconhecidos. Adicionalmente, assumimos que a matriz de derivadas $\mathbf{D}_\beta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$ tem posto $p\,(p < n)$ para todo $\boldsymbol{\beta} \in \Omega_\beta \subset I\!R^p$, com Ω_β um conjunto compacto com pontos interiores. O modelo com componente sistemática (2.2) é considerado não-linear quando a matriz $\mathbf{D}_\beta$ envolve o vetor de parâmetros $\boldsymbol{\beta}$. Assim, os modelos simétricos de regressão são definidos pela componente aleatória (2.1) e pela componente sistemática (2.2). Entretanto, é possível formular o modelo descrito acima usando as propriedades das distribuições pertencentes à classe simétrica. Isto é feito escrevendo o valor da variável de interesse na seguinte forma

$$y_i \;=\; \mu_i(\boldsymbol{\beta}) + \sqrt{\phi}\,\epsilon_i, \qquad i = 1, \ldots, n, \tag{2.3}$$

em que $\epsilon_1, \ldots, \epsilon_n$ são variáveis aleatórias independentes seguindo distribuição $S(0, 1)$. Esta formulação é a mais popular na literatura estatística, pois, da mesma maneira que em modelos de resposta normal, os ϵ_i podem ser interpretados como erros aleatórios, permitindo estender diretamente aos modelos simétricos de regressão muitas das idéias e dos conceitos nos modelos lineares e não-lineares de resposta normal.

A seguir, apresentamos alguns exemplos de componentes sistemáticas não-lineares usadas freqüentemente em modelos de regressão aplicados nas áreas biológica e química:

- Curva de Gompertz

$$\mu_i(\boldsymbol{\beta}) = \beta_1 \exp\{-\beta_2 \exp(-\beta_3 x_i)\}, \tag{2.4}$$

- Curva de Michaelis-Menten

$$\mu_i(\boldsymbol{\beta}) = \frac{\beta_1 x_i}{x_i + \beta_2}. \tag{2.5}$$

A curva de Gompertz permite descrever o crescimento de um objeto ou indivíduo em relação ao tempo, sendo β_1 o tamanho máximo que pode ter o indivíduo e β_3 a taxa de crescimento. Por outro lado, a curva de Michaelis-Menten é usada na área química para descrever o comportamento cinético das enzimas em relação a sua concentração, sendo β_1 a velocidade máxima da reação e β_2 uma constante cinética. No entanto, devido a sua flexibilidade, estas curvas também são utilizadas para descrever grande variedade de fenômenos em áreas como engenharia e economia. Nas Figuras 2.5 e 2.6 ilustramos o comportamento destas curvas para algumas combinações dos parâmetros.

Figura 2.5 *Comportamento da família de curvas de Gompertz para diferentes valores de* β_3 *com* $\beta_1 = 10$ *e* $\beta_2 = 7$.

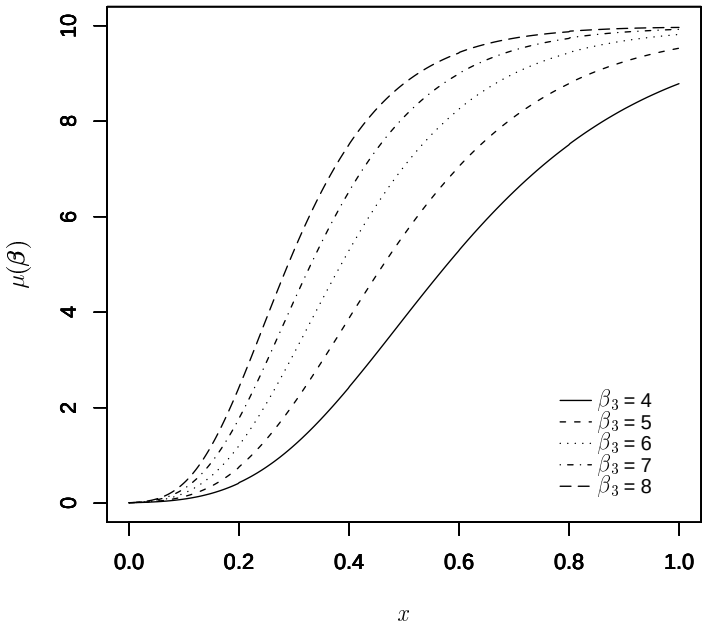
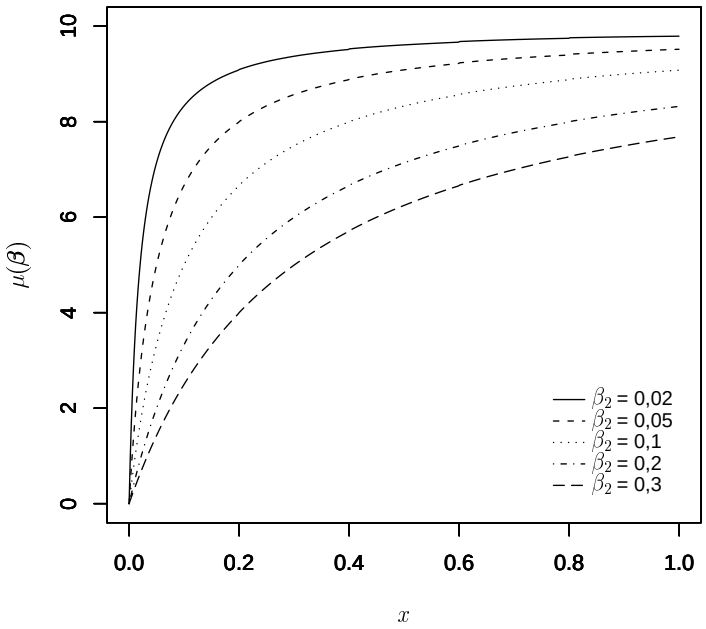


Figura 2.6 *Comportamento da família de curvas de Michaelis-Menten para diferentes valores de* β_2 *com* $\beta_1 = 10$.



Para a estimação dos parâmetros nos modelos simétricos de regressão é possível aplicar o método de máxima verossimilhança. Este método de estimação consiste em maximizar uma função que expresse a chance de observar os dados que compõem a amostra em função dos parâmetros do modelo. A utilização deste método apresenta muitas vantagens, pois, sob certas condições, os estimadores obtidos desta maneira possuem propriedades estatísticas desejáveis, como consistência, eficiência e normalidade assintóticas. Para a aplicação deste método de estimação podemos considerar o logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$, o qual é dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} l(y_i; \mu_i, \phi),$$

em que

$$l(y_i; \mu_i, \phi) = \log\big[g(z_i^2)\big] - \frac{1}{2}\log(\phi), \tag{2.6}$$

com $z_i = (y_i - \mu_i(\boldsymbol{\beta}))/\phi^{1/2}$. O estimador de máxima verossimilhança de $\boldsymbol{\theta}$ pode ser escrito como $\hat{\boldsymbol{\theta}} = \arg\max_{\theta} L(\boldsymbol{\theta})$. Quando o parâmetro de dispersão ϕ é conhecido,

o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ pode ser obtido na seguinte forma

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} MQ(\boldsymbol{\beta}), \tag{2.7}$$

em que $MQ(\boldsymbol{\beta})$ é uma função não-negativa e duplamente diferenciável em relação ao vetor de parâmetros $\boldsymbol{\beta}$, podendo ser expressa na forma abaixo

$$MQ(\boldsymbol{\beta}) = 2\sum_{i=1}^{n} \log\left[g(0)/g(z_i^2)\right], \tag{2.8}$$

com $g(\cdot)$ a função geradora de densidades. Então, a estimativa de $\boldsymbol{\beta}$ é obtida minimizando uma expressão da forma $\sum \Psi(z_i^2)$, em que a função $\Psi(\cdot)$ depende da função $g(\cdot)$, a qual é induzida pela distribuição assumida para a variável resposta. Para a distribuição normal temos que $g(u) = \exp(-u/2)$, caso em que $\Psi(\cdot)$ é a função identidade e (2.7) corresponde ao estimador de mínimos quadrados ordinários. O processo de estimação de $\boldsymbol{\theta}$ quando ϕ é desconhecido será discutido na página 19.

De (2.6) podemos obter a função de escore para $\boldsymbol{\theta}$, a qual pode ser escrita na seguinte forma

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_{\beta}(\boldsymbol{\theta}) \\ \\ \mathrm{U}_{\phi}(\boldsymbol{\theta}) \end{bmatrix},$$

em que

$$\mathbf{U}_{\beta}(\boldsymbol{\theta}) = \frac{1}{\phi}\mathbf{D}_{\beta}^{T}\mathbf{D}(\mathbf{v})(\mathbf{y}-\boldsymbol{\mu}) \tag{2.9}$$

e

$$\mathrm{U}_{\phi}(\boldsymbol{\theta}) = -\frac{n}{2\phi} + \frac{1}{2\phi^2}(\mathbf{y}-\boldsymbol{\mu})^{T}\mathbf{D}(\mathbf{v})(\mathbf{y}-\boldsymbol{\mu}), \tag{2.10}$$

com $\mathbf{D}(\mathbf{v}) = \mathrm{diag}\{v_1,\ldots,v_n\}$, $\mathbf{y} = (y_1,\ldots,y_n)^{T}$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^{T}$, $\mu_i = \mu_i(\boldsymbol{\beta})$, $v_i = v(z_i)$, $v(z) = -2g'(z^2)/g(z^2)$, e $g'(z^2) = \partial g(u)/\partial u|_{u=z^2}$. A matriz de informação observada de Fisher para $\boldsymbol{\theta}$ pode ser denotada como $-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}} = -\ddot{\mathbf{L}}_{\theta\theta}|_{\hat{\theta}}$. Podemos escrever a matriz $\ddot{\mathbf{L}}_{\theta\theta}$ da seguinte forma

$$\ddot{\mathbf{L}}_{\theta\theta} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\phi} \\ \ddot{\mathbf{L}}_{\phi\beta} & \ddot{\mathrm{L}}_{\phi\phi} \end{bmatrix},$$

em que

$$\ddot{\mathbf{L}}_{\beta\beta} = -\frac{1}{\phi}\Big\{\sum_{i=1}^{n} 2s_i\mathbf{D}_{\beta\beta}(i) + \mathbf{D}_{\beta}^{T}\mathbf{D}(\mathbf{a})\mathbf{D}_{\beta}\Big\},$$
$$\ddot{\mathbf{L}}_{\beta\phi} = \frac{2}{\phi^2}\mathbf{D}_{\beta}^{T}\mathbf{b},$$
$$\ddot{\mathrm{L}}_{\phi\phi} = \frac{1}{\phi^2}\Big\{\frac{n}{2} + \mathbf{u}^{T}\mathbf{D}(\mathbf{c})\mathbf{u} - \frac{1}{\phi}\mathbf{e}^{T}\mathbf{D}(\mathbf{v})\mathbf{e}\Big\},$$

com $\mathbf{D}_{\beta\beta}(i) = \partial^2\mu_i/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$, $\mathbf{D}(\mathbf{a}) = \mathrm{diag}\{a_1,\ldots,a_n\}$, $\mathbf{D}(\mathbf{c}) = \mathrm{diag}\{c_1,\ldots,c_n\}$, $\mathbf{b} = (b_1,\ldots,b_n)^T$, $\mathbf{u} = (u_1,\ldots,u_n)^T$, $a_i = v(z_i) + v'(z_i)z_i$, $c_i = -v'(z_i)/4z_i$, $b_i = -\{v(z_i)/2+v'(z_i)z_i/4\}e_i$, $u_i = z_i^2$, $e_i = (y_i-\mu_i)$, $s_i = -v(z_i)e_i/2$, $i = 1,\ldots,n$. A inversa de $\ddot{\mathbf{L}}_{\theta\theta}$ pode ser escrita na forma abaixo

$$\ddot{\mathbf{L}}_{\theta\theta}^{-1} = \begin{bmatrix} -\phi\mathbf{M}^{-1} + \frac{\mathbf{A}\mathbf{A}^T}{\mathrm{E}} & \frac{\mathbf{A}}{\mathrm{E}} \\ \frac{\mathbf{A}^T}{\mathrm{E}} & \frac{1}{\mathrm{E}} \end{bmatrix}, \tag{2.11}$$

em que $\mathbf{M}=\mathbf{D}_{\beta}^{T}\mathbf{D}(\mathbf{a})\,\mathbf{D}_{\beta}+\sum_{i=1}^{n} 2s_i\mathbf{D}_{\beta\beta}(i)$, $\mathbf{A} = \frac{2}{\phi}\mathbf{M}^{-1}\mathbf{D}_{\beta}^{T}\mathbf{b}$ e $\mathrm{E} = \ddot{\mathrm{L}}_{\phi\phi}+\frac{2}{\phi^2}\mathbf{b}^T\mathbf{D}_{\beta}\mathbf{A}$.

Dado que o logaritmo da função de verossimilhança é assumido regular com respeito ao vetor de parâmetros $\boldsymbol{\theta}$ segundo as condições descritas em Cox e Hinkley (1974), podemos assumir que o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ é consistente e segue distribuição normal assintoticamente, com matriz de variâncias e covariâncias dependendo da matriz de informação esperada $\mathbf{K}_{\theta\theta}$, a qual pode ser escrita como

$$\mathbf{K}_{\theta\theta} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathrm{K}_{\phi\phi} \end{bmatrix}, \tag{2.12}$$

em que $\mathbf{K}_{\beta\beta} = \dfrac{4d_g}{\phi}(\mathbf{D}_{\beta}^{T}\mathbf{D}_{\beta})$ e $\mathrm{K}_{\phi\phi} = \dfrac{n}{4\phi^2}(4f_g - 1)$, com $4d_g = E(v^2(z)z^2)$ e $4f_g = E(v^2(z)z^4)$, sendo z uma variável aleatória com distribuição $S(0,1)$. No Quadro 2.2 apresentamos expressões para $v(z)$, $4d_g$ e $4f_g$ para algumas das distribuições da classe simétrica. Da expressão (2.12) podemos concluir que os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ são assintoticamente ortogonais, o que conjuntamente com a suposição de normalidade, implica que estes estimadores são assintoticamente independentes. É possível mostrar que $\hat{\mathbf{K}}_{\beta\beta}^{-1} = \frac{\hat{\phi}}{4d_g}(\mathbf{D}_{\hat{\beta}}^{T}\mathbf{D}_{\hat{\beta}})$ é um estimador consistente da matriz de variâncias e covariâncias assintótica do

Quadro 2.2 *Expressões para $v(z)$, $4d_g$, $4f_g$ e ξ para algumas distribuições pertencentes à classe simétrica.*

Distribuição	$v(z)$	$4d_g$	$4f_g$	ξ
Normal	1	1	3	1
t-Student	$\dfrac{\nu+1}{\nu+z^2}$	$\dfrac{\nu+1}{\nu+3}$	$\dfrac{3(\nu+1)}{\nu+3}$	$\dfrac{\nu}{\nu-2}$, $\nu>2$
t-Student generalizada	$\dfrac{r+1}{s+z^2}$	$\dfrac{r(r+1)}{s(r+3)}$	$\dfrac{3(r+1)}{r+3}$	$\dfrac{s}{r-2}$, $r>2$
Logística tipo I	$2\tanh(z^2/2)$	$1,47724$	$4{,}01378$	$0{,}79569$
Logística tipo II	$\dfrac{1-\exp(-\lvert z\rvert)}{\lvert z\rvert(1+\exp(-\lvert z\rvert))}$	$\dfrac{1}{3}$	$2{,}42996$	$\dfrac{\pi^2}{3}$
Logística Generalizada	$\dfrac{\alpha m[1-\exp(-\alpha\lvert z\rvert)]}{\lvert z\rvert[1+\exp(-\alpha\lvert z\rvert)]}$	$\dfrac{\alpha^2m^2}{2m+1}$	$\dfrac{(2+m^2\psi^{(1)}(m))}{(2m+1)/2m}$	$2\psi^{(1)}(m)$
Exponencial potência	$(1+k)^{-1}\lvert z\rvert^{-2k/(k+1)}$	$\dfrac{2^{1-k}\,\Gamma((3-k)/2)}{(1+k)^2\,\Gamma((k+1)/2)}$	$\dfrac{k+3}{k+1}$	$2^{(1+k)}\frac{\Gamma(3(1+k)/2)}{\Gamma((1+k)/2)}$

estimador de máxima verossimilhança de $\boldsymbol{\beta}$. De maneira análoga, pode-se mostrar que $\hat{\mathrm{K}}^{-1}_{\phi\phi} = \frac{4\hat{\phi}^2}{n(4f_g-1)}$ é um estimador consistente da variância assintótica do estimador de máxima verossimilhança de ϕ. Vale salientar que as condições de regularidade para o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$ não são satisfeitas para distribuições como, por exemplo, exponencial dupla, Kotz, e Kotz generalizada. Neste trabalho somente consideramos as distribuições para as quais são satisfeitas estas condições.

A estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ pode ser obtida resolvendo a equação $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Se a variável resposta segue distribuição Exponencial potência mostra-se facilmente que as estimativas de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ podem ser obtidas separadamente. Nesse caso, o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ pode ser escrito na forma abaixo

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \sum_{i=1}^{n} \left| y_i - \mu_i(\boldsymbol{\beta}) \right|^{2/1+k}.$$

Por outro lado, o estimador de máxima verossimilhança do parâmetro de dispersão ϕ apresenta forma fechada, podendo ser expresso na seguinte forma

$$\hat{\phi} = \left\{ \frac{1}{n(1+k)} \sum_{i=1}^{n} \left| y_i - \mu_i(\hat{\boldsymbol{\beta}}) \right|^{2/1+k} \right\}^{1+k}.$$

Para $k = 1$ temos que $\hat{\boldsymbol{\beta}}$ é obtido minimizando a quantidade denominada L_1−$norm$, a qual tem sido estudada na literatura estatística como alternativa robusta ao estimador de mínimos quadrados (veja Huber, 1981; Sun e Wei, 2004). De forma similar, quando k tende para -1, a distribuição exponencial potência tende a uma distribuição uniforme no intervalo $\left(\mu_i - \sqrt{3\phi},\, \mu_i + \sqrt{3\phi}\right)$, caso em que $\hat{\boldsymbol{\beta}}$ é obtido minimizando a quantidade denominada na literatura estatística como L_∞−$norm$. Para $k = 0$ a distribuição exponencial potência é equivalente à distribuição normal, caso em que $\hat{\boldsymbol{\beta}}$ corresponde ao estimador de mínimos quadrados ordinários.

Contudo, no caso geral das distribuições pertencentes à classe simétrica, as estimativas de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ devem ser obtidas através de um processo iterativo conjunto. Por exemplo, $\hat{\boldsymbol{\theta}}$ pode ser obtido através de um processo iterativo da seguinte forma

$$\boldsymbol{\beta}^{(m+1)} = \left\{ \mathbf{D}_{\beta}^{T(m)} \mathbf{W}^{(m)} \mathbf{D}_{\beta}^{(m)} \right\}^{-1} \mathbf{D}_{\beta}^{T(m)} \mathbf{W}^{(m)} \tilde{\mathbf{Z}}^{(m)}, \tag{2.13}$$

$$\phi^{(m+1)} = \frac{1}{n} \mathbf{Q}(\boldsymbol{\beta}^{(m+1)}, \phi^{(m)}), \qquad m = 0, 1, 2, \ldots, \tag{2.14}$$

em que $\tilde{\mathbf{Z}} = \left[\mathbf{D}_{\beta} \boldsymbol{\beta} + (1/\phi) \mathbf{W}^{-1} \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}) \right]$ é uma variável resposta *local* modificada em cada iteração do processo, $\mathbf{W}$ é uma matriz simétrica positiva definida chamada matriz de pesos, e $\mathbf{Q}(\boldsymbol{\beta}, \phi) = (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu})$. O processo iterativo da equação (2.13) é chamado de *algoritmo delta* (veja Jørgensen, 1984). Este algoritmo é análogo ao algoritmo Newton-Raphson em que a matriz $-\ddot{\mathbf{L}}_{\beta\beta}$ é substituída por uma aproximação da forma

$$-\ddot{\mathbf{L}}_{\beta\beta} \approx (\mathbf{D}_{\beta}^{T} \mathbf{W} \mathbf{D}_{\beta}).$$

A matriz $\mathbf{W}$ deve ser selecionada de tal forma que $(\mathbf{D}_{\hat{\beta}}^{T} \hat{\mathbf{W}} \mathbf{D}_{\hat{\beta}})$ seja assintoticamente equivalente à matriz de informação esperada de $\boldsymbol{\beta}$. Por simplicidade, a matriz $\mathbf{W}$ pode ser assumida como diagonal, isto é, $\mathbf{W} = \text{diag}\{w_1, \ldots, w_n\}$ com $w_i > 0$ para todo i, garantindo que $\mathbf{W}$ é positiva definida. Neste trabalho consideramos a matriz $\mathbf{W} = -E(\partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T) = (4d_g/\phi)\mathbf{I}$, em que $\mathbf{I}$ é a matriz identidade de ordem n. Sendo assim, o processo da equação (2.13) é equivalente ao algoritmo *scoring* de Fisher, o qual é usado como algoritmo padrão no processo de estimação em muitos modelos de regressão. Além disso, com esta matriz $\mathbf{W}$, o algoritmo de estimação pode ser reescrito de tal maneira, que é possível estudar o efeito da especificação da componente aleatória do modelo no processo de estimação. Assim, podemos reescrever o processo iterativo dado em (2.13) como um procedimento de mínimos quadrados ordinários modificado, em que temos

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \left\{ \mathbf{D}_{\beta}^{T(m)} \mathbf{D}_{\beta}^{(m)} \right\}^{-1} \mathbf{D}_{\beta}^{T(m)} \left\{ \mathbf{D}_{\beta}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{D}(\boldsymbol{\rho}^{(m)}) \left[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(m)}) \right] \right\} \\ &= \left\{ \mathbf{D}_{\beta}^{T(m)} \mathbf{D}_{\beta}^{(m)} \right\}^{-1} \mathbf{D}_{\beta}^{T(m)} \tilde{\mathbf{Z}}^{(m)}, \qquad m = 0, 1, 2, \ldots \end{aligned} \tag{2.15}$$

com $\mathbf{D}(\boldsymbol{\rho}) = \text{diag}\{\rho(z_1), \ldots, \rho(z_n)\}$ e $\rho(z_i) = v(z_i)/4d_g$. No processo iterativo de estimação de $\boldsymbol{\theta}$ é possível observar que quando $\boldsymbol{\theta}^{(m)}$ converge para $\hat{\boldsymbol{\theta}}$, o valor de $\hat{\boldsymbol{\beta}}$ pode ser considerado como a estimativa de mínimos quadrados dos parâmetros do seguinte modelo linear

$$\begin{cases} E(\tilde{\mathbf{Z}}) = \mathbf{D}_{\hat{\beta}} \boldsymbol{\beta}, \\ Var(\tilde{\mathbf{Z}}) = (\phi/4d_g)\mathbf{I}. \end{cases} \tag{2.16}$$

Além disso, o efeito da especificação da componente aleatória do modelo no processo de estimação pode ser observado através da matriz $\mathbf{D}(\boldsymbol{\rho})$ no processo (2.15) para $\boldsymbol{\beta}$, e através da matriz $\mathbf{D}(\mathbf{v})$ no processo (2.14) para ϕ. Pode-se observar nas expressões (2.14) e (2.15) que as funções $v(\cdot)$ e $\rho(\cdot)$ atribuem pesos às observações no processo iterativo, os quais podem ser considerados como medidas da qualidade da informação a respeito do vetor de parâmetros $\boldsymbol{\theta}$ contida em cada uma das observações. Assim, os valores de $v^{(m)}(z)$ e $\rho^{(m)}(z)$ são o critério através do qual o modelo considerado determina quais das observações contêm as maiores e melhores informações a respeito dos parâmetros de interesse. Podemos observar que, se a função $g(u)$ é monotonicamente decrescente para $u > 0$, então $v(\cdot)$ e $\rho(\cdot)$ são funções positivas, isto é, $v(z) > 0$ e $\rho(z) > 0$ para todo $z \in I\!R$. Além disso, $v(\cdot)$ e $\rho(\cdot)$ são funções pares de z, isto é, $v(z) = v(-z)$ e $\rho(z) = \rho(-z)$ para todo $z \in I\!R$, o que implica que os valores de $v(z)$ e $\rho(z)$ dependem da magnitude e não do sinal de z.

Os modelos em que a distribuição assumida para a resposta tem caudas mais pesadas do que a normal atribuem pesos "pequenos" às observações nas quais o valor de $|z|$ é "grande". Desta maneira, tais modelos podem reduzir e controlar a influência das observações extremas. Em contraste, o modelo de resposta normal atribui pesos iguais a todas as observações, ou seja, não considera a informação fornecida por z, sendo, em conseqüência, muito sensível às observações aberrantes. Portanto, estudar a robustez do modelo considerado frente às observações extremas ou aberrantes é equivalente a estudar o comportamento das funções $v(\cdot)$ e $\rho(\cdot)$ induzidas pela distribuição assumida para a variável resposta. É fácil mostrar que em distribuições como t-Student, logística tipo II, t-Student generalizada ($s \geq r - 2 > 0$) e exponencial potência ($k > 0$), os valores das funções $v(\cdot)$ e $\rho(\cdot)$ decrescem à medida que $|z|$ aumenta. Nas Figuras 2.7 a 2.11 apresentamos os gráficos da função $\rho(\cdot)$ para estas distribuições; em todos os casos a função de pesos considerada é comparada com a função de pesos da distribuição normal, a qual é representada pela linha pontilhada. Nas Figuras 2.8 a 2.11 é possível observar que o peso $\rho(z)$ decresce à medida que $|z|$ aumenta. Portanto, em modelos baseados nestas distribuições, as observações com valor de $|z|$ "grande" terão peso "pequeno" no processo de estimação e de inferência. Para a distribuição t-Student este comportamento foi descrito em Arellano-Valle (1994) e Kowalski et *al* (1999).

Na Figura 2.7 é possível observar que no caso da distribuição exponencial potência com $k < 0$, o peso $\rho(z)$ cresce com o valor de $|z|$. Vale salientar que quando $k < 0$ a distribuição exponencial potência tem caudas mais leves do que a normal. Para a distribuição t-Student observamos que, à medida que o número de graus de liberdade aumenta, o peso $\rho(z)$ tende para 1 para todo z, o que indica que nesses casos o modelo é mais sensível a observações extremas. Este comportamento é esperado pelo fato da distribuição t-Student tender para a distribuição normal quando o número de graus de liberdade tende para infinito. Para a distribuição exponencial potência é observado um comportamento análogo ao da distribuição t-Student quando k decresce para zero, caso em que esta distribuição tende para a normal.

Desta maneira, o estudo do comportamento das funções $v(\cdot)$ e $\rho(\cdot)$ pode ser uma ferramenta para o pesquisador na etapa da formulação do modelo, pois o comportamento destas funções descreve a robustez do modelo considerado frente a observações extremas ou aberrantes. Além disso, na classe simétrica existe grande variedade de distribuições, o que permite ao pesquisador selecionar uma distribuição para modelar a resposta que possua ao mesmo tempo, robustez frente às observações extremas e grande poder descritivo do comportamento da variável de interesse na população de estudo.

Figura 2.7 *Comportamento da função de pesos $\rho(z)$ para modelos em que o erro segue distribuição exponencial potência de parâmetro k.*

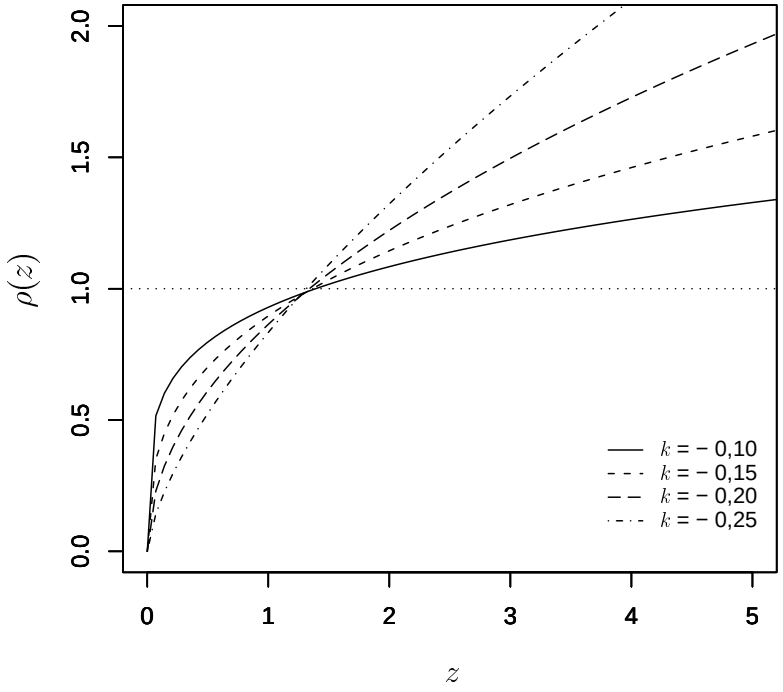


Figura 2.8 *Comportamento da função de pesos $\rho(z)$ para modelos em que o erro segue distribuição t-Student com ν graus de liberdade.*

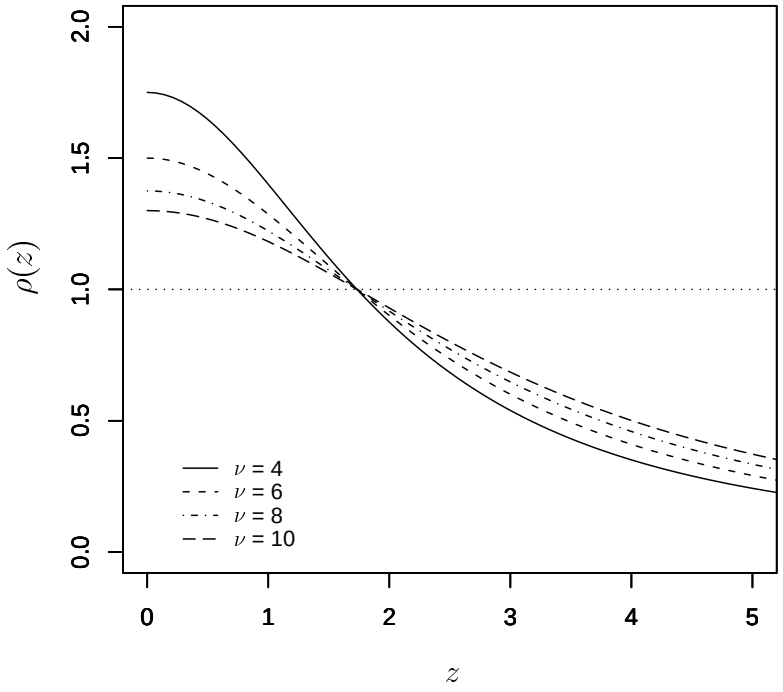


Figura 2.9 *Comportamento da função de pesos $\rho(z)$ para modelos em que o erro segue distribuição t-Student generalizada com parâmetros $s = 4$ e r.*

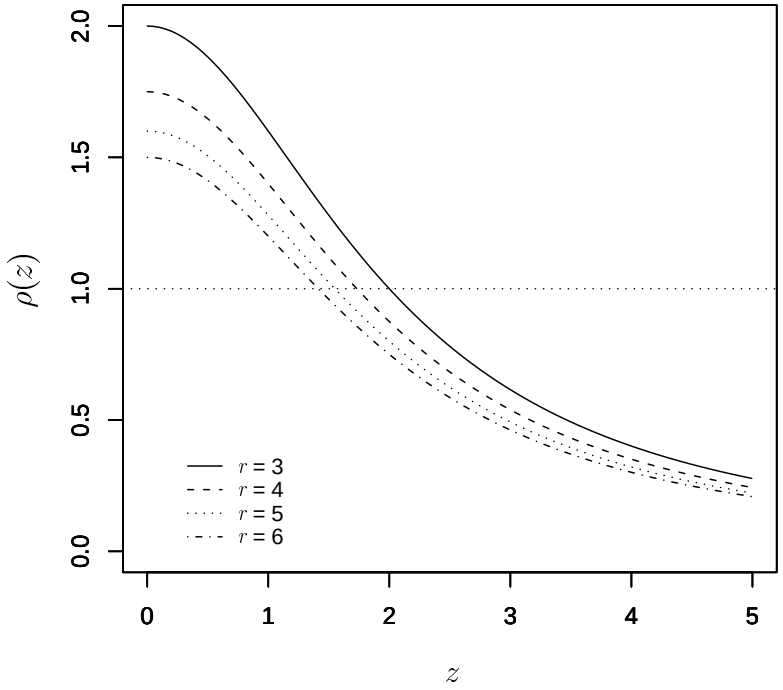


Figura 2.10 *Comportamento da função de pesos $\rho(z)$ para modelos em que o erro segue distribuição exponencial potência de parâmetro k.*

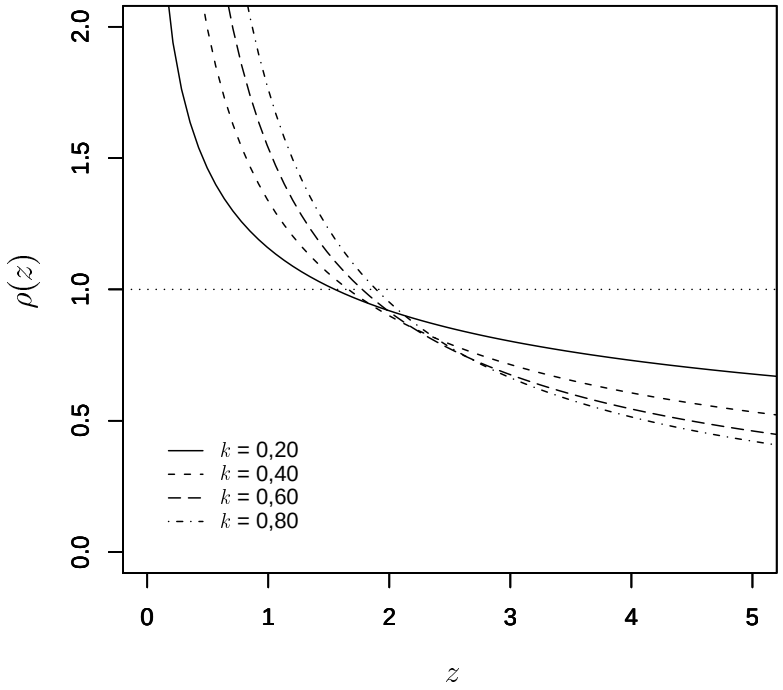
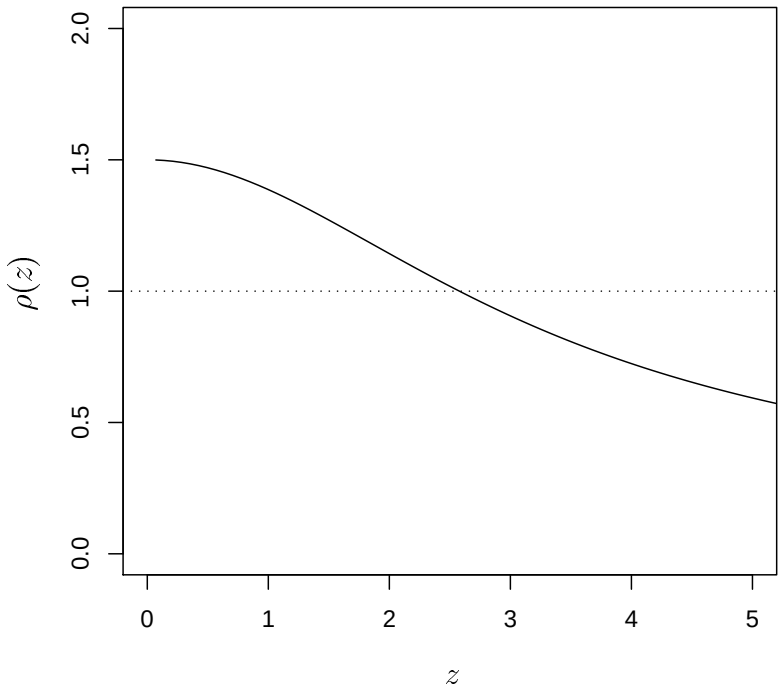


Figura 2.11 *Comportamento da função de pesos $\rho(z)$ para modelos em que o erro segue distribuição logística tipo II.*



CAPÍTULO 3

Resíduos

Os métodos de diagnóstico são usados para avaliar a adequabilidade de modelos de regressão aplicados com o objetivo de descrever e/ou explicar o comportamento de uma variável de interesse. Para esta avaliação podem ser usadas medidas denominadas resíduos, as quais têm como objetivo medir a qualidade do ajuste do modelo para cada observação na amostra. Os resíduos proporcionam evidência a respeito de observações cujo comportamento não é descrito completamente pelo modelo ajustado, avaliando se a diferença entre o observado e o explicado pelo modelo é estatisticamente significativa. Embora os resíduos meçam a qualidade do ajuste para cada observação separadamente, estes podem ser usados conjuntamente para avaliar a qualidade do ajuste global do modelo considerado. Por exemplo, podem ser utilizados para avaliar o efeito da inclusão ou exclusão de parâmetros do modelo ou, para estudar os possíveis efeitos não-lineares das variáveis explicativas. Além disso, os resíduos podem ser usados para construir testes estatísticos cujo objetivo é avaliar a adequabilidade global do modelo (veja Cox e Snell, 1971).

