MÉTODOS ROBUSTOS EM ANÁLISE DE REGRESSÃO

Maria de Fátima Brant Drumond

**DISSERTAÇÃO APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA**

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE EM ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO : ESTATÍSTICA

ORIENTADOR: PROF. DR. JÚLIO DA MOTTA SINGER

São Paulo, novembro de 1989

Defesa: 09.03.1990

MINHA GRATIDÃO ESPECIAL A

Prof. Dr. Júlio da Motta Singer

*Departamento de Estatística da Universidade Federal de
Minas Gerais*

CAPES e PICD/UFG

Prof. Dr. Oscar Bustos

Prof. Dr. José Francisco Soares

Prof. Dr. Subhash Narula

Marina Alves de Souza

Roberto Guimarães

Maria de Fátima Brant Drumond

ERRATA: MÉTODOS ROBUSTOS EM ANÁLISE DE REGRESSÃO.

Página	Linha(s)*	Onde se lê	Leia-se
04	-11	Capítulo 1	Capítulo 2
06	1	parâmetro	paramétrico
07	16	ajuste do	ajuste de
07	-10, -11	limitadas	limitada
19	15	$a_n(i)$	$a_N(i)$
20	-1	$a_n(R_i)$	$a_n(r_i)$
26	Eq. 2.20	ψ	Ψ
27	-6	$\tilde{x}^T \tilde{\beta}$	$\tilde{x}^T \tilde{\beta}$
31	3	Jurečková	Jurečková
31	-1	$(\sum_{i=1}^n x_{ij}^2)$	$(\sum_{i=1}^n x_{ij}^2)^{-1/2}$
32	3	$\tilde{X}'\tilde{X}$	$\tilde{X}_n^T \tilde{X}_n$
33	-10	computado	computador
33	-10	Roussevw	Rousseevw
38	12	$\hat{\beta}(0)$	$\tilde{\beta}(0)$
46	6	simultânea e	simultânea de β e σ
46	7	(3.4):	(3.4));
46	8	escla	escala
46	-8	g	σ
51	9	práticos;	práticos);
61	-6	$\sum_{i=1}^n \psi \frac{r_i}{\hat{\sigma}})^2/(n-p)$	$\sum_{i=1}^n \psi \left(\frac{r_i}{\hat{\sigma}}\right)^2/(n-p)$

*0 sinal - indica contagem de baixo para cima.

ÍNDICE

	<u>Página</u>
1. <u>ESTIMAÇÃO EM MODELOS DE REGRESSÃO</u>	01
1.1. INTRODUÇÃO	01
1.2. CONSIDERAÇÕES GERAIS SOBRE MÉTODOS ROBUSTOS ...	05
1.2.1. <u>Conceito e Objetivos dos Métodos Robustos</u>	05
1.2.2. <u>Tipos de Desvios do Modelo Paramétrico</u>	07
1.3. PRINCIPAIS ABORDAGENS TEÓRICAS PARA ROBUSTEZ ..	08
1.4. MEDIDAS DE ROBUSTEZ	11
1.4.1. <u>Funcionais</u>	11
1.4.2. <u>Robustez Qualitativa</u>	12
1.4.3. <u>Ponto de Ruptura</u>	12
1.4.4. <u>Função de Influência</u>	13
2. <u>CLASSES DE ESTIMADORES EM ANÁLISE DE REGRESSÃO</u>	16
2.1. INTRODUÇÃO	16
2.2. ESTIMADORES L	16
2.3. ESTIMADORES R	19
2.4. ESTIMADORES M	21
2.5. ALGUNS ESTIMADORES ALTERNATIVOS	30
2.5.1. <u>Estimadores M Tipo Huber Adaptativos e Invariantes em Relação à Escala</u>	31
2.5.2. <u>Regressão de Mínima Mediana de Quadrados ("least median of squares")</u>	33
2.5.3. <u>Estimadores MM</u>	34

