
MODELOS LINEARES GENERALIZADOS SIMÉTRICOS HETEROSCEDÁSTICOS

MARIA LÍDIA COCO TERRA

Orientador: Prof. Dr. Francisco José A. Cysneiros

Área de Concentração: Estatística Matemática

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

Recife, fevereiro de 2009

Terra, Maria Lídia Coco
Modelos lineares generalizados simétricos
heteroscedásticos / Maria Lídia Coco Terra - Recife :
O Autor, 2009.
ix, 48 folhas : il., fig., tab.

Dissertação (mestrado) – Universidade Federal
de Pernambuco. CCEN. Estatística, 2009.

Inclui bibliografia e apêndice.

1. Estatística matemática. I. Título.

519.536 CDD (22.ed.) MEI2009-032

Universidade Federal de Pernambuco

Pós-Graduação em Estatística

26 de fevereiro de 2009
(data)

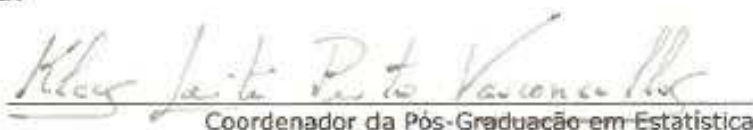
Nós recomendamos que a dissertação de mestrado de autoria de

Maria Lídia Coco Terra

intitulada

"Modelos Lineares Generalizados Simétricos Heteroscedásticos"

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


Coordenador da Pós-Graduação em Estatística

Banca Examinadora:


Francisco José de Azevedo Cysneiros orientador


Juvêncio Santos Nobre (UFCE)


Audrey Helen Mariz de Aquino Cysneiros

Este documento será anexado à versão final da dissertação.

*Aos meus pais, Paulo e Lia, e a minha irmã, Lúcia
pelo apoio incondicional.*

Agradecimentos

Aos meus pais e minha irmã, pelo carinho, dedicação, confiança, paciência e por terem contribuído para a minha formação moral e acadêmica.

À minha família e amigos que ficaram na Bahia, Camila, Dorinny, Tarsilla, Larissa e Iano pela enorme paciência, apoio e compreensão com a minha ausência.

Ao professor Francisco José A. Cysneiros, pela incrível paciência, atenção e excelente orientação.

Aos professores do mestrado, os quais me ensinaram sobre Estatística e me receberam bem aqui no mestrado.

Aos meus amigos do mestrado, Juliana, Larissa, Cecílio, Olga, Raphael, Marcelo, Carlos, Hemílio, Fábio, Abraão, Valmir, Marcelle, Wallace, Wilton, Andréia, Izabel, Alice e Wagner pela amizade, ajuda, apoio e pelos momentos de alegria.

À Valéria Bittencourt, pelo enorme carinho, paciência e amizade com que sempre tratou a mim e aos demais alunos do mestrado.

Em especial a minha irmã Lígia, Juliana, Cecílio e Larissa pelo tempo maravilhoso que convivemos aqui em Recife.

Ao CNPq, pelo apoio financeiro.

Resumo

A classe de modelos lineares generalizados simétricos heteroscedásticos é caracterizada pela resposta pertencente à classe simétrica de distribuições, um preditor linear relacionado com o parâmetro de posição e um preditor linear relacionado com o parâmetro de escala, ambos relacionados através de funções de ligação. Estes modelos são uma extensão natural dos modelos lineares generalizados simétricos (MLGS) proposta por Lobos (2004), em que apenas o parâmetro de posição é relacionado por um preditor linear através de uma função de ligação duplamente diferenciável. Neste trabalho, inicialmente, apresentamos alguns resultados nos MLGS heteroscedásticos. Em seguida, testes assintóticos para avaliar homoscedasticidade são propostos bem como métodos de diagnósticos sob o enfoque de influência local. Alguns exemplos com dados reais serão discutidos a fim de ilustrar a teoria desenvolvida.

Palavras-chave: modelos lineares generalizados simétricos heteroscedásticos, influência local, parâmetro de escala.

Abstract

The heteroscedastic symmetrical generalized linear models class is describe for a response that belongs to a symmetrical class of distributions, a linear predictor relate to a location parameter and a linear predictor associate to a dispersion parameter, both link functions. These models are natural extensions of symmetrical generalized linear models propose by Lobos (2004), where only a location parameter is associate to a one-to-one continuously differentiable positive link function. First of all, in this work we present some results for heteroscedastic symmetrical generalized linear models. Then, asymptotic tests to evaluate homogeneity are proposed as well as diagnostic methods based on local influence. Some examples with real data are discuss to illustrate the developed theory.

Keywords: heteroscedastic symmetrical generalized linear models, local influence, dispersion parameter.

Lista de Figuras	vii
Lista de Tabelas	ix
1 Introdução	1
1.1 Introdução	1
2 Modelos Lineares Generalizados Simétricos Heteroscedásticos	6
2.1 Introdução	6
2.2 Função de Máxima Verossimilhança e Função Escore	7
2.3 Informação de Fisher	9
2.4 Estimação dos parâmetros	12
2.5 Teste de Homogeneidade	14
2.6 Resíduos	17
2.7 Métodos de Diagnóstico	20
2.7.1 Alavancagem Generalizada	20
2.7.2 Influência local	23
3 Aplicação do Método de Influência Local	28

4	Conclusões	43
A	Escolaridade	44
B	Crescimento de crianças	45
	Referências Bibliográficas	46

Lista de Figuras

1.1	Gráfico da f.d.p. das distribuições normal padrão, t -Student com 5 graus de liberdade, Logística e Exponencial potência, respectivamente.	3
1.2	Gráfico de dispersão dos dados de escolaridade e renda.	4
1.3	Gráfico de dispersão do crescimento de crianças.	5
3.1	Gráfico de dispersão de escolaridade e renda.	29
3.2	Gráficos de diagnósticos para o modelo normal linear simples.	30
3.3	Resíduos para a distribuição normal.	32
3.4	Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição Normal.	32
3.5	Resíduos para a distribuição t -Student(3).	33
3.6	Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição t -Student(3).	34
3.7	Gráfico de dispersão do crescimento de crianças e idade (meses).	36
3.8	Gráficos de diagnósticos para o modelo normal linear simples.	37
3.9	Resíduos para a distribuição normal.	38
3.10	Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição Normal.	39
3.11	Resíduos para a distribuição t -Student(5).	40

3.12 Gráfico de índice do C_i com perturbação de caso ponderado para a distribuição t -Student(5).	40
3.13 Gráficos de índice do C_i com perturbação de caso resposta aditiva para a distribuição Normal, a esquerda, e t -Student(5), a direita.	41

Lista de Tabelas

3.1	Critério de informação de Akaike.	31
3.2	Estimação dos parâmetros e erros padrão associados.	31
3.3	Razão de mudança para os estimadores máxima verossimilhança.	34
3.4	Estimativas de máxima verossimilhança dos estimadores.	35
3.5	Critério de informação de Akaike.	37
3.6	Estimação dos parâmetros e erros padrão associados.	38
3.7	Razão de mudança para os estimadores máxima verossimilhança.	41
3.8	Estimativas de máxima verossimilhança dos estimadores.	42
A.1	Conjunto de dados sobre Escolaridade e Renda.	44
B.1	Conjunto de dados sobre Razão(peso/altura) e idade(meses) de crianças.	45

1.1 Introdução

É muito comum ao realizar-se uma análise de regressão a suposição de normalidade para os erros. Mas nem sempre estes modelos se ajustam bem aos dados, principalmente quando existem observações discrepantes. Assim, alternativas à suposição de erros normais têm sido propostas na literatura.

Quando a suposição de normalidade se tornava muito inadequada, tentava-se algum tipo de transformação para se atingir a normalidade. A transformação mais conhecida foi proposta por Box e Cox (1964). Mas esta transformação nem sempre produzia um efeito satisfatório. Com o avanço computacional, outros modelos foram criados para modelagem de dados que não possuem características de normalidade.

Em 1972, Nelder e Wedderburn propuseram os modelos lineares generalizados, em que a distribuição da variável resposta pertence à família exponencial, e a relação entre a média da variável resposta e o preditor linear η é feita por uma função monótona e diferenciável, denominada função de ligação. Eles também propuseram um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio, que é muito utilizado na avaliação do ajuste dos modelos, e contribuíram com o desenvolvimento de resíduos e medidas de diagnósticos.

O problema de se modelar variâncias foi amplamente discutido na literatura estatística, prin-

principalmente na área de econometria (ver Park (1966); Harvey (1976)). Por exemplo, sob erros normais Cook e Weisberg (1983) e Atkinson (1985) apresentam alguns métodos gráficos para se detectar heteroscedasticidade. Atkinson (1987) apresenta macros GLIM para se modelar a variância heterogênea. Carrol e Rupert (1988) desenvolvem procedimentos de diagnósticos usando métodos de influência local para estimar o parâmetro da variância em vários modelos não lineares para a média, enquanto que Verbyla (1993) compara os estimadores de máxima verossimilhança residual com e sem a retirada de uma observação através do desvio da verossimilhança. Já Smyth (1989) descreve um método que permite modelar o parâmetro de dispersão em alguns modelos lineares generalizados enquanto que Botter e Cordeiro (1998) apresentam expressões de correção de Bartlett para melhorar o teste de razão de verossimilhanças em modelos lineares generalizados com covariáveis de dispersão. Barroso e Vasconcellos (2002) apresentam expressões para aperfeiçoamento do teste escore em modelos t -Student heterocedásticos e mais recentemente Taylor e Verbyla (2004) propuseram uma modelagem conjunta de parâmetros de localização e escala em modelos t -Student.

Definição 1.1.1. *Uma variável aleatória y tem distribuição simétrica, ou seja, $Y \sim S(\mu, \phi, a)$, se sua função de densidade é dada da seguinte forma:*

$$f_Y(y, \mu, \phi) = \frac{1}{\sqrt{\phi}} a\left(\frac{(y - \mu)^2}{\phi}\right), \quad (1.1)$$

em que $\mu \in \mathbb{R}$, $\phi > 0$ e a função $a(\cdot)$ é uma função conhecida como geradora de densidade, em que $a : \mathbb{R} \rightarrow [0, \infty)$ satisfazendo $\int_0^\infty a(u) du < \infty$.

As distribuições simétricas mais conhecidas são a normal ou gaussiana e a t -Student, mas também existem outras distribuições simétricas, como por exemplo, a exponencial potência, a Cauchy, a Logística entre outras.

Estas distribuições são chamadas de simétrica porque o gráfico de sua função densidade de probabilidade é uma função simétrica em torno de seu parâmetros de localização μ , isto é, $\forall x \in \mathbb{R}$, $f_Y(\mu - y, \mu, \phi) = f_Y(\mu + y, \mu, \phi)$. Alguns exemplos de gráficos de densidades se encontram na Figura 1.1.

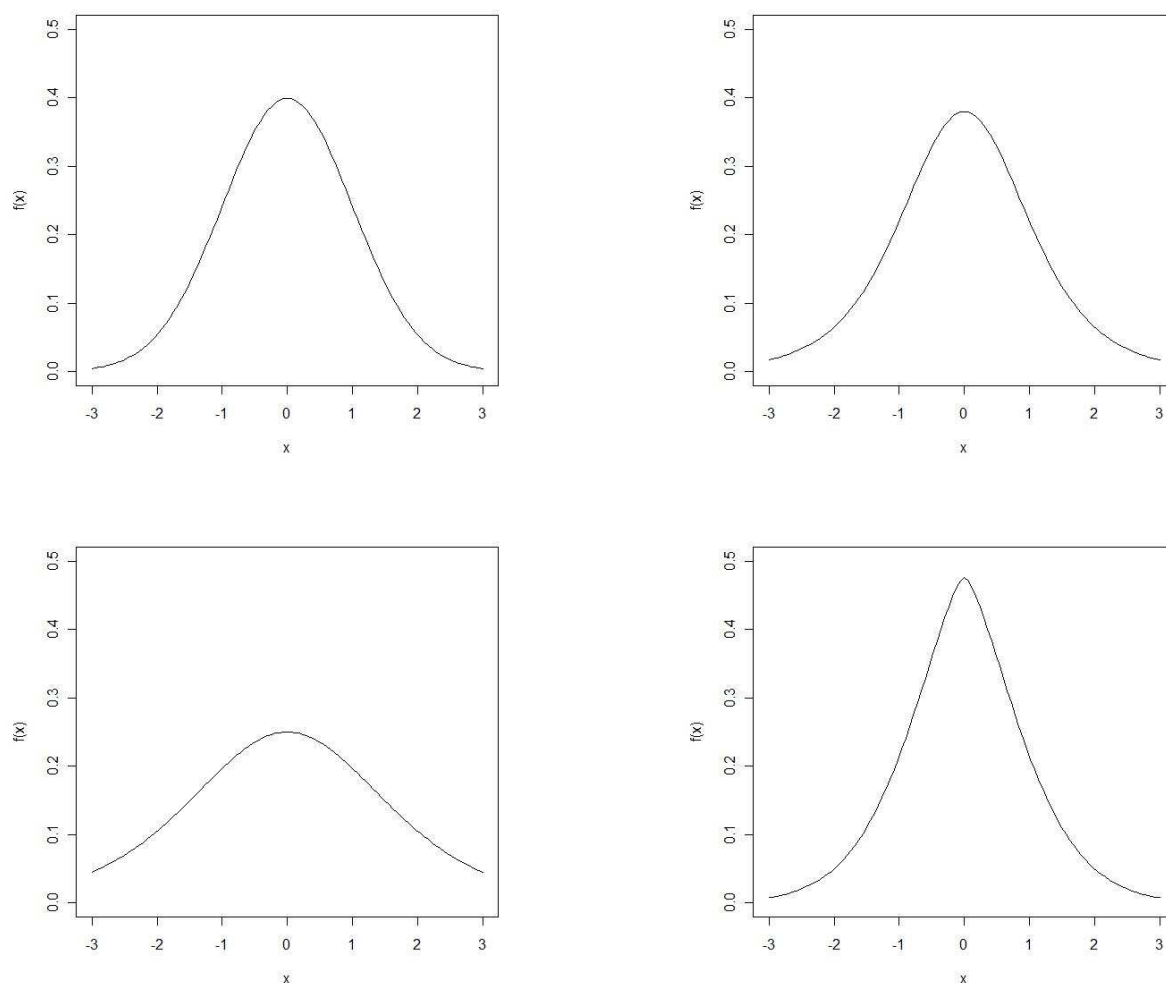


Figura 1.1: Gráfico da f.d.p. das distribuições normal padrão, t -Student com 5 graus de liberdade, Logística e Exponencial potência, respectivamente.

Uma outra suposição que também fazemos ao ajustarmos um modelo a um conjunto de dados é a suposição de que os erros possuem uma variância constante igual a ϕ , ou seja, homoscedasticidade, isto é, $\text{Var}(\epsilon_i) = \phi, \forall i$. Quando as variâncias dos erros não são constantes se diz haver heteroscedasticidade. A violação desta suposição pode causar adversas consequências para a eficiência dos estimadores.

Nesta dissertação nós propomos uma generalização dos MLGS em que o parâmetro de escala é relacionado por um preditor linear através de uma função de ligação positiva duplamente dife-

renciável, denominada função de dispersão. Isso é feito para podermos solucionar um problema comum em análise de regressão que é a presença de heteroscedasticidade.

A presença de heteroscedasticidade pode influenciar nas estimativas dos parâmetros do modelo e o estimador usual da matriz de variâncias e covariâncias possivelmente passa a ser não consistente, o que compromete as inferências resultantes. Para contornar este problema podemos ajustar um modelo que leva em conta a possível presença de heteroscedasticidade, incluindo um parâmetro de dispersão para cada observação.

Este trabalho se divide em quatro capítulos. No capítulo 2, apresentaremos o modelo linear generalizado simétrico heteroscedástico, bem como o desenvolvimento de um resíduo padronizado via Cox e Snell (1968) e medidas de diagnóstico sob enfoque de influência local.

No capítulo 3, apresentaremos duas aplicações do modelo proposto, como primeira aplicação trataremos, para cada unidade da federação, a relação entre o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio, estes dados foram analisados em Paula (2004). Na figura abaixo temos o gráfico de dispersão destes dados.