Em resumo, os resíduos são uma ferramenta muito importante no processo de validação do modelo, o que motiva o estudo das propriedades estatísticas das definições de resíduo existentes, bem como a definição de novos resíduos que sejam mais eficazes para identificar falta de ajuste no modelo estimado. Galea, Paula e Cysneiros (2005) apresentam um resíduo para os modelos simétricos de regressão baseado na definição geral de resíduos descrita em Cox e Snell (1968). A metodologia descrita por Cox e Snell permite definir resíduos para modelos em

que as respostas são assumidas independentes. Embora esta metodologia possa ser aplicada em diversos métodos de estimação, os autores focam sua atenção em situações em que os parâmetros são estimados através do método de máxima verossimilhança, caso para o qual é apresentado um procedimento para calcular os dois primeiros momentos da distribuição dos resíduos. Para a aplicação desta definição de resíduo assumimos que os valores da variável de interesse podem ser escritos na seguinte forma

$$y_i = m_i(\boldsymbol{\theta}, \epsilon_i), \qquad i = 1, \ldots, n, \tag{3.1}$$

em que $m_1, \ldots, m_n$ são funções conhecidas e $\epsilon_1, \ldots, \epsilon_n$ são variáveis aleatórias contínuas, independentes e identicamente distribuídas. Adicionalmente, assumimos que a distribuição das ϵ_i é conhecida e completamente especificada. Supondo uma solução única para ϵ_i em (3.1) podemos escrever

$$\epsilon_i = k_i(y_i, \boldsymbol{\theta}), \qquad i = 1, \ldots, n.$$

Sendo assim, o resíduo de Cox e Snell pode ser definido como

$$\hat{\epsilon}_i = k_i(y_i, \hat{\boldsymbol{\theta}}), \qquad i = 1, \ldots, n.$$

Da expressão (2.3) temos que para os modelos simétricos de regressão o resíduo de Cox e Snell corresponde a $\hat{z}_i = (y_i - \mu_i(\hat{\boldsymbol{\beta}}))/\hat{\phi}^{1/2}$. Galea, Paula e Cysneiros (2005) apresentam uma versão padronizada deste resíduo, denotada t_{r_i}, que pode ser expressa na seguinte forma

$$t_{r_i} = \frac{\hat{z}_i}{\xi^{1/2}[1 - (4d_g\xi)^{-1}\hat{h}_{ii}]^{1/2}},$$

em que h_{ii} é o i-ésimo elemento da diagonal da matriz $\mathbf{H} = \mathbf{D}_\beta(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{D}_\beta^T$. No entanto, neste capítulo, propomos uma definição geral de resíduos para os modelos simétricos de regressão, em que o resíduo t_{r_i} é um caso particular. As propriedades estatísticas mais importantes dos resíduos propostos são estudadas analiticamente usando o procedimento descrito em Cox e Snell (1968) e empiricamente através de simulação de Monte Carlo.

3.1 Definição de resíduos

Consideramos resíduos definidos como $t(\hat{z}_i)$, para alguma função $t(\cdot)$ ímpar e duplamente diferenciável, em que $\hat{z}_i$ corresponde ao resíduo obtido aplicando aos modelos simétricos de regressão a definição geral de resíduos descrita em Cox e Snell (1968). A transformação $t(\cdot)$ preserva as principais propriedades estatísticas de $\hat{z}_i$; por exemplo, se $\hat{z}_i$ segue distribuição simétrica em torno de zero, então o resíduo $t(\hat{z}_i)$ também seguirá distribuição simétrica em torno de zero. Dado que o resíduo $t(\hat{z}_i)$ somente depende da estrutura do modelo de regressão através dos valores de $\mu_i(\hat{\boldsymbol{\beta}})$ e $\hat{\phi}$, este resíduo é invariante sob transformações invertíveis do vetor de parâmetros $\boldsymbol{\beta}$, pois $\mu_i(\hat{\boldsymbol{\beta}})$ e $\hat{\phi}$ são invariantes a essas transformações. Além disso, é possível selecionar uma função $t(\cdot)$ que permita, pelo menos para grandes amostras, que a distribuição de $t(\hat{z}_i)$ apresente uma forma conhecida e fácil de interpretar em gráficos de resíduos. Assim, para uma adequada escolha da transformação $t(\cdot)$, o resíduo $t(\hat{z}_i)$ pode apresentar propriedades estatísticas desejáveis.

A forma da função $t(\cdot)$ pode ser escolhida com base na estatística do teste para avaliar se a i-ésima observação pode ser considerada como outlier. Se esta estatística assume valores "grandes" na região de rejeição, é claro que quando aumenta o valor desta estatística, diminui a probabilidade de errar ao concluir que o comportamento desta observação não é descrito adequadamente pelo modelo ajustado. Assim, o valor desta estatística pode ser usado como uma medida da qualidade do ajuste do modelo nesta observação, desempenhando o mesmo papel que um resíduo. Desta maneira, podemos definir resíduos como funções da estatística do teste para identificar outliers. Tais funções, além de preservar a informação fornecida pela estatística do teste, devem permitir que o resíduo obtido apresente características desejáveis, como, por exemplo, média zero e variância um. No caso dos modelos simétricos de regressão pode-se mostrar que resíduos obtidos desta forma podem ser escritos como $t(\hat{z}_i)$, com $t(\cdot)$ uma função ímpar e duplamente diferenciável.

De maneira geral, testar se uma observação é outlier consiste em avaliar se o modelo ajustado descreve adequadamente o comportamento desta observação, o que é equivalente a testar se esta observação pode ser considerada como "gerada" pela distribuição assumida para a variável resposta, que, neste caso, é

completamente especificada pelos valores de μ e ϕ. Sendo assim, testar se a i-ésima observação pode ser considerada como outlier é avaliar a seguinte hipótese

$$\begin{aligned} H_0 &: \ \mu_i = \mu_i^\circ \\ H_1 &: \ \mu_i \neq \mu_i^\circ \end{aligned} \tag{3.2}$$

Para avaliar a hipótese descrita em (3.2) podemos construir um teste com um nível de significância α $(0 < \alpha < 1)$ para rejeitar H_0 quando $\delta(y_i) = 1$, em que $\delta(\cdot)$ é definido como

$$\delta(y_i) = \begin{cases} 1, & \text{se } y_i \notin (a, b), \\ 0, & \text{em outro caso.} \end{cases}$$

Então, a função de poder para este teste é expressa da seguinte forma

$$\pi(\Delta) = P[\, y_i \notin (a, b) \,|\, \mu_i = \mu_i^\circ + \Delta \,], \qquad \Delta \in I\!R,$$

com a restrição $\pi(0) = \alpha$, que permite que o teste possa ser aplicado com o nível de significância α. Desta maneira, para cada intervalo (a, b) que satisfaz $\pi(0) = \alpha$, temos um teste de nível α para avaliar a hipótese (3.2). Para selecionar o intervalo (a, b) no qual estará baseado o teste, podemos usar por exemplo, a estatística da razão de verossimilhanças. Entretanto, podemos também definir um teste da seguinte maneira

$$\delta(y_i) = \begin{cases} 1, & \text{se } y_i \notin (q_{(\alpha/2)}, q_{(1-\alpha/2)}) \ , \\ 0, & \text{em outro caso,} \end{cases} \tag{3.3}$$

em que $q_{(1-\alpha/2)}$ é o quantil $100(1-\alpha/2)\%$ da distribuição $S(\mu_i^\circ, \phi)$. Assim, podemos definir resíduos como funções da estatística do teste para avaliar a hipótese em (3.2), em que μ_i° é substituído por $\mu_i(\hat{\boldsymbol{\beta}})$ e ϕ por $\hat{\phi}$; ou seja, μ_i° é representado pelo ajuste do modelo. A idéia por trás desta definição de resíduo é medir a diferença entre os valores observado e predito para a i-ésima observação usando como métrica a estatística do teste para avaliar a hipóteses em (3.2). Como a distribuição de y_i é simétrica, o teste da razão de verossimilhanças e o teste definido em (3.3) determinam o mesmo intervalo (a, b). Porém, os resíduos baseados nestes testes apresentam características diferentes.

3.1.1 Resíduo componente de desvio

Aplicando o raciocínio descrito acima aos modelos lineares generalizados, e usando a estatística da razão de verossimilhanças para avaliar a hipótese em

(3.2), obtemos o resíduo *componente de desvio* proposto por Pregibon (1981), que devido a suas propriedades estatísticas é o mais utilizado nesta classe de modelos (veja Pierce e Schafer, 1986; Davison e Gigli, 1989; McCullagh e Nelder, 1989). Além disso, este resíduo tem sido estendido a outros modelos de regressão, por exemplo, Svetliza e Paula (2003) estudam o comportamento do resíduo *componente de desvio* em modelos não-lineares de resposta binomial negativa, encontrando grande proximidade entre a distribuição deste resíduo e a normal padrão. A seguir, definimos um resíduo para os modelos simétricos de regressão baseado na estatística da razão de verossimilhanças generalizada (LR) (veja Gigli, 1987). A estatística LR para avaliar a hipótese em (3.2) pode ser expressa como

$$LR = 2\big[\, l(y_i\,;\tilde{\mu}_i,\phi) - l(y_i\,;\mu_i^{\circ},\phi)\,\big],$$

em que $l(y_i\,;\mu_i,\phi)$ é a contribuição da i-ésima observação ao logaritmo da função de verossimilhança de $\boldsymbol{\theta}$, e $\tilde{\mu}_i$ é o valor de μ que maximiza $l(y_i\,;\mu,\phi)$. Mostra-se facilmente que o valor de $\tilde{\mu}_i$ é igual a y_i. Assim, da equação (2.6) e substituindo o valor de $\tilde{\mu}_i$ na equação acima temos o seguinte

$$LR = 2\Big\{\, \log\big[g(0)\big] - \log\big[g(z_i^{\circ\,2})\big]\Big\},$$

em que $g(\cdot)$ é a função geradora de densidades e $z_i^{\circ} = (y_i - \mu_i^{\circ})/\phi^{1/2}$. Então, podemos definir o resíduo $t_D(\hat{z}_i)$ baseado na estatística LR, onde ϕ é substituído pela sua estimativa de máxima verossimilhança no modelo de interesse, obtendo

$$t_D(\hat{z}_i) = \text{sinal}(\hat{z}_i)\Big\{2\log\big[g(0)\big] - 2\log\big[g(\hat{z}_i^2)\big]\Big\}^{\frac{1}{2}},$$

em que $\hat{z}_i = (y_i - \hat{\mu}_i)/\hat{\phi}^{1/2}$ e $\text{sinal}(x) = I_{[\,0\,,\,\infty)}(x) - I_{(-\infty\,,\,0\,)}(x)$, com $I_A(x)$ a função indicadora. No Quadro 3.1 apresentamos as expressões para o cálculo do resíduo $t_D(\hat{z}_i)$ para algumas distribuições pertencentes à classe simétrica. Podemos observar que o resíduo $t_D(\hat{z}_i)$ assume valores no intervalo $(-\infty,\infty)$. Temos que $t_D(\cdot)$ é uma função ímpar e duplamente diferenciável; em conseqüência, podemos usar o Lema 1 do Apêndice A para mostrar que a simetria de $t_D(\hat{z}_i)$ depende da simetria de $\hat{z}_i$, ou seja, se $\hat{z}_i$ é simétrico em torno de zero, então o resíduo $t_D(\hat{z}_i)$ também é simétrico em torno de zero. Notamos também que a quantidade $MQ(\hat{\boldsymbol{\beta}})$ pode ser escrita na seguinte forma

$$MQ(\hat{\boldsymbol{\beta}}) = \sum_{i\,=\,1}^{n} t_D^2(\hat{z}_i),$$

Quadro 3.1 *Expressões para o resíduo $t_D(\hat{z}_i)$ em algumas distribuições pertencentes à classe simétrica.*

Distribuição	$t_D(\hat{z}_i)$
Normal	$\hat{z}_i$
t-Student	$\text{sinal}(\hat{z}_i)\left\{(\nu+1)\log\left(1+\frac{\hat{z}_i^2}{\nu}\right)\right\}^{\frac{1}{2}}$
t-Student generalizada	$\text{sinal}(\hat{z}_i)\left\{(r+1)\log\left(1+\frac{\hat{z}_i^2}{s}\right)\right\}^{\frac{1}{2}}$
Logística tipo II	$\text{sinal}(\hat{z}_i)\left\{4\log\left[\,\frac{1}{2}(1+\mathrm{e}^{\lvert\hat{z}_i\rvert})\,\right]-2\lvert\hat{z}_i\rvert\right\}^{\frac{1}{2}}$
Exponencial potência	$\text{sinal}(\hat{z}_i)\,\lvert\hat{z}_i\rvert^{1/(k+1)}$

sendo $MQ(\cdot)$ a função definida na expressão (2.8). Desta maneira, a quantidade $MQ(\hat{\boldsymbol{\beta}})$, quando apropriadamente padronizada, pode ser usada como uma medida da qualidade do ajuste global do modelo considerado.

Dado que a função $t_D(\cdot)$ é contínua e os estimadores de máxima verosimilhança de $\boldsymbol{\theta}$ são consistentes, podemos usar o Lema 3 do Apêndice A para mostrar que o resíduo $t_D(\hat{z}_i)$ converge em distribuição para $t_D(z_i)$, em que $z_i = (y_i - \mu_i)/\phi^{1/2}$. De acordo com o Lema 4 do Apêndice A, as funções de distribuição e de densidade de $t_D(z_i)$ são dadas por

$$F_{t_D}(z) = \begin{cases} 1 - F(z_g(z)) & \text{se } z < 0. \\ F(z_g(z)) & \text{se } z \geq 0, \end{cases} \tag{3.4}$$

e

$$f_{t_D}(z) = \frac{g(0)\exp(-z^2/2)\,\lvert z\rvert}{z_g(z)\,v(z_g(z))}, \quad z \neq 0, \tag{3.5}$$

em que $F(\cdot)$ é a função de distribuição acumulada de z_i e $z_g(z) > 0$ é uma constante tal que $g(z_g^2(z)) = g(0)\exp(-z^2/2)$. Expressões de $z_g(z)$ para algumas distribuições pertencentes à classe simétrica podem ser encontradas no Quadro 3.2. Do Lema 1 do Apêndice A, podemos concluir que a distribuição de $t_D(z_i)$ é simétrica em torno de zero. Além disso, no Quadro 3.1 pode-se observar que $t_D(z_i)$ segue distribuição normal padrão se y_i é normalmente distribuída. Contudo, no caso em que y_i segue outra distribuição simétrica, o comportamento de $t_D(z_i)$ depende das funções $z_g(\cdot)$ e $v(\cdot)$. Pode-se mostrar que em modelos em que

o erro segue distribuição t-Student generalizada de parâmetros r e s, a densidade de $t_D(z_i)$ não depende de s (veja Lema 5 do Apêndice A), o que implica que a distribuição deste resíduo é equivalente à distribuição de $t_D(z_i)$ quando o erro segue distribuição t-Student com $\nu = r$ graus de liberdade. Este resultado segue do fato da distribuição t-Student generalizada ser equivalente à distribuição t-Student com ν graus de liberdade quando $r = s = \nu$. Como ilustração, apresentamos nas Figuras 3.1 a 3.3 o gráfico da função de densidade de $t_D(z_i)$ para algumas distribuições da classe simétrica. Em todos os casos, a densidade considerada é comparada com a densidade da distribuição normal padrão, a qual é representada pela linha pontilhada.

Quadro 3.2 *Expressões para $z_g(z)$ em algumas distribuições pertencentes à classe simétrica.*

Distribuição	$z_g(z)$
Normal	$\lvert z\rvert$
t-Student	$\left\{\nu\left[\exp\left(z^2/(\nu+1)\right)-1\right]\right\}^{1/2}$
t-Student generalizada	$\left\{s\left[\exp\left(z^2/(r+1)\right)-1\right]\right\}^{1/2}$
Logística tipo II	$-z^2/2 - 2\log\left\{1-\left[1-\exp(-z^2/2)\right]^{1/2}\right\}$
Exponencial potência	$\lvert z\rvert^{1+k}$

Figura 3.1 *Função de densidade de $t_D(z_i)$ em modelos em que o erro segue distribuição t-Student generalizada com parâmetro $r = 5$ (esquerda) e $r = 8$ (direita).*

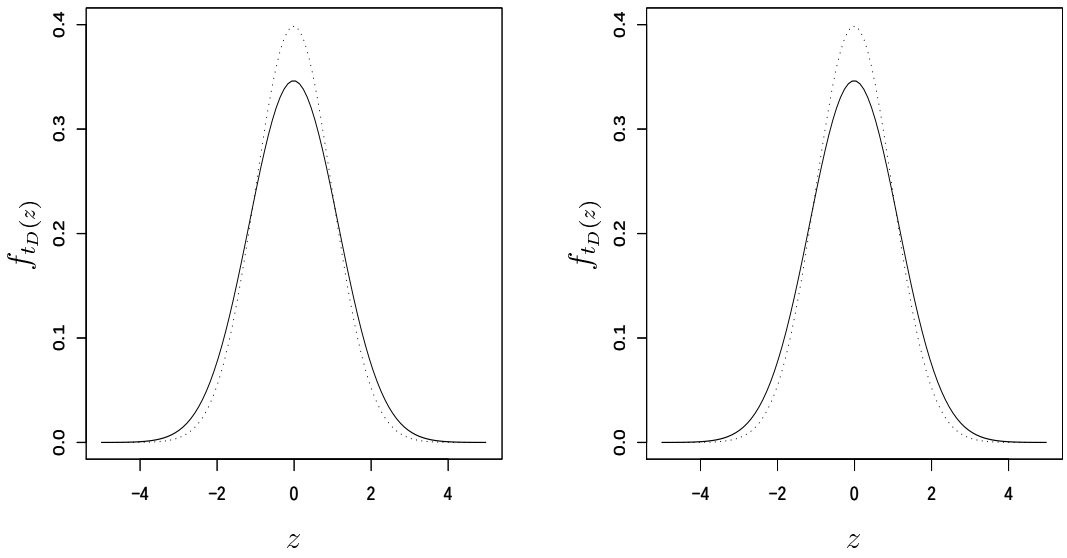


Figura 3.2 *Função de densidade de $t_D(z_i)$ em modelos em que o erro segue distribuição t-Student com $\nu = 4$ (esquerda) e $\nu = 10$ (direita) graus de liberdade.*

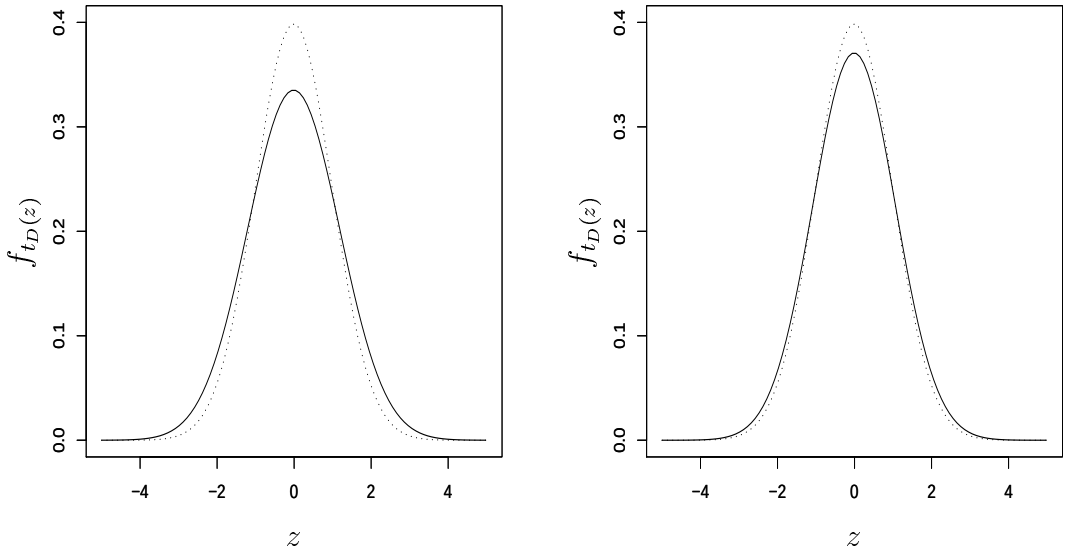
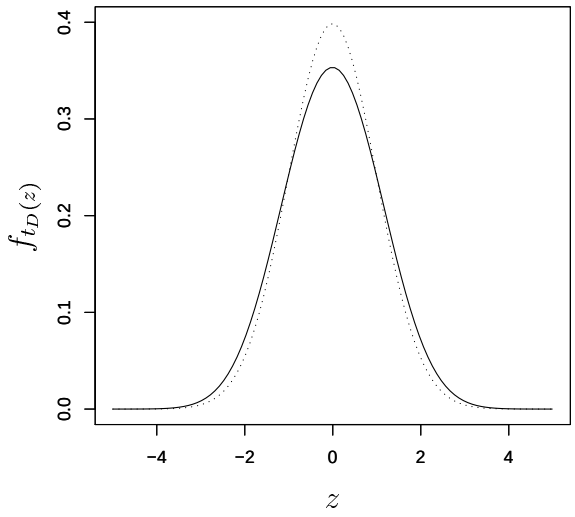


Figura 3.3 *Função de densidade de $t_D(z_i)$ em modelos em que o erro segue distribuição logística tipo II.*



Apesar de que, em geral, $t_D(z_i)$ não segue distribuição normal, em alguns casos as caudas desta distribuição podem ser usadas como referência para o resíduo $t_D(\hat{z}_i)$. Por exemplo, quando o erro do modelo segue distribuições como t-Student, t-Student generalizada, exponencial potência, e logística tipo II, $t_D(z_i)$ pode ser padronizado para que as propriedades da sua distribuição estejam mais próximas da normal. Esta padronização é aplicada dividindo pelo desvio padrão σ_g, em que $\sigma_g^2 = E\big(2\log[g(0)/g(z^2)]\big)$, com $z \sim S(0,1)$. No Quadro 3.3 apresentamos uma comparação das caudas da distribuição de $t_D^*(z_i) = t_D(z_i)/\sigma_g$ com as caudas da normal. Em todos os casos apresentamos o valor de $F^*_{t_D}(q_{1-\alpha})$, em que $F^*_{t_D}(\cdot)$ é a função de distribuição acumulada de $t_D^*(z_i)$, com $q_{1-\alpha}$ uma constante tal que $\Phi(q_{1-\alpha}) = 1-\alpha$, sendo $\Phi(\cdot)$ a função de distribuição acumulada da normal padrão. Vale salientar que os valores de $F^*_{t_D}(\cdot)$ foram calculados com base na expressão dada em (3.4). No Quadro 3.3 podemos observar que as caudas da distribuição de $t_D^*(z_i)$ e da normal estão muito próximas, o que indica que, para grandes tamanhos da amostra, o resíduo $t_D^*(\hat{z}_i) = t_D(\hat{z}_i)/\sigma_g$ pode ser aplicado usando os quantis da distribuição normal como referência.

Quadro 3.3 *Comparação das caudas da distribuição de* $t_D(z_i)/\sigma_g$ *com as caudas da distribuição normal padrão.*

Distribuição		σ_g^2	$F^*_{t_D}(q_{1-\alpha})$		
			α =0,05	α =0,025	α =0,005
t-Student(ν)	4	1,40186	0,949947	0,975162	0,995199
	5	1,31777	0,949962	0,975099	0,995127
	6	1,26261	0,949972	0,975066	0,995088
	7	1,22369	0,949979	0,975047	0,995064
	8	1,19478	0,949983	0,975036	0,995048
	9	1,17247	0,949989	0,975028	0,995038
	10	1,15473	0,949989	0,975022	0,995030
	11	1,14029	0,949991	0,975018	0,995025
Exponencial Potência(k)	0,0	1,0	0,950000	0,975000	0,995000
	0,1	1,1	0,952009	0,977099	0,995951
	0,2	1,2	0,953716	0,978831	0,996555
	0,3	1,3	0,955556	0,980806	0,997302
	0,4	1,4	0,957249	0,982282	0,997760
	0,5	1,5	0,958563	0,983395	0,998168
	0,6	1,6	0,960433	0,984823	0,998509
	0,7	1,7	0,961857	0,986039	0,998794
	0,8	1,8	0,963444	0,987287	0,998963
	0,9	1,9	0,965167	0,988502	0,999196
t-Student generalizada (r, s)	r =4	1,40186	0,949947	0,975162	0,995199
	5	1,31777	0,949962	0,975099	0,995127
	6	1,26261	0,949972	0,975066	0,995088
	7	1,22369	0,949979	0,975047	0,995064
	8	1,19478	0,949983	0,975036	0,995048
	9	1,17247	0,949989	0,975028	0,995038
	10	1,15473	0,949989	0,975022	0,995030
	11	1,14029	0,949991	0,975018	0,995025
Logística tipo II		1,22701	0,949953	0,975729	0,995714

Nota: Os valores de σ_g^2 foram calculados via integração numérica.

3.1.2 Resíduo quantal

Como a distribuição de y_i é simétrica, o teste dado na equação (3.3) pode ser reescrito como

$$\delta(y_i) = \begin{cases} 1, & \text{se } T \notin (\alpha/2, 1-\alpha/2), \\ 0, & \text{em outro caso,} \end{cases}$$

em que $T = F(z_i^\circ)$. Sob a hipótese nula, T segue distribuição uniforme no intervalo $(0,1)$, porém, é possível usar qualquer distribuição contínua como referência para a estatística T. Por exemplo, se desejamos que a estatística T se comporte de acordo com uma lei cuja função de distribuição acumulada é denotada por $G(\cdot)$, podemos reescrever o teste $\delta(y_i)$ da seguinte forma

$$\delta(y_i) = \begin{cases} 1, & \text{se } T \notin (q^*_{\alpha/2}, q^*_{1-\alpha/2}) \ , \\ 0, & \text{em outro caso,} \end{cases}$$

com $T = G^{-1}(F(z_i^\circ))$, e $q^*_{\alpha/2}$ uma constante tal que $G(q^*_{\alpha/2}) = \alpha/2$. Então, sob a hipótese nula, a estatística T segue uma distribuição caracterizada pela função acumulada $G(\cdot)$. De fato, podemos selecionar a distribuição do erro do modelo como distribuição de referência para a estatística T, o que é feito definindo $G(\cdot) = F(\cdot)$, caso em que o resíduo pode ser escrito como $\hat{z}_i$. No entanto, para facilitar a aplicação e interpretação deste teste, escolhemos a distribuição normal padrão como referência, ou seja, definimos $G(\cdot) = \Phi(\cdot)$. Desta maneira, é possível definir um resíduo baseado na estatística T, substituindo o valor de ϕ pela sua estimativa de máxima verossimilhança no modelo de interesse, obtendo o resíduo $t_Q(\hat{z}_i)$, o qual é dado por

$$t_Q(\hat{z}_i) = \Phi^{-1}[F(\hat{z}_i)].$$

O resíduo $t_Q(\hat{z}_i)$ foi proposto por Dunn e Smyth (1996) para os modelos lineares generalizados. Podemos notar que $t_Q(\cdot)$ é uma função ímpar duplamente diferenciável, portanto, se $\hat{z}_i$ é simétrica em torno de zero então o resíduo $t_Q(\hat{z}_i)$ também é simétrico em torno de zero. Como a função $t_Q(\cdot)$ é contínua e os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ são consistentes, podemos usar o Lema 3 do Apêndice A para mostrar que o resíduo $t_Q(\hat{z}_i)$ converge em distribuição para $t_Q(z_i)$, ou seja, $t_Q(\hat{z}_i)$ segue assintoticamente distribuição normal padrão.

3.2 Avaliação analítica dos resíduos

Na seção anterior foram estudadas as principais características das distribuições assintóticas dos resíduos $t_D(\hat{z}_i)$ e $t_Q(\hat{z}_i)$, as quais podem ser usadas como distribuições de referência para os resíduos quando o tamanho da amostra é "grande". Nesta seção, avaliamos o efeito da variabilidade amostral de $\hat{\boldsymbol{\theta}}$ nos dois primeiros momentos das distribuições dos resíduos, o que é feito usando expansões de Cox e Snell (1968). Através destas expansões obtemos uma padronização para os resíduos com o objetivo de permitir uma maior comparabilidade entre os mesmos. Para isto, notamos que os resíduos podem ser escritos como $t(\hat{z}_i)$, para alguma função $t(\cdot)$ ímpar e duplamente diferenciável. De fato, o resíduo $\hat{z}_i$ também pode ser escrito desta maneira, com $t(z) = z$. Em conseqüência, estudar o comportamento destes resíduos é equivalente a estudar as funções $t(\cdot)$ que os definem. Inicialmente observamos que, como o resíduo $t(z)$ é uma função ímpar de z temos

$$t(z) = -t(-z),$$

derivando ambos os lados desta equação obtemos

$$t'(z) = t'(-z),$$

o que indica que a função $t'(z)$ é par. Derivando de novo ambos os lados da equação temos que

$$t''(z) = -t''(-z),$$

isto é, $t''(z)$ é uma função ímpar. Se z tem distribuição simétrica em torno de zero podemos concluir que $t(z)$ e $t''(z)$ também têm distribuição simétrica em torno de zero, o que implica que, quando existem, $E(t(z)) = E(t''(z)) = 0$. Procedendo de maneira análoga a Cox e Snell (1968), expandimos o resíduo $t(\hat{z}_i)$ em série de Taylor até segunda ordem em torno de $\boldsymbol{\theta}$, obtendo

$$\begin{aligned} t(\hat{z}_i) = t(z_i) &+ \sum_{r=1}^{p}(\hat{\beta}_r - \beta_r)M_r^{(i)} + \frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p}(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)M_{rs}^{(i)} \\ &+ (\hat{\phi} - \phi)M_{\phi}^{(i)} + \sum_{r=1}^{p}(\hat{\beta}_r - \beta_r)(\hat{\phi} - \phi)M_{r\phi}^{(i)} + \frac{1}{2}(\hat{\phi} - \phi)^2 M_{\phi\phi}^{(i)}, \end{aligned} \tag{3.6}$$

em que

$$M_r^{(i)} = \frac{\partial t(z_i)}{\partial \beta_r} = -\frac{t'(z_i)}{\sqrt{\phi}}\frac{\partial \mu_i}{\partial \beta_r}, \tag{3.7}$$

$$M_\phi^{(i)} = \frac{\partial t(z_i)}{\partial \phi} = -\frac{t'(z_i)z_i}{2\phi}, \tag{3.8}$$

$$M_{r\phi}^{(i)} = \frac{\partial^2 t(z_i)}{\partial \beta_r \partial \phi} = \frac{1}{2\phi^{3/2}}\frac{\partial \mu_i}{\partial \beta_r}\big(t''(z_i)z_i + t'(z_i)\big),$$

$$M_{rs}^{(i)} = \frac{\partial^2 t(z_i)}{\partial \beta_r \partial \beta_s} = \frac{t''(z_i)}{\phi}\frac{\partial \mu_i}{\partial \beta_r}\frac{\partial \mu_i}{\partial \beta_s} - \frac{t'(z_i)}{\sqrt{\phi}}\frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}, \tag{3.9}$$

$$M_{\phi\phi}^{(i)} = \frac{\partial^2 t(z_i)}{\partial^2 \phi} = \frac{1}{4\phi^2}\big[t''(z_i)z_i^2 + 3t'(z_i)z_i\big]. \tag{3.10}$$

De maneira similar à expressão (25) em Cox e Snell (1968), aplicamos esperança a ambos os lados de (3.6), obtendo

$$\begin{aligned} E(t(\hat{z}_i)) = &\sum_{r=1}^{p} E\big(\hat{\beta}_r - \beta_r\big)E\big(M_r^{(i)}\big) + \sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}E\Big(M_r^{(i)}u_s^{(i)} + \frac{1}{2}M_{rs}^{(i)}\Big) \\ &+ E\big(\hat{\phi} - \phi\big)E\big(M_\phi^{(i)}\big) + \mathrm{K}_{\phi\phi}^{-1}E\Big(M_\phi^{(i)}u_\phi^{(i)} + \frac{1}{2}M_{\phi\phi}\Big), \end{aligned} \tag{3.11}$$

em que $E\big(\hat{\beta}_r - \beta_r\big)$ e $E\big(\hat{\phi} - \phi\big)$ são os viéses de segunda ordem dos estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ, respectivamente (veja Cordeiro et *al*, 2000), I^{rs} é o (r,s)-elemento da inversa da matriz de informação esperada de $\boldsymbol{\beta}$, e $u_\phi^{(i)}$ e $u_s^{(i)}$ são dados por

$$u_s^{(i)} = \frac{v(z_i)z_i}{\sqrt{\phi}}\frac{\partial \mu_i}{\partial \beta_s},$$

e

$$u_\phi^{(i)} = -\frac{1}{2\phi} + \frac{v(z_i)z_i^2}{2\phi}.$$

Podemos usar o Lema 2 do Apêndice A para mostrar que $M_\phi^{(i)}$, $M_\phi^{(i)}u_\phi^{(i)}$, $M_r^{(i)}u_s^{(i)}$ e $M_{\phi\phi}^{(i)}$ são funções ímpares de z_i. Além disso, como z_i tem distribuição simétrica em torno de zero podemos concluir que, quando existem,

$$E\big(M_\phi^{(i)}\big) = E\big(M_\phi^{(i)}u_\phi^{(i)}\big) = E\big(M_{\phi\phi}^{(i)}\big) = E\big(M_r^{(i)}u_s^{(i)}\big) = 0. \tag{3.12}$$

Então, substituindo as quantidades em (3.12) na equação (3.11) obtemos

$$E(t(\hat{z}_i)) = \sum_{r=1}^{p} E\big(\hat{\beta}_r - \beta_r\big)E\big(M_r^{(i)}\big) + \frac{1}{2}\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}E\Big(M_{rs}^{(i)}\Big). \tag{3.13}$$

Das equações (3.7) e (3.9) temos que

$$E\big(M_r^{(i)}\big) = -\frac{E(t'(z_i))}{\sqrt{\phi}}\frac{\partial \mu_i}{\partial \beta_r} \tag{3.14}$$

e

$$E\Big(M_{rs}^{(i)}\Big) = -\frac{E(t'(z_i))}{\sqrt{\phi}}\frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}. \tag{3.15}$$

Substituindo (3.14) e (3.15) na equação (3.13) obtemos

$$E(t(\hat{z}_i)) = \frac{E(t'(z_i))}{\sqrt{\phi}}\sum_{r=1}^{p} E\big(\hat{\beta}_r - \beta_r\big)\frac{\partial \mu_i}{\partial \beta_r} - \frac{E(t'(z_i))}{2\sqrt{\phi}}\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}\frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}. \tag{3.16}$$

Finalmente, substituindo na equação (3.16) o viés de segunda ordem dos estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ dado no Apêndice B, temos que até ordem n^{-1} a esperança de $t(\hat{z}_i)$ pode ser escrita como

$$E(t(\hat{z}_i)) = \frac{E(t'(z_i))}{\sqrt{\phi}}\Big[\eta_i - \mathbf{d}_i^T(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{D}_\beta^T\eta_i\Big],$$

em que $\eta_i = -\dfrac{\phi}{8d_g}\mathrm{tr}\{(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{D}_{\beta\beta}(i)\}$ e $\mathbf{d}_i = \Big(\dfrac{\partial \mu_i}{\partial \beta_1}, \ldots, \dfrac{\partial \mu_i}{\partial \beta_p}\Big)^T$, com $\mathbf{D}_{\beta\beta}(i) = \partial^2\mu_i/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$. Assim, em forma matricial podemos escrever

$$\begin{aligned} E(\mathbf{t}) &= \frac{E\big(t'(z)\big)}{\sqrt{\phi}}(\mathbf{I}-\mathbf{H})\boldsymbol{\eta}, \\ &= \frac{E\big(t'(z)\big)}{\sqrt{\phi}}E(\mathbf{r}), \end{aligned} \tag{3.17}$$

com $\mathbf{t} = (t(\hat{z}_1), \ldots, t(\hat{z}_n))^T$, $\mathbf{r} = (\mathbf{y}-\hat{\boldsymbol{\mu}})$ o resíduo ordinário, $\mathbf{H} = \mathbf{D}_\beta(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{D}_\beta^T$ e $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$. Em (3.17) podemos observar que a média dos resíduos depende da estrutura do modelo de regressão através da matriz $\mathbf{H}$ e do vetor $\boldsymbol{\eta}$. Além disso, depende também da distribuição da variável resposta por meio dos valores de $E\big(t'(z)\big)$ e $4d_g$. Para modelos com componentes sistemáticas lineares temos que $\mathbf{D}_{\beta\beta}(i)=0$ para todo i, o que implica que $E(t(\hat{z}_i)) = 0$ para toda função $t(\cdot)$ ímpar e duplamente diferenciável. Desta maneira, os resíduos $t_D(\hat{z}_i)$, $t_Q(\hat{z}_i)$ e $\hat{z}_i$ têm média zero. Entretanto, para modelos com componentes sistemáticas não-lineares, a média dos resíduos depende da curvatura intrínseca do modelo através das matrizes $\mathbf{D}_{\beta\beta}(i)$, $i = 1, \ldots, n$. Finalmente, de (3.17) é possível concluir que,

se $\hat{r}_i = (y_i - \hat{\mu}_i)$ tem média zero, então para toda função $t(\cdot)$ ímpar, o resíduo $t(\hat{z}_i)$ também tem média zero.

