Modelos Simétricos Aplicados

Francisco José A. Cysneiros
Departamento de Estatística
Universidade Federal de Pernambuco, Brasil
cysneiros@de.ufpe.br

Gilberto A. Paula
Instituto de Matemática e Estatística
Universidade de São Paulo, Brasil
giapaula@ime.usp.br

e

Manuel Galea
Departamento de Estadistica
Universidad de Valparaíso, Chile
Manuel.Galea@uv.cl

9ª Escola de Modelos de Regressão
Fevereiro-2005
Águas de São Pedro - SP

Dedicado a: Audrey e Rafael (*Francisco Cysneiros*)
Marlene, Natália e Alexandre (*Gilberto A. Paula*)
Patrícia, Rodrigo e Felipe (*Manuel Galea*)

Prefácio

A suposição de normalidade sempre foi muito atrativa para os erros de modelos de regressão com resposta contínua e, mesmo quando não era alcançada, procurava-se alguma transformação na resposta no sentido de obter-se pelo menos a simetria. Contudo, com o passar do tempo, verificou-se que as estimativas obtidas para os coeficientes dos modelos normais mostraram-se sensíveis a observações extremas, comumente chamadas de observações aberrantes, incentivando o desenvolvimento de metodologias robustas contra tais observações. Dentre essas metodologias, destacam-se os métodos robustos e modelos robustos. Estes últimos serão discutidos neste trabalho.

Na linha de modelos robustos, alternativas à suposição de erros normais têm sido propostas na literatura. Uma dessas alternativas é assumir para os erros distribuições com caudas mais pesadas do que a normal, a fim de tentar reduzir a influência de pontos aberrantes nas estimativas dos coeficientes. Neste contexto, podemos citar Lange, Little e Taylor (1989) que propõem o modelo $t-$Student com ν graus de liberdade. Na última década, diversos resultados de natureza teórica e aplicada surgiram como alternativas à modelagem com erros normais como, por exemplo, o uso de distribuições simétricas (ou elípticas). Grande parte desses resultados podem ser encontrados em Fang, Kotz e Ng (1990) e Fang e Anderson (1990).

Este trabalho teve início no IME-USP quando o primeiro autor estava desenvolvendo sua tese de doutorado. O objetivo geral do texto é reunir alguns resultados sobre modelagem de dados simétricos, focando em particular o desenvolvimento da análise inferencial e de diagnóstico na classe de modelos lineares e não-lineares com erros simétricos independentes. A classe simétrica reúne distribuições com caudas mais leves e mais pesadas do que a normal, tais como normal contaminada, $t-$Student, $t-$Student generalizada, logística-I, logística-II, logística generalizada, exponencial potência, dentre outras. O texto é dividido em quatro capítulos. No Capítulo 1 é apresentada uma coletânea de resultados teóricos sobre distribuições simétricas. O Capítulo 2 resume os principais resultados inferenciais em modelos de regressão linear e não-linear simétricos e discute também a aplicação de técnicas de diagnóstico nos modelos apresentados. Exemplos são ilustrados no Capítulo 3 e analisados pela `library elliptical` desenvolvida para o ajuste de modelos simétricos em S-Plus e R. No último capítulo são discutidas algumas extensões para a análise de dados correlacionados simétricos. Como é um texto ainda em desenvolvimento, ficamos desde já abertos a críticas e sugestões que podem ser enviadas para **cysneiros@de.ufpe.br**.

Universidade Federal de Pernambuco, Brasil — Francisco Cysneiros
Universidade de São Paulo, Brasil — Gilberto A. Paula
Universidad de Valparaíso, Chile — Manuel Galea

Fevereiro, 2005.

Conteúdo

Lista de Figuras

Lista de Tabelas

CAPÍTULO 1

Distribuições simétricas

1.1 Motivação

A distribuição normal tem sido largamente utilizada no estudo de variáveis aleatórias contínuas simétricas, havendo um grande desenvolvimento inferencial em diversas áreas da Estatística. Isto é particularmente o caso de análise multivariada e regressão linear.

Em outras áreas do conhecimento, tais como Finanças, a suposição de normalidade é também comumente adotada. Por exemplo, nos modelos de valorização de ativos de capital, `CAPM` ("Capital Asset Pricing Model") em que é estabelecida uma relação funcional entre a rentabilidade esperada de um título, o retorno livre de risco e um prêmio por risco. Esses modelos que foram desenvolvidos independentemente por Sharpe (1964), Lintner (1965) e Mossin (1966) assumem o seguinte:

$$\mathrm{E}(y_r) = r_f + \beta\{\mathrm{E}(r_m) - r_f\}, \tag{1.1}$$

em que y_r denota o retorno de um título, r_f a taxa de retorno livre de risco, β é um risco sistemático do ativo sob estudo e r_m é o retorno fornecido pelo mercado medido por algum índice, por exemplo, no caso do Brasil o `IBOVESPA`. O risco sistemático é uma medida importante de risco tanto para analistas financeiros como para administradores de carteiras. Este parâmetro tem grande importância para o cálculo do custo de capital dos fundos próprios, que é básico na avaliação de qualquer projeto ou mesmo na valorização de uma empresa (ver, por exemplo, Campbell, Lo e MacKinlay, 1997).

Para estimar o parâmetro β utiliza-se a regressão linear simples. Ou seja, para um conjunto de n rentabilidades de uma determinada ação do mercado e para um

ativo livre de risco, o seguinte modelo tem sido utilizado:

$$y_{rt} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \epsilon_t, \tag{1.2}$$

em que y_{rt} denota o retorno da ação durante o t-ésimo período, r_{mt} é o retorno do mercado no período t, r_{ft} indica a taxa livre de risco durante o t-ésimo período e ϵ_t são erros independentes de média zero (quando existe) e parâmero de escala ϕ. Se denotarmos por $y_t = y_{rt} - r_{ft}$ e $x_t = r_{mt} - r_{ft}$ teremos um modelo de regressão linear simples com parâmetros α e β.

Utiliza-se em geral o método de mínimos quadrados para estimar os parâmetros do modelo (1.2). Além disso, a inferência é feita assumindo que os erros ϵ_t seguem distribuição aproximadamente normal (ver, por exemplo, Elton e Gruber, 1995; Campbell, Lo e MacKinlay, 1997). Contudo, como é conhecido, este método é bastante sensível a rentabilidades atípicas muito comuns na prática, particularmente nos mercados latino-americanos. Estas observações aberrantes podem distorcer a estimativa de β. Por outro lado, existem evidências empíricas de que as rentabilidades das ações tenham distribuições com caudas mais pesadas do que a normal (ver, por exemplo, Fama, 1965; Blattberg e Gonedes, 1975; Zhou, 1993). Lange, Little e Taylor (1989) propõem o uso da distribuição t de Student como alternativa robusta à distribuição normal e apresentam aplicações em análise multivariada e de regressão.

Para ilustrar um exemplo, vamos considerar as rentabilidades mensais das ações da empresa Concha & Toro, denotadas por y_{rt}, uma companhia do setor vinícola do mercado chileno. Como índice da rentabilidade do mercado será utilizado o Índice de Preços Seletivos de Ações (`IPSA`), r_{mt}, e como taxa livre de risco utilizaremos a taxa de juros em venda, base mensal, do banco central chileno, r_{ft}. Os dados correspondem ao período compreendido entre janeiro de 1990 a junho de 2004 e são apresentados no Apêndice. Na Figura 1.1 tem-se o diagrama de dispersão de y_t contra x_t e as retas ajustadas supondo erros normais, t de Student com 5 graus de liberdade e exponencial potência com parâmetro de forma $k = 0,9$. Os graus de liberdade da distribuição t de Student e o parâmetro de forma da

distribuição exponencial potência foram obtidos através do procedimento de seleção de Akaike. Para a t de Student e exponencial potência os coeficientes de curtose são, respectivamente, $\gamma_2 = 9$ e $\gamma_2 = 5,6$, enquanto que para a normal tem-se $\gamma_2 = 3$. Nota-se pelo gráfico de dispersão alguns retornos atípicos que são menos influentes nas estimativas dos parâmetros sob erros t de Student e exponencial potência.

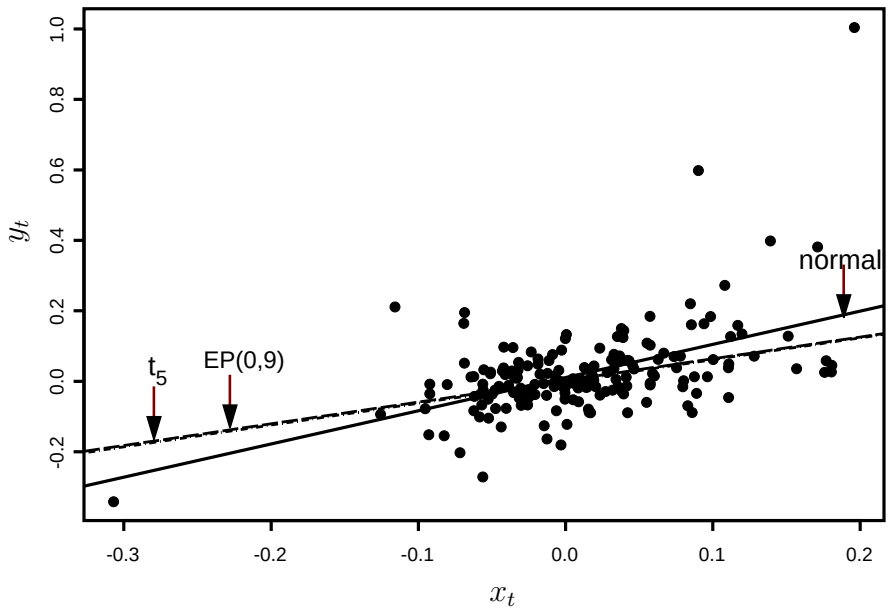

Figura 1.1 *Retas ajustadas aos dados sobre retornos das ações da empresa Concha & Toro no período de 1990 a 2004.*

Neste exemplo, o interesse principal é na estimativa do parâmetro β que representa o risco sistemático. Sob a suposição de normalidade dos erros temos que a estimativa de máxima verossimilhança de β é 0,942 com desvio padrão aproximado de 0,122 enquanto que sob a suposição de erros $t-$Student com 5 graus de liberdade e exponencial potência com k=0,9 obtemos as estimativas 0,618 e 0,688 com desvios padrão aproximados de 0,085 e 0,078, respectivamente. Nas Figuras 1.2a, 1.2b e 1.2c, os gráficos de influência local total $C_i(\beta)$ sob perturbações de casos

mostram que a influência das observações em $\hat{\beta}$ é bem menor para o modelo baseado na suposição de erros $t-$Student com 5 graus de liberdade do que no modelo baseado na suposição de normalidade e erros exponencial potência com $k = 0,9$.

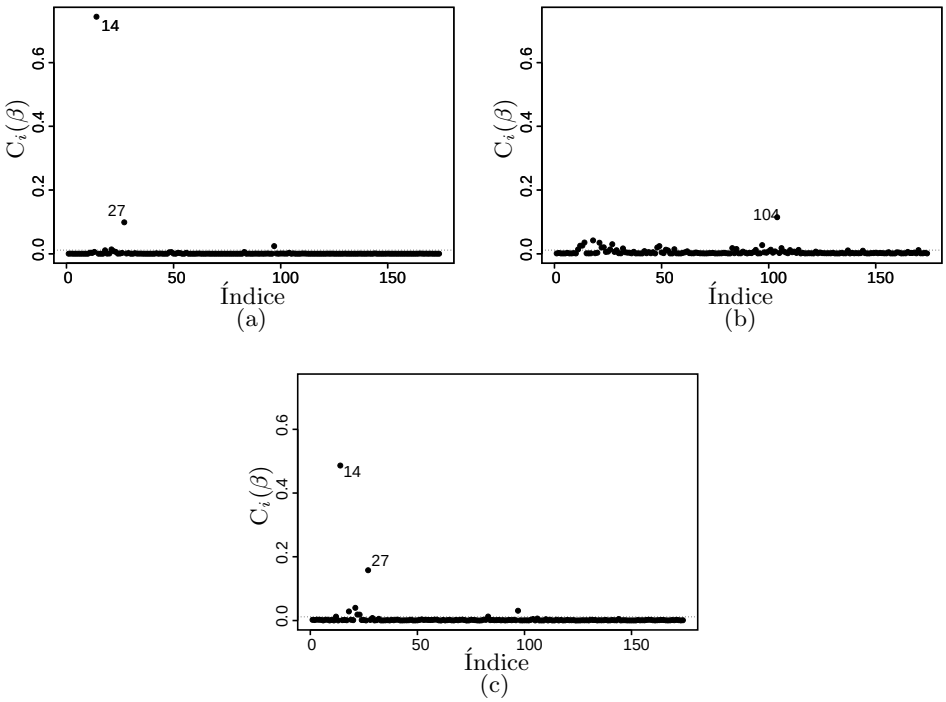

Figura 1.2 *Gráficos de influência local total* $\mathrm{C}_i(\beta)$ *sob perturbações de casos para o modelo (*1.2*) sob erros normais (a), t−Student com 5 g.l. (b) e exponencial potência com k=0,9 (c).*

A Tabela 1.1 mostra as variações percentuais nas estimativas de máxima verossimilhança dos modelos ajustados quando eliminamos as observações 14, 27 e 104. A variação na estimativa é definida como: $(\hat{\beta}_{(I)} - \hat{\beta})/\hat{\beta}$, em que $\hat{\beta}_{(I)}$ denota a estimativa de máxima verossimilhança para β depois de retirarmos as observações pertencentes ao conjunto de índices I. Como esperado, as variações maiores são observadas sob modelos normais e menores sob modelos simétricos com caudas mais pesadas.

Tabela 1.1 *Variações (em %) na estimativa de máxima verossimilhança de β do modelo dado em (1.2) ajustado aos dados de retornos das ações quando eliminanos o conjunto I de observações.*

I	Normal	t_5	EP(0,9)
14	-20	-3	-7
27	-5	-2	-3
14 e 27	-26	-4	-8
104	-3	-11	-10

Tabela 1.2 *Variações (em %) nas estimativas dos desvios padrão assintóticos da estimativa de máxima verossimilhança de β do modelo dado em (1.2) ajustado aos dados de retornos das ações quando eliminanos o conjunto I de observações.*

I	Normal	t_5	EP(0,9)
14	-16	0	-5
24	-6	-2	-4
14 e 27	-24	-2	-10
104	7	5	6

1.2 Algumas distribuições simétricas

Definimos a seguir a classe simétrica univariada e em seguida apresentamos as principais propriedades das distribuições mais conhecidas.

Definição 1.1 *Seja a variável aleatória y com suporte em $\mathbb{R}$, com parâmetro de locação $\mu \in \mathbb{R}$ e de escala $\phi > 0$ com função de densidade de probabilidade dada por*

$$f(y;\mu,\phi) = \frac{1}{\sqrt{\phi}} g\left\{\frac{(y-\mu)^2}{\phi}\right\}, \qquad y \in \mathbb{R}, \tag{1.3}$$

para alguma função $g(\cdot)$ denominada função geradora de densidades, com $g(u) > 0$, para $u > 0$ e $\int_0^\infty u^{-1/2} g(u) du = 1$. Esta condição é necessária para que $f(y;\mu,\phi)$ seja uma função de densidade de probabilidade. Denotamos por $y \sim S(\mu,\phi)$ e denominamos de variável aleatória simétrica.

Como distribuições pertencentes a esta classe podemos citar a normal, normal contaminada, t-Student, t-Student generalizada, logística tipos I e II, logística generalizada, Kotz, Kotz generalizada, exponencial potência, dentre outras.

Teorema 1.1 *Seja y uma variável aleatória que tem distribuição simétrica com parâmetro de locação μ, parâmetro de escala ϕ e função geradora de densidades $g(\cdot)$. Então*

(i) y tem uma representação estocástica dada por

$$y \stackrel{d}{=} \mu + \sqrt{\phi}\, r\mathtt{u}\,, \tag{1.4}$$

em que $\stackrel{d}{=}$ denota mesma distribuição, $r \stackrel{d}{=} |z|$, com $z \sim S(0,1)$ sendo uma variável aleatória radial $\in \mathbb{R}^+$ e $\mathtt{u}$ uma variável aleatória uniforme em $D = \{-1, 1\}$, isto é, $Pr(\mathtt{u} = -1) = Pr(\mathtt{u} = 1) = \frac{1}{2}$. Além disso, r e $\mathtt{u}$ são variáveis aleatórias independentes. Para mais detalhes ver Fang, Kotz e Ng (1990).

(ii) As distribuições de r e $t = r^2$ ficam dadas por:

$$f_r(r) = 2\, g(r^2) \quad \text{e} \tag{1.5}$$

$$f_t(t) = \frac{1}{\sqrt{t}}\, g(t)\,. \tag{1.6}$$

(iii) De (1.4) temos que

$$\mathrm{E}(y) = \mu \quad \text{se} \quad \mathrm{E}(r) < \infty \quad \text{e} \tag{1.7}$$

$$\mathrm{Var}(y) = \phi\, \mathrm{E}(r^2) \quad \text{se} \quad \mathrm{E}(r^2) < \infty. \tag{1.8}$$

Vale salientar que, para encontrar o primeiro momento da variável aleatória y, deve existir o primeiro momento de r; e para encontrar o segundo momento da variavel aleatória y, deve existir o segundo momento de r. Para mais detalhes ver Fang, Kotz e Ng (1990). A distribuição de $t = r^2$ pode ser obtida para algumas distribuições simétricas, conforme descrito em Arellano-Valle, Galea e Iglesias, (2003) (ver Tabela 1.3).

Tabela 1.3 *Distribuição da variável r^2 para algumas distribuições simétricas.*

Distribuição simétrica	Distribuição de r^2
Normal	$\chi^2(1)$
Cauchy	$F(1,1)$
t-Student	$F(1,\nu)$
t-Student generalizada	$\frac{s}{r}F(1,r)$
Exponencial potência	$G^{\frac{1}{s}}(\frac{1}{s},\frac{1}{2})$
Kotz generalizada	$G^{\frac{1}{s}}(\frac{2m-1}{2s},\frac{r}{2})$

Nota: $G^{\frac{1}{s}}(\alpha,\lambda)$ significa que $r^{2s} \sim G(\alpha,\lambda)$ (a distribuição gama com parâmetros α e λ)

Algumas propriedades da distribuição normal podem ser estendidas para a classe simétrica de distribuições. Podemos ver que, se $y \sim S(\mu,\phi)$ então a função característica de y, $\varsigma_y(t) = \mathrm{E}(e^{ity})$ é dada por $e^{it\mu}\varphi(t^2\phi)$, $t \in I\!R$ para alguma função φ, com $\varphi(u) \in I\!R$ para $u > 0$. Quando existem, $\mathrm{E}(y) = \mu$ e $\mathrm{Var}(y) = \xi\phi$, em que $\xi > 0$ é uma constante que pode ser obtida pelo valor esperado do quadrado da variável radial ou pela derivada da função característica avaliada em zero dada por $\xi = -2\varphi'(0)$, com $\varphi'(0) = d\varphi(u)/du|_{u=0}$ e que não depende dos parâmetros μ e ϕ (Fang, Kotz e Ng, 1990, p.43). Kelker (1970) observa que se $u^{-\frac{1}{2}(k+1)}g(u)$ for integrável então o k-ésimo momento de y existe.

Temos também que, se $y \sim S(\mu,\phi)$ então $a + by \sim S(a + b\mu, b^2\phi)$, em que $a, b \in I\!R$ com $b \neq 0$, isto é, a distribuição de qualquer combinação linear de uma variável aleatória com distribuição simétrica é também simétrica. Como exemplo, se $y \sim S(\mu,\phi)$ então $z = (y-\mu)/\sqrt{\phi} \sim S(0,1)$, com função de densidade $f(z) = f(z;0,1) = g(z^2), z \in I\!R$ e chamaremos z de simétrica padrão.

Berkane e Bentler (1986) considerando uma distribuição simétrica padrão e que seus momentos existem, mostram que a função característica de z pode ser expandida como

$$\varsigma_z(t) = \sum_{k=0}^{\infty} i^k \mu'_k \frac{t^k}{k!},$$

em que $\mu'_k = \mathrm{E}(y^k) = i^{-k}\varsigma_z^{(k)}(0)$, com $\varsigma_z^{(k)}(0)$ denotando a $k-$ésima derivada de $\varsigma_z(t)$ avaliada em $t = 0$. Portanto, $\mu'_k = 0$ para k ímpar e para $k = 2m$, $m = 1, 2, \ldots,$

temos que

$$\mu'_{2m} = \frac{(2m)!}{2^m m!}(\mu'_2)^m\{k(m)+1\} \text{ e } k(m) = \frac{\varphi^{(m)}(0)}{\{\varphi^{(1)}(0)\}^m} - 1,$$

em que $\varphi^{(r)}(0)$ é a r-ésima derivada da função φ, avaliada em zero. Os coeficientes $k(m), m = 1, 2, \ldots$ são conhecidos como parâmetros de momentos e generalizam o coeficiente de curtose $\gamma_2 = 3\{k(2)+1\}$ de uma distribuição $S(\mu, \phi)$ (Muirhead, 1982). Cambanis, Huang e Simons (1981) observam que a família de distribuições simétricas coincide com a classe de distribuições elípticas univariadas. Contribuições importantes surgiram a partir dos trabalhos de Kelker (1970) para as distribuições elípticas univariadas e multivariadas. Podemos citar algums trabalhos que discutem propriedades dessas distribuições, tais como Berkane e Bentler (1986), Muirhead (1980 e 1982), Rao (1990), Cambanis, Huang e Simons (1981) e Anderson e Fang (1987). Na literatura podemos encontrar excelentes livros, tais como Fang, Kotz e Ng (1990), Fang e Anderson (1990) e Fang e Zhang (1990) e as teses de doutorado de Arellano-Valle (1994), de Uribe-Opazo (1997) e de Cysneiros (2004).

A seguir apresentaremos algumas distribuições simétricas com suporte na reta real para $u = (y-\mu)^2/\phi$, em que $y \sim S(\mu, \phi)$.

1.2.1 Distribuição Normal

A normal é a distribuição pertencente à classe simétrica mais utilizada devido a todo o desenvolvimento teórico e aplicado estabelecido no decorrer dos anos. Alguns resultados devidos a Muirhead (1982), Devlin, Gnanadesikan e Kettenring (1976) caracterizam a distribuição normal, chamada de normal composta, dentro da classe de distribuições simétricas.

Se $y \sim S(\mu, \phi)$ e a função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{1}{\sqrt{2\pi}}\exp\{-u/2\}, \qquad u > 0,$$

então y tem uma distribuição normal denotada por $y \sim \mathrm{N}(\mu, \phi)$, e sua função

característica é dada por

$$\varsigma_y(t) = e^{it\mu}\exp\{-t^2\phi/2\}, \qquad t \in I\!R.$$

Se $y \sim \mathrm{N}(\mu, \phi)$ então $\mathrm{E}(y) = \mu$, $\mathrm{Var}(y) = \phi$ e os momentos centrais de ordem r são

$$\mu_r = \mathrm{E}\{(y-\mu)^r\} = \begin{cases} 0, & r \text{ ímpar} \\ \phi^{r/2} r!/\{2^{r/2}(r/2)!\}, & r \text{ par,} \end{cases}$$

portanto o coeficiente de curtose é $\gamma_2 = 3$.

1.2.2 Distribuição de Cauchy

Dizemos que a variável aleatória $y \sim S(\mu, \phi)$ tem distribuição de Cauchy se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{1}{\pi}(1+u)^{-1}, \qquad u > 0.$$

Denotamos por $y \sim \mathrm{C}(\mu, \phi)$ e sua função característica é dada por

$$\varsigma_y(t) = \exp\{it\mu - |t|\sqrt{\phi}\}, \qquad t \in I\!R.$$

Em particular, os momentos e os cumulantes para essa distribuição não existem. Sua mediana e moda são iguais a μ, os quartis superior e inferior iguais a $\mu \pm \sqrt{\phi}$. Os pontos de inflexão da função de densidade são $\mu \pm \sqrt{3\phi}$, e os valores da função de distribuição acumulada nos pontos de inflexão são 0,273 e 0,723 que são próximos aos correspondentes da distribuição normal (0,159 e 0,841). A diferença mais importante é que a distribuição de Cauchy tem caudas mais pesadas do que a normal. Um resultado interessante é que para $a_j \neq 0$, $\sum_{j=1}^n a_j y_j$ e $y_j \sim \mathrm{C}(\mu_j, \phi_j)$ independentes temos uma distribuição de Cauchy com parâmetros de locação $\mu = \sum_{i=1}^n a_j \mu_j$ e escala, $\phi = \sum_{i=1}^n a_j^2 \phi_j$. Em particular, se y_j são i.i.d. então $\bar{y} = n^{-1}\sum_{i=1}^n y_j \sim \mathrm{C}(\mu, \phi)$. A distribuição de Cauchy padronizada reduz-se ($\mu = 0$ e $\phi = 1$) à distribuição central $t-$Student com um grau de liberdade. Temos ainda a relação $y = \mu + \phi N_1/N_2$ em que $N_i \sim \mathrm{N}(0,1)$ para $i = 1, 2$ independentes. Com essa relação é possível definir um gerador de números aleatórios para a distribuição de Cauchy.

1.2.3 Distribuição t−Student

A variável aleatória y tem distribuição $t-$Student com ν graus de liberdade se $y \sim S(\mu, \phi)$ e a sua função geradora de densidades for da forma

$$g(u) = \frac{\nu^{\nu/2}}{\mathrm{B}(1/2, \nu/2)}(\nu + u)^{-\frac{\nu+1}{2}}, \qquad \nu > 0, u > 0,$$

em que $\mathrm{B}(\cdot, \cdot)$ é a função beta e denotamos $y \sim t(\mu, \phi, \nu)$. Logo, a função de densidade de y é obtida de (1.3) aplicando a função $g(\cdot)$ acima. Podemos encontrar a sua função característica definida em Fang, Kotz e Ng (1990, p.87). Relacionando algumas propriedades temos que se y é definido por $y = \theta^{1/2}z$, em que $\theta \sim \mathrm{GI}(\nu/2, \nu/2)$ (gama inversa), $\nu > 0$ e $z \sim \mathrm{N}(0, 1)$ independentes, então $y \sim t(0, 1, \nu)$.

Se $t(0, 1, \nu)$ temos o seguinte :

(i) Para $\nu > r$, seus momentos de ordem r existem e são dados por

$$\mathrm{E}(y^r) = \begin{cases} 0, & r \text{ ímpar} \\ \nu^{r/2}\Gamma(\frac{r+1}{2})\Gamma(\frac{\nu-r}{2})/\{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})\}, & r \text{ par,} \end{cases}$$

em que $\Gamma(\cdot)$ denota a função gama. Logo, $\mathrm{E}(y) = 0$ para $\nu > 1$ e $\mathrm{Var}(y) = \nu/(\nu - 2)$ para $\nu > 2$. Se $r \geqslant \nu$ e r par temos que o momento de ordem r é infinito;

(ii) o desvio médio é dado por

$$\mathrm{E}(|y|) = \frac{\nu^{1/2}\Gamma(\frac{\nu-1}{2})}{\Gamma(1/2)\Gamma(\nu/2)};$$

(iii) o coeficiente de curtose é dado por $\gamma_2 = 3 + 6/(\nu - 4)$, para $\nu > 4$. Observe que este coeficiente é maior do que o coeficiente da distribuição normal.