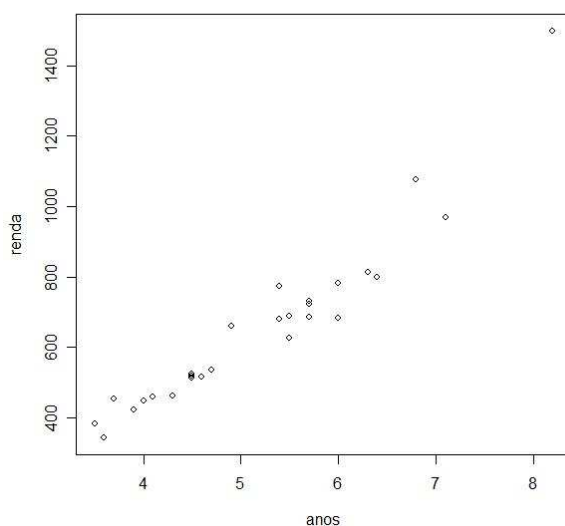


Figura 1.2: Gráfico de dispersão dos dados de escolaridade e renda.

Como segunda aplicação temos dados analisados por Tarpey (2001) que avalia o crescimento de crianças com a razão peso/altura versus idade de 72 crianças desde o nascimento até 71,5

meses de idade. Na Figura 1.3 apresentaremos o gráfico de dispersão destes dados.

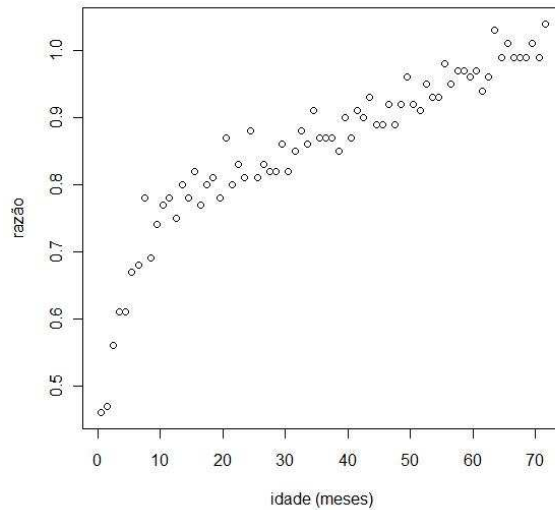


Figura 1.3: Gráfico de dispersão do crescimento de crianças.

Os gráficos e ajuste desta dissertação foram produzidos utilizando o software **R**, que se encontra disponível gratuitamente em www.r-project.org. Os exemplos foram analisados pela `library gamlss` desenvolvida para o ajuste de modelos aditivos generalizados para locação escala e forma, para maiores detalhes ver Rigby e Stasinopoulos (2001). E no último capítulo apresentamos as considerações finais e faremos algumas propostas futuras de trabalho.

Modelos Lineares Generalizados Simétricos Heteroscedásticos

2.1 Introdução

Na modelagem de dados reais, é muito comum a suposição de que a variância dos erros é constante. No entanto, nem sempre esta suposição é verificada. Como dissemos anteriormente, variâncias não constantes podem comprometer os resultados inferenciais. Para se contornar este problema, podemos aplicar uma transformação na variável resposta a fim de estabilizar a variância, mas nem sempre esta técnica traz resultados satisfatórios. A modelagem da variância foi discutida em Cysneiros et al. (2007), que se propõe a modelagem da heteroscedasticidade em modelos lineares simétricos.

Neste capítulo iremos apresentar o modelo linear generalizado simétrico, em que um parâmetro de dispersão é atribuído a cada observação e é relacionado com as variáveis explicativas através de uma função de ligação. Vamos agora mostrar alguns resultados importantes para os modelos lineares generalizados simétricos heteroscedásticos.

Considere o modelo de regressão linear generalizado

$$y_i = \mu_i + \sqrt{\phi_i} \epsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

sendo a densidade de y_i dada por

$$f_{y_i}(y, \mu, \phi) = \frac{1}{\sqrt{\phi_i}} a\left(\frac{(y_i - \mu_i)^2}{\phi_i}\right), \quad (2.2)$$

em que $y_1, \dots, y_n$ são variáveis resposta observadas, a componente sistemática $\mu_i = g(\eta_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ com $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ um vetor $p \times 1$ de parâmetros desconhecidos, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ um vetor de variáveis explicativas, $\phi_i = h(\tau_i)$ e $\tau_i = \mathbf{z}_i^T \boldsymbol{\gamma}$, $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ é um vetor $q \times 1$ de parâmetros desconhecidos e $\epsilon_i \sim S(0, 1, a)$. A função $g(\cdot)$ é uma função monótona e diferenciável, denominada de função de ligação e $h(\cdot)$ é chamada de ligação de dispersão sendo uma função continuamente diferenciável e bijetiva, além de ser conhecida e positiva. Os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são funcionalmente independentes. Quando existem, $E(y_i) = \mu_i$ e $\text{Var}(y_i) = \xi\phi$, em que ξ é uma constante maior que zero dada por $\xi = -2\varphi'(0)$ e $\varphi'(0) = d\varphi(u)/du|_{u=0}$, em que $u = (y - \mu)^2/\phi_i$.

Temos, por exemplo, que a função geradora de densidade da distribuição normal é dada por $a(u) = \frac{1}{\sqrt{2\pi}}e^{-u/2}$ e a função geradora de densidade da distribuição t -Student com grau de liberdade ν é $a(u) = \frac{\nu^{\nu/2}}{B(1/2, \nu/2)}(\nu + u)^{-(\nu+1)/2}$ com $\nu > 0$ e $B(\cdot, \cdot)$ é a função beta.

Este modelo proposto é uma extensão natural dos modelos lineares generalizados simétricos (MLGS) proposta por Lobos (2004) em que apenas o parâmetro de locação é relacionado por preditor linear através de uma ligação monótona e diferenciável.

2.2 Função de Máxima Verossimilhança e Função Escore

Sejam $y_1, \dots, y_n$ variáveis aleatórias independentes tal que $y_i \sim S(\mu_i, \phi_i, a)$, com f.d.p. dada por (2.2). A função de log-verossimilhança é dada por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n \log \left\{ \frac{a(u_i)}{\sqrt{\phi_i}} \right\},$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ e $u_i = \phi_i^{-1}(y_i - \mu_i)^2$. A função escore para $\boldsymbol{\theta}$ é denotada por

$$U(\boldsymbol{\theta}) = (U^T(\boldsymbol{\beta}), U^T(\boldsymbol{\gamma}))^T, \text{ sendo } U(\boldsymbol{\beta}) = (U(\beta_1), \dots, U(\beta_p))^T \text{ e } U(\boldsymbol{\gamma}) = (U(\gamma_1), \dots, U(\gamma_q))^T,$$

com

$$\begin{aligned}
 U(\beta_j) &= \frac{\partial L(\theta)}{\partial \beta_j} \quad j = 1, \dots, p, \\
 &= - \sum_{i=1}^n \left[\frac{\sqrt{\phi_i}}{a(u_i)} \frac{1}{\sqrt{\phi_i}} a'(u_i) \phi_i^{-1} 2(y_i - \mu_i) \frac{\partial \mu_i}{\partial \beta_j} \right] \\
 &= -2 \sum_{i=1}^n \left[\frac{a'(u_i)}{a(u_i)} \phi_i^{-1} 2(y_i - \mu_i) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right] \\
 &= -2 \sum_{i=1}^n \left[W_a(u_i) \phi_i^{-1} 2(y_i - \mu_i) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} x_{ij} \right]
 \end{aligned}$$

em que $W_a(u_i) = \frac{a'(u_i)}{a(u_i)}$. Em notação matricial,

$$U(\beta) = \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\mathbf{G}) \mathbf{D}^{-1}(\phi) (\mathbf{y} - \boldsymbol{\mu}), \quad (2.3)$$

em que $\mathbf{X}$ é uma matriz $n \times p$ de posto coluna completo com linhas $\mathbf{x}_i^T$, $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, \dots, v_n\}$ com $v_i = -2W_a(u_i)$, $\mathbf{D}(\phi) = \text{diag}\{\phi_1, \dots, \phi_n\}$ com $\phi_i = h(\tau_i) = \mathbf{z}_i^T \boldsymbol{\gamma}$, $\mathbf{D}(\mathbf{G}) = \text{diag}\{G_1, \dots, G_n\}$ com $G_i = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

Analogamente, temos que a função $U(\gamma_\ell)$ é dada por

$$\begin{aligned}
 U(\gamma_\ell) &= \frac{\partial L(\theta)}{\partial \gamma_\ell} \quad \ell = 1, \dots, q, \\
 &= \sum_{i=1}^n \frac{\partial}{\partial \gamma_j} \left\{ \log a(u_i) - \frac{1}{2} \log \phi_i \right\} \\
 &= - \sum_{i=1}^n \left[\frac{a'(u_i)}{a(u_i)} \phi_i^{-2} (y_i - \mu_i)^2 h'(\tau_i) \frac{\partial \tau_i}{\partial \gamma_j} - \frac{1}{2} \phi_i^{-1} h'(\tau_i) \frac{\partial \tau_i}{\partial \gamma_j} \right] \\
 &= \sum_{i=1}^n \left\{ \frac{h'(\tau_i)}{2\phi_i} \left[-2 \frac{a'(u_i)}{a(u_i)} \frac{1}{\phi_i} (y_i - \mu_i)^2 - 1 \right] \frac{\partial \tau_i}{\partial \gamma_j} \right\} \\
 &= \sum_{i=1}^n \left\{ z_{ij} \frac{h'(\tau_i)}{2\phi_i} \left[-2 \frac{a'(u_i)}{a(u_i)} \frac{1}{\phi_i} (y_i - \mu_i)^2 - 1 \right] \right\} \\
 &= \sum_{i=1}^n \{ z_{ij} w_i [-2W_a(u_i) u_i - 1] \} \\
 &= \sum_{i=1}^n z_{ij} w_i (v_i u_i - 1) \\
 &= \sum_{i=1}^n z_{ij} w_i (m_i - 1)
 \end{aligned}$$

em que $w_i = \frac{h'(\tau_i)}{2\phi_i}$ e $m_i = v_i u_i$.

Em notação matricial,

$$U(\gamma) = \mathbf{Z}^T \mathbf{D}(\mathbf{w})(\mathbf{m} - \mathbf{1}), \quad (2.4)$$

em que $\mathbf{Z}$ é uma matriz $n \times q$ com linhas $\mathbf{z}_i^T$, $\mathbf{D}(\mathbf{w}) = \text{diag}\{w_1, \dots, w_n\}$ com $w_i = \frac{h'(\tau_i)}{2\phi_i}$, $\mathbf{m} = (m_1, \dots, m_n)^T$ com $m_i = v_i u_i$ e $\mathbf{1}$ um vetor $n \times 1$ de uns.

As expressões de W_a para as distribuições normal e t -Student são, respectivamente, $-\frac{1}{2}$ e $-\frac{\nu+1}{2(\nu+u)}$. E as expressões de W'_a para as distribuições normal e t -Student são, respectivamente, 0 e $\frac{\nu+1}{(\nu+u)^2}$.

2.3 Informação de Fisher

Nesta seção, calcularemos a matriz de Informação de Fisher, que será utilizada no processo de estimação através do método scoring de Fisher. A matriz de Informação de Fisher pode ser expressa como:

$$\mathbf{K}_{\theta\theta} = E(U(\theta)U^T(\theta)) = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\gamma\gamma} \end{bmatrix},$$

em que $\mathbf{K}_{\beta\beta} = E(U(\beta)U^T(\beta))$ e $\mathbf{K}_{\gamma\gamma} = E(U(\gamma)U^T(\gamma))$.

O $(j\ell)$ -ésimo elemento, $1 \leq j, \ell \leq p$, de $\mathbf{K}_{\beta\beta}$ é dado por:

$$\begin{aligned} E[U(\beta_j)U(\beta_\ell)] &= E\left[\frac{\partial L(\theta)}{\partial \beta_j} \frac{\partial L(\theta)}{\partial \beta_\ell}\right] \\ &= \sum_{i=1}^n \frac{4}{\phi_i} E\left(W_a^2(u_i) \frac{(y_i - \mu_i)^2}{\phi_i}\right) \left[\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i}\right]^2 x_{ij}x_{i\ell} \\ &= \sum_{i=1}^n \frac{4}{\phi_i} E(W_a^2(u_i)u_i) \left[\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i}\right]^2 x_{ij}x_{i\ell} \\ &= 4 \sum_{i=1}^n \frac{1}{\phi_i} E(W_a^2(u_i)u_i) \left[\frac{\partial g^{-1}}{\partial \eta_i}\right]^2 x_{ij}x_{i\ell}, \end{aligned}$$

em que $u_i = \frac{(y_i - \mu_i)^2}{\phi_i}$. Logo, em notação matricial temos

$$\mathbf{K}_{\beta\beta} = 4d_a \mathbf{X}^T \mathbf{D}^2(\mathbf{G}) \mathbf{D}^{-1}(\phi) \mathbf{X}, \quad (2.5)$$

em que $d_a = E(W_a^2(U)U)$, $\mathbf{D}^2(\mathbf{G}) = \text{diag}\{G_1^2, \dots, G_n^2\}$ com $G_i^2 = \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i}\right)^2$ e $U \sim S(0, 1, a)$.

O $(j\ell)$ -ésimo elemento, $1 \leq j, \ell \leq q$, de $\mathbf{K}_{\gamma\gamma}$ pode ser escrito por

$$\begin{aligned}
 U(\gamma_j)U(\gamma_\ell) &= E \left[\frac{\partial L(\boldsymbol{\theta})}{\partial \gamma_j} \frac{\partial L(\boldsymbol{\theta})}{\partial \gamma_\ell} \right] \\
 &= \sum_{i=1}^n E \left[w_i^2 z_{ij} z_{i\ell} (m_i - 1)^2 \right] \\
 &= \sum_{i=1}^n E \left[w_i^2 z_{ij} z_{i\ell} (v_i u_i - 1)^2 \right] \\
 &= \sum_{i=1}^n E \left[w_i^2 z_{ij} z_{i\ell} (-2W_a(u_i) - 1)^2 \right] \\
 &= \sum_{i=1}^n E \left\{ [w_i^2 z_{ij} z_{i\ell} [4W_a^2(u_i)u_i^2 + 4W_a(u_i)u_i + 1]] \right\} \\
 &= \sum_{i=1}^n w_i^2 z_{ij} z_{i\ell} [4E(W_a^2(u_i)u_i^2) + 4E(W_a(u_i)u_i) + 1] \\
 &= \sum_{i=1}^n w_i^2 z_{ij} z_{i\ell} [4f_a - 3] \\
 &= \sum_{i=1}^n w_i^2 z_{ij} z_{i\ell} (4f_a - 1),
 \end{aligned}$$

dado que $E(W_a(U)U) = -\frac{1}{2}$ e com $f_a = E[(W_a(U)U)^2]$, rescrevendo na forma matricial:

$$\mathbf{K}_{\gamma\gamma} = (4f_a - 1) \mathbf{Z}^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}. \quad (2.6)$$

Assim, $\mathbf{K}_{\theta\theta}$ pode ser escrita como

$$\mathbf{K}_{\theta\theta} = \begin{bmatrix} 4d_a \mathbf{X}^T \mathbf{D}^2(\mathbf{G}) \mathbf{D}^{-1}(\boldsymbol{\phi}) \mathbf{X} & \mathbf{0} \\ \mathbf{0} & (4f_a - 1) \mathbf{Z}^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z} \end{bmatrix}. \quad (2.7)$$

A matriz de informação observada será utilizada nos métodos de detecção de pontos de alavanca e influência local. A matriz de informação observada é denotada por $\ddot{L}_{\theta\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, e pode ser expressa da seguinte forma

$$\ddot{L}_{\theta\theta} = \begin{bmatrix} \ddot{L}_{\beta\beta} & \ddot{L}_{\beta\gamma} \\ \ddot{L}_{\gamma\beta} & \ddot{L}_{\gamma\gamma} \end{bmatrix}.$$