	<u>Página</u>
3. <u>ESTIMADORES M</u>	36
3.1. INTRODUÇÃO	36
3.2. ESTIMADORES M INVARIANTES EM RELAÇÃO À ESCALA .	36
3.2.1. <u>Introdução</u>	36
3.2.2. <u>Algoritmos para Cálculo de Estimativas M</u> <u>Invariantes em Relação à Escala</u>	37
3.3. ALGUNS PACOTES COMPUTACIONAIS	42
3.4. ESTIMADORES PARA A MATRIZ DE COVARIÂNCIA DE $\hat{\beta}$.	44
4. <u>COMPARAÇÃO DE ALGUNS ESTIMADORES ROBUSTOS EM</u> <u>MODELOS DE REGRESSÃO</u>	46
4.1. INTRODUÇÃO	46
4.2. DEFINIÇÃO DOS ESTIMADORES E ASPECTOS COMPUTACIONAIS	46
4.3. ESTUDO DE MONTE CARLO	50
4.4. EXEMPLO	65
5. <u>CONCLUSÕES E SUGESTÕES PARA FUTUROS TRABALHOS</u>	75
6. <u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	76
<u>APÊNDICE A</u>	81

1. ESTIMAÇÃO EM MODELOS DE REGRESSÃO

1.1. INTRODUÇÃO

Pesquisadores de diversas áreas do conhecimento enfrentam, com frequência, a tarefa de investigar e modelar a relação entre uma variável resposta y e uma ou mais variáveis explicativas $x_1, x_2, \dots, x_p$. Em um experimento químico, por exemplo, a resposta y poderia ser o rendimento de uma reação para produção de ácido sulfúrico e x_1, x_2 e x_3 seriam a temperatura, tempo e pressão em que a reação foi conduzida. As técnicas estatísticas apropriadas para este tipo de problema são conhecidas sob a denominação geral de Análise de Regressão.

O modelo estatístico usual é:

$$\underline{y} = \underline{X} \underline{\beta} + \underline{\varepsilon} \quad (1.1)$$

onde $\underline{y}$ é um vetor ($n \times 1$) cujos elementos são as n observações da variável resposta; $\underline{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)^T$ é uma matriz ($n \times p$) onde a i -ésima linha contém os valores das p variáveis explicativas correspondentes à i -ésima observação ($i = 1, \dots, n$), os elementos do vetor $\underline{\beta}$ são os p parâmetros desconhecidos e os n elementos do vetor $\underline{\varepsilon}$ são erros aleatórios não-observáveis. Assumimos, usualmente, que os elementos de $\underline{X}$ são fixos e que $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ são independentes e identicamente distribuídos com função de distribuição F , média zero e variância σ^2 finita.

Em geral, o interesse está em estimar $\underline{\beta}$ e testar hipóteses envolvendo seus componentes. Um procedimento clássico de estimação é baseado no popular método dos mínimos quadrados (MQ). Neste caso, a estimativa $\hat{\underline{\beta}}$ é escolhida como o valor de $\underline{\beta}$ que minimiza a soma de quadrados residual

$$\sum_{i=1}^n (y_i - \underline{x}_i^T \underline{\beta})^2 = (\underline{y} - \underline{X} \underline{\beta})^T (\underline{y} - \underline{X} \underline{\beta}) \quad (1.2)$$

Este problema de minimização é equivalente à solução do seguin

te sistema de equações, obtido por diferenciação de (1.2) em relação a $\underline{\beta}$

$$\underline{X}^T (\underline{y} - \underline{X} \underline{\beta}) = \underline{0} \quad (1.3)$$

Assumindo que $\underline{X}$ tem posto completo p , a solução é única e pode ser escrita de forma explícita como

$$\underline{\hat{\beta}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y} \quad (1.4)$$