(iv) $y^2 \sim F_{(1,\nu)}$ em que $F_{(1,\nu)}$ denota a distribuição $F-$Snedecor com 1 e ν graus de liberdade;

(v) se $w = (\nu + 1)/(\nu + y^2)$ então

$$\mathrm{E}(y^{2k}w^{\ell}) = \frac{(-\frac{\nu+1}{2})^{\ell}}{\nu^{\ell-k}} \frac{\mathrm{B}[(2k+1)/2, \{\nu + 2(\ell - k)\}/2]}{\mathrm{B}(1/2, \nu/2)},$$

para $\ell = 0, 1, 2$ e $k = 1, 2, \ldots$;

(*vi*) a função de densidade de y tem pontos de inflexão em $\pm\{\nu/(\nu+2)\}^{1/2}$;

(*vii*) a variável aleatória $u = (1+\nu/y^2)^{-1}$ tem distribuição beta com parâmetros $a = 1/2$ e $b = \nu/2$ (Manoukian, 1985, p.41);

(*viii*) $y|\theta \sim \mathrm{N}(0,\nu)$;

(*ix*) $v|\theta \sim \mathrm{GI}\{(\nu+1)/2, (\nu+y^2)/2\}$.

Baseados nessas propriedades podemos ver que a distribuição $t-$Student de parâmetros (μ,ϕ,ν) tende a um distribuição normal com média μ e variância ϕ quando $\nu \to \infty$. Quando $\nu = 1$ temos a distribuição de Cauchy com parâmetros μ e ϕ.

1.2.4 Distribuição t−Student Generalizada

Uma variável aleatória $y \sim S(\mu,\phi)$ com a função geradora de densidades definida por

$$g(u) = \frac{s^{r/2}}{\mathrm{B}(1/2, r/2)}(s+u)^{-\frac{r+1}{2}}, \qquad s, r > 0, u > 0,$$

é chamada $t-$Student generalizada com parâmetros (μ,ϕ,s,r) (Dickey, 1967). Como membro desta família de distribuições temos a $t-$Student($s = r = \nu$) e Cauchy ($s = r = 1$). Quando $\sqrt{s} = c$ e $(r+1)/2 = m$, com $m > 1/2$ temos a distribuição Pearson VII (Fang, Kotz e Ng, 1990).

Suponha $y|\theta \sim \mathrm{N}(\mu, \nu\phi)$, em que $\theta \sim \mathrm{GI}(r/2, s/2)$, independentes com $s, r > 0$ podendo não ser inteiros. Podemos relacionar algumas propriedades :

(*i*) $y \sim t\mathrm{G}(\mu,\phi,s,r)$;

(*ii*) $\mathrm{E}(y) = \mu$ para $r > 1$, $\mathrm{Var}(y) = \{s/(r-2)\}\phi$ para $r > 2$ e o coeficiente de curtose $\gamma_2 = 3 + 6/(r-4)$ para $r > 4$. Vale salientar que o coeficiente de curtose não depende do parâmetro s e é maior do que o coeficiente de curtose da normal;

(*iii*) $\theta|y \sim \mathrm{GI}\{(r+1)/2, (s+z^2)/2\}$, em que $z^2 = (y-\mu)^2/\phi$;

(*iv*) $u^2 = rz^2/s \sim F_{(1,r)}$;

(*v*) se $w = (r+1)/(s+z^2)$ então

$$\mathrm{E}(z^{2k}w^{\ell}) = \frac{(-\frac{r+1}{2})^{\ell}}{s^{\ell-k}}\frac{\mathrm{B}[(2k+1)/2, \{r+2(\ell-k)\}/2]}{\mathrm{B}(1/2, r/2)},$$

para $\ell = 0, 1, 2$ e $k = 1, 2, \ldots$;

(*vi*) os parâmetros s e r tem uma relação com o parâmetro de curtose e o segundo momento central (Johnson e Kotz, 1970, p.116) dados por

$$r = \frac{2(2\gamma_2 - 3)}{\gamma_2 - 3} \qquad \text{e} \qquad s = \frac{2\mu_2\gamma_2}{\gamma_2 - 3};$$

(*vii*) o ℓ-ésimo momente existe se e somente se $r > \ell$;

(*viii*) para a variável aleatória $y = \theta^{-1/2}z$, z e θ variáveis aleatórias independentes, em que $z \sim \mathrm{N}(0, 1)$ e $\theta \sim \mathrm{GI}(r/2, s/2)$ então $y \sim t\mathrm{G}(0, 1, s, r)$.

1.2.5 Distribuição Logística-I

Dizemos que a variável aleatória $y \sim S(\mu, \phi)$ tem distribuição logística-I (Fang, Kotz e Ng, 1990) se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = c\frac{e^{-u}}{(1 + e^{-u})^2}, \qquad u > 0,$$

em que c é a constante normalizadora obtida da relação $\int_0^\infty u^{-1/2}g(u) = 1$, logo $c \approx 1,484300029$ e é denotada por $y \sim \mathrm{LI}(\mu, \phi)$. Temos que $\mathrm{E}(y) = \mu$, $\mathrm{Var}(y) \approx 0,79569\phi$ e $\gamma_2 \approx 2,385165$. Observe que o coeficiente de curtose da distribuição logística-I é menor do que o coeficiente de curtose da distribuição normal.

Se $v = (e^{-z^2} - 1)/(1 + e^{-z^2})$, com $z^2 = (y - \mu)^2/\phi$, então

$$\mathrm{E}(z^{2r}v^\ell) = \frac{c}{2}(-1)^\ell \int_0^1 \{\log(1+s) - \log(1-s)\}^{r-1/2} s^\ell ds, \;\; \ell = 0, 1, 2, \ldots \text{ e } r = 1, 2, \ldots$$

1.2.6 Distribuição Logística-II

Dizemos que a variável aleatória $y \sim S(\mu, \phi)$ tem distribuição logística-II se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{e^{-u^{1/2}}}{(1 + e^{-u^{1/2}})^2}, \qquad u > 0,$$

denotada por $y \sim \mathrm{LII}(\mu, \phi)$. A função característica é dada por $\varsigma_y(t) = \frac{2(\mathrm{e}^{it\mu}\pi\phi^{1/2}t)}{(\mathrm{e}^{\pi\phi^{1/2}t} - \mathrm{e}^{-\pi\phi^{1/2}t})}$, $t \in I\!R$. Temos que $\mathrm{E}(y) = \mu, \mathrm{Var}(y) = \pi^2\phi/3$ e $\gamma_2 = 4,2$. E ainda, tem-se

que a mediana e moda são iguais à média. Uma relação bastante útil para gerar amostras aleatórias é dada por Hastings e Peacock (1975). Seja $u \sim \mathrm{U}(0,1)$ e $y = \mu + \sqrt{\phi}\log\{u/(1-u)\}$, então $y \sim \mathrm{LII}(\mu, \phi)$. A função de distribuição logística-II é comumente usada para representar curvas de crescimento em economia e demografia (Johnson e Kotz, 1970).

1.2.7 Distribuição Logística Generalizada

Uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição logística generalizada se a sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{\alpha}{\mathrm{B}(m,m)} \left\{ \frac{\mathrm{e}^{-\alpha\sqrt{u}}}{(1+\mathrm{e}^{-\alpha\sqrt{u}})^2} \right\}^m, \qquad m > 0,\ u > 0,$$

em que $\alpha = \alpha(m)$ com $\alpha(\cdot)$ definida em $\mathbb{R}^+$ e $\alpha(m) > 0$, para $m > 0$, e é denotada por $y \sim \mathrm{LG}(\mu, \phi, m)$. Esta distribuição pertence à família de distribuições de Perks (veja Johnson e Kotz, 1970). Se $\alpha(m) = 1, \forall m > 0$ e $m = 1$ temos a distribuição logística-II. Gumbel (1944) utiliza a distribuição logística generalizada com uma função particular $\alpha(\cdot)$ para a distribuição da $m-$ésima amplitude (média entre o maior e o menor valor de uma amostra aleatória de tamanho n) para uma classe de distribuições simétricas. Temos que $\mathrm{E}(y) = \mu, \mathrm{Var}(y) = 2\psi'(m)\phi/\alpha(m)$ e $\gamma_2 = 3 + \frac{\psi'''(m)}{2\psi'(m)^2}$, em que $\psi'(\cdot)$ e $\psi'''(\cdot)$ são a primeira e a terceira derivadas da função digama, respectivamente e $\forall m > 0$ temos que $\gamma_2 > 0$. Quando $m \to \infty$ temos que $\gamma_2 \to 3$, ou seja, o coeficiente de curtose da logística generalizada converge para o coeficiente de curtose da normal.

1.2.8 Distribuição Exponencial Dupla

Uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição exponencial dupla (Laplace) se a sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{1}{2}\exp\{-\sqrt{u}\}, \qquad u > 0,$$

e denotamos por $y \sim \mathrm{ED}(\mu, \phi)$. A função característica é dada por $\varsigma_y(t) = \frac{e^{it\mu}}{1+t^2\phi}$, $t \in \mathbb{R}$. Se $z \sim \mathrm{ED}(0,1)$ temos os momentos μ'_r dados por

$$\mu'_r = \mathrm{E}(z^r) = \begin{cases} 0, & r \text{ ímpar} \\ r!, & r \text{ par.} \end{cases}$$

Portanto, $\mathrm{E}(y) = \mu$, $\mathrm{Var}(y) = 2\phi$, a mediana e a moda são iguais a μ e ainda o coeficiente de curtose $\gamma_2 = 6$. Os quartis superior e inferior são $\mu \pm 0,534\sqrt{\phi}$.

1.2.9 Distribuição Exponencial Potência

Uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição exponencial potência (Box e Tiao, 1973, Cap. 3) se a sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = C(k)\exp\{-\frac{1}{2}u^{1/(1+k)}\}, \qquad -1 < k \leqslant 1,\ u > 0,$$

em que $C(k)^{-1} = \Gamma(1 + \frac{1+k}{2})2^{1+(1+k)/2}$ e denotamos por $y \sim \mathrm{EP}(\mu, \phi, k)$.

Temos ainda que

$$\mathrm{E}(y) = \mu,\ \mathrm{Var}(y) = 2^{(1+k)}\left[\frac{\Gamma\{\frac{3(1+k)}{2}\}}{\Gamma(\frac{1+k}{2})}\right]\phi \text{ e } \gamma_2 = \frac{\Gamma\{\frac{5}{2}(1+k)\}\Gamma(\frac{1+k}{2})}{\Gamma^2\{\frac{3}{2}(1+k)\}}.$$

Observe que para $k > 0$, temos que $\gamma_2 > 3$, ou seja, a distribuição é leptocúrtica e para $k < 0$, temos $\gamma_2 < 3$, ou seja, a distribuição é platicúrtica. Podemos ver o parâmetro k como uma medida de curtose, ou mesmo, uma medida de não normalidade pois quando $k = 0$ temos a distribuição normal. Em particular, quando $k = 1$ temos a distribuição exponencial dupla. Se k tende a -1, a distribuição tende a uma distribuição uniforme no intervalo $(\mu - \sqrt{3\phi}, \mu + \sqrt{3\phi})$.

Se $y = (2w)^{1/r}v$ em que $v \sim \mathrm{U}(-1,1)$, $w \sim \mathrm{G}(1 + 1/r, 1)$ e $r = 2/(1+k)$ independentes (veja Devroye, 1986, pp.174-175), então $y \sim \mathrm{EP}(0,1,k)$. Essa relação é suficiente para gerar amostras de uma distribuição $\mathrm{EP}(0,1,k)$.

1.2.10 Distribuição Potência Estendida

Uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição potência estendida (Albert, Delampady e Polasek, 1991) se a sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = C(c,\lambda)\exp\left[-\frac{1}{2}c\rho_\lambda\{1 + u/(c-1)\}\right],$$

denotamos por $y \sim \mathrm{PE}(\mu, \phi, \lambda)$ em que $C(c, \lambda)$ é uma constante normalizadora, $c > 1, \lambda \geqslant 0$, $u > 0$ e

$$\rho_\lambda(v) = \begin{cases} \dfrac{v^\lambda - 1}{\lambda}, & \text{se } \lambda > 0 \\ \lim\limits_{\lambda \to 0} \dfrac{v^\lambda - 1}{\lambda}, & \text{se } \lambda = 0. \end{cases}$$

Podemos citar alguns casos particulares, quando $\lambda = 1$ temos a distribuição $\mathrm{N}(\mu, \phi\{c-1\}/c)$, se $\lambda = 0$ temos a distribuição $t-$Student $(\mu, \phi, c-1)$ e quando $\lambda = 1/2$ temos a distribuição exponencial dupla. Se $\lambda > 0$, os momentos $\mathrm{E}(y^k)$ existem para $k > 0$.

1.2.11 Distribuição de Kotz

Dizemos que uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição de Kotz (Kotz, 1975) se a sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{r^{(2N-1)/2}}{\Gamma(\frac{2N-1}{2})} u^{N-1} \mathrm{e}^{-ru}, \qquad r > 0, N \geqslant 1,\ u > 0,$$

e denotamos por $y \sim \mathrm{K}(\mu, \phi, N, r)$. Quando $N = 1$ temos a distribuição normal com média μ e variância $\phi/(2r)$. Ainda se $N > 1$, a distribuição é bimodal com modas em $y = \mu \pm \sqrt{(N-1)/(r\phi)}$. Temos que $\mathrm{E}(y) = \mu$, $\mathrm{Var}(y) = \{(2N-1)/(2r)\}\phi$, o coeficiente de curtose $\gamma_2 = (2N+1)/(2N-1)$ e os momentos centrais de ordem $2m$ dados por

$$\mu_{2m} = \mathrm{E}\{(y-\mu)^{2m}\} = \frac{\Gamma\{(2N+2m-1)/2\}}{r^m \Gamma\{(2N-1)/2\}} \phi^m,\ m > 0.$$

Se $z^2 = (y-\mu)^2/\phi$ então $z^2 \sim \mathrm{G}(\{2N-1\}/2, r)$. Em particular, se $N = 1$ e $r = 1/2$ então temos que $z^2 \sim \chi^2_1$.

1.2.12 Distribuição de Kotz Generalizada

Seja $y \sim S(\mu, \phi)$ com a função geradora de densidades $g(\cdot)$ dada por

$$g(u) = \frac{s r^{(2N-1)/2s}}{\Gamma(\frac{2N-1}{2s})} u^{N-1} \mathrm{e}^{-ru^s}, \qquad r, s > 0, N \geqslant 1,\ u > 0.$$

Então y tem distribuição de Kotz generalizada e denotamos por $y \sim \mathrm{KG}(\mu, \phi, N, r, s)$. Quando $s = 1$ a distribuição reduz a $\mathrm{K}(\mu, \phi, N, r)$ e, quando $N = 1, s = 1$ e $r = 1/2$ temos a distribuição normal $\mathrm{N}(\mu, \phi)$. Ainda, se $N = 1, r = 1/2$ e $s = 1/(1 + k)$ temos a distribuição exponencial potência.

Temos que

$$\mathrm{E}(y) = \mu,\ \mathrm{Var}(y) = \frac{\Gamma\{(2N-1)/2s\}}{r^{1/s}\Gamma\{(2N-1)/2s\}}\phi \text{ e } \gamma_2 = \frac{\Gamma\{(2N-1)/2s\}\Gamma\{(2N+3)/2s\}}{\Gamma^2\{(2N+1)/2s\}}$$

e os momentos centrais de ordem $2m$ são dados por

$$\mu_{2m} = \mathrm{E}\{(y-\mu)^{2m}\} = \frac{\Gamma\{(2N+2m-1)/2s\}}{r^{m/s}\Gamma\{(2N-1)/2s\}}\phi^m, \qquad m > 0.$$

1.2.13 Distribuição Normal Contaminada

Considere uma variável aleatória $y \sim S(\mu, \phi)$ com a função geradora de densidades $g(\cdot)$ dada por

$$g(u) = (1-\epsilon)\frac{1}{\sqrt{2\pi}}\exp\{-u/2\} + \epsilon\frac{1}{\sqrt{2\pi\sigma}}\exp\{-u/(2\sigma^2)\},$$

em que $u > 0, \sigma > 0$ e $0 \leqslant \epsilon \leqslant 1$ e denotaremos $y \sim \mathrm{NC}(\mu, \phi, \epsilon, \sigma^2)$. Temos que $\mathrm{E}(y) = \mu$ e $\mathrm{Var}(y) = \{1 + \epsilon(\sigma^2 - 1)\}\phi$. O coeficiente de curtose fica dado por (Berkane e Bentler, 1986)

$$\gamma_2 = \frac{3\{1 + \epsilon(\sigma^4 - 1)\}}{\{1 + \epsilon(\sigma^2 - 1)\}^2}.$$

Little (1988) incorpora parâmetros adicionais para ajustar a curtose utilizando esta distribuição.

Como ilustração, temos nas Figuras 1.3 a 1.6 os gráficos da função de densidade de várias distribuições simétricas (linha cheia) comparando com a função de densidade da distribuição normal (linha pontilhada). Para todas as distribuições aqui consideradas, o parâmetro de locação e escala são fixados em $\mu = 0$ e $\phi = 1$, respectivamente.

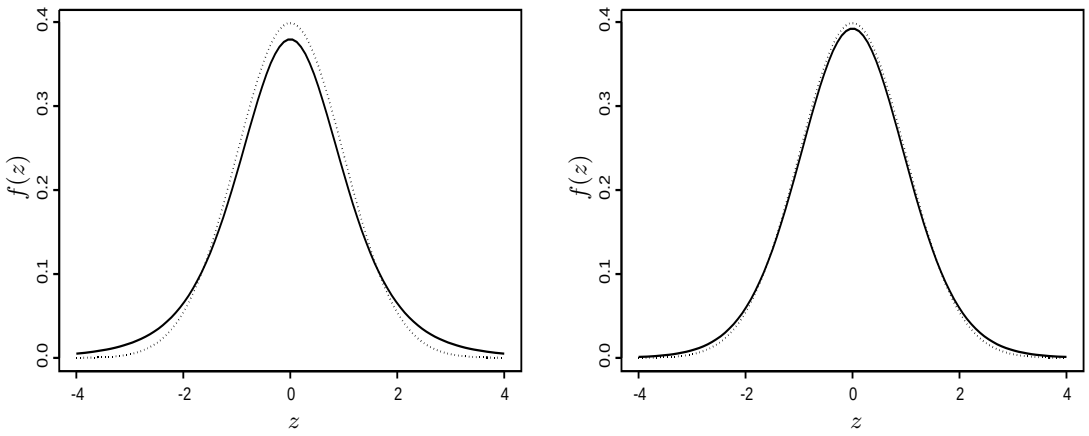

Figura 1.3 *Gráficos da função de densidade da distribuição t-Student com* $\nu = 5$ *(esquerda) e com* $\nu = 15$ *(direita).*

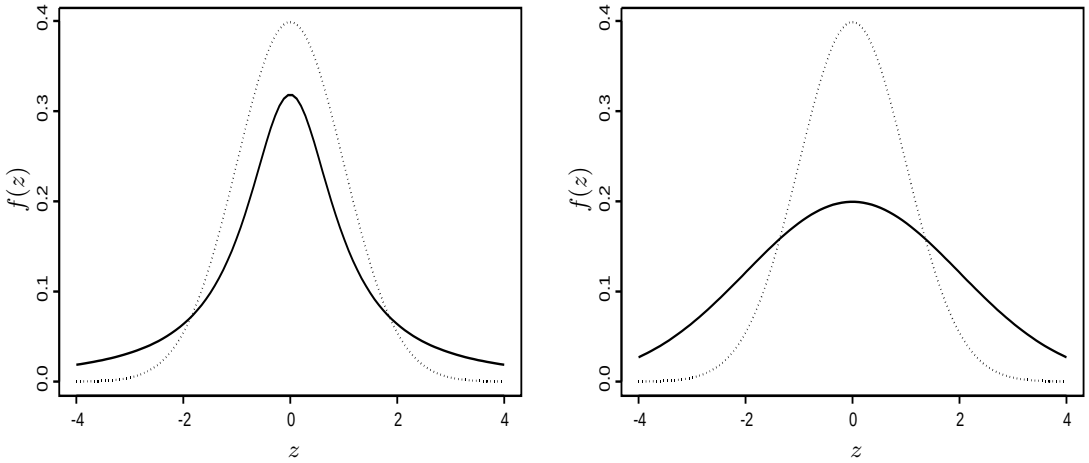

Figura 1.4 *Gráficos da função de densidade da distribuição t-Student com* $\nu = 1$ *(esquerda) e normal contaminada com* $\epsilon = 0,7$ *e* $\sigma = 2$ *(direita).*

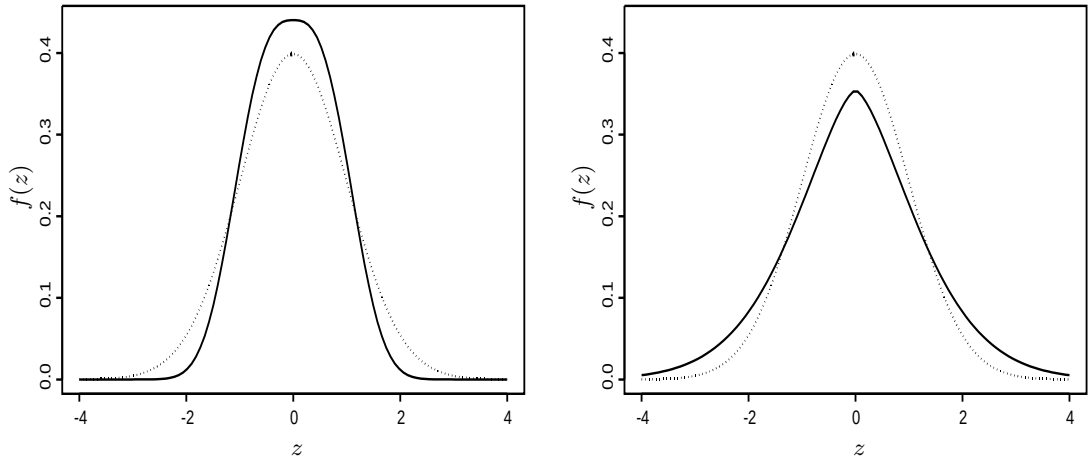

Figura 1.5 *Gráficos da função de densidade da distribuição exponencial potência com* $k = -0,3$ *(esquerda) e com* $k = 0,3$ *(direita).*

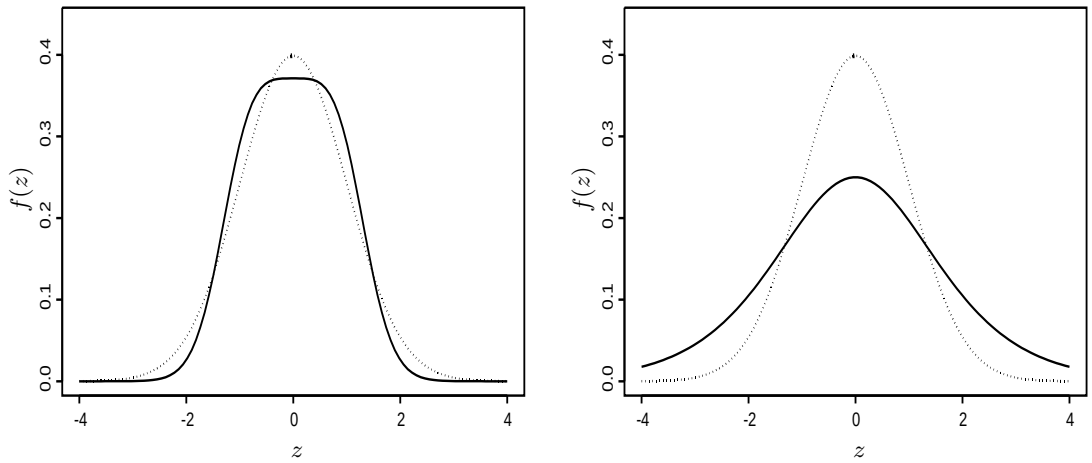

Figura 1.6 *Gráficos da função de densidade da distribuição logística-I (esquerda) e logística-II (direita).*

CAPÍTULO 2

Modelos de regressão com erros simétricos

2.1 Introdução

A classe de distribuições simétricas tem recebido uma crescente atenção na literatura estatística nos últimos anos (veja por exemplo, Fang, Kotz e Ng, 1990; Fang e Zhang, 1990; Fang e Anderson, 1990 e Gupta e Varga, 1993). Uma revisão de diferentes áreas em que as distribuições simétricas são aplicadas é descrita em Chmielewski (1981). Em muitas situações da modelagem estatística há necessidade de procurar modelos cujas estimativas sejam menos sensíveis a observações aberrantes. É bem conhecido que as estimativas de máxima verossimilhança sob erros normais são altamente sensíveis a observações aberrantes. Como alternativa robusta, Lange, Little e Taylor (1989) propõem obter as estimativas de máxima verossimilhança dos coeficientes da regressão linear e não-linear em modelos com erros $t-$Student, enquanto Little (1988) e Yamaguchi (1990) utilizam a distribuição normal contaminada para os erros. Em ambos os modelos incorpora-se parâmetros adicionais os quais permitem ajustar a curtose da distribuição aos dados. No caso da $t-$Student os graus de liberdade são usados para controlar a curtose. Taylor (1992) propõe o ajuste de um modelo de regressão linear supondo erros distribuídos como exponencial potência com um parâmetro extra de forma. Albert, Delampady e Polasek (1991) estendem resultados para a família potência estendida estudando propriedades robustas no enfoque de estimação dos parâmetros do modelo de regressão. Arellano–Valle (1994) apresenta vários resultados para a $t-$Student com aplicações em modelos com erros nas variáveis. Ferrari e Arellano–Valle (1996) desenvolvem correções de Bartlett e tipo-Bartlett para teste de hipóteses em modelos de regressão linear com erros $t-$Student e Uribe–Opazo (1997) e Ferrari e Uribe–Opazo (2001) estendem esses resultados para modelos de regressão linear com

erros simétricos. Uribe–Opazo, Ferrari e Cordeiro (2003) desenvolvem correções tipo-Bartlett para modelos de regressão linear com erros simétricos e Cordeiro (2004) desenvolve correções de Bartlett para os modelos de regressão não-lineares simétricos.

2.2 Modelos simétricos de regressão

Para definir a classe de modelos de regressão com erros simétricos suponha que $\epsilon_1, \ldots, \epsilon_n$ são variáveis aleatórias independentes com função de densidade definida como

$$f_{\epsilon_i}(\epsilon) = \frac{1}{\sqrt{\phi}} g\{\epsilon^2/\phi\}, \tag{2.1}$$

$\epsilon \in I\!R$ and $g(\cdot)$ definida como na Seção 1.2. O modelo simétrico não-linear é definido aqui por

$$y_i = \mu_i(\boldsymbol{\beta}; \mathbf{x}_i) + \epsilon_i, \tag{2.2}$$

em que $\mu_i = \mu_i(\boldsymbol{\beta}; \mathbf{x}_i)$ é uma função não-linear contínua e diferenciável de $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ tal que a matriz de derivadas $\mathbf{D}_\beta = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}$ tenha posto p ($p < n$) para todo $\boldsymbol{\beta}$ com $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, $\mathbf{y} = (y_1, \ldots, y_n)^T$ é o vetor de respostas observadas, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ contém valores de p variáveis explanatórias e $\epsilon_i \sim S(0, \phi)$. No caso linear tem-se que $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ com $\mathbf{X}$ uma matriz $n \times p$ de posto completo cujas linhas são denotadas por $\mathbf{x}_i^T$, $i = 1, \ldots n$. A densidade de y_i é dada por

$$f_{y_i}(y_i) = \frac{1}{\sqrt{\phi}} g(u_i), \tag{2.3}$$

em que $u_i = (y_i - \mu_i)^2/\phi$ e $y_i \sim S(\mu_i, \phi)$. Quando existem, $\mathrm{E}(y_i) = \mu_i$ e $\mathrm{Var}(y_i) = \xi\phi$. O modelo definido por (2.2) e (2.3) é dito modelo simétrico de regressão não-linear. O logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ é dado por

$$\mathrm{L}(\boldsymbol{\theta}) = -\frac{n}{2}\log\phi + \sum_{i=1}^{n} \log\{g(u_i)\}.$$

A função $\mathrm{L}(\boldsymbol{\theta})$ é assumida ser regular (Cox e Hinkley, 1974, Cap. 9) com respeito a $\boldsymbol{\beta}$ e ϕ. Condições regulares são encontradas, também, em Serfling (1980, p. 144). Para obter a função escore e as matrizes de informação de Fisher precisamos

derivar $\mathrm{L}(\boldsymbol{\theta})$ com respeito aos parâmetros desconhecidos e então calcular alguns momentos dessas derivadas. Supomos aqui que tais derivadas existem. Contudo, algumas distribuições simétricas não satisfazem as condições de regularidade, por exemplo, a exponencial dupla. Esses casos não serão considerados.