O $(j\ell)$ -ésimo elemento da matriz $\ddot{L}_{\beta\beta}$ é dado por

$$\begin{aligned}
 \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} &= -2 \sum_{i=1}^n \frac{x_{ij}}{\phi_i} \left[-2W'_a(u_i)u_i \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 x_{i\ell} + W_a(u_i) \left\{ - \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 x_{i\ell} \right. \right. \\
 &\quad \left. \left. + (y_i - \mu_i(\beta)) \frac{\partial^2 g^{-1}(\eta_i)}{\partial \eta_i^2} x_{i\ell} \right\} \right] \\
 &= -2 \sum_{i=1}^n \frac{x_{ij}}{\phi_i} [-2W'_a(u_i)u_i G_i^2 - W_a(u_i)G_i^2 + W_a(u_i)(y_i - \mu_i)G'_i] x_{i\ell} \\
 &= - \sum_{i=1}^n \frac{x_{ij}}{\phi_i} [-4W'_a(u_i)u_i G_i^2 - 2W_a(u_i)G_i^2 + 2W_a(u_i)(y_i - \mu_i)G'_i] x_{i\ell} \\
 &= - \sum_{i=1}^n \frac{x_{ij}}{\phi_i} [-4W'_a(u_i)u_i G_i^2 + v_i G_i^2 - v_i(y_i - \mu_i)G'_i] x_{i\ell} \\
 &= - \sum_{i=1}^n x_{ij} \left[v_i \left(1 - (y_i - \mu_i) \frac{G'_i}{G_i^2} \right) - 4W'_a(u_i)u_i \right] \frac{1}{\phi_i} G_i^2 \\
 &= - \sum_{i=1}^n x_{ij} b_i \frac{1}{\phi_i} G_i^2,
 \end{aligned}$$

em que $b_i = v_i \left(1 - (y_i - \mu_i(\beta)) \frac{G'_i}{G_i^2} \right) - 4W'_a(u_i)u_i$. Em forma matricial temos,

$$\ddot{L}_{\beta\beta} = -\mathbf{X}^T \mathbf{D}(\mathbf{b}) \mathbf{D}^2(\mathbf{G}) \mathbf{X}. \quad (2.8)$$

O $(j\ell)$ -ésimo elemento da matriz $\ddot{L}_{\beta\gamma}$ é dado por

$$\begin{aligned}
 \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_\ell} &= -2 \sum_{i=1}^n (y_i - \mu_i) x_{ij} G_i \left[-W'_a(u_i) \frac{(y_i - \mu_i)^2}{\phi_i^2} h'(\tau_i) z_{i\ell} \frac{1}{\phi_i} - W_a(u_i) \frac{1}{\phi_i^2} h'(\tau_i) z_{i\ell} \right] \\
 &= 2 \sum_{i=1}^n x_{ij} G_i \left[W'_a(u_i) u_i \frac{(y_i - \mu_i)}{\phi_i^2} h'(\tau_i) z_{i\ell} + W_a(u_i) \frac{(y_i - \mu_i)}{\phi_i} \frac{h'(\tau_i) z_{i\ell}}{\phi_i} \right] \\
 &= 2 \sum_{i=1}^n x_{ij} G_i \left[\frac{h'(\tau_i)}{\phi_i^2} (y_i - \mu_i) \right] \{ W'_a(u_i) u_i + W_a(u_i) \} z_{i\ell} \\
 &= 2 \sum_{i=1}^n x_{ij} G_i d_i z_{i\ell}
 \end{aligned}$$

em que $d_i = \left[\frac{h'(\tau_i)}{\phi_i^2} (y_i - \mu_i) \right] \{ W'_a(u_i) u_i + W_a(u_i) \}$. Em forma matricial, temos

$$\ddot{L}_{\beta\gamma} = 2\mathbf{X} \mathbf{D}(\mathbf{d}) \mathbf{D}(\mathbf{G}) \mathbf{Z}. \quad (2.9)$$

O $(j\ell)$ -ésimo elemento da matriz $\ddot{L}_{\gamma\gamma}$ é dado por

$$\begin{aligned}
 \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} &= - \sum_{i=1}^n \left[\frac{z_{ij}}{2} \left(\frac{h''(\tau_i)}{\phi} - \frac{(h'(\tau_i))^2}{\phi_i^2} \right) z_{i\ell} - z_{ij} \left\{ h''(\tau_i) \frac{W_a(u_i)u_i}{\phi_i} + h'(\tau_i) \left[-W'_a(u_i) \frac{(y_i - \mu_i)^2}{\phi_i^2} \right. \right. \right. \\
 &\quad \left. \left. \times h'(\tau_i) \frac{u_i}{\phi_i} + W_a(u_i) \left(-\frac{(y_i - \mu_i)^2}{\phi_i^2} \frac{h'(\tau_i)}{\phi_i} - u_i \frac{h'(\tau_i)}{\phi_i^2} \right) \right] \right\} z_{i\ell} \right] \\
 &= - \sum_{i=1}^n z_{ij} \left\{ \frac{h''(\tau_i)}{2\phi_i} - \frac{(h'(\tau_i))^2}{2\phi_i^2} - \frac{h''(\tau_i)W_a(u_i)u_i}{\phi_i} \frac{(h'(\tau_i))^2 W'_a(u_i)u_i^2}{\phi_i^2} \right. \\
 &\quad \left. + \frac{2W_a(u_i)u_i h'(\tau_i)^2}{\phi_i^2} \right\} z_{i\ell} \\
 &= - \sum_{i=1}^n z_{ij} \left[\frac{1}{2\phi_i} \left(h''(\tau_i) - \frac{h'(\tau_i)^2}{\phi_i} \right) - \frac{h'(\tau_i)^2 u_i \{W'_a(u_i)u_i + 2W_a(u_i)u_i\}}{\phi_i^2} \right. \\
 &\quad \left. + \frac{h''(\tau_i)W_a(u_i)u_i}{\phi_i} \right] z_{i\ell} \\
 &= - \sum_{i=1}^n z_{ij} c_i z_{i\ell}
 \end{aligned}$$

em que $c_i = \frac{1}{2\phi_i} \left(h''(\tau_i) - \frac{h'(\tau_i)^2}{\phi_i} \right) - \frac{h'(\tau_i)^2 u_i}{\phi_i^2} \{W'_a(u_i)u_i + 2W_a(u_i)u_i\} + \frac{h''(\tau_i)W_a(u_i)u_i}{\phi_i}$. Em forma matricial temos,

$$\ddot{L}_{\gamma\gamma} = -\mathbf{Z}^T \mathbf{D}(\mathbf{c}) \mathbf{Z}. \quad (2.10)$$

Assim, $\ddot{L}_{\theta\theta}$ pode ser escrita por

$$\ddot{L}_{\theta\theta} = \begin{bmatrix} -\mathbf{X}^T \mathbf{D}(\mathbf{b}) \mathbf{D}^2(\mathbf{G}) \mathbf{X} & 2\mathbf{X} \mathbf{D}(\mathbf{d}) \mathbf{D}(\mathbf{G}) \mathbf{Z} \\ 2\mathbf{X} \mathbf{D}(\mathbf{d}) \mathbf{D}(\mathbf{G}) \mathbf{Z} & -\mathbf{Z}^T \mathbf{D}(\mathbf{c}) \mathbf{Z} \end{bmatrix}. \quad (2.11)$$

Os valores de d_a , f_a e ξ para a distribuição normal são, respectivamente, $\frac{1}{4}$, $\frac{3}{4}$ e 1. E os valores de d_a , f_a e ξ para a distribuição t -Student, $\frac{\nu+1}{4(\nu+3)}$, $\frac{3(\nu+1)}{4(\nu+3)}$ e $\frac{\nu}{\nu-2}$, $\nu > 2$. Para encontrar valores de W_a , W'_a , d_a , f_a e ξ para outras distribuições simétricas, ver Cysneiros (2004).

2.4 Estimação dos parâmetros

Para a obtenção da estimativa de máxima verossimilhança é necessário um processo iterativo, já que pelo método da máxima verossimilhança não é possível obter uma forma analítica para o estimador. O processo iterativo utilizado foi o de Newton-Raphson, que consiste em expandir a

função escore $U(\boldsymbol{\theta})$ em torno de um valor inicial $\boldsymbol{\theta}^{(0)}$, tal que

$$\begin{aligned} U(\boldsymbol{\theta}) &\cong U(\boldsymbol{\theta}^{(m)}) + U'(\boldsymbol{\theta}^{(m)}) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(m)} \right), \\ \boldsymbol{\theta}^{(m+1)} &= \boldsymbol{\theta}^{(m)} + \left\{ -U'(\boldsymbol{\theta}^{(m)}) \right\}^{-1} U(\boldsymbol{\theta}^{(m)}) \end{aligned}$$

com $m = 0, 1, \dots$ e $U'(\boldsymbol{\theta})$ denota a primeira derivada de $U(\boldsymbol{\theta})$. Como a matriz $-U'(\boldsymbol{\theta}^{(m)})$ pode não ser positiva definida, aplica-se o método de scoring de Fisher, ou seja, substitui-se a matriz $-U'(\boldsymbol{\theta})$ pelo seu valor esperado. Isso resulta no seguinte processo iterativo:

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} U(\boldsymbol{\theta}^{(m)}),$$

Na forma matricial temos:

$$\begin{pmatrix} \boldsymbol{\beta}^{(m+1)} \\ \boldsymbol{\gamma}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^{(m)} \\ \boldsymbol{\gamma}^{(m)} \end{pmatrix} + \begin{bmatrix} \mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} \end{bmatrix} \begin{pmatrix} U(\boldsymbol{\beta}^{(m)}) \\ U(\boldsymbol{\gamma}^{(m)}) \end{pmatrix}, \quad (2.12)$$

ou seja,

$$\begin{cases} \boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} U(\boldsymbol{\beta}^{(m)}) \\ \boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)} + \mathbf{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} U(\boldsymbol{\gamma}^{(m)}) \end{cases} \quad (2.13)$$

Substituindo $\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}$, $\mathbf{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$, $U(\boldsymbol{\gamma})$ e $U(\boldsymbol{\beta})$ nas equações acima temos:

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + \left[4d_a \mathbf{X}^T \mathbf{D}^2(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)}) \mathbf{X} \right]^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{v}^{(m)}) \mathbf{D}(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)}) (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ &= \left[4d_a \mathbf{X}^T \mathbf{D}^2(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)}) \mathbf{X} \right]^{-1} \left\{ \left[4d_a \mathbf{X}^T \mathbf{D}^2(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)}) \mathbf{X} \right] \boldsymbol{\beta}^{(m)} \right. \\ &\quad \left. + \mathbf{X}^T \mathbf{D}(\mathbf{v}^{(m)}) \mathbf{D}(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)}) (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \right\}, \end{aligned}$$

sendo $\mathbf{W}_1 = 4d_a \mathbf{D}^2(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)})$ temos $\mathbf{Q}^{(m)} = \mathbf{D}(\mathbf{v}^{(m)}) \mathbf{D}(\mathbf{G}^{(m)}) \mathbf{D}^{-1}(\boldsymbol{\phi}^{(m)})$,

$$\boldsymbol{\beta}^{(m+1)} = [\mathbf{X}^T \mathbf{W}_1 \mathbf{X}]^{-1} \left[(\mathbf{X}^T \mathbf{W}_1 \mathbf{X}) \boldsymbol{\beta}^{(m)} + \mathbf{X}^T \mathbf{Q}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \right].$$

Seja $\mathbf{P}^{(m)} = \mathbf{W}_1^{-1} \mathbf{Q}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$,

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= [\mathbf{X}^T \mathbf{W}_1 \mathbf{X}]^{-1} \left\{ (\mathbf{X}^T \mathbf{W}_1) (\mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{P}^{(m)}) \right\} \\ \boldsymbol{\beta}^{(m+1)} &= (\mathbf{X}^T \mathbf{W}_1 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_1 \mathbf{z}_{\boldsymbol{\beta}}^{(m)}, \end{aligned}$$

em que $\mathbf{z}_{\boldsymbol{\beta}}^{(m)} = \mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{P}^{(m)}$ é um vetor $(p \times 1)$ com o i -ésimo elemento dado por

$$z_{\beta_i} = \mu_i + \frac{v_i}{4d_a} (y_i - \mu_i).$$

Analogamente, substituindo $\mathbf{K}_{\gamma\gamma}$ e $U(\gamma)$ em (2.13) obtemos:

$$\gamma^{(m+1)} = \gamma^{(m)} + \left[(4f_a - 1) \mathbf{Z}^T \mathbf{D}^2(\mathbf{w}^{(m)}) \mathbf{Z} \right]^{-1} \mathbf{Z}^T \mathbf{D}(\mathbf{w}^{(m)}) (\mathbf{m}^{(m)} - 1).$$

Definido $\mathbf{W}_2 = (4f_a - 1) \mathbf{D}^2(\mathbf{w}^{(m)})$, $\mathbf{M} = \mathbf{D}(\mathbf{w}^{(m)}) (\mathbf{m}^{(m)} - 1)$,

$$\begin{aligned} \gamma^{(m+1)} &= \gamma^{(m)} + [\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}]^{-1} \mathbf{Z}^T \mathbf{M} \\ &= (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z})^{-1} \left\{ (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}) \gamma^{(m)} + \mathbf{Z}^T \mathbf{M} \right\} \quad \text{seja } \mathbf{T} = \mathbf{W}_2^{-1} \mathbf{M} \\ &= (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z})^{-1} \left\{ (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}) \gamma^{(m)} + \mathbf{Z}^T \mathbf{W}_2 \mathbf{M} \right\} \\ &= (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z})^{-1} \left\{ \mathbf{Z}^T \mathbf{W}_2 \left(\mathbf{Z} \gamma^{(m)} + \mathbf{T}^{(m)} \right) \right\}. \end{aligned}$$

Fazendo $\mathbf{z}_{\gamma}^{(m)} = \mathbf{Z} \gamma^{(m)} + \mathbf{T}^{(m)} = \boldsymbol{\tau}^{(m)} + \mathbf{T}^{(m)}$,

$$\gamma^{(m+1)} = (\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z})^{-1} \left\{ \mathbf{Z}^T \mathbf{W}_2 \mathbf{z}_{\gamma}^{(m)} \right\},$$

$$(\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}) \gamma^{(m+1)} = \mathbf{Z}^T \mathbf{W}_2 \mathbf{z}_{\gamma}^{(m)},$$

em que $\mathbf{z}_{\gamma} = z_{\gamma_1}, \dots, z_{\gamma_n}^t$ com

$$\begin{aligned} z_{\gamma_i} &= \tau_i + \frac{M_i}{W_{2i}} = \tau_i + \left[\frac{h'(\tau_i)/2\phi_i(v_i u_i - 1)}{(4f_a - 1)h'(\tau_i)^2/4\phi_i^2} \right] \\ &= \tau_i + \frac{2\phi_i(v_i u_i - 1)}{(4f_a - 1)h'(\tau_i)} \\ &= \tau_i + \frac{(v_i u_i - 1)}{(4f_a - 1) \frac{h'(\tau_i)}{2\phi_i}}. \end{aligned}$$

2.5 Teste de Homogeneidade

A variância de uma variável aleatória simétrica, $y \sim S(\mu_i, \phi_i, a)$, é dada por $\text{Var}(y_i) = \xi \phi_i$, em que ξ é uma constante maior que zero dada por $\xi = -2\varphi'(0)$ e $\varphi'(0) = d\varphi(u)/du|_{u=0}$, em que $\varsigma_y(t) = E(e^{ity}) = e^{it\mu_i} \varphi(t^2 \phi)$. No modelo (2.2) testar $H_0 : \phi_1 = \phi_2 = \dots = \phi_n = \phi$ contra $H_1 : \text{pelo menos um } \phi_i \neq \phi_j, i \neq j$ é equivalente a testar a homogeneidade das variâncias. Seja o vetor $\boldsymbol{\gamma} = (\gamma_1, \boldsymbol{\gamma}^{*T})^T$ em que $\boldsymbol{\gamma}^* = (\gamma_2, \dots, \gamma_q)^T$. Seja também $(\mathbf{Z}_2 \mathbf{Z}_1)$ uma partição de $\mathbf{Z}$ tal que, $\mathbf{Z}_2 = \mathbf{1}_n$, ou seja, um vetor $n \times 1$ de uns, e $\mathbf{Z}_1 = [\mathbf{z}_1 \dots \mathbf{z}_q]$ em que $\mathbf{z}_i$ é a i -ésima coluna de $\mathbf{Z}$.