Os estimadores de mínimos quadrados são ótimos no sentido estabelecido pelo Teorema de Gauss-Markov. Mais especificamente se $\underline{y}$ for um vetor aleatório tal que

$$E(\underline{y}) = \underline{X} \underline{\beta}$$

$$\text{Var}(\underline{y}) = \sigma^2 \underline{I}$$

então, toda função estimável $\underline{c}^T \underline{\beta}$, onde $\underline{c}$ ($p \times 1$) é um vetor de constantes conhecidas, tem uma única estimativa linear que possui variância mínima na classe de todas estimativas lineares não-viciadas. Essa estimativa é dada por $\underline{c}^T \underline{\hat{\beta}}$, onde $\underline{\hat{\beta}}$ é qualquer estimativa de mínimos quadrados. Além disso, se a distribuição dos erros for normal, esses estimadores têm variância mínima entre todos estimadores não-viciados. Neste caso, os estimadores de mínimos quadrados coincidem com os estimadores de máxima verossimilhança.

Embora esse resultado seja importante, convém notar as seguintes considerações apresentadas em Hampel et alii (1986, Cap.6):

(i) a linearidade exigida pelo Teorema de Gauss-Markov é uma séria restrição para a forma dos estimadores; muitos estimadores de máxima verossimilhança não são lineares, como aqueles obtidos sob as distribuições Logística, Cauchy e t;

(ii) o estimador de mínimos quadrados é ótimo na classe de

todos estimadores não-viciados somente se os erros forem normalmente distribuídos. Portanto, a restrição de linearidade pode ser justificada apenas sob normalidade;

- (iii) o modelo normal, em geral, não é exatamente verdadeiro e, na presença de pequenos desvios, o procedimento de mínimos quadrados pode ser bastante ineficiente (Huber, 1973, 1981).

Valores discrepantes ("outliers") gerados por distribuições com caudas mais pesadas que a normal, por variâncias não-homogêneas entre os ϵ_i ou, simplesmente, por dados errados, alteram de maneira acentuada os estimadores de mínimos quadrados e suas propriedades. A Figura 1.1, correspondente a um exemplo apresentado em Hampel et alii (1986), mostra o efeito de uma observação discrepante no ajuste de mínimos quadrados.

Figura 1.1. Efeito de um ponto discrepante (P_1) no ajuste de Mínimos Quadrados (MQ).

Em função dessas considerações é fácil perceber que há interesse em procedimentos robustos, isto é, insensíveis a pequenos desvios das suposições sobre as quais se baseiam, mesmo que não sejam ótimos sob o modelo probabilístico adotado.

Um grande número de métodos robustos em Análise de Regressão foi proposto nos últimos anos como alternativa ao procedimento de mínimos quadrados. No entanto, a utilização desses métodos por estatísticos aplicados é ainda muito reduzida; a maioria continua a usar técnicas de mínimos quadrados, às vezes associadas a medidas de diagnóstico, principalmente pelas facilidades de cálculo inerentes às mesmas (ver por exemplo, Draper & Smith (1981) e Belsley et alii (1980)). O recente desenvolvimento na área de computação eletrônica eliminou as dificuldades computacionais associadas aos métodos alternativos e sua utilização precisa ser incentivada. Uma outra razão para a reduzida utilização de métodos robustos em Estatística Aplicada parece-nos bem descrita por Hogg (1979): "há uma quantidade enorme de esquemas robustos e os estatísticos aplicados simplesmente não sabem por onde começar".

Estudos comparativos de métodos robustos em Análise de Regressão divulgados na literatura tornam-se, pois, interessantes para orientação na prática estatística. Este é um dos objetivos do presente trabalho.

Nas seções seguintes deste capítulo são introduzidos alguns conceitos básicos da Teoria de Robustez. No Capítulo 1 apresentamos os principais estimadores robustos para Análise de Regressão e discutimos suas propriedades, concentrando-nos especialmente nos chamados estimadores M. No Capítulo 3 abordamos o problema da invariância de estimadores M em relação à escala e indicamos algumas estratégias para solucioná-lo com os respectivos algoritmos para cálculo dos estimadores. No Capítulo 4 comparamos através de simulação algumas das propostas discutidas no Capítulo 3; além disso aplicamos esses métodos em problemas práticos. No Capítulo 5 resumimos as conclusões deste estudo e propomos alguns tópicos para futuros trabalhos.