Da mesma maneira, podemos estudar o comportamento do segundo momento da distribuição dos resíduos. Assim, até ordem n^{-1} temos que

$$\begin{aligned} E\big(t^2(\hat{z}_i)\big) = & E\big(t^2(z_i)\big) + 2\sum_{r=1}^{p} E\big(\hat{\beta}_r - \beta_r\big) E\big(t(z_i) M_r^{(i)}\big) + 2E\big(\hat{\phi} - \phi\big) E\big(t(z_i) M_\phi^{(i)}\big) \\ & + 2\sum_{r=1}^{p}\sum_{s=1}^{p} I^{\,rs} E\Big(t(z_i) M_r^{(i)} u_s^{(i)} + \frac{1}{2} M_r^{(i)} M_s^{(i)} + \frac{1}{2} t(z_i) M_{rs}^{(i)}\Big) \\ & + 2\,\mathrm{K}_{\phi\phi}^{-1} E\Big(t(z_i) M_\phi^{(i)} u_\phi^{(i)} + \frac{1}{2} M_\phi^{(i)\,2} + \frac{1}{2} t(z_i) M_{\phi\phi}^{(i)}\Big). \end{aligned} \tag{3.18}$$

Do Lema 2 do Apêndice A podemos concluir que $t(z_i) M_r^{(i)}$ e $t(z_i) t'(z_i)$ são funções ímpares de z_i, o que implica que, quando existem,

$$E\big(t(z_i) M_r^{(i)}\big) = E\big(t(z_i) t'(z_i)\big) = 0. \tag{3.19}$$

Substituindo (3.19), o viés de segunda ordem do estimador de máxima verossimilhança de ϕ (veja Apêndice B) e o valor de $\mathrm{K}_{\phi\phi}$ na equação (3.18) obtemos

$$\begin{aligned} E\big(t^2(\hat{z}_i)\big) = & \, Var\big(t(z_i)\big) - \frac{2}{\phi}\sum_{r=1}^{p}\sum_{s=1}^{p} I^{\,rs} \frac{\partial \mu_i}{\partial \beta_r}\frac{\partial \mu_i}{\partial \beta_s} E\Big(t(z_i) t'(z_i) v(z_i) z_i\Big) \\ & + \frac{1}{\phi}\sum_{r=1}^{p}\sum_{s=1}^{p} I^{\,rs} \frac{\partial \mu_i}{\partial \beta_r}\frac{\partial \mu_i}{\partial \beta_s} E\Big(t'^{\,2}(z_i) + t''(z_i) t(z_i)\Big) + \frac{c_g}{n}, \end{aligned} \tag{3.20}$$

em que c_g é uma constante que depende do número de parâmetros no modelo, da distribuição da resposta e da função $t(\cdot)$. De uma forma geral, c_g pode ser expresso como

$$\begin{aligned} c_g = & -a_g E\Big(t(z_i) t'(z_i) z_i\Big) + \frac{1}{4f_g - 1} E\Big(t(z_i) t'(z_i) z_i\big(5 - 2v(z_i) z_i^2\big)\Big) \\ & + \frac{1}{4f_g - 1} E\Big(\big(t'^{\,2}(z_i) + t''(z_i) t(z_i)\big) z_i^2\Big), \end{aligned}$$

sendo a_g uma constante definida no Apêndice B. Quando $t(\cdot)$ é a função identidade temos que c_g adota a seguinte expressão

$$c_g = -a_g \xi + \frac{6\xi + 2\alpha_{1,3}}{4f_g - 1}.$$

Por outro lado, para $t(\cdot) = t_D(\cdot)$ temos que

$$c_g = -a_g + \frac{5 - 8f_g - \alpha_{2,2}}{4f_g - 1},$$

com $\alpha_{2,2}$ como definido no Apêndice B. Assim, quando y_i segue distribuição normal, a constante c_g é igual à dimensão do vetor de parâmetros $\boldsymbol{\beta}$.

Substituindo o valor de I^{rs} em (3.20) temos que

$$\begin{aligned} E\big(t^2(\hat{z}_i)\big) &= Var\big(t(z_i)\big) - \tau \,\mathrm{tr}\Big\{(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{d}_i\mathbf{d}_i^T\Big\} + \frac{c_g}{n} \\ &= Var\big(t(z_i)\big) - \Big(\tau\, h_{ii} - \frac{c_g}{n}\Big), \end{aligned} \tag{3.21}$$

com $\tau = (4d_g)^{-1}E\Big(2t(z_i)t'(z_i)v(z_i)z_i - t'^2(z_i) - t''(z_i)t(z_i)\Big)$. Então, reescrevendo a equação (3.21) temos

$$E\big(t^2(\hat{z}_i)\big) \approx Var\big(t(z_i)\big)\Big\{1 - \tau^*\, h_{ii}\Big\}, \tag{3.22}$$

em que $\tau^* = \dfrac{\tau}{Var(t(z_i))}$ e $h_{ii} = \mathbf{d}_i^T(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{d}_i$.

Na equação (3.22) é possível observar que o efeito da estrutura do modelo de regressão no segundo momento da distribuição dos resíduos depende do valor de h_{ii}, o qual é usado freqüentemente para identificar observações atípicas no espaço gerado pelas colunas da matriz $\mathbf{D}_\beta$. É fácil mostrar que a matriz $\mathbf{H}$ é simétrica e idempotente, o que implica que $0 \leq h_{ii} \leq 1$, sendo os valores "altos" de h_{ii} os que indicam que a observação está localizada numa região remota no espaço gerado pelas colunas da matriz $\mathbf{D}_\beta$. Além disso, segundo a equação (3.22) valores "altos" de h_{ii} indicam que a variância de $t(\hat{z}_i)$ é significativamente menor que a variância de $t(z_i)$. Entretanto, é possível observar também que o efeito de h_{ii} na variância dos resíduos depende da distribuição da variável resposta através do valor de τ^*, que é igual a 1 quando y_i tem distribuição normal. Do Lema 6 do Apêndice A podemos concluir que o valor do fator de correção τ para o resíduo $t_D(\hat{z}_i)$ é igual a 1, enquanto que, para o resíduo $\hat{z}_i$ temos que $\tau = 1/4d_g$, o que implica que $E(\hat{z}_i^2) = \xi(1 - (4d_g\xi)^{-1}h_{ii})$ (veja Galea, Paula e Cysneiros, 2005). Além disso, é possível observar que o valor de $Var\big(t(z_i)\big)$ é igual a 1 quando $t(\cdot) = t_Q(\cdot)$, pois a distribuição assintótica do resíduo $t_Q(\hat{z}_i)$ é normal padrão. No Quadro 3.4 apresentamos os resíduos $t_D(\hat{z}_i)$, $t_Q(\hat{z}_i)$ e $\hat{z}_i$ bem como suas versões padronizadas, denotadas por $t_D^*(\hat{z}_i)$, $t_Q^*(\hat{z}_i)$ e t_{r_i}, respectivamente.

Quadro 3.4 *Expressões gerais para as versões padronizadas dos resíduos* $t_D(\hat{z}_i)$, $t_Q(\hat{z}_i)$ *e* $\hat{z}_i$.

Resíduo	$t(z)$	Padronização
Componente de desvio	$\text{sinal}(z)\Big\{2\log\big[g(0)/g(z^2)\big]\Big\}^{\frac{1}{2}}$	$t_D^*(\hat{z}_i) = \dfrac{t_D(\hat{z}_i)}{(\sigma_g^2 - \hat{h}_{ii})^{\frac{1}{2}}}$
Quantal	$\Phi^{-1}[F(z)]$	$t_Q^*(\hat{z}_i) = \dfrac{t_Q(\hat{z}_i)}{(1 - \tau\hat{h}_{ii})^{\frac{1}{2}}}$
Cox e Snell	Identidade	$t_{r_i} = \dfrac{\hat{z}_i}{\xi^{\frac{1}{2}}(1 - (4d_g\xi)^{-1}\hat{h}_{ii})^{\frac{1}{2}}}$

Como ilustração, nos Quadros 3.5 e 3.6 apresentamos os valores da variância e do fator de correção τ^* para os resíduos $t_D(\hat{z}_i)$ e $t_Q(\hat{z}_i)$, respectivamente. Nestes quadros consideramos modelos em que o erro segue distribuição t-Student. Estes valores foram obtidos via integração numérica. Notamos que, à medida que aumenta o número de graus de liberdade da distribuição t-Student, o valor de τ^* está mais próximo do fator de correção para o resíduo quando o erro do modelo segue distribuição normal. Além disso, é possível observar que à medida que a distribuição da variável resposta apresenta caudas mais pesadas do que a normal, o valor do fator de correção τ^* decresce, indicando que em modelos baseados nessas distribuições a variância do resíduo $t_D(\hat{z}_i)$ tende a ser menos afetada pelo valor de h_{ii}. Além disso, podemos concluir que o resíduo proposto por Galea, Paula e Cysneiros (2005) é equivalente a uma padronização do resíduo $\hat{z}_i$ usando a expressão descrita em (3.22).

Quadro 3.5 *Valores da variância e do fator de correção para o resíduo* $t_D(\hat{z}_i)$ *quando o erro do modelo segue distribuição t-Student com* ν *graus de liberdade.*

ν	$Var\big(t(u_i)\big)$	τ^*	ν	$Var\big(t(u_i)\big)$	τ^*
4	1,401862	0,7133	18	1,084832	0,9218
5	1,317766	0,7588	19	1,080294	0,9256
6	1,262606	0,7920	20	1,076217	0,9291
7	1,223688	0,8172	21	1,072534	0,9323
8	1,194779	0,8369	22	1,069190	0,9352
9	1,172467	0,8529	23	1,066141	0,9379
10	1,154730	0,8660	24	1,063349	0,9404
11	1,140294	0,8769	25	1,060783	0,9426
12	1,128318	0,8862	26	1,058417	0,9448
13	1,118222	0,8942	27	1,056228	0,9467
14	1,109597	0,9012	28	1,054197	0,9485
15	1,102144	0,9073	29	1,052308	0,9502
16	1,095639	0,9127	30	1,050546	0,9518
17	1,089912	0,9175	31	1,049486	0,9528

Quadro 3.6 *Valores do fator de correção para o resíduo* $t_Q(\hat{z}_i)$ *quando o erro do modelo segue distribuição t-Student com* ν *graus de liberdade.*

ν	τ^*	ν	τ^*	ν	τ^*
4	0,7113	13	0,8978	22	0,9377
5	0,7611	14	0,9011	23	0,9371
6	0,8003	15	0,9022	24	0,9284
7	0,8065	16	0,9195	25	0,9402
8	0,8444	17	0,9261	26	0,9448
9	0,8492	18	0,9128	27	0,9442
10	0,8535	19	0,9253	28	0,9401
11	0,8784	20	0,9377	29	0,9402
12	0,8892	21	0,9279	30	0,9447

.

Por outro lado, temos que a medida da qualidade do ajuste global $MQ(\hat{\boldsymbol{\beta}})$ pode ser padronizada usando o resultado na equação (3.22). Isto é feito definindo a medida $\overline{MQ}(\hat{\boldsymbol{\beta}}) = MQ(\hat{\boldsymbol{\beta}})/(n\sigma_g^2 - p)$, a qual pode ser reescrita na forma abaixo

$$\overline{MQ}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} l_i \, t_D^{*2}(\hat{z}_i),$$

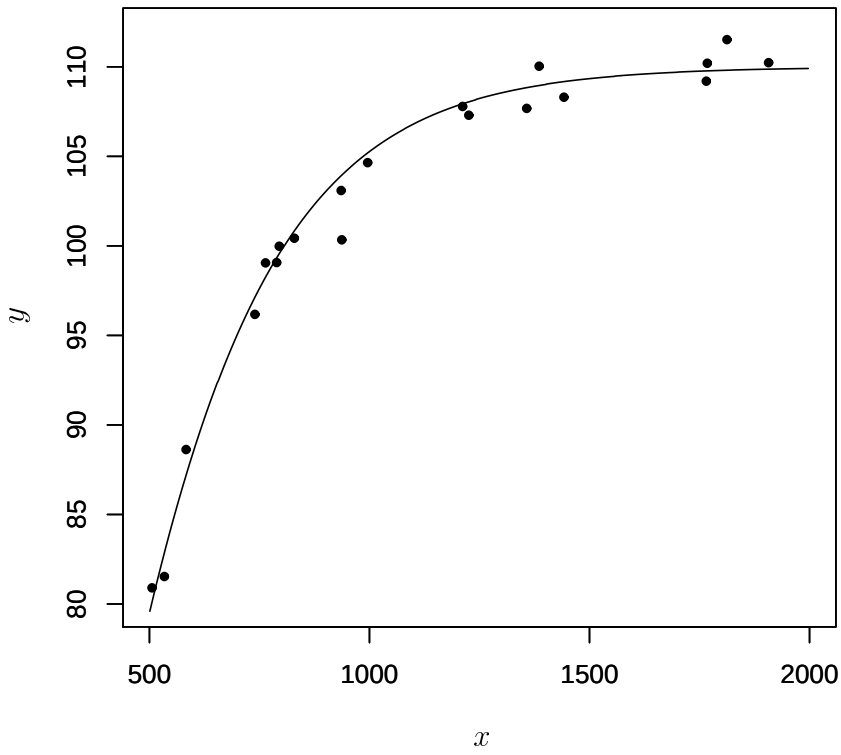
em que $l_i = (\sigma_g^2 - \hat{h}_{ii})/(n\sigma_g^2 - p)$ é o peso da i-ésima observação na medida da qualidade do ajuste global. Podemos notar que $0 < l_i < 1$, $i = 1, \ldots, n$, e $\sum l_i = 1$. Além disso, temos que l_i é "pequeno" se $\hat{h}_{ii}$ é "grande", indicando que resíduos de observações localizadas em regiões remotas do subespaço gerado pelas colunas da matriz $\mathbf{D}_{\hat{\beta}}$ terão peso "pequeno" na medida da qualidade do ajuste global $\overline{MQ}(\hat{\boldsymbol{\beta}})$.

3.3 Avaliação numérica dos resíduos

Com o objetivo de avaliar e comparar o comportamento dos resíduos propostos neste capítulo com o resíduo t_{r_i} proposto por Galea, Paula e Cysneiros (2005), desenvolvemos um estudo de simulação no qual consideramos modelos em que a componente sistemática é dada pela curva de Gompertz (veja expressão 2.4), e a componente aleatória está baseada nas distribuições t-Student com $\nu = 5$ graus de liberdade e logística tipo II. Os valores dos parâmetros do modelo são $\beta_1 = 110$; $\beta_2 = 2,4$; $\beta_3 = 0,004$ e $\phi = 0,1; 0,5$. Os valores da variável explicativa x são mantidos fixos ao longo do processo de simulação e são gerados seguindo distribuição uniforme no intervalo $(500, 2000)$. Em cada uma das 10000 réplicas de Monte Carlo, os parâmetros do modelo são estimados usando o algoritmo descrito nas expressões (2.14) e (2.15). Com base nas estimativas dos parâmetros são calculadas as versões padronizadas dos resíduos, denotadas por $t_D^*(\hat{z}_i)$, $t_Q^*(\hat{z}_i)$ e t_{r_i} (veja Quadro 3.4). O tamanho da amostra considerado é $n = 20$. Como ilustração, apresentamos na Figura 3.4 um gráfico que descreve o comportamento da resposta y em relação à variável explicativa para uma das amostras geradas. Neste caso o erro do modelo foi gerado seguindo distribuição logística tipo II com parâmetro de escala $\phi = 0,5$. Como uma segunda parte do estudo de simulação, consideramos modelos em que a componente sistemática é dada pela curva de Michaelis-Menten (veja expressão 2.5), e a componente aleatória está baseada

nas distribuições t-Student com $\nu = 5$ graus de liberdade e logística tipo II. Os valores dos parâmetros do modelo são $\beta_1 = 212$; $\beta_2 = 0,0641$; e $\phi = 1;2$. Os valores da variável explicativa x são mantidos fixos ao longo do processo de simulação e são gerados seguindo distribuição uniforme no intervalo $(0, 1)$.

Figura 3.4 *Relação entre a variável resposta y e a variável explicativa x para uma amostra gerada do modelo de Gompertz com erros logística tipo II e parâmetro de escala ϕ =0,5.*



Nos Quadros 3.7 a 3.18 apresentamos os valores da média, da variância e dos coeficientes de assimetria e curtose da distribuição empírica dos resíduos. Nestes quadros pode-se observar que em todos os casos os valores da média e da variância estão próximos de zero e um, respectivamente. Além disso, os valores dos coeficientes de assimetria estão muito próximos de zero. Em relação ao coeficiente de curtose pode-se observar que para os resíduos $t_D^*(\hat{z}_i)$ e $t_Q^*(\hat{z}_i)$, este valor está próximo de três, indicando grande proximidade com o coeficiente de curtose da distribuição normal. O resíduo t_{r_i} apresenta uma curtose que acompa-

nha a curtose da distribuição da variável resposta. O comportamento observado nas distribuições empíricas dos resíduos não varia significativamente para diferentes valores do parâmetro de dispersão, nem para diferentes distribuições da resposta. Os resultados do estudo de simulação indicam que mesmo para pequenas amostras, a distribuições dos resíduos $t_D^*(\hat{z}_i)$ e $t_Q^*(\hat{z}_i)$ estão muito próximas da distribuição normal padrão. Isto indica que a distribuição normal padrão pode ser usada como distribuição de referência para estes resíduos. Entretanto, o resíduo $t_D^*(\hat{z}_i)$ pode ser mais adequado na prática, pois o cálculo deste resíduo não requer a função de distribuição acumulada de z_i. Por outro lado, sugerimos que a análise destes resíduos seja complementada pela construção de gráficos de envelope, como foi descrito por Atkinson (1981).

Quadro 3.7 *Medidas descritivas para a distribuição empírica do resíduo* $t_Q^*(\hat{z}_i)$ *no modelo de Gompertz com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 0,1$				$\phi = 0,5$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0022	1,1554	0,0016	2,9176	-0,0212	1,1186	-0,0426	2,8918
2	-0,0112	1,1457	-0,0314	2,8409	0,0041	1,1414	-0,0226	2,7543
3	-0,0026	1,1404	-0,0261	2,8274	-0,0081	1,1060	0,0488	2,8818
4	0,0029	1,0590	-0,0163	2,8767	0,0073	1,1541	0,0134	3,0264
5	-0,0012	1,1527	0,0356	2,8727	0,0377	1,1076	-0,0148	3,0023
6	0,0310	1,1026	-0,0107	2,9131	-0,0003	1,1390	0,0257	3,0899
7	0,0229	1,1380	0,0081	2,8760	0,0099	1,1227	0,1035	2,9527
8	-0,0057	1,1243	0,1319	3,1857	0,0038	1,1622	0,0445	2,8913
9	0,0099	1,1530	0,0273	2,8819	0,0181	1,1150	0,0011	2,9598
10	0,0075	1,1354	0,0582	2,8645	-0,0286	1,1518	0,0345	2,9393
11	-0,0134	1,1567	0,0104	2,8234	0,0219	1,1335	0,0038	2,8952
12	0,0158	1,1538	0,0170	2,8134	0,0056	1,1435	0,0259	2,9798
13	0,0057	1,1249	0,0359	2,9100	0,0047	1,1991	0,0235	3,3228
14	-0,0035	1,3331	0,0082	3,2923	0,0452	1,1599	-0,0404	2,8220
15	0,0148	1,1192	-0,0707	3,2898	-0,0318	1,1199	0,0336	2,9189
16	-0,0278	1,1346	0,0694	2,8672	-0,0042	1,1820	-0,0423	3,0550
17	0,0100	1,1702	-0,0479	2,9179	-0,0424	1,1563	-0,0006	3,0167
18	-0,0257	1,1671	-0,0342	2,8963	0,0119	1,0368	-0,0302	3,0715
19	0,0038	1,1330	0,0340	3,1398	-0,0141	1,1087	-0,0237	2,9171
20	-0,0030	1,1194	-0,0207	2,9009	0,0078	1,1622	0,0222	2,9999

Quadro 3.8 *Medidas descritivas para a distribuição empírica do resíduo* $t^*_D(\hat{z}_i)$ *no modelo de Gompertz com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 0,1$				$\phi = 0,5$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0022	1,1554	0,0016	2,8968	-0,0212	1,1189	-0,0422	2,8732
2	-0,0112	1,1460	-0,0311	2,8230	0,0041	1,1418	-0,0225	2,7357
3	-0,0026	1,1407	-0,0261	2,8076	-0,0081	1,1062	0,0480	2,8618
4	0,0029	1,0507	-0,0163	2,8760	0,0073	1,1540	0,0136	3,0043
5	-0,0013	1,1527	0,0353	2,8525	0,0377	1,1077	-0,0145	2,9764
6	0,0310	1,1028	-0,0106	2,8887	-0,0003	1,1389	0,0256	3,0675
7	0,0229	1,1381	0,0079	2,8571	0,0098	1,1229	0,1020	2,9314
8	-0,0058	1,1243	0,1297	3,1598	0,0037	1,1622	0,0436	2,8690
9	0,0099	1,1531	0,0267	2,8605	0,0181	1,1152	0,0008	2,9368
10	0,0075	1,1356	0,0574	2,8430	-0,0287	1,1518	0,0345	2,9184
11	-0,0134	1,1568	0,0101	2,8035	0,0219	1,1338	0,0033	2,8753
12	0,0158	1,1540	0,0162	2,7951	0,0056	1,1435	0,0251	2,9590
13	0,0057	1,1251	0,0352	2,8894	0,0046	1,1986	0,0235	3,2930
14	-0,0035	1,1320	0,0086	2,9220	0,0452	1,1601	-0,0406	2,8031
15	0,0149	1,1191	-0,0693	3,2626	-0,0318	1,1201	0,0337	2,8997
16	-0,0278	1,1348	0,0691	2,8481	-0,0041	1,1817	-0,0415	3,0310
17	0,0101	1,1701	-0,0474	2,8961	-0,0424	1,1562	0,0002	2,9963
18	-0,0256	1,1671	-0,0333	2,8761	0,0119	1,0837	-0,0300	2,9215
19	0,0038	1,1330	0,0335	3,1153	-0,0141	1,1089	-0,0227	2,8966
20	-0,0030	1,1196	-0,0198	2,8802	0,0077	1,1621	0,0221	2,9791

Quadro 3.9 *Medidas descritivas para a distribuição empírica do resíduo* t_{r_i} *no modelo de Gompertz com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 0,1$				$\phi = 0,5$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0018	1,1779	0,0939	6,4374	-0,0232	1,1006	0,0007	6,2107
2	-0,0131	1,1232	-0,0741	5,3950	0,0030	1,1422	0,2350	8,8910
3	-0,0038	1,1431	0,2421	9,1463	-0,0038	1,1041	0,2223	6,0230
4	0,0024	1,3743	-0,0166	5,9187	0,0074	1,1741	-0,1275	6,5980
5	0,0011	1,1745	-0,0400	6,1165	0,0352	1,1565	-0,1915	6,4737
6	0,0291	1,1463	-0,1967	9,4615	0,0012	1,1531	0,0054	6,2304
7	0,0231	1,1382	-0,0060	5,3861	0,0174	1,1294	0,3471	7,0421
8	0,0042	1,1521	0,6036	9,4369	0,0075	1,1994	0,1091	7,5952
9	0,0123	1,1801	0,1022	6,7957	0,0177	1,1394	-0,2362	9,8034
10	0,0115	1,1581	-0,0775	9,3483	-0,0258	1,1629	0,0519	6,0640
11	-0,0122	1,1694	0,0438	5,9028	0,0221	1,1183	0,0791	5,8022
12	0,0175	1,1418	0,1317	5,2889	0,0079	1,1529	0,1390	5,8382
13	0,0087	1,1283	0,1751	6,0118	0,0055	1,2528	-0,1534	6,7163
14	-0,0038	1,4488	-0,2503	9,0923	0,0421	1,1607	-0,0248	5,6673
15	0,0083	1,1704	-0,1539	9,3614	-0,0291	1,1174	0,0076	5,8188
16	-0,0228	1,1335	0,0781	5,6036	-0,0078	1,2242	-0,1777	7,1280
17	0,0062	1,2054	-0,0901	6,7322	-0,0423	1,1648	-0,1699	6,2322
18	-0,0285	1,1839	-0,1835	6,0337	0,0098	1,0907	-0,1270	5,5865
19	0,0062	1,1515	0,0934	7,5473	-0,0166	1,1107	-0,3146	6,9533
20	-0,0055	1,1331	-0,2307	6,2802	0,0089	1,1808	0,0127	5,8141

Quadro 3.10 *Medidas descritivas para a distribuição empírica do resíduo* $t_Q^*(\hat{z}_i)$ *no modelo de Gompertz com erros logística tipo II.*

	$\phi = 0,1$				$\phi = 0,5$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0029	1,1516	0,0048	2,7892	0,0060	1,1099	-0,0132	3,1014
2	0,0177	1,1469	-0,0417	2,8012	-0,0025	1,1332	-0,0127	2,9799
3	-0,0032	1,1308	-0,0007	2,7606	0,0061	1,1273	-0,0298	2,7927
4	-0,0041	1,1106	-0,0162	2,9308	0,0317	1,1792	-0,0217	2,9250
5	-0,0056	1,1490	0,0091	2,8934	-0,0256	1,1213	-0,0104	2,8024
6	0,0105	1,1466	0,0051	2,7617	-0,0201	1,1264	0,0021	2,7282
7	0,0198	1,0939	-0,0521	2,7196	-0,0056	1,1426	0,0331	2,8619
8	0,0034	1,1286	0,0212	2,8430	-0,0190	1,1676	-0,0066	2,8681
9	0,0313	1,1656	-0,0317	2,8900	-0,0025	1,1802	-0,0528	2,8552
10	-0,0005	1,1608	-0,0580	2,9884	0,0109	1,0699	-0,0100	3,0299
11	0,0035	1,1630	0,0143	2,8176	0,0292	1,1588	-0,0669	2,8419
12	-0,0167	1,1278	0,0130	2,9305	0,0080	1,1177	-0,0147	2,8311
13	-0,0446	1,1352	0,0142	2,9481	0,0148	1,1554	0,0110	2,8122
14	-0,0028	1,1341	-0,0031	2,8734	-0,0165	1,1495	0,0414	2,8405
15	0,0197	1,1348	-0,0058	2,7875	-0,0001	1,1449	-0,0341	2,8627
16	0,0256	1,1852	-0,0281	2,8593	-0,0176	1,1304	0,0633	2,9978
17	0,0073	1,1778	-0,0215	2,8698	-0,0165	1,1847	0,0338	2,9393
18	-0,0174	1,1128	-0,0161	2,9560	-0,0126	1,1688	0,0161	2,8453
19	-0,0192	1,1403	0,0551	2,7514	0,0259	1,0198	0,0162	2,7137
20	-0,0217	1,1418	0,0519	2,8792	0,0232	1,1173	-0,0175	2,8723

Quadro 3.11 *Medidas descritivas para a distribuição empírica do resíduo* $t_D^*(\hat{z}_i)$ *no modelo de Gompertz com erros logística tipo II.*

	$\phi = 0,1$				$\phi = 0,5$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0030	1,1505	0,0046	2,7012	0,0061	1,1080	-0,0122	2,9793
2	0,0179	1,1461	-0,0399	2,7069	-0,0252	1,1337	-0,0111	2,8790
3	-0,0032	1,1309	-0,0005	2,6672	0,0062	1,1271	-0,0287	2,7034
4	-0,0040	1,1099	-0,0143	2,8251	0,0317	1,1763	-0,0201	2,8104
5	-0,0056	1,1475	0,0091	2,7878	-0,0255	1,1208	-0,0071	2,7117
6	0,0104	1,1460	0,0041	2,6696	-0,0201	1,1265	0,0012	2,6415
7	0,0200	1,0389	-0,0508	2,9801	-0,0058	1,1405	0,0307	2,7647
8	0,0033	1,1278	0,0184	2,7443	-0,0190	1,1655	-0,0038	2,7703
9	0,0315	1,1634	-0,0320	2,7940	-0,0022	1,1774	-0,0489	2,7527
10	-0,0002	1,1586	-0,0549	2,8787	0,0109	1,0696	-0,0086	2,9217
11	0,0034	1,1618	0,0146	2,7235	0,0295	1,1564	-0,0630	2,7433
12	-0,0168	1,1264	0,0159	2,8250	0,0081	1,1175	-0,0162	2,7356
13	-0,0446	1,1341	0,0176	2,8429	0,0147	1,1538	0,0090	2,7187
14	-0,0028	1,1327	-0,0029	2,7691	-0,0168	1,1484	0,0418	2,7376
15	0,0197	1,1344	-0,0072	2,6969	0,0001	1,1430	-0,0310	2,7526
16	0,0257	1,1826	-0,0280	2,7554	-0,0180	1,1285	0,0602	2,8832
17	0,0074	1,1759	-0,0207	2,7731	-0,0166	1,1818	0,0313	2,8322
18	-0,0174	1,1119	-0,0126	2,8487	-0,0127	1,1668	0,0160	2,7416
19	-0,0195	1,1398	0,0523	2,6605	0,0259	1,0239	0,0133	2,6466
20	-0,0219	1,1407	0,0506	2,7792	0,0233	1,1164	-0,0179	2,7687

Quadro 3.12 *Medidas descritivas para a distribuição empírica do resíduo* t_{r_i} *no modelo de Gompertz com erros logística tipo II.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0027	1,1549	-0,0003	3,5981	0,0055	1,1250	-0,0237	4,3056
2	0,0161	1,1506	-0,0599	3,7099	-0,0255	1,1238	-0,0232	3,9953
3	-0,0033	1,1272	0,0064	3,7192	0,0049	1,1254	-0,0415	3,6248
4	-0,0047	1,1145	-0,0321	3,9908	0,0312	1,2022	-0,0606	4,1550
5	-0,0052	1,1604	0,0148	3,9725	-0,0261	1,1209	-0,0392	3,6256
6	0,0108	1,1490	0,0136	3,6374	-0,0202	1,1216	0,0164	3,5270
7	0,0183	1,1382	-0,0645	3,7123	-0,0042	1,1562	0,0573	3,7857
8	0,0042	1,1333	0,0492	3,8172	-0,0194	1,1810	-0,0368	3,7927
9	0,0302	1,1776	-0,0327	3,7998	-0,0049	1,2008	-0,1025	3,8885
10	-0,0028	1,1755	-0,0946	4,0927	0,0105	1,0701	-0,0283	4,0973
11	0,0041	1,1684	0,0119	3,7145	0,0266	1,1755	-0,1056	3,7764
12	-0,0162	1,1378	-0,0110	3,9761	0,0073	1,1173	0,0132	3,7850
13	-0,0440	1,1419	-0,0189	3,9951	0,0154	1,1641	0,0316	3,6949
14	-0,0030	1,1448	-0,0052	3,9146	-0,0148	1,1569	0,0366	3,8602
15	0,0196	1,1342	0,0062	3,6451	-0,0015	1,1608	-0,0697	4,0041
16	0,0246	1,2049	-0,0215	3,9114	-0,0152	1,1460	0,0926	4,1970
17	0,0064	1,1883	-0,0395	3,7882	-0,0152	1,2053	0,0584	3,9909
18	-0,0180	1,1187	-0,0537	4,0260	-0,0120	1,1832	0,0211	3,8844
19	-0,0170	1,1407	0,0873	3,6263	0,0260	0,9826	0,0492	3,3723
20	-0,0196	1,1480	0,0644	3,8298	0,0226	1,1234	-0,0237	3,9216

Quadro 3.13 *Medidas descritivas para a distribuição empírica do resíduo* $t^*_Q(\hat{z}_i)$ *no modelo de Michaelis-Menten com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0003	1,0997	-0,0017	2,8539	-0,0111	1,0767	-0,0044	2,9478
2	0,0050	1,0786	0,0149	2,8378	-0,0132	1,0966	0,0040	2,8486
3	-0,0089	1,0701	-0,0072	2,9346	0,0062	1,0848	-0,0131	2,9313
4	-0,0072	1,0746	0,0097	2,9249	-0,0048	1,0862	0,0321	2,8112
5	-0,0043	1,0885	0,0243	3,0008	-0,0129	1,0689	-0,0185	2,8721
6	0,0054	1,0951	0,0331	2,9178	-0,0031	1,0765	-0,0088	2,9357
7	-0,0099	1,0150	-0,0334	3,4862	-0,0131	1,0867	0,0704	2,8994
8	0,0076	1,0842	-0,0397	2,8661	0,0217	1,0819	-0,0227	2,9619
9	0,0085	1,0751	-0,0042	2,8892	-0,0017	1,0917	-0,0075	2,9994
10	0,0118	1,0576	0,0103	2,9227	0,0121	1,0650	-0,0149	2,8541
11	-0,0171	1,0786	0,0319	2,8395	-0,0088	1,0872	-0,0072	2,9952
12	0,0234	1,0672	-0,0002	2,8346	0,0047	1,0904	-0,0292	2,8740
13	-0,0195	1,0943	0,0028	2,9324	-0,0065	1,0466	-0,0309	2,9290
14	-0,0033	1,0784	-0,0081	2,9004	-0,0135	1,0946	-0,0215	3,0363
15	-0,0158	1,0086	-0,0505	2,9740	-0,0012	1,1072	-0,0402	2,9115
16	-0,0029	1,0706	0,0107	2,9353	0,0017	1,0584	0,0141	2,9225
17	-0,0065	1,0887	-0,0064	2,9884	0,0090	1,0805	-0,0207	2,8538
18	0,0036	1,0951	-0,0081	2,9141	0,0025	1,1090	0,0039	2,8307
19	0,0027	1,1052	-0,0215	2,7737	0,0189	1,0867	0,0024	2,8730
20	0,0198	1,0823	-0,0534	2,8420	-0,0017	1,0829	0,0396	2,9368