A variância de γ^* pode ser expressa por

$$\text{Var}(\hat{\gamma}) = \begin{bmatrix} \text{Var}(\hat{\gamma}_1) & \text{Cov}(\hat{\gamma}^*, \hat{\gamma}_1) \\ \text{Cov}(\hat{\gamma}^*, \hat{\gamma}_1) & \text{Var}(\hat{\gamma}^*) \end{bmatrix}.$$

Assumindo que $\gamma \in \Omega_\gamma \subset \mathbb{R}^q$, em que Ω_γ é um conjunto aberto. É possível mostrar que $\hat{\gamma}$, o estimador de máxima verossimilhança de γ , é um estimador consistente de γ e que

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N_q(\mathbf{0}, J_{\gamma\gamma}^{-1}), \text{ em que } J_{\gamma\gamma} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{K}_{\gamma\gamma},$$

então, $\hat{\mathbf{K}}_{\gamma\gamma}^{-1}$ é um estimador consistente da matriz de variância-covariância assintótica de $\hat{\gamma}$.

A matriz $\mathbf{K}_{\gamma\gamma}$ é dada da seguinte forma,

$$\mathbf{K}_{\gamma\gamma} = \begin{bmatrix} \mathbf{K}_{\gamma_1\gamma_1} & \mathbf{K}_{\gamma_1\gamma^*} \\ \mathbf{K}_{\gamma^*\gamma_1} & \mathbf{K}_{\gamma^*\gamma^*} \end{bmatrix}.$$

onde

$$\begin{aligned} \mathbf{K}_{\gamma_1\gamma_1} &= (4f_a - 1) \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2; & \mathbf{K}_{\gamma_1\gamma^*} &= (4f_a - 1) \mathbf{Z}_1^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2; \\ \mathbf{K}_{\gamma^*\gamma_1} &= (4f_a - 1) \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_1; & \mathbf{K}_{\gamma^*\gamma^*} &= (4f_a - 1) \mathbf{Z}_1^T \mathbf{D}^1(\mathbf{w}) \mathbf{Z}_1. \end{aligned}$$

Logo, temos que

$$\mathbf{K}_{\gamma\gamma}^{-1} = \begin{bmatrix} \mathbf{K}_{\gamma_1\gamma_1}^{-1} + \mathbf{F} \mathbf{E}^{-1} \mathbf{F}^T & -\mathbf{F} \mathbf{E} \\ -\mathbf{E}^{-1} \mathbf{F}^T & \mathbf{E}^{-1} \end{bmatrix}$$

onde

$$\begin{aligned} \mathbf{E} &= \mathbf{K}_{\gamma^*\gamma^*} - \mathbf{K}_{\gamma_1\gamma^*} \mathbf{K}_{\gamma_1\gamma_1}^{-1} \mathbf{K}_{\gamma^*\gamma_1} \\ \mathbf{F} &= \mathbf{K}_{\gamma_1\gamma_1}^{-1} \mathbf{K}_{\gamma^*\gamma_1}. \end{aligned}$$

Logo, $\text{Var}(\hat{\gamma}^*) = \mathbf{E}^{-1} = [\mathbf{K}_{\gamma^*\gamma^*} - \mathbf{K}_{\gamma_1\gamma^*} \mathbf{K}_{\gamma_1\gamma_1}^{-1} \mathbf{K}_{\gamma^*\gamma_1}]^{-1}$. Daí temos,

$$\begin{aligned} \text{Var}(\hat{\gamma}^*) &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_1 - \mathbf{Z}_1^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_1 \right]^{-1} \\ &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \mathbf{D}(\mathbf{w}) \left\{ \mathbf{D}(\mathbf{w}) \mathbf{Z}_1 - \mathbf{D}(\mathbf{w}) \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_1 \right\} \right]^{-1} \\ &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \mathbf{D}(\mathbf{w}) \left\{ \mathbf{I} - \mathbf{D}(\mathbf{w}) \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}(\mathbf{w}) \right\} \mathbf{D}(\mathbf{w}) \mathbf{Z}_1 \right]^{-1} \\ &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \mathbf{D}(\mathbf{w}) (\mathbf{I} - \mathbf{H}_2) \mathbf{D}(\mathbf{w}) \mathbf{Z}_1 \right]^{-1}, \end{aligned}$$

em que $\mathbf{H}_2 = \mathbf{D}(\mathbf{w})\mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}(\mathbf{w})$. Podemos rescrever a variância de γ^* por

$$\text{Var}(\hat{\gamma}^*) = (4f_a - 1)^{-1} [\mathbf{R}^T \mathbf{D}^2(\mathbf{w})\mathbf{R}]^{-1}, \quad (2.14)$$

em que $\mathbf{R} = \mathbf{Z}_1 - \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_1$. Mas, $\mathbf{Z}_2 = \mathbf{1}_n$, logo

$$\begin{aligned} \mathbf{R} &= \mathbf{Z}_1 - \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_1 \\ &= \mathbf{Z}_1 - \mathbf{1}_n (\mathbf{1}_n^T \mathbf{D}^2(\mathbf{w})\mathbf{1}_n)^{-1} \mathbf{1}_n^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_1. \end{aligned}$$

O teste de homogeneidade pode ser escrito como $H_0 : \gamma^* = 0$ contra $H_1 : \text{pelo menos um } i \neq 0 \forall i = 2, \dots, q$. Então sob H_0 , temos $\gamma_2 = \gamma_3 = \dots = \gamma_q = 0$ portanto, para todo i , $\phi_i = h(\tau_i) = h(\tau_1)$. Logo, $\phi_i = h(\tau_i) = h(\tau_1)$ para todo i . Assim $w_i = \frac{h'(\tau_1)}{2h(\tau_1)} \forall i$ e temos

$$\mathbf{D}^2(\mathbf{w}) = \frac{h'(\gamma_1)}{2h(\gamma_1)} \mathbf{I}_n,$$

Portanto,

$$\begin{aligned} \mathbf{R} &= \mathbf{Z}_1 - \mathbf{D}^{-2}(\mathbf{w})\mathbf{1}_n (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}_n^T \mathbf{D}^2(\mathbf{w})\mathbf{Z}_1 \\ &= \mathbf{Z}_1 - \mathbf{1}_n (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}_n^T \mathbf{Z}_1 \\ &= \mathbf{Z}_1 - \mathbf{1}_n \mathbf{1}_n^T \frac{1}{n} \mathbf{Z}_1 \\ &= \mathbf{Z}_1 - \mathbf{J}_n \frac{1}{n} \mathbf{Z}_1, \end{aligned}$$

em que $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$, com $\mathbf{1}_n$ representando um vetor de dimensão $n \times 1$ com todos os elementos iguais a 1.

Logo, variância de $\hat{\gamma}^*$ sob H_0 é dada por:

$$\begin{aligned} \text{Var}_0(\hat{\gamma}^*) &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \mathbf{D}(\mathbf{w}) \left(\mathbf{I} - \mathbf{J}_n \frac{1}{n} \right) \mathbf{D}(\mathbf{w})\mathbf{Z}_2 \right]^{-1} \\ &= (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \left(\mathbf{I} - \mathbf{J}_n \frac{1}{n} \right) \mathbf{Z}_1 \right]^{-1} \mathbf{D}^{-2}(\mathbf{w}). \end{aligned}$$

A estatística Escore será então dada por

$$\begin{aligned} \xi_{SR} &= U_\gamma \left(\hat{\boldsymbol{\theta}}^0 \right)^T \widehat{\text{Var}_0(\hat{\gamma}^*)} U_\gamma \left(\hat{\boldsymbol{\theta}}^0 \right) \\ &= [\mathbf{Z}_1^T \mathbf{D}(\mathbf{w})(\hat{\mathbf{m}}^0 - \mathbf{1})]^T (4f_a - 1)^{-1} \left[\mathbf{Z}_1^T \left(\mathbf{I} - \mathbf{J}_n \frac{1}{n} \right) \mathbf{Z}_1 \right]^{-1} \mathbf{D}^{-2}(\mathbf{w}) [\mathbf{Z}_1^T \mathbf{D}(\mathbf{w})(\hat{\mathbf{m}}^0 - \mathbf{1})] \end{aligned}$$

$$\xi_{SR} = \frac{1}{(4f_a - 1)} (\hat{\mathbf{m}}^0 - \mathbf{1})^T \mathbf{Z}_1^T \left[\mathbf{Z}_1^T \left(\mathbf{I} - \frac{\mathbf{J}_n}{n} \right) \mathbf{Z}_1 \right]^{-1} \mathbf{Z}_1 (\hat{\mathbf{m}}^0 - \mathbf{1}). \quad (2.15)$$

em que $\mathbf{m}^0$ é avaliado em H_0 .

A estatística do teste da Razão de Verossimilhanças é dada por:

$$\xi_{RV} = 2\{L(\hat{\gamma}, \mathbf{y}) - L(\hat{\gamma}^0, \mathbf{y})\}, \quad (2.16)$$

em que, $\hat{\gamma}$ é o estimador de máxima verossimilhança de γ e $\hat{\gamma}^0$ é o estimador de máxima verossimilhança de γ sob H_0 .

A estatística do teste Wald é definido, neste caso, por

$$\begin{aligned} \xi_W &= [\hat{\gamma} * -\mathbf{0}]^T \widehat{\text{Var}}^{-1}(\hat{\gamma} *) [\hat{\gamma} * -\mathbf{0}] \\ &= (4f_a - 1) [\hat{\gamma} *]^T [\mathbf{R}^T \mathbf{D}^2(\mathbf{w}) \mathbf{R}] \hat{\gamma} *, \end{aligned}$$

em que $\mathbf{R} = \mathbf{Z}_1 - \mathbf{Z}_2 (\mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{D}^2(\mathbf{w}) \mathbf{Z}_1$.

Sob a hipótese nula, ξ_{SR} , ξ_{RV} e ξ_W têm distribuição χ^2 com $q - 1$ graus de liberdade para n suficientemente grande.

2.6 Resíduos

Após ajustarmos um modelo ao conjunto de dados devemos saber se ele está bem ajustado, pois um modelo não apropriado pode levar a conclusões erradas. Uma forma de se verificar a adequação do modelo aos dados é através da análise de resíduos. Com esta técnica podemos observar se as suposições feitas para os erros estão corretas e se existem observações aberrantes.

O resíduo pode ser definido como a diferença entre o valor observado e o valor estimado, este resíduo é chamado de resíduo ordinário. Vamos utilizar a metodologia apresentada por Cox e Snell (1968) para calcularmos os dois primeiros momentos do resíduo ordinário em modelos lineares generalizados simétricos heteroscedásticos. Considerando o resíduo ordinário com ϕ_i conhecido ou fixo expresso na seguinte forma

$$r_i(y_i, \hat{\mu}_i, \phi_i) = y_i - \hat{\mu}_i, \quad i = 1, \dots, n, \quad (2.17)$$

em que $\hat{\mu}_i = g(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$, $y_i = \mu_i + \epsilon_i$ e $\epsilon_i \sim S(0, \phi_i, a)$.

Geralmente, estes resíduos são viesados e têm distribuição não normal o que dificulta a verificação da adequabilidade do modelo. Cook e Tsai (1985) propuseram um resíduo para modelos de regressão não linear, mas com parâmetro de dispersão constante, que têm distribuição aproximadamente normal de média zero e variância dependendo de ϕ .

Cysneiros e Vanegas (2008) propuseram resíduos padronizados na classe dos modelos não-lineares simétricos. Eles verificaram através de estudos de simulação que o resíduo componente do desvio e o resíduo quantal têm média zero e variância 1, assimetria desprezível e curtose 3 de acordo com os momentos da distribuição normal padrão.

Propomos corrigir, até a ordem n^{-1} , os dois primeiros momentos de r_i para obtermos propriedades similares às do i -ésimo erro $\epsilon_i = y_i - \mu_i$. A expansão em série de Taylor de $r_i - \epsilon_i$, sendo $r_i = f(y_i, \hat{\boldsymbol{\beta}}, \phi_i)$ e $\epsilon_i = f(y_i, \boldsymbol{\beta}, \phi_i)$ em termos de $\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r$, $r = 1, \dots, p$, até a ordem n^{-1} , é dada pela seguinte expressão:

$$r_i = \epsilon_i + (\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r) H_r^i + \frac{1}{2} (\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r) (\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r) H_{rs}^i, \quad (2.18)$$

em que

$$H_r^i = \frac{\partial f(y_i, \boldsymbol{\beta}, \phi_i)}{\partial \beta_r} = \frac{\partial f(y_i, \boldsymbol{\beta}, \phi_i)}{\partial \mu_i} \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r} \quad (2.19)$$

e

$$H_{rs}^i = \frac{\partial^2 f(y_i, \boldsymbol{\beta}, \phi_i)}{\partial \beta_r \partial \beta_s} = \frac{\partial f(y_i, \boldsymbol{\beta}, \phi_i)}{\partial \mu_i} \frac{\partial^2 g^{-1}(\eta_i)}{\partial \eta_i^2} \frac{\partial^2 \eta_i}{\partial \beta_r \partial \beta_s} \quad (2.20)$$

Seja $f(y_i, \boldsymbol{\beta}, \phi_i) = y_i - \mu_i$, então temos por (2.19) e (2.20) que $H_r^i = -d_{ir}$ e $H_{rs}^i = g_{irs}$, em que $d_{ir} = \partial \mu_i \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r} = G_i x_{ir}$ e $E(H_{rs}^i) = \frac{\partial^2 g^{-1}(\eta_i)}{\partial \eta_i^2} \frac{\partial^2 \eta_i}{\partial \beta_r \partial \beta_s} = G_i^2 \frac{\partial^2 \eta_i}{\partial \beta_r \partial \beta_s} = 0$.

Assim, aplicando a esperança na expressão (2.18) e procedendo de forma análoga a Cox e Snell obtemos

$$E(r_i) = E(\epsilon_i) + E(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r) E(H_r^i) + I^{rs} E(H_r^i U_s^i + \frac{1}{2} H_{rs}^i) \quad (2.21)$$

sendo que no lado direito de (2.21) aplica-se a convenção da soma para $r, s = 1, \dots, p$, $\epsilon_i = f(y_i, \boldsymbol{\beta}, \phi_i) = y_i - \mu_i$, $E(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r)$ é o viés de ordem n^{-1} , que nesse caso é nulo, U_s^i é a função escore $\frac{\partial L(\boldsymbol{\theta})}{\partial \beta_s}$ e I^{rs} denota o (rs) -ésimo elemento da matriz $\mathbf{K}_{\boldsymbol{\beta}}^{-1}$.

Então, encontramos $E(\epsilon_i) = 0$, $E(H_r^i U_s^i) = E\left[\frac{2}{\phi_i} W_a(u_i)(y_i - \mu_i) G_i^2 x_{ir} x_{is}\right] = 0$, $E(H_r^i) = -G_i x_{ir}$ e $E(H_{rs}^i) = 0 \forall (r, s)$ de modo que $E(r_i) = 0$.

Vamos calcular agora $E(r_i^2)$. Assim,

$$E(r_i^2) = E(\epsilon_i^2) + 2E(\hat{\beta}_r - \beta_r)E(\epsilon_i H_r^i) + 2I^{rs} E\left[\epsilon_i H_r^i U_s^i \frac{1}{2} H_r^i H_s^i + \frac{1}{2} \epsilon_i H_{rs}^i\right], \quad (2.22)$$

com $\epsilon_i^2 = (y_i - \mu_i)^2$, $\epsilon_i H_r^i = -(y_i - \mu_i) G_i x_{ir}$, $\epsilon_i H_r^i U_s^i = 2W_a(u_i) u_i G_i^2 x_{ir} x_{is}$, $\frac{1}{2} H_r^i H_s^i = \frac{1}{2} G_i^2 x_{ir} x_{is}$ e $\frac{1}{2} \epsilon_i H_{rs}^i = 0$.