1.2. CONSIDERAÇÕES GERAIS SOBRE MÉTODOS ROBUSTOS

Nesta seção estudamos os métodos robustos no seu sentido amplo, discutimos suas relações com diversas áreas da Estatística e introduzimos alguns conceitos básicos. Ao leitor interessado em maiores detalhes sugerimos consultar Huber (1981) e Hampel et alii (1986).

1.2.1. Conceito e Objetivos dos Métodos Robustos

Um dos objetivos da Estatística é extrair informações úteis de dados empíricos. Uma maneira de obtenção dessas informações é através da abordagem que utiliza os chamados "métodos paramétricos". Essa abordagem se baseia em modelos probabilísticos específicos o que a torna excelente para resumir a informação contida num conjunto de dados; no entanto, as técnicas correspondentes são eficientes apenas sob o modelo adotado. Uma abordagem alternativa utiliza os chamados "métodos não-paramétricos" que se baseiam em modelos probabilísticos mais gerais e, conseqüentemente, são limitados quanto a capacidade de extração de informação dos dados; no entanto, são eficientes sobre uma classe mais ampla de modelos probabilísticos.

A grosso modo, os métodos robustos combinam as virtudes de ambas as abordagens. Modelos paramétricos são usados como veículos de informação em conjunto com procedimentos que não são criticamente dependentes das suposições inerentes a esses modelos.

A Figura 1.2, extraída de Hampel et alii (1986, Cap. 1), apresenta de forma esquemática os espaços de distribuições de probabilidade associados com modelos utilizados em Estatística Não-paramétrica, Paramétrica e Robusta. A Estatística Não-paramétrica permite uma ampla variedade de distribuições de probabilidade; a Estatística Paramétrica se baseia somente em subconjuntos de distribuições de probabilidade de baixa dimensão-

Figura 1.2. Representação esquemática de distribuições de probabilidade em alguns espaços amostrais (Hampel et alii, 1986).

o modelo paramétrico. A Estatística Robusta considera uma vizinhança completa do modelo paramétrico, o que a torna mais realista e fornece, além disso, muitas das vantagens do modelo paramétrico exato.

A palavra "robustez" em estatística é usada em diferentes sentidos por diversos autores. Nós a usaremos em um sentido relativamente restrito, como em Huber (1981) e Hampel et alii (1986), onde robustez significa insensibilidade a desvios do modelo probabilístico subjacente. Mais especificamente, Huber (1981) sugere que qualquer procedimento estatístico deve apresentar as seguintes características:

- (i) boa eficiência sob o modelo suposto;
- (ii) robustez no sentido de que pequenos desvios das hipóteses do modelo afetem seu desempenho apenas ligeiramente, isto é, seu comportamento deve estar próximo daquele calculado sob o modelo (em termos de variância assintótica, por exemplo);
- (iii) desvios maiores do modelo não devem causar catástrofes.

1.2.2. Tipos de Desvios do Modelo Paramétrico

Hampel et alii (1986) distinguem quatro tipos principais de desvios dos modelos paramétricos exatos:

- (i) **Ocorrência de erros grosseiros.** Erros grosseiros são valores ocasionais resultantes, por exemplo, da transcrição ou digitação dos dados. Eventualmente valores discrepantes (valores afastados do conjunto formado pela maioria dos dados) podem ser gerados por fenômenos transientes e, neste caso, parte dos dados não se ajusta ao modelo original; eles podem, também, ser gerados por uma distribuição com cauda pesada (por exemplo, Cauchy) ou, ainda, pela escolha errada do modelo. Sem dúvida, é o tipo de desvio mais óbvio e mais discutido. Se não for tratado convenientemente pode inutilizar completamente uma análise estatística como é o caso, por exemplo, do ajuste do mínimos quadrados apresentado na Figura 1.1. Considerando que na prática a freqüência de erros grosseiros é da ordem de 1 - 10% (Hampel et alii, 1986), os valores discrepantes podem facilmente escapar à atenção se não forem tomadas precauções. Técnicas robustas são, nesses casos, definitivamente necessárias.
- (ii) **Arredondamento e agrupamento de dados.** Devemos lembrar que na prática todos os dados têm precisão limitada e são basicamente discretos; além disso, muitas vezes são arredondados, agrupados ou classificados ainda mais grosseiramente. Esses efeitos locais, segundo Hampel et alii (1986, Cap. 1), parecem se constituir no tipo mais inofensivo de desvio e podem ser muitas vezes desprezados; porém, não devem ser esquecidos.
- (iii) **O modelo pode ter sido formulado como uma aproximação.** É o que acontece, por exemplo, quando utilizamos o Teorema Limite Central. Sabe-se que grandes conjuntos de dados de boa qualidade (sem erros grosseiros) ten-

dem a mostrar pequenos mas visíveis desvios do modelo normal. Este fato já era conhecido desde o início do século 19, mas foi conscientemente ignorado pouco após a invenção do método dos mínimos quadrados. Mais tarde importantes estatísticos como Student (1927) coletaram amostras de alta qualidade e observaram que elas apresentavam, usualmente, distribuições com caudas mais pesadas que a normal.

(iv) **Hipótese de independência.** Alguns modelos paramétricos exigem independência ou outra estrutura de correlação que pode ser apenas aproximadamente satisfeita; pouco ainda se sabe a respeito de robustez contra este tipo de desvio. Esse assunto não será abordado neste trabalho; o leitor interessado no tema pode consultar Hampel et alii (1986, Cap. 8) para breve discussão e referências bibliográficas.

1.3. PRINCIPAIS ABORDAGENS TEÓRICAS PARA ROBUSTEZ

Métodos robustos existem desde o início do desenvolvimento da Estatística Aplicada. O costume de olhar os dados e identificar as observações discrepantes é um passo na direção de robustez; a eliminação de valores altamente destoantes da maioria dos dados é um procedimento robusto informal.

Os fundamentos da teoria de robustez moderna foram apresentados por Huber (1964). Ele introduziu os estimadores M, fornecendo uma estrutura flexível para o estudo de estimação robusta do parâmetro de locação. Esses estimadores são inspirados nos estimadores de máxima verossimilhança, porém, são mais gerais. No caso do modelo de locação ($y_i = \theta + \epsilon_i$), Huber propõe a minimização de $\sum \rho(y_i - \theta)$ em lugar da minimização de $-\sum \ln f(y_i - \theta)$, onde f é a função densidade dos y_i e ρ é uma função não-constante. Huber determinou estimadores M que são ótimos em um sentido minimax. Para este fim ele considerou o

"modelo de erros grosseiros", que essencialmente consiste de uma espécie de "vizinhança" do modelo simétrico com distribuição F . A idéia é a seguinte: em vez de fixar um modelo paramétrico exato (F), assume-se que uma fração conhecida dos dados ϵ ($0 \leq \epsilon < 1$) pode consistir de erros grosseiros com uma distribuição simétrica desconhecida (H). Então o modelo de erros grosseiros é definido por

$$P_\epsilon = \{(1 - \epsilon) F + \epsilon H; \quad H \text{ é uma distribuição simétrica}\} \quad (1.5)$$

Huber (1964) propôs um estimador que otimiza o pior que pode acontecer sobre a vizinhança do modelo, medido pela variância assintótica do estimador. Em outras palavras, ele definiu o estimador M , associado à função ρ_0 ou Ψ_0 (onde $\Psi_0 = \partial \rho_0 / \partial \theta$, denominada função escore) como o estimador que minimiza a variância assintótica máxima sobre P_ϵ

$$\sup_{G \in P_\epsilon} V(\Psi_0, G) = \min_{\Psi} \sup_{G \in P_\epsilon} V(\Psi, G), \quad (1.6)$$

onde $V(\Psi, G)$ é a variância assintótica do estimador M definido pela função escore Ψ , sob uma distribuição G pertencente à classe de distribuições P_ϵ , que é dada por $V(\Psi, G) = \int \Psi^2 dG / (\int \Psi' dG)^2$.