Quadro 3.14 *Medidas descritivas para a distribuição empírica do resíduo* $t^*_D(\hat{z}_i)$ *no modelo de Michaelis-Menten com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0003	1,1001	-0,0015	2,8344	-0,0111	1,0770	-0,0046	2,9286
2	0,0050	1,0791	0,0146	2,8196	-0,0133	1,0664	0,0038	2,8152
3	-0,0089	1,0704	-0,0067	2,9134	0,0062	1,0851	-0,0129	2,9115
4	-0,0073	1,0750	0,0099	2,9028	-0,0049	1,0867	0,0316	2,7941
5	-0,0044	1,0887	0,0243	2,9778	-0,0129	1,0695	-0,0177	2,8522
6	0,0054	1,0954	0,0326	2,8979	-0,0031	1,0768	-0,0089	2,9148
7	-0,0099	1,0153	-0,0325	3,4601	-0,0132	1,0871	0,0697	2,8786
8	0,0077	1,0846	-0,0392	2,8456	0,0217	1,0823	-0,0228	2,9429
9	0,0085	1,0755	-0,0038	2,8680	-0,0016	1,0919	-0,0076	2,9761
10	0,0118	1,0581	0,0097	2,9030	0,0121	1,0655	-0,0149	2,8350
11	-0,0172	1,0790	0,0318	2,8202	-0,0088	1,0873	-0,0069	2,9713
12	0,0234	1,0678	-0,0005	2,8153	0,0048	1,0907	-0,0289	2,8542
13	-0,0195	1,0945	0,0033	2,9121	-0,0065	1,0472	-0,0302	2,9095
14	-0,0033	1,0787	-0,0080	2,8800	-0,0135	1,0948	-0,0208	3,0153
15	-0,0157	1,0090	-0,0497	2,9468	-0,0012	1,1075	-0,0400	2,8904
16	-0,0029	1,0710	0,0103	2,9152	0,0017	1,0588	0,0141	2,9018
17	-0,0065	1,0889	-0,0060	2,9648	0,0091	1,0809	-0,0203	2,8335
18	0,0036	1,0954	-0,0086	2,8940	0,0025	1,1094	0,0037	2,8125
19	0,0027	1,1057	-0,0213	2,7559	0,0189	1,0871	0,0019	2,8538
20	0,0199	1,0828	-0,0530	2,8232	-0,0018	1,0832	0,0393	2,9169

Quadro 3.15 *Medidas descritivas para a distribuição empírica do resíduo* t_{r_i} *no modelo de Michaelis-Menten com erros t-Student de* $\nu = 5$ *graus de liberdade.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0006	1,0962	-0,0210	6,2451	-0,0107	1,0531	0,0765	5,6657
2	0,0059	1,0529	0,0214	5,5257	-0,0120	0,9448	-0,0690	9,2814
3	-0,0098	1,0767	-0,0348	7,8141	0,0048	1,0747	-0,1222	6,2115
4	-0,0069	1,0858	-0,0668	7,6214	-0,0022	1,0538	0,1285	5,1064
5	-0,0031	1,1085	-0,0544	7,7924	-0,0146	1,0472	-0,2382	6,6685
6	0,0076	1,0855	0,0517	5,6856	-0,0032	1,0784	0,1384	7,5152
7	-0,0113	0,9703	-0,1722	7,4580	-0,0079	1,0778	0,1824	7,2236
8	0,0046	1,0818	-0,0609	7,2844	0,0198	1,0547	-0,0032	5,5338
9	0,0076	1,0769	-0,0411	7,8204	-0,0018	1,1127	0,0692	7,5454
10	0,0126	1,0350	0,1426	6,0945	0,0110	1,0465	0,0098	6,2247
11	-0,0146	1,0650	0,1208	6,4799	-0,0093	1,1203	0,0146	9,3896
12	0,0229	1,0494	-0,0246	6,3911	0,0026	1,0854	-0,0568	5,9541
13	-0,0195	1,0887	-0,0771	6,0823	-0,0087	1,0244	-0,2327	8,5232
14	-0,0038	1,0745	-0,0046	6,3413	-0,0150	1,0820	-0,1449	6,1917
15	-0,0170	0,9668	-0,0330	8,5146	-0,0034	1,1116	0,1074	7,5398
16	-0,0017	1,0486	0,1299	6,4273	0,0023	1,0452	0,0366	7,2277
17	-0,0072	1,1154	-0,1294	8,5866	0,0070	1,0777	-0,1087	7,9623
18	0,0038	1,0927	0,1295	5,9718	0,0029	1,0901	0,0091	5,5450
19	0,0011	1,0873	-0,0749	6,4353	0,0190	1,0747	0,0896	5,8242
20	0,0156	1,0560	-0,1934	6,2266	0,0007	1,0695	0,0446	5,7675

Quadro 3.16 *Medidas descritivas para a distribuição empírica do resíduo* $t_Q^*(\hat{z}_i)$ *no modelo de Michaelis-Menten com erros logística tipo II.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0041	1,0675	-0,0105	2,7824	-0,0104	1,0709	0,0084	2,8126
2	0,0002	1,0707	-0,0145	2,8437	0,0197	1,0955	0,0175	2,8257
3	-0,0161	1,0303	-0,0115	2,9033	0,0017	1,0781	0,0015	2,8832
4	-0,0225	1,0639	0,0336	2,8440	0,0070	1,0797	-0,0112	2,7692
5	-0,0141	1,0821	0,0180	2,8451	0,0254	1,1190	0,0110	2,8087
6	0,0070	1,0673	0,0191	2,8462	-0,0085	1,1035	-0,0074	2,9116
7	0,0123	1,0472	-0,0055	3,2031	-0,0296	1,0869	-0,0064	3,0300
8	-0,0044	1,1003	-0,0313	3,0216	-0,0209	1,0779	0,0285	2,7630
9	-0,0090	1,1049	0,0259	2,8813	-0,0018	1,0535	0,0065	2,8444
10	0,0180	1,0910	0,0093	2,8097	-0,0067	1,0595	-0,0333	3,1687
11	-0,0010	1,1023	0,0037	2,7405	0,0063	1,0577	0,0312	3,1622
12	0,0124	1,0935	0,0202	2,8356	-0,0313	1,0624	-0,0255	2,8972
13	0,0150	1,0918	0,0333	2,7795	0,0171	1,0898	-0,0006	2,8497
14	0,0064	1,0896	-0,0197	2,8665	0,0166	1,1228	-0,0033	2,7900
15	-0,0149	1,0807	0,0412	2,8022	-0,0083	1,0923	0,0272	2,8344
16	-0,0064	1,1009	-0,0268	2,8090	-0,0006	1,0844	-0,0147	2,7652
17	-0,0027	1,0834	-0,0117	2,8124	0,0025	1,0656	0,0251	2,7896
18	0,0249	1,0902	0,0097	2,7992	0,0098	1,1051	-0,0012	2,7858
19	-0,0067	1,0846	0,0302	2,8305	0,0046	1,1068	0,0460	2,9428
20	0,0133	1,1042	-0,0062	2,8756	0,0108	1,0608	0,0124	2,7876

Quadro 3.17 *Medidas descritivas para a distribuição empírica do resíduo* $t_D^*(\hat{z}_i)$ *no modelo de Michaelis-Menten com erros logística tipo II.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0040	1,0691	-0,0101	2,6903	-0,0105	1,0719	0,0096	2,7164
2	0,0003	1,0718	-0,0129	2,7437	0,0196	1,0953	0,0139	2,7293
3	-0,0162	1,0375	-0,0100	3,1890	0,0017	1,0781	0,0005	2,7835
4	-0,0226	1,0655	0,0323	2,7452	0,0070	1,0804	-0,0114	2,6787
5	-0,0142	1,0831	0,0177	2,7488	0,0253	1,1182	0,0077	2,7138
6	0,0069	1,0687	0,0176	2,7510	-0,0085	1,1027	-0,0056	2,8051
7	0,0123	1,0525	-0,0050	3,0914	-0,0295	1,0862	-0,0044	2,9146
8	-0,0042	1,1036	-0,0281	2,9200	-0,0211	1,0787	0,0279	2,6744
9	-0,0092	1,1054	0,0257	2,7794	-0,0019	1,0543	0,0065	2,7486
10	0,0179	1,0921	0,0070	2,7119	-0,0066	1,0595	-0,0292	3,0481
11	-0,0010	1,1034	0,0027	2,6522	0,0062	1,0579	0,0264	3,0425
12	0,0123	1,0941	0,0174	2,7337	-0,0311	1,0627	-0,0215	2,7941
13	0,0148	1,0927	0,0309	2,6870	0,0171	1,0899	-0,0026	2,7519
14	0,0065	1,0905	-0,0195	2,7659	0,0166	1,1221	-0,0053	2,6969
15	-0,0151	1,0821	0,0396	2,7091	-0,0084	1,0922	0,0251	2,7363
16	-0,0063	1,1021	-0,0253	2,7148	-0,0005	1,0853	-0,0136	2,6735
17	-0,0026	1,0846	-0,0115	2,7197	0,0023	1,0665	0,0234	2,6977
18	0,0248	1,0911	0,0057	2,7042	0,0098	1,1049	-0,0020	2,6958
19	-0,0069	1,0859	0,0279	2,7361	0,0044	1,1063	0,0415	2,8398
20	0,0133	1,1043	-0,0064	2,7708	0,0108	1,0619	0,0107	2,6941

Quadro 3.18 *Medidas descritivas para a distribuição empírica do resíduo* t_{r_i} *no modelo de Michaelis-Menten com erros logística tipo II.*

	$\phi = 1$				$\phi = 2$			
i	Média	Variância	Assimetria	Curtose	Média	Variância	Assimetria	Curtose
1	-0,0045	1,0600	-0,0081	3,6471	-0,0101	1,0618	-0,0093	3,7811
2	-0,0002	1,0689	-0,0368	3,8233	0,0205	1,0960	0,0518	3,7555
3	-0,0163	1,0337	-0,0312	4,3996	0,0017	1,0765	0,0125	3,8494
4	-0,0212	1,0593	0,0500	3,8157	0,0065	1,0724	-0,0118	3,6189
5	-0,0134	1,0819	0,0239	3,7506	0,0260	1,1231	0,0392	3,7260
6	0,0077	1,0636	0,0350	3,7304	-0,0088	1,1098	-0,0257	3,9722
7	0,0121	1,0497	-0,0170	4,2765	-0,0298	1,0921	-0,0285	4,2264
8	-0,0055	1,1052	-0,0602	3,9613	-0,0199	1,0689	0,0350	3,5783
9	-0,0080	1,1116	0,0284	3,8485	-0,0016	1,0465	0,0048	3,7557
10	0,0184	1,0901	0,0385	3,7776	-0,0077	1,0584	-0,0820	4,4111
11	-0,0009	1,0984	0,0186	3,5753	0,0071	1,0548	0,0867	4,4193
12	0,0133	1,0965	0,0494	3,8544	-0,0322	1,0604	-0,0718	3,9369
13	0,0165	1,0894	0,0517	3,6697	0,0170	1,0877	0,0226	3,7774
14	0,0056	1,0916	-0,0185	3,8521	0,0165	1,1251	0,0208	3,6927
15	-0,0133	1,0763	0,0624	3,6774	-0,0072	1,0920	0,0517	3,8070
16	-0,0076	1,1006	-0,0402	3,7228	-0,0012	1,0754	-0,0200	3,6535
17	-0,0032	1,0806	-0,0122	3,6957	0,0035	1,0570	0,0391	3,6657
18	0,0253	1,0891	0,0485	3,7168	0,0098	1,1040	0,0037	3,6231
19	-0,0056	1,0825	0,0571	3,7434	0,0063	1,1079	0,0984	3,9695
20	0,0131	1,1120	-0,0088	3,9215	0,0113	1,0511	0,0291	3,6649

CAPÍTULO 4

Influência

A identificação de observações com uma influência desproporcional nas estimativas dos parâmetros é uma componente fundamental do processo de validação do modelo, pois a presença deste tipo de observações pode evidenciar, por exemplo, que a inferência a partir do modelo ajustado pode ser inadequada. Por outro lado, é possível que o exame da estrutura das medidas de influência permita caracterizar as observações potencialmente influentes, ou seja, pode estabelecer os cenários nos quais o modelo considerado pode ser altamente influenciado. Por exemplo, no caso do modelo de regressão linear de resposta normal, podemos examinar as medidas de influência (veja Cook e Weisberg, 1982), concluindo que este modelo pode ser altamente influenciado por observações extremas e/ou localizadas em regiões remotas do espaço gerado pelas colunas da matriz modelo $\mathbf{X}$. De forma similar, é possível que nos modelos simétricos de regressão o exame das medidas de influência permita caracterizar as observações potencialmente influentes, podendo estudar a robustez do modelo considerado frente às observações extremas ou aberrantes. Portanto, o desenvolvimento e estudo das medidas de influência é muito importante.

Para os modelos simétricos de regressão, Galea, Paula e Uribe-Opazo (2003) estudam influência local em modelos com componentes sistemáticas lineares, enquanto que, Galea, Paula e Cysneiros (2005) avaliam influência local em modelos com componentes sistemáticas não-lineares. Entretanto, neste trabalho avaliamos influência seguindo o enfoque de influência global.

4.1 Modelo de deleção de casos

Uma aproximação importante para a identificação de observações influentes, baseia-se na metodologia denominada modelo de deleção de casos (CDM), proposta por Cook (1977) para os modelos lineares de resposta normal. Esta aproximação consiste em avaliar através de alguma medida apropriada, o efeito da exclusão de observações da análise nas estimativas dos parâmetros do modelo. Cook (1977) sugere utilizar a medida denominada *Distância de Cook*, que corresponde à norma para uma métrica especificada do vetor diferença $(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})$, sendo $\hat{\boldsymbol{\theta}}_{(i)}$ a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ quando a i-ésima observação tem sido excluída do modelo. Contudo, Cook e Weisberg (1982) definem uma medida mais geral para avaliar influência, denominada "afastamento da verossimilhança". Dado que a metodologia CDM é fácil de aplicar e de interpretar, tem sido estudada e estendida a muitos modelos de regressão. Por exemplo, Pregibon (1981) estende os resultados de Cook (1977) aos modelos lineares generalizados. Christensen, Pearson e Johnson (1992) avaliam influência em modelos mistos usando o modelo de deleção de casos (CDM). Galea, Riquelme e Paula (2000) usam o enfoque de influência global para avaliar a influência das observações nas estimativas dos parâmetros de locação e escala em modelos elípticos lineares. Tang, Wei e Wang (2000) estendem os resultados de Pregibon (1981) aos modelos não-lineares reprodutivos. Fung et *al* (2002) aplicam a metodologia CDM em modelos mistos semi-paramétricos. Sun e Wei (2004) avaliam através do enfoque de influência global a robustez frente a observações extremas da estimação baseada na norma L_1.

O modelo de deleção de casos (CDM) pode ser enunciado na seguinte forma

$$y_j = \mu_j(\boldsymbol{\beta}) + \epsilon_j, \qquad j \neq i. \tag{4.1}$$

Estudamos a influência da i-ésima observação na estimativa de $\boldsymbol{\theta}$, usando a medida denominada afastamento da verossimilhança, que pode ser expressa como

$$LD(\hat{\boldsymbol{\theta}}_{(i)}) = 2\big\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{(i)})\big\}, \tag{4.2}$$

em que $L(\boldsymbol{\theta})$ é o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$. Dado que o cálculo de $\hat{\boldsymbol{\theta}}_{(i)}$, $i = 1, \ldots, n$, pode ser computacionalmente custoso, especialmente quando n é grande, podemos obter uma aproximação destes valores maximizando uma

aproximação quadrática de $L_{(i)}(\boldsymbol{\theta})$, sendo esta última o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$ no modelo (4.1). Assim, expandindo $L_{(i)}(\boldsymbol{\theta})$ em Taylor até segunda ordem em torno de $\hat{\boldsymbol{\theta}}$, obtemos

$$L_{(i)}(\boldsymbol{\theta}) \approx L_{(i)}(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{U}_{(i)}(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T(-\ddot{\mathbf{L}}^{(i)}_{\hat{\theta}\hat{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

em que $\mathbf{U}_{(i)}(\boldsymbol{\theta}) = \partial L_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ e $\ddot{\mathbf{L}}^{(i)}_{\theta\theta} = \partial^2 L_{(i)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$. Se a matriz $-\ddot{\mathbf{L}}^{(i)}_{\hat{\theta}\hat{\theta}}$ é positiva definida, então a expressão acima atinge o máximo em

$$\hat{\boldsymbol{\theta}}^I_{(i)} = \hat{\boldsymbol{\theta}} - \left[\ddot{\mathbf{L}}^{(i)\,-1}_{\theta\theta}\mathbf{U}_{(i)}(\boldsymbol{\theta})\right]\Big|_{\theta=\hat{\theta}}.$$

Substituindo $-\ddot{\mathbf{L}}^{(i)}_{\hat{\theta}\hat{\theta}}$ pela sua esperança $\mathbf{K}^{(i)}_{\theta\theta}$, temos que $\hat{\boldsymbol{\theta}}^I_{(i)}$ pode ser escrito na seguinte forma

$$\hat{\boldsymbol{\theta}}^I_{(i)} = \hat{\boldsymbol{\theta}} + \left[\mathbf{K}^{(i)\,-1}_{\theta\theta}\mathbf{U}_{(i)}(\boldsymbol{\theta})\right]\Big|_{\theta=\hat{\theta}}. \tag{4.3}$$

Esta alternativa para o cálculo de $\hat{\boldsymbol{\theta}}_{(i)}$ é conhecida como aproximação a um passo e foi introduzida por Pregibon (1981). Esta aproximação tem sido utilizada em modelos não-lineares, como por exemplo, em Tang, Wei e Wang (2000). Podemos interpretar que, se $\hat{\boldsymbol{\theta}}_{(i)}$ não é muito diferente de $\hat{\boldsymbol{\theta}}$ e $L_{(i)}(\boldsymbol{\theta})$ é localmente quadrática, a aproximação a um passo deveria estar próxima do valor de $\hat{\boldsymbol{\theta}}$.

Teorema 1. *Para o modelo (4.1), a aproximação a um passo para* $\hat{\boldsymbol{\theta}}_{(i)}$ *pode ser expressa como*

$$\hat{\boldsymbol{\theta}}^I_{(i)} = \begin{bmatrix}\hat{\boldsymbol{\beta}}^I_{(i)}\\ \hat{\phi}^I_{(i)}\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{\beta}} & -\hat{\phi}^{1/2}\rho(\hat{z}_i)\hat{z}_i(\mathbf{D}^T_{\hat{\beta}}\mathbf{D}_{\hat{\beta}})^{-1}\hat{\mathbf{d}}_i/(1-\hat{h}_{ii})\\ \hat{\phi} & -2\hat{\phi}(v(\hat{z}_i)\hat{z}_i^2 - 1)/((n-1)(4f_g - 1))\end{bmatrix} \tag{4.4}$$

Prova: A matriz de informação esperada de $\boldsymbol{\theta}$ no modelo (4.1) pode ser escrita na seguinte forma

$$\mathbf{K}^{(i)\,-1}_{\theta\theta} = \begin{bmatrix}\mathbf{K}^{(i)\,-1}_{\beta\beta} & \mathbf{0}\\ \mathbf{0} & \mathrm{K}^{(i)\,-1}_{\phi\phi}\end{bmatrix},$$

em que $\mathbf{K}^{(i)}_{\beta\beta} = \frac{4d_g}{\phi}(\mathbf{D}^T_\beta\Delta_i\mathbf{D}_\beta)$, $\mathrm{K}^{(i)}_{\phi\phi} = \frac{n-1}{4\phi^2}(4f_g - 1)$, e $\Delta_i = \mathrm{diag}\{\delta_1, \ldots, \delta_n\}$, com $\delta_i = 1$ e $\delta_j = 0$ para todo $j \neq i$. Então, a inversa da matriz $\mathbf{K}^{(i)}_{\beta\beta}$ pode ser expressa

como

$$\begin{aligned}\mathbf{K}_{\beta\beta}^{(i)\,-1} &= \frac{\phi}{4d_g}(\mathbf{D}_\beta^T\Delta_i\mathbf{D}_\beta)^{-1} \\ &= \frac{\phi}{4d_g}\Big[\,\mathbf{D}_\beta^T\mathbf{D}_\beta - \mathbf{d}_i\mathbf{d}_i^T\,\Big]^{-1} \\ &= \frac{\phi}{4d_g}\left[\,(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1} + \frac{(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{d}_i\mathbf{d}_i^T(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}}{1-h_{ii}}\,\right], \qquad (4.5)\end{aligned}$$

em que $\mathbf{d}_i = \partial\mu_i/\partial\boldsymbol{\beta}$ e $h_{ii} = \mathbf{d}_i^T(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{d}_i$. Por outro lado, temos que $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{(i)}$ verificam as seguintes equações

$$\mathbf{U}(\hat{\boldsymbol{\theta}}) = \sum_{j\neq i}\mathbf{U}^j(\hat{\boldsymbol{\theta}}) + \mathbf{U}^i(\hat{\boldsymbol{\theta}}) = \mathbf{0} \qquad (4.6)$$

e

$$\mathbf{U}_{(i)}(\hat{\boldsymbol{\theta}}_{(i)}) = \sum_{j\neq i}\mathbf{U}^j(\hat{\boldsymbol{\theta}}_{(i)}) = \mathbf{0}, \qquad (4.7)$$

sendo $\mathbf{U}^i(\boldsymbol{\theta}) = \partial\, l(y_i;\mu_i,\phi)/\partial\boldsymbol{\theta}$. Então, das expressões (4.6) e (4.7) temos que $\mathbf{U}_{(i)}(\hat{\boldsymbol{\theta}}) = -\mathbf{U}^i(\hat{\boldsymbol{\theta}})$, com $\mathbf{U}^i(\boldsymbol{\theta}) = \big(\mathbf{U}_\beta^{i\,T}(\boldsymbol{\theta}), \mathbf{U}_\phi^i(\boldsymbol{\theta})\big)^T$, em que $\mathbf{U}_\beta^i(\boldsymbol{\theta}) = \frac{4d_g}{\sqrt{\phi}}\rho(z_i)z_i\mathbf{d}_i$ e $\mathbf{U}_\phi^i(\boldsymbol{\theta}) = -1/2\phi + v(z_i)z_i^2/2\phi$ (veja expressões 2.9 e 2.10). Assim, substituindo (4.5) e $\mathbf{U}_{(i)}(\hat{\boldsymbol{\theta}})$ na equação (4.3) obtemos

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{(i)}^I = \begin{bmatrix}\hat{\boldsymbol{\beta}}_{(i)}^I \\ \hat{\phi}_{(i)}^I\end{bmatrix} &= \begin{bmatrix}\hat{\boldsymbol{\beta}} - \mathbf{K}_{\hat\beta\hat\beta}^{(i)\,-1}\mathbf{U}_\beta^i(\hat{\boldsymbol{\theta}}) \\ \hat{\phi} - \mathrm{K}_{\hat\phi\hat\phi}^{(i)\,-1}\mathbf{U}_\phi^i(\hat{\boldsymbol{\theta}})\end{bmatrix} \\ &= \begin{bmatrix}\hat{\boldsymbol{\beta}} \;-\hat{\phi}^{1/2}\rho(\hat{z}_i)\hat{z}_i(\mathbf{D}_{\hat\beta}^T\mathbf{D}_{\hat\beta})^{-1}\hat{\mathbf{d}}_i/(1-\hat{h}_{ii}) \\ \hat{\phi} \;-2\hat{\phi}(v(\hat{z}_i)\hat{z}_i^2-1)/((n-1)(4f_g-1))\end{bmatrix}\end{aligned}$$

□

Assim, podemos calcular uma aproximação do afastamento da verossimilhança $LD(\hat{\boldsymbol{\theta}}_{(i)})$, aplicando diretamente o Teorema 1 e substituindo (4.4) em (4.2), obtendo a seguinte expressão

$$LD(\hat{\boldsymbol{\theta}}_{(i)}^I) = 2\big\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{(i)}^I)\big\}.$$

4.1.1 Distância de Cook

Cook (1977) introduziu uma medida bem popular para identificar observações influentes, esta medida é conhecida como Distância de Cook. Supondo que $LD(\hat{\boldsymbol{\theta}}_{(i)})$ pode ser bem representada por uma função quadrática, consideramos uma aproximação de segunda ordem para esta função. Para isto, expandimos $L(\hat{\boldsymbol{\theta}}_{(i)})$ em Taylor até segunda ordem, obtendo

$$L(\hat{\boldsymbol{\theta}}_{(i)}) \approx L(\hat{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T \mathbf{U}(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T(-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}})(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Assim, podemos obter uma aproximação de $LD(\hat{\boldsymbol{\theta}}_{(i)})$ substituindo a expressão acima na equação (4.2). Esta aproximação, denominada $DG(\hat{\boldsymbol{\theta}}_{(i)})$, adota a seguinte expressão

$$DG(\hat{\boldsymbol{\theta}}_{(i)}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T\,[-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}\,]\,(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Entretanto, de uma maneira mais geral, $DG(\hat{\boldsymbol{\theta}}_{(i)})$ pode ser expressa na seguinte forma, a qual tem sido chamada de *Distância de Cook generalizada*

$$DG(\hat{\boldsymbol{\theta}}_{(i)}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T\mathbf{C}\,(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}), \tag{4.8}$$

em que $\mathbf{C}$ é uma matriz que além de ser positiva definida é assintoticamente equivalente à matriz de informação esperada de $\boldsymbol{\theta}$. Neste trabalho consideramos dois casos para $\mathbf{C}$, $-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}$ e $\mathbf{K}_{\theta\theta}$, as matrizes de informação observada e esperada de Fisher, respectivamente. Substituindo (4.4) em (4.8) obtemos uma aproximação para $DG(\hat{\boldsymbol{\theta}}_{(i)})$ que pode ser expressa como

$$DG(\hat{\boldsymbol{\theta}}^I_{(i)}) = (\hat{\boldsymbol{\theta}}^I_{(i)} - \hat{\boldsymbol{\theta}})^T\mathbf{C}\,(\hat{\boldsymbol{\theta}}^I_{(i)} - \hat{\boldsymbol{\theta}}).$$

Por outro lado, $DG(\hat{\boldsymbol{\theta}}_{(i)})$ pode ser modificada para que possa ser aplicada em situações em que um subconjunto do vetor de parâmetros $\boldsymbol{\theta}$ é de particular interesse. Então, se nosso interesse principal é avaliar a influência das observações na estimativa de $\boldsymbol{\beta}$, $DG(\hat{\boldsymbol{\theta}}_{(i)})$ pode ser expressa na seguinte forma

$$DG_\beta(\hat{\boldsymbol{\theta}}_{(i)}) = (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^T\big\{(\,\mathbf{I}_p, 0)\mathbf{C}^{-1}(\,\mathbf{I}_p, 0)^T\big\}^{-1}\,(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}), \tag{4.9}$$

em que $\mathbf{I}_p$ é a matriz identidade de ordem p. Similarmente a (4.8), podemos calcular uma aproximação para $DG_\beta(\hat{\boldsymbol{\theta}}_{(i)})$ substituindo $\hat{\boldsymbol{\beta}}_{(i)}$ por $\hat{\boldsymbol{\beta}}^I_{(i)}$.

Da forma similar a Cook (1977), notamos que é possível avaliar a magnitude de $DG(\hat{\boldsymbol{\theta}}_{(i)})$ observando que, assintoticamente

$$\left\{ \boldsymbol{\theta} : (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{T} \mathbf{C} \, (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^{2}_{p',q} \right\}$$

é uma região de confiança de $100q\%$ para $\boldsymbol{\theta}$, em que p' é a dimensão do vetor de parâmetros $\boldsymbol{\theta}$ (veja Cox e Hinkley 1974, cap. 9). Sendo assim podemos concluir que, se $DG(\hat{\boldsymbol{\theta}}_{(i)}) = \chi^{2}_{p',1-\alpha}$ para algum α, então a exclusão da i-ésima observação desloca a estimativa de $\boldsymbol{\theta}$ do centro da região de confiança descrita acima à elipse $\{\boldsymbol{\theta} : DG(\boldsymbol{\theta}) = \chi^{2}_{p',1-\alpha}\}$. Portanto, podemos definir uma versão padronizada de $DG(\hat{\boldsymbol{\theta}}_{(i)})$ na seguinte forma

$$DG^{*}(\hat{\boldsymbol{\theta}}_{(i)}) = \aleph\big[DG(\hat{\boldsymbol{\theta}}_{(i)})\big],$$

em que $\aleph[\,\cdot\,]$ é a função de distribuição acumulada de uma variável aleatória com distribuição qui-quadrado de p' graus de liberdade. Desta maneira, temos que $DG^{*}(\hat{\boldsymbol{\theta}}_{(i)}) \in (0,1)$. Além disso, $DG^{*}(\hat{\boldsymbol{\theta}}_{(i)})$ pode ser aproximada substituindo $\hat{\boldsymbol{\theta}}_{(i)}$ pela aproximação a um passo dada no Teorema 1.

No caso em que $\mathbf{C}$ é a matriz de informação esperada de $\boldsymbol{\theta}$, temos que $DG(\hat{\boldsymbol{\theta}}^{I}_{(i)})$ pode ser escrito na seguinte forma

$$\begin{aligned} DG(\hat{\boldsymbol{\theta}}^{I}_{(i)}) &= (\hat{\boldsymbol{\theta}}^{I}_{(i)} - \hat{\boldsymbol{\theta}})^{T} \mathbf{K}_{\hat{\theta}\hat{\theta}} \, (\hat{\boldsymbol{\theta}}^{I}_{(i)} - \hat{\boldsymbol{\theta}}) \\ &= \frac{(4d_g)\rho^{2}(\hat{z}_i)\hat{z}_i^{2}}{(1-\hat{h}_{ii})^{2}} \hat{\mathbf{d}}_i^{T} (\mathbf{D}_{\hat{\beta}}^{T} \mathbf{D}_{\hat{\beta}})^{-1} \hat{\mathbf{d}}_i + \frac{4\hat{\phi}^{2} \left(v(\hat{z}_i)\hat{z}_i^{2} - 1\right)^{2}}{(n-1)^{2}(4f_g - 1)^{2}} \mathrm{K}_{\hat{\phi}\hat{\phi}} \\ &= \frac{(4d_g)\rho^{2}(\hat{z}_i)\hat{z}_i^{2} \, \hat{h}_{ii}}{(1-\hat{h}_{ii})^{2}} + \frac{n\big(v(\hat{z}_i)\hat{z}_i^{2} - 1\big)^{2}}{(n-1)^{2}(4f_g - 1)}. \end{aligned}$$

Dado que os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ são assintoticamente ortogonais, podemos usar a expressão (4.9) para mostrar que o efeito da exclusão da i-ésima observação na estimativa de $\boldsymbol{\beta}$ pode ser bem representada pela seguinte expressão

$$DG_{\beta}(\hat{\boldsymbol{\theta}}^{I}_{(i)}) = \frac{(4d_g)\rho^{2}(\hat{z}_i)\hat{z}_i^{2} \, \hat{h}_{ii}}{(1-\hat{h}_{ii})^{2}}. \tag{4.10}$$

De maneira similar, podemos concluir que o efeito da retirada da i-ésima observação na estimativa de ϕ pode ser avaliada usando

$$DG_{\phi}(\hat{\boldsymbol{\theta}}^{I}_{(i)}) = \frac{n\big(v(\hat{z}_i)\hat{z}_i^{2} - 1\big)^{2}}{(n-1)^{2}(4f_g - 1)}. \tag{4.11}$$

Analogamente, podemos avaliar a influência da i-ésima observação na estimativa de $\boldsymbol{\beta}$ usando a matriz $\mathbf{C} = -\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}$. Então, de (4.9) e (2.11) temos que a distância de Cook generalizada pode ser expressa como

$$DG^{\star}_{\beta}(\hat{\boldsymbol{\theta}}^{I}_{(i)}) = \frac{\hat{\phi}\rho^2(\hat{z}_i)\hat{z}_i^2}{(1-\hat{h}_{ii})^2}\hat{\mathbf{d}}_i^T(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}\mathbf{C}_{\hat{\beta}}^{-1}(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}\hat{\mathbf{d}}_i, \tag{4.12}$$

com $\mathbf{C}_{\beta} = \phi\,\mathbf{M}^{-1} - \frac{\mathbf{A}\mathbf{A}^T}{\mathrm{E}}$, em que $\mathbf{M}$, $\mathbf{A}$ e E foram definidas no Capítulo 2.

Segundo as medidas baseadas em $DG(\hat{\boldsymbol{\theta}}^{I}_{(i)})$ (veja expressões 4.10 e 4.12), a influência da i-ésima observação na estimativa de $\boldsymbol{\beta}$ depende da estrutura do modelo de regressão através do valor de $\hat{h}_{ii}$, e da distribuição da variável resposta através da função $\rho(\cdot)$. Também é possível notar que, observações localizadas em regiões remotas do espaço gerado pelas colunas da matriz $\mathbf{D}_{\hat{\beta}}$ tenderão a assumir valores "grandes" nestas medidas. Além disso, é possível observar que o comportamento das medidas de influência depende do comportamento de $\rho(z)z$, a qual é uma função ímpar de z. Expressões para a função $\rho(z)z$ para algumas distribuições da classe simétrica podem ser encontradas no Quadro 4.1.

Quadro 4.1 *Expressões para $\rho(z)z$ em algumas distribuições da classe simétrica.*

Distribuição	$\rho(z)z$
Normal	z
t-Student	$(\nu+3)z/(\nu+z^2)$
t-Student generalizada	$s(r+3)z/(r(s+z^2))$
Logística tipo II	$\text{sinal}(z)\frac{3[1-\exp(-\lvert z\rvert)]}{1+\exp(-\lvert z\rvert)}$
Exponencial potência	$\text{sinal}(z)\frac{(1+k)\Gamma((k+1)/2)}{2^{1-k}\Gamma((3-k)/2)}\lvert z\rvert^{\frac{1-k}{k+1}}$

Como ilustração, apresentamos nas Figuras 4.1 a 4.4 os gráficos da função $\rho(z)z$ para algumas das distribuições pertencentes à classe simétrica. Em todos os casos a função considerada é comparada com a função $\rho(z)z$ da distribuição normal, a qual é representada pela linha pontilhada. Nestes gráficos podemos observar que para a distribuição t-Student a função $\rho(z)z$ tende para zero à medida que $|z|$ cresce, o que indica que, segundo $DG(\hat{\boldsymbol{\theta}}^{I}_{(i)})$, em modelos em que o

erro segue distribuição t-Student a influência da i-ésima observação na estimativa de $\boldsymbol{\beta}$ tende para zero à medida que $|\hat{z}_i|$ cresce. Um comportamento análogo é observado para a distribuição t-Student generalizada. Por outro lado, para a distribuição logística tipo II temos que embora a função $\rho(z)z$ seja crescente, esta é majorada, ou seja, $|\rho(z)z| < 3$ para todo $z \in I\!R$. Do Quadro 4.1 e da Figura 4.4 temos que para os modelos de resposta exponencial potência, a função $\rho(z)z$ aumenta quando $|z|$ cresce, porém, pode-se mostrar facilmente que à medida que k tende para 1, a taxa de crescimento da função $\rho(z)z$ tende para zero.

Por outro lado, podemos observar que segundo a expressão (4.11), a estimativa de ϕ é afetada pelas observaçoes em que o valor de $|z|$ é "grande" ou está próximo de zero. Além disso, observamos que o comportamento de $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$ depende da função $v(z)z^2$. Inicialmente, notamos que $v(z)z^2$ é uma função par de z, ou seja, esta função depende da magnitude e não do sinal de z. Para modelos em que o erro segue distribuição normal temos que $v(z)z^2 = z^2$, indicando que a estimativa de ϕ é altamente afetada pelas observações em que $|\hat{z}_i|$ é "grande". Para modelos em que o erro segue distribuição t-Student temos que $v(z)z^2 = [(\nu+3)/(\nu+z^2)]z^2$, podendo observar que $v(z)z^2$ tende para $(\nu+3)$ à medida que z^2 tende para infinito. Isto indica que segundo $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$, a influência da i-esima observação na estimativa de ϕ tem um limite máximo. Um comportamento análogo é observado para a distribuição t-Student generalizada, em que $v(z)z^2 = [(r+3)/(s+z^2)]z^2$ tende para $(r+3)$ à medida que z^2 tende para infinito. Para modelos em que o erro segue distribuições como logística tipo II e exponencial potência ($k > 0$), pode-se concluir que a função $v(z)z^2$ tende para infinito à medida que z^2 também tende para infinito, porém, a taxa de crescimento da função $v(z)z^2$ para estas distribuições é menor do que no modelo de resposta normal. Além disso, dado que $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$ somente depende da estrutura do modelo de regressão através dos valores de $\hat{\mu}_i$ e $\hat{\phi}$, esta medida é invariante sob transformações invertíveis do vetor de parâmetros $\boldsymbol{\beta}$, pois $\hat{\mu}_i$ e $\hat{\phi}$ são invariantes a essas transformações. Finalmente, pode-se observar que segundo (4.11), a influência da i-ésima observação na estimativa de ϕ tende para zero à medida que o tamanho da amostra cresce.