As funções escore para $\boldsymbol{\beta}$ e ϕ tomam, respectivamente, as formas

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

e

$$\mathbf{U}_{\phi}(\boldsymbol{\theta}) = (2\phi)^{-1}\{\phi^{-1}Q_{\mathbf{v}}(\boldsymbol{\beta}) - n\},$$

em que $Q_V(\boldsymbol{\beta}) = \{\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})\}^T\mathbf{D}(\mathbf{v})\{\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})\}$, $\mathbf{D}(\mathbf{v}) = \mathrm{diag}\{v_1,\ldots,v_n\}$ com $v_i = -2W_g(u_i)$. Expressões para $W_g(u)$ e $W_g'(u)$ para algumas distribuições simétricas são dadas na Tabela 2.1. Algoritmos de estimação são discutidos em Smyth (1996). Um processo iterativo para obter as estimativas de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ pode ser desenvolvido usando, por exemplo, o método *scoring* de Fisher. O processo iterativo conjunto é dado por

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (4d_g)^{-1}\{\mathbf{D}_{\beta}^{T(m)}\mathbf{D}_{\beta}^{(m)}\}^{-1}\mathbf{D}_{\beta}^{(m)T}\mathbf{D}(\mathbf{v}^{(m)})\{\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta}^{(m)})\} \qquad (2.4)$$

e

$$\phi^{(m+1)} = \frac{1}{n}Q_V(\boldsymbol{\beta}^{(m+1)}) \qquad (m=0,1,2,\ldots), \qquad (2.5)$$

em que $d_g = \mathrm{E}\{W_g^2(U^2)U^2\}$ com $U \sim S(0,1)$. Alguns valores de d_g podem ser encontrados na Tabela 2.2. Note que em (2.5) tem-se sempre uma solução positiva para $\phi^{(m+1)}$.

No caso linear temos uma simplificação na função escore $\mathbf{U}_{\beta}(\boldsymbol{\theta})$ e conseqüentemente no processo iterativo, visto que $\mathbf{D}_{\beta} = \mathbf{X}$. A função escore fica dada por $\mathbf{U}_{\beta}(\boldsymbol{\theta}) = \frac{1}{\phi}\mathbf{X}^T\mathbf{D}(\mathbf{v})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ e o processo iterativo assume a forma alternativa

$$\boldsymbol{\beta}^{(m+1)} = \{\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{y} \qquad (2.6)$$

e

$$\phi^{(m+1)} = \frac{1}{n}\{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{(m+1)}\}^T\mathbf{D}(\mathbf{v}^{(m+1)})\{\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{(m+1)}\} \qquad (m=0,1,2,\ldots). \quad (2.7)$$

Em (2.6) o peso $v_i^{(m)}$ é inversamente proporcional à distância entre o valor observado y_i e o seu valor predito $\mu_i^{(m)}$ (a menos da normal que é uma função constante e da logística-I que é diretamente proporcional), de forma que observações mais distantes tendem a ter pesos menores no processo de estimação (veja discussão, por exemplo, em Lange, Little e Taylor, 1989). No caso normal linear os estimadores de máxima verossimilhança tomam expressões em forma fechada, pois $v_i = 1$, para todo i. Para a distribuição $t-$Student com ν graus de liberdade, temos que $g(u) = c(1+u/\nu)^{-(\nu+1)/2}, \nu > 0$ e $u > 0$ de forma que $W_g(u_i) = -(\nu+1)/2(\nu+u_i)$ e $v_i = (\nu+1)/(\nu+u_i)$, para todo i. Para a distribuição exponencial potência com parâmetro de forma $\gamma = 1/(1+k)$ fixado, $g(u) = ce^{-0,5u^{\gamma-1}}$, $u > 0$ e $\gamma \geqslant 1/2$, então $W_g(u_i) = -\frac{1}{2}\gamma u_i^{\gamma-1}$ e $v_i = \gamma u_i^{\gamma-1}$.

A Figura 2.1 descreve o comportamento de v contra u para alguns valores de graus de liberdade da distribuição t de Student, enquanto na Figura 2.2 tem-se o gráfico de v contra u para alguns valores de k para a distribuição exponencial potência. Nota-se que pontos extremos têm peso menor no processo iterativo (2.6) a medida que os graus de liberdade diminuem sob erros t de Student e o parâmetro k aumenta sob erros exponencial potência.

Tabela 2.1 *Expressões para $W_g(u)$ e $W_g'(u)$ para algumas distribuições simétricas.*

Distribuição	$W_g(u)$	$W_g'(u)$
Normal	$-\frac{1}{2}$	0
$t-$Student	$-\frac{\nu+1}{2(\nu+u)}$	$\frac{(\nu+1)}{2(\nu+u)^2}$
$t-$Student generalizada	$-\frac{(r+1)}{2(s+u)}$	$\frac{(r+1)}{2(s+u)^2}$
Logística-I	$-\tanh(\frac{u}{2})$	$-\mathrm{sech}(\frac{u}{2})/2$
Logística-II	$-\frac{\exp(-\sqrt{u})-1}{(-2\sqrt{u})[1+\exp(-\sqrt{u})]}$	$\frac{2\exp(-\sqrt{u})\sqrt{u}+\exp(-2\sqrt{u})-1}{-4u^{3/2}[1+\exp(-\sqrt{u})]^2}$
Logística generalizada	$\frac{-\alpha m[\exp(-\alpha\sqrt{u})-1]}{(-2\sqrt{u})[1+\exp(-\alpha\sqrt{u})]}$	$-\frac{\alpha m}{4}\frac{2\alpha\exp(-\alpha\sqrt{u})\sqrt{u}+\exp(-2\alpha\sqrt{u})-1}{u^{3/2}[1+\exp(-\alpha\sqrt{u})]^2}$
Exponencial potência	$-\frac{1}{2(1+k)u^{k/(k+1)}}$	$\frac{k}{(1+k)^2 2u^{(2k+1)/(1+k)}}$

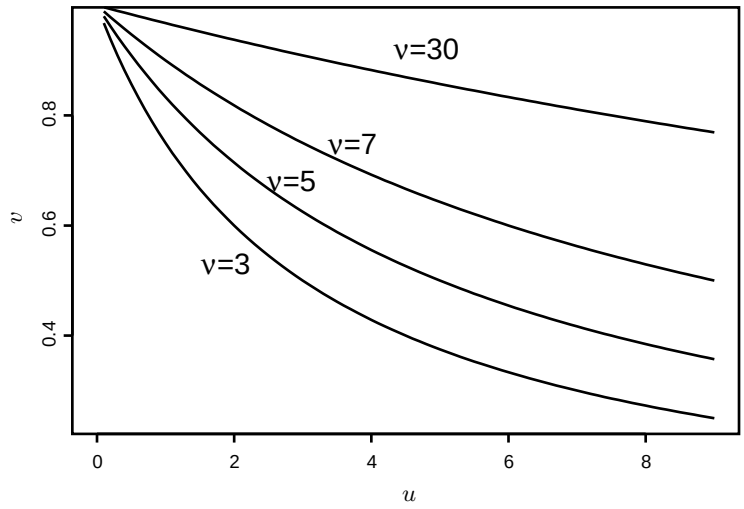

Figura 2.1 *Comportamento de v contra u para alguns graus de liberdade da distribuição t de Student.*

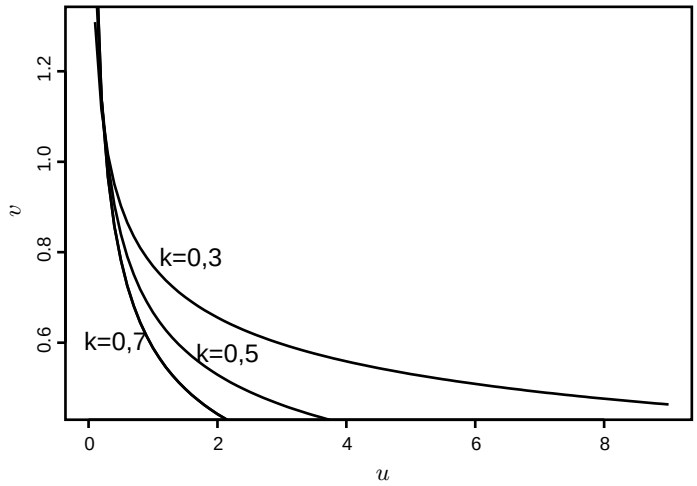

Figura 2.2 *Comportamento de v contra u para alguns valores de k da distribuição exponencial potência.*

2.2.1 Informação de Fisher

Seja $-\ddot{\mathbf{L}}_{\theta\theta}\,|_{\hat{\boldsymbol{\theta}}}$ a matriz de informação observada de Fisher para $\boldsymbol{\theta}$. Após algumas manipulações algébricas, encontramos o seguinte :

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

$$\begin{aligned} \ddot{\mathbf{L}}_{\beta\beta} &= -\frac{1}{\phi}\left\{\sum_{i=1}^{n} 2s_i\mathbf{D}_{\beta\beta}(i) + \mathbf{D}_\beta^T\mathbf{D}(\mathbf{a})\mathbf{D}_\beta\right\} \\ &= -\frac{1}{\phi}\{[2\mathbf{s}^T][\mathbf{D}_{\beta\beta}] + \mathbf{D}_\beta^T\mathbf{D}(\mathbf{a})\mathbf{D}_\beta\}, \\ \ddot{\mathbf{L}}_{\beta\phi} &= \frac{2}{\phi^2}\mathbf{D}_\beta^T\mathbf{b} \ \text{ e} \\ \ddot{\mathrm{L}}_{\phi\phi} &= \frac{1}{\phi^2}\left\{\frac{n}{2} + \mathbf{u}^T\mathbf{D}(\mathbf{c})\mathbf{u} - \frac{1}{\phi}\boldsymbol{\epsilon}^T\mathbf{D}(\mathbf{v})\boldsymbol{\epsilon}\right\}, \end{aligned}$$

sendo $\mathbf{D}_{\beta\beta}(i) = \partial^2\mu_i/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$, $\mathbf{D}(\mathbf{a}) = \mathrm{diag}\{a_1, \ldots, a_n\}$, $\mathbf{D}(\mathbf{c}) = \mathrm{diag}\{c_1, \ldots, c_n\}$, $\mathbf{b}^T = (b_1, \ldots, b_n)$, $\mathbf{u} = (u_1, \ldots, u_n)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$, $a_i = -2\{W_g(u_i) + 2u_iW_g'(u_i)\}$, $c_i = W_g'(u_i)$, $b_i = \{W_g(u_i) + u_iW_g'(u_i)\}\epsilon_i$, $\epsilon_i = y_i - \mu_i$, $s_i = W_g(u_i)\epsilon_i$, $i = 1, \ldots, n$ e a notação entre colchetes está definida em Bates e Watts (1988). No caso linear temos que $\mathbf{D}_{\beta\beta}(i) = 0$, para todo i, coincidindo com as expressões dadas em Galea, Paula e Uribe–Opazo (2003). Essas expressões simplicadas são dada por

$$\ddot{\mathbf{L}}_{\beta\beta} = -\frac{1}{\phi}\{\mathbf{X}^T\mathbf{D}(\mathbf{a})\mathbf{X}\} \quad \text{e} \quad \ddot{\mathbf{L}}_{\beta\phi} = \frac{2}{\phi^2}\mathbf{X}^T\mathbf{b}.$$

A inversa de $\ddot{\mathbf{L}}_{\theta\theta}$ para o caso não-linear pode ser expressa na forma

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

em que $\mathbf{M} = 2[\mathbf{s}^T][\mathbf{D}_{\beta\beta}] + \mathbf{D}_\beta^T\mathbf{D}(\mathbf{a})\mathbf{D}_\beta$, $\mathbf{A} = \frac{2}{\phi}\mathbf{M}^{-1}\mathbf{D}_\beta^T\mathbf{b}$ e $\mathrm{E} = \ddot{\mathbf{L}}_{\phi\phi} + \frac{2}{\phi^2}\mathbf{b}^T\mathbf{D}_\beta\mathbf{A}$. E para o caso linear, temos $\mathbf{M} = \mathbf{X}^T\mathbf{D}(\mathbf{a})\mathbf{X}$, $\mathbf{A} = \frac{2}{\phi}\mathbf{M}^{-1}\mathbf{X}^T\mathbf{b}$ e $\mathrm{E} = \ddot{\mathbf{L}}_{\phi\phi} + \frac{2}{\phi^2}\mathbf{b}^T\mathbf{X}\mathbf{A}$.

A matriz de informação de Fisher para $\boldsymbol{\theta}$ pode ser expressa na forma

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

Tabela 2.2 *Valores de d_g, f_g e ξ para algumas distribuições simétricas.*

Distribuição	d_g	f_g	ξ
Normal	$\frac{1}{4}$	$\frac{3}{4}$	1
$t-$Student	$\frac{(\nu+1)}{4(\nu+3)}$	$\frac{3(\nu+1)}{4(\nu+3)}$	$\frac{\nu}{\nu-2}, \quad \nu > 2$
$t-$Student generalizada	$\frac{r(r+1)}{4s(r+3)}$	$\frac{3(r+1)}{4(r+3)}$	$\frac{s}{r-2}, \quad s > 0, r > 2$
Logística-I	$0,369310044$	1,003445984	0,79569
Logística-II	$\frac{1}{12}$	0,60749	$\pi^2/3$
Logística generalizada	$\frac{\alpha^2 m^2}{4(2m+1)}$	$\frac{2m(2+m^2\psi'(m))}{4(2m+1)}$	$2\psi'(m)$
Exponencial potência	$\frac{\Gamma\{(3-k)/2\}}{4(2^{k-1})(1+k)^2\Gamma\{(k+1)/2\}}$	$\frac{(k+3)}{4(k+1)}$	$2^{(1+k)}\frac{\Gamma\{3(k+1)/2\}}{\Gamma\{(k+1)/2\}}$

sendo que para o caso não-linear temos $\mathbf{K}_{\beta\beta} = \frac{4d_g}{\phi}\mathbf{D}_\beta^T\mathbf{D}_\beta$ com $\mathrm{K}_{\phi\phi} = \frac{n}{4\phi^2}(4f_g - 1)$, e para o caso linear $\mathbf{K}_{\beta\beta} = \frac{4d_g}{\phi}\mathbf{X}^T\mathbf{X}$ com $f_g = \mathrm{E}\{W_g^2(U^2)U^4\}$ e $U \sim S(0,1)$ (veja Tabela 2.2). Portanto, temos ortogonalidade entre $\boldsymbol{\beta}$ e ϕ. Por exemplo, para a distribuição $t-$Student com ν graus de liberdade segue que $d_g = (\nu+1)/\{4(\nu+3)\}$ e $f_g = 3(\nu+1)/\{4(\nu+3)\}$.

Assumimos que $\boldsymbol{\beta} \in \Omega_\beta \subset I\!R^p$, em que Ω_β é um conjunto aberto com pontos interiores. É possível mostrar que $\hat{\boldsymbol{\beta}}$, o estimador de máxima verossimilhança de $\boldsymbol{\beta}$, é um estimador consistente de $\boldsymbol{\beta}$, e

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}, \mathbf{J}_{\beta\beta}^{-1}), \quad \text{em que} \quad \mathbf{J}_{\beta\beta} = \lim_{n\to\infty}\frac{1}{n}\mathbf{K}_{\beta\beta}.$$

Então, $\hat{\mathbf{K}}_{\beta\beta}^{-1} = \frac{\hat{\phi}}{4d_g}(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}$ é um estimador consistente da matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$. Observe que no caso linear a matriz de correlação assintótica não depende de parâmetros desconhecidos pois $\hat{\mathbf{K}}_{\beta\beta}^{-1} = \frac{\hat{\phi}}{4d_g}(\mathbf{X}^T\mathbf{X})^{-1}$. De forma similar o estimador de máxima verossimilhança $\hat{\phi}$ é um estimador consistente

de ϕ, e

$$\sqrt{n}(\hat{\phi}-\phi) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathrm{J}_{\phi\phi}^{-1}), \quad \text{em que} \quad \mathrm{J}_{\phi\phi} = \lim_{n\to\infty} \frac{1}{n}\mathrm{K}_{\phi\phi}.$$

Então, $\hat{\mathrm{K}}_{\phi\phi}^{-1} = \frac{4\hat{\phi}^2}{n(4f_g-1)}$ é um estimador consistente da variância assintótica de $\hat{\phi}$.

2.3 Teste de hipóteses

Hipóteses envolvendo os coeficientes $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ podem ser expressas na forma geral H: $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ contra A: $\mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$, em que $\mathbf{C}$ é uma matriz $k \times p$ de posto completo k ($k \leqslant p$) e $\mathbf{d}$ é um vetor $k \times 1$ de constantes. A hipótese nula pode contemplar situações bastante simples, como por exemplo testar H: $\beta_j = 0$ contra A: $\beta_j \neq 0$. Nestes casos a estimação dos parâmetros pode ser feita pelos processos iterativos (2.4)-(2.5) ou (2.6)-(2.7). Contudo, a hipótese nula pode envolver situações mais complexas, tais como testar H: $\beta_1 + \beta_2 + \beta_3 = 0$ contra A: $\beta_1 + \beta_2 + \beta_3 \neq 0$, as quais podem requerer o desenvolvimento de algum processo de estimação sob H. O problema aqui é maximizar o logaritmo da função de verossimilhança $\mathrm{L}(\boldsymbol{\theta})$ sujeito a restrições lineares $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$, em que $\mathbf{C} = (\mathbf{C}_1^T, \ldots, \mathbf{C}_k^T)^T$ e $\mathbf{d} = (\mathrm{d}_1, \ldots, \mathrm{d}_k)^T$. Para resolver o problema acima podemos aplicar a metodologia da função penalizada considerando, por exemplo, a função penalizada quadrática

$$\mathrm{P}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \mathrm{L}(\boldsymbol{\theta}) - \frac{1}{2}\sum_{j=1}^{k} \delta_j(\mathrm{d}_j - \mathbf{C}_j^T\boldsymbol{\beta})^2,$$

em que $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)^T$. O procedimento consiste em encontrar a solução de $\max_{\{\beta,\phi\}}\mathrm{P}(\boldsymbol{\theta}, \boldsymbol{\delta})$ para valores fixados e positivos de δ_j, $j = 1, \ldots, k$. A solução para $\boldsymbol{\beta}$ com $\boldsymbol{\delta}$ fixado será denotada por $\boldsymbol{\beta}(\boldsymbol{\delta})$. O estimador restrito de igualdades lineares é dado por

$$\hat{\boldsymbol{\beta}}^0 = \lim_{\delta_1,\ldots,\delta_k\to\infty} \boldsymbol{\beta}(\boldsymbol{\delta}).$$

Nesse sentido, desenvolvemos um processo iterativo para o caso linear ($\mathbf{D}_\beta = \mathbf{X}$) que é descrito abaixo (vide Cysneiros e Paula, 2005)

$$\begin{aligned}\boldsymbol{\beta}^{0(m+1)} &= \left\{\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{y} + \left\{\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{X}\right\}^{-1}\mathbf{C}^T \times \\ &\quad \left[\mathbf{C}\left\{\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{X}\right\}^{-1}\mathbf{C}^T\right]^{-1}\left[\mathbf{d} - \mathbf{C}\left\{\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{X}\right\}^{-1}\right. \\ &\quad \left.\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(m)})\mathbf{y}\right], \qquad (2.8)\end{aligned}$$

para $m = 0, 1, \ldots$, sendo $\phi^{(m)}$ obtido de (2.7) cuja solução será denotada por $\hat{\phi}_0$.

Se denotarmos por $\hat{\boldsymbol{\beta}}^0$ a estimativa de máxima verossimilhança sob H, segue sob certas condições de regularidade que $\hat{\boldsymbol{\beta}}^0$ é um estimador consistente de $\boldsymbol{\beta}$, e

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}, (\mathbf{J}^{\mathbf{0}}_{\beta\beta})^{-\mathbf{1}}),$$

sendo

$$\mathbf{J}^0_{\beta\beta} = \lim_{\delta_1,\ldots,\delta_k\to\infty}\left[\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left\{-\frac{\partial \mathrm{P}(\boldsymbol{\theta},\boldsymbol{\delta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\}\right]$$

e

$$\mathrm{E}\left\{-\frac{\partial \mathrm{P}(\boldsymbol{\theta},\boldsymbol{\delta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\} = \frac{4d_g}{\phi}\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{D}(\boldsymbol{\delta})\mathbf{C},$$

com $\mathbf{D}(\boldsymbol{\delta}) = \mathrm{diag}\{\delta_1, \ldots, \delta_k\}$. Então, um estimador consistente da matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}^0$ fica dado por

$$\lim_{\delta_1,\ldots,\delta_k\to\infty}\left\{\frac{4d_g}{\phi}\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{D}(\boldsymbol{\delta})\mathbf{C}\right\}^{-1} = \mathbf{K}^{-1}_{\beta\beta}\{\mathbf{I}_p - \mathbf{C}^T(\mathbf{C}\mathbf{K}^{-1}_{\beta\beta}\mathbf{C}^T)^{-1}\mathbf{C}\mathbf{K}^{-1}_{\beta\beta}\},$$

que pode ser avaliado em alguma estimativa consistente, tais como $\hat{\boldsymbol{\beta}}$ ou $\hat{\boldsymbol{\beta}}^0$.

De uma forma geral a estatística da razão de verossimilhanças para testar H: $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ contra A: $\mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$ fica dada por

$$\begin{aligned}\xi_{RV} &= 2\{\mathrm{L}(\hat{\boldsymbol{\beta}}, \hat{\phi}) - \mathrm{L}(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0)\} \\ &= 2\left[\frac{n}{2}\log\left(\frac{\hat{\phi}_0}{\hat{\phi}}\right) + \sum_{i=1}^{n}\log\left\{\frac{g\{(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2/\hat{\phi}\}}{g\{(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}^0)^2/\hat{\phi}_0\}}\right\}\right].\end{aligned}$$

Analogamente, as estatísticas de Wald e de escore podem também ser desenvolvidas, sendo dadas, respectivamente, por

$$\begin{aligned}
\xi_W &= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T \hat{\mathrm{Var}}^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) \\
&= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T (\mathbf{C}\hat{\mathbf{K}}_{\beta\beta}^{-1}\mathbf{C}^T)^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) \\
&= \frac{4d_g}{\hat{\phi}}(\mathbf{C}\hat{\boldsymbol{\beta}} - d)^T \{\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\}^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - d) \quad \text{e} \\
\xi_{SR} &= \{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\hat{\boldsymbol{\beta}}, \hat{\phi})\}^T \hat{\mathrm{Var}}_0(\hat{\boldsymbol{\beta}})\{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\hat{\boldsymbol{\beta}}, \hat{\phi})\} \\
&= \mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0)^T (\hat{\mathbf{K}}_{\beta\beta}^0)^{-1}\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0^2) \\
&= \frac{\hat{\phi}_0}{4d_g}\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0)^T (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0),
\end{aligned}$$

em que $\hat{\mathbf{K}}_{\beta\beta}$ e $\hat{\mathbf{K}}_{\beta\beta}^0$ são as matrizes de informação de Fisher avaliadas em $(\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$ e $(\hat{\boldsymbol{\beta}}^{0^T}, \hat{\phi}_0)^T$, respectivamente. Para grandes amostras tem-se sob certas condições de regularidade que a distribuição nula das três estatísticas acima é aproximadamente uma qui-quadrado central com k graus de liberdade. De maneira similar podemos desenvolver as estatísticas dos testes para testar H: $\phi = \phi_0$ contra A: $\phi \neq \phi_0$.

2.4 Modelos simétricos heteroscedásticos

Não é incomum aparecerem indícios de heteroscedasticidade através das técnicas de diagnóstico desenvolvidas para os modelos normais homocedásticos, particularmente o gráfico de resíduos contra os valores ajustados. Se comprovados esses indícios após a aplicação de algum teste apropriado (veja por exemplo, Goldfeld e Quandt, 1965; White 1980; e Breusch e Pagan, 1979), uma possibilidade é tentar ajustar algum modelo do tipo abaixo

$$y_i = \mu_i + \sqrt{\phi_i}\epsilon_i, \tag{2.9}$$

$i = 1, \ldots, n$, em que $\epsilon_i \sim S(0, 1)$. Além disso, podemos assumir que o parâmetro de dispersão ϕ_i seja parametrizado como $\phi_i = h(\tau_i)$, em que $h(\cdot)$ é uma função monótona conhecida, contínua e diferenciável e $\tau_i = \mathbf{z}_i^T\boldsymbol{\gamma}$, sendo $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})^T$ formado por valores de q variáveis explicativas e $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$. A função $h(\cdot)$

é usualmente chamada de função de ligação de dispersão e deve ser uma função positiva. Uma possível escolha seria $h(\tau) = \exp(\tau)$. As variáveis de dispersão $\mathbf{z}_i$'s não são necessariamente as mesmas variáveis de locação $\mathbf{x}_i$'s.

Pode-se mostrar que $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são parâmetros globalmente ortogonais e a matriz de informação de Fisher $\mathbf{K}_{\theta\theta}$ para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ é bloco-diagonal, isto é, $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_\beta, \mathbf{K}_\gamma\}$. As matrizes de informação de Fisher $\mathbf{K}_\beta$ e $\mathbf{K}_\gamma$ para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são dadas por $\mathbf{K}_\beta = \mathbf{X}^T\mathbf{W}_1\mathbf{X}$ e $\mathbf{K}_\gamma = \mathbf{Z}^T\mathbf{W}_2\mathbf{Z}$, respectivamente, em que $\mathbf{W}_1 = \text{diag}\{4d_g/\phi_i\}$ e $\mathbf{W}_2 = \text{diag}\{\frac{(4f_g-1)h_i'^2}{4\phi_i^2}\}$, $\mathbf{X}$ é uma matriz $n \times p$ de linhas $\mathbf{x}_i^T$ e $\mathbf{Z}$ é uma matriz $n \times q$ de linhas $\mathbf{z}_i^T$, para $i = 1, \ldots, n$.

Um processo iterativo para obter as estimativas de máxima verossimilhança de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ pode ser desenvolvido usando, por exemplo, o método *scoring* de Fisher (vide Cysneiros, 2004), que nos leva ao seguinte sistema de equações:

$$\mathbf{X}^T\mathbf{W}_1^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T\mathbf{W}_1^{(m)}\mathbf{z}_\beta^{(m)} \qquad \text{e} \qquad \mathbf{Z}^T\mathbf{W}_2^{(m)}\mathbf{Z}\boldsymbol{\gamma}^{(m+1)} = \mathbf{Z}^T\mathbf{W}_2^{(m)}\mathbf{z}_\gamma^{(m)},$$

em que $\mathbf{z}_\beta$ e $\mathbf{z}_\gamma$ são vetores $n \times 1$ cujas componentes tomam as formas

$$z_{\beta_i} = \mu_i + \frac{v_i}{4d_g}(y_i - \mu_i) \qquad \text{e} \qquad z_{\gamma_i} = \tau_i + \frac{2\phi_i}{(4f_g - 1)h_i'}(v_i u_i - 1),$$

para $i = 1, \ldots, n$. Assim, sob certas condições de regularidade e para amostras grandes os estimadores de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ têm distribuição aproximadamente normal de médias $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ e matrizes de variância-covariância dadas por $(\mathbf{X}^T\mathbf{W}_1\mathbf{X})^{-1}$ e $(\mathbf{Z}^T\mathbf{W}_2\mathbf{Z})^{-1}$, respectivamente.