Aplicando a esperança obtemos $E(\epsilon_i^2) = \xi \phi_i$, $E(\epsilon_i H_r^i) = 0$, $E(\epsilon_i H_r^i U_s^i) = -G_i x_{ir} x_{is}$, $E(\epsilon_i H_{rs}^i) = 0$ e $E(H_r^i H_s^i) = G_i^2 x_{ir} x_{is}$. Assim, até a ordem n^{-1}

$$\begin{aligned} E(r_i^2) &= \xi \phi_i - I^{rs} G_i^2 x_{ir} x_{is} \\ Var(r_i) &= \xi \phi_i - (4d_a)^{-1} (\mathbf{X}^T \mathbf{D}(\mathbf{G}^2) \mathbf{D}(\phi)^{-1} \mathbf{X})^{-1} G_i^2 x_{ir} x_{is} \\ &= \xi \phi_i \{1 - (4d_a \xi)^{-1} a_{rs} G_i^2 x_{ir} x_{is}\} \\ &= \xi \phi_i \{1 - (4d_a \xi)^{-1} h_{ii}\}, \end{aligned}$$

em que a_{rs} é o (r, s) -ésimo elemento de $(\mathbf{X}^T \mathbf{D}(\mathbf{G}^2) \mathbf{D}(\phi)^{-1} \mathbf{X})^{-1}$, $D(\phi)^{-\frac{1}{2}} = diag\{\phi_1^{-\frac{1}{2}}, \dots, \phi_n^{-\frac{1}{2}}\}$ e $h_{ij} = \phi_i^{-\frac{1}{2}} G_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{D}(\mathbf{G}^2) \mathbf{D}(\phi)^{-1} \mathbf{X})^{-1} \mathbf{x}_j G_j \phi_j^{-\frac{1}{2}}$. Ainda, obtemos até ordem n^{-1} , $E(r_i r_j) = -I^{rs} G_i x_{ir} G_j x_{js}$. Sendo assim, $Cov(r_i, r_j) = -\phi_i \xi (4f_a \xi)^{-1} h_{ij}$ para $i \neq j$. Portanto em notação matricial temos que a matriz de variância-covariância do vetor de resíduos ordinários pode ser expressa na forma

$$Var(\mathbf{r}) = \xi \mathbf{D}(\phi) \{ \mathbf{I}_n - (4d_a \xi)^{-1} \mathbf{H} \}, \quad (2.23)$$

em que $\mathbf{H} = \mathbf{D}(\phi)^{-\frac{1}{2}} \mathbf{D}(\mathbf{G}) \mathbf{X} (\mathbf{X}^T \mathbf{D}(\mathbf{G}^2) \mathbf{D}(\phi)^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\phi)^{-\frac{1}{2}}$.

Como os r_i 's podem ter variâncias diferentes, é conveniente expressá-las em forma padronizada, para podermos comparar os resíduos. O resíduo padronizado é calculado subtraindo pela média e dividindo pelo desvio-padrão, obtendo a expressão

$$\begin{aligned} t_{r_i} &= \frac{r_i}{(\xi \phi_i)^{\frac{1}{2}} \left\{ 1 - (4d_a \xi)^{-1} \hat{h}_{ii} \right\}^{\frac{1}{2}}} \\ &= \frac{y_i - \hat{\mu}_i}{(\xi \phi_i)^{\frac{1}{2}} \left\{ 1 - (4d_a \xi)^{-1} \hat{h}_{ii} \right\}^{\frac{1}{2}}}, \quad i = 1, \dots, n. \end{aligned} \quad (2.24)$$

Usaremos este resíduo para analisarmos conjuntos de dados reais no capítulo 3.

2.7 Métodos de Diagnóstico

Como dissemos na seção anterior, é importante verificar se o modelos que ajustamos se adequa bem aos dados e se existem observações aberrantes. Esta análise é feita assumindo o modelo como correto e investigando se as conclusões feitas são coerentes.

Uma observação é dita influente se ela alterar o valor das estimativas quando a retirarmos do conjunto de dados ou quando ela é submetida a perturbações. E é dita ponto de alavancagem se quando a retiramos, ela interfere na variância dos valores ajustados.

Em modelos normais lineares, uma das medidas de diagnósticos mais utilizadas é a análise da matriz de projeção $\mathbf{H}$, ela é também conhecida como matriz chapéu ou hat. A diagonal dessa matriz representa a influência do valor observado sobre o respectivo valor predito, e este valor representa o peso desproporcional que tal observação exerce sobre seu valor estimado. Nesta seção iremos apresentar um extensão dessa alavancagem para modelos lineares generalizados simétricos heteroscedásticos.

Mas os elementos do conjunto que efetivamente controlam aspectos da análise são os pontos influentes. Estes pontos mudam a estimativa da inclinação da reta estimada e podem alterar completamente a inferência dos resultados. Iremos, nesta seção, também apresentar um método para detectarmos influência local para modelos lineares generalizados simétricos heteroscedásticos.

2.7.1 Alavancagem Generalizada

Os pontos de alavanca são pontos que elevam a variância dos valores ajustados, o que prejudica na previsão de dados. Para a detecção desses pontos vamos estudar a influência de cada observação em seu próprio valor ajustado. Segundo Wei et al. (1998), a matriz $\partial \hat{\mathbf{y}} / \partial \mathbf{y}$ pode ser obtida da forma geral

$$GL(\boldsymbol{\theta}) = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}} = \left\{ \mathbf{D}_{\boldsymbol{\theta}} \left(-\ddot{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} \right)^{-1} \ddot{L}_{\boldsymbol{\theta}\mathbf{y}} \right\} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (2.25)$$

em que $\mathbf{D}_{\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}$, $\ddot{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ e $\ddot{L}_{\boldsymbol{\theta}\mathbf{y}} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \mathbf{y}^T}$.

Como foi feito na seção (2.3), temos já calculado $\ddot{L}_{\theta\theta}$. Sendo $\mathbf{D}_{\theta} = (\mathbf{D}_{\beta}, \mathbf{0})$, temos que

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}.$$

Em forma matricial,

$$\mathbf{D}_{\beta} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = \mathbf{D}(\mathbf{G})\mathbf{X}. \quad (2.26)$$

Daí, temos que

$$\mathbf{D}_{\theta} = (\mathbf{D}(\mathbf{G})\mathbf{X}, \mathbf{0}). \quad (2.27)$$

Também temos que o $(j - \ell)$ -ésimo elemento de $\ddot{L}_{\beta\mathbf{y}}$ é dado por

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial y_{\ell}} &= -2 \sum_{i=1}^n \frac{x_{ij}}{\phi_i} G_i [2W'_a(u_i)u_i + W_a(u_i)] \\ &= \sum_{i=1}^n \frac{x_{ij}}{\phi_i} G_i [-2W_a(u_i) - 4W'_a(u_i)u_i] \\ &= \sum_{i=1}^n \frac{x_{ij}}{\phi_i} G_i [v_i - 4W'_a(u_i)u_i] \\ &= \sum_{i=1}^n x_{ij} G_i [v_i - 4W'_a(u_i)u_i] \frac{1}{\phi_i} \\ &= \sum_{i=1}^n x_{ij} G_i a_i, \end{aligned}$$

em que $a_i = [v_i - 4W'_a(u_i)u_i] \frac{1}{\phi_i}$. Em forma matricial temos,

$$\ddot{L}_{\beta\mathbf{y}} = \mathbf{X}^T \mathbf{D}(\mathbf{G})\mathbf{D}(\mathbf{a}). \quad (2.28)$$

E, o $(j - \ell)$ -ésimo elemento de $\ddot{L}_{\gamma\mathbf{y}}$ é dado por

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \gamma_j \partial y_{\ell}} &= z_{ij} w_i \left(-2W'_a(u_i)u_i \frac{2}{\phi_i} (y_i - \mu_i) - \frac{2}{\phi_i} W_a(u_i) 2(y_i - \mu_i) \right) \\ &= z_{ij} w_i \left(-4W'_a(u_i)u_i \frac{(y_i - \mu_i)}{\phi_i} - 4 \frac{(y_i - \mu_i)}{\phi_i} W_a(u_i) \right) \\ &= -2z_{ij} w_i \frac{(y_i - \mu_i)}{\phi_i} (2W'_a(u_i)u_i + 2W_a(u_i)) \\ &= -2z_{ij} \frac{h'(\tau_i)}{\phi_i^2} (y_i - \mu_i) (W'_a(u_i)u_i + W_a(u_i)) \\ &= -2z_{ij} d_i, \end{aligned}$$

Em forma matrcial temos,

$$\ddot{L}_{\gamma y} = -2\mathbf{Z}^T \mathbf{D}(\mathbf{d}). \quad (2.29)$$

Das equações (2.28) e (2.29) temos

$$\ddot{L}_{\theta y} = \begin{bmatrix} \mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\mathbf{a}) \\ -2\mathbf{Z}^T \mathbf{D}(\mathbf{d}) \end{bmatrix}. \quad (2.30)$$

Substituindo (2.30), (2.27) e (2.3) em (2.25) temos que a matriz de pontos de alavanca toma a forma

$$GL^\mu(\boldsymbol{\theta}) = \mathbf{D}(\mathbf{G}) \mathbf{X} \left\{ \ddot{L}_{\beta\beta}^{-1} + \mathbf{F} \mathbf{E} \mathbf{F}^T \right\} \ddot{L}_{\beta y} - \mathbf{D}(\mathbf{G}) \mathbf{X} \mathbf{F} \mathbf{E} \ddot{L}_{\gamma y} \quad (2.31)$$

em que $\mathbf{F} = \ddot{L}_{\beta\beta}^{-1} \ddot{L}_{\beta\gamma}$, $\mathbf{E} = \left(\ddot{L}_{\gamma\gamma} - \ddot{L}_{\beta\gamma} \ddot{L}_{\beta\beta}^{-1} \ddot{L}_{\beta\gamma} \right)^{-1}$, e $\ddot{L}_{\beta\beta}$, $\ddot{L}_{\beta\gamma}$ são elementos da matriz de informação observada (2.11).

Podemos, rescrever (2.31), da seguinte forma,

$$GL^\mu(\boldsymbol{\theta}) = GL_\beta^\mu(\boldsymbol{\theta}) + GL_\gamma^\mu(\boldsymbol{\theta}),$$

em que

$$GL_\beta^\mu(\boldsymbol{\theta}) = \mathbf{D}(\mathbf{G}) \mathbf{X} (\mathbf{X}^T \mathbf{D}(\mathbf{b}) \mathbf{D}^2(\mathbf{G}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\mathbf{a})$$

e

$$GL_\gamma^\mu(\boldsymbol{\theta}) = \mathbf{D}(\mathbf{G}) \mathbf{X} (\mathbf{F} \mathbf{E}^{-1} \mathbf{F}^T)^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\mathbf{a}) + 2\mathbf{D}(\mathbf{G}) \mathbf{X} \mathbf{F} \mathbf{E} \mathbf{Z}^T \mathbf{D}(\mathbf{d}).$$

Também é interessante observarmos a influência de y_i em $\hat{\phi}_i$, denotada por $\partial \hat{\phi}_i / \partial y_i$, em que $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$. Então, similarmente a $GL^\mu(\boldsymbol{\theta})$, a matriz de pontos de alavanca $\partial \hat{\boldsymbol{\phi}} / \partial \mathbf{y}$ pode ser expressa como

$$GL^\phi(\boldsymbol{\theta}) = GL_\gamma^\phi(\boldsymbol{\theta}) + GL_\beta^\phi(\boldsymbol{\theta}),$$

em que

$$GL_\gamma^\phi(\boldsymbol{\theta}) = -2\mathbf{Z} (\mathbf{Z}^T \mathbf{D}(\mathbf{c}) \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{D}(\mathbf{d})$$

e

$$GL_\beta^\phi(\boldsymbol{\theta}) = 2\mathbf{Z} \mathbf{A} \mathbf{B}^{-1} \mathbf{A}^T \mathbf{Z}^T \mathbf{D}(\mathbf{d}) - \mathbf{Z} \mathbf{A} \mathbf{B}^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\mathbf{a}),$$

em que $\mathbf{A} = 2(\mathbf{Z}^T \mathbf{D}(\mathbf{c}) \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{D}(\mathbf{d}) \mathbf{D}(\mathbf{G}) \mathbf{X}^T$ e $\mathbf{B} = -\mathbf{X}^T \mathbf{D}^2(\mathbf{G}) \mathbf{D}(\mathbf{a}) \mathbf{X} + 2\mathbf{X}^T \mathbf{D}(\mathbf{G}) \mathbf{D}(\mathbf{d}) \mathbf{Z} \mathbf{A}$.

Como sugestão o gráfico $\widehat{GL}(\boldsymbol{\theta})_{ii}$ versus índice pode revelar observações com alta alavanca-gem.

2.7.2 Influência local

Um método muito conhecido na detecção de pontos influentes no modelo é o método de influência local proposto por Cook (1986), que avalia a sensibilidade dos estimadores de máxima verossimilhança com respeito a perturbações no modelo ou nos dados.

Para um conjunto de dados observados seja $L(\boldsymbol{\theta})$ a log-verossimilhança do modelo onde $\boldsymbol{\theta}$ é um vetor $(p+q) \times 1$ de parâmetros desconhecidos. Podemos introduzir um vetor de perturbação $\boldsymbol{\delta}$ de dimensão $n \times 1$ cuja log-verossimilhança é $L(\boldsymbol{\theta}_{\boldsymbol{\delta}})$, e comparar os estimadores de máxima verossimilhança para se determinar influência local. O método de Cook calcula LD , ou seja, *desvio local*, que nada mais é do que a diferença das verossimilhanças avaliadas em seus estimadores ($\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}$). Grandes valores de LD indicam que as estimações são altamente sensíveis à perturbação. Em linguagem matemática temos que,

$$LD(\boldsymbol{\delta}) = 2 \left\{ L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}) \right\}. \quad (2.32)$$

Um gráfico de $LD(\boldsymbol{\delta})$ versus $\boldsymbol{\delta}$ contém informação sobre a influência do tipo de perturbação escolhido. Este gráfico forma uma superfície chamada de *gráfico de influência* (denotaremos por $\boldsymbol{\alpha}(\boldsymbol{\delta})$). Um estudo de influência local consiste em analisar como a superfície se desvia de seu plano tangente em $\boldsymbol{\delta}_0$, onde $\boldsymbol{\delta}_0$ é um vetor de não perturbação com mesma dimensão de $\boldsymbol{\delta}$ ou menor, tal que $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}_{\boldsymbol{\delta}_0})$. Esta análise pode ser feita se estudando as curvaturas das seções normais à superfície $\boldsymbol{\alpha}(\boldsymbol{\delta})$ em $\boldsymbol{\delta}_0$, que são interseções de $\boldsymbol{\alpha}(\boldsymbol{\delta})$ com o plano que contém o vetor que é normal ao seu plano tangente em $\boldsymbol{\delta}_0$.

Para construir uma seção normal, considere uma linha aumentada (*lifted line*) em $\boldsymbol{\delta}$ que passa através de $\boldsymbol{\delta}_0$. Esta linha pode ser representada por:

$$\boldsymbol{\delta}(\boldsymbol{a}) = \boldsymbol{\delta}_0 + \boldsymbol{a}\boldsymbol{t},$$

onde $\boldsymbol{a} \in \mathbb{R}$ e $\boldsymbol{t}$ é um vetor unitário, ou seja, tem comprimento 1.

A curvatura $\boldsymbol{C}_{\boldsymbol{t}}$, que é a curvatura da seção normal de $\boldsymbol{\alpha}(\boldsymbol{\delta})$ em $\boldsymbol{\delta}_0$ na direção de $\boldsymbol{t}$, é usada para caracterizar o comportamento do gráfico de influência em uma vizinhança de $\boldsymbol{\delta}_0$. A curvatura é

dada por:

$$C_t(\theta) = 2 \left| t^T \Delta^T \ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} \Delta t \right|, \quad (2.33)$$

em que $\ddot{L}_{\hat{\theta}\hat{\theta}}^{-1} = \frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^T}$ e Δ é uma matriz $(p+q) \times n$ que depende do esquema de perturbação e cujos elementos são dados por $\Delta_{ij} = \frac{\partial^2 L(\theta|\delta)}{\partial \theta_i \partial \delta_j}$, com $i = 1, \dots, p+q$ e $j = 1, \dots, n$ avaliado em $\hat{\theta}$ e δ_0 .

Perturbação é qualquer modificação nas suposições do modelo ou nos dados que causam diferenças substanciais nos resultados da análise. Os esquemas mais conhecidos são: perturbação de casos ponderados, perturbação aditiva na resposta, perturbação multiplicativa na resposta, perturbação de escala e perturbação do preditor. Neste trabalho vamos nos ater aos dois primeiros casos.