Figura 4.1 *Comportamento da função $\rho(z)z$ para modelos em que o erro segue distribuição t-Student com ν graus de liberdade.*

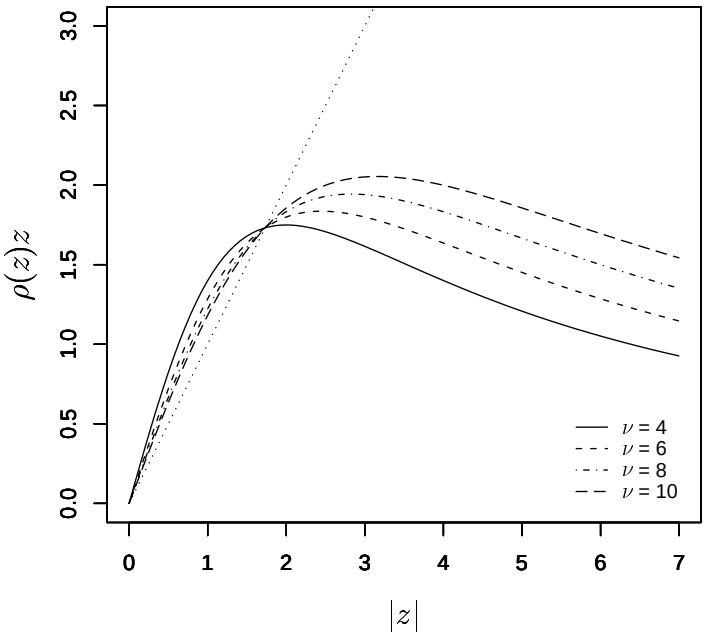


Figura 4.2 *Comportamento da função $\rho(z)z$ para modelos em que o erro segue distribuição t-Student generalizada de parâmetros s =4 e r.*

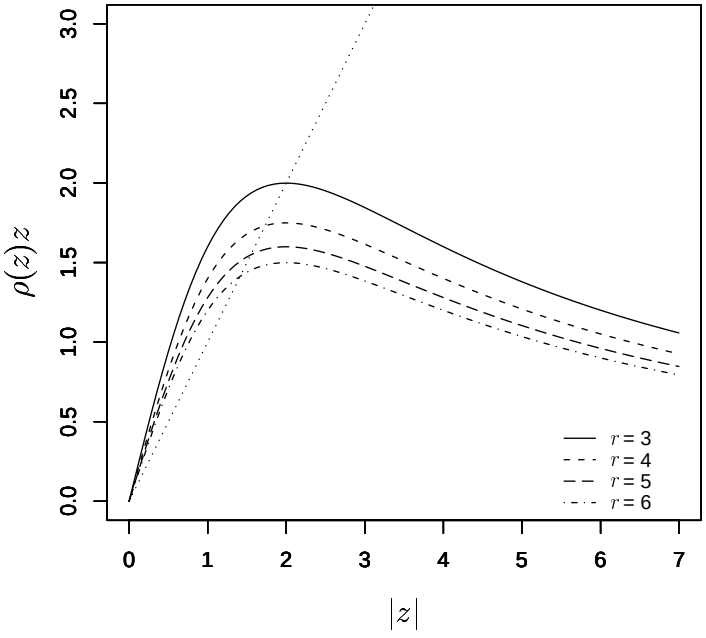


Figura 4.3 *Comportamento da função $\rho(z)z$ para modelos em que o erro segue distribuição logística tipo II.*

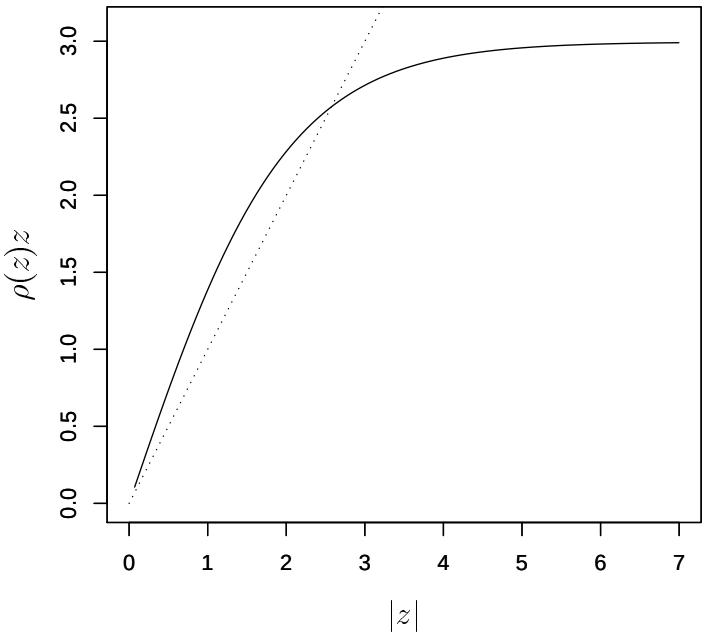
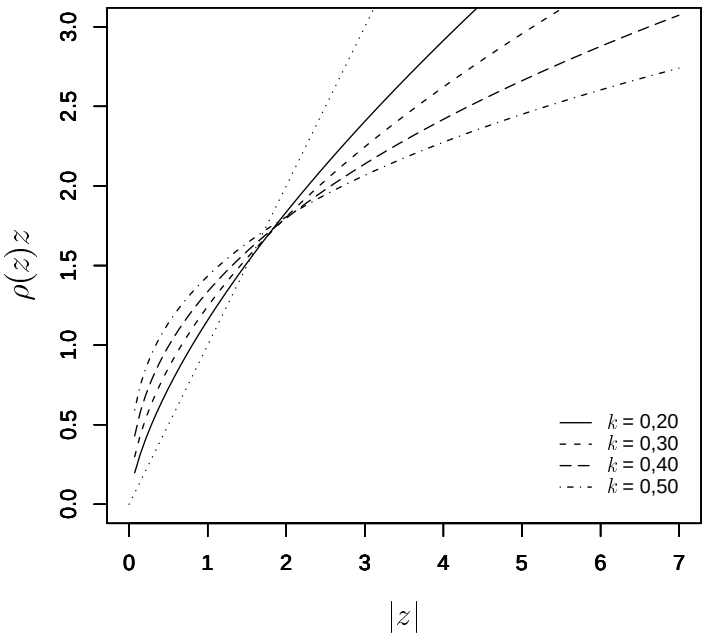


Figura 4.4 *Comportamento da função $\rho(z)z$ para modelos em que o erro segue distribuição exponencial potência de parâmetro k.*



4.1.2 Estatística W-K

Para medir a influência da i-ésima observação na estimativa do r-ésimo elemento do vetor de parâmetros $\boldsymbol{\beta}$, $r = 1, \ldots, p$, podemos considerar uma versão univariada da medida denominada distância de Cook, discutida anteriormente. Então, medimos a diferença entre $\hat{\beta}_r$ e $\hat{\beta}_{r\,(i)}$ usando, por exemplo, a métrica baseada na matriz de informação esperada de $\boldsymbol{\theta}$. Assim, da expressão (4.9) e do resultado enunciado no Teorema 1 temos que

$$\begin{aligned} \mathrm{WK}_i(\hat{\beta}_r^I) &= \frac{\left(\hat{\beta}_r - \hat{\beta}^I_{r\,(i)}\right)^2}{Var(\hat{\beta}_r)} \\ &= \frac{(4d_g)\rho^2(\hat{z}_i)\hat{z}_i^2}{(1-\hat{h}_{ii})^2} \frac{\left(\mathbf{a}^T(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}\hat{\mathbf{d}}_i\right)^2}{\mathbf{a}^T(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}\mathbf{a}}, \end{aligned}$$

com $Var(\hat{\beta}_r)$ a variância assintótica do estimador de máxima verossimilhança de β_r, e $\mathbf{a} = (a_1, \ldots, a_p)^T$, em que $a_r = 1$ e $a_s = 0$ para todo $s \neq r$. Valores "altos" de $\mathrm{WK}_i(\hat{\beta}_r^I)$ indicam que a i-ésima observação tem uma influência desproporcional na estimativa de β_r. A estatística $\mathrm{WK}_i(\hat{\beta}_r^I)$ foi considerada por Pregibon (1981) e Tang, Wei e Wang (2000) para os modelos lineares generalizados e os modelos não-lineares reprodutivos, respectivamente.

4.2 Modelo de mudança de médias

Como foi descrito na primeira seção do Capítulo 3, avaliar se a i-ésima observação na amostra pode ser considerada como outlier (observação extrema ou aberrante) consiste em testar a seguinte hipótese

$$\begin{aligned} H_0 &: \ \mu_i = \mu_i^\circ \\ H_1 &: \ \mu_i \neq \mu_i^\circ \end{aligned}$$

Nesta seção avaliamos a hipótese descrita acima levando em conta a estrutura do modelo de regressão, o que é feito testando hipóteses sobre a inclusão de parâmetros no modelo de interesse. Ou seja, testar se a i-ésima observação é um outlier consiste em avaliar a adequabilidade do seguinte modelo

$$\begin{cases} y_j = \mu_j(\boldsymbol{\beta}) + \epsilon_j, & j \neq i, \ j = 1, \ldots, n, \\ y_i = \mu_i(\boldsymbol{\beta}, \gamma) + \epsilon_i, \end{cases} \tag{4.13}$$

em que a função $\mu_i(\boldsymbol{\beta}, \gamma)$ é contínua e duplamente diferenciável com respeito ao vetor de parâmetros $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^T, \gamma)^T$. Adicionalmente, assumimos que existe γ° no espaço paramétrico tal que $\mu_i(\boldsymbol{\beta}, \gamma^\circ) = \mu_i(\boldsymbol{\beta})$. Assim, avaliar se a i-ésima observação pode ser considerada como outlier consiste em testar a seguinte hipótese

$$\begin{aligned} H_0 &: \ \gamma = \gamma^\circ \\ H_1 &: \ \gamma \neq \gamma^\circ \end{aligned} \tag{4.14}$$

A metodologia para identificar outliers descrita acima é denominada *modelo de mudança de médias* (MSOM), e tem sido utilizada por muitos autores, como, por exemplo, Cook e Weisberg (1982), Wei e Fung (1999), Tang, Wei e Wang (2000), Fung et *al* (2002), Sun e Wei (2004), entre outros. A hipótese em (4.14) pode ser avaliada usando o teste da razão de verossimilhanças. Este teste não depende da relação funcional entre μ_i e γ, ou seja, é invariante à parametrização de γ. Portanto, podemos assumir qualquer forma para a função $\mu_i(\boldsymbol{\beta}, \gamma)$, como, por exemplo, a seguinte

$$\begin{aligned} \mu_j &= \mu_j(\boldsymbol{\beta}), \qquad j \neq i, \ j = 1, \ldots, n, \\ \mu_i &= \mu_i(\boldsymbol{\beta}) + \gamma, \end{aligned} \tag{4.15}$$

ou

$$\boldsymbol{\mu} = \mu(\boldsymbol{\beta}) + b_i\gamma,$$

em que $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, e b_i é um vetor coluna de dimensão n com "um" na i-ésima posição e "zero" nos outros elementos. A componente sistemática (4.15) pode ser usada para aplicar o modelo de mudança de médias em modelos com componentes sistemáticas lineares. Desta maneira, γ é um parâmetro extra que indica a presença de outliers. Se o valor de γ é diferente de zero então a i-ésima observação é um outlier.

De Cox e Hinkley (1974), a estatística do teste da razão de verossimilhanças para avaliar a hipótese (4.14) pode ser escrita como

$$\xi_L = 2\{L(\hat{\boldsymbol{\theta}}_{mi}) - L(\hat{\boldsymbol{\theta}}^\circ)\},$$

em que $\hat{\boldsymbol{\theta}}_{mi} = (\hat{\boldsymbol{\beta}}_{mi}^T, \hat{\gamma}_{mi}, \hat{\phi}_{mi})^T$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ no modelo (4.13), e $\hat{\boldsymbol{\theta}}^\circ = (\hat{\boldsymbol{\beta}}^T, \gamma^\circ, \hat{\phi})^T$, sendo $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$ a estimativa de máxima

verossimilhança de $\boldsymbol{\theta}$ no modelo de interesse. Valores "grandes" na estatística ξ_L indicam que γ é significativamente diferente de γ°, evidenciando que a i-ésima observação é um outlier. Para grandes tamanhos da amostra, a estatística ξ_L pode ser comparada com a distribuição qui-quadrado com um grau de liberdade.

Teorema 2. *Para o modelo de deleção de casos (4.1) e o modelo para identificar outliers (4.13) se os respectivos estimadores de máxima verossimilhança são únicos, então para grandes tamanhos da amostra temos que* $\hat{\boldsymbol{\beta}}_{mi} = \hat{\boldsymbol{\beta}}_{(i)}$ *e* $\hat{\phi}_{mi} = \hat{\phi}_{(i)}$.

Prova: No modelo de mudança de médias (MSOM) temos que os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ verificam as seguintes equações

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \frac{1}{\sqrt{\phi}} \sum_{j \neq i} v(z_j) z_j \mathbf{d}_j + \frac{1}{\sqrt{\phi}} v(z_i) z_i \mathbf{d}_i = \mathbf{0}, \tag{4.16}$$

$$\mathbf{U}_\phi(\boldsymbol{\theta}) = -\frac{n}{2\phi} + \frac{1}{2\phi} \sum_{j \neq i} v(z_j) z_j^2 + \frac{1}{2\phi} v(z_i) z_i^2 = 0, \tag{4.17}$$

$$\mathbf{U}_\gamma(\boldsymbol{\theta}) = \frac{1}{\sqrt{\phi}} v(z_i) z_i \mathrm{d}_\gamma = 0, \tag{4.18}$$

em que $\mathrm{d}_\gamma = \partial \mu_i(\boldsymbol{\beta}, \gamma)/\partial \gamma$. Supondo que d_γ é diferente de zero em alguma região temos de (4.18) que $\mu_i(\hat{\boldsymbol{\beta}}_{mi}, \hat{\gamma}_{mi}) = y_i$, o que implica que $\hat{z}_i = 0$. Então, substituindo (4.18) nas equações (4.17) e (4.16) temos que $\hat{\boldsymbol{\beta}}_{mi}$ e $\hat{\phi}_{mi}$ verificam as seguintes equações, as quais não dependem da relação funcional entre $\mu_i(\boldsymbol{\beta}, \gamma)$ e o parâmetro γ

$$\sum_{j \neq i} v(z_j) z_j \mathbf{d}_j = \mathbf{0},$$
$$\frac{1}{n} \sum_{j \neq i} v(z_j) z_j^2 = 1.$$

Por outro lado, para o modelo de deleção de casos (CDM) temos que $\hat{\boldsymbol{\theta}}_{(i)}$ verifica as equações $\partial L_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \mathbf{0}$. Estas equações podem ser expressas na seguinte forma

$$\sum_{j \neq i} v(z_j) z_j \mathbf{d}_j = \mathbf{0},$$
$$\frac{1}{n} \sum_{j \neq i} v(z_j) z_j^2 = 1 - \frac{1}{n}.$$

Então, como os estimadores de máxima verossimilhança nos modelos de deleção de casos (CDM) e de mudança de médias (MSOM) são únicos, e para grandes tamanhos da amostra as equações que determinam estes estimadores são equivalentes, podemos concluir que para amostras grandes $\hat{\boldsymbol{\beta}}_{mi} = \hat{\boldsymbol{\beta}}_{(i)}$ e $\hat{\phi}_{mi} = \hat{\phi}_{(i)}$. □

4.2.1 Estudo de simulação

No Capítulo 3 deste trabalho observamos que em modelos onde o erro segue distribuições como t-Student, t-Student generalizada, exponencial potência e logística tipo II, as caudas da distribuição de $t_D(z_i)/\sigma_g$ estão muito próximas das caudas da distribuição normal padrão. Além disso, podemos usar o Lema 3 do Apêndice A para mostrar que a estatística $t_D^2(\hat{z}_i)$ converge em distribuição para $t_D^2(z_i)$. Assim, para grandes tamanhos da amostra, a estatística $t_D^{*2}(\hat{z}_i)$ pode ser aplicada para identificar outliers usando como referência as caudas da distribuição qui-quadrado com um grau de liberdade. Contudo, existem distribuições pertencentes à classe simétrica nas quais a proximidade entre as caudas da distribuição de $t_D(z_i)/\sigma_g$ e da normal não é verificada. Nesses casos podemos usar a estatística $t_Q^{*2}(\hat{z}_i)$, pois do Lema 3 do Apêndice A podemos concluir que $t_Q^2(\hat{z}_i)$ converge em distribuição para $t_Q^2(z_i)$. Assim, para grandes tamanhos da amostra, a estatística $t_Q^{*2}(\hat{z}_i)$ pode ser aplicada para identificar outliers usando como referência as caudas da distribuição qui-quadrado com um grau de liberdade.

Com o objetivo de avaliar e comparar o desempenho dos testes para identificar outliers baseados nas estatísticas ξ_L, $t_D^{*2}(\hat{z}_i)$ e $t_Q^{*2}(\hat{z}_i)$, desenvolvemos um estudo de simulação. Para este estudo consideramos o modelo de mudança de médias (MSOM) com a componente sistemática descrita pela expressão (4.15). Para a função $\mu_i(\boldsymbol{\beta})$ consideramos a curva de Michaelis-Menten (veja expressão 2.5). A componente aleatória está baseada na distribuição t-Student com $\nu = 4$ graus de liberdade. Os valores para $\gamma_s = \gamma/\phi^{1/2}$ são 0,5; 0,8; 1,38; 2,07; 2,77; 4,15; 5,54 e 6,92. Os parâmetros do modelo são $\beta_1 = 212$, $\beta_2 = 0,0641$ e $\phi = 1,5$. Os valores da variável explicativa são mantidos fixos ao longo do processo de simulação e são gerados seguindo distribuição uniforme no intervalo $(0, 1)$. Os tamanhos das amostras considerados são $n = 10, 20, 30$ e 50. O número de réplicas de Monte Carlo é 10000.

Inicialmente, comparamos o tamanho dos testes baseados nas estatísticas consideradas, isto é feito supondo que as caudas da distribuição qui-quadrado com um grau de liberdade podem ser usadas como referência. No Quadro 4.2 apresentamos os resultados desta comparação. Neste quadro podemos observar que para as estatísticas $t_D^{*2}(\hat{z}_i)$ e $t_Q^{*2}(\hat{z}_i)$ o tamanho do teste está muito próximo do nível nominal. Entretanto, para a estatística ξ_L as taxas empíricas de rejeição são significativamente maiores que os níveis nominais, indicando que para amostras pequenas as caudas da distribuição qui-quadrado não podem ser usadas como referência para esta estatística.

Quadro 4.2 *Tamanho dos testes para identificar outliers no modelo de Michaelis-Menten quando o erro segue distribuição t-Student com ν =4 graus de liberdade.*

n	α	$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L
10	0,10	0,1237	0,1244	0,2510
	0,05	0,0702	0,0702	0,1674
	0,01	0,0188	0,0184	0,0652
20	0,10	0,1167	0,1173	0,2087
	0,05	0,0672	0,0672	0,1312
	0,01	0,0154	0,0149	0,0468
30	0,10	0,1086	0,1091	0,1897
	0,05	0,0560	0,0560	0,1161
	0,01	0,0145	0,0145	0,0359
50	0,10	0,1055	0,1056	0,1822
	0,05	0,0522	0,0519	0,1079
	0,01	0,0123	0,0119	0,0349

Para garantir uma maior comparabilidade entre as estatísticas estudadas, obtemos os percentis da distribuição empírica de ξ_L, os quais são usados para construir a região de rejeição para o teste baseado nesta estatística. Para as estatísticas $t_D^{*2}(\hat{z}_i)$ e $t_Q^{*2}(\hat{z}_i)$ usamos as caudas da distribuição qui-quadrado com um grau de liberdade. Nos Quadros 4.3 a 4.6 apresentamos o poder empírico do teste para todos os valores de γ_s considerados. Notamos que na maioria dos casos, o poder do teste cresce à medida que o tamanho da amostra aumenta. Este

comportamento também é observado quando o valor de γ_s se afasta de zero. A principal conclusão do estudo de simulação é que o desempenho dos testes para identificar outliers baseados nas estatísticas $t_D^{*2}(\hat{z}_i)$, $t_Q^{*2}(\hat{z}_i)$ e ξ_L é muito similar. Entretanto, propomos usar os testes baseados nas estatísticas $t_D^{*2}(\hat{z}_i)$ e $t_Q^{*2}(\hat{z}_i)$, pois o teste da razão de verossimilhanças é mais "custoso" computacionalmente. Além disso, como foi observado no estudo de simulação, a comparação da estatística ξ_L com a distribuição qui-quadrado com um grau de liberdade pode ser inadequada para amostras pequenas, o qual dificulta a avaliação da magnitude da estatística ξ_L e em conseqüência a identificação de outliers.

Quadro 4.3 *Poder empírico do teste para identificar outliers no modelo de Michaelis-Menten para γ_s =0,5 e 0,8 quando o erro segue distribuição t-Student com ν =4 graus de liberdade.*

n	α	$\gamma_s = 0,5$			$\gamma_s = 0,8$		
		$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L	$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L
10	0,10	0,1398	0,1402	0,1303	0,1668	0,1675	0,1503
	0,05	0,0775	0,0775	0,0552	0,0926	0,0926	0,0655
	0,01	0,0220	0,0215	0,0114	0,0266	0,0261	0,0132
20	0,10	0,1331	0,1341	0,1246	0,1652	0,1657	0,1598
	0,05	0,0752	0,0751	0,0526	0,0875	0,0874	0,0645
	0,01	0,0183	0,0177	0,0112	0,0227	0,0220	0,0132
30	0,10	0,1232	0,1239	0,1132	0,1530	0,1537	0,1516
	0,05	0,0642	0,0642	0,0576	0,0786	0,0783	0,0695
	0,01	0,0171	0,0168	0,0114	0,0191	0,0185	0,0127
50	0,10	0,1234	0,1238	0,1163	0,1491	0,1493	0,1424
	0,05	0,0624	0,0621	0,0595	0,0764	0,0758	0,0719
	0,01	0,0144	0,0136	0,0107	0,0169	0,0159	0,0125

Quadro 4.4 *Poder empírico do teste para identificar outliers no modelo de Michaelis-Menten para γ_s =1,38 e 2,07 quando o erro segue distribuição t-Student com ν =4 graus de liberdade.*

n	α	$\gamma_s = 1,38$			$\gamma_s = 2,07$		
		$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L	$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L
10	0,10	0,2526	0,2535	0,2057	0,3991	0,4000	0,3827
	0,05	0,1489	0,1489	0,1002	0,2539	0,2539	0,2219
	0,01	0,0387	0,0385	0,0206	0,0733	0,0710	0,0692
20	0,10	0,2462	0,2470	0,2121	0,4069	0,4080	0,3945
	0,05	0,1489	0,1489	0,1046	0,2610	0,2607	0,2567
	0,01	0,0358	0,0346	0,0211	0,0689	0,0667	0,0583
30	0,10	0,2516	0,2520	0,2323	0,4335	0,4344	0,4131
	0,05	0,1402	0,1400	0,1215	0,2658	0,2650	0,2609
	0,01	0,0280	0,0268	0,0177	0,0554	0,0536	0,0532
50	0,10	0,2576	0,2580	0,2466	0,4679	0,4683	0,4649
	0,05	0,1317	0,1307	0,1249	0,2668	0,2651	0,2593
	0,01	0,0237	0,0229	0,0177	0,0429	0,0413	0,0410

Quadro 4.5 *Poder empírico do teste para identificar outliers no modelo de Michaelis-Menten para γ_s =2,77 e 4,15 quando o erro segue distribuição t-Student com ν =4 graus de liberdade.*

n	α	$\gamma_s = 2,77$			$\gamma_s = 4,15$		
		$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L	$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L
10	0,10	0,5728	0,5736	0,4960	0,8325	0,8333	0,7862
	0,05	0,3991	0,3989	0,2947	0,7000	0,7000	0,5926
	0,01	0,1323	0,1305	0,0658	0,3225	0,3182	0,1882
20	0,10	0,5986	0,5995	0,5498	0,8741	0,8745	0,8549
	0,05	0,4236	0,4235	0,3420	0,7693	0,7693	0,6865
	0,01	0,1363	0,1323	0,0765	0,3787	0,3717	0,2524
30	0,10	0,6375	0,6386	0,6132	0,9027	0,9029	0,8931
	0,05	0,4532	0,4522	0,4180	0,8027	0,8026	0,7773
	0,01	0,1161	0,1122	0,0669	0,3862	0,3769	0,2535
50	0,10	0,6886	0,6894	0,6736	0,9295	0,9297	0,9261
	0,05	0,4797	0,4787	0,4619	0,8457	0,8448	0,8353
	0,01	0,0909	0,0866	0,0656	0,3577	0,3453	0,2678

Quadro 4.6 *Poder empírico do teste para identificar outliers no modelo de Michaelis-Menten para γ_s =5,54 e 6,92 quando o erro segue distribuição t-Student com ν =4 graus de liberdade.*

n	α	$\gamma_s = 5,54$			$\gamma_s = 6,92$		
		$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L	$t_Q^{*2}(\hat{z}_i)$	$t_D^{*2}(\hat{z}_i)$	ξ_L
10	0,10	0,9461	0,9464	0,9292	0,9843	0,9843	0,9792
	0,05	0,8811	0,8811	0,8159	0,9588	0,9587	0,9322
	0,01	0,5762	0,5695	0,3784	0,7737	0,7676	0,5985
20	0,10	0,9696	0,9697	0,9662	0,9917	0,9917	0,9908
	0,05	0,9357	0,9356	0,9047	0,9837	0,9837	0,9764
	0,01	0,6943	0,6876	0,5355	0,8935	0,8904	0,7999
30	0,10	0,9811	0,9813	0,9786	0,9951	0,9951	0,9947
	0,05	0,9557	0,9555	0,9488	0,9904	0,9904	0,9893
	0,01	0,7274	0,7189	0,5840	0,9227	0,9193	0,8496
50	0,10	0,9836	0,9836	0,9829	0,9949	0,9949	0,9946
	0,05	0,9660	0,9659	0,9646	0,9908	0,9908	0,9905
	0,01	0,7470	0,7355	0,6596	0,9390	0,9347	0,9051

4.3 Gráfico da variável adicionada

Um dos objetivos principais da aplicação de modelos de regressão para a análise estatística de dados consiste na obtenção de um modelo que descreva adequadamente a relação entre a variável resposta e as variáveis explicativas. Este modelo, além de considerar a menor quantidade de parâmetros possível, deve ter grande descrição do fenômeno em estudo. Para a obtenção deste modelo podemos estudar a adequabilidade da componente sistemática avaliando a entrada e saída de parâmetros do modelo. Assim, avaliamos se a inclusão de um novo parâmetro incrementa significativamente a capacidade descritiva do modelo. Por outro lado, avaliamos se o modelo pode ser expresso de uma maneira mais simples, ou seja, avaliamos se um modelo com um número de parâmetros menor, apresenta uma capacidade descritiva similar a do modelo atual. Entretanto, estes procedimentos não podem identificar diretamente as observações que estão contribuindo para a entrada ou saída de parâmetros do modelo, ou seja, não podem determinar se as conclusões obtidas dos testes de significância são atribuíveis ao efeito de umas poucas observações ou, se pelo contrário, são determinadas pela totalidade da

amostra. Cook e Weisberg (1982) e Wang (1985) descrevem métodos gráficos para identificar as observações que estão contribuindo e aquelas que estão se desviando da relação descrita pela hipótese avaliada. Assim, estes métodos podem ser usados conjuntamente com os testes de hipóteses para incrementar a eficácia do processo de procura do "melhor" modelo. Nesta seção apresentamos uma extensão para os modelos simétricos de regressão dos métodos gráficos descritos por Wang (1985) para os modelos lineares generalizados.

4.3.1 Testando a inclusão de parâmetros

Consideramos o modelo simétrico de regressão com a seguinte estrutura

$$y_i = \mu_i(\boldsymbol{\beta}) + \epsilon_i, \qquad i = 1, \ldots, n. \tag{4.19}$$

O estimador de máxima verossimilhança de $\boldsymbol{\theta}$ neste modelo é denotado como $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$. Assumimos que é de interesse avaliar a inclusão de um novo parâmetro na componente sistemática, ou seja, queremos avaliar a adequabilidade do seguinte modelo

$$y_i = \mu_i^*(\boldsymbol{\beta}, \gamma) + \epsilon_i, \qquad i = 1, \ldots, n. \tag{4.20}$$

Neste modelo, o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$ é denotado por $\hat{\boldsymbol{\theta}}_c = (\hat{\boldsymbol{\beta}}_c^T, \hat{\gamma}_c, \hat{\phi}_c)^T$. Para avaliar a adequabilidade deste modelo, supomos que existe γ°, ponto interior do espaço paramétrico, tal que $\mu_i^*(\boldsymbol{\beta}, \gamma^\circ) = \mu_i(\boldsymbol{\beta})$. Então, avaliamos estatisticamente a adequabilidade do modelo (4.20) testando a seguinte hipótese

$$\begin{aligned} H_0 &: \ \gamma = \gamma^\circ \\ H_1 &: \ \gamma \neq \gamma^\circ \end{aligned} \tag{4.21}$$

Se γ é significativamente diferente de γ° então o modelo (4.20) será mais adequado que o modelo (4.19) para descrever o conjunto de dados em estudo. De (2.9) e (2.12) temos que a função de escore para γ e a matriz de informação esperada para $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^T, \gamma)^T$ no modelo (4.20) podem ser escritas como

$$\mathbf{U}_\gamma(\boldsymbol{\theta}) = \frac{1}{\phi}\mathbf{d}_\gamma^T \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu})$$

e

$$\mathbf{K}_{\beta^*\beta^*} = \frac{4d_g}{\phi}\big(\mathbf{D}_{\beta^*}^T\mathbf{D}_{\beta^*}\big),$$

em que $\mathbf{D}_{\beta^*} = (\,\mathbf{D}_\beta, \mathbf{d}_\gamma)$, sendo $\mathbf{d}_\gamma = \Big(\frac{\partial\mu_1^*}{\partial\gamma}, \ldots, \frac{\partial\mu_n^*}{\partial\gamma}\Big)^T$. Em conseqüência, a variância assintótica do estimador de máxima verossimilhança de γ pode ser expressa na seguinte forma

$$Var(\hat{\gamma}) = \frac{\phi}{4d_g}(\mathbf{R}_\gamma^T\mathbf{R}_\gamma)^{-1},$$

em que $\mathbf{R}_\gamma = (\mathbf{I} - \mathbf{H})\mathbf{d}_\gamma$ e $\mathbf{H} = \mathbf{D}_\beta(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{D}_\beta^T$. Assim, o teste escore para avaliar a hipóteses em (4.21) pode ser escrito na forma abaixo

$$\begin{aligned}
\xi_S &= \big\{\mathbf{U}_\gamma^T(\boldsymbol{\theta})Var(\hat{\gamma})\mathbf{U}_\gamma(\boldsymbol{\theta})\big\}\big|_{\theta=\hat{\theta}_c^\circ} \\
&= \left\{\frac{1}{\phi}(\mathbf{y}-\boldsymbol{\mu})^T\mathbf{D}(\mathbf{v})\mathbf{d}_\gamma\left[\frac{\phi}{4d_g}(\mathbf{R}_\gamma^T\mathbf{R}_\gamma)^{-1}\right]\frac{1}{\phi}\mathbf{d}_\gamma^T\mathbf{D}(\mathbf{v})(\mathbf{y}-\boldsymbol{\mu})\right\}\Big|_{\theta=\hat{\theta}_c^\circ} \\
&= \left\{\phi^{-1}(4d_g)(\mathbf{y}-\boldsymbol{\mu})^T\,\mathbf{D}(\boldsymbol{\rho})\mathbf{d}_\gamma(\mathbf{R}_\gamma^T\mathbf{R}_\gamma)^{-1}\mathbf{d}_\gamma^T\,\mathbf{D}(\boldsymbol{\rho})(\mathbf{y}-\boldsymbol{\mu})\right\}\Big|_{\theta=\hat{\theta}_c^\circ},
\end{aligned}$$

em que $\hat{\boldsymbol{\theta}}_c^\circ = (\hat{\boldsymbol{\beta}}^T, \gamma^\circ, \hat{\phi})^T$. Da convergência do processo iterativo de estimação de $\boldsymbol{\theta}$ no modelo (4.19) temos que

$$\big\{(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}\big\}|_{\theta=\hat{\theta}_c^\circ} = \big\{\mathbf{D}(\boldsymbol{\rho})(\mathbf{y}-\boldsymbol{\mu})\big\}|_{\theta=\hat{\theta}_c^\circ}, \tag{4.22}$$

em que $\tilde{\mathbf{Z}} = \mathbf{D}(\boldsymbol{\rho})(\mathbf{y}-\boldsymbol{\mu}) + \mathbf{D}_\beta\boldsymbol{\beta}$. Substituindo (4.22) na expressão para a estatística ξ_S obtemos

$$\xi_S = \frac{4d_g}{\hat{\phi}}\,\tilde{\mathbf{Z}}^T(\mathbf{I}-\hat{\mathbf{H}})\hat{\mathbf{d}}_\gamma(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{d}}_\gamma^T(\mathbf{I}-\hat{\mathbf{H}})\tilde{\mathbf{Z}},$$

em que $\hat{\mathbf{R}}_\gamma = (\mathbf{I}-\hat{\mathbf{H}})\hat{\mathbf{d}}_\gamma$, com $\hat{\mathbf{d}}_\gamma = \mathbf{d}_\gamma|_{\hat{\theta}^\circ}$. Dado que a matriz $\hat{\mathbf{H}}$ é simétrica e idempotente temos que

$$\xi_S = \frac{(4d_g)[(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_z]\,[(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_z]}{\hat{\phi}(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}},$$

com $\hat{\mathbf{R}}_z = \big(\rho(\hat{z}_1)(y_1-\hat{\mu}_1), \ldots, \rho(\hat{z}_n)(y_n-\hat{\mu}_n)\big)^T$. Assim, a estatística ξ_S pode ser considerada como a estatística de Wald para testar a significância do parâmetro

γ no seguinte modelo linear passando pela origem

$$\begin{cases} \hat{\mathbf{R}}_z = \gamma \hat{\mathbf{R}}_\gamma + \epsilon, \\ E(\epsilon) = 0, \\ Var(\epsilon) = \dfrac{\hat{\phi}}{4d_g}\mathbf{I}, \end{cases} \tag{4.23}$$

onde $\hat{\gamma} = (\hat{\mathbf{R}}_\gamma^T \hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{R}}_\gamma^T \hat{\mathbf{R}}_z$. Então, ξ_S é equivalente à estatística de Wald para avaliar a hipótese $H_0 : \gamma = 0$ vs. $H_1 : \gamma \neq 0$ no modelo (4.23). Portanto, um gráfico de $\hat{\mathbf{R}}_\gamma$ contra $\hat{\mathbf{R}}_z$ pode fornecer informações sobre a evidência dessa regressão, indicando quais observações estão contribuindo e quais estão se desviando da relação descrita pela hipótese (4.21).