2.5 Métodos de diagnóstico

2.5.1 Resíduos

Uma pergunta comum após o ajuste de um modelo proposto é a seguinte : "será que o modelo se ajusta bem aos dados ?" É importante responder a essa pergunta pois se o modelo não estiver bem ajustado, o mesmo pode fornecer conclusões errôneas. Uma técnica que pode ajudar a responder esta pergunta é a análise de resíduos. Esta técnica verifica, por exemplo, se há afastamentos sérios

das suposições feitas para os erros e se existem observações aberrantes. Uma definição natural de resíduo é a diferença entre a resposta observada e o valor predito, denominado resíduo ordinário. É importante conhecer algumas propriedades desse resíduo. Nesse sentido, podemos utilizar a metodologia apresentada em Cox e Snell (1968) para determinar os momentos do resíduo ordinário em modelos simétricos. Consideraremos o resíduo ordinário com ϕ conhecido ou fixo, expresso na forma abaixo

$$r_i(y_i, \hat{\mu}_i, \phi) = y_i - \hat{\mu}_i, \tag{2.10}$$

$i = 1, \ldots, n$, em que $\mu_i = \mu(\mathbf{x}_i, \boldsymbol{\beta})$, $y_i = \mu_i + \epsilon_i$ e $\epsilon_i \sim S(0, \phi)$.

Esses resíduos são, em geral, viesados e têm distribuição não normal, mesmo assintoticamente, dificultando a verificação da adequacidade dos modelos pelos métodos tradicionais. Em modelos de regressão normais não-lineares Cook e Tsai (1985) propuseram o resíduo projetado obtido num sub-espaço dos resíduos ordinários. Esses novos resíduos têm distribuição aproximadamente normal de média zero e variância dependendo de σ^2. Contudo, árduas álgebras podem ser necessárias para obter tais resíduos.

Paula, Cysneiros e Galea (2003) propõem corrigir, até ordem n^{-1}, os dois primeiros momentos de r_i a fim de obter propriedades próximas às do $i-$ésimo erro $\epsilon_i = y_i - \mu_i$. Após algumas manipulações algébricas obtemos

$$\mathrm{E}(r_i) = -\mathbf{d}_i^T (\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_\beta^T \boldsymbol{\eta} + \eta_i, \tag{2.11}$$

em que $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, $\eta_i = -\dfrac{\phi}{8d_g} \mathrm{tr}\{(\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_{\beta\beta}(i)\}$ e $\mathbf{d}_i = (d_{i1}, \ldots, d_{ip})^T$ com $d_{ij} = \partial \mu_i / \partial \beta_j$. Conseqüentemente, em forma matricial

$$\mathrm{E}(\mathbf{r}) = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\eta}, \tag{2.12}$$

em que $\mathbf{H} = \mathbf{D}_\beta (\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_\beta^T$ e $\mathbf{I}_n$ é a matriz identidade de ordem n, generalizando as expressões dadas em Cook, Tsai e Wei (1986) que encontraram essa relação para os modelos normais não-lineares. Fazendo desenvolvimento similar para $\mathrm{E}(r_i^2)$ e $\mathrm{E}(r_i, r_j)$ obtemos

$$\mathrm{Var}(r_i) \;=\; \phi\xi\{1 - (4d_g\xi)^{-1} h_{ii}\}, \tag{2.13}$$

e

$$\mathrm{Cov}(r_i, r_j) \;=\; -\phi\xi(4d_g\xi)^{-1}h_{ij}, \quad i \neq j, \tag{2.14}$$

em que $h_{ij} = \mathbf{d}_i^T(\mathbf{D}_\beta^T\mathbf{D}_\beta)^{-1}\mathbf{d}_j$. Portanto, em notação matricial temos que a matriz de variância-covariância aproximada do vetor de resíduos ordinários pode ser expressa na forma

$$\mathrm{Var}(\mathbf{r}) = \phi\xi\{\mathbf{I}_n - (4d_g\xi)^{-1}\mathbf{H}\}. \tag{2.15}$$

No caso em que podemos estabelecer a relação linear, $\mu_i = \mathbf{x}_i^T\boldsymbol{\beta}$, encontramos simplicações interessantes nas expressões acima. Devido ao fato de que o viés de ordem n^{-1} de $\hat{\boldsymbol{\beta}}$ é nulo quando temos um relação linear nos parâmetros, segue o seguinte :

$$\mathrm{E}(\mathbf{r}) = \mathbf{0} \qquad \text{e} \qquad \mathrm{Var}(\mathbf{r}) = \phi\xi\{\mathbf{I}_n - (4d_g\xi)^{-1}\mathbf{H}\},$$

em que $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Como os r_i's têm variâncias diferentes, é conveniente expressá-los em forma padronizada, a fim de permitir uma comparabilidade entre os mesmos. Uma definição natural do resíduo padronizado é subtrair pela média e dividir pelo respectivo desvio-padrão, obtendo-se a expressão abaixo

$$t_{r_i} \;=\; \frac{y_i - \hat{y}_i}{\{\xi\hat{\phi}\}^{1/2}\{1 - (4d_g\xi)^{-1}\hat{h}_{ii}\}^{1/2}}, \qquad i = 1, \ldots, n. \tag{2.16}$$

Estudos de simulação desenvolvidos pelos autores indicam que o resíduo proposto acima tem média e variância aproximadamente zero e um, respectivamente, uma assimetria desprezível e uma curtose acompanhando a curtose da distribuição do erro. Portanto, muitas das técnicas usuais de análise de resíduos desenvolvidas para o caso normal linear podem ser estendidas para os modelos simétricos. Em particular, pode-se gerar bandas empíricas de confiança através do modelo ajustado para o resíduo t_{r_i}, também conhecidas como envelope (Atkinson, 1981, 1985). Tais bandas podem auxiliar na avaliação da simetria para os erros, adequação das caudas, presença de observações aberrantes e adequação da relação funcional entre a média e os parâmetros da regressão.

2.5.2 Influência local

A idéia principal de influência local é verificar, através de alguma medida apropriada de influência, o efeito de pequenas perturbações no modelo ou nos dados nos principais resultados do ajuste. Se essas perturbações causarem efeitos desproporcionais em determinados resultados podem ser indícios de que o modelo está mal ajustado ou que existem afastamentos importantes das suposições feitas para o mesmo. A identificação das observações responsáveis por essas discrepâncias pode ajudar na escolha de um modelo mais adequado. A medida de influência mais conhecida é o afastamento da verossimilhança $\mathrm{LD}(\boldsymbol{\omega}) = 2\{\mathrm{L}(\hat{\boldsymbol{\theta}}) - \mathrm{L}(\hat{\boldsymbol{\theta}}_\omega)\}$, em que $\hat{\boldsymbol{\theta}}_\omega$ denota a estimativa de máxima verossimilhança sob o modelo perturbado e $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_s)^T$ é o vetor de perturbações aplicadas no modelo. A proposta de Cook (1986) é estudar o comportamento de $\mathrm{LD}(\boldsymbol{\omega})$, ou de alguma outra medida de influência, em torno do vetor de não-perturbação $\boldsymbol{\omega}_0$. Tem-se que $\mathrm{L}(\hat{\boldsymbol{\theta}}_{\omega_0}) = \mathrm{L}(\hat{\boldsymbol{\theta}})$ e conseqüentemente $\mathrm{LD}(\boldsymbol{\omega}_0) = 0$. Logo, desde que $\mathrm{LD}(\boldsymbol{\omega}) \geqslant 0$, $\boldsymbol{\omega}_0$ é um ponto de mínimo da função $\mathrm{LD}(\boldsymbol{\omega})$. A sugestão de Cook (1986) é investigar a curvatura normal da linha projetada $\mathrm{LD}(\omega_0 + a\boldsymbol{\ell})$, em que $a \in I\!R$, em torno de $a = 0$ para alguma direção arbitrária $\boldsymbol{\ell}$, $||\boldsymbol{\ell}|| = 1$. Mostra-se que a curvatura normal pode ser expressa numa forma geral $\mathrm{C}_\ell(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^T\boldsymbol{\Delta}^T\ddot{\mathbf{L}}_{\theta\theta}^{-1}\boldsymbol{\Delta}\boldsymbol{\ell}|$, em que $\boldsymbol{\Delta}$ é uma matriz $(p+q) \times s$ com elementos $\Delta_{ij} = \partial^2\mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\theta_i\partial\omega_j$, $i = 1, \ldots, p+q$ e $j = 1, \ldots, s$, com todas as quantidades sendo avaliadas em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ e $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Cook sugere tomar a direção correspondente à maior curvatura, denotada por $\boldsymbol{\ell}_{max}$, o maior autovetor correspondente ao maior autovalor da matriz $\mathbf{B} = -\boldsymbol{\Delta}^T\ddot{\mathbf{L}}_{\theta\theta}^{-1}\boldsymbol{\Delta}$. O gráfico de índices de $\boldsymbol{\ell}_{max}$ pode mostrar como se deve perturbar, por exemplo, o parâmetro de escala para obter maiores mudanças nas estimativas de $\boldsymbol{\theta}$. Contudo, se o interesse é somente no vetor $\boldsymbol{\beta}$, a curvatura normal na direção $\boldsymbol{\ell}$ é dada por $\mathrm{C}_{\boldsymbol{\ell}}(\boldsymbol{\beta}) = 2|\boldsymbol{\ell}^T\boldsymbol{\Delta}^T(\ddot{\mathbf{L}}_{\theta\theta}^{-1} - \mathbf{L}_1)\boldsymbol{\Delta}\boldsymbol{\ell}|$ (veja Cook, 1986), em que

$$\mathbf{L}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{\phi\phi}^{-1} \end{bmatrix},$$

com $-\ddot{\mathbf{L}}_{\phi\phi}\mid_{\hat{\theta}}$ sendo a informação observada de Fisher para ϕ. O gráfico de índices do maior autovetor de $\boldsymbol{\Delta}^T(\ddot{\mathbf{L}}_{\theta\theta}^{-1} - \mathbf{L}_1)\boldsymbol{\Delta}$ pode revelar quais observações são influentes em $\hat{\boldsymbol{\beta}}$. Similarmente, a curvatura normal para o parâmetro de escala ϕ na direção $\boldsymbol{\ell}$ é dada por $\mathrm{C}_{\boldsymbol{\ell}}(\phi) = 2|\boldsymbol{\ell}^T\boldsymbol{\Delta}^T(\ddot{\mathbf{L}}_{\theta\theta}^{-1} - \mathbf{L}_2)\boldsymbol{\Delta}\boldsymbol{\ell}|$, em que

$$\mathbf{L}_2 = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

com $-\ddot{\mathbf{L}}_{\beta\beta}\mid_{\hat{\theta}}$ sendo a matriz de informação observada de Fisher para $\boldsymbol{\beta}$. A influência local das observações em $\hat{\phi}$ pode ser avaliada considerando-se o gráfico de índices de $\boldsymbol{\ell}_{max}$ para a matriz $|\boldsymbol{\Delta}^T(\ddot{\mathbf{L}}_{\theta\theta}^{-1} - \mathbf{L}_2)\boldsymbol{\Delta}|$.

Escobar e Meeker (1992) sugerem tomar como medida de influência os elementos da diagonal principal da matriz $\mathbf{B} = -\boldsymbol{\Delta}^T\ddot{\mathbf{L}}_{\theta\theta}^{-1}\boldsymbol{\Delta}$, enquanto Lesaffre e Verbeke (1998) sugerem avaliar a curvatura normal na direção da $i-$ésima observação, que consiste na avaliação de $\mathrm{C}_\ell(\boldsymbol{\theta})$ no vetor $(n \times 1)$ $\boldsymbol{\ell}_i$ formado por zeros com um na $i-$ésima posição. Esta curvatura, denotada por C_i ou $\mathrm{C}_i(\theta)$ que é igual a $2|b_{ii}|$. É sugerido que as observações tais que $\mathrm{C}_i > 2\bar{\mathrm{C}}$ tenham uma atenção especial.

Em particular, fazendo uma perturbação aditiva no $i-$ésimo valor da resposta, $y_{i\omega} = y_i + \sigma\omega_i$, em que $\omega_i \in I\!R$ e σ é o desvio padrão de y_i, podemos considerar a mudança instantânea no $i-$ésimo valor predito (quando $\omega_i \to 0$) como uma medida de influência da $i-$ésima observação no seu próprio valor predito. Podemos citar outros esquemas de perturbação de interesse, como por exemplo :

- supor que se deseja verificar a possibilidade das respostas possuírem variâncias distintas, isto é, $\mathrm{Var}(y_i) = \xi\phi/\omega_i$, ou seja, a possibilidade de termos um modelo heteroscedástico;
- interesse em perturbar a t-ésima variável explicativa, com $(x_{i1}, \ldots, x_{it}+s_t\omega_i, \ldots x_{ip})$, em que s_t é um fator de escala, que pode ser a norma da $t-$ésima coluna da matriz $\mathbf{X}$.

É possível perturbar o modelo proposto de diversas outras maneiras, porém é importante escolher esquemas de perturbação e medidas de influência que permitam interpretações fáceis. Galea, Bolfarine e Vilca–Labra (2002) estudam influência lo-

cal nos modelos com erros nas variáveis sob a distribuição $t-$Student, enquanto Galea, Paula e Bolfarine (1997) e Galea, Paula e Uribe–Opazo (2003) investigam a influência das observações nas estimativas dos parâmetros usando o enfoque de influência local na classe dos modelos simétricos lineares.

Perturbação na escala

Considere agora o modelo heteroscedástico

$$f_{y_i}(y_i|\omega_i) = \sqrt{\frac{\omega_i}{\phi}} g(\omega_i u_i), \tag{2.17}$$

em que ω_i denota o peso correspondente ao $i-$ésimo caso, $i = 1, \ldots, n$. Quando $\omega_i = 1$, o modelo perturbado (2.17) reduz-se ao modelo postulado (2.2). Além disso, estamos perturbando o parâmetro de escala pela mudança do seu valor para ϕ/ω_i para a $i-$ésima observação. A matriz $(p+1) \times n$ $\mathbf{\Delta}$ fica neste caso dada por

$$\mathbf{\Delta} = \begin{pmatrix} -2\hat{\phi}^{-1}\mathbf{D}(\hat{\mathbf{b}})\mathbf{D}_{\hat{\beta}} \\ -\hat{\phi}^{-2}\mathbf{D}(\hat{\mathbf{b}})\mathbf{r} \end{pmatrix},$$

em que $\mathbf{D}(\hat{\mathbf{b}}) = \text{diag}\{\hat{b}_1, \ldots, \hat{b}_n\}$ e $\mathbf{r} = (r_1, \ldots, r_n)^T$ com $r_i = y_i - \hat{y}_i$.

Perturbação de casos

Considere o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$ expresso na forma

$$\mathbf{L}(\boldsymbol{\theta}|\omega_i) = \sum_{i=1}^{n} \omega_i \log\left\{\frac{g(u_i)}{\sqrt{\phi}}\right\}, \tag{2.18}$$

em que $0 \leqslant \omega_i \leqslant 1$. Sob esse esquema de perturbação a matriz $\mathbf{\Delta}$ assume a forma

$$\mathbf{\Delta} = \begin{pmatrix} \hat{\phi}^{-1}\mathbf{D}(\hat{\mathbf{v}})\mathbf{D}(\mathbf{r})\mathbf{D}_{\hat{\beta}} \\ -2\hat{\phi}\hat{\mathbf{s}} \end{pmatrix},$$

em que $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \ldots, r_n\}$, $\mathbf{D}(\hat{\mathbf{v}}) = \text{diag}\{\hat{v}_1, \ldots, \hat{v}_n\}$ e $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_n)^T$ com $\hat{s}_i = 1 - \hat{v}_i\hat{u}_i$.

2.5.3 Influência local na predição

Seja $\mathbf{q}$ um vetor $p \times 1$ de valores das variáveis explanatórias no modelo simétrico linear, para o qual não temos necessariamente uma resposta observada. Então,

a predição em $\mathbf{q}$ é dada por $\hat{\mu}(\mathbf{q}) = \sum_{j=1}^{p} q_j \hat{\beta}_j$. Analogamente, o ponto predito em $\mathbf{q}$ baseado no modelo perturbado é dado por $\hat{\mu}(\mathbf{q}, \omega) = \sum_{j=1}^{p} q_j \hat{\beta}_{jw}$, em que $\hat{\boldsymbol{\beta}}_\omega = (\hat{\beta}_{1\omega}, \ldots, \hat{\beta}_{p\omega})^T$ denota a estimativa de máxima verossimilhança do modelo perturbado. Thomas e Cook (1990) têm investigado o efeito de pequenas perturbações na predição em algum particular ponto $\mathbf{q}$ em modelos lineares generalizados contínuos assumindo ϕ conhecido ou estimado separadamente de $\hat{\boldsymbol{\beta}}$. Contudo, como não é tão claro definir o afastamento da verossimilhança para predições para as quais não se tem nenhuma resposta observada, três funções objetivo baseadas em diferentes resíduos foram definidas. A função objetivo $f(\mathbf{q}, \boldsymbol{\omega}) = \{\hat{\mu}(\mathbf{q}) - \hat{\mu}(\mathbf{q}, \boldsymbol{\omega})\}^2$ tem sido escolhida devido à simplicidade e invariância com respeito a outras medidas de influência.

Similarmente, concentraremos nossos estudos na investigação da curvatura normal na superfície formada pelo vetor $\boldsymbol{\omega}$ e a função $f(\mathbf{q}, \boldsymbol{\omega})$ em torno de $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, em que $\boldsymbol{\omega}_0$ é tal que $\hat{\boldsymbol{\beta}}_{\omega_0} = \hat{\boldsymbol{\beta}}$. A curvatura normal na direção unitária $\boldsymbol{\ell}$ assume, neste caso, a forma $\mathrm{C}_{\boldsymbol{\ell}} = |\boldsymbol{\ell}^T \ddot{\mathbf{f}} \boldsymbol{\ell}|$, em que $\ddot{\mathbf{f}} = \partial^2 f / \partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T$ é avaliada em $\boldsymbol{\omega}_0$ e $\hat{\boldsymbol{\beta}}$. Seguindo Thomas e Cook (1990), obtemos

$$\ddot{\mathbf{f}} = -2\boldsymbol{\Delta}^T (\ddot{\mathbf{L}}_{\beta\beta}^{-1} \mathbf{q}\mathbf{q}^T \mathbf{L}_{\beta\beta}^{-1}) \boldsymbol{\Delta},$$

em que $\boldsymbol{\Delta} = \partial^2 \mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T$ é avaliado em $(\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$. Conseqüentemente,

$$\boldsymbol{\ell}_{max}(\mathbf{q}) \propto \boldsymbol{\Delta}^T \ddot{\mathbf{L}}_{\beta\beta}^{-1} \mathbf{q}.$$

A seguir descrevemos a matriz $\boldsymbol{\Delta}$ e $\boldsymbol{\ell}_{max}(\mathbf{q})$ sob dois esquemas de perturbação, a perturbação aditiva na resposta e em cada variável explanatória.

Suponha inicialmente que a $i-$ésima resposta é perturbada tal que $y_{i\omega} = y_i + \omega_i,\ \omega_i \in I\!R$. Mostra-se neste caso que a matriz $\boldsymbol{\Delta}$ pode ser expressa na forma

$$\boldsymbol{\Delta} = \hat{\phi}^{-1} \mathbf{X}^T \mathbf{D}(\hat{\mathbf{a}}),$$

em que $\mathbf{D}(\hat{\mathbf{a}}) = \mathrm{diag}\{\hat{a}_1, \ldots \hat{a}_n\}$. O vetor $\boldsymbol{\ell}_{max}(\mathbf{z})$ é avaliado aqui em $\mathbf{z} = \mathbf{x}_i$, ficando dado por

$$\boldsymbol{\ell}_{max}(\mathbf{x}_i) \propto \mathbf{D}(\hat{\mathbf{a}}) \mathbf{X} \{\mathbf{X}^T \mathbf{D}(\hat{\mathbf{a}}) \mathbf{X}\}^{-1} \mathbf{x}_i.$$

Assim, um valor alto para o $i-$ésimo componente de $\boldsymbol{\ell}_{max}(\mathbf{x}_i)$ indica uma influência local desproporcional da $i-$ésima observação no próprio valor ajustado. Isto sugere considerar o gráfico de índices de $\boldsymbol{\ell}_{max_i}(\mathbf{x}_i)$.

Similarmente podemos perturbar os valores da $t-$ésima variável explicativa, assumindo que a mesma seja contínua, tal que $x_{it\omega} = x_{it} + \omega_i$, $\omega_i \in I\!R$. Mostra-se que a matriz $\boldsymbol{\Delta}$ assume neste caso a forma abaixo

$$\boldsymbol{\Delta} = \hat{\phi}^{-1}\{\mathbf{F}\mathbf{D}(\mathbf{r})\mathbf{D}(\hat{\mathbf{v}}) - \hat{\beta}_t\mathbf{X}^T\mathbf{D}(\hat{\mathbf{a}})\},$$

em que $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \ldots, r_n\}$, $\mathbf{D}(\hat{\mathbf{v}}) = \text{diag}\{\hat{v}_1, \ldots, \hat{v}_n\}$ e $\mathbf{F}$ é uma matriz $p \times n$ de zeros com uns na $t-$ésima linha. Similarmente ao caso anterior de perturbação na resposta, $\boldsymbol{\ell}_{max}(\mathbf{z})$ deve ser avaliado em $\mathbf{z} = \mathbf{x}_i$, ficando expresso na forma

$$\boldsymbol{\ell}_{max}(\mathbf{x}_i) \propto \{\mathbf{F}\mathbf{D}(\mathbf{r}) - \hat{\beta}_t\mathbf{X}^T\mathbf{D}(\hat{\mathbf{a}})\}\{\mathbf{X}^T\mathbf{D}(\hat{\mathbf{a}})\mathbf{X}\}^{-1}\mathbf{x}_i.$$

Novamente, o gráfico de índices de $\boldsymbol{\ell}_{max_i}(\mathbf{x}_i)$ pode revelar quais observações exercem influência desproporcional na própria predição sob pequenas perturbações no valor da $t-$ésima variável explicativa.

2.5.4 Ponto de alavanca generalizado

Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ o vetor de respostas observadas as quais têm função de probabilidade de densidade $f(\mathbf{y}; \boldsymbol{\theta})$, sendo $\boldsymbol{\theta}$ um vetor q-dimensional. Se denotarmos por $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\mathbf{y})$ a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ e por $\boldsymbol{\mu}$ o vetor de valores esperados, então $\hat{\mathbf{y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ é o vetor de respostas preditas. A principal idéia por trás do conceito de ponto de alavanca (veja, por exemplo, Hoaglin e Welsch, 1978; Cook e Weisberg, 1982; Emerson, Hoaglin e Kempthorne, 1984; St. Laurent e Cook, 1992 e Wei, Hu e Fung, 1998) é conhecer a influência de y_i no próprio valor predito. Esta influência pode ser bem representada pela derivada $\partial\hat{y}_i/\partial y_i$ que é igual a h_{ii} no caso normal linear, em que h_{ii} é o $i-$ésimo elemento da diagonal principal da matriz de projeção $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ e $\mathbf{X}$ é a matriz modelo. Extensões para modelos de regressão mais gerais têm sido propostas, por exemplo, por St. Laurent e Cook (1992) e Wei, Hu e Fung, (1998) quando $\boldsymbol{\theta}$ é irrestrito e por Paula (1993,1995,1999) quando $\boldsymbol{\theta}$ é restrito em desigualdades lineares.

Em particular, se denotarmos por $\mathrm{L}(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança de $\boldsymbol{\theta} \in I\!R^q$ e por $\hat{\boldsymbol{\theta}}(\mathbf{y})$ a estimativa que maximiza $\mathrm{L}(\boldsymbol{\theta})$, segue de Wei, Hu e Fung (1998) que a matriz $(n \times n)$ $(\partial\hat{\mathbf{y}}/\partial\mathbf{y}^T)$ de pontos de alavanca pode ser expressa na forma

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \{(\mathbf{D}_\theta)(-\ddot{\mathbf{L}}_{\theta\theta})^{-1}(\ddot{\mathbf{L}}_{\theta\mathbf{y}})\} \mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(\mathbf{y})}}, \tag{2.19}$$

em que $\mathbf{D}_\theta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^T$, $\ddot{\mathbf{L}}_{\theta\theta} = \partial^2\mathrm{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ e $\ddot{\mathbf{L}}_{\theta\mathbf{y}} = \partial^2\mathrm{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{y}^T$. A expressão (2.19) generaliza a definição de pontos de alavanca generalizados dada em St. Laurent e Cook (1992).

Sendo $\mathbf{D}_\theta = (\mathbf{D}_\beta, \mathbf{0})$, e desde que

$$\ddot{\mathbf{L}}_{\beta y} = \frac{1}{\phi}\mathbf{D}_\beta^T\mathbf{D}(\mathbf{a}) \;\text{ e }\; \ddot{\mathbf{L}}_{\phi y} = -\frac{2}{\phi^2}\mathbf{b}^T,$$

então usando a expressão (2.19) a matriz generalizada de pontos de alavanca na classe de modelos simétricos não-lineares assume a forma

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \mathbf{GL}_\beta(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_\phi(\hat{\boldsymbol{\theta}}) \tag{2.20}$$

com

$$\mathbf{GL}_\beta(\hat{\boldsymbol{\theta}}) = \mathbf{D}_{\hat{\beta}}\hat{\mathbf{M}}^{-1}\mathbf{D}_{\hat{\beta}}^T\mathbf{D}(\hat{\mathbf{a}}) \quad \text{e}$$

$$\mathbf{GL}_\phi(\hat{\boldsymbol{\theta}}) = \frac{4}{\hat{\mathrm{E}}\hat{\phi}^3}\mathbf{D}_{\hat{\beta}}\hat{\mathbf{M}}^{-1}\mathbf{D}_{\hat{\beta}}^T\hat{\mathbf{b}}\hat{\mathbf{b}}^T\{\mathbf{I}_n - \mathbf{GL}_\beta(\hat{\boldsymbol{\theta}})\}$$

com $\hat{\mathbf{M}} = \mathbf{D}_{\hat{\beta}}^T\mathbf{D}(\hat{\mathbf{a}})\mathbf{D}_{\hat{\beta}} + 2[\hat{\mathbf{s}}^T][\mathbf{D}_{\hat{\beta}\hat{\beta}}]$, $\mathbf{I}_n$ sendo a matriz identidade de ordem n, $\mathbf{D}(\mathbf{a})$ e E definidos na Seção 2.2.1. Uma interpretação interessante para (2.20) pode ser obtida se considerarmos o procedimento de estimação de mínimos quadrados ao invés de máxima verossimilhança, considerando a função objetivo

$$Q(\boldsymbol{\beta}) = \frac{1}{2\sigma^2}\sum_{i=1}^{n} a_i\{y_i - \mu_i(\boldsymbol{\beta})\}^2,$$

em que $\mathrm{Var}(y_i) = \frac{\sigma^2}{a_i}$ e os a_i's são constantes positivas. Então, usando a expressão geral (2.2) de Wei, Hu e Fung (1998) encontramos $\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \mathbf{GL}_\beta(\hat{\boldsymbol{\theta}})$ com $s_i = -a_i e_i$. Isto é, o procedimento de mínimos quadrados leva em conta somente a influência da estimativa do parâmetro de locação na medida de alavanca, enquanto

o de máxima verossimilhança também tende a considerar a influência da estimativa do parâmetro de escala. Quando o parâmetro de dispersão ϕ é conhecido é fácil mostrar que $\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \mathbf{GL}_{\beta}(\hat{\boldsymbol{\theta}})$. Contudo, para o caso normal, desde que $\mathbf{D}_{\hat{\beta}}^T\hat{\mathbf{b}} = \mathbf{0}$ a influência de $\hat{\phi}$ na matriz generalizada de pontos de alavanca anula-se e $\mathbf{GL}(\hat{\boldsymbol{\theta}})$ reduz-se à matriz Jacobiana de pontos de alavanca

$$\hat{\mathbf{J}} = \mathbf{D}_{\hat{\beta}} \left\{ \mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}} - [\mathbf{r}^T][\mathbf{D}_{\hat{\beta}\hat{\beta}}] \right\}^{-1} \mathbf{D}_{\hat{\beta}}^T. \tag{2.21}$$

St. Laurent and Cook (1992) comparam (2.21) com a matriz de pontos de alavanca do plano tangente definida por $\hat{\mathbf{H}} = \mathbf{D}_{\hat{\beta}}(\mathbf{D}_{\hat{\beta}}^T\mathbf{D}_{\hat{\beta}})^{-1}\mathbf{D}_{\hat{\beta}}^T$, que é a matriz de projeção ortogonal no subespaço gerado pelas colunas da matriz $\mathbf{D}_{\hat{\beta}}$. Neste caso, seguem as propriedades $0 \leqslant \hat{h}_{ii} \leqslant 1$, $\sum \hat{h}_{ii} = p$ e que $\hat{h}_{kk} = 1$ implica em $\hat{h}_{ik} = 0$ para $i \neq k$. Essas propriedades não são garantidas para $\hat{j}_{ii}$, o $i-$ésimo elemento da diagonal de $\hat{\mathbf{J}}$. Podemos ter, por exemplo, $\hat{j}_{ii} > 1$ chamado superalavanca.