Huber mostrou que sob condições bastante gerais a solução de (1.6), no caso do modelo de erros grosseiros definido em (1.5), é obtida encontrando a "distribuição menos favorável" F_0 , isto é, a distribuição que minimiza a informação de Fisher sobre toda a distribuição $G \in P_\epsilon$. O estimador M obtido, correspondente a $\Psi_0 = F_0''/F_0'$, é o estimador de máxima verossimilhança para essa distribuição menos favorável.

No caso onde $F = \Phi$ (distribuição normal padrão) a informação de Fisher é minimizada por

$$\begin{aligned} f_0(x) &= \frac{1 - \epsilon}{\sqrt{2\pi}} e^{-x^2/2}, \quad \text{para } |x| \leq k \\ &= \frac{1 - \epsilon}{\sqrt{2\pi}} e^{k^2/2 - k|x|}, \quad \text{para } |x| > k, \end{aligned} \quad (1.7)$$

com k e ϵ relacionados através de

$$\frac{2 \phi(k)}{k} - 2 \Phi(-k) = \frac{\epsilon}{1 - \epsilon} \quad (1.8)$$

onde $\phi = \Phi'$ é a função densidade da distribuição normal padrão. Nesse caso, a função escore ψ_0 é dada por

$$\psi_0(x) = \max[-k; \min(k, x)], \quad (1.9)$$

conhecida como função escore de Huber. A distribuição menos favorável f_0 (normal no centro e exponencial dupla nas caudas) e a função escore ψ_0 estão esboçadas na Figura 1.3.

Figura 1.3. Distribuição menos favorável no modelo de erros grosseiros e seu estimador de máxima verossimilhança.

Outra abordagem não menos importante e desenvolvida quase simultaneamente à proposta por Huber foi considerada por Hampel (1968) e seus principais resultados são amplamente discutidos

em Hampel et alii (1986). Ela se baseia em três conceitos centrais: robustez qualitativa, função de influência e ponto de ruptura, que serão definidos na Seção 1.4. Essas ferramentas descrevem diferentes aspectos do comportamento dos procedimentos estatísticos e podem ser utilizados tanto como orientação na escolha entre vários procedimentos estatísticos disponíveis quanto na construção de novos métodos com características especificadas.

A abordagem baseada em funções de influência é relacionada com a abordagem minimax de Huber, porém, é matematicamente mais simples. Em geral, ela pode ser aplicada sempre que estimadores de máxima verossimilhança ou testes de razão de verossimilhança façam sentido.

1.4. MEDIDAS DE ROBUSTEZ

O conhecimento de algumas medidas de robustez é necessário para a descrição das propriedades de um conjunto de estimadores. Há interesse em medir, por exemplo, a proporção de dados contaminados que um estimador pode suportar sem que se torne inútil (ponto de ruptura) ou o efeito de um dado discrepante no estimador (função de influência).

As principais medidas de robustez são apresentadas a seguir. Inicialmente, porém, introduzimos o conceito de funcionais, útil na caracterização precisa de robustez.

1.4.1. Funcionais

Muitos dos estimadores mais comuns podem ser escritos como funcionais $T_n = T(F_n)$ da função de distribuição empírica F_n ; esses funcionais são chamados funções estatísticas. Por exemplo, para a variância σ^2 o funcional apropriado é:

$$T(F) = \int \{x - \int x dF(x)\}^2 dF(x)$$

$T_n = T(F_n)$ é a variância amostral. Em geral, temos

$$T_n \rightarrow T(F) \text{ com } n \rightarrow \infty \text{ e} \quad (1.10)$$

$$L_F(\sqrt{n} [T_n - T(F)]) \rightarrow N(0, V(T, F))$$

com $n \rightarrow \infty$,

onde $L_F(E)$ significa "a distribuição da estatística E sob F " e $V(T, F)$ é chamada variância assintótica de $\{T_n; n \geq 1\}$ em F .