4.3.2 Testando a exclusão de parâmetros do modelo

Seja $\mu_i = \mu_i(\boldsymbol{\beta}^*, \gamma)$ a componente sistemática do modelo em estudo. Nosso interesse é avaliar se o modelo pode ser formulado em uma maneira mais simples, em particular, queremos saber se a exclusão do parâmetro γ da componente sistemática do modelo afeta significativamente a capacidade descritiva do mesmo. Para isto, assumimos que existe γ°, ponto interior do espaço paramétrico, tal que $\mu_i(\boldsymbol{\beta}^*, \gamma^\circ) = \mu_i(\boldsymbol{\beta}^*)$. Então, testamos a seguinte hipótese

$$\begin{aligned} H_0 &: \ \gamma = \gamma^\circ \\ H_1 &: \ \gamma \neq \gamma^\circ \end{aligned} \tag{4.24}$$

Se γ é significativamente diferente de γ° então o modelo com a componente sistemática $\mu_i(\boldsymbol{\beta}^*, \gamma)$ será mais adequado para descrever o conjunto de dados em estudo. A hipótese (4.24) pode ser avaliada usando, por exemplo, a estatística de Wald. Da convergência do processo iterativo de estimação de $\boldsymbol{\theta}$ (veja expressão 2.16) temos que

$$\hat{\boldsymbol{\beta}} = (\mathbf{D}_{\hat{\beta}}^T \mathbf{D}_{\hat{\beta}})^{-1}\mathbf{D}_{\hat{\beta}}^T \tilde{\mathbf{Z}}, \tag{4.25}$$

em que $\mathbf{D}_\beta = (\mathbf{D}_{\beta^*}, \mathbf{d}_\gamma)$, $\mathbf{D}_{\beta^*} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}^*$, $\mathbf{d}_\gamma = \left(\dfrac{\partial\mu_1}{\partial\gamma}, \ldots, \dfrac{\partial\mu_n}{\partial\gamma}\right)^T$, e $\boldsymbol{\beta} = ({\boldsymbol{\beta}^*}^T, \gamma)^T$. Da expressão (4.25) temos que

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}^* \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} (\mathbf{D}_{\hat{\beta}^*}^T \mathbf{D}_{\hat{\beta}^*})^{-1}\mathbf{D}_{\hat{\beta}^*}^T\left[\mathbf{I} - (\hat{\mathbf{R}}_\gamma^T \hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{d}}_\gamma \hat{\mathbf{d}}_\gamma^T(\mathbf{I} - \hat{\mathbf{H}}^*)\right]\tilde{\mathbf{Z}} \\ -(\hat{\mathbf{R}}_\gamma^T \hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{d}}_\gamma^T(\mathbf{I} - \hat{\mathbf{H}}^*)\tilde{\mathbf{Z}} \end{bmatrix}, \tag{4.26}$$

com $\mathbf{H}^* = \mathbf{D}_{\beta^*}(\mathbf{D}_{\beta^*}^T\mathbf{D}_{\beta^*})^{-1}\mathbf{D}_{\beta^*}^T$. A estatística de Wald para testar a hipótese (4.24) pode ser escrita na seguinte forma

$$\xi_W = \frac{(\hat{\gamma} - \gamma^\circ)^2}{Var(\hat{\gamma})}. \tag{4.27}$$

Susbtituindo (4.26) em (4.27)

$$\xi_W = \frac{(4d_g)\big[(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{d}}_\gamma^T(\mathbf{I} - \mathbf{H}^*)\tilde{\mathbf{Z}} + \gamma^\circ\big]^2}{\hat{\phi}(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}}.$$

Como a matriz $\mathbf{H}^*$ é simétrica e idempotente podemos reescrever ξ_W como

$$\xi_W = \frac{(4d_g)[\,(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_z^*\,]^2}{\hat{\phi}(\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}},$$

em que $\hat{\mathbf{R}}_z^* = [\,(\mathbf{I} - \hat{\mathbf{H}}^*)\tilde{\mathbf{Z}} + \gamma^\circ\hat{\mathbf{R}}_\gamma\,]$. Assim, ξ_W pode ser considerada como a estatística de Wald para testar a significância do parâmetro γ no seguinte modelo linear passando pela origem

$$\begin{cases} \hat{\mathbf{R}}_z^* = \gamma\hat{\mathbf{R}}_\gamma + \epsilon, \\ E(\epsilon) = 0, \\ Var(\epsilon) = \dfrac{\hat{\phi}}{4d_g}\mathbf{I}, \end{cases} \tag{4.28}$$

onde $\hat{\gamma} = (\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_\gamma)^{-1}\hat{\mathbf{R}}_\gamma^T\hat{\mathbf{R}}_z^*$. Então, ξ_W é equivalente à estatística de Wald para avaliar a hipótese $H_0 : \gamma = 0$ vs. $H_1 : \gamma \neq 0$ no modelo (4.28). Portanto, um gráfico de $\hat{\mathbf{R}}_\gamma$ contra $\hat{\mathbf{R}}_z^*$ pode fornecer informações sobre a evidência dessa regressão, indicando quais observações estão contribuindo e quais estão se desviando da relação descrita pela hipótese (4.24).

CAPÍTULO 5

Exemplos

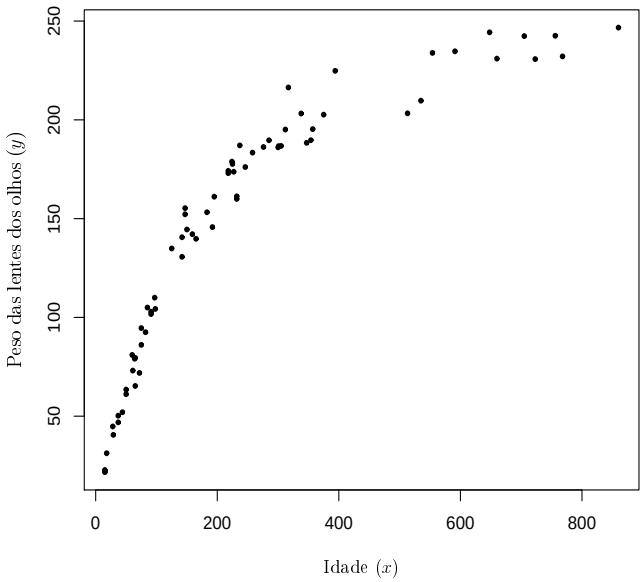
5.1 Coelhos Europeus na Austrália

Para ilustrar a aplicação das medidas estudadas neste trabalho, consideramos o conjunto de dados descrito em Ratkowsky (1983), cujo interesse principal é relacionar o peso das lentes dos olhos de coelhos europeus, y (em mg) e a idade do animal, x (em dias), em uma amostra de 71 observações. Este conjunto de dados pode ser encontrado no Apêndice C. A motivação para o uso deste conjunto de dados é a suspeita da presença de outliers quando este é analisado com o modelo de resposta normal. Este conjunto de dados tem sido analisado em Cysneiros, Paula e Galea (2005) e Galea, Paula e Cysneiros (2005) sob o enfoque de influência local. Estes autores propuseram analisar este conjunto de dados usando o seguinte modelo

$$y_i = \exp\left(\beta_1 - \frac{\beta_2}{x_i + \beta_3}\right) e^{\epsilon_i}, \qquad i = 1, \ldots, 71,$$

em que $\epsilon_i \sim S(0, \phi)$ são variáveis aleatórias mutuamente independentes. Consideramos três distribuições para o erro além da normal, que são as distribuições t-Student com $\nu = 4$ e 10 graus de liberdade e a distribuição logística tipo II. Na Figura 5.1 apresentamos um gráfico que descreve a relação entre a idade e o peso das lentes dos olhos dos coelhos europeus. Neste gráfico é possível observar que a variabilidade na resposta cresce com a idade do animal, justificando o uso de um modelo em que o erro é multiplicativo. No Quadro 5.1 apresentamos as estimativas de máxima verossimilhança dos parâmetros do modelo para todas as

Figura 5.1 *Gráfico de dispersão do peso das lentes dos olhos contra idade de coelhos europeus.*



distribuições do erro consideradas.

Quadro 5.1 *Estimativas de máxima verossimilhança (erro padrão) para alguns modelos simétricos ajustados aos dados dos coelhos europeus.*

	Parâmetro			
	β_1	β_2	β_3	ϕ
Normal	5,6399	130,5837	37,6028	0,0037
	(0,019)	(5,602)	(2,273)	(0,0006)
t-Student ($\nu = 10$)	5,6333	127,5395	36,0785	0,0028
	(0,018)	(5,096)	(2,061)	(0,0005)
t-Student ($\nu = 4$)	5,6313	126,2899	35,2958	0,0020
	(0,016)	(4,646)	(1,871)	(0,0004)
Logística tipo II	5,6330	127,2577	35,8639	0,0010
	(0,017)	(4,992)	(2,015)	(0,0002)

Neste quadro podemos observar que as estimativas de máxima verossimilhança dos parâmetros de locação são muito similares para todos os modelos considerados. Observamos também que à medida que a distribuição para o erro do modelo tem caudas mais pesadas, a estimativa do parâmetro de dispersão ϕ diminui, indicando que parte da variabilidade observada na resposta está sendo explicada pela distribuição assumida para o erro do modelo. Por outro lado, observamos que os parâmetros são altamente significativos em todos os modelos, porém, os erros padrões das estimativas dos parâmetros de locação são menores para o modelo de resposta t-Student com $\nu = 4$ graus de liberdade.

Com o objetivo de avaliar a qualidade do ajuste do modelo, bem como de avaliar graficamente possíveis afastamentos da suposição de homoscedasticidade, apresentamos na Figura 5.2 um gráfico dos valores ajustados contra o resíduo $t_D^*(\hat{z}_i)$ para todos os modelos considerados. Nestes gráficos não é possível observar um comportamento sistemático de $|t_D^*(\hat{z}_i)|$ em relação a $\hat{\mu}_i$, indicando que a variância da resposta não depende da média e que a suposição de variância constante é razoável. Por outro lado, notamos que as observações 4, 5, 16, e 17 aparecem em destaque para todos os modelos considerados, indicando que nenhum dos modelos descreve adequadamente o comportamento destas observações. Além disso, notamos que a qualidade do ajuste das observações 1, 2 e 3 é maior nos modelos de resposta t-Student e logística tipo II do que no modelo de resposta normal. Para as observações restantes concluímos que a qualidade do ajuste é muito similar em todos os modelos considerados.

Para avaliar conjuntamente a qualidade do ajuste para todas as observações na amostra, apresentamos na Figura 5.3 o gráfico normal de probabilidades com envelope do resíduo $t_D^*(\hat{z}_i)$ para todos os modelos considerados. Nestes gráficos podemos observar que a qualidade do ajuste global para os modelos de resposta normal e t-Student(10) é muito similar. Além disso, notamos que para os modelos de resposta t-Student(4) e logística tipo II todas as observações estão dentro das bandas de confiança. Entretanto, o comportamento do gráfico normal de probabilidades com envelope para o modelo de resposta t-Student(4) mostra uma leve superioridade, indicando que este modelo pode ser mais adequado para descrever o conjunto de dados em estudo.

Para a identificação de outliers consideramos as estatísticas ξ_L e $t_D^{*2}(\hat{z})$ discutidas no capítulo 4 deste trabalho. Na Figura 5.4 apresentamos um gráfico no qual

Figura 5.2 *Gráfico de resíduos $t_D^*(\hat{z}_i)$ contra os valores ajustados para o exemplo dos coelhos europeus.*

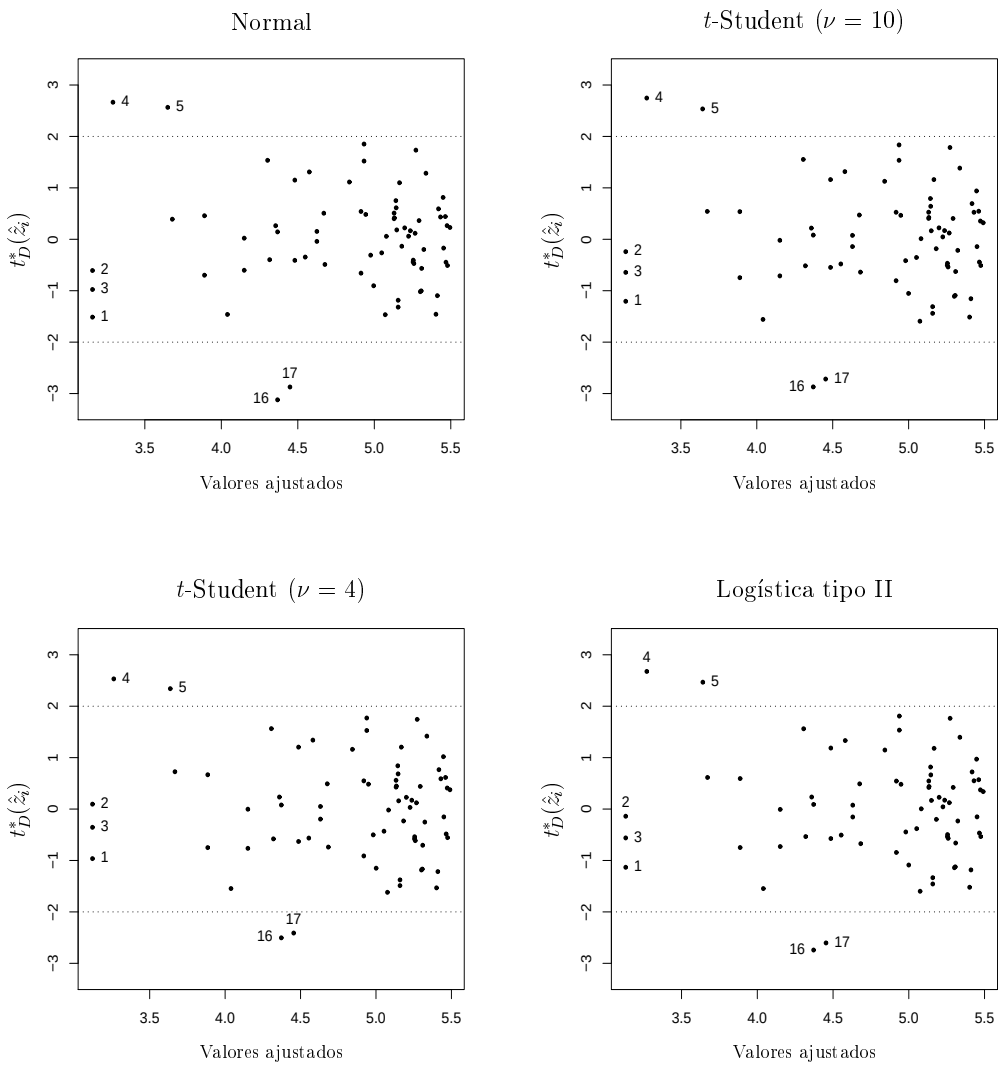
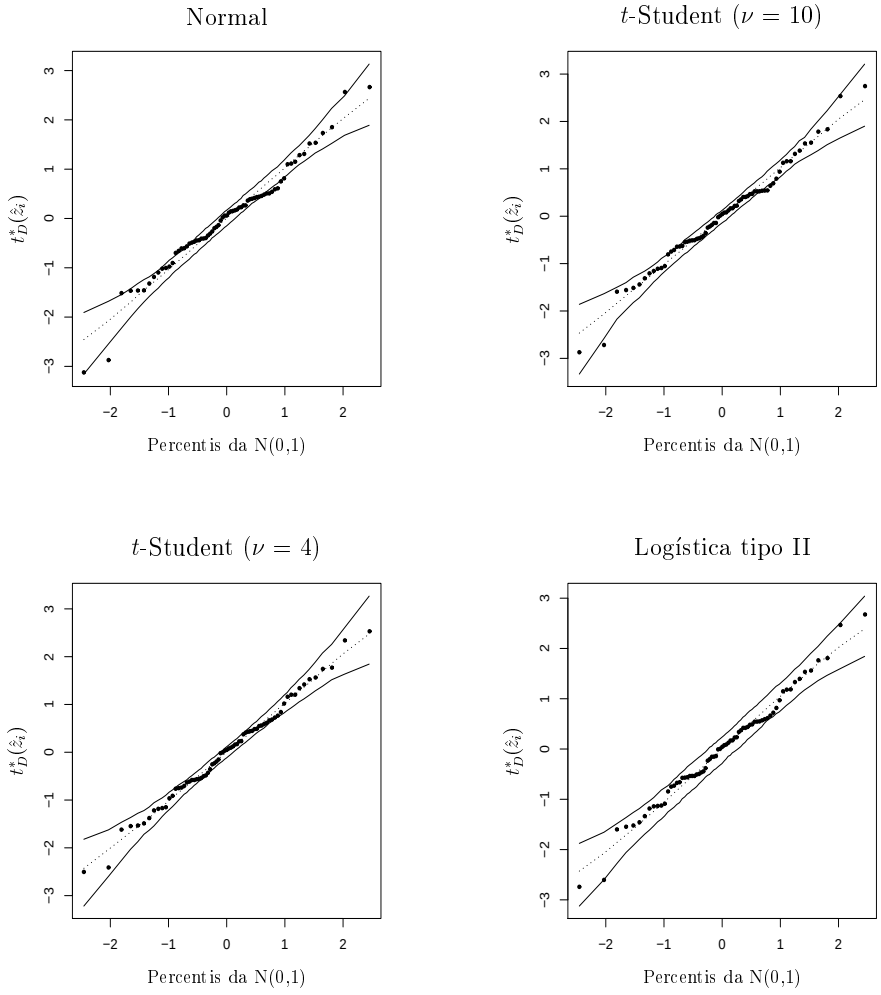
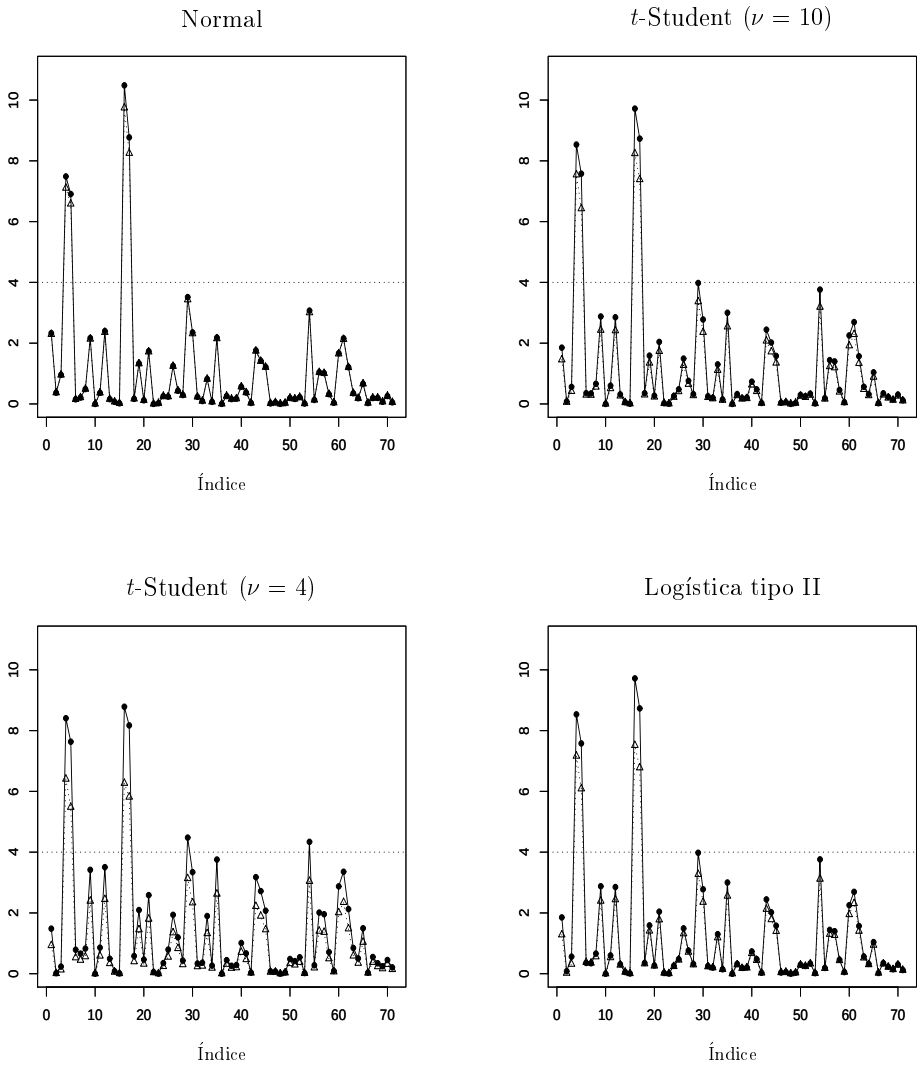


Figura 5.3 *Gráfico normal de probabilidades com envelope para o exemplo dos coelhos europeus.*



comparamos estas duas estatísticas. Inicialmente, observamos que as estatísticas ξ_L e $t_D^{*2}(\hat{z})$ são muito similares em todos os modelos considerados, indicando que a informação fornecida por estas é equivalente. Em segundo lugar, notamos que as observações 4, 5, 16, e 17 podem ser consideradas como outliers, pois estas observações aparecem destacadas em todos os gráficos.

Figura 5.4 *Gráfico das estatísticas ξ_L e $t_D^{*2}(\hat{z})$ contra o índice das observações para o exemplo dos coelhos europeus. As estatísticas ξ_L e $t_D^{*2}(\hat{z})$ são representadas por $\bullet$ e $\triangle$, respectivamente.*



A seguir, avaliamos a influência das observações nas estimativas dos parâmetros do modelo, para isto, usamos a medida denominada $DG(\hat{\boldsymbol{\theta}}^I_{(i)})$, desenvolvida no Capítulo 4 deste trabalho. Para avaliar a influência das observações na estimativa de $\boldsymbol{\beta}$, apresentamos na Figura 5.5 o gráfico de $DG_\beta(\hat{\boldsymbol{\theta}}^I_{(i)})$ contra o índice das observações para todos os modelos considerados. Podemos notar que as observações 1, 3, 4, 5, 16 e 17 aparecem destacadas para o modelo de resposta normal, evidenciando que a influência das mesmas na estimativa de $\boldsymbol{\beta}$ é significativamente maior que a influência das outras observações. Galea, Paula e Cysneiros (2005) identificam que os pontos 1, 2 e 3 são pontos de alavanca. Pode-se notar também que a influência destas observações é menor para os modelos em que a distribuição para o erro tem caudas mais pesadas do que a normal. Em particular, o modelo de resposta t-Student com 4 graus de liberdade se apresenta mais robusto, pois este modelo consegue reduzir a influência destas observações. Em segundo lugar, estudamos a influência das observações na estimativa do parâmetro de dispersão ϕ, para isto, apresentamos na Figura 5.6 o gráfico de $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$ contra o índice das observações para todos os modelos considerados. Nesta figura podemos notar que as observações 4, 5, 16 e 17 aparecem destacadas em todos os gráficos, o que evidencia que estas observações têm uma influência desproporcional na estimativa de ϕ. Entretanto, notamos também que a influência destas observações é menor para os modelos em que a distribuição assumida para o erro tem caudas mais pesadas do que a normal. Além disso, observamos que no modelo de resposta t-Student(4) a estimativa de ϕ é menos afetada pelas observações extremas do que nos outros modelos considerados. A principal conclusão desta análise de diagnóstico, é que o modelo de resposta t-Student(4) é mais robusto frente às observações aberrantes, portanto, este modelo pode ser mais adequado para descrever o conjunto de dados em estudo.

No Quadro 5.2 apresentamos a variação das estimativas de máxima verossimilhança dos parâmetros do modelo quando as observações 4, 5, 16 e 17 são excluídas. Vale salientar que a variação considerada corresponde a $100 \times (\hat{\beta}_{(I)} - \hat{\beta})/\hat{\beta}$, $I = \{4, 5, 16, 17\}$. Neste quadro podemos observar que, como esperado, a eliminação das observações extremas produz uma redução na estimativa do parâmetro de dispersão ϕ em todos os modelos. Notamos também que a eliminação destas observações produz menores mudanças nas estimativas do modelo de resposta t-Student(4) do que nas estimativas dos outros modelos considerados.

Figura 5.5 *Gráfico de* $DG_\beta(\hat{\boldsymbol{\theta}}^I_{(i)})$ *contra o índice das observações para o exemplo dos coelhos europeus.*

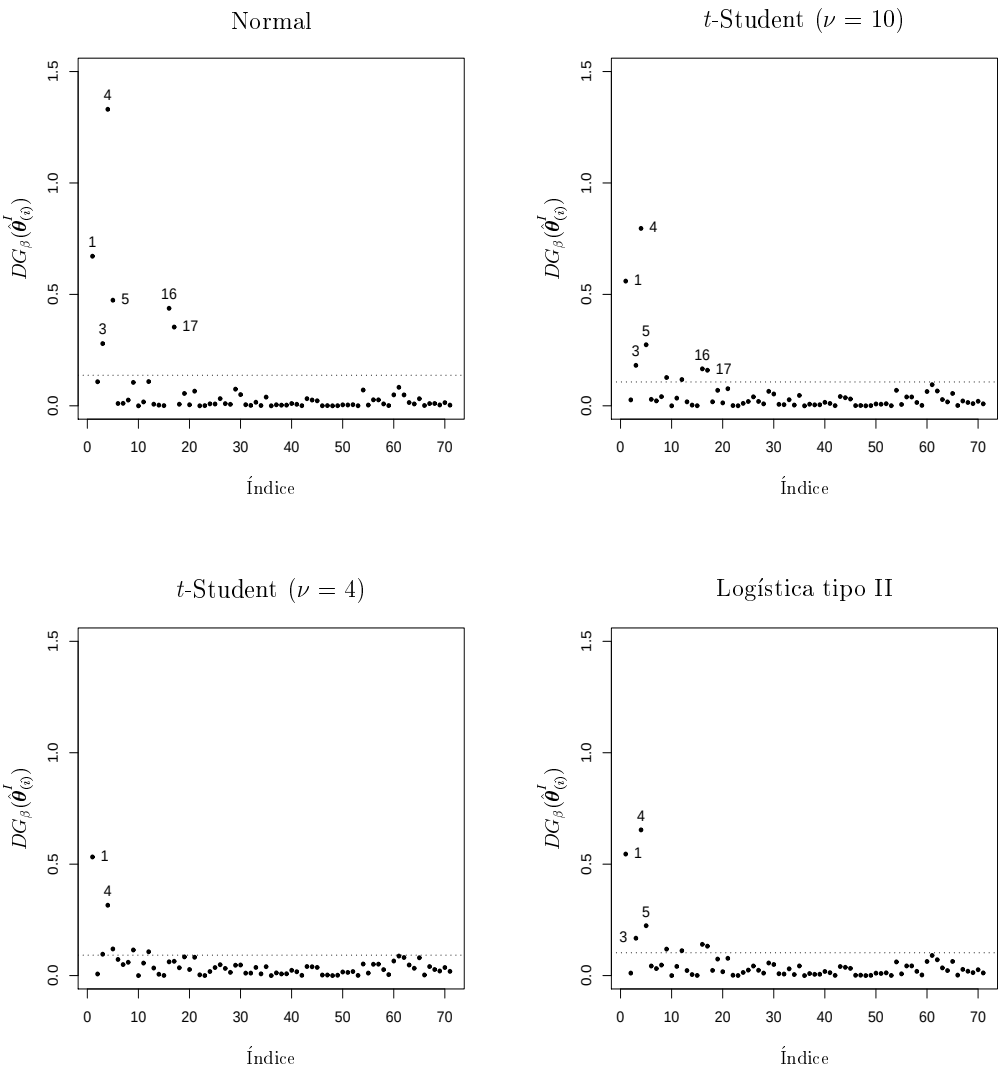
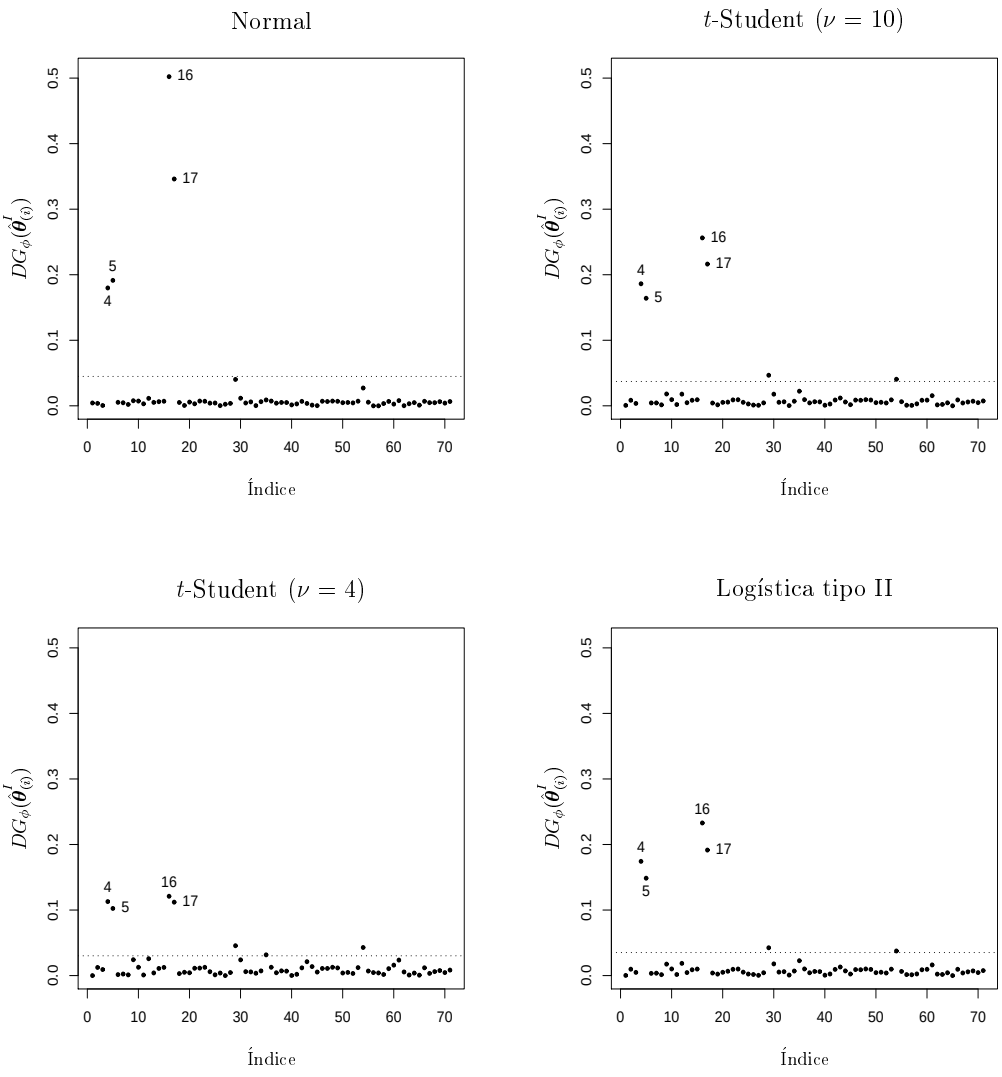


Figura 5.6 *Gráfico de* $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$ *contra o índice das observações para o exemplo dos coelhos europeus.*



Quadro 5.2 *Mudanças percentuais nas estimativas dos parâmetros nos modelos ajustados aos dados dos coelhos depois de excluídas as observações (4, 5, 16, 17).*

	Parâmetro			
	β_1	β_2	β_3	ϕ
Normal	-0,28	-5,63	-10,03	-43,13
t-Student ($\nu = 10$)	-0,13	-2,95	-5,62	-35,06
t-Student ($\nu = 4$)	-0,05	-1,32	-2,65	-26,98
Logística tipo II	-0,10	-2,43	-4,66	-34,23

Para avaliar a influência das observações na significância estatística dos parâmetros do modelo, utilizamos as ferramentas gráficas desenvolvidas no Capítulo 4 deste trabalho. Como ilustração, avaliamos a influência das observações na significância estatística de β_3. Para isto, apresentamos na Figura 5.7 um gráfico de $\hat{\mathbf{R}}_{\beta_3}$ contra $\hat{\mathbf{R}}_z^*$ em que podemos observar que para todos os modelos considerados a maioria das observações estão contribuindo para a significância estatística de β_3. Além disso, notamos que embora as observações 1, 2, 3 e 4 apresentem valores atípicos em $\hat{\mathbf{R}}_{\beta_3}$, estas contribuem para a significância estatística de β_3 em todos os modelos. Notamos também que no modelo de resposta normal, as observações 16 e 17 não estão contribuindo para a significância estatística de β_3. Concluímos que para todos os modelos considerados o parâmetros β_3 contribui significativamente para a descrição do comportamento da maioria das observações, o que justifica a presença deste parâmetro na componente sistemática dos modelos ajustados.

Concluímos que, dado que o modelo de resposta t-Student com $\nu = 4$ graus de liberdade é mais robusto frente às observações extremas do que os outros modelos considerados, este modelo pode ser mais adequado para descrever o fenômeno em estudo. Como ilustração, apresentamos na Figura 5.8 uma comparação das medidas de influência baseadas em $\hat{\boldsymbol{\theta}}_{(i)}$ com as baseadas na aproximação a um passo $\hat{\boldsymbol{\theta}}_{(i)}^I$ para o modelo de resposta t-Student(4). Nesta figura é possível observar que, na maioria dos casos, a apromixação a um passo apresenta um bom desempenho, pois a informação fornecida pelas medidas baseadas em $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}_{(i)}^I$ é equivalente. O programa na linguagem `R` utilizado para o ajuste e o cálculo das medidas de diagnóstico para o modelo de resposta t-Student(4) pode ser encontrado no Apêndice D.

Figura 5.7 *Gráfico do parâmetro excluído (β_3) para o exemplo dos coelhos europeus.*

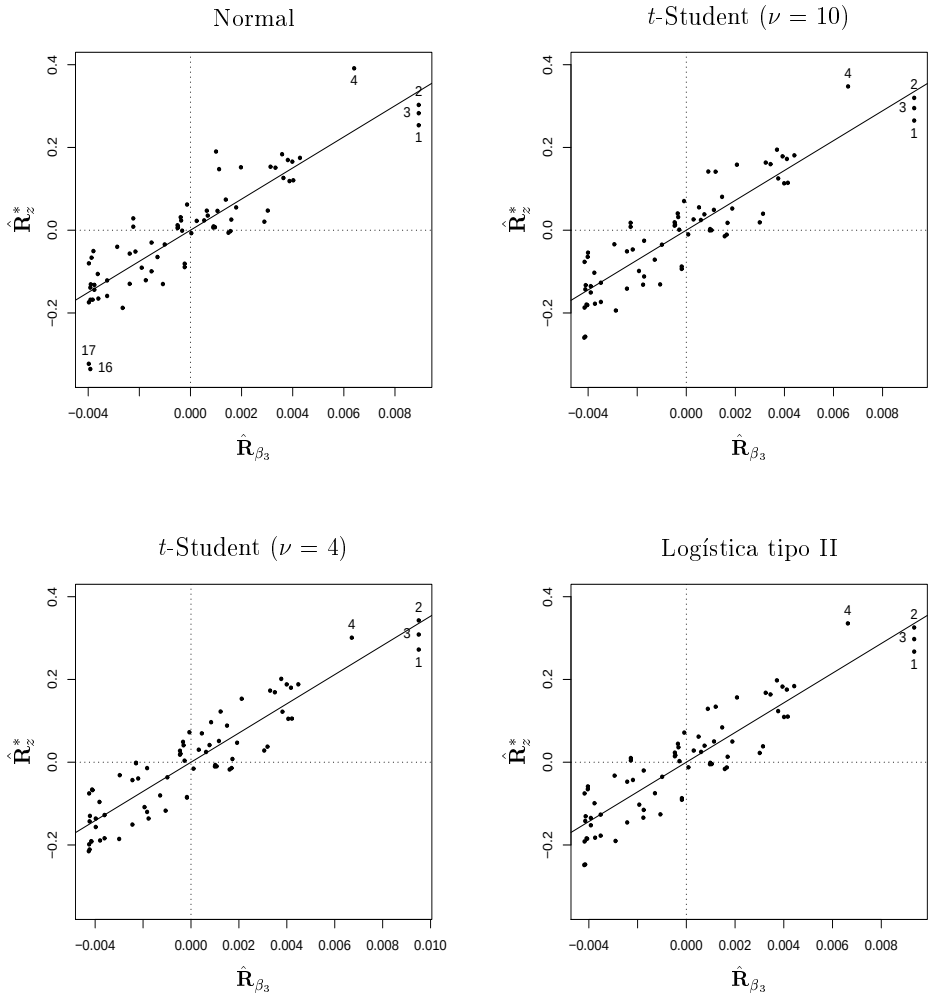
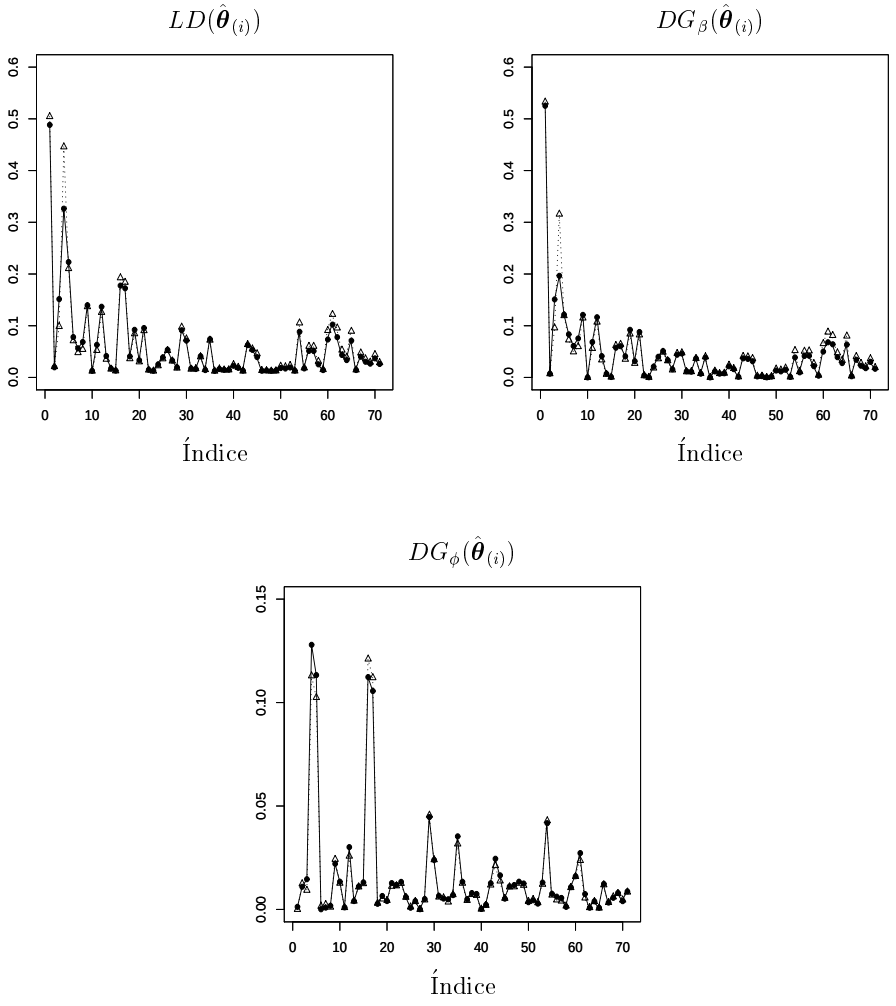


Figura 5.8 *Comparação das medidas de influência baseadas em* $\hat{\boldsymbol{\theta}}_{(i)}$ *com as baseadas na aproximação a um passo* $\hat{\boldsymbol{\theta}}^{I}_{(i)}$ *para o modelo de resposta t-Student(4) no exemplo dos coelhos. As medidas baseadas em* $\hat{\boldsymbol{\theta}}_{(i)}$ *e* $\hat{\boldsymbol{\theta}}^{I}_{(i)}$ *são representadas por* $\bullet$ *e* $\triangle$*, respectivamente.*



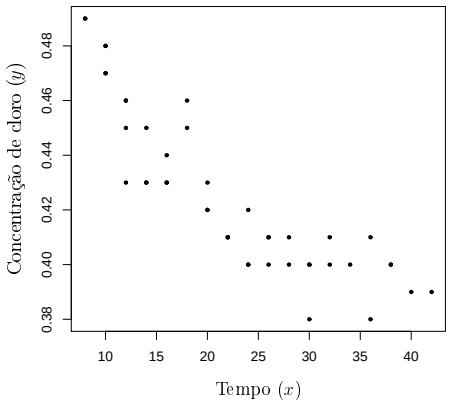
5.2 Concentração de Cloro

Como um segundo exemplo, consideramos o conjunto de dados analisado em Draper e Smith (1998), que corresponde a um experimento desenvolvido com o objetivo de explicar a concentração de cloro disponível num produto, y em relação ao tempo desde que o mesmo foi fabricado, x (em semanas), em uma amostra de 44 observações. Este conjunto de dados pode ser encontrado no Apêndice E. Smith e Dubey (1964) propuseram analisar este conjunto de dados usando o seguinte modelo

$$y_i = \beta_1 + (0,49 - \beta_1)\exp\big(-\beta_2(x_i - 8)\big) + \epsilon_i, \qquad i = 1, \ldots, 44,$$

em que $x_i \geq 8$, $\beta_1 < 0,49$ e $\epsilon_i \sim N(0, \phi)$ são variáveis aleatórias mutuamente independentes. O parâmetro β_2 representa o grau e a direção da associação entre a variável resposta e a variável explicativa. Se $\beta_2 > 0$, β_1 representa uma assíntota horizontal que corresponde ao valor mínimo que pode assumir a resposta. A motivação para o uso deste conjunto de dados é a presença de outliers quando o mesmo é analisado com o modelo de resposta normal. Sendo assim, consideramos duas distribuições para o erro além da normal, que são a distribuição t-Student com $\nu = 3$ graus de liberdade e a distribuição exponencial potência com $k = 0,5$. Como ilustração, apresentamos na Figura 5.9 um gráfico que descreve a relação entre a concentração de cloro disponível no produto e o tempo em semanas desde sua fabricação para uma amostra de 44 observações.