Caso linear homocedástico

Considere agora o caso linear homocedástico em que $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \epsilon_i$ e seja $\mathbf{X}$ a matriz modelo com linhas $\mathbf{x}_i^T$, $i = 1, \ldots, n$. Segue que $\mathbf{D}_{\beta} = \mathbf{X}$ e $\mathbf{D}_{\beta\beta} = \mathbf{0}$ de modo que a matriz generalizada de pontos de alavanca assume uma forma simplificada

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{H}} + \frac{4}{\hat{\mathrm{E}}\hat{\phi}^3}\hat{\mathbf{H}}\mathbf{D}^{-1}(\hat{\mathbf{a}})\hat{\mathbf{b}}\hat{\mathbf{b}}^T\{\mathbf{I}_n - \hat{\mathbf{H}}\},$$

em que $\hat{\mathbf{H}} = \mathbf{X}\{\mathbf{X}^T\mathbf{D}(\hat{\mathbf{a}})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{D}(\hat{\mathbf{a}})$. Entretanto, se os $a_i's$ são constantes positivas, $\hat{\mathbf{H}}$ pode ser interpretada como a matriz de projeção ortogonal em $C(\mathbf{X}\mathbf{D}^{1/2}(\hat{\mathbf{a}}))$, que denota o subespaço gerado pelas colunas da matriz $\mathbf{X}\mathbf{D}^{1/2}(\hat{\mathbf{a}})$. Quando $a_i = 1$, $\forall i$, tem-se $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Relação entre influência local e alavanca generalizada

Usando o esquema de perturbação aditiva na resposta e a medida de influência $\mathbf{LD}(\boldsymbol{\omega})$ encontramos que $\boldsymbol{\Delta}^T = [(1/\hat{\phi})\mathbf{D}(\hat{\mathbf{a}})\mathbf{D}_{\hat{\beta}}, -(2/\hat{\phi}^2)\hat{\mathbf{b}}]$. Então, podemos expressar

$$\mathbf{B} = \frac{1}{\hat{\phi}}\left[\mathbf{D}(\hat{\mathbf{a}})\mathbf{GL}(\hat{\boldsymbol{\theta}}) + \frac{4\hat{\mathbf{b}}\hat{\mathbf{b}}^T}{\hat{\phi}^3\hat{\mathrm{E}}}\left\{\mathbf{I}_n - \mathbf{GL}_{\beta}(\hat{\boldsymbol{\theta}})\right\}\right].$$

Em particular, quando ϕ é fixado, a matriz generalizada de pontos de alavanca $\mathbf{GL}(\hat{\boldsymbol{\theta}})$ reduz-se a

$$\begin{aligned}
\mathbf{GL}(\hat{\boldsymbol{\theta}}) &= -\mathbf{D}_{\hat{\beta}}\ddot{\mathbf{L}}^{-1}_{\hat{\beta}\hat{\beta}}\mathbf{D}^T_{\hat{\beta}}\mathbf{D}(\hat{\mathbf{a}}), \quad \text{e} \\
\mathbf{B} &= -\boldsymbol{\Delta}^T\ddot{\mathbf{L}}^{-1}_{\hat{\beta}\hat{\beta}}\boldsymbol{\Delta} \\
&= -\frac{1}{\hat{\phi}^2}\mathbf{D}(\hat{\mathbf{a}})\mathbf{D}_{\hat{\beta}}\ddot{\mathbf{L}}^{-1}_{\hat{\beta}\hat{\beta}}\mathbf{D}^T_{\hat{\beta}}\mathbf{D}(\hat{\mathbf{a}}) \\
&= \frac{1}{\hat{\phi}}\mathbf{D}(\hat{\mathbf{a}})\mathbf{GL}(\hat{\boldsymbol{\theta}}).
\end{aligned}$$

Neste caso, a medida de influência C_i assume a forma simples

$$\mathrm{C}_i = \frac{2|\hat{a}_i|}{\hat{\phi}}\mathrm{GL}_{ii}(\hat{\boldsymbol{\theta}}), \tag{2.22}$$

em que $a_i = -2\{W_g(u_i)+2u_iW'_g(u_i)\}$. Então, pela Tabela 2.1 temos que $a_i = 1$ para o caso normal e $a_i = (\nu+1)(\nu-3u_i)/(\nu+u_i)^2$ para a distribuição $t-$Student com ν graus de liberdade. A expressão (2.22) pode ser usada para avaliar a influência local total da $i-$ésima observação na estimativa $\hat{\boldsymbol{\beta}}$.

CAPÍTULO 3

Aplicações

É importante que toda a teoria desenvolvida aqui esteja disponível em algum software. Com esta preocupação foram desenvolvidos macros nos softwares S-Plus (Becker, Chambers e Wilks, 1988 e Chambers e Hastie, 1992) e R (Ihaka e Gentleman, 1996). O S-Plus consiste de uma linguagem de programação orientada a objeto baseada na linguagem S, permitindo o uso interativo de comandos bem como a implementação de funções em C e Fortran. O aplicativo possui também um conjunto de ferramentas estatísticas para a análise de dados. Sua versão comercial pode ser encontrada no site `www.insightful.com`. Similar ao S-Plus, o R é um ambiente para computação estatística e análise de dados. Devido ao seu código fonte aberto, o mesmo tem recebido inúmeras contribuições de várias comunidades científicas. O R encontra-se disponível em `www.r-project.org`, bem como diversos macros, que são implementações das mais variadades áreas de estudo.

Com o pensamento de difundir a modelagem estatística com erros simétricos, desenvolvemos a `library elliptical`, que consiste em um conjunto de rotinas computacionais que permite a definição de distribuições pertencentes à classe simétrica, ajuste dos parâmetros de locação e escala pelo método de máxima verossimilhança para modelos lineares e não-lineares simétricos, cálculos das medidas de curvatura intrínseca e paramétrica dos modelos não-lineares simétricos, cálculo de resíduos, cálculo de medidas de diagnóstico bem como a construção de gráficos e ainda a geração de dados de algumas distribuições simétricas. É possível também fazer o ajuste dos parâmetros de locação e escala pelo método da máxima verossimilhança para modelos lineares e não-lineares simétricos restritos a igualdades lineares. Este conjunto de rotinas encontra-se disponível gratuitamente para uso acadêmico em `http://www.de.ufpe.br/~cysneiros/elliptical/elliptical.html`.

Inicialmente, vamos apresentar a sintaxe do comando de ajuste de um modelo simétrico que se assemelha aos comandos utilizados para ajustar modelos lineares generalizados.

```
elliptical(formula = formula(data), DerB = NULL, parmB = NULL,
family = Normal,  data = sys.parent(), dispersion= NULL,
weights,subset, na.action = "na.fail",  method = "elliptical.fit",
control = glm.control(maxit = 100, trace = F),  model = F, x = F,
y = T, contrasts = NULL, linear = T, restrict = F, Cres =  NULL,
sol = NULL, offset, ...)
```

Para o uso da library `elliptical` é necessário chamar inicialmente a library `Matrix`. Após o ajuste de um modelo simétrico utilizando a library `elliptical` ficará dispónivel uma lista de objetos gerados, tais como

`coefficients` : coeficientes de locação do modelo ajustado;

`dispersion` : coeficiente de dispersão do modelo ajustado;

`residuals` : resíduo $(y-\mu)/\sqrt{\phi}$;

`fitted.values`: valores ajustados;

`loglik` : o valor do logaritmo da verossimilhança do modelo ajustado;

`Wg` : os valores da função $W_g(u)$;

`Wgder` : os valores da função $W_g'(u)$;

`v` : os valores da função v_i;

`iter` : número de iterações;

`scale` : $4d_g$;

`scaledispersion` : $4f_g - 1$;

`scalevariance` : ξ;

`linear` : assume T, se for um modelo de regressão linear e F, se for não-linear;

`Xmodel` : é a matriz modelo no caso linear e no caso não-linear é a matriz $\mathbf{D}_\beta$;

`DerBB` : é o `array` $\mathbf{D}_{\beta\beta}$;

`nfunc` : a função não-linear.

Os demais objetos seguem a estrutura dos objetos gerados pela função `glm`. Na opção `family`, define-se a família de distribuição a ser ajustada. Esta `library` atualmente está definida para a classe distribuições abaixo :

Normal : `family`=Normal()

$t-$Student, por exemplo, com 3 g.l. : `family`=Student(3)

$t-$Student Generalizada, por exemplo, com s=2 e r=3 : `family`=Gstudent(c(2,3))

Logística-I : `family`=LogisI()

Logística-II : `family`=LogisII()

Logística Generalizada, por exemplo, com $\alpha = 1$ e $m = 2$: `family`=Glogis(c(1,2))

Exponencial Potência, por exemplo, com k=0,5 : `family`=Powerexp(0.5).

Também é possível usar esta `library` para definir outras distribuições através do comando `make.family.elliptical`. Por exemplo, criar a família `Cauchy` através dos comandos abaixo.

```
Cauchy <- function()
{
  make.family.elliptical("Cauchy")
} #  Um objeto '.deriv' (de mesmo nome que a funcao geradora!)
Cauchy.deriv<-structure(
.Data=list(
           g0=function(z,...) log((1/pi)*(1+z^2)^(-1)),
           g1=function(z,...) -1/(1+z^2),
           g2=function(z,...) 1/8,
           g3=function(z,...) 3/8,
           g4=function(z,...) 1, ## nao definido
           g5=function(z,...) 1/((1+z^2)^2)),
           .Dim=c(6,1),
           .Dimnames=list(c("g0","g1","g2","g3","g4","g5"),
                  c("Cauchy")))
```

Alguns gráficos de diagnóstico discutidos no capítulo anterior já estão implementados através do comando `elliptical.diag.plots(obj)`, em que `obj` é o objeto proveniente do ajuste do modelo. Sendo assim, após chamada a função será aberta uma janela na qual aparecerá o seguinte:

```
 Make a plot selection (or 0 to exit)


1: plot: All
2: plot: Response residual against fitted values
3: plot: Response residual against index
4: plot: Standardized residual against fitted values
5: plot: Standardized residual against index
6: plot: QQ-plot of  response residuals
7: plot: QQ-plot of  Standardized residuals
8: plot: Generalized Leverage
9: plot: Ci against index
10: plot: |Lmax| against index (local influence on coefficients)
11: plot: Bii against index

Selection:
```

As demais funções implementadas serão discutidas através de exemplos.

3.1 Estudo da luminosidade de um novo produto alimentício

A preocupação atual com a qualidade nutricional dos alimentos vem estimulando o desenvolvimento de novos produtos alimentícios, com baixos teores de gordura e de outros elementos possivelmente nocivos à saúde. Tendo esse objetivo em mente foi desenvolvido no Departamento de Nutrição da Faculdade de Saúde Pública da Universidade de São Paulo um produto do tipo "snack", que possui baixo teor de gordura saturada e de ácidos graxos. Neste novo produto optou-se por substi-

tuir, totalmente ou parcialmente, o agente responsável pela fixação do aroma do produto, a gordura vegetal hidrogenada por óleo de canola. Foram produzidos 8 diferentes formas do novo produto, sendo que cada uma delas recebeu os mesmos ingredientes aromatizantes, sob mesmas concentrações e mesmas condições ambientais, com exceção do veículo lipídico usado como fixador do aroma, cuja variação se deu na proporção de óleo de canola e gordura vegetal hidrogenada. Em particular, há interesse em comparar 5 dessas formas: A (22% de gordura, 0% de óleo de canola), B (0% de gordura, 22% de óleo de canola), C (17% de gordura, 5% de óleo de canola), D (11% de gordura, 11% de óleo de canola) e E (5% de gordura, 17% de óleo de canola) segundo os níveis de ácidos graxos e a cor do produto, tais como tom, luminosidade e croma, ao longo do tempo. Um experimento foi conduzido durante 20 semanas em que nas semanas ímpares 3 embalagens de cada um dos produtos A, B, C, D e E foram analisadas em laboratório e observadas diversas variáveis, dentre as quais a luminosidade do produto (quanto maior o valor mais claro o produto) na escala de 0 a 100 (ver, por exemplo, Beering, 1999), cujos resultados serão discutidos a seguir e os dados estão disponíveis no Apêndice. Uma análise estatística completa deste experimento está descrita em Paula, de Moura e Yamaguchi (2004).

Na Figura 3.1 tem-se o comportamento da luminosidade (para todos os grupos) ao longo das 20 semanas. Nota-se um decrescimento do grau de luminosidade ao longo do tempo havendo uma estabilidade a partir da 11ª semana em torno do valor 68,5. Como o objetivo principal deste estudo é comparar os 5 grupos (A a E) segundo a luminosidade média do produto e como o comportamento ao longo das semanas é muito similar entre os grupos, trataremos a variável tempo como uma covariável assumindo, em princípio, a mesma tendência para os 5 grupos. Assim, denotando por y_{ijk} a luminosidade do k-ésimo produto do i-ésimo tipo na j-ésima semana, em que $i = 1(\text{A}), 2(\text{B}), 3(\text{C}), 4(\text{D})$ e $5(\text{E})$; $j = 1, 3, 5, 7, 9, 11, 13, 15, 17, 19$ e $k = 1, 2, 3$, propomos o seguinte modelo:

$$y_{ijk} = \alpha + \beta_i + \gamma_1 x_j + \gamma_2 x_j^2 + \epsilon_{ijk} \ , \tag{3.1}$$

em que $\alpha + \beta_i$ é o efeito (controlado pela semana) do i-ésimo grupo (assumiremos $\beta_1 = 0$), x_j : j-ésima semana e $\epsilon_{ijk} \sim S(0, \phi)$ são erros mutuamente independentes. Iniciaremos as análises supondo erros normais e em seguida sob outros erros simétricos.

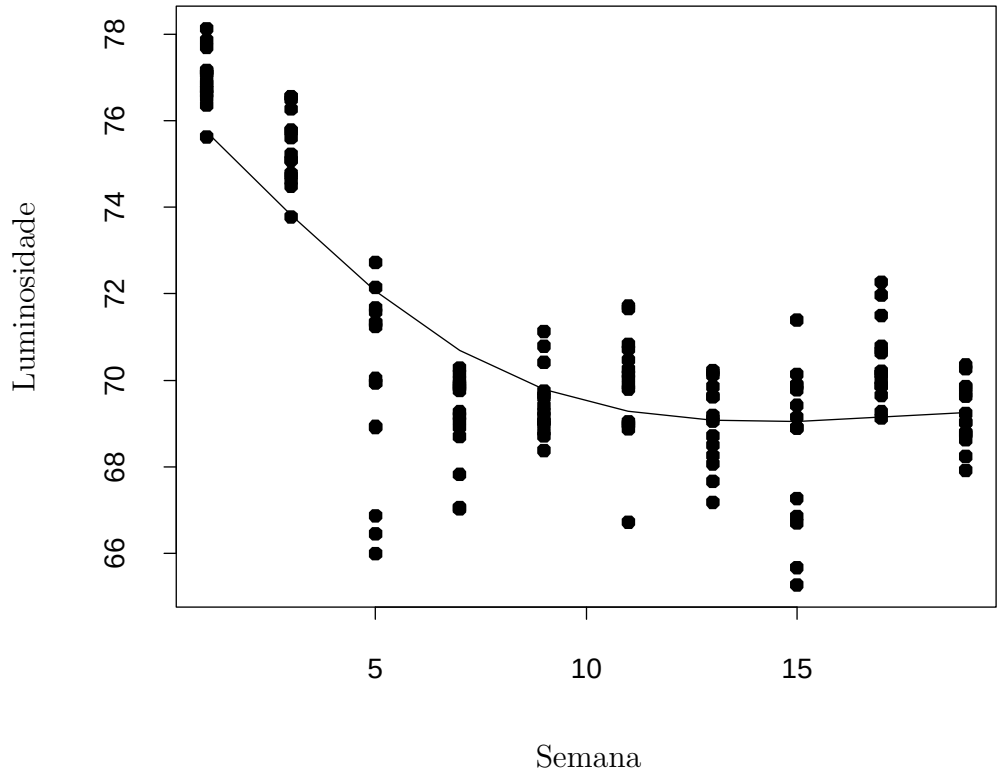

Figura 3.1 *Comportamento da luminosidade dos produtos ao longo das semanas.*

3.1.1 Análise sob erros normais

Na Tabela 3.1 tem-se as estimativas de máxima verossimilhança sob erros normais independentes e nota-se pelos níveis descritivos (p-valores) que apenas o grupo D parece ter uma luminosidade média inferior ao grupo A. O efeito semana parece ter sido controlado como pode-se notar pela Figura 3.2a, onde tem-se o gráfico dos resíduos contra o tempo, embora as observações A5.1, A5.2 e A5.3 para o grupo A referentes à 5ª semana apareçam como aberrantes. Pelos gráficos de $C_i(\boldsymbol{\theta})$ sob

perturbações de casos e na escala apresentados nas Figuras 3.4a e 3.5a, respectivamente, as mesmas observações são destacadas como influentes. No gráfico de envelope (Figura 3.3a) as três observações aparecem fora da banda gerada. A eliminação desses pontos, além de causar mudanças desproporcionais nas estimativas dos efeitos (vide Tabela 3.5), causam mudanças inferenciais, como pode-se notar pela Tabela 3.1. Com a eliminação das três observações apenas o grupo B parece não diferir do grupo A. Os demais grupos parecem ter um nível médio de luminosidade menor do que o grupo A. Fica então a dúvida: qual modelo ajustado considerar, com ou sem as três observações discrepantes? A fim de tentar reduzir a influência desses pontos nos resultados do modelo ajustado, assumiremos a seguir erros com caudas mais pesadas do que a normal.

Tabela 3.1 *Estimativas de máxima verossimilhança dos parâmetros do modelo (3.1) ajustado aos dados de luminosidade sob erros normais.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	t-valor	p-valor	Estimativa	t-valor	p-valor
Constante	78,13	164,26	0,00	78,92	193,81	0,00
Grupo A	0,00	-	-	0,00	-	-
Grupo B	0,31	0,76	0,45	-0,35	-1,00	0,32
Grupo C	-0,17	-0,41	0,68	-0,83	-2,37	0,02
Grupo D	-0,94	-2,29	0,02	-1,61	-4,56	0,00
Grupo E	-0,49	-1,20	0,23	-1,16	-3,29	0,00
γ_1	-1,44	-15,60	0,00	-1,44	-18,69	0,00
γ_2	0,05	12,32	0,00	0,05	14,46	0,00
ϕ	2,54	8,66	0,00	2,58	8,57	0,00
R^2	0,73	-	-	0,81	-	-

3.1.2 Análise sob erros simétricos de caudas pesadas

Inicialmente vamos considerar erros t de Student com ν graus de liberdade para o modelo (3.1). Assim, aplicando um procedimento de seleção tipo Akaike que consiste em minimizar a função AIC$= -\mathrm{L}_\nu(\boldsymbol{\theta}) + p$, encontramos o menor valor para AIC quando $\nu \cong 5$. Ajustamos então o modelo (3.1) aos dados com erros independentes t de Student com 5 graus de liberdade, cujas estimativas são apresentadas na Tabela 3.2 e os gráficos de diagnóstico nas Figuras 3.2b, 3.3b, 3.4b e 3.5b. Para ajustar este modelo pela library `elliptical` deve-se usar os comandos dados a seguir.

Tabela 3.2 *Estimativas de máxima verossimilhança dos parâmetros do modelo (3.1) ajustado aos dados de luminosidade sob erros t de Student com 5 g.l..*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
Constante	78,87	191,25	0,00	79,07	200,91	0,00
Grupo A	0,00	-	-	0,00	-	-
Grupo B	-0,36	-1,02	0,31	-0,55	-1,61	0,11
Grupo C	-0,77	-2,17	0,03	-0,95	-2,78	0,00
Grupo D	-1,50	-4,21	0,00	-1,67	-4,90	0,00
Grupo E	-1,02	-2,86	0,00	-1,18	-3,48	0,00
γ_1	-1,43	-17,94	0,00	-1,43	-19,19	0,00
γ_2	0,05	13,85	0,00	0,05	14,71	0,00
ϕ	1,42	6,85	0,00	1,23	6,78	0,00

```
library(Matrix)
luz.dat <- scan(what=list(y=0,x1=0,x2=0,rot=""))
77.86   1   1   A1.1
77.70   1   1   A1.2
.
.
.
```

```
69.69   5   19  E19.3

 attach(luz.dat)
 y <- luz.dat$y
 x1 <- luz.dat$x1
 x1 <- factor(x1)
 x1 <- C(x1,treatment)
 x2 <- luz.dat$x2
 x3 <- (luz.dat$x2)^2
 luz <- data.frame(y,x1,x2,x3)
 luzt.elpt <- elliptical(formula=y~x1+x2+x3,family=Student(5)
 ,data=luz)
summary(luzt.elpt)
Call: elliptical(formula = y ~ x1 + x2 + x3, family = Student(5),
data = luz)

Coefficients:
                  Value Std. Error    z-value      p-value
(Intercept)  78.8741789 0.41022365  192.27116 0.00000e+000
        x12  -0.3638037 0.35512226   -1.02445 3.05625e-001
        x13  -0.7751104 0.35512226   -2.18266 2.90610e-002
        x14  -1.5025361 0.35512226   -4.23104 2.32614e-005
        x15  -1.0206426 0.35512226   -2.87406 4.05232e-003
         x2  -1.4299810 0.07970797  -17.94025 5.71962e-072
         x3   0.0535159 0.00386368   13.85103 1.25395e-043

Scale parameter for Student : 1.41876  ( 0.207223 )

Degrees of Freedom: 150 Total; 143 Residual
-2*Log-Likelihood 542.794
```

```
Number  Iterations: 11

Correlation of Coefficients:
    (Intercept)       x12       x13       x14       x15        x2
x12 -0.432840
x13 -0.432840    0.500000
x14 -0.432840    0.500000  0.500000
x15 -0.432840    0.500000  0.500000  0.500000
 x2 -0.728639    0.000000  0.000000  0.000000  0.000000
 x3  0.631037    0.000000  0.000000  0.000000  0.000000 -0.969458
```

Notamos pela Figura 3.2b que o efeito tempo foi controlado, contudo as observações A5.1, A5.2 e A5.3 continuam aparecendo como aberrantes. Pelas Figuras 3.4b e 3.5b não detectamos observações localmente influentes nos coeficientes estimados $\hat{\boldsymbol{\theta}}$ sob perturbações de casos e na escala, respectivamente. O gráfico de envelope dado na Figura 3.3b acomoda melhor as observações aberrantes do que sob erros normais. Quando eliminamos os três pontos aberrantes não notamos mudanças inferenciais ao nível de significância de 5% (vide Tabela 3.2) e as estimativas variam bem menos do que sob erros normais (vide Tabela 3.5). Pelo modelo t de Student com 5 graus de liberdade não detectamos, ao nível de 5%, diferenças significativas entre os valores médios dos grupos A e B, havendo fortes indícios de que os demais grupos têm níveis médios de luminosidade inferiores ao grupo A. Esses resultados não mudam quando as observações aberrantes são eliminadas, confirmando a robustez das estimativas de máxima verossimilhança sob erros t de Student contra pontos extremos. Para ajustar o modelo t de Student com 5 graus de liberdade no S-Plus e R sem as três observações aberrantes que estão nas posições 7,8 e 9 deve-se fazer o seguinte :

```
 luzts <- elliptical(formula=y~x1+x2+x3,subset=-c(7,8,9),
family=Student(5),data=luz)
```

Ajustamos também o modelo (3.1) aos dados sob erros independentes logístico tipo II e exponencial potência com parâmetro k. Usamos novamente o critério de Akaike, agora para encontrar o valor k que minimiza a quantidade AIC$= -\mathrm{L}_k(\boldsymbol{\theta}) + p$ no modelo exponencial potência. O menor valor de AIC foi obtido para $k \cong 0,7$. As estimativas de máxima verossimilhança são apresentadas nas Tabelas 3.3 e 3.4, respectivamente, e os gráficos de diagnóstico nas Figuras 3.2c, 3.3c, 3.4c e 3.5c para o modelo logístico-II e nas Figuras 3.2d, 3.3d, 3.4d e 3.5d para o modelo exponencial potência com $k = 0,7$. Para ajustar os dois modelos em S-Plus e R com a library `elliptical` deve-se proceder conforme os comandos dados abaixo.

```
luzlII <- elliptical(formula=y~x1+x2+x3,family=LogisII(),data=luz)
luzpe <- elliptical(formula=y~x1+x2+x3,family=Powerexp(0.7),
data=luz)
```

Para ambos os modelos ajustados o efeito tempo foi controlado, embora os pontos A5.1, A5.2 e A5.3 continuem aparecendo como aberrantes. Essas três observações aparecem com alguma influência local nas estimativas de máxima verossimilhança dos dois modelos ajustados e os envelopes gerados são parecidos com o envelope sob erros normais. Os resultados inferenciais para ambos os modelos são os mesmos sob erros t de Student, conforme pode-se observar pelas Tabelas 3.3 e 3.4. Quando eliminamos os pontos aberrantes o modelo logístico-II comporta-se de forma similar ao modelo t de Student com 5 graus de liberdade, não havendo mudanças inferenciais ao nível de 5% (vide Tabela 3.3).

Contudo, sob erros exponencial potência há indícios agora de que o nível médio de luminosidade do grupo B seja inferior ao nível médio do grupo A, embora os demais resultados inferencias não mudem (vide Tabela 3.4). Olhando a Tabela 3.5 notamos qua as variações nas estimativas de máxima verossimilhança, quando as observações aberrantes são retiradas, variam mais sob erros normais. Essas variações diminuem substancialmente sob erros com caudas mais pesadas, variando

menos sob erros t de Student e exponencial potência. Porém, se considerarmos também as variações nas estimativas dos desvios padrão assintóticos (vide Tabela 3.6), notamos que as menores variações ocorrem sob erros t de Student. Assim, como uma conclusão preliminar para este exemplo, podemos dizer que o modelo t de Student com 5 graus de liberdade parece se ajustar melhor aos dados do que os demais modelos considerados no sentido de robustez das estimativas obtidas contra os pontos aberrantes.