1.4.2. Robustez Qualitativa

Robustez qualitativa está relacionada com uma forma de continuidade do funcional T . Dizemos que a seqüência $\{T_n; n \geq 1\}$ é qualitativamente robusta para $F = F_0$ se, dado $\epsilon > 0$, existem $\delta > 0$ e $n_0 > 0$ tais que, para toda F e $n \geq n_0$

$$d(F_0, F) < \delta \Rightarrow d\{L_{F_0}(T_n), L_F(T_n)\} < \epsilon \quad (1.11)$$

onde d é uma função distância conveniente.

1.4.3. Ponto de Ruptura

Robustez quantitativa pode ser medida pela variação na distribuição assintótica $L_F(T_n)$ induzida por uma variação em F . Nesse contexto, a medida de robustez mais comum é o ponto de ruptura assintótico ϵ^* definido por

$$\epsilon^* = \sup \{ t: b(t) < b(1) \} \quad (1.12)$$

onde:

$$b(t) = \lim_{n \rightarrow \infty} \sup_{F \in P_t} |\text{mediana}[L_F(T_n - T(F_0))]|$$

é o vício assintótico e P_t é alguma vizinhança de F_0 , por exem

plo, a vizinhança de Lévy

$$P_t(F_0) = \{F: \forall x, F_0(x-t) - t \leq F(x) \leq F_0(x+t) + t\}$$

ou uma vizinhança "contaminada"

$P_t(F_0) = \{F: F = (1-t)F_0 + tH\}$ e H é uma medida de probabilidade.

O ponto de ruptura corresponde à fração limite de valores discrepantes que o estimador definido por T pode suportar e, portanto, é uma medida global de robustez.

1.4.4. Função de Influência

Outra importante ferramenta no estudo de robustez é a função de influência, definida por:

$$I(x; F, T) = \lim_{t \rightarrow 0} \frac{\{T[(1-t)F + t\delta_x] - T(F)\}}{t} \quad (1.13)$$

onde δ_x indica um ponto com probabilidade 1 em x . Essencialmente, ela mede a influência de uma única observação no valor de uma função estatística $T(F_n)$. Segundo Hampel et alii (1986, Cap. 2), a importância da função de influência está em sua interpretação heurística: ela descreve o efeito na estimativa produzido por uma contaminação infinitesimal no ponto x , padronizado pela massa da contaminação. Do ponto de vista de robustez devemos dar preferência aos estimadores que possuam $I(x; F, T)$ próxima de zero para todo valor de x .

Algumas medidas de robustez de $T(F_n)$ podem ser construídas a partir da função de influência. Entre elas temos:

1. **sensibilidade a erros grosseiros.** É definida por:

$$\gamma^* = \sup_x |I(x; F, T)| \quad (1.14)$$

e, portanto, mede aproximadamente o maior efeito que uma observação contaminada produz no estimador. Quando γ^* é finita dizemos que T é B-robusto sob a distribuição F , onde B se refere ao vício ("bias") assintótico. Um dos primeiros passos na robustificação de um estimador é limitar γ^* e isto, muitas vezes, entra em conflito com o objetivo de eficiência assintótica elevada;

2. **sensibilidade a deslocamentos locais.** É definido por:

$$\lambda^* = \sup_{x \neq y} |I(y; F, T) - I(x; F, T)| / |y - x| \quad (1.15)$$

mede os efeitos de arredondamento ou agrupamento de observações;

3. **ponto de rejeição.** É definido por:

$$\rho^* = \inf \{ r > 0 : I(x; F, T) = 0 \text{ para } |x| > r \} \quad (1.16)$$

quando a distribuição subjacente é simétrica em torno de zero. O ponto de rejeição é, pois, a distância do centro de simetria da distribuição acima da qual todas as observações são rejeitadas. Portanto, é desejável que os estimadores tenham ρ^* finito.