Perturbação de casos ponderados

Com este tipo de perturbação se deseja avaliar se a contribuição das observações com ponderações afetam o estimador de máxima verossimilhança. A função de log-verossimilhança do modelo perturbado é dado por:

$$L(\theta|\delta) = \sum_{i=1}^n \delta_i \log \left\{ a(u_i) / \sqrt{\phi_i} \right\}, \quad (2.34)$$

onde $0 \leq \delta_i \leq 1$. Vamos assumir que a matriz $\hat{\Delta}^T = (\hat{\Delta}_1^T, \hat{\Delta}_2^T)$ onde o $(j\ell)$ -ésimo elemento de $\hat{\Delta}_1^T$ é dado por $\frac{\partial^2 L(\theta|\delta)}{\partial \beta_j \partial \delta_\ell} \Big|_{(\beta=\hat{\beta}, \delta=\delta_0)}$, analogamente temos que o $(j\ell)$ -ésimo elemento de $\hat{\Delta}_2^T$ é dado por $\frac{\partial^2 L(\theta|\delta)}{\partial \gamma_j \partial \delta_\ell} \Big|_{(\gamma=\hat{\gamma}, \delta=\delta_0)}$.

Assim, temos que

$$\begin{aligned} \frac{\partial L(\theta|\delta)}{\partial \beta_j} &= - \sum_{i=1}^n \left[\delta_i \frac{a'(u_i)}{a(u_i)} \frac{2(y_i - \mu_i)}{\phi_i} \frac{\partial g^{-1}(\eta_i)}{\partial \eta_j} x_{ij} \right] \\ &= -2 \sum_{i=1}^n \delta_i W_a(u_i) \frac{(y_i - \mu_i)}{\phi_i} G_i x_{ij} \\ &= \sum_{i=1}^n \delta_i v_i \frac{(y_i - \mu_i)}{\phi_i} G_i x_{ij}. \end{aligned}$$

Portanto,

$$\frac{\partial^2 L(\theta|\delta)}{\partial \beta_j \partial \delta_\ell} = v_\ell \frac{(y_\ell - \mu_\ell)}{\phi_\ell} G_\ell x_{\ell j} \quad \forall \ell = 1, \dots, n.$$

Em forma matricial, temos

$$\hat{\Delta}_1^T = D(\hat{v})D^{-1}(\hat{\phi})D(r)D(G)X, \quad (2.35)$$

onde $D(r) = \text{diag}\{r_1, \dots, r_n\}$ com $r_i = y_i - \hat{\mu}_i$. Para $\hat{\Delta}_2^T$ temos que seu $(j\ell)$ -ésimo elemento é dado por

$$\begin{aligned} \frac{\partial L(\theta|\delta)}{\partial \gamma_j} &= \sum_{i=1}^n \delta_i \frac{h'(\tau_i)}{2\phi_i} \left[-2 \frac{a'(u_i)}{a(u_i)} \frac{1}{\phi_i} (y_i - \mu_i)^2 - 1 \right] z_{ij} \\ &= \sum_{i=1}^n \delta_i w_i (-2W_a(u_i)u_i - 1) z_{ij} \\ &= \sum_{i=1}^n \delta_i w_i (v_i u_i - 1) z_{ij} \\ &= \sum_{i=1}^n \delta_i w_i (m_i - 1) z_{ij} \end{aligned}$$

Assim,

$$\frac{\partial^2 L(\theta|\delta)}{\partial \gamma_j \partial \delta_\ell} = w_\ell (m_\ell - 1) z_{\ell j}, \quad \forall \ell = 1, \dots, n.$$

Na forma matricial,

$$\hat{\Delta}_2^T = D(\hat{w}) [D(\hat{m}) - I_q] Z, \quad (2.36)$$

onde I_q é a matriz identidade de ordem q .

Perturbação aditiva na resposta

Considere uma perturbação aditiva sobre a i -ésima resposta, isto é, $y_{it} = y_i + \delta s_i$, onde s_i pode ser uma estimativa do desvio padronizado de y_i e $\delta \in \mathbb{R}$. Então, temos o modelo:

$$L(\theta|\delta) = \sum_{i=1}^n \log \left\{ a(u_{it}) / \sqrt{\phi_i} \right\} \quad (2.37)$$

onde $u_{it} = (y_{it} - \mu_i)^2 / \phi_i$.

Ao derivarmos em relação a β_j

$$\frac{\partial L(\theta|\delta)}{\partial \beta_j} = \sum_{i=1}^n v_{it} \frac{(y_i + \delta s_i - \mu_i)}{\phi_i} G_i x_{ij}.$$

Para obtermos $\hat{\Delta}_1^T$ temos que seu $(j\ell)$ -ésimo elemento é dado por

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\delta})}{\partial \beta_j \partial \delta_\ell} &= \sum_{i=1}^n \frac{G_i x_{ij}}{\phi_i} [-2W'_a(u_{it})2u_{it}s_i - 2W_a(u_{it})s_i] \\ &= \sum_{i=1}^n \frac{G_i x_{ij}}{\phi_i} [-2W_a(u_{it}) - 4W'_a(u_{it})u_{it}] s_i \\ &= \sum_{i=1}^n \frac{1}{\phi_i} [v_{it} - 4W'_a(u_{it})u_{it}] s_i G_i x_{ij} \\ &= \sum_{i=1}^n a_{it} G_i s_i x_{ij}. \end{aligned}$$

Na forma matricial,

$$\hat{\Delta}_1^T = \mathbf{D}(\hat{\mathbf{a}}_t) \mathbf{D}(\hat{\mathbf{G}}) \mathbf{D}(\mathbf{s}) \mathbf{X}. \quad (2.38)$$

Ao derivarmos $L(\boldsymbol{\theta}|\boldsymbol{\delta})$ em relação a γ_j , temos

$$\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{\delta})}{\partial \gamma_j} = \sum_{i=1}^n \frac{h'(\tau_i)}{2\phi_i} [-2W'_a(u_{it})u_{it} - 1] z_{ij}.$$

Para obtermos $\hat{\Delta}_2^T$ temos que seu $(j\ell)$ -ésimo elemento é dado por

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\delta})}{\partial \gamma_j \partial \delta_\ell} &= \sum_{i=1}^n w_i \left[-2W'_a(u_{it})u_{it}2\frac{(y_i + t_i s_i - \mu_i)}{\phi_i} s_i - 4W_a(u_{it})\frac{(y_i + t_i s_i - \mu_i)}{\phi_i} s_i \right] z_{ij} \\ &= -2 \sum_{i=1}^n \frac{h'(\tau_i)}{2\phi_i} (y_i + t_i s_i - \mu_i) [W'_a(u_{it})u_{it} + W_a(u_{it})] s_i z_{ij} \\ &= -2 \sum_{i=1}^n d_{it} s_i z_{ij}, \end{aligned}$$

onde $d_{it} = \frac{h'(\tau_i)}{2\phi_i} (y_i + t_i s_i - \mu_i) [W'_a(u_{it})u_{it} + W_a(u_{it})]$. Na forma matricial temos,

$$\hat{\Delta}_2^T = -2\mathbf{D}(\hat{\mathbf{d}}_t) \mathbf{D}(\mathbf{s}) \mathbf{Z}. \quad (2.39)$$

Utilizando as derivadas da matriz Δ para os esquemas de perturbação e calculando a curvatura da equação (2.33), podemos avaliar quais pontos são influentes no modelo proposto de acordo com cada caso. Um gráfico interessante de se observar é o gráfico $\mathbf{C}_t$ versus índice.

Escobar e Meeker (1992) sugerem que os elementos da diagonal principal da matriz $\mathbf{B} = -\Delta^T \ddot{L}_{\theta\theta}^{-1} \Delta$ devem ser tomados como medida de influência. Já Lesaffre e Verbeke (1998) sugerem

que se deve avaliar a curvatura na direção da i -ésima observação, ou seja, avaliamos $\mathbf{C}_t$ no vetor $\mathbf{t}_i$ de zeros com um na i -ésima posição. Assim, obtemos que $C_i = 2|b_{ii}|$, em que b_{ii} é o elemento ii da diagonal da matriz $\mathbf{B}$. As observações tais que $C_i > 2\bar{C}$, onde $\bar{C}$ é a média das curvaturas, devem ser estudadas com cautela.

Aplicação do Método de Influência Local

Neste capítulo ilustraremos os modelos generalizados simétricos heteroscedásticos com dois exemplos de dados reais e aplicaremos a teoria de influência local apresentada no capítulo 2.

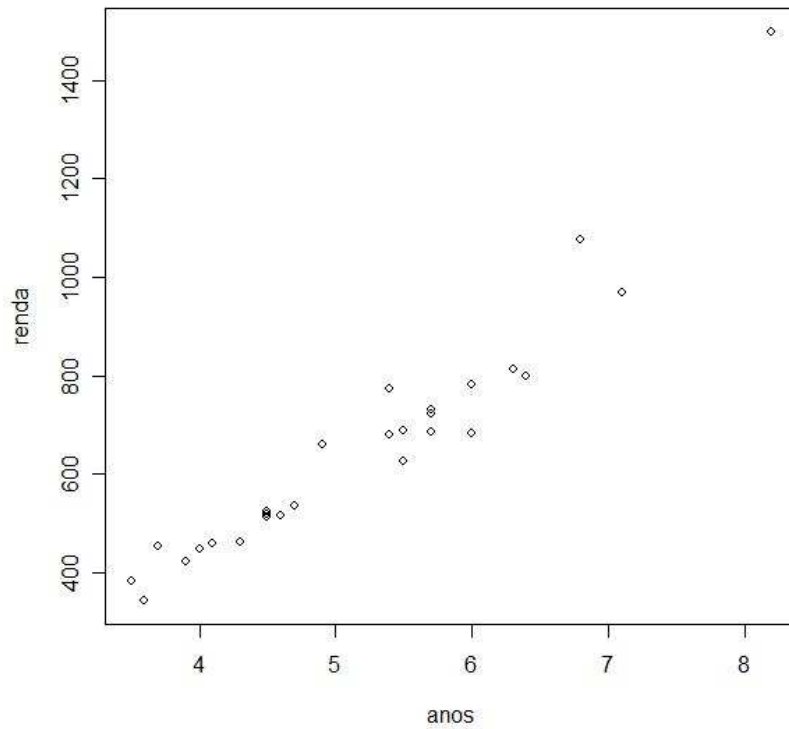
Escolaridade e Renda

Esse dados foram extraídos do censo do IBGE de (2000). Para cada unidade da federação, temos o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio. Em Paula (2004), é feita uma análise dos dados ajustando-se um modelo normal simples e também ajusta-se um modelo em que a variável resposta é transformada pela função logarítmica na tentativa de se resolver a heteroscedasticidade.

A Figura 3.1 apresenta o gráfico de dispersão dos dados. Nossa variável resposta é a renda média mensal e o número médio de anos de estudo é a variável explicativa. Inicialmente, propomos um modelo normal linear simples em que Y denota a renda e X a escolaridade. O modelo é o seguinte

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, 27,$$

onde $\epsilon_i \sim N(\mu, \sigma^2)$.



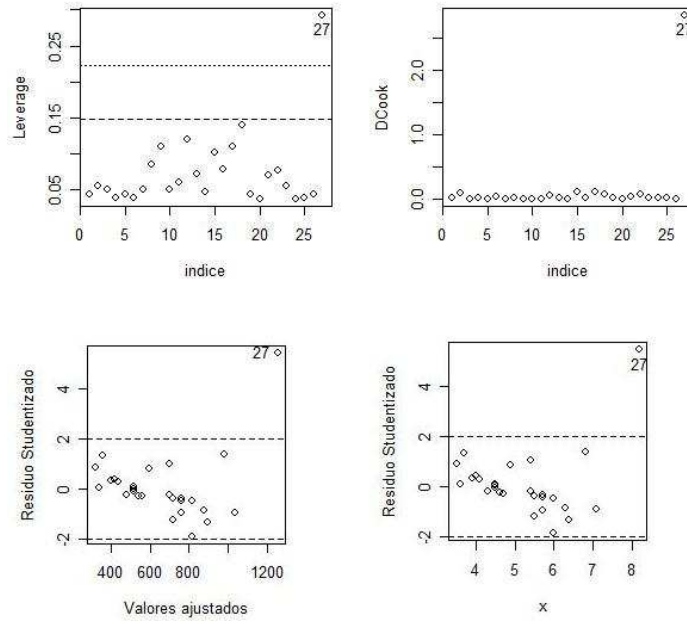


Figura 3.2: Gráficos de diagnósticos para o modelo normal linear simples.

Como foi verificado a presença de heteroscedasticidade e a presença de um ponto influente nós investigamos a adequacidade de alguns modelos lineares generalizados simétricos na tentativa de reduzir a heteroscedasticidade.

$$\mu_i = \beta_0 + \beta_1 x_i,$$

$$\phi_i = \exp(\gamma_0 + \gamma_1 z_i), \quad i = 1, \dots, 27,$$

em que $z_i = x_i - \bar{x}$ e $y_i \sim S(\mu_i, \phi_i, a)$.

Ajustamos o modelo acima para duas distribuições, a distribuição Normal e a distribuição t -Student que é uma distribuição de cauda pesada para tentar reduzir a influência de observações atípicas. Para decidirmos que grau de liberdade iremos usar na análise para a distribuição t -Student utilizaremos o critério de informação de Akaike (AIC). A Tabela seguinte apresenta os valores AIC dos modelos ajustados.

Baseado nestes resultados, vamos utilizar uma distribuição t -Student com 3 graus de liberdade, pois esta distribuição possui o menor AIC. Então com estes dois modelos, normal e

Tabela 3.1: Critério de informação de Akaike.

Modelo	Critério de informação de Akaike (AIC)
Normal	294,774
t -Student ($\nu = 3$)	294,266
t -Student ($\nu = 4$)	294,432
t -Student ($\nu = 5$)	294,539
t -Student ($\nu = 6$)	294,608

t -Student(3), vamos fazer nossa análise e verificar qual dos dois melhor se ajusta aos dados. Na Tabela abaixo se encontram a estimação dos parâmetros com seus respectivos erros padrão para os dois modelos.

Tabela 3.2: Estimação dos parâmetros e erros padrão associados.

Modelo	Normal	t -Student ($\nu = 3$)
$\hat{\beta}_0$	-210,7328 (44,6600)	-198,4205 (35,0000)
$\hat{\beta}_1$	164,4353 (10,1900)	160,1135 (8,0800)
$\hat{\gamma}_0$	7,7833 (0,2728)	3,5136 (0,1925)
$\hat{\gamma}_1$	0,9481 (0,1925)	0,5244 (0,1687)

Modelo normal

Iniciaremos a análise com a variável resposta seguindo uma distribuição Normal. Na primeira coluna da Tabela 3.2 encontram-se as estimativas dos parâmetros do modelo normal heteroscedástico ajustado. Observe que todos os parâmetros são significativos para o modelo, e o coeficiente angular para a média (β_1) é positivo, o que pode ser interpretado que se aumentarmos em um ano o tempo médio de escolaridade, a renda média domiciliar aumenta 164,4353 reais.

Agora vamos fazer uma análise de diagnóstico para o conjunto de dados. Vamos começar com a análise dos resíduos, utilizando o resíduo que proposto no capítulo 2.

Observe que a observação 15, correspondente ao estado de Alagoas se destaca na Figura (3.3) como um ponto aberrante, ou seja, ele interfere no intercepto do modelo e consequentemente altera os valores ajustados.

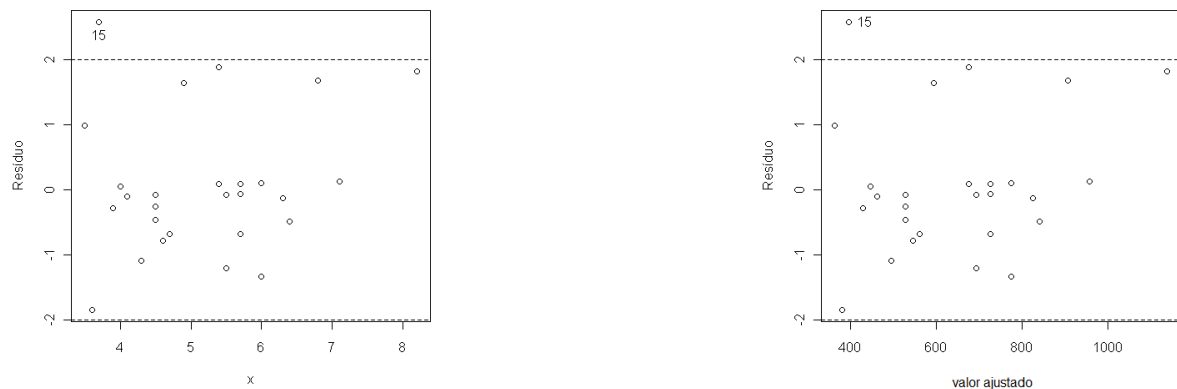


Figura 3.3: Resíduos para a distribuição normal.