Os objetos gerados pelo ajustes pertencem à classe `elliptical`. Associada a esta classe temos várias funções implementadas tais como `summary.elliptical` que produz a saída descrita acima, `anova.elliptical` quando chamada gera o teste da razão de verossimilhanças, e `elliptical.diag` que calcula as medidas de diagnóstico descritas no Capítulo 2.

Pode-se extrair três tipos de resíduos, utilizando a função `resid`, que com a opção `type=‘‘response’’`, `type=‘‘pearson’’` e `type=‘‘stand’’` produz o resíduo ordinário, o resíduo $\frac{(y-\mu)}{\phi\sqrt{\xi}}$ e o resíduo t_{r_i}, respectivamente. Para uma análise residual, podemos construir alguns gráficos baseados nos valores ajustados que são

Tabela 3.3 *Estimativas de máxima verossimilhança dos parâmetros do modelo (3.1) ajustado aos dados de luminosidade sob erros logístico-II.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
Constante	78,72	183,48	0,00	79,04	200,32	0,00
Grupo A	0,00	-	-	0,00	-	-
Grupo B	-0,21	-0,57	0,57	-0,51	-1,49	0,13
Grupo C	-0,64	-1,73	0,08	-0,92	-2,71	0,00
Grupo D	-1,38	-3,72	0,00	-1,66	-4,87	0,00
Grupo E	-0,90	-2,43	0,00	-1,18	-3,45	0,00
γ_1	-1,43	-17,28	0,00	-1,43	-19,18	0,00
γ_2	0,05	13,41	0,00	0,05	14,73	0,00
ϕ	0,68	7,32	0,00	0,55	7,25	0,00

Tabela 3.4 *Estimativas de máxima verossimilhança dos parâmetros do modelo (3.1) ajustado aos dados de luminosidade sob erros exponencial potência.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
Constante	79,09	201,05	0,00	79,27	223,76	0,00
Grupo A	0,00	-	-	0,00	-	-
Grupo B	-0,51	-1,49	0,13	-0,72	-2,36	0,02
Grupo C	-0,71	-2,11	0,03	-0,89	-2,91	0,00
Grupo D	-1,60	-4,73	0,00	-1,80	-5,89	0,00
Grupo E	-1,08	-3,19	0,00	-1,27	-4,16	0,00
γ_1	-1,44	-19,09	0,00	-1,43	-21,39	0,00
γ_2	0,05	14,66	0,00	0,05	16,27	0,00
ϕ	0,61	6,64	0,00	0,47	6,57	0,00

obtidos através da função `fitted(objeto)` ou extraído do `objeto`. Para ilustrar a construção do gráfico de resíduo t_{r_i} contra os valores ajustados, linhas pontilhadas em -2 e 2 e três observações identificadas com os rótulos dados em `rot`, temos o seguinte :

```
tdt <- resid(luzt.elpt,type="stand")
fitt <- fitted(luzt.elpt)
plot(fitt,tdt,xlab = "Valores Ajustados",
ylab = "Resíduo Padronizado")
abline(-2,0,lty = 2)
abline(2,0,lty = 2)
identify(fitt,tdt,n=3,labels=rot)
```

Para construir as bandas de confiança no gráfico normal de probabilidades correspondente ao modelo t de Student com 5 graus de liberdade basta chamar a função envelope.

```
arg <- 5
testet <- envelope(luzt.elpt,B=100,arg=arg)
```

Tabela 3.5 *Variações (em %) nas estimativas de máxima verossimilhança dos modelos ajustados aos dados de luminosidade quando eliminamos os pontos aberrantes A5.1,A5.2 e A5.3.*

Efeito	Normal	t_5	EP$(0,7)$	Logístico-II
Constante	1	0	0	0
Grupo B	-212	50	42	140
Grupo C	391	22	24	44
Grupo D	70	11	12	20
Grupo E	135	16	18	30
γ_1	-0	-0	-1	0
γ_2	-2	-1	-2	-1
ϕ	-31	-13	-22	-19

Tabela 3.6 *Variações (em %) nas estimativas dos desvios padrão assintóticos das estimativas de máxima verossimilhança dos modelos ajustados aos dados de luminosidade quando eliminamos os pontos aberrantes A5.1,A5.2 e A5.3.*

Efeito	Normal	t_5	EP$(0,7)$	Logístico-II
Constante	-14	-4	-9	-8
Grupo B	-15	-4	-9	-8
Grupo C	-15	-4	-9	-8
Grupo D	-15	-4	-9	-8
Grupo E	-15	-4	-9	-8
γ_1	-17	-7	-12	-10
γ_2	-17	-7	-12	-10
ϕ	-30	-12	-21	-19

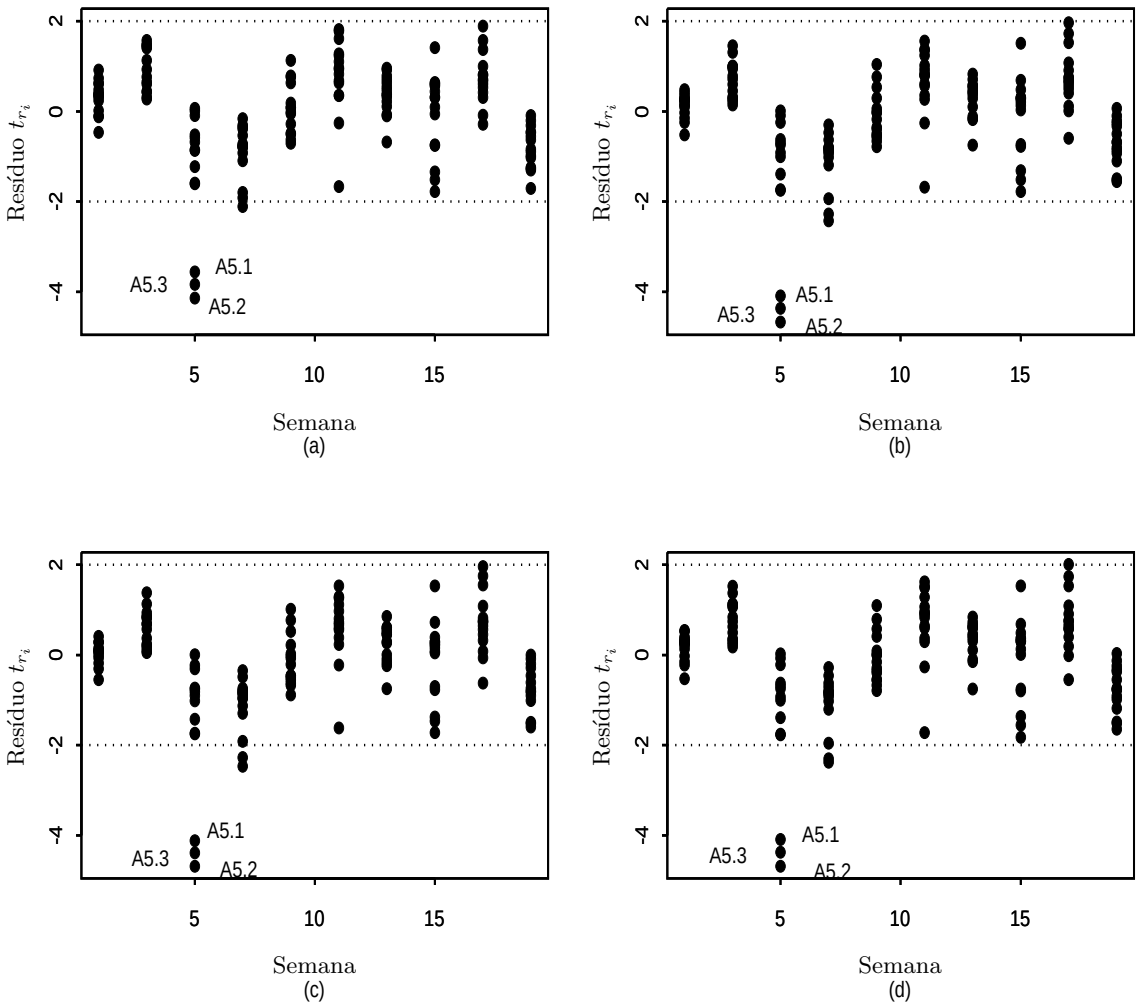

Figura 3.2 *Gráficos de t_{r_i} contra o tempo para o modelo (3.1) sob erros normais (a), t−Student com 5 g.l. (b), exponencial potência com k=0,7 (c) e logístico-II (d).*

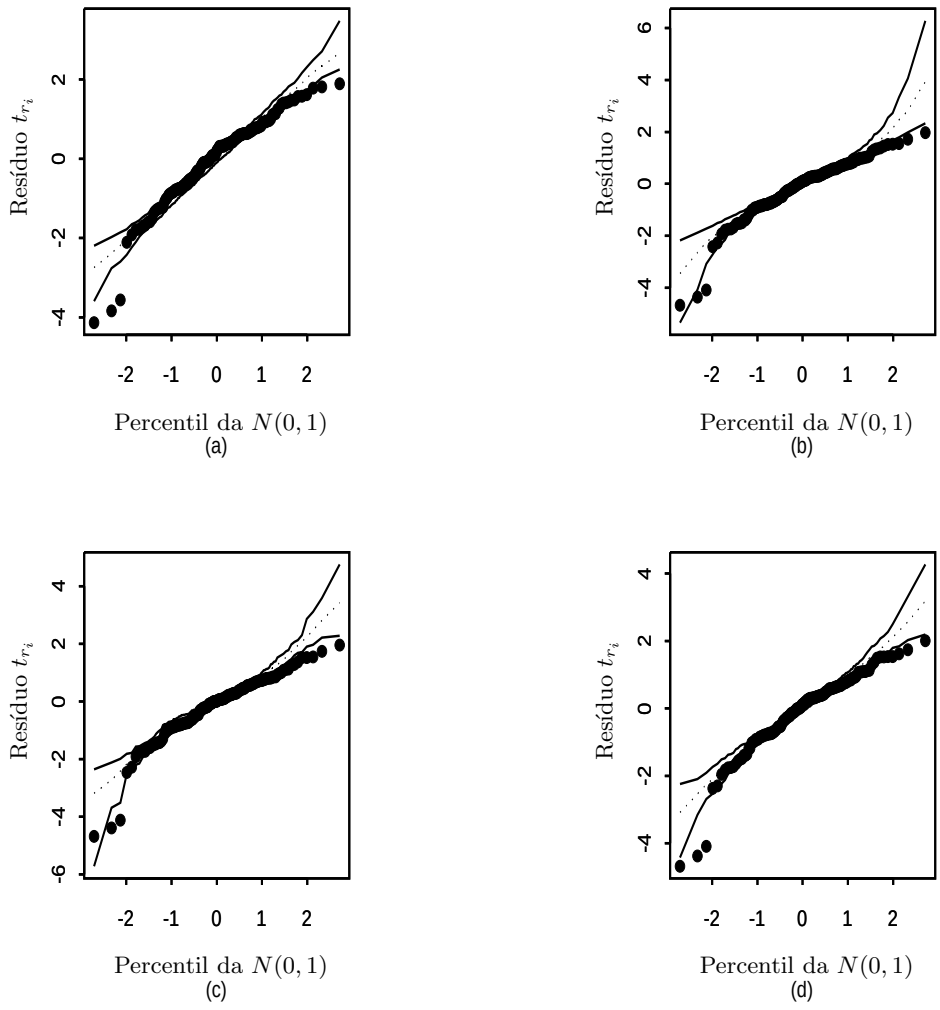

Figura 3.3 *Gráficos normais de probabilidades com envelopes para o resíduo* t_{r_i} *para o modelo (3.1) sob erros normais (a), t−Student com 5 g.l. (b), exponencial potência com k=0,7 (c) e logístico-II (d).*

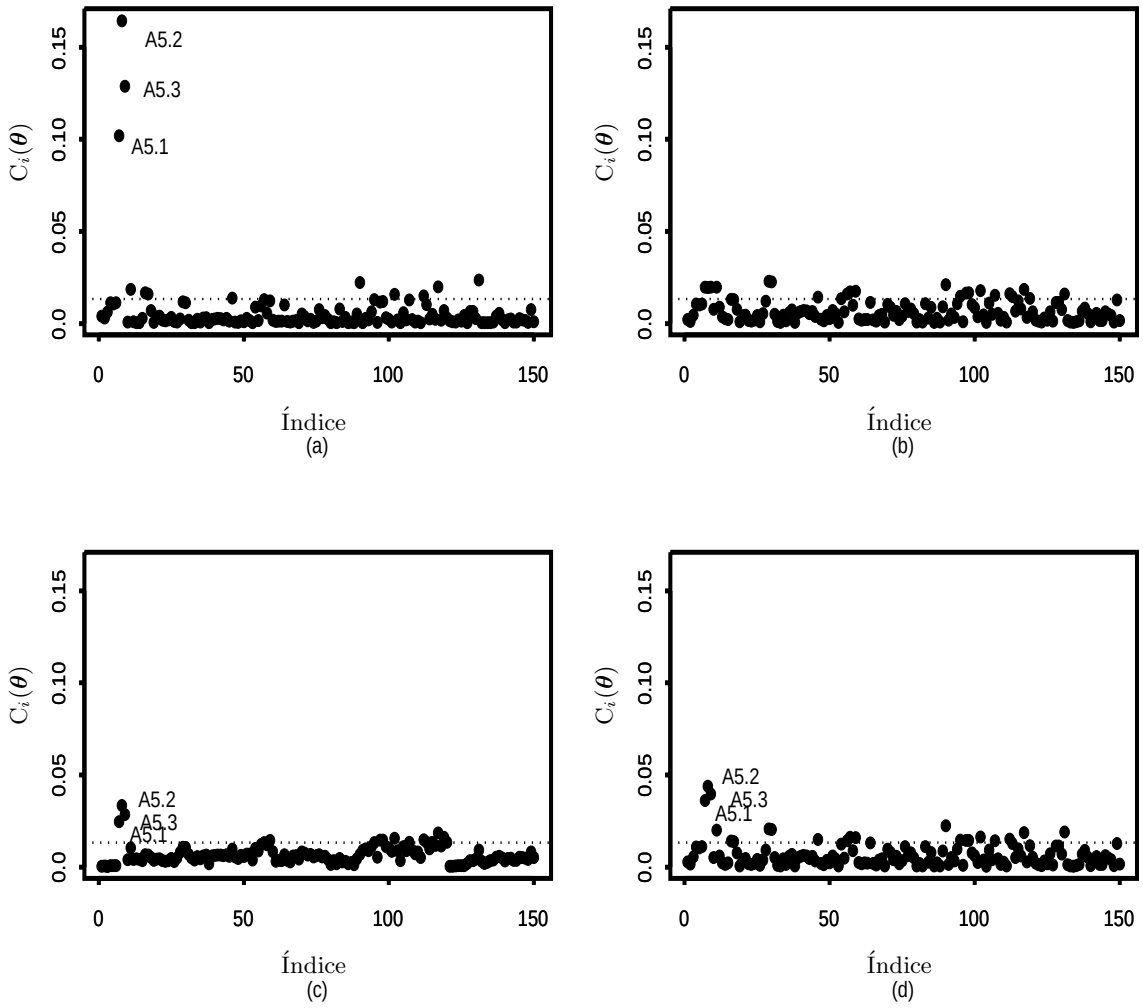

Figura 3.4 *Gráficos de influência local total* $C_i(\boldsymbol{\theta})$ *sob perturbação de casos para o modelo (3.1) sob erros normais (a), t−Student com 5 g.l. (b), exponencial potência com k=0,7 (c) e logístico-II (d).*

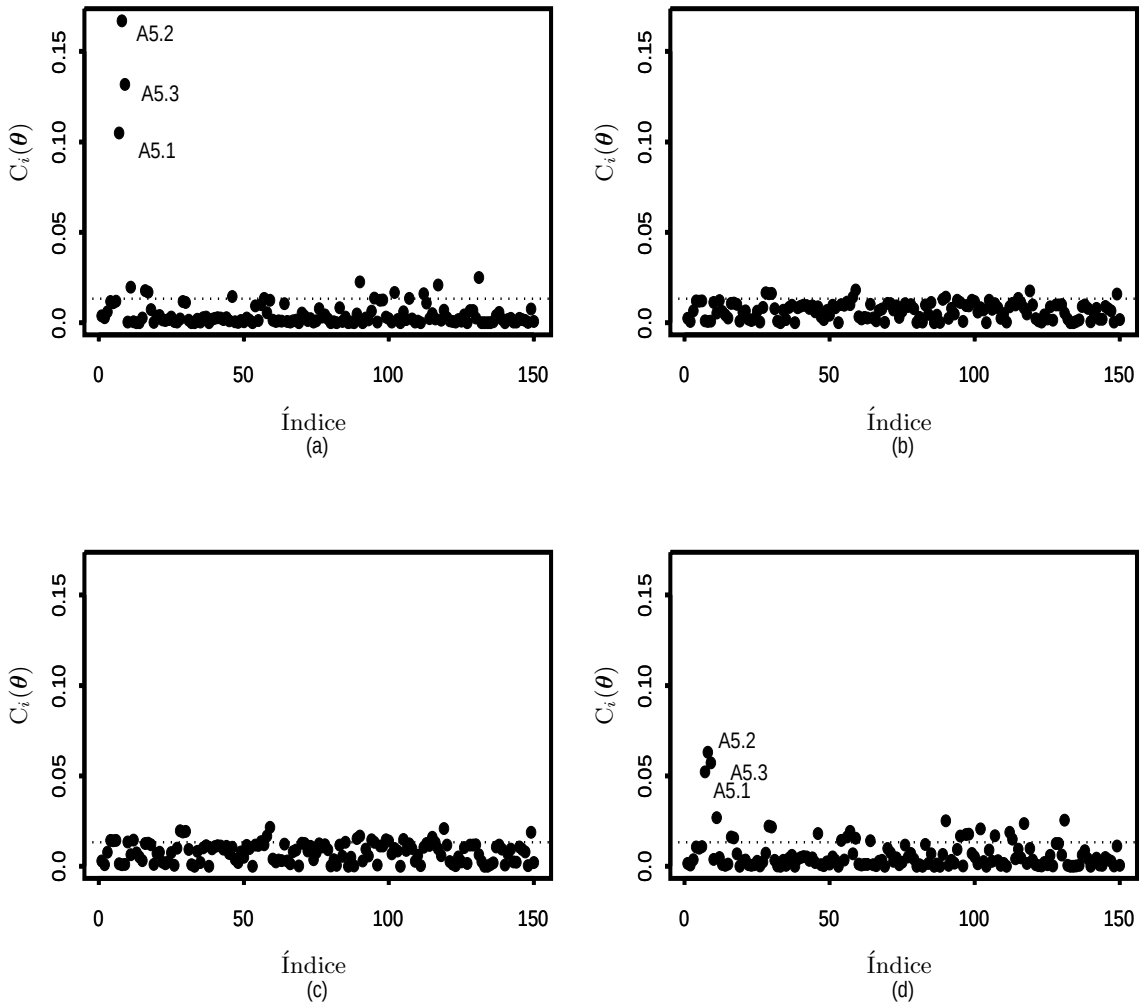

Figura 3.5 *Gráficos de influência local total* $\mathrm{C}_i(\boldsymbol{\theta})$ *sob perturbação na escala para o modelo (3.1) sob erros normais (a), t−Student com 5 g.l. (b), exponencial potência com k=0,7 (c) e logístico-II (d).*

3.2 Coelhos europeus na Austrália

Para ilustrar uma aplicação com preditor não-linear consideraremos o conjunto de dados descrito em Ratkowsky (1983, Tabela 6.1) apresentado no Apêndice, cujo interesse principal é relacionar o peso das lentes dos olhos de coelhos europeus, y

(em mg) (*Oryctolagus cuniculus)* e a idade do animal, x (em dias), uma amostra de 71 observações. Este animal é largamente distribuído na população selvagem da Austrália. Um aspecto interessante para este conjunto de dados, que suporta o uso de erros com distribuição com caudas mais pesadas do que a normal, é a suspeita de dois pontos aberrantes sob estimação de mínimos quadrados (vide, Wei, 1998, Exemplo 6.8). Então, para reanalizar os dados, propomos o seguinte modelo :

$$y_i = \exp\left(\alpha - \frac{\beta}{x_i + \gamma}\right) e^{\epsilon_i}, \tag{3.2}$$

$i = 1, \ldots, 71$, em que $\epsilon_i \sim S(0, \phi)$ são erros mutuamente independentes.

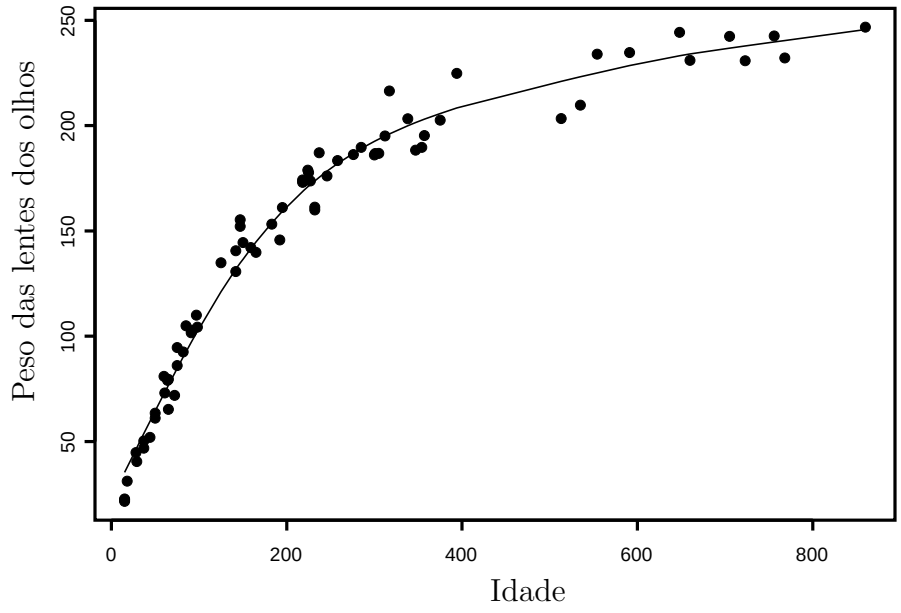

Figura 3.6 *Gráfico de dispersão do peso das lentes dos olhos contra idade de coelhos europeus.*

Várias distribuições com caudas mais pesadas do que a normal foram assumidas, porém, somente dois modelos parecem ajustar-se aos dados tão bem quanto ou melhor do que o modelo normal, o modelo $t-$Student com 4 graus de liberdade e

o modelo logístico-II. Para determinar os graus de liberdade do modelo $t-$Student encontramos o menor valor para AIC quando $\nu \cong 4$. Ajustamos então o modelo (3.2) aos dados com erros independentes normal, logístico-II e t de Student com 4 graus de liberdade. Para ajustar modelos não-lineares simétricos usando a `library elliptical` é necessário fornecer a estrutura funcional da matriz de primeiras derivadas e segundas derivadas (se o objetivo for fazer análise de diagnóstico). Neste exemplo temos que introduzir os comandos dados abaixo.

```
DerBrabbits <- function(parm,X)
{
         alpha <- parm[1]
         beta <- parm[2]
         nu <- parm[3]
         x0 <- X[,1]
         x <- X[,2]
        .grad <- array(0, c(length(x0), 3),
        list(NULL, c("alpha","beta","nu")))
        .grad[, "alpha"] <- x0
        .grad[, "beta"] <- -1/(x+nu)
        .grad[, "nu"] <- beta/((x+nu)^2)
        .value <- alpha -beta/(x+nu)
        .hess <- array(0, c(3, 3,length(x0)),
        list( c("alpha","beta","nu"), c("alpha","beta","nu"),NULL))
        #necessary, if you want to analyse diagnostic
        .hess["alpha","alpha", ] <- rep(0,length(x0))
        .hess["alpha","beta", ] <- rep(0,length(x0))
        .hess["alpha","nu",  ] <- rep(0,length(x0))
        .hess["beta","alpha", ] <- .hess["alpha","beta", ]
        .hess["beta","beta", ] <- rep(0,length(x0))
        .hess["beta","nu", ] <- 1/((x+nu)^2)
```

```
        .hess["nu","alpha", ] <- .hess["alpha","nu", ]
        .hess["nu","beta", ] <- .hess["beta","nu", ]
        .hess["nu","nu",] <- -2*beta/((x+nu)^3)
        fit <- list(value=.value,gradient=.grad,hessian=.hess)
        fit
}
```

Existem pequenas diferenças no uso da `library elliptical` quando pretendemos ajustar um modelo não-linear. Estas diferenças podem ser vistas nas seguintes opções:

`Linear=F` : indica que ajustaremos um modelo não-linear;

`formula` : devemos escrever no lado esquerdo de $\sim$ a variável resposta e do lado direito as variáveis explicativas;

`DerB` : recebe a função que define as primeiras e segundas derivadas e

`parmB` : será a opção que recebe o vetor de valores iniciais.

Para o exemplo dos coelhos podemos por exemplo, escrever o programa para ajuste dos dados pelo modelo logístico-II da seguinte forma :

```
library(Matrix)
rabbits.dat<-scan(what=list(idade=0,peso=0))
 15  21.66
.
.
860 246.70
attach(rabbits.dat)
y<-log(rabbits.dat$peso)
x<-rabbits.dat$idade
X <- cbind(1,x)
rabbits <- data.frame(y,x)
rabbits.nonelpt <- elliptical(formula=y~x,linear=F,DerB=DerBrabbits,
parmB=c(5,130,37),family=LogisII(),data=rabbits)
```

A saída descritiva com os resultados do ajuste do modelo logístico-II não-linear aplicado aos dados dos coelhos fica dada por

```
summary(rabbits.nonelpt)
Call: elliptical(formula = y ~ x, DerB = DerBrabbits,
parmB = c(5, 130, 37),family = LogisII(), data = rabbits, linear = F)

Coefficients:
         Value Std. Error  z-value      p-value
alpha   5.6330  0.0178231 316.0498 0.00000e+000
 beta 127.2577  4.9923644  25.4905 2.51434e-143
   nu  35.8639  2.0158432  17.7910 8.29297e-071

Scale parameter for LogisII : 0.00108729  ( 0.000215816 )

Degrees of Freedom: 71 Total; 68 Residual
-2*Log-Likelihood -198.44

Number  Iterations: 10

Correlation of Coefficients:
        alpha     beta
beta 0.889047
  nu 0.782776 0.960375
```

Semelhante ao caso linear, podemos gerar envelopes no gráfico normal de probabilidades do modelo não-linear ajustado (veja, `envelope(rabbits.nonelpt,B=100)`). Podemos também calcular algumas medidas de diagnóstico utilizando `elliptical.diag(rabbits.nonelpt)`. Se pretendemos utilizar os gráficos pré-definidos podemos chamar `elliptical.diag.plots(rabbits.nonelpt)`. Uma série de objetos

é disponibilizada quando a função `elliptical.diag` é chamada. Temos a seguir alguns deles:

`resid` : resíduo $(y - \mu)/\sqrt{\phi}$;

`rs` : resíduo padronizado (t_{r_i});

`dispersion`: estimativa do parâmetro de dispersão;

`GL` : $\mathbf{GL}_{ii}(\hat{\theta})$;

`GLbeta` : $\mathbf{GL}_{\beta_{ii}}(\hat{\theta})$;

`GLphi` : $\mathbf{GL}_{\phi_{ii}}(\hat{\theta})$;

`h` : $\hat{h}_{ii}$;

`Om` : matriz de informação observada;

`IOm` : inversa da matriz de informação observada;

`a` : pesos a_i;

`b` : pesos b_i;

`c` : pesos c_i;

`Cic` : C_i (pertubação de casos) ,

`Cih` : C_i (pertubação na escala) e

`Ci` : C_i (pertubação na resposta),

dentre outras medidas.