Figura 5.9 *Gráfico de dispersão da concentração de cloro disponível no produto e o tempo em semanas desde sua fabricação.*



No Quadro 5.3 apresentamos as estimativas de máxima verossimilhança dos parâmetros do modelo para todas as distribuições do erro consideradas. Neste quadro é possível observar que os parâmetros são "altamente" significativos em todos os modelos. Também podemos notar que os erros padrões das estimativas de máxima verossimilhança são menores para o modelo de resposta t-Student(3). Além disso, notamos que a estimativa de β_2 é maior para o modelo de resposta t-Student(3), indicando que a estimativa da força da associação entre a variável explicativa e a variável resposta é maior neste modelo.

Quadro 5.3 *Estimativas de máxima verossimilhança (erro padrão) para alguns modelos simétricos ajustados aos dados da concentração de cloro.*

Distribuição	Parâmetro		
	β_1	β_2	ϕ
Normal	0,3900	0,1000	0,00010
	(0,0049)	(0,0127)	(0,00002)
t-Student ($\nu = 3$)	0,3916	0,1064	0,00005
	(0,0037)	(0,0107)	(0,00002)
Exponencial potência ($k = 0,5$)	0,3911	0,1045	0,00004
	(0,0041)	(0,0115)	(0,00001)

Para avaliar a qualidade do ajuste de cada observação em cada modelo considerado, apresentamos na Figura 5.10 um gráfico dos valores ajustados contra o resíduo $t_D^*(\hat{z}_i)$. Nesta figura podemos notar que as observações 10, 17 e 18 aparecem destacadas em todos os gráficos, evidenciando que os modelos ajustados não descrevem adequadamente o comportamento destas observações. Além disso, para o modelo de resposta normal a observação 35 foi destacada. Entretanto, os modelos de resposta t-Student e exponencial potência conseguem uma melhor descrição destas observações do que o modelo de resposta normal. Para as observações restantes concluímos que a qualidade do ajuste é muito similar em todos os modelos considerados. Notamos também que nos gráficos de resíduos não é possível observar uma tendência de $|t_D^*(\hat{z}_i)|$ em relação aos valores ajustados, o que indica que a suposição de homoscedasticidade é razoável. Para avaliar a qualidade do ajuste conjuntamente para todas as observações, apresentamos

na Figura 5.11 o gráfico normal de probabilidades com envelope para todos os modelos considerados. No caso da distribuição normal pode-se observar que este modelo não descreve adequadamente o conjunto de dados. Para os outros modelos considerados os gráficos de envelope não apresentam nenhum comportamento não usual.

Figura 5.10 *Gráfico de resíduos* $t_D^*(\hat{z}_i)$ *contra os valores ajustados para o exemplo da concentração de cloro.*

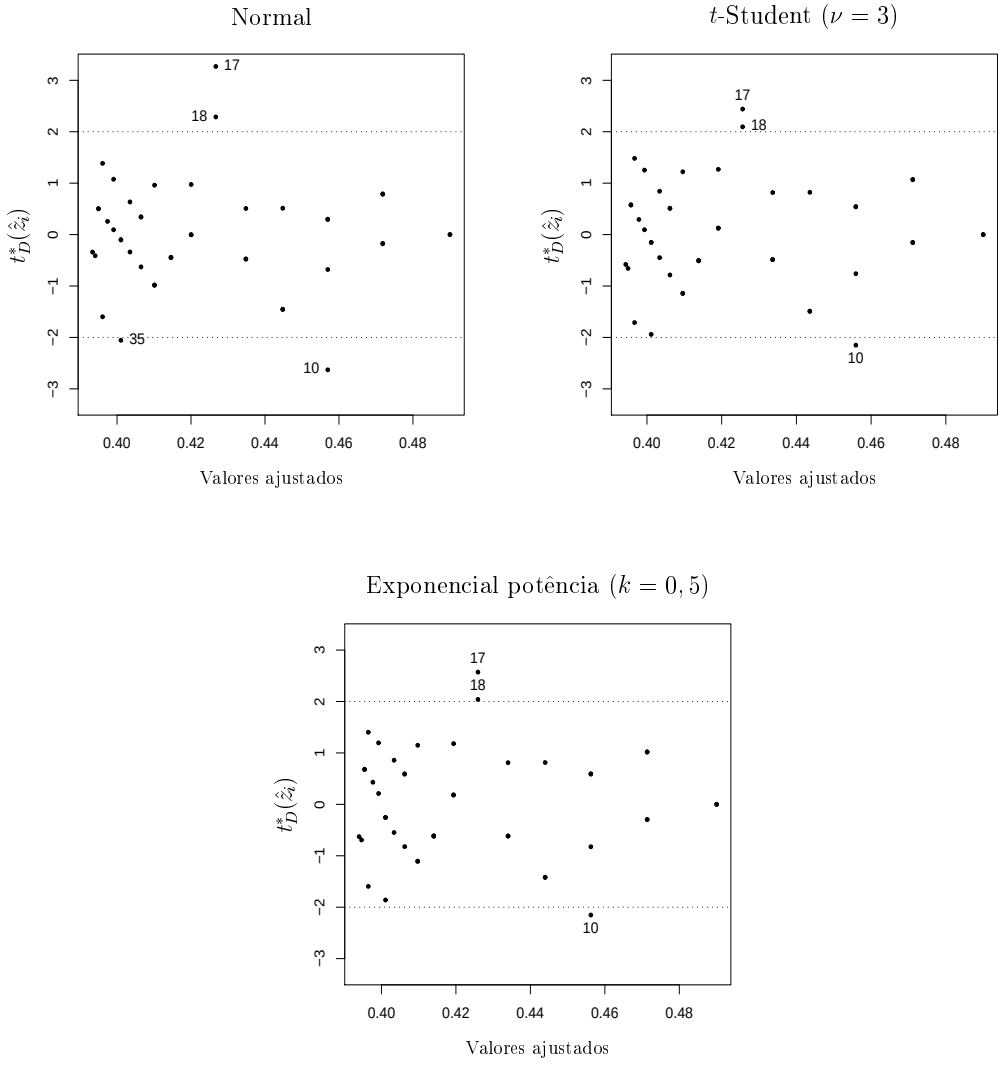
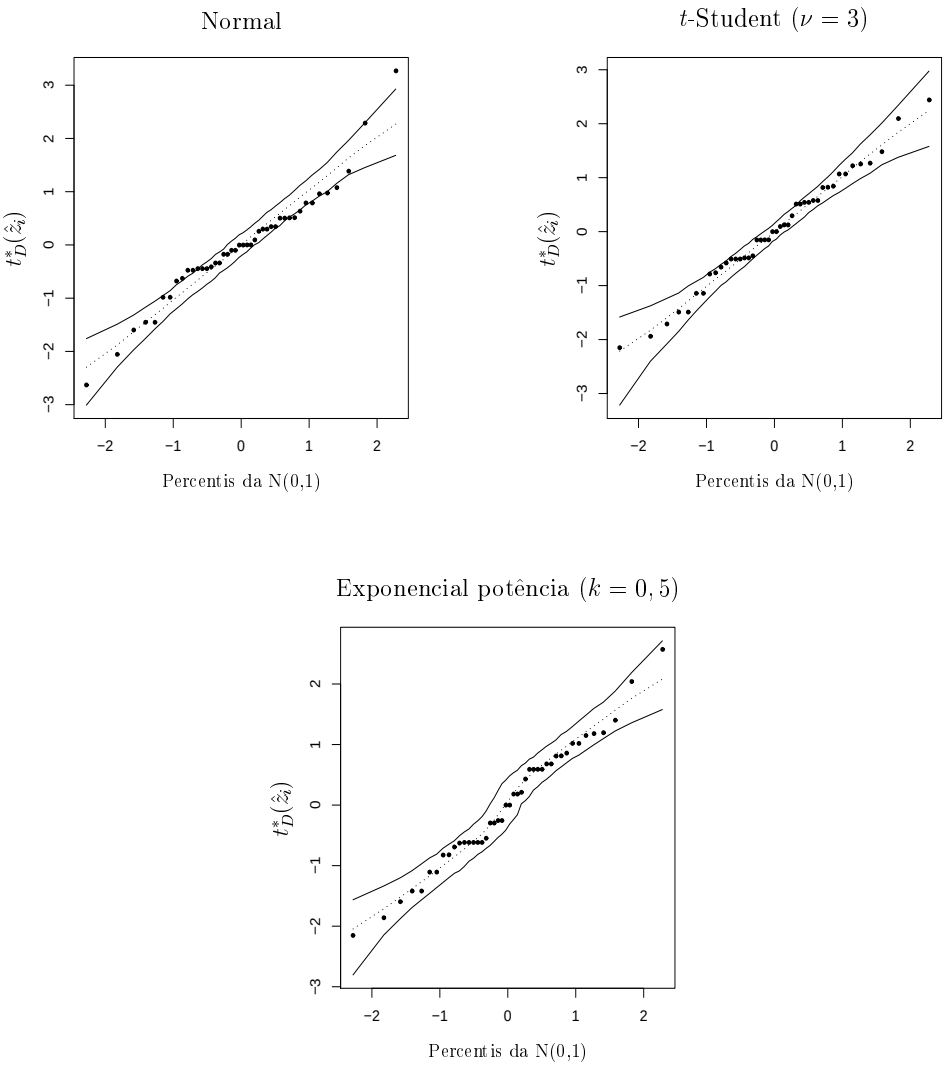


Figura 5.11 *Gráfico normal de probabilidades com envelope para o exemplo da concentração de cloro.*



Na Figura 5.12 podemos notar que para o modelo de resposta normal as observações 10, 17, 18, 35, 39 e 40 aparecem destacadas, evidenciando uma influência desproporcional destas observações na estimativa de $\boldsymbol{\beta}$. Além disso, para os modelos com caudas mais pesadas, os pontos destacados tem uma menor influência do que aqueles destacados pelo modelo normal. Vale salientar que as observações 39 e 40 também aparecem destacadas para os modelos de resposta t-Student e exponencial potência. A principal conclusão desta análise de diagnóstico é que os modelos em que o erro segue uma distribuição com caudas mais pesadas do que a normal são mais robustos. Em particular, o modelo de resposta t-Student(3) consegue controlar a influência das observações extremas na estimativa de $\boldsymbol{\beta}$. Em segundo lugar, estudamos a influência das observações na estimativa do parâmetro de dispersão ϕ, onde uma análise similar pode ser verificada (veja Figura 5.13). Sendo assim, o modelo t-Student(3) é o mais robusto que os outros modelos considerados.

Para avaliar a influência das observações na estimativa de β_2, apresentamos na Figura 5.14 um gráfico da estatística WK_i contra o índice das observações para todos os modelos considerados. Nesta Figura é possível observar que os modelos de resposta t-Student e exponencial potência são mais robustos do que o modelo normal. No Quadro 5.4 apresentamos a variação percentual das estimativas dos parâmetros quando as observações 10, 17 e 18 são excluídas. Neste quadro notamos que a retirada destas observações produz menores mudanças no modelo de resposta t-Student(3) do que nos outros modelos considerados. Um comportamento similar é obtido quando são excluídas as observações 10, 17, 18, 35, 39 e 40 (veja Quadro 5.5). Sendo assim, concluímos que o modelo t-Student(3) é mais adequado para descrever o conjunto de dados em estudo. Como ilustração, apresentamos na Figura 5.15 uma comparação das medidas de influência baseadas em $\hat{\boldsymbol{\theta}}_{(i)}$ com as baseadas na aproximação a um passo $\hat{\boldsymbol{\theta}}^I_{(i)}$ para o modelo de resposta t-Student(3). Nesta figura é possível observar que, na maioria dos casos, a apromixação a um passo apresenta um bom desempenho, pois a informação fornecida pelas medidas baseadas em $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}^I_{(i)}$ é equivalente.

Figura 5.12 *Gráfico de* $DG_\beta(\hat{\boldsymbol{\theta}}^I_{(i)})$ *contra o índice das observações para o exemplo da concentração de cloro.*

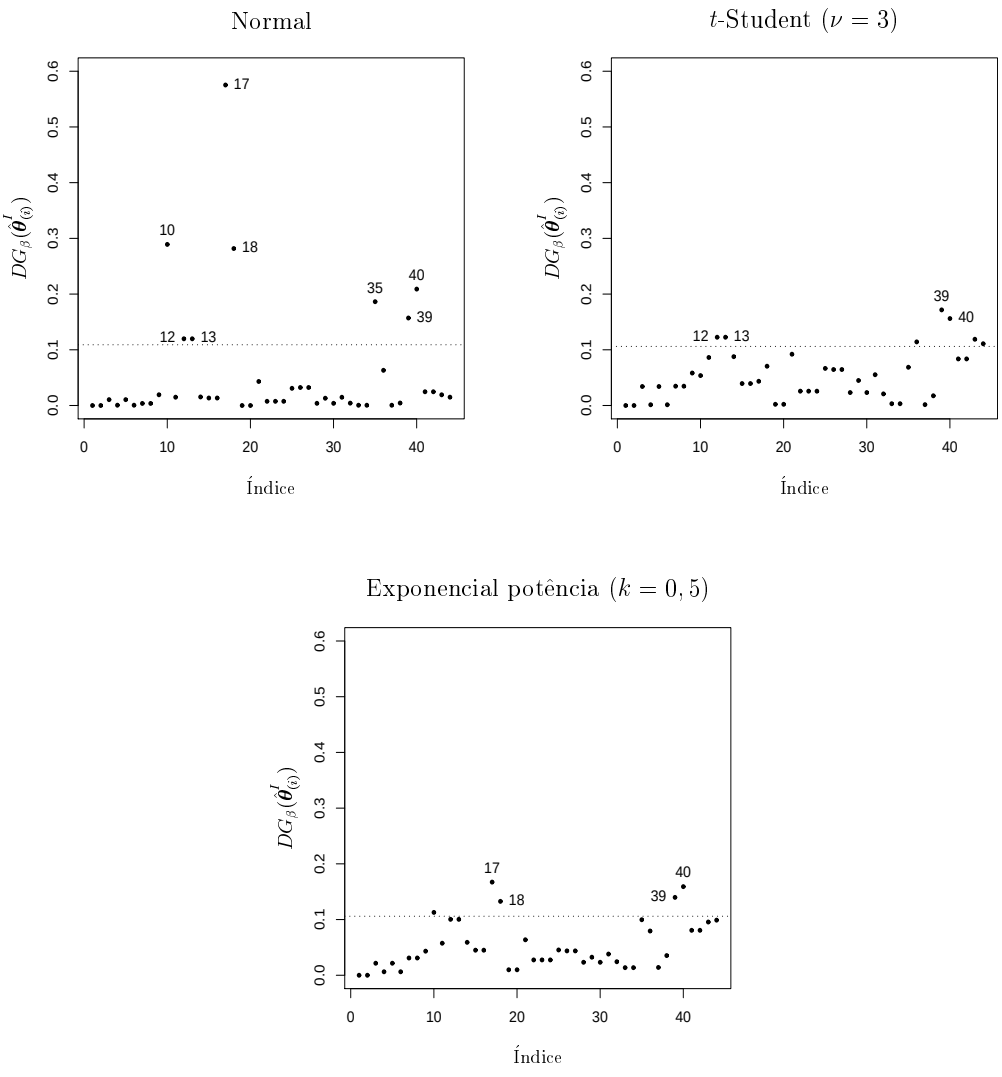


Figura 5.13 *Gráfico de* $DG_\phi(\hat{\boldsymbol{\theta}}^I_{(i)})$ *contra o índice das observações para o exemplo da concentração de cloro.*

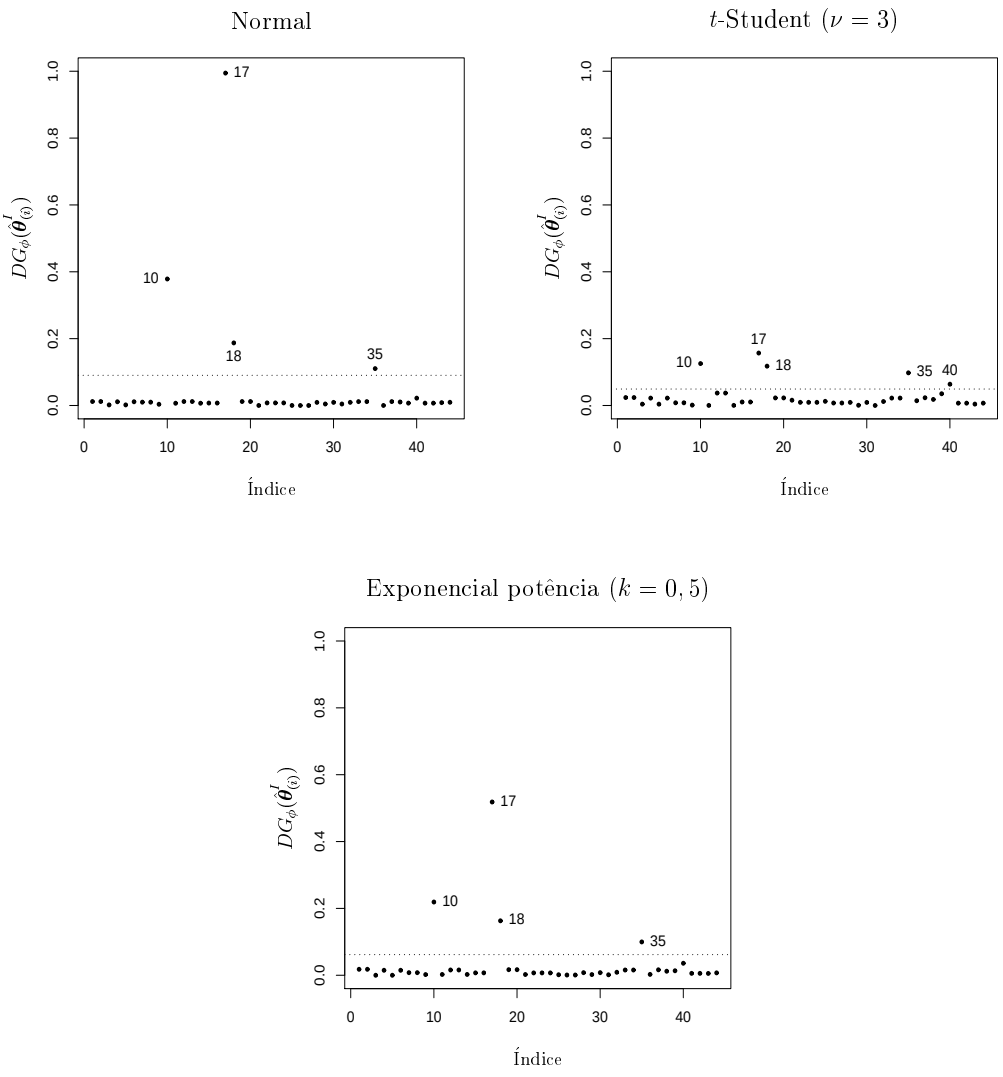
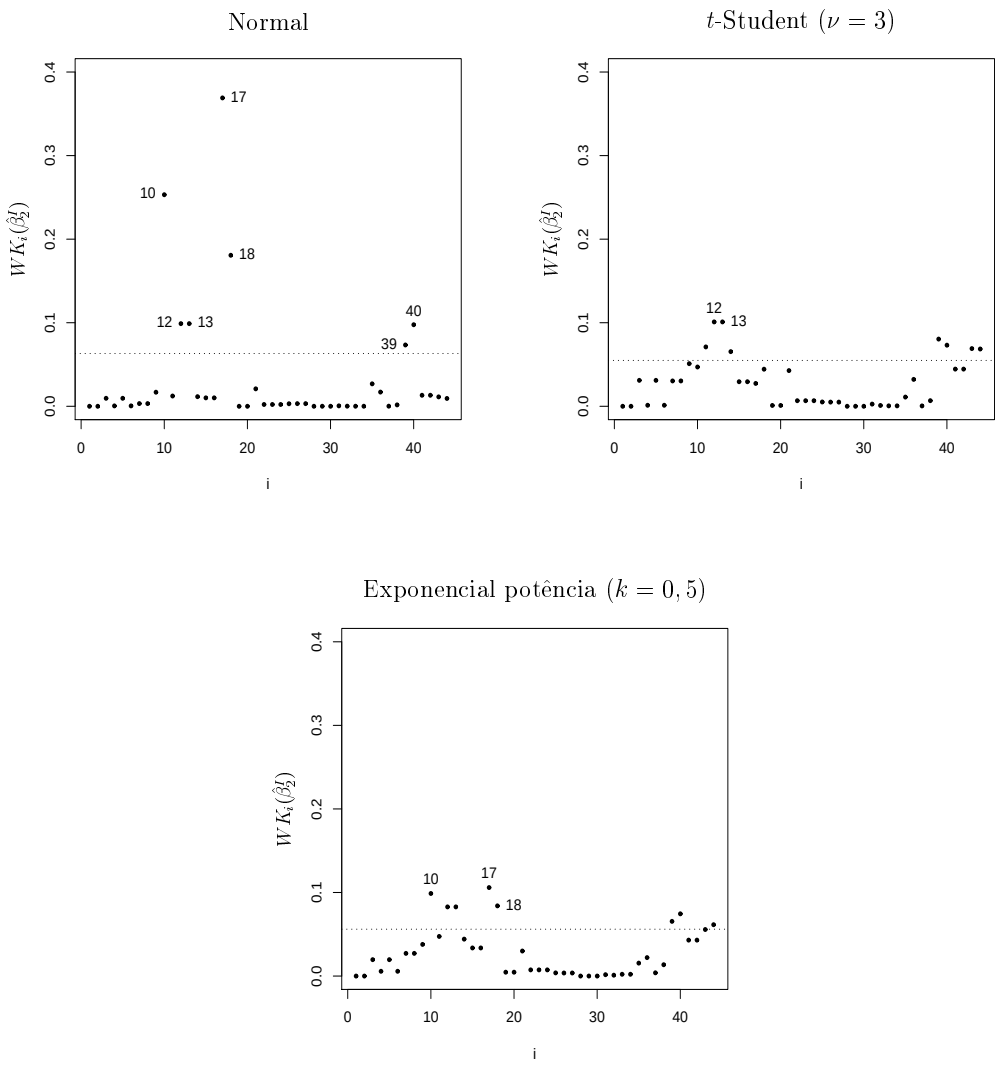


Figura 5.14 *Gráfico de* $WK_i(\hat{\beta}_2^I)$ *contra o índice das observações para o exemplo da concentração de cloro.*



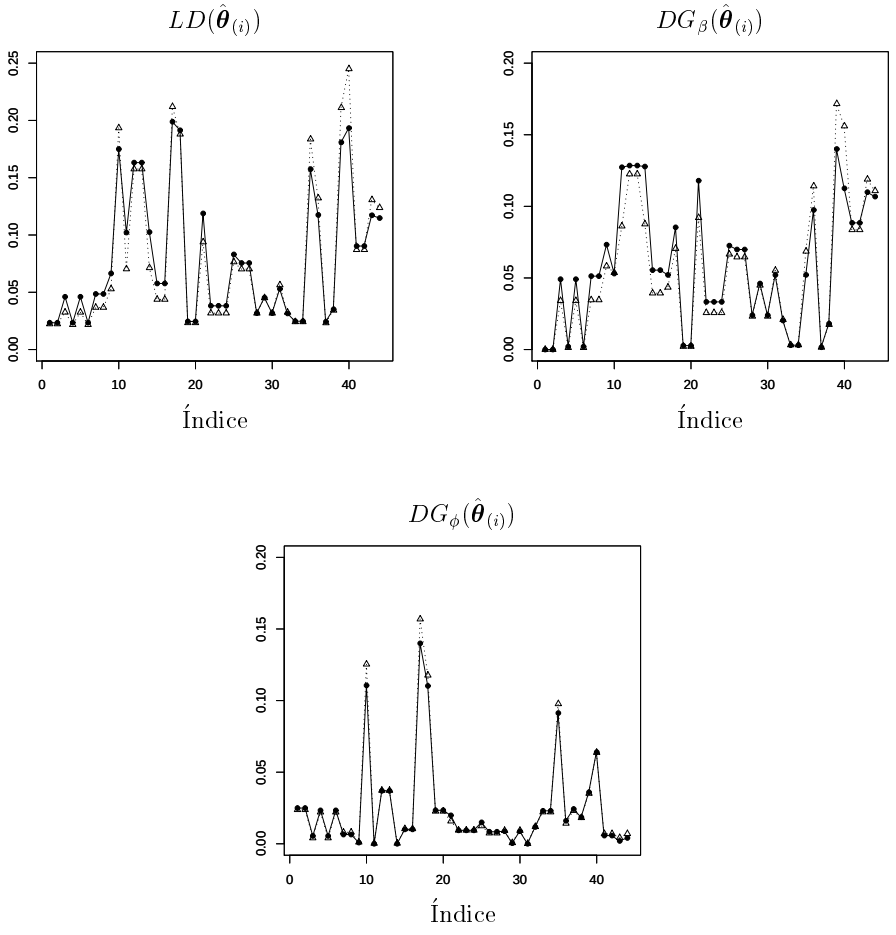
Quadro 5.4 *Mudanças percentuais nas estimativas dos parâmetros nos modelos ajustados aos dados da concentração de cloro depois de excluídas as observações (10, 17, 18).*

Distribuição	Parâmetro		
	β_1	β_2	ϕ
Normal	0,27	7,58	-47,27
t-Student ($\nu = 3$)	0,06	1,77	-28,63
Exponencial potência ($k = 0,5$)	0,23	5,87	-42,09

Quadro 5.5 *Mudanças percentuais nas estimativas dos parâmetros nos modelos ajustados aos dados da concentração de cloro depois de excluídas as observações (10, 17, 18, 35, 39, 40).*

Distribuição	Parâmetro		
	β_1	β_2	ϕ
Normal	0,79	11,00	-64,53
t-Student ($\nu = 3$)	0,17	2,32	-45,92
Exponencial potência ($k = 0,5$)	0,54	7,54	-59,08

Figura 5.15 *Comparação das medidas de influência baseadas em* $\hat{\boldsymbol{\theta}}_{(i)}$ *com as baseadas na aproximação a um passo* $\hat{\boldsymbol{\theta}}^{I}_{(i)}$ *para o modelo de resposta t-Student(3) no exemplo da concentração de cloro. As medidas baseadas em* $\hat{\boldsymbol{\theta}}_{(i)}$ *e* $\hat{\boldsymbol{\theta}}^{I}_{(i)}$ *são representadas por* $\bullet$ *e* $\triangle$*, respectivamente.*



CAPÍTULO 6

Conclusões

Inicialmente neste trabalho, discutimos o processo de estimação e de inferência nos modelos simétricos de regressão, o qual está baseado na metodologia de máxima verossimilhança. Observamos que nestes modelos as estimativas dos parâmetros de locação e escala devem ser obtidas através de um processo iterativo conjunto. Entretanto, no caso dos modelos de resposta exponencial potência, o estimador de máxima verossimilhança dos parâmetros de locação não depende do parâmetro de escala, sendo possível escrever o estimador do parâmetro de escala em forma fechada e como uma função do estimador dos parâmetros de locação. Vale salientar que como um caso particular da distribuição exponencial potência temos a distribuição normal.

Por outro lado, estudamos o efeito da especificação da componente aleatória do modelo no processo de estimação, encontrando que no caso dos parâmetros de locação este efeito pode ser avaliado através dos pesos $\rho(z_i^{(m)})$, em que $z_i^{(m)}$ é a diferença padronizada entre os valores observado e predito para a i-ésima observação no passo m do processo iterativo. Observamos que a função $\rho(\cdot)$, que atribui pesos às observações no processo iterativo, depende da magnitude e não do sinal de $z_i^{(m)}$. Observamos também que, se a função geradora de densidades $g(u)$ é monotonicamente decrescente para $u > 0$, então $\rho(\cdot)$ é uma função positiva. Além disso, observou-se que em modelos onde a resposta segue uma distribuição com caudas mais pesadas do que a normal, a função $\rho(z)$ decresce quando $|z|$ aumenta, indicando que nestes modelos as observações extremas têm um peso "pequeno" no processo de estimação. Vale salientar que para a distribuição normal a função $\rho(z) = 1$ para todo z.

Na segunda etapa deste trabalho, propusemos resíduos como medidas da qualidade do ajuste do modelo para cada observação na amostra. Consideramos resíduos definidos como $t(\hat{z}_i)$, para alguma função $t(\cdot)$ ímpar e duplamente diferenciável, em que $\hat{z}_i$ corresponde ao resíduo obtido aplicando aos modelos simétricos de regressão a definição geral de resíduos descrita em Cox e Snell (1968). Esta *família* de resíduos compreende, por exemplo, extensões para os modelos simétricos de regressão dos resíduos *componente de desvio* e *quantal*, definidos para os modelos lineares generalizados por Pregibon (1981) e Dunn e Smyth (1996), respectivamente. O resíduo proposto por Galea, Paula e Cysneiros (2005) é também um caso particular desta definição. Discutimos duas funções $t(\cdot)$ para as quais os resíduos obtidos apresentam propriedades estatísticas desejáveis. A primeira, com a que é possível definir o resíduo $t_D(\hat{z}_i)$, permite que para grandes tamanhos da amostra a versão padronizada deste resíduo possa ser aplicada utilizando os quantis da distribuição normal padrão como referência. Em segundo lugar, definimos o resíduo $t_Q(\hat{z}_i)$, que segue assintoticamente distribuição normal padrão.

Usando expansões de Cox e Snell (1968), observamos que até ordem n^{-1} em modelos com componentes sistemáticas lineares, o resíduo $t(\hat{z}_i)$ tem média zero para toda função $t(\cdot)$ ímpar e duplamente diferenciável. Para modelos com componentes sistemáticas não-lineares observamos que, se $r_i = (y_i - \hat{\mu}_i)$ tem media zero então $t(\hat{z}_i)$ também tem média zero. Por outro lado, obtimos que até ordem n^{-1} a variância de $t(\hat{z}_i)$ pode ser escrita como $Var(t(\hat{z}_i)) \approx Var(t(z_i))\{1-\tau^* h_{ii}\}$, sendo τ^* um fator de correção que depende da função $t(\cdot)$ e da distribuição do erro do modelo. Além disso, avaliamos as propriedades estatísticas dos resíduos propostos através de simulação de Monte Carlo, concluindo que as versões padronizadas dos resíduos $t_D(\hat{z}_i)$ e $t_Q(\hat{z}_i)$ podem ser aplicadas usando a distribuição normal padrão como distribuição de referência.

Na parte final deste trabalho, desenvolvemos medidas de diagnóstico nos modelos simétricos de regressão usando dois modelos bem conhecidos na literatura estatística, o modelo de deleção de casos (CDM) e o modelo de mudança de médias (MSOM). Através do modelo de deleção de casos (CDM) obtivemos medidas para identificar observações influentes tais como Distância de Cook Generalizada, Estatística W-K, aproximação a um passo e afastamento da verossimilhança. Para a

Distância de Cook Generalizada consideramos as métricas baseadas nas matrizes de informação observada e esperada de Fisher. Quando a métrica considerada é a matriz de informação esperada, a Distância de Cook Generalizada pode ser decomposta em duas partes, uma devida aos parâmetros de locação e outra devida ao parâmetro de escala. Observamos, que nas aplicações consideradas neste trabalho, as medidas de influência baseadas na aproximação a um passo apresentaram bom desempenho. Foram apresentados também, alguns métodos gráficos para avaliar a influência das observações na significância estatística dos parâmetros do modelo. Através do modelo de mudança de médias (MSOM) construímos um teste para identificar outliers. Utilizando simulação de Monte Carlo, este teste foi estudado e comparado com os testes para identificar outliers baseados nos resíduos propostos no Capítulo 3, encontrando que o desempenho destes testes é muito similar. Além disso, mostramos que para grandes tamanhos da amostra, os estimadores de máxima verossimilhança dos parâmetros nos modelos CDM e MSOM são equivalentes.

Em resumo, neste trabalho propusemos algumas ferramentas para a validação de modelos simétricos de regressão tais como

- Resíduos componente desvio e quantal, bem como uma padronização para os mesmos.
- Medidas para identificar observações influentes nas estimativas dos parâmetros de locação e escala.
- Testes para identificar outliers.
- Alguns métodos gráficos para identificar observações influentes na significância estatística dos parâmetros do modelo.

Além disso, este trabalho contribui em alguma medida à tarefa de difundir a modelagem estatística com erros simétricos. Como trabalhos futuros poderíamos desenvolver resíduos e medidas de diagnóstico em modelos com erros assimétricos, como por exemplo, skew-normal.