As observações influentes podem acarretar em mudança nos estimadores dos parâmetros e até mesmo causar mudanças inferenciais, por isso vamos detectar tais observações e verificar se elas alteram muito os resultados do modelo ajustado. Pela Figura 3.4, podemos ver que as observações 15 e 27, respectivamente os estados de Alagoas e Distrito Federal, apresentam curvaturas mais elevadas, o que revela uma possível influência desproporcional dessas observações nas estimativas dos parâmetros.

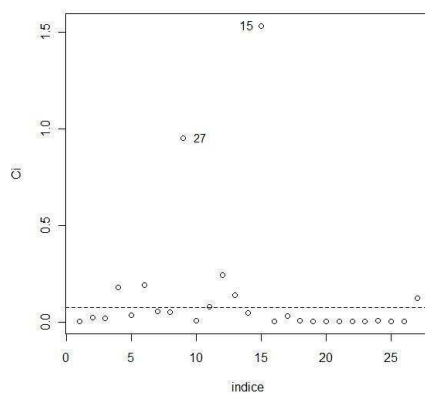


Figura 3.4: Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição Normal.

Modelo t -Student

Agora vamos analisar o mesmo modelo mas assumindo que a variável resposta seguirá uma distribuição t -Student com 3 graus de liberdade como foi decidido anteriormente. Na segunda coluna da Tabela 3.2 encontram-se as estimativas dos parâmetros do modelo t -Student(3) heteroscedástico ajustado. Observe que todos os parâmetros são significativos para o modelo, e o coeficiente angular para a média (β_1) é positivo, o que pode ser interpretado que se aumentarmos em um ano o tempo médio de escolaridade, a renda média domiciliar aumenta 160,1135 reais.

Agora vamos fazer uma análise de diagnóstico dos dados utilizando o resíduo que foi proposto no capítulo 2.

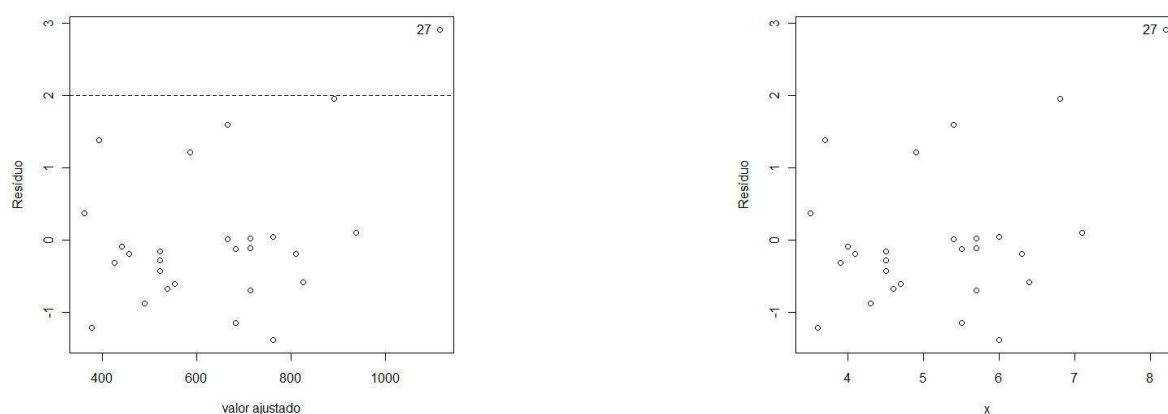


Figura 3.5: Resíduos para a distribuição t -Student(3).

Observe que a observação 27, correspondente ao Distrito Federal se destaca na Figura (3.5) como um ponto aberrante, ou seja, ele interfere no intercepto do modelo e consequentemente altera os valores ajustados.

Pela Figura 3.6, podemos ver que apenas a observação 27, referente ao Distrito Federal, apresenta curvatura mais elevada, o que revela uma maior estabilidade da distribuição t -Student do que da Normal.

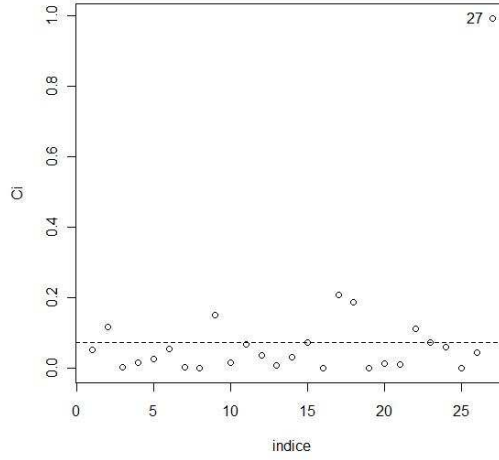


Figura 3.6: Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição t -Student(3).

Com isso, percebemos que para o esquema de perturbação proposto, os estimadores são mais sensíveis a perturbações na distribuição normal do que na t -Student. Precisamos verificar se as observações destacadas anteriormente afetam muito na estimação dos parâmetros.

Para avaliarmos o efeito que se ao eliminarmos uma ou mais observações sobre os estimadores de máxima verossimilhança vamos utilizar a razão de mudança. Esta razão de mudança é expressa por:

$$RM_j = \left| \frac{\hat{\theta} - \hat{\theta}_j}{\hat{\theta}} \right| \times 100, \quad j = 1, \dots, n,$$

em que $\hat{\theta}_j$ é o estimador de máxima verossimilhança quando se elimina a observação j . A Tabela seguinte mostra os RM's dos estimadores se eliminarmos os pontos 27 e 15.

Tabela 3.3: Razão de mudança para os estimadores máxima verossimilhança.

Caso eliminado	Modelo	$RM_{\hat{\beta}_0}$	$RM_{\hat{\beta}_1}$	$RM_{\hat{\gamma}_0}$	$RM_{\hat{\gamma}_1}$
27	Normal	2,6501	0,8738	2,9792	23,6084
27	t -Student ($\nu = 3$)	1,6915	0,4115	1,5454	21,1109
15	Normal	12,4416	2,7819	2,7956	19,2793

Podemos observar pela Tabela 3.3, nas duas distribuições possuem razão de mudança baixa para os estimadores de β , tendo um valor maior quando retiramos a observação 15 (Alagoas)

e ajustamos uma distribuição normal, ou seja, este ponto muda em 12% o estimador de β_1 e 19% o estimador de $\hat{\gamma}_1$. Ao retirarmos o ponto 27, a distribuição Normal possui o maior índice de mudança para o estimador de γ_1 , 23,6084, enquanto que a distribuição t -Student obteve um índice de mudança de 21,1109.

Na Tabela 3.4 podemos observar os valores dos estimadores com todos os pontos e com a retirada dos pontos 27 e 15 da distribuição Normal e 27 da distribuição t -Student.

Tabela 3.4: Estimativas de máxima verossimilhança dos estimadores.

	Normal	t -Student(3)	Normal sem 27	Normal sem 15	t -Student (3) sem 27
$\hat{\beta}_0$	-210,7328	-198,4205	-205,1481	-236,9514	-201,7769
$\hat{\beta}_1$	164,3530	160,1135	162,9169	168,9253	160,7724
$\hat{\gamma}_0$	7,7833	3,5136	7,5514	7,5657	3,4593
$\hat{\gamma}_1$	0,9482	0,5244	0,7243	1,1310	0,4137

A distribuição t -Student se apresenta mais estável a variações nos estimadores do que a distribuição normal, e apresentou apenas um ponto influente enquanto que a normal apresentou dois pontos influentes.

Crescimento de crianças

Eubank (1988) apresenta dados razão(peso/altura) e idade (do nascimento até 71,5 meses) de 72 crianças, e estes dados foram analisados por Tarpey (2001). Nossa variável resposta é a razão(peso/altura) e idade (em meses) é a variável explicativa. Inicialmente, propomos um modelo normal linear simples em que Y denota a resposta e X a variável explicativa. A Figura 3.7 apresenta o gráfico de dispersão deste conjunto de dados.

Inicialmente, propomos um modelo normal linear simples em que Y denota a resposta e X a variável explicativa. O modelo é dado por

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, 72,$$

em que $\epsilon_i \sim N(\mu, \sigma^2)$.

As estimativas dos parâmetros são dadas por $\hat{\beta}_0 = 66,5623$ e $\hat{\beta}_1 = 0,5276$ e seus desvios padrão são respectivamente 1,2176 e 0,0293, o que indica que β_1 é altamente significativo. A

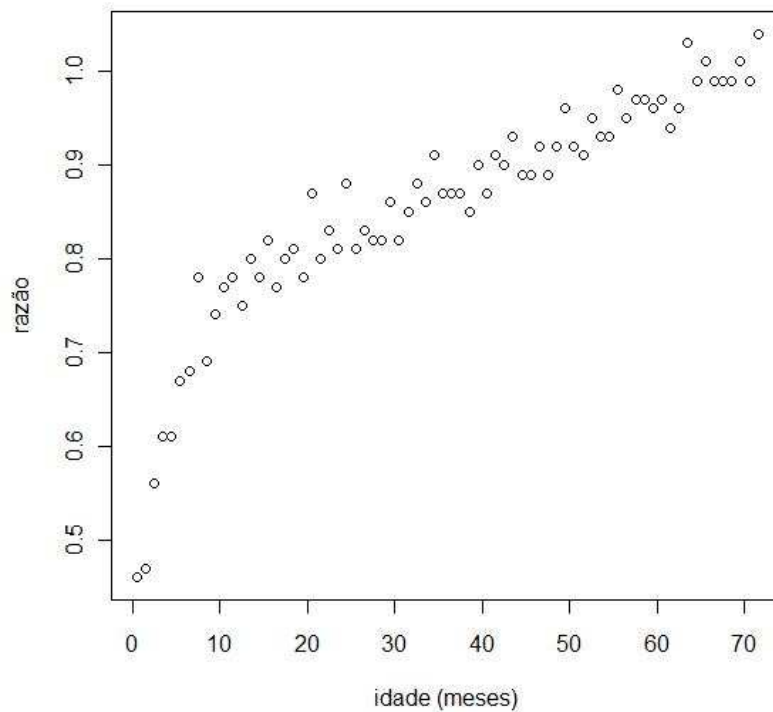


Figura 3.7: Gráfico de dispersão do crescimento de crianças e idade (meses).

estimativa de σ^2 foi de 5,165 e critério de Informação de Akaike de $AIC = 444,741$.

Observe que no gráfico de resíduo studentizado versus valores ajustados (Figura 3.8) nota-se que há indícios de heteroscedasticidade, ou seja, há uma relação não linear entre o resíduo e a variável explicativa. Note também que os pontos 1 e 2, que representam as duas primeiras razões (peso/altura) para o momento do nascimento e o primeiro mês de vida, aparecem como pontos influentes e aberrantes.

Como foi verificado a presença de heteroscedasticidade e a presença de pontos influentes nós investigamos a adequacidade de alguns modelos lineares generalizados simétricas na tentativa de reduzir o impacto desse ponto influente.

$$(\mu_i)^{1/3} = \beta_0 + \beta_1 x_i,$$

$$\phi_i = \exp(\gamma_0 + \gamma_1 z_i), \quad i = 1, \dots, 27,$$

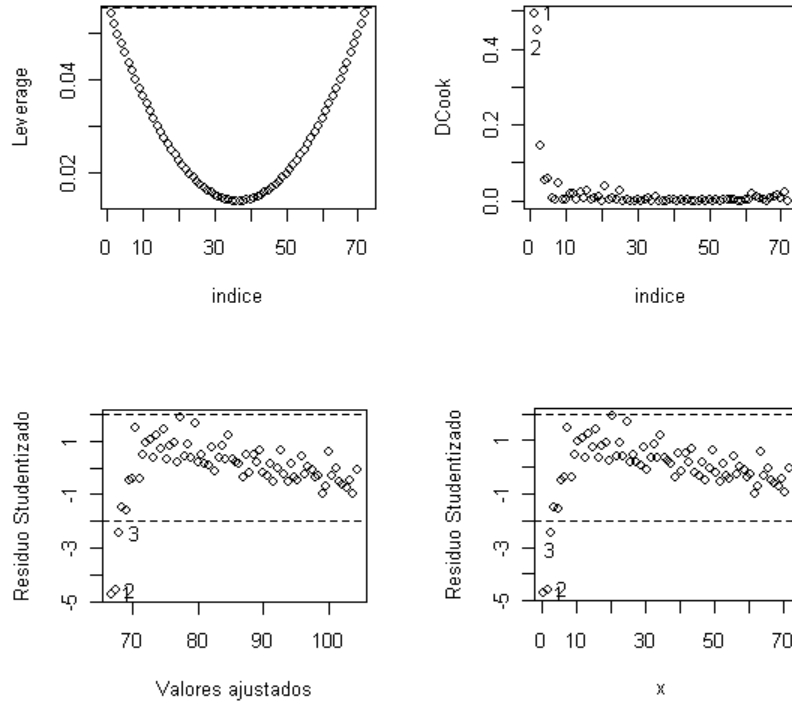


Figura 3.8: Gráficos de diagnósticos para o modelo normal linear simples.

onde $z_i = x_i - \bar{x}$ e $y_i \sim S(\mu_i, \phi_i, a)$.

Ajustamos o modelo acima para duas distribuições, a distribuição Normal e a distribuição t -Student que é uma distribuição de cauda pesada para tentar reduzir a influência de observações atípicas. Para decidirmos que grau de liberdade iremos usar na análise para a distribuição t -Student utilizaremos o critério de informação de Akaike (AIC). A Tabela seguinte apresenta os valores AIC dos modelos ajustados.

Tabela 3.5: Critério de informação de Akaike.

Modelo	Critério de informação de Akaike (AIC)
Normal	390,6184
t -Student ($\nu = 3$)	385,7104
t -Student ($\nu = 4$)	385,7617
t -Student ($\nu = 5$)	386,0822
t -Student ($\nu = 6$)	386,4420

Baseado nestes resultados, vamos utilizar uma distribuição t -Student com 5 graus de liberdade, pois esta distribuição possui o menor AIC. Então com estes dois modelos, normal e t -Student(5), vamos fazer nossa análise e verificar qual dos dois melhor se ajusta aos dados. Na Tabela abaixo se encontram a estimação dos parâmetros com seus respectivos erros padrão para os dois modelos.

Tabela 3.6: Estimação dos parâmetros e erros padrão associados.

Modelo	Normal	t -Student ($\nu = 5$)
$\hat{\beta}_0$	4,1618 (0,0201)	4,1700 (0,0178)
$\hat{\beta}_1$	0,0071 (0,0003)	0,0069 (0,0003)
$\hat{\gamma}_0$	2,4760 (0,1666)	0,9832 (0,1054)
$\hat{\gamma}_1$	-0,0577 (0,0081)	-0,0274 (0,0051)

Modelo normal

Iniciaremos a análise com a variável resposta seguindo uma distribuição Normal. Na primeira coluna da Tabela 3.6 encontram-se as estimativas dos parâmetros do modelo normal heteroscedástico ajustado. Observe que todos os parâmetros são significativos para o modelo.

Agora vamos fazer uma análise de diagnóstico para o conjunto de dados. Vamos começar com a análise dos resíduos, utilizando o resíduo que proposto no capítulo 2.