A Figura 3.6 indica que a variabilidade da resposta cresce quando a idade do animal cresce, justificando o uso de um modelo multiplicativo. As estimativas de máxima verossimilhança são apresentadas nas Tabelas 3.7, 3.8 e 3.9, respectivamente, supondo erros normais, $t-$Student com 4 graus de liberdade e logístico-II. As estimativas em geral são parecidas, embora os desvios padrão aproximados das estimativas dos modelos $t-$Student e logístico-II são, quase sempre, menores do que as estimativas dos desvios padrão do modelo normal. As curvaturas intrínseca e paramétrica são desprezíveis nos três modelos, e o viés relativo das estimativas dos parâmetros tende a ser menor nos modelos com curtose maior. Os gráficos de

resíduos contra os valores ajustados mostram que as observações 4, 5, 16 e 17 aparecem com destaque como aberrantes em todos os modelos ajustados e os gráficos normais de probabilidades com envelope gerado para o resíduo t_{r_i} não apresentam nenhum comportamento não usual (vide Figuras 3.7, 3.8 e 3.9).

Tabela 3.7 *Estimativas de máxima verossimilhança dos parâmetros do modelo dado em (3.2) ajustado aos dados de coelhos sob erros normais.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
α	5,64	288,62	0,00	5,62	389,11	0,00
β	130,58	23,30	0,00	123,23	30,61	0,00
γ	37,60	16,54	0,00	33,83	20,64	0,00
ϕ	0,004	5,96	0,00	0,002	5,79	0,00

Tabela 3.8 *Estimativas de máxima verossimilhança dos parâmetros do modelo dado em (3.2) ajustado aos dados de coelhos sob erros $t-$Student com 4 graus de liberdade.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
α	5,63	336,82	0,00	5,63	390,78	0,00
β	126,28	27,17	0,00	124,62	30,84	0,00
γ	35,29	18,86	0,00	34,36	20,91	0,00
ϕ	0,002	4,50	0,00	0,001	4,37	0,00

Tabela 3.9 *Estimativas de máxima verossimilhança dos parâmetros do modelo dado em (3.2) ajustado aos dados de coelhos sob erros logístico-II.*

	Com todos os pontos			Sem pontos aberrantes		
Efeito	Estimativa	z-valor	p-valor	Estimativa	z-valor	p-valor
α	5,63	316,04	0,00	5,63	388,93	0,00
β	127,26	25,49	0,00	124,16	30,66	0,00
γ	35,86	17,79	0,00	34,19	20,75	0,00
ϕ	0,001	5,04	0,00	0,0007	4,88	0,00

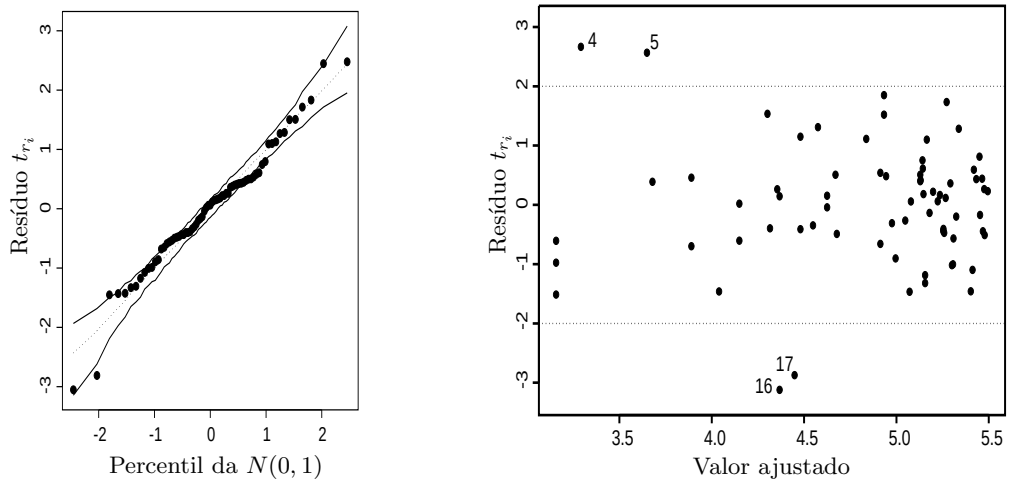

Figura 3.7 *Gráfico normal de probabilidades com envelope para t_{r_i} (esquerda) e gráfico de resíduos t_{r_i} contra os valores ajustados (direita) para o modelo normal ajustado aos dados de coelhos.*

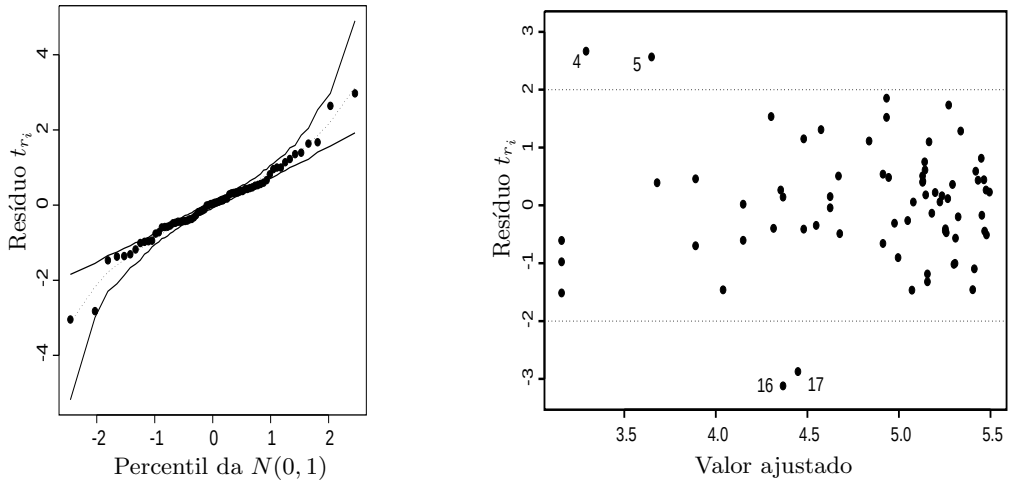

Figura 3.8 *Gráfico normal de probabilidades com envelope para t_{r_i} (esquerda) e gráfico de resíduos t_{r_i} contra os valores ajustados (direita) para o modelo t−Student com 4 g.l. ajustado aos dados de coelhos.*

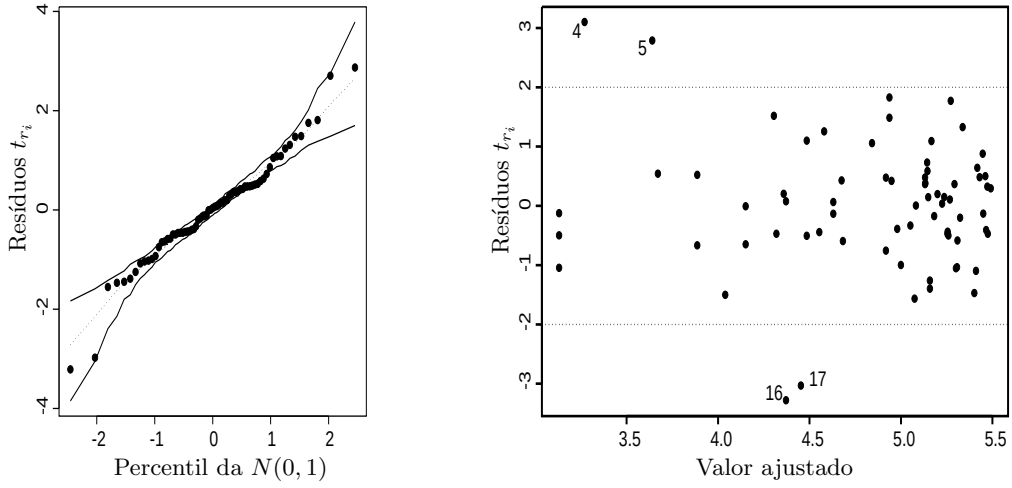

Figura 3.9 *Gráfico normal de probabilidades com envelope para t_{r_i} (esquerda) e gráfico de resíduos t_{r_i} contra os valores ajustados para o modelo logístico-II (direita) ajustado aos dados de coelhos.*

O modelo $t-$Student destaca menos observações nos gráficos de índices de $C_i(\boldsymbol{\theta})$ do que os modelos logístico-II e normal (vide Figuras 3.10, 3.11, e 3.12). Aqui, os animais jovens tendem a ser mais influentes nas estimativas dos parâmetros. Paula, Cysneiros e Galea (2003) observam que os pontos 1, 2 e 3 aparecem como pontos de alavanca nos três modelos mostrando a dificuldade de predição na resposta para animais mais jovens (vide Figura 3.13). A linha pontilhada nos gráficos de GL_{ii} representa o gráfico de índices de $\hat{h}_{ii}$ (ponto de alavanca do plano tangente) cujos valores são negligenciáveis, como esperado para o caso normal, pois a curvatura intrínseca é não significativa, mas diferem de valores nos modelos $t-$Student e logístico-II.

A eliminação das observações 4,5,16 e 17 produz maiores mudanças nas estimativas do modelo normal do que nas estimativas dos modelo $t-$Student e logístico-II (vide Tabelas 3.10 e 3.11). Eliminando os pontos aberrantes, influentes e de alta alavanca (vide Tabelas 3.12 e 3.13) ocorrem mais variações sob erros normais do que sob os erros $t-$Student e logístico-II. Nossa principal conclusão, após esta análise de diagnóstico, é que os modelos $t-$Student com 4 graus de liberdade e logístico-II parecem produzir estimativas para os parâmetros mais robustas contra os pontos discrepantes do que o modelo normal. Comparando-se os modelos t de Student com 4 g.l. e logístico-II notamos menores variações para o modelo t de Student tanto para as estimativas dos parâmetros quanto para os desvios padrão estimados.

Tabela 3.10 *Variações (em %) nas estimativas de máxima verossimilhança dos modelos ajustados aos dados de coelhos quando eliminamos os pontos aberrantes 4,5,16 e 17.*

Parâmetro	Normal	t_4	Logístico-II
α	0	0	0
β	-6	-1	-2
γ	-10	-3	-5
ϕ	-43	-27	-34

Tabela 3.11 *Variações (em %) nas estimativas dos desvios padrão assintóticos das estimativas de máxima verossimilhança dos modelos ajustados aos dados de coelhos quando eliminamos os pontos 4,5,16 e 17.*

Parâmetro	Normal	t_4	Logístico-II
α	-26	-14	-19
β	-28	-13	-19
γ	-28	-12	-18
ϕ	-41	-25	-32

Tabela 3.12 *Variações (em %) nas estimativas de máxima verossimilhança dos modelos ajustados aos dados de coelhos quando eliminamos os pontos 1,2,3,4,5,16 e 17.*

Parâmetro	Normal	t_4	Logístico-II
α	0	0	0
β	-6	0	-2
γ	-11	0	-3
ϕ	-41	-23	-31

Tabela 3.13 *Variações (em %) nas estimativas dos desvios padrão assintóticos das estimativas de máxima verossimilhança dos modelos ajustados aos dados de coelhos quando eliminamos os pontos 1,2,3,4,5,16 e 17.*

Parâmetro	Normal	t_4	Logístico-II
α	-7	-12	3
β	11	-11	28
γ	48	-12	70
ϕ	-38	-19	-28

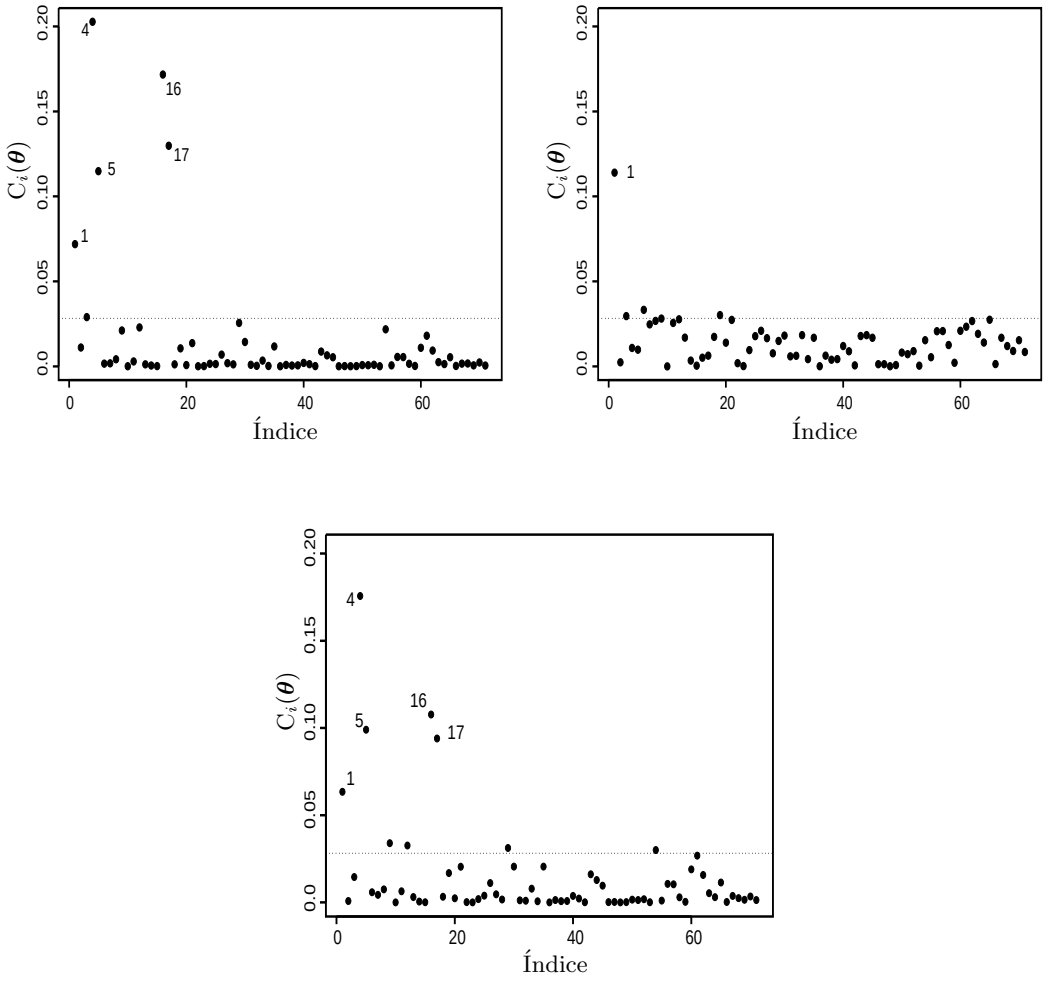

Figura 3.10 *Gráficos de índices de* $\mathrm{C}_i(\boldsymbol{\theta})$ *para o modelo normal (esquerda), t−Student com 4 g.l. (direita) e logístico-II (abaixo) ajustados aos dados de coelhos.*

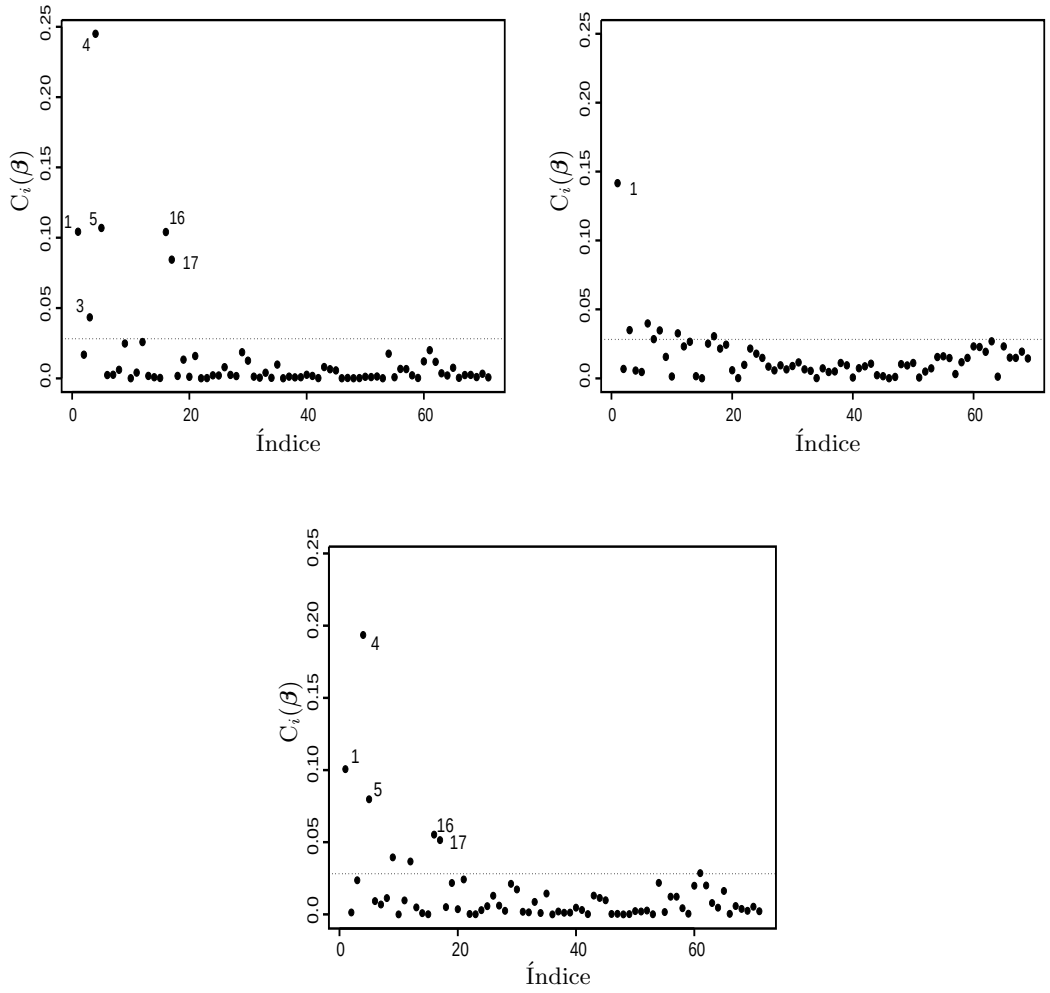

Figura 3.11 *Gráficos de índices de* $C_i(\boldsymbol{\beta})$ *para o modelo normal (esquerda), t−Student com 4 g.l. (direita) e logístico-II (abaixo) ajustados aos dados de coelhos.*

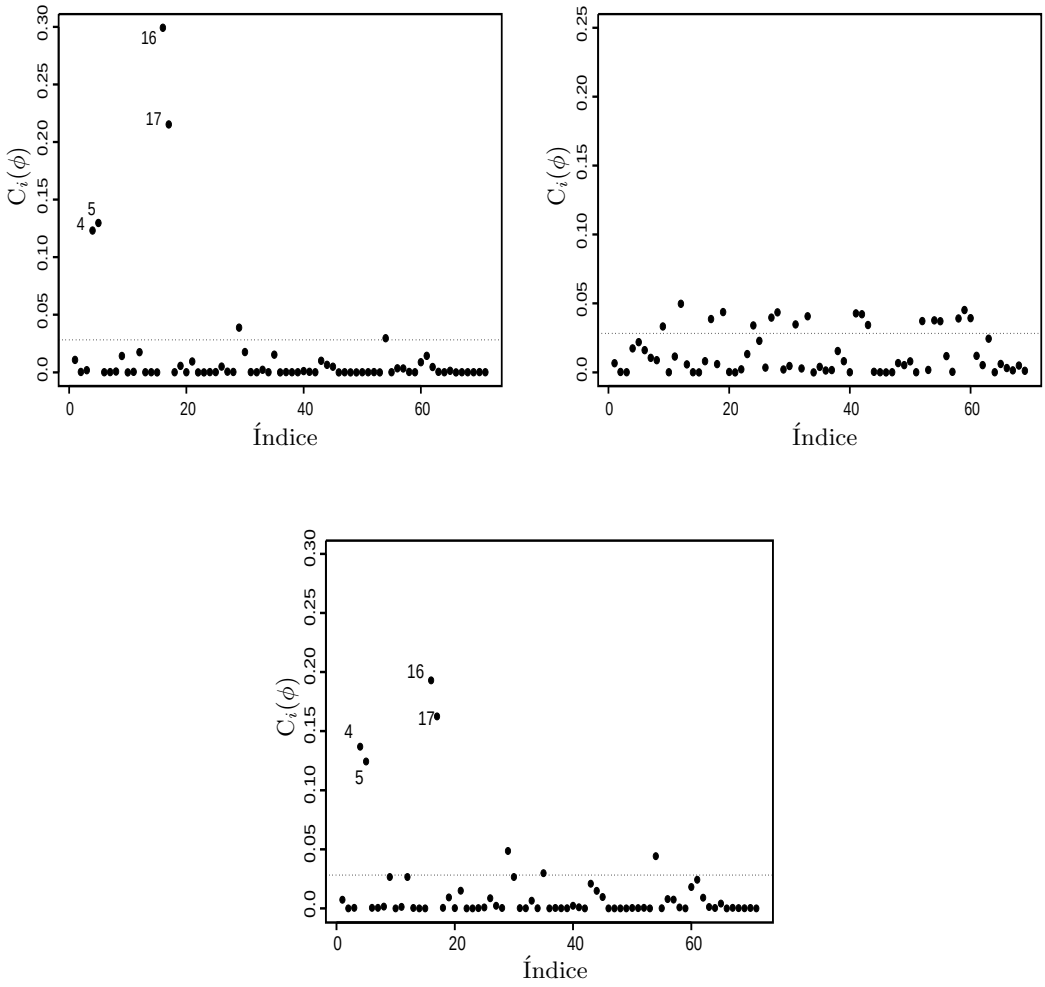

Figura 3.12 *Gráficos de índices de* $C_i(\phi)$ *para o modelo normal (esquerda), t−Student com 4 g.l. (direita) e logístico-II (abaixo) ajustados aos dados de coelhos.*

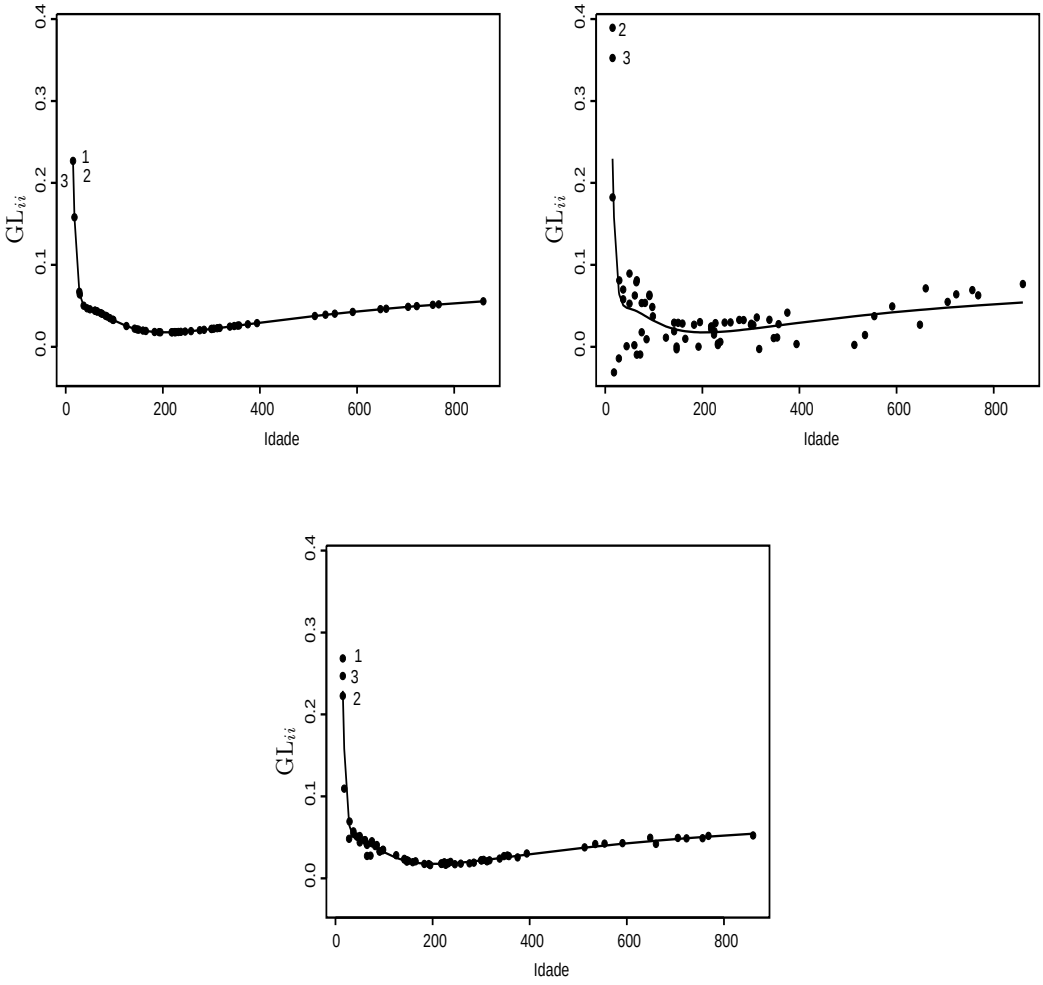

Figura 3.13 *Gráficos de pontos de alavanca generalizados contra a idade para o modelo normal (esquerda), t−Student com 4 g.l. (direita) e logístico-II (abaixo) ajustados aos dados de coelhos.*

CAPÍTULO 4

Extensões

4.1 Introdução

Extensões de modelos elípticos para dados correlacionados têm sido propostas por vários autores nos últimos anos. Por exemplo, sob erros com distribuição t de Student e estrutura longitudinal, Lange, Little e Taylor (1989) apresentam procedimentos de estimação e alguns resultados inferenciais com graus de liberdade conhecidos e desconhecidos, Arellano-Valle (1994) estuda o caso de modelos com erros nas variáveis, Welsch e Richardson (1997) estudam o caso de modelos mistos marginais, Kowalski, Mendoza-Blanco, Tu e Gleser (1999) comparam procedimentos inferenciais clássicos e bayesianos, Fernandez e Steel (1999) chamam a atenção de alguns cuidados necessários na estimação dos graus de liberdade sob esses dois enfoques e mais recentemente Pinheiro, Liu e Wu (2001) propõem uma estrutura hierárquica em que os erros e os efeitos aleatórios seguem distribuição t de Student com graus de liberdade desconhecidos sendo os parâmetros estimados através de algoritmo tipo EM, Galea, Bolfarine e Vilca-Labra (2002) desenvolvem métodos de diagnóstico em modelos com erros nas variáveis enquanto Cysneiros and Paula (2004) discutem estimação e testes com restrições nos parâmetros na forma de igualdades e desigualdades lineares. Sob outros erros elípticos, por exemplo, Huggins (1993) propõe estimadores robustos em modelos elípticos multivariados com aplicações na área de genética, Lindsey (1999) propõe modelos com erros exponencial potência para a análise de dados com medidas repetidas, Galea, Paula e Bolfarine (1997), Liu (2000, 2002) e Días-Garcia, Galea e Leiva-Sánchez (2003) derivam métodos de diagnóstico em modelos elípticos multivariados, enquanto Liu (2004) desenvolve métodos de influência local em modelos de séries temporais com estrutura heteroscedástica e Savalli, Paula e Cysneiros (2004) discutem a aplicação de

um teste tipo escore para testar os componentes de variância em modelos elípticos mistos.