APÊNDICE A

Lemas e Resultados

Lema 1. *Seja Z uma variável aleatória com distribuição simétrica em torno de zero, e $h(\cdot)$ uma função contínua ímpar. Então, a variável aleatória $h(Z)$ tem distribuição simétrica em torno de zero.*

Prova: A variável aleatória Z tem distribuição simétrica em torno de zero se, e somente se, $P[Z \leq z] = P[-Z \leq z]$ para todo z; ou seja, Z tem distribuição simétrica se, e somente se, Z e $-Z$ são identicamente distribuídas. Portanto, $h(Z)$ e $h(-Z)$ também são identicamente distribuídas, então

$$\begin{aligned} P[h(Z) \leq u] &= P[h(-Z) \leq u], \qquad \forall z \in I\!R \\ &= P[-h(Z) \leq z] \quad \text{(pois } h(\cdot) \text{ é ímpar).} \end{aligned}$$

Então $P[h(Z) \leq z] = P[-h(Z) \leq z]$ para todo z, o que implica que as variáveis aleatórias $h(Z)$ e $-h(Z)$ são identicamente distribuídas. Logo podemos concluir que $h(Z)$ tem distribuição simétrica em torno de zero. (veja James, 1981) □

Lema 2. *Sejam $g_1(\cdot)$, $g_2(\cdot)$, $g_3(\cdot)$ e $g_4(\cdot)$ funções contínuas tais que: $g_1(\cdot)$, $g_3(\cdot)$ são funções ímpares, e $g_2(\cdot)$, $g_4(\cdot)$ são funções pares. Então temos que:*

i) $g_1(\cdot)g_2(\cdot)$ *é uma função ímpar,*

ii) $g_1(\cdot)g_3(\cdot)$ *é uma função par,*

iii) $g_2(\cdot)g_4(\cdot)$ *é uma função par.*

Prova: Para $g_1(\cdot)$ e $g_3(\cdot)$ funções ímpares temos que $g_1(-x) = -g_1(x)$ e $g_3(-x) = -g_3(x)$ para todo x. Da mesma maneira, para $g_2(\cdot)$ e $g_4(\cdot)$ temos que $g_2(-x) = g_2(x)$ e $g_4(-x) = g_4(x)$ para todo x. Assim, para o item i) observamos que:

$$g_1(-x)g_2(-x) = [-g_1(x)]g_2(x) = -g_1(x)g_2(x) \quad \forall x,$$

então $g_1(x)g_2(x)$ é ímpar. Para o item ii) temos que:

$$g_1(-x)g_3(-x) = [-g_1(x)][-g_3(x)] = g_1(x)g_2(x) \quad \forall x,$$

então $g_1(x)g_3(x)$ é par. Analogamente para o item iii):

$$g_2(-x)g_4(-x) = [g_2(x)][g_4(x)] = g_2(x)g_4(x), \quad \forall x,$$

então $g_2(x)g_4(x)$ é par. □

Lema 3. *Seja* $\hat{z}_i = (y_i - \mu_i(\hat{\boldsymbol{\beta}}))/\hat{\phi}^{1/2}$, *em que* $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$ *é o estimador de máxima verossimilhança do vetor de parâmetros* $\boldsymbol{\theta}$ *no modelo (2.2). Então,* $t(\hat{z}_i) \xrightarrow{p} t(z_i)$ *para toda função* $t(\cdot)$ *contínua.*

Prova: Como os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ são consistentes, temos que $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ e $\hat{\phi} \xrightarrow{p} \phi$. Além disso, a função $\mu_i(\cdot)$ é contínua em $\boldsymbol{\beta}$ (verdadeiro valor do parâmetro). Desta maneira, podemos mostrar que

i) $\mu_i(\hat{\boldsymbol{\beta}}) \xrightarrow{p} \mu_i(\boldsymbol{\beta})$,

ii) $\sqrt{\hat{\phi}} \xrightarrow{p} \sqrt{\phi}$.

A variável aleatória $\hat{z}_i$ pode ser reescrita como

$$\hat{z}_i = \frac{\sqrt{\phi}}{\sqrt{\hat{\phi}}}\left(z_i + \frac{\mu_i(\boldsymbol{\beta}) - \mu_i(\hat{\boldsymbol{\beta}})}{\sqrt{\phi}}\right). \tag{A.1}$$

Assim, usando os resultados em i) e ii) acima, obtemos que $\hat{z}_i = z_i + o_p(1)$. Então, como a função $t(\cdot)$ é contínua em todo ponto, de Serfling (1980, pág. 24) temos que $t(\hat{z}_i)$ converge em probabilidade para $t(z_i)$. □

Lema 4. *Para o resíduo* $t_D(z_i)$ *definido no Capítulo 2, as funções de distribuição* $F_{t_D}(\cdot)$ *e de densidade* $f_{t_D}(\cdot)$, *são dadas por*

$$F_{t_D}(z) = \begin{cases} 1 - F(z_g(z)) & se\ z < 0. \\ F(z_g(z)) & se\ z \geq 0, \end{cases}$$

e

$$f_{t_D}(z) = \frac{g(0)\exp(-z^2/2)\,|z|}{z_g(z)\,v(z_g(z))}, \quad z \neq 0,$$

em que $F(\cdot)$ *é a função de distribuição acumulada de* $z_i = (y_i - \mu_i)/\phi^{1/2}$ *e* $z_g(z) > 0$ *é uma constante tal que* $g(z_g^2(z)) = g(0)\exp(-z^2/2)$, *sendo* z_i *uma variável aleatória com distribuição* $S(0,1)$ *e função geradora de dendidades* $g(\cdot)$.

Prova: Do Lema 1 temos que a distribuição de $t_D(z_i)$ é simétrica em torno de zero, portanto temos

$$\begin{aligned} P[-z \leq t_D(z_i) \leq z] &= F_{t_D}(z) - Ft_D(-z) \\ &= 2F_{t_D}(z) - 1, \quad \forall\, z > 0. \end{aligned}$$

Mas, $P[-z \leq t_D(z_i) \leq z] = P[\,t_D^2(z_i) \leq z^2\,] = F_{t_D^2}(z^2)$, com $F_{t_D^2}(\cdot)$ a função de distribuição acumulada de $t_D^2(z_i)$. Em conseqüência temos que

$$F_{t_D}(z) = \begin{cases} \dfrac{1 + F_{t_D^2}(z^2)}{2}, & \text{se } z \geq 0, \\ \dfrac{1 - F_{t_D^2}(z^2)}{2}, & \text{se } z < 0. \end{cases} \tag{A.2}$$

Logo,

$$\begin{aligned} F_{t_D^2}(z^2) &= P[\,t_D^2(z_i) \leq z^2\,] \\ &= P\Big[\,2\log\Big[\frac{g(0)}{g(z_i^2)}\Big] \leq z^2\,\Big] \\ &= P\Big[\,g(z^2) \geq g(0)\exp(-z^2/2)\,\Big]. \end{aligned}$$

Seja $z_g(z) > 0$ uma constante tal que $g(z_g^2(z)) = g(0)\exp(-z^2/2)$, então, como $g(z^2)$ é uma função decrescente de z^2 temos que:

$$\begin{aligned} F_{t_D^2}(z^2) &= P\Big[g(z^2) \geq g(0)\exp(-z^2/2)\Big] \\ &= P[\,|z_i| \leq z_g(z)\,] \\ &= F(z_g(z)) - F(-z_g(z)) \\ &= 2F(z_g(z)) - 1. \end{aligned} \tag{A.3}$$

Susbtituindo (A.3) em (A.2) obtemos

$$F_{t_D}(z) = \begin{cases} F(z_g(z)), & \text{se } z \geq 0, \\ 1 - F(z_g(z)), & \text{se } z < 0. \end{cases} \tag{A.4}$$

A função de densidade de $t_D(z_i)$ pode ser escrita como

$$f_{t_D}(z) = g(z_g^2(z)) \left| \frac{\partial z_g(z)}{\partial z} \right|, \quad z \neq 0. \tag{A.5}$$

Derivando ambos os lados da equação $g(z_g^2(z)) = g(0)\exp(-z^2/2)$, temos que

$$\begin{aligned} g'(z_g^2(z))2z_g(z)\frac{\partial z_g(z)}{\partial z} &= -g(0)\exp(-z^2/2)\, z, \quad \text{o que implica que,} \\ \frac{\partial z_g(z)}{\partial z} &= -\frac{g(0)\exp\{-z^2/2\}z}{2g'(z_g^2(z))z_g(z)} \\ &= -\frac{g(z_g^2(z))z}{2g'(z_g^2(z))z_g(z)} \\ &= \frac{z}{v(z_g(z))z_g(z)}. \end{aligned} \tag{A.6}$$

Substituindo (A.6) em (A.5) obtemos

$$\begin{aligned} f_{t_D}(z) &= \frac{g(z_g^2(z))|z|}{v(z_g(z))z_g(z)} \\ &= \frac{g(0)\exp(-z^2/2)|z|}{v(z_g(z))z_g(z)}, \quad z \neq 0. \end{aligned}$$

□

Lema 5. *A distribuição de $t_D(z_i)$ quando z_i segue distribuição t-Student generalizada de parâmetros r e s, não depende do parâmetro s.*

Prova: Do Lema 4 temos que a função de densidade de $t_D(z_i)$ pode ser escrita na seguinte forma

$$f_{t_D}(z) = \frac{g(0)\exp(-z^2/2)|z|}{v(z_g(z))z_g(z)}, \quad z \neq 0.$$

Quando z_i segue distribuição t-Student generalizada com parâmetros r e s temos o seguinte

$$z_g^2(z) = s\Big\{ \exp\Big[z^2/(r+1)\Big] - 1\Big\}$$

e

$$v(z_g(z)) = \frac{r+1}{s+z^2}.$$

Logo, a função de densidade $f_{t_D}(z)$ pode ser expressa como

$$\begin{aligned} f_{t_D}(z) =& \frac{\exp(-z^2/2)|z|s^{-1/2}\exp(z^2/(r+1))s}{(r+1)s^{1/2}\big\{\exp[z^2/(r+1)]-1\big\}^{1/2}\mathbf{B}(1/2,r/2)} \\ =& \frac{\exp(-z^2/2)|z|\exp(z^2/(r+1))}{(r+1)\big\{\exp[z^2/(r+1)]-1\big\}^{1/2}\mathbf{B}(1/2,r/2)}, \end{aligned}$$

em que $\mathbf{B}(\cdot,\cdot)$ é a função Beta. Assim, podemos concluir que a distribuição de $t_D(z_i)$ não depende do parâmetro s. □

Lema 6. *Seja* $\tau = E\big(2t(z)t'(z)v(z)z - t'^2(z) - t''(z)t(z)\big)$ *o fator de correção para a variância do resíduo* $t(\hat{z}_i)$ *definido no Capítulo 3. Então*

$$\tau = \begin{cases} 4d_g, & se \quad t(z) = t_D(z), \\ 1, & se \quad t(z) = z. \end{cases}$$

Prova: Inicialmente consideramos o caso em que $t(z) = t_D(z)$. Derivando ambos os lados da equação $t_D^2(z) = 2\log\Big[\dfrac{g(0)}{g(z^2)}\Big]$, obtemos

$$2t_D(z)t'_D(z) = \frac{\partial[t_D^2(z)]}{\partial z} = -2\frac{\partial[\log(g(z^2))]}{\partial z} = 2v(z)z.$$

Então $E\big(2t(z)t'(z)v(z)z\big) = 2E\big(v^2(z)z^2\big) = 8d_g$. Derivando de novo ambos os lados da equação acima temos

$$2t_D(z)t''_D(z) + 2t'^2_D(z) = \frac{\partial^2[t_D^2(z)]}{\partial z^2} = -2\frac{\partial^2[\log(g(z^2))]}{\partial z^2}.$$

Daí, $E\big(t_D(z)t''_D(z) + t'^2_D(z)\big) = -E\Big(\dfrac{\partial^2[\log(g(z^2))]}{\partial z^2}\Big) = 4d_g$. Assim, podemos concluir que para o resíduo $t_D(\hat{z}_i)$ o valor do fator de correção τ é dado por $4d_g$. Por outro lado, para $t(z) = z$ temos que

$$E\big(2t(z)t'(z)v(z)z - t'^2(z) - t''(z)t(z)\big) = 2E\big(z^2v(z)\big) - 1,$$

pois $t'(z) = 1$ e $t''(z) = 0$. De Fang, Kotz e Ng (1990) sabemos que $E(z^2 v(z)) = 1$, portanto, para o resíduo $\hat{z}_i$ o valor do fator de correção τ é igual a 1. □

APÊNDICE B

Viés de segunda ordem

Cordeiro, Ferrari, Uribe-Opazo e Vasconcellos (2000) obtiram os viéses de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros nos modelos simétricos de regressão. No caso do vetor de parâmetros $\boldsymbol{\beta}$, o viés pode ser expresso na seguinte forma

$$\mathbf{B}(\hat{\boldsymbol{\beta}}) = (\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_\beta^T \boldsymbol{\eta},$$

em que $\boldsymbol{\eta}$ é um vetor coluna de dimensão n cujo i-ésimo elemento pode ser escrito como $\eta_i = -\dfrac{\phi}{8d_g}\{(\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_{\beta\beta}(i)\}$, em que $\mathbf{D}_{\beta\beta}(i) = \partial^2 \mu_i / \partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$. Para modelos com componentes sistemáticas lineares $\mathbf{D}_{\beta\beta}(i) = 0$ para todo i, o que implica que nestes modelos $\mathbf{B}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$. Por outro lado, o viés do estimador de máxima verossimilhança de ϕ pode ser reescrito como $\mathbf{B}(\hat{\phi}) = a_g \dfrac{\phi}{n}$, em que a_g é uma constante que pode ser expressa na forma abaixo

$$a_g = \frac{1}{\alpha_{2,2} - 1}\left\{ p\left(\frac{\alpha_{3,1}}{\alpha_{2,0}} + 2\right) + \frac{\alpha_{3,3} + \alpha_{2,2} + 1}{(\alpha_{2,2} - 1)} \right\},$$

sendo $\alpha_{r,s} = E\left(\dfrac{\partial^{(r)}[\log(g(z^2))]}{\partial z^{(r)}} z^s\right)$, com z uma variável aleatória com distribuição $S(0,1)$ (veja Cordeiro et *al*, 2000). No caso da distribuição normal a_g é igual à dimensão do vetor de parâmetros $\boldsymbol{\beta}$.

APÊNDICE C

Coelhos Europeus

Quadro C.1 *Pesos das lentes dos olhos de coelhos europeus (y), em miligramas, e idade (x) em dias numa amostra de 71 observações.*

x	y	x	y
15	21,66	218	174,18
15	22,75	218	173,03
15	22,30	219	173,54
18	31,25	224	178,86
28	44,79	225	177,68
29	40,55	227	173,73
37	50,25	232	159,98
37	46,88	232	161,29
44	52,03	237	187,07
50	63,47	246	176,13
50	61,13	258	183,40
60	81,00	276	186,26
61	73,09	285	189,66
64	79,09	300	186,09
65	79,51	301	186,70
65	65,31	305	186,80
72	71,90	312	195,10
75	86,10	317	216,41
75	94,60	338	203,23
82	92,50	347	188,38
85	105,00	354	189,70
91	101,70	357	195,31
91	102,90	375	202,63
97	110,00	394	224,82
98	104,30	513	203,30
125	134,90	535	209,70
142	130,68	554	233,90
142	140,58	591	234,70
147	155,30	648	244,30
147	152,20	660	231,00
150	144,50	705	242,40
159	142,15	723	230,77
165	139,81	756	242,57
183	153,22	768	232,12
192	145,72	860	246,70
195	161,10		

APÊNDICE D

Programa Coelhos Europeus

Neste Apêndice apresentamos o código na linguagem R utilizado para o cálculo das medidas de diagnóstico no modelo de resposta t-Student(4) ajustado aos dados do exemplo dos Coelhos Europeus. Vale salientar, que este programa usa a rotina `elliptical` (veja Cysneiros, Paula e Galea, 2005).

```
# Função que calcula o vetor gradiente e a matriz de segundas derivadas para o
# modelo de interesse
DerBrabbits <- function(parm,X)
{
         beta_1 <- parm[1]
         beta_2 <- parm[2]
         beta_3 <- parm[3]
         x0 <- X[,1]
         x <- X[,2]
        .grad <- array(0, c(length(x0), 3), list(NULL, c("beta_1","beta_2","beta_3")))
        .grad[, "beta_1"] <- x0
        .grad[, "beta_2"] <- -1/(x+nu)
        .grad[, "beta_3"] <- beta_2/((x+beta_3)^2)
        .value <- beta_1 -beta_2/(x+beta_3)
        .hess <- array(0, c(3, 3,length(x0)),
        list( c("beta_1","beta_2","beta_3"), c("beta_1","beta_2","beta_3"),NULL))
.hess["beta_1","beta_1", ] <- rep(0,length(x0))
.hess["beta_1","beta_2", ] <- rep(0,length(x0))
.hess["beta_1","beta_3", ] <- rep(0,length(x0))
.hess["beta_2","beta_1", ] <- .hess["beta_1","beta_2", ]
.hess["beta_2","beta_2", ] <- rep(0,length(x0))
.hess["beta_2","beta_3", ] <- 1/((x+beta_3)^2)
.hess["beta_3","beta_1", ] <- .hess["beta_1","beta_3", ]
.hess["beta_3","beta_2", ] <- .hess["beta_2","beta_3", ]
.hess["beta_3","beta_3", ] <- -2*beta_2/((x+beta_3)^3)
fit <- list(value=.value,gradient=.grad, hessian=.hess)
fit
}

# Função que calcula o vetor gradiente e a matriz de segundas derivadas para o
# modelo de mudança de médias
DerBrabbits3 <- function(parm,X)
{
         beta_1 <- parm[1]
         beta_2 <- parm[2]
```

```
         beta_3 <- parm[3]
         gamma  <- parm[4]
         x0   <- X[,1]
         x    <- X[,2]
         b_i  <- X[,3]
        .grad <- array(0, c(length(x0), 4), list(NULL, c("beta_1","beta_2","beta_3","gamma")))
        .grad[, "beta_1"]  <- x0
        .grad[, "beta_2"]  <- -1/(x+beta_3)
        .grad[, "beta_3"]  <- beta/((x+beta_3)^2)
        .grad[, "gamma"]   <- b_i
        .value <- beta_1 -beta_2/(x+beta_3) + gamma*b_i
        .hess <- array(0, c(4, 4,length(x0)),
        list( c("beta_1","beta_2","beta_3","gamma"), c("beta_1","beta_2","beta_3","gamma"),NULL))
.hess["beta_1","beta_1", ] <- rep(0,length(x0))
.hess["beta_1","beta_2", ] <- rep(0,length(x0))
.hess["beta_1","beta_3", ] <- rep(0,length(x0))
.hess["beta_1","gamma",  ] <- rep(0,length(x0))
.hess["beta_2","beta_1", ] <- .hess["beta_1","beta_2", ]
.hess["beta_2","beta_2", ] <- rep(0,length(x0))
.hess["beta_2","beta_3", ] <- 1/((x+beta_3)^2)
.hess["beta_2","gamma",  ] <- rep(0,length(x0))
.hess["beta_3","beta_1", ] <- .hess["beta_1","beta_3", ]
.hess["beta_3","beta_2", ] <- .hess["beta_2","beta_3", ]
.hess["beta_3","beta_3", ] <- -2*beta_2/((x+beta_3)^3)
.hess["beta_3","gamma",  ] <- rep(0,length(x0))
.hess["gamma","beta_1",  ] <- rep(0,length(x0))
.hess["gamma","beta_2",  ] <- rep(0,length(x0))
.hess["gamma","beta_3",  ] <- rep(0,length(x0))
.hess["gamma","gamma",   ] <- rep(0,length(x0))
fit <- list(value=.value,gradient=.grad, hessian=.hess)
fit
}

# Leitura dos dados
rabbits.dat <-  scan("C:\\coelhos.dat" , what=list(idade=0,peso=0))
attach(rabbits.dat)
y  <- log(rabbits.dat$peso)
x  <- rabbits.dat$idade
X  <- cbind(1,x)
rabbits <- data.frame(y,x)

T4  <- elliptical( formula = y~x, linear = F, DerB = DerBrabbits, parmB = c(5,127,35),
family = Student(4), data=rabbits )
TT  <- "T-Student(4)"

# Sumario do modelo
summary(T4)

# Plotando os pesos
vp_T <- function(x, gl){
s <- (gl + 3)/(gl + x^2)
s}
fitt_T4 <- fitted(T4)
phi_T4  <- T4$dispersion
rest_T4 <- (y-fitt_T4)/sqrt(phi_T4)
plot(vp_T(rest_T4,4), main="TT" ,xlab="Indice", ylab="pesos", cex=0.3, lwd=3)
identify(vp_T(rest_T4,4), n=10)

# Resíduos
p  <- length(T4$coeff)
D_T4 <-T4$Xmodel
H_T4 <- D_T4%*%solve(t(D_T4)%*%D_T4)%*%t(D_T4)
sdiag_T4 <- diag(H_T4)
phi_T4 <- T4$dispersion
gl <- 4
```

```
g_0_T4 <- gamma((gl+1)/2)*gl^(gl/2)*((gl)^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))
g_u_T4 <- gamma((gl+1)/2)*gl^(gl/2)*((gl + rest_T4^(2))^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))

# Resíduo quantal padronizado
tQ_T4 <- qnorm(pt(rest_T4, 4))/sqrt(1 - 0.7113*sdiag_T4)

# Resíduo componente de desvio padronizado
tD_T4 <- sqrt(2)*(1-2*(y<fitt_T4))*sqrt(log(g_0_T4/g_u_T4))/sqrt(1.4018-sdiag_T4)

# Resíduo Cox e Snell padronizado
tr_i <- resid(T4, type="stand")

plot(fitt_T4, tD_T4, main="TT",xlab="Valores Ajustados", ylab="tD", ylim=c(-3.25,3.25),
cex=0.3, lwd=3)
abline(-2,0,lty=3)
abline(2,0,lty=3)
identify(fitt_T4, tD_T4, n=10)

# Médida da qualidade do ajuste global baseada no resíduo tD_T4
MQ <- sum(2*log(g_0_T4/g_u_T4))/(length(y)*1.4018-p)

# Calculando a estatística da razão de verossimilhanças para o teste
# para identificar outliers
E_L <- matrix(0,71,1)
for(i in 1:71){
b_i <- matrix(0,71,1)
b_i[i] <- 1
x2 <- cbind(x,b_i)
X <- cbind(1,x2)
rabbits <- data.frame(y,x)
haber <- elliptical(formula=y~x2,linear=F, DerB=DerBrabbits3,parmB=c(5,130,37,0),
family=Student(4),data=rabbits)
E_L[i] <- -2*(T4$loglik-haber$loglik)
}

plot(tD_T4^2, main="",xlab="", ylab="", ylim=c(0,11), cex=0.8, pch=24)
lines(tD_T4^2, lty=3)
par(new=T)
plot(E_L, main="TT",xlab="Indice", ylab="tD", ylim=c(0,11), cex=0.8, pch=19)
lines(E_L, lty=1, cex=0.05)
abline(4,0,lty=3)

# Gráfico Normal de probabilidades com Envelope
k <- 500
alfa <- 0.05
re <- matrix(0, n, k)
alfa1 <- ceiling(k*alfa)
alfa2 <- ceiling(k*(1-alfa))
n <- nrow(D_T4)
for(i in 1:k){
nresp <- fitt_T4 + rt(n,4)*sqrt(phi_T4)
rabbits2 <- data.frame(nresp,x)
T4_1 <- elliptical(formula=nresp~x,linear=F, DerB=DerBrabbits,parmB=c(5,130,37),
            family=Student(4),data=rabbits2)

fitt_T4_1 <-fitted(T4_1)
D_T4_1 <- T4_1$Xmodel
H_T4_1 <- D_T4_1%*%solve(t(D_T4_1)%*%D_T4_1)%*%t(D_T4_1)
sdiag_T4_1 <- diag(H_T4_1)
phi_T4_1 <- T4_1$dispersion
rest_T4_1 <- (nresp-fitt_T4_1)/sqrt(phi_T4_1)
gl <- 4
g_u_T4_1 <- gamma((gl+1)/2)*gl^(gl/2)*((gl + rest_T4_1^(2))^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))
sign <- (1-2*(nresp<fitt_T4_1))
re[,i] <- sort(sqrt(2)*sign*sqrt(log(g_0_T4/g_u_T4_1))/sqrt(1.4018-sdiag_T4_1))}
```

```
e1<-numeric(n)
e2<-numeric(n)
for(i in 1:n){
eo<-sort(re[i,])
e1[i]<-eo[alfa1]
e2[i]<-eo[alfa2]}
xb<-apply(re,1,mean)
faixa<-range(tD_T4,e1,e2)
par(pty="s")
qqnorm(e1,axes=F,xlab="",type="l",ylab="",ylim=faixa,lty=1,main="")
par(new=T)
qqnorm(e2,axes=F,xlab="",type="l",ylab="",ylim=faixa,lty=1,main="")
par(new=T)
qqnorm(xb,axes=F,xlab="",type="l",ylab="",ylim=faixa,lty=3,main="")
par(new=T)
qqnorm(tD_T4,xlab="PN",ylab="tD", main="TT", ylim=faixa, cex=0.3, lwd=3)

# Distância de Cook
DGB_I <- T4$scale*(vp_T(rest_T4,4)*rest_T4)^2*sdiag_T4/((1-sdiag_T4)^2)
DGP_I <- 71*(T4$v*rest_T4^2 - 1)^2/(70*70*(T4$scaledispersion))

plot(DGB_I, main="TT",xlab="Indice", ylim=c(0,1.5), ylab="DC", cex=0.3, lwd=3)
abline(2*mean(DGB_I),0,lty=3)
identify(DGB_I, n=10)

plot(DGP_I, main="TT",xlab="Indice", ylim=c(0,0.51), ylab="Df", cex=0.3, lwd=3)
abline(2*mean(DGP_I),0,lty=3)
identify(DGP_I, n=10)

# Avaliando o desempenho das medidas baseadas na aproximação
# a um passo
valor <- function(beta){
beta[1] -beta[2]/(x+beta[3])
}
LD_I <- matrix(0,71,1)
LD <- matrix(0,71,1)
DGB <- matrix(0,71,1)
DGP <- matrix(0,71,1)
gl <- 4
beta_hat <- T4$coeff
for(i in 1:71){
nano2 <- elliptical(formula=y~x,linear=F, DerB=DerBrabbits, parmB=c(5,130,37),
family=Student(4),data=rabbits, subset=-c(i))
zeta <- (y[i]-fitt_T4[i])/sqrt(phi_T4)
mar <- solve(t(D_T4)%*%D_T4)%*%D_T4[i,]
beta_I <- beta_hat - sqrt(phi_T4)*vp_T(zeta,gl)*zeta*mar/(1-sdiag_T4[i])
phi_I <- phi_T4 - 2*phi_T4*(T4$v[i]*zeta^2-1)/(70*(T4$scaledispersion))
rest_T4_I <- (y - valor(beta_I))/sqrt(phi_I)
rest_T4_i <- (y - valor(nano2$coeff))/sqrt(nano2$dispersion)
g_u_T4_I <- gamma((gl+1)/2)*gl^(gl/2)*((gl + rest_T4_I^(2))^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))
g_u_T4_i <- gamma((gl+1)/2)*gl^(gl/2)*((gl + rest_T4_i^(2))^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))
LD_I[i] <- 2*(T4$loglik + (71/2)*log(phi_I) - sum(log(g_u_T4_I)))
rest_T4_I <- (y - D_T4%*%nano2$coeff)/sqrt(nano2$dispersion)
g_u_T4 <- gamma((gl+1)/2)*gl^(gl/2)*((gl + rest_T4_I^(2))^(-(gl+1)/2))/(gamma(1/2)*gamma(gl/2))
LD[i] <- 2*(T4$loglik + (71/2)*log(nano2$dispersion) - sum(log(g_u_T4_i)))
DGB[i] <- (T4$scale/phi_T4)*t(beta_hat-nano2$coeff)%*%(t(D_T4)%*%D_T4)%*%(beta_hat-nano2$coeff)
DGP[i] <- ((phi_T4-nano2$dispersion)^2)*(71*T4$scaledispersion)/(4*phi_T4^2)
}

plot(DGB_I, main="",xlab="", ylab="", ylim=c(0,0.6), cex=0.8, pch=24)
lines(DGB_I, lty=3)
par(new=T)
plot(DGB, main="DGB",xlab="Indice", ylab="", ylim=c(0,0.6), cex=0.8, pch=19)
lines(DGB, lty=1, cex=0.05)
```

```
plot(DGP_I, main="",xlab="", ylab="", ylim=c(0,0.15), cex=0.8, pch=24)
lines(DGP_I, lty=3)
par(new=T)
plot(DGP, main="DGP",xlab="Indice", ylab="", ylim=c(0,0.15), cex=0.8, pch=19)
lines(DGP, lty=1, cex=0.05)

plot(LD_I, main="",xlab="", ylab="", ylim=c(0,0.6), cex=0.8, pch=24)
lines(LD_I, lty=3)
par(new=T)
plot(LD, main="LD",xlab="Indice", ylab="", ylim=c(0,0.6), cex=0.8, pch=19)
lines(LD, lty=1, cex=0.05)

# Mudanças nas estimativas excluindo as observações (4,5,16,17)
nano <- elliptical(formula=y~x,linear=F, DerB=DerBrabbits,parmB=c(5,130,37),
        family=Student(4),data=rabbits, subset=-c(4,5,16,17))

100*(nano$coeff-beta_hat)/beta_hat
100*(nano$dispersion-T4$dispersion)/T4$dispersion

# Gráfico do parâmetro excluido (beta_3)
D_e <- cbind(T4$Xmodel[,1],T4$Xmodel[,2])
zeta <- (T4$v/T4$scale)*(y-fitted(T4)) + D_T4%*%T4$coef
gamma <- T4$Xmodel[,3]
H_e <- D_e%*%solve(t(D_e)%*%D_e)%*%t(D_e)
RG <- gamma - H_e%*%gamma
RZ <- zeta - H_e%*%zeta
plot(RG, RZ, main="TT", ylim=c(-0.35,0.4), cex=0.3, lwd=3)
abline(lm(RZ~-1 +RG), lty=1)
abline(h=0, lty=3)
abline(v=0, lty=3)
identify(RG,RZ, n=10)
```

APÊNDICE E

Concentração de Cloro

Quadro E.1 *Concentração de cloro disponível num produto, (y) e tempo desde que o mesmo foi produzido, (x) em semanas numa amostra de 44 observações.*

x	y	x	y
8	0,49	22	0,41
8	0,49	22	0,41
10	0,48	24	0,42
10	0,47	24	0,40
10	0,48	24	0,40
10	0,47	26	0,41
12	0,46	26	0,40
12	0,46	26	0,41
12	0,45	28	0,41
12	0,43	28	0,40
14	0,45	30	0,40
14	0,43	30	0,40
14	0,43	30	0,38
16	0,44	32	0,41
16	0,43	32	0,40
16	0,43	34	0,40
18	0,46	36	0,41
18	0,45	36	0,38
20	0,42	38	0,40
20	0,42	38	0,40
20	0,43	40	0,39
22	0,41	42	0,39

Referências

Arellano-Valle, R.B. (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão.* Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil.

Atkinson, A.C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika*, **68**, 13-20.

Berkane, M. e Bentler, P.M. (1986). Moments of elliptical distributed random variates. *Statistics and Probability Letters*, **4**, 333-335.

Cambanis, S.; Huang, S. e Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368-385.

Christensen, R.; Pearson, L.M. e Johnson, W. (1992). Case deletion diagnostics for mixed models. *Technometrics*, **34**, 38-45.

Cook, R.D. (1986). Assesment of local influence (with discussion). *Journal of the Royal Statistical Society* B, **48**, 133-169.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15-18.

Cook, R.D. e Weisberg, S. (1982). *Residuals and influence in Regression.* New York: Chapman and Hall.

Cordeiro, G.M.; Ferrari, S.L.P.; Uribe-Opazo, M.A. e Vasconcellos, K.L.P. (2000). Corrected maximum likelihood estimation in a class of symmetric nonlinear regression models, *Statistics and Probability Letters*, **46**, 317-328.

Cordeiro, G.M. (2004). Corrected LR tests in symmetric nonlinear regression models. *Journal Statistical Computation & Simulation*, **74**, 609-620.

Cox, D.R. e Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cox, D.R. e Snell, E.J. (1968). A General Definition of Residuals (with discussion). *Journal of the Royal Statistical Society* B, **30**, 248-275.

Cox, D.R. e Snell, E.J. (1971). On Test Statistics Computed From Residuals. *Biometrika*, **58**, 589-594.

Cysneiros, F.J.A. (2004). *Métodos Restritos e Validação de Modelos Simétricos de Regressão*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil.

Cysneiros, F.J.A. e Paula, G.A. (2004). One-Sided Test in Linear Models with Multivariate t-distribution. *Communications in Statistics, Simulation and Computation*, **33(3)**, 747-772.

Cysneiros, F.J.A.; Paula, G.A. e Galea, M. (2005). *Modelos Simétricos Aplicados*. São Paulo: ABE - Associação Brasileira de Estatística.

Davison, A.C. e Gigli, A. (1989). Deviance residuals and normal score plots. *Biometrika*, **76**, 211-221.

Draper, N.R. e Smith H. (1998). *Applied Regression Analysis*. New York: Willey.

Dunn, K.P. e Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.

Fang, K.T. e Anderson, T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions*. New York: Allerton Press.

Fang, K.T.; Kotz, S. e Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman and Hall.

Fang, K.T. e Zhang, Y.T. (1990). *Generalized Multivariate Analysis.* New York: Springer-Verlag.

Ferrari, S.L.P. e Uribe-Opazo, M.A. (2001). Correted likelihood ratio tests in class of symmetric linear regression models. *Brazilian Journal of Probability and Statistical*, **15**, 49-67.

Fung, W.H.; Zhu, Z.Y.; Wei, B.C. e He X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society* B **64**, 565-579.

Galea, M.; Paula, G.A. e Cysneiros, F.J.A. (2005). On diagnostic in symmetrical nonlinear models. *Statistics and Probability Letters*, **73**, 459-467.

Galea, M.; Paula, G.A. e Uribe-Opazo, M. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers*, **44**, 23-45.

Galea, M.; Riquelme, M. e Paula G.A. (2000). Diagnostics methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistical*, **14**, 167-184.

Gigli, A. (1987). *A comparison between Cox e Snell residuals and deviance residuals.* Tese de mestrado, Imperial College, London.

Hoaglin, D.C. e Welsch, R.E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, **32**, 17-22.

Huber, P.Y. (1981). *Robust Statistics.* New York: Willey.

James, B. (1981). *Probabilidade: Um curso em nível Intermediário.* Rio de Janeiro: IMPA.

Jørgensen, B. (1984). The delta algorithm and GLIM. *International Statistical Review*, **52**, 283-300.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya A*, **32**, 419-430.

Kowalski, J.; Mendoza-Blanco, J.; Tu, X.M. e Gleser, L.J. (1999). On the difference in inference and prediction betwwen the joint and independent t-error models for seemingly unrelated regressions. *Communications in Statistics, Theory and Methods*, **28**, 2119-2140.

Landsman, Z. e Valdez, E.A. (2002). Tail Conditional Expectations for Elliptical Distributions. *North American Actuarial Journal*, **2**, 1-28.

McCullagh, P. e Nelder, J.A. (1989). *Generalized Linear Models, 2nd. Edition.* London: Chapman and Hall.

Pierce, D.A. e Shafer, D.W. (1986). Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, **81**, 977-986.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705-724.

Rao, B.L.S.P. (1990). Remarks on univariate symmetric distributions. *Statistics and Probability Letters*, **10**, 307-315.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modelling.* New York: Marcel Dekker.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Willey.

Smith, H. e Dubey, S.A. (1964). Some reliability problems in the chemical industry. *Industrial Quality Control*, **21**, 64-70.

Svetliza, C.F. e Paula, G.A. (2003). Diagnostics in nonlinear negative binomial models. *Communications in Statistics, Theory and Methods*, **32**, 1227-1250.

Sun, B.R. e Wei, B.C. (2004). On influence assessment for LAD regression. *Statistics and Probability Letters*, **67**, 97-110.

Tang, N.S.; Wei, B.C e Wang X.R. (2000). Influence diagnostics in nonlinear reproductive dispersion models. *Statistics and Probability Letters*, **46**, 59-68.

Wang, P.C. (1985). Adding a variable in generalized linear models. *Technometrics*, **27**, 273-276.

Wei, W.H.; Hu, Y.Q. e Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, **25**, 25-37.

Wei, W.H. e Fung, W.K. (1999). The mean-shift outlier model in general weighted regression and its applications . *Computational Statistical & Data Analysis*, **30**, 429-441.