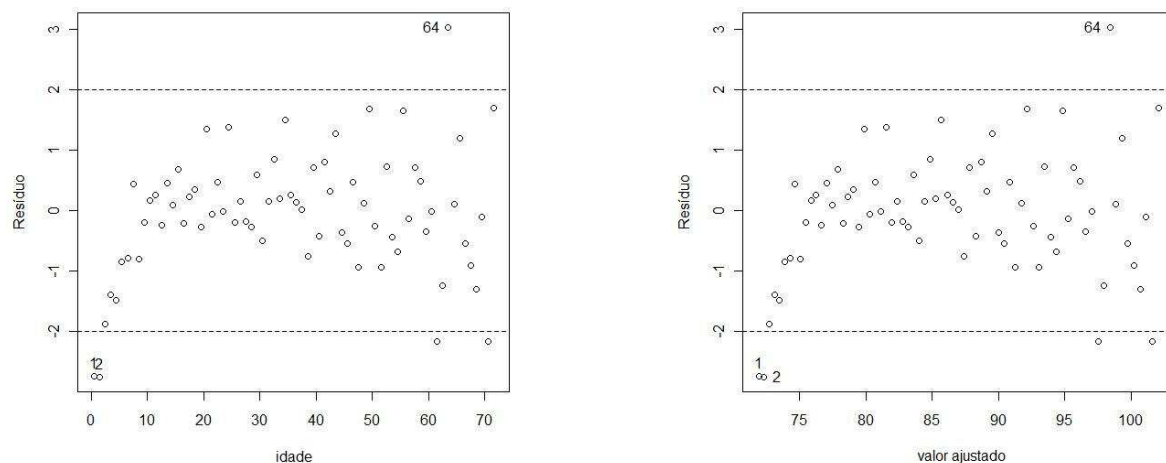


Figura 3.9: Resíduos para a distribuição normal.

Observe que as observações 1, 2 e 64, se destacam na Figura (3.9) como pontos aberrantes, ou seja, eles interferem no intercepto do modelo e consequentemente alteram os valores ajustados.

Pela Figura 3.10, podemos ver que a observação 64, apresenta curvatura mais elevada, o que revela uma possível influência desproporcional dessa observação na estimativa dos parâmetros.

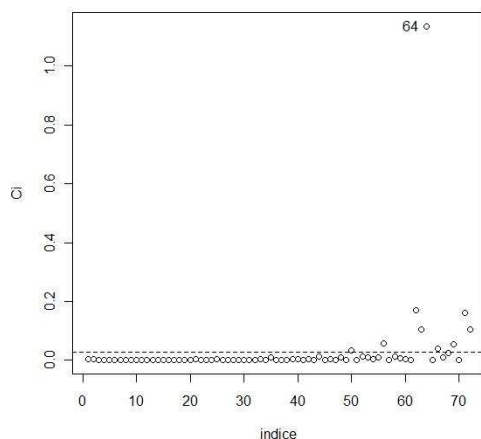


Figura 3.10: Gráficos de índice do C_i com perturbação de caso ponderado para a distribuição Normal.

Modelo t -Student

Agora vamos analisar o mesmo modelo mas a variável resposta seguirá a distribuição t -Student com 5 graus de liberdade como foi decidido anteriormente. Na segunda coluna da Tabela 3.6 encontram-se as estimativas dos parâmetros do modelo t -Student(5) heteroscedástico ajustado. Observe que todos os parâmetros são significativos para o modelo.

Agora vamos fazer uma análise de diagnóstico dos dados utilizando o resíduo que foi proposto no capítulo 2.

Observe que as observações 1 e 2 se destacam na Figura (3.11) como pontos aberrantes, ou seja, eles interferem no intercepto do modelo e consequentemente alteram os valores ajustados.

Pela Figura 3.12, podemos ver que apenas a observação 72 apresenta curvatura mais elevadas. Observe também que a observação 72 é a única que se destaca, enquanto que na Normal outras

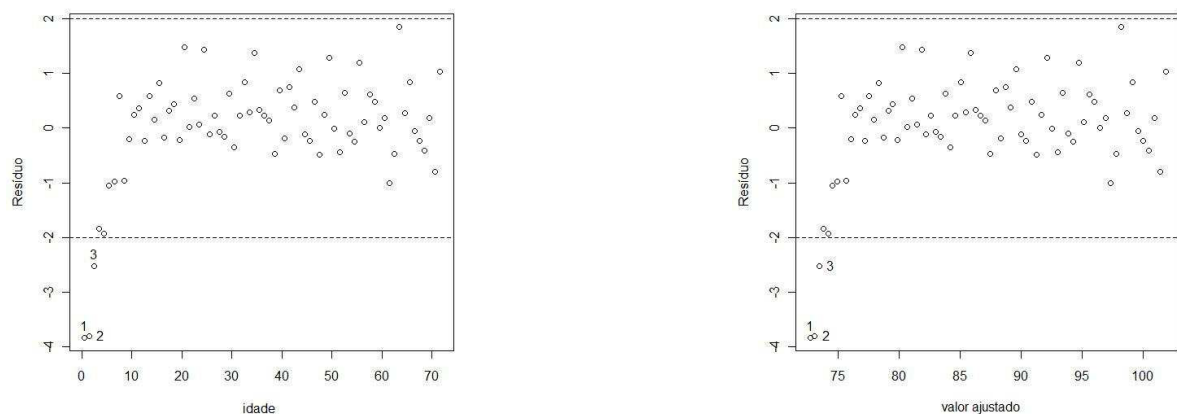


Figura 3.11: Resíduos para a distribuição t -Student(5).

observações tendem a ter curvaturas um pouco elevadas, e a curvatura de 72 na t -Student é menor do que a curvatura de 64 na normal. O que revela uma maior estabilidade da distribuição t -Student do que da Normal.

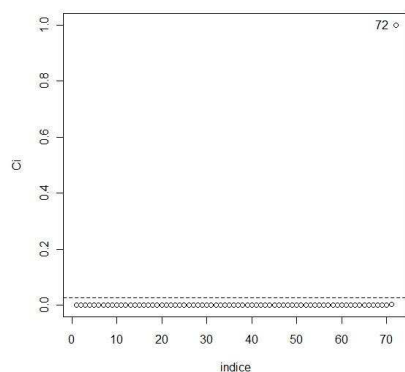


Figura 3.12: Gráfico de índice do C_i com perturbação de caso ponderado para a distribuição t -Student(5).

Para o esquema de perturbação do caso de resposta aditiva, podemos observar pela Figura 3.13 que para a distribuição Normal, a observação 72 se destaca das demais com curvatura elevada e para a distribuição t -Student a observação 70 tem curvatura mais elevada, mas a distribuição t -Student possui curvaturas menores do que a distribuição Normal.

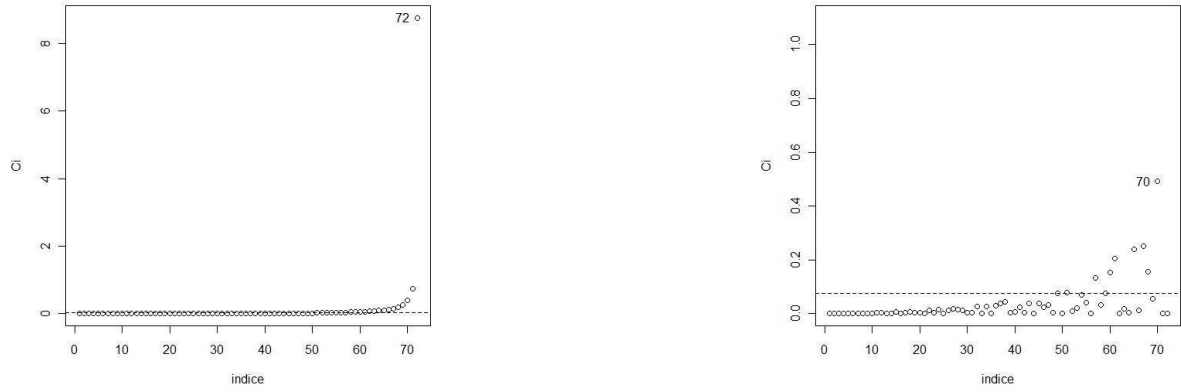


Figura 3.13: Gráficos de índice do C_i com perturbação de caso resposta aditiva para a distribuição Normal, a esquerda, e t -Student(5), a direita.

Com isso, percebemos que para os dois esquemas de perturbação propostos, os estimadores são mais sensíveis a perturbações. Precisamos verificar se as observações destacadas anteriormente afetam muito na estimação dos parâmetros. Para avaliarmos o efeito que se ao eliminarmos uma ou mais observações sobre os estimadores de máxima verossimilhança vamos utilizar a mesma razão de mudança utilizada na seção anterior. A Tabela seguinte mostra os RM's dos estimadores ao eliminarmos os pontos 1 e 2 que apareceram nas duas distribuições como pontos aberrantes.

Tabela 3.7: Razão de mudança para os estimadores máxima verossimilhança.

Caso eliminado	Modelo	$RM_{\hat{\beta}_0}$	$RM_{\hat{\beta}_1}$	$RM_{\hat{\gamma}_0}$	$RM_{\hat{\gamma}_1}$
1	Normal	0,0473	0,3918	5,4082	8,6101
1	t -Student ($\nu = 5$)	0,0184	0,1201	6,0786	6,9343
2	Normal	0,0579	0,4983	8,3117	8,5479
2	t -Student ($\nu = 5$)	0,0207	0,1416	5,9862	7,6642

Podemos observar pela Tabela 3.7, nas duas distribuições possuem razão de mudança baixa para os estimadores de β , tendo um valor maior quando retiramos a observação 64 e ajustamos uma distribuição normal, ou seja, este ponto muda em 2% o estimador de β_1 . Para os estimadores de γ , os maiores valores ocorrem para os estimadores de γ_1 . A normal possui os maiores índices de mudança para o estimador de γ_1 , 8,5479 e 8,6101 ou retirarmos 2 e 1, respectivamente. Enquanto que a t -Student o maior índice de mudança foi de 7,6642 ou retirarmos o ponto 2.

Na Tabela 3.8 podemos observar os valores dos estimadores com todos os pontos e com a retirada dos pontos 1 e 2 da distribuição normal e distribuição t -Student.

Tabela 3.8: Estimativas de máxima verossimilhança dos estimadores.

	Normal	t -Student(5)	Normal s/1	Normal s/2	t -Student(5) s/1	t -Student(3) s/2
$\hat{\beta}_0$	4,1618	4,1700	4,1637	4,1642	4,1707	4,1708
$\hat{\beta}_1$	0,0071	0,0069	0,0071	0,0071	0,0069	0,0069
$\hat{\gamma}_0$	2,4760	0,9832	2,3423	2,2702	0,9234	1,8241
$\hat{\gamma}_1$	-0,0577	-0,0274	-0,0528	-0,0528	-0,0255	-0,0253

A distribuição t -Student se mais estável a variações nos estimadores do que a distribuição normal apresentando índices de mudanças menores do que a distribuição ao eliminarmos pontos aberrantes.

Neste trabalho apresentamos uma classe geral de modelos lineares generalizados simétricos heteroscedásticos e desenvolvemos a metodologia de detecção de pontos de alavanca e influência local, com dois esquemas de perturbação, bem como desenvolvemos um processo iterativo para a estimação dos parâmetros deste modelo e testes de razão de verossimilhanças, escore e Wald para testarmos a homoscedasticidade do parâmetro de dispersão. Propomos um resíduo padronizado para essa classe de modelos.

Ajustamos os modelos lineares generalizados simétricos heteroscedásticos para dois conjuntos de dados e observamos que o problema de heteroscedasticidade quando não foi resolvido, foi suavizado, e para distribuições de caudas pesadas tendem a diminuir o número de observações aberrantes.

Várias linhas de pesquisa podem ser tratadas como se desenvolver um estudo de simulação para verificar a distribuição assintótica do resíduo proposto, avaliar a consistência dos estimadores de máxima verossimilhança de β e γ , assim como fazer estudos de simulações para se verificar o poder e o tamanho dos testes propostos neste trabalho.

APÊNDICE A

Escolaridade

Tabela A.1: Conjunto de dados sobre Escolaridade e Renda.

Estado	Escolaridade (anos)	Renda (reais)	Estado	Escolaridade (anos)	Renda (reais)
RR	5.7	685	AL	3.7	454
AP	6.0	683	CE	4.0	448
AC	4.5	526	SP	6.8	1076
RO	4.9	662	RJ	7.1	970
PA	4.7	536	ES	5.7	722
AM	5.5	627	MG	5.4	681
TO	4.5	520	SC	6.3	814
PB	3.9	423	RS	6.4	800
MA	3.6	343	PR	6.0	782
RN	4.5	513	MT	5.4	775
SE	4.3	462	GO	5.5	689
PI	3.5	383	MS	5.7	731
BA	4.1	460	DF	8.2	1499
PE	4.6	517			

APÊNDICE B

Crescimento de crianças

Tabela B.1: Conjunto de dados sobre Razão(peso/altura) e idade(meses) de crianças.

Razão	idade	Razão	idade	Razão	idade	Razão	idade
46	0.5	87	20.5	87	40.5	97	60.5
47	1.5	80	21.5	91	41.5	94	61.5
56	2.5	83	22.5	90	42.5	96	62.5
61	3.5	81	23.5	93	43.5	103	63.5
61	4.5	88	24.5	89	44.5	99	64.5
67	5.5	81	25.5	89	45.5	101	65.5
68	6.5	83	26.5	92	46.5	99	66.5
78	7.5	82	27.5	89	47.5	99	67.5
69	8.5	82	28.5	92	48.5	99	68.5
74	9.5	86	29.5	96	49.5	101	69.5
77	10.5	82	30.5	92	50.5	99	70.5
78	11.5	85	31.5	91	51.5	104	71.5
75	12.5	88	32.5	95	52.5		
80	13.5	86	33.5	93	53.5		
78	14.5	91	34.5	93	54.5		
82	15.5	87	35.5	98	55.5		
77	16.5	87	36.5	95	56.5		
80	17.5	87	37.5	97	57.5		
81	18.5	85	38.5	97	58.5		
78	19.5	90	39.5	96	59.5		

Referências Bibliográficas

- [1] Atkinson, A. C. (1985). *Plots, Transformation and regression*. Oxford: Clarendon Press.
- [2] Atkinson, A. C. (1987). Two graphical display for outlying and influential observations in regression. *Biometrika* **68**, 13 – 20.
- [3] Barroso, L. P. e Vasconcellos, K. L. P. (2002) Second-order asymptotic for score tests in heterocedastic t regression models. *Comm. Statist.-Theory Methods* **31**, 1515 – 1529.
- [4] Botter, D. A. e Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Journal of Statistical Comput. Simulation* **62**, 91 – 104.
- [5] Box, G. E. P. e Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal Royal Statistics Society* **B26**, 211 – 252.
- [6] Carrol, R. J. e Ruppert, D. (1988). On robust test for heteroscedasticity. *Ann. Statist.* **48**, 133 – 169.
- [7] Cook, R. D. (1986). Assessment of Local Influence (with discussion). *Journal Royal Statistics Society* **B48**, 133 – 169.
- [8] Cook, R. D. e Tsai, C. L. (1985). Residuals in nonlinear regression. *Biometrika* **72**, 23 – 29.
- [9] Cook, R. D. e Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1 – 10.

- [10] Cox, D. R. e Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society* **B30**, 284 – 275.
- [11] 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.
- [12] Cysneiros, F. J. A.; Galea, M. e Paula, G. A. (2007). Heteroscedastic symmetrical linear models. *Statistic and Probability Letters* **77**, 1084 – 1090.
- [13] Cysneiros, F. J. A. e Vanegas, L. (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics and Probability Letters* **78**, 3209 – 3273.
- [14] Escobar, L. A. e Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507 – 528.
- [15] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker, Inc.
- [16] Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrika* **41**, 461 – 465.
- [17] Lesaffre, F. e Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **38**, 963 – 974.
- [18] Lobos, Cristian Villegas (2004). *Influencia Local en Modelos Lineales Generalizados Elípticos*. Tese de mestrado, Universidad de Valparaíso, Chile.
- [19] Park, R. E. (1966). Estimating with heteroscedastic terms. *Econometrika* **34**, 888.
- [20] Paula, G. A. (2004). *Modelos de Regressão com apoio computacional*. São Paulo: Instituto de Matemática e Estatística - USP.
- [21] Rigby, B. e Stasinopoulos, M. (2001). The GAMLSS project: a flexible approach to statistical modelling. *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical modeling*. Odense, Denmark: Klein, B. Korsholm, L.

- [22] Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society* **B51**, 47 – 60.
- [23] Tarpey, T. (2001). A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician* **542**, 116 – 118.
- [24] Taylor, J. e Verbyla, A. P. (2004). Joint modelling of location and scale parameter of the t distribution. *Statist. Model.* **4**, 91 – 112.
- [25] Verbyla, A. P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society* **B55**, 493 – 508.
- [26] Wei, B. C.; Hu, Y. Q. e Fung, W. K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25 – 37.