4.2 Modelos elípticos mistos

Os modelos mistos são em geral expressos na forma abaixo

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{4.1}$$

$i = 1, \ldots, n$, em que $\mathbf{y}_i$ denota um vetor m_i-dimensionnal com as respostas observadas para a i-ésima unidade experimental, $\mathbf{X}_i$ é uma matriz $(m_i \times p)$ que contém os valores das variáveis explicativas, $\boldsymbol{\beta}$ denota o vetor de parâmetros fixos e $\mathbf{Z}_i$ é uma matriz $(m_i \times q)$ que contém a especificação dos efeitos aleatórios. É usual assumir que $\mathbf{b}_i \sim \mathrm{N}_q(\mathbf{0}, \mathbf{D})$ e $\boldsymbol{\epsilon}_i \sim \mathrm{N}_{m_i}(\mathbf{0}, \sigma^2\mathbf{I}_{m_i})$. Contudo, devido à falta de robustez da estimativa de máxima verossimilhança sob erros normais contra observações aberrantes, uma extensão natural dos modelos propostos no Capítulo 2 é assumir distribuições multivariadas para os erros com caudas mais pesadas do que a normal, tais como t de Student, exponencial potência, logística-II, dentre outras. Por exemplo, poderemos assumir que os erros $\mathbf{b}_i$ e $\boldsymbol{\epsilon}_i$ são tais que $(\mathbf{b}_i^T, \boldsymbol{\epsilon}_i^T)^T$ segue distribuição elíptica de média zero e matriz de dispersão $\mathbf{V}_i = \mathrm{diag}\{\mathbf{D}, \sigma^2\mathbf{I}_{m_i}\}$, denotaremos $(\mathbf{b}_i^T, \boldsymbol{\epsilon}_i^T)^T \sim \mathrm{El}_{m_i+q}(\mathbf{0}, \mathbf{V}_i)$. Isto significa que $\mathbf{b}_i$ e $\boldsymbol{\epsilon}_i$ são não correlacionados, porém não necessariamente independentes (exceto para o caso normal). Assim, podemos expressar

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim \mathrm{El}_{m_i+q}\left\{ \begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}; \begin{bmatrix} \sigma^2\mathbf{I}_{m_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^T & \mathbf{D} \end{bmatrix} \right\}, \; i = 1, \ldots, n. \tag{4.2}$$

Similar aos modelos normais mistos pode-se fazer inferências para os parâmetros do modelo misto elíptico através do modelo marginal. Segue de Fang, Kotz e Ng (1990) que a distribuição marginal de $\mathbf{y}_i$ é também elíptica assumindo a forma

$$\mathbf{y}_i \sim \mathrm{El}_{m_i}(\mathbf{X}_i\boldsymbol{\beta}; \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{m_i}). \tag{4.3}$$

A função densidade de $\mathbf{y}_i$ é dada por

$$f(\mathbf{y}_i) = |\boldsymbol{\Sigma}_i|^{-1/2}\, g(u_i),$$

$i = 1, \ldots, n$, em que $u_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$ com $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}_{m_i}$, $g(\cdot) : I\!R \to [0, \infty]$ tal que $\int_0^\infty u^{m_i/2-1} g(u) du < \infty$, chamada de função geradora de densidades como no caso univariado, $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ e $\boldsymbol{\Sigma}_i$ é proporcional à matriz de variância-covariância de $\mathbf{y}_i$. Por simplicidade é usual assumir que $\mathbf{D} = \text{diag}\{\tau_1, \ldots, \tau_q\}$. Assim, os parâmetros a serem estimados são dados por $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\tau}^T)^T$, em que $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^T$. Um processo iterativo conjunto para estimar os parâmetros fixos e os componentes de variância é dado por

$$\boldsymbol{\beta}^{(m+1)} = \left[\sum_{i=1}^{n} v_i^{(m)} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-(m)} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^{n} v_i^{(m)} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-(m)} \mathbf{y}_i \right]$$

e

$$\boldsymbol{\alpha}^{(m+1)} = \text{argmax}_{\boldsymbol{\alpha}} \{ L(\boldsymbol{\beta}^{(m+1)}, \boldsymbol{\alpha}) \},$$

para $m = 0, 1, 2, \ldots$, em que $\boldsymbol{\alpha} = (\sigma^2, \boldsymbol{\tau}^T)^T$, $v_i = -2 \frac{g'(u_i)}{g(u_i)}$ e $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ denota o logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$. A matriz de informação de Fisher para $\boldsymbol{\theta}$ assume a forma bloco diagonal

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\alpha}) \end{pmatrix}, \tag{4.4}$$

em que

$$\mathbf{K}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{4 d_g}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \text{ e } \mathbf{K}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{K}_i(\boldsymbol{\alpha}).$$

O elemento (r, s) da matriz $\mathbf{K}_i(\boldsymbol{\alpha})$ é dado por

$$\text{K}_{i,rs}(\boldsymbol{\alpha}) = \frac{b_{rs}}{4} \Big(\frac{4 f_g}{m_i(m_i+2)} - 1 \Big) + \frac{2 f_g}{m_i(m_i+2)} \text{tr} \Big(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \alpha_r} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \alpha_s} \Big),$$

em que $d_g = E\{W_g^2(U) U\}$, $f_g = E\{W_g^2(U) U^2\}$ com $U = ||\mathbf{Z}||^2$, $\mathbf{Z} \sim \text{El}_{m_i}(\mathbf{0}, \mathbf{I}_{\mathbf{m_i}})$ e $b_{rs} = \text{tr}(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \alpha_r)$ $\text{tr}(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \alpha_s)$. Expressões em forma fechada para d_g e f_g podem ser obtidas para algumas distribuições elípticas multivariadas. Como a inferência para os modelos elípticos mistos é similar à inferência para os modelos normais mistos é razoável supor, por exemplo, que sob certas condições de regularidade e para amostras grandes $\hat{\boldsymbol{\beta}}$ tenha distribuição aproximadamente normal de média $\boldsymbol{\beta}$ e matriz de variância-covariância dada por $\mathbf{K}^{-1}(\boldsymbol{\beta})$.

4.3 Modelos elípticos multivariados

Dizemos que uma matriz aleatória $(n \times p)$ $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_p)^T$ tem uma distribuição elíptica, com matriz de locação $(n \times p)$ $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_p)^T$ e matriz de escala $(np \times np)$ $\boldsymbol{\Sigma} \otimes \boldsymbol{\Phi}$, $\boldsymbol{\Sigma} > \mathbf{0}$ e $\boldsymbol{\Phi} > \mathbf{0}$, com $\boldsymbol{\Sigma}$ sendo uma matriz $(p \times p)$, $\boldsymbol{\Phi}$ uma matriz $(n \times n)$, se sua densidade é dada por

$$f(\mathbf{Y}) = |\boldsymbol{\Sigma}|^{-n/2}|\boldsymbol{\Phi}|^{-p/2}\, g\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})^T\boldsymbol{\Phi}^{-1}(\mathbf{Y}-\boldsymbol{\mu}))\}, \tag{4.5}$$

em que a função $g : I\!R \longrightarrow [0, \infty)$ é tal que $\int_0^\infty u^{np/2-1}g(u)\, du < \infty$ e $\otimes$ denota o produto usual de Kronecker. A função $g(\cdot)$ é conhecida como função geradora de densidades. Neste caso usamos a notação $\mathbf{Y} \sim \mathrm{EC}_{(n\times p)}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$. Note que $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ pode ser visto como uma amostra de uma população elíptica p-dimensional. Esta classe de distribuições inclui a normal, t de Student, normal contaminada, logísticas I e II e exponencial potência, dentre outras.

Consideremos agora o modelo de regressão linear multivariado

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4.6}$$

em que $\mathbf{Y}$ é uma matriz $(n \times p)$ de respostas, $\mathbf{X}$ é uma matriz conhecida $(n \times q)$ de posto q, $\boldsymbol{\beta}$ é uma matriz $(q \times p)$ de parâmetros e $\boldsymbol{\epsilon}$ é uma matriz de erros $(n \times p)$ com distribuição $\mathrm{EC}_{(n\times p)}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, em que $\boldsymbol{\Sigma} > \mathbf{0}$ é a matriz de escala $(p \times p)$ e sua densidade é dada por (4.5), com $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. Este é denominado modelo de regressão linear elíptico multivariado. Se $g(\cdot)$ é uma função contínua e decrescente, então os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}$ são dados por (ver Gupta e Varga, 1993, pp. 285-286)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad \text{e} \quad \widehat{\boldsymbol{\Sigma}} = u_0\, \mathbf{Q}(\hat{\boldsymbol{\beta}}), \tag{4.7}$$

em que $\mathbf{Q}(\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ e u_0 maximiza a função $h_n(u) = u^{-np/2}\, g(p/u)$, $u \geq 0$. Se $g(\cdot)$ é uma função contínua e decrescente, então seu máximo u_0 existe é finito e positivo. É facil ver que para as distribuições normal e t de Student $u_0 = 1/n$. Para a exponencial potência $(g(u) = c\, \exp(-u^s/2))$ $u_0 = (s\, p^{s-1}/n)^{1/s}$.

No entanto, para as distribuições normal contaminada e logísticas I e II, u_0 deve ser obtido numericamente. Usando propriedades das distribuições elípticas temos que

$$\hat{\boldsymbol{\beta}} \sim \mathrm{EC}_{p\times q}(\boldsymbol{\beta}, \boldsymbol{\Sigma} \otimes (\mathbf{X}^T\mathbf{X})^{-1}). \tag{4.8}$$

Além disso, o teste da razão de verosimilhanças para testar H : $\mathbf{A}\,\boldsymbol{\beta} = \mathbf{M}$, em que $\mathbf{A}$ é uma matriz $(t \times q)$ de posto t e $\mathbf{M}$ é uma matriz $(t \times p)$ de constantes, é dado por (ver Gupta e Varga, 1993, pp. 297-299)

$$\lambda = \left(\frac{\det(\mathbf{Q}(\hat{\boldsymbol{\beta}}))}{\det(\mathbf{Q}(\hat{\boldsymbol{\beta}}) + \mathbf{T})} \right)^{-n/2}, \tag{4.9}$$

e tem a mesma distribuição do caso normal, em que

$$\mathbf{T} = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{M})^T(\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{M}).$$

4.4 Modelos elípticos assimétricos

Embora os modelos de regressão com erros elípticos representem uma boa proposta alternativa ao modelo normal, existem situações práticas em que a resposta é assimétrica (vide, por exemplo, Hill e Dixon, 1982). Nestes casos, uma nova classe de distribuições denominada *skew-normal* (normal assimétrica) proposta por Azzalini (1985, 1986) poderia ser postulada para os erros. A função de densidade de probabilidade da *skew-normal* possui um parâmetro que permite controlar a assimetria e, conseqüentemente, se ajustar melhor a dados de natureza assimétrica. Azzalini e Dalla Valle (1996) estenderam os resultados da *skew-normal* univariada para o caso multivariado e diversos autores têm estendido esses resultados para situações mais gerais. Uma excelente revisão de distribuições elípticas assimétricas pode ser encontrada, por exemplo, em Genton (2004).

APÊNDICE A

Arquivos de Dados

Tabela A.1 *Rentabilidades mensais das ações da empresa Concha & Toro, IPSA e Taxas de juros mensais do banco central chileno.*

Ano	y_{rt}	r_{mt}	r_{ft}	Ano	y_{rt}	r_{mt}	r_{ft}
1990	0,028	0,025	0,044	1991	0,035	0,009	0,190
	0,081	0,032	0,143		1,009	0,006	0,202
	0,015	0,016	0,013		0,078	0,006	0,084
	0,000	0,023	-0,00		0,033	0,009	0,026
	0,000	0,025	-0,00		0,071	0,013	0,056
	0,050	0,021	-0,02		0,397	0,016	0,187
	0,017	0,024	0,009		0,085	0,014	0,142
	0,012	0,021	0,000		0,052	0,014	0,125
	0,055	0,023	-0,00		0,410	0,012	0,151
	0,023	0,030	0,036		0,206	0,011	-0,058
	0,089	0,031	0,208		-0,255	0,017	-0,039
	0,290	0,017	0,125		-0,113	0,013	-0,002
1992	0,104	0,008	-0,028	1993	0,131	0,005	0,117
	0,044	0,008	0,165		-0,022	0,003	-0,020
	0,604	0,006	0,096		-0,046	0,004	-0,021
	-0,116	0,006	0,007		-0,037	0,006	-0,056
	-0,169	0,012	0,008		0,017	0,010	0,015
	0,088	0,012	0,044		-0,025	0,010	0,099
	-0,008	0,010	0,023		-0,051	0,008	0,016
	-0,195	0,008	-0,064		-0,053	0,007	0,062
	0,002	0,010	-0,083		-0,014	0,011	0,043
	0,016	0,015	0,095		0,140	0,014	0,049
	0,000	0,010	-0,071		0,054	0,019	0,065
	-0,007	0,009	0,030		0,054	0,009	0,190

Ano	y_{rt}	r_{mt}	r_{ft}	Ano	y_{rt}	r_{mt}	r_{ft}
1994	0,030	0,005	0,181	1995	-0,100	0,005	-0,048
	0,093	0,010	-0,014		0,103	0,007	-0,035
	-0,086	0,007	-0,118		0,050	0,006	-0,019
	-0,078	0,011	0,096		0,069	0,007	0,070
	-0,061	0,009	0,092		0,045	0,007	0,082
	-0,026	0,011	-0,039		-0,083	0,006	0,023
	-0,007	0,008	0,006		-0,071	0,007	-0,041
	-0,040	0,006	0,117		-0,007	0,007	-0,038
	0,021	0,009	0,050		0,018	0,010	-0,023
	0,069	0,008	0,108		0,033	0,008	0,036
	-0,013	0,007	-0,024		-0,028	0,009	-0,049
	-0,078	0,006	-0,056		0,022	0,007	0,067
1996	0,064	0,006	-0,006	1997	0,190	0,007	0,105
	-0,077	0,006	0,000		0,067	0,007	0,049
	-0,018	0,007	-0,041		-0,035	0,008	-0,020
	0,045	0,009	0,042		0,150	0,006	0,046
	-0,036	0,011	-0,020		0,019	0,006	0,093
	0,082	0,010	0,046		-0,015	0,004	0,013
	0,062	0,009	-0,023		-0,075	0,004	0,019
	0,026	0,007	-0,034		-0,061	0,006	-0,051
	0,019	0,007	0,026		0,127	0,005	0,005
	0,012	0,008	0,001		-0,147	0,007	-0,075
	0,173	0,008	-0,061		-0,023	0,009	-0,026
	0,000	0,008	-0,048		-0,041	0,005	-0,020
1998	0,220	0,008	-0,108	1999	-0,032	0,007	0,030
	0,020	0,007	0,104		0,187	0,003	0,061
	0,106	0,004	0,061		-0,012	0,004	0,084
	0,057	0,006	-0,063		0,141	0,008	0,009
	-0,144	0,007	-0,086		0,028	0,006	-0,045
	0,022	0,008	-0,054		0,162	0,004	0,121
	0,161	0,011	0,049		0,000	0,004	0,008
	-0,333	0,008	-0,299		0,079	0,003	-0,006
	-0,019	0,017	0,057		-0,029	0,004	-0,024
	0,231	0,011	0,096		-0,047	0,004	0,004
	0,146	0,011	0,131		0,085	0,005	0,072
	-0,094	0,007	-0,051		0,026	0,004	0,063

Ano	y_{rt}	r_{mt}	r_{ft}	Ano	y_{rt}	r_{mt}	r_{ft}
2000	0,048	0,004	0,023	2001	0,050	0,004	0,050
	-0,125	0,005	-0,039		0,043	0,005	-0,037
	-0,010	0,007	0,021		0,013	0,002	-0,031
	0,039	0,007	-0,049		-0,010	0,005	0,041
	-0,021	0,008	0,046		0,168	0,007	0,093
	0,053	0,005	-0,014		-0,005	0,006	-0,032
	0,014	0,004	-0,031		0,098	0,004	0,027
	-0,010	0,004	0,045		-0,024	0,005	0,032
	0,025	0,005	-0,012		-0,031	0,005	-0,088
	0,033	0,007	-0,044		-0,034	0,005	-0,008
	0,014	0,008	0,024		0,000	0,005	0,063
	0,023	0,007	0,000		-0,158	0,005	-0,008
2002	-0,064	0,005	-0,025	2003	-0,014	0,002	0,002
	0,027	0,005	-0,007		0,006	0,002	0,013
	0,043	0,004	0,017		-0,013	0,002	-0,006
	-0,064	0,004	-0,022		0,130	0,002	0,153
	0,032	0,003	-0,033		0,110	0,002	0,058
	0,015	0,003	-0,062		0,033	0,002	-0,002
	-0,050	0,002	0,007		0,073	0,002	0,076
	0,066	0,002	-0,017		0,008	0,002	0,037
	-0,076	0,002	-0,093		-0,087	0,002	0,044
	0,063	0,002	0,034		-0,008	0,002	0,060
	-0,031	0,002	0,002		-0,074	0,002	-0,038
	0,126	0,002	0,041		-0,086	0,002	0,018
2004	-0,049	0,002	-0,051				
	0,165	0,001	0,095				
	-0,023	0,001	-0,055				
	-0,036	0,001	-0,019				
	0,090	0,001	-0,001				
	0,058	0,001	0,038				

Tabela A.2 *Grau de luminosidade dos produtos A,B,C,D e E durante 20 semanas.*

Semana	A	B	C	D	E
	77,86	76,36	77,09	75,63	76,86
1	77,70	76,91	77,16	76,72	76,73
	78,13	77,09	77,14	76,53	76,62
	76,55	75,61	76,27	75,08	74,73
3	76,49	75,10	74,78	75,76	74,49
	76,56	75,78	74,61	73,78	75,23
	66,87	71,32	71,34	68,93	71,67
5	65,99	72,72	71,24	68,91	69,94
	66,45	71,58	72,15	70,04	69,94
	70,28	69,91	68,91	68,70	69,06
7	67,82	69,84	69,17	69,27	67,03
	70,15	69,87	69,76	67,06	70,01
	69,60	68,98	68,38	69,65	69,41
9	69,75	71,13	69,03	68,81	69,27
	70,79	69,17	68,71	70,41	69,05
	71,71	66,72	70,72	68,97	69,00
11	71,66	69,80	70,04	70,47	69,88
	70,83	68,87	70,25	69,04	70,14
	69,05	69,85	69,63	68,26	67,18
13	70,12	69,61	68,71	68,50	68,07
	70,18	70,22	69,11	67,66	69,19
	69,87	70,14	68,89	65,27	67,27
15	69,78	69,15	66,70	65,67	68,91
	69,90	71,39	69,42	66,85	68,90
	70,79	70,69	70,63	70,15	70,15
17	70,21	71,97	69,27	69,65	70,00
	69,13	72,27	69,87	71,50	69,88
	69,82	69,63	70,35	69,03	69,86
19	68,75	69,02	69,23	68,24	68,62
	68,79	70,26	67,92	68,77	69,69

Tabela A.3 *Pesos das lentes dos olhos de coelhos europeus (y), em miligramas e idade (x) em dias numa amostra de 71 observações (Ratkowsky, 1983, Tabela 6.1).*

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

Referências

Albert, J.; Delampady, M. e Polasek, W. (1991). A class of distributions for robustness studies. *Journal of Statistical Planning and Inference*, **28**, 291-304.

Anderson, T.W. e Fang, K.T. (1987). Cochran's theorem for elliptically contourned distributions. *Sankhya* A, **49**, 305-315.

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.

Arellano-Valle, R.; Galea, M. e Iglesias, P. (2003). *Bayesian analysis in elliptical CAPM in the Chilean Stock Market*. (Submetido a publicação)

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

Atkinson, A.C. (1985). *Plots, Transformation and Regression*. Oxford: Clarendon Press.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46**, 199-208.

Azzalini, A. e Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Bates, D.M. e Watts, D.G. (1988). *Nonlinear Regression Analysis and its Applications*. New York: John Wiley.

Blattberg R.C. e Gonedes N.J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, **47**, 244-280.

Becker, R.A.; Chambers, J.M. e Wilks, A.R. (1988). *The New S Language*. New York: Chapman and Hall.

Beering, M. (1999). Techniques for measuring color. *Metal Finishing*, **97**, 552-557.

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

Breusch, T.S. e Pagan, A.R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287-1294.

Box, M.J. e Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley.

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

Campbell, J.; Lo, A. e MacKinlay, A. (1997). *Econometrics of Financial Markets*. New Jersey: Princeton University Press.

Chambers, J.M. e Hastie, T.J. (1992). *Statistical Models in S*. New York: Chapman and Hall.

Chmielewski, M.A. (1981). Elliptically symmetric distributions: a review and bibliography. *International Statistical Review*, **49**, 67-74.

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

Cook, R.D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. London: Chapman and Hall.

Cook, R.D. e Tsai, C.L. (1985). Residuals in nonlinear regression. *Biometrika*, **72**, 23-29.

Cook, R.D.; Tsai, C.L. e Wei, B.C. (1986). Bias in nonlinear regression. *Biometrika*, **73**, 615-623.

Cordeiro, G.M. (2004). Corrected LR tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*. (Aceito para publicação)

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. *Journal of the Royal Statistical Society* B, **30**, 248-275.

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. (2004). Local influence and residual analysis in heteroscedastic symmetrical linear models. *Statistical Modelling: Proceedings of the 19th International Workshop on Statistical Modelling*, Biggeri, A.; Dreassi, E.; Lagazio, M.M. (Eds). Firenze: Firenze University Press, pp. 376-380.

Cysneiros, F.J.A. e Paula, G.A. (2004). One-sided test in linear models with multivariate $t-$ distribution. *Communication in Statistics-Simulation and Computation*, **33**, 747-772.

Cysneiros, F.J.A. e Paula, G.A. (2005). Restricted methods in symmetrical linear regression models. *Computational Statistics and Data Analysis*. (Aceito para publicação)

Devlin, S.J.; Gnanadesikan, R. e Kettenring, J.R. (1976). Some multivariate applications of elliptical distributions. *Essays in Probability and Statistics*, **24**, 365-393. Tokyo: Shinko Tsusho Co.

Devroye, L. (1986). *Non-Uniform Random Variable Generator*. New York: Springer-Verkag.

Díaz-García, J.A.; Galea, M. e Leiva-Sánchez, V. (2003). Infuence diagnostics for elliptical multivariate linear regression models. *Communications in Statistics - Theory and Methods*, **32**, 625-641.

Dickey, J.M. (1967). Multivariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Annals of Mathematical Statistics*, **38**, 511-518.

Elton E. e Gruber M. (1995). *Modern Portfolio Theory and Investment Analysis*. New York: Wiley.

Emerson, J.D.; Hoaglin, D.C. e Kempthorne, P.J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables. *Journal of the American*

Statistical Association, **79**, 329-335.

Escobar, L.A. e Meeker, W.Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507-528.

Fama E. (1965). The behavior of stock market prices. *Journal of Business*, **38**, 34-105.

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

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

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

Fernandez, C. e Steel, M.F.J. (1999). Multivariate Student-t regression models: pitfalls and inference. *Biometrika*, **86**, 153-167.

Ferrari, S.L.P. e Arellano–Valle, R.B. (1996). Bartlett corrected tests for regression models with Student-t independent errors . *Brazilian Journal of Probability and Statistics*, **10**, 15-33.

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

Galea, M.; Bolfarine, H. e Vilca–Labra, F. (2002). Influence diagnostics for the structural error-in-variables model under the Student-t distribution. *Journal of Applied Statistics*, **29**, 1191-1204.

Galea, M.; Paula, G.A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71-79.

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

Genton, M.G. (2004). *Skew-Elliptical Distribution and Their Application: A Journey Beyond Normality*. London: Chapman and Hall/CRC.

Goldfeld, S.M. e Quandt, R.E. (1965). Some test for homoscedasticity. *Journal of the American Statistical Association*, **60**, 539-547.

Gumbel, E. (1944). Ranges and midranges. *Annals of Mathematical Statistics*, **15**, 414-422.

Gupta, A. K. e Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Kluwer Academic Publishers.

Hastings, N.A.J. e Peacock, J.B. (1975). *Statistical Distributions.* New York: John Wiley.

Hill, M.A. e Dixon, W.J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrika*, **38**, 377-396.

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

Huggins, R.M. (1993). On the robust analysis of variance components models for pedigree data. *Australian Journal of Statistics*, **35**, 43-57.

Ihaka, R. e Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational Graphical and Statistics*, **5**, 299-314.

Johnson, R. e Kotz, S. (1970). *Continuous Univariate Distributions V.2*. Boston: Houghton Mifflin.

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

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

Kotz S. (1975). Multivariate distributions at a cross-road. *Statistical Distributions in Scientific Work, 1* Ed. G.P. Patil, S. Kotz and J.K. Ord., pp. 247-270. Dordrecht, Reiden.

Lange, K.L.; Little, R.J.A. e Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881-896.

Lesaffre, F. e Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **38**, 963-974.

Lindsey, J.K. (1999). Multivariate elliptically-contoured distributions for repeated measurements. *Biometrics*, **55**, 1277-1280.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **41**, 13-37.

Little, R.J.A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, **37**, 23-39.

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers*, **41**, 211-224.

Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra and Its Applications*, **354**, 159-174.

Liu, S. (2004). On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *Applied Probability*, **41A**, 393-405.

Manoukian, E.B. (1985). *Modern Concepts and Theorems of Mathematical Statistics*. New York: Springer-Verlag.

Mossin, J. (1966). Equilibrium in capital asset market. *Econometrica*, **35**, 768-783.

Muirhead, R. (1980). The effects of symmetric distributions on some standard procedures involving correlation coefficients. In: *Multivariate Statistical Analysis* (Ed. R.P. Gupta) North-Holland, pp. 143-159.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley.

Paula, G.A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis*, **16**, 63-79.

Paula, G.A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation*, **51**, 315-331.

Paula, G.A. (1999). Leverage in inequality constrained regression models. *The Statistician*, **48**, 529-538.

Paula, G.A., Cysneiros, F.J.A. e Galea, M. (2003). Local influence and leverage in elliptical nonlinear regression models. In: *Proceedings of the 18th International Workshop on Statistical Modelling*, Verbeke, G., Molenberghs, G., Aerts, A. and

Fieuws, S. (Eds.). Leuven: Katholieke Universiteit Leuven, pp. 361-365.

Paula, G.A.; de Moura, A.S. e Yamaguchi, A.M. (2004). Relatório de Análise Estatística sobre o Projeto: *Estabilidade Sensorial de Snacks Aromatizados com Óleo de Canola e Gordura Vegetal Hidrogenada*. RAE-CEA 04105, IME-USP.

Pinheiro, J.C.; Liu, C. e Wu, Y.N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**, 249-276.

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.

Savalli, C.; Paula, G.A. e Cysneiros, F.J.A. (2004). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling: Proceedings of the 19th International Workshop on Statistical Modelling*, Biggeri, A.; Dreassi, E.; Lagazio, M.M. (Eds). Firenze: Firenze University Press, pp. 144-148.

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

Sharpe, W. (1964). Capital asset prices : A theory of markets equilibrium under conditions of risk. *Journal of Finance*, **19**, 425-442.

Smyth, G.K. (1996). Partitioned algorithms for maximum likelihood and other nonlinear estimation. *Statistics and Computing*, **6**, 201-216.

St. Laurent, R.T. e Cook, R.D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, **87**, 985-990.

Taylor, J.M.G. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics and Data Analysis*, **13**, 33-46.

Thomas, W. e Cook. R.D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics*, **32**, 59-65.

Uribe–Opazo, M.A. (1997). *Aperfeiçoamento de Testes Estatísticos em Várias Famílias de Distribuições*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil.

Uribe–Opazo, M.A.; Ferrari, S.L.P e Cordeiro, G.M. (2003). Improved score tests in symmetric linear regression models. *Relatório Técnico* RT-MAE 2003-05, IME-USP.

Welsh, A.H. e Richardson, A.M. (1997). *Approaches to the Robust Estimation of Mixed Models*. Vol. 15 of Maddala and Rao (1997), pp. 343-384.

Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer-Verlag.

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817-838.

Yamaguchi, K. (1990). Generalized EM algorithm for model with contaminated error term. In: *Proceedings of the Seven Japan and Korea Joint Conference of Statistics*, pp. 107-114

Zhou G. (1993). Asset-pricing test under alternative distributions. *The Journal of Finance*, **48**, 1927-1942.