Um esboço dessas medidas para uma função de influência particular é apresentado na Figura 1.4.

A função de influência pode, também, ser usada para o cálculo da variância assintótica de estimadores (ver por exemplo, Hampel et alii (1986, Cap. 2))

$$V(T, F) = \int I(x; F, T)^2 dF(x) \quad (1.17)$$

Figura 1.4. Uma função de influência e respectivas medidas de robustez.

As medidas (1.14) a (1.17) demonstram a importância da função de influência na teoria de robustez. A abordagem devida a Hampel tem nesta medida seu ponto de apoio; ela contrabalança eficiência no modelo ideal, medida pela variância assintótica e robustez a pequenos desvios do modelo, medida pela sensibilidade a erros grosseiros, ambas determinadas pela função de influência.

2. CLASSES DE ESTIMADORES EM ANÁLISE DE REGRESSÃO

2.1. INTRODUÇÃO

Neste capítulo apresentamos algumas propostas de estimadores utilizados em Análise de Regressão como alternativa para o procedimento de mínimos quadrados. Introduzimos três classes de estimadores correspondendo a combinações lineares de estatísticas de ordem (estimadores L), estimativas construídas a partir de testes de postos (estimadores R) e estimadores tipo máxima verossimilhança (estimadores M). Citaremos, brevemente, outros estimadores encontrados na literatura recente de métodos robustos sendo que alguns deles são modificações dos estimadores M.

Para facilitar a interpretação em alguns casos, introduziremos o estimador através de modelos de locação unidimensionais e, então, ampliaremos o conceito para o modelo de regressão.

As referências básicas são Huber (1981) e Hampel et alii (1986) que podem ser consultadas para obtenção de fundamentos rigorosos sobre o tema e informações detalhadas a respeito da literatura.

2.2. ESTIMADORES L

Vamos considerar uma amostra aleatória de uma distribuição contínua com parâmetro de locação θ . As estatísticas de ordem da amostra são denotadas por $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Um estimador L para θ é uma combinação linear dessas estatísticas de ordem

$$T_n(X_1, \dots, X_n) = \sum_{i=1}^n a_i X_{(i)} \quad (2.1)$$

onde

a_i , $i = 1, 2, \dots, n$ são constantes (pesos).

Exemplos de estimadores L são:

i. mediana amostral;

ii. média α -aparada (" α -trimmed mean"):

$$\bar{x}_\alpha = \frac{1}{n - 2[\alpha]} \sum_{i=[\alpha]+1}^{n-[\alpha]} x_{(i)}$$

onde a notação $[x]$ indica a parte inteira de x ;

iii. estimador de Gastwirth: média ponderada dos percentis $33 \frac{1}{3}^\circ$, 50° e $66 \frac{2}{3}^\circ$ com pesos 0,3; 0,4 e 0,3, respectivamente; e

iv. tri-média ("trimean") de Tukey: média ponderada do 1° , 2° e 3° quartis com pesos $1/4$, $1/2$ e $1/4$, respectivamente.

Estes e outros estimadores L para o parâmetro de locação são descritos detalhadamente em Andrews et alii (1972).

A generalização de estimadores L para a situação de regressão não é óbvia. Uma abordagem interessante tem como ponto de partida uma definição elementar de quantis amostrais. De uma maneira geral podemos definir o q -ésimo quantil amostral, $0 < q < 1$, como qualquer solução para o problema de minimização

$$\min_{\theta \in R} \left[\sum_{i \in \{i: y_i \geq \theta\}} q |y_i - \theta| + \sum_{i \in \{i: y_i < \theta\}} (1 - q) |y_i - \theta| \right] \quad (2.2)$$

onde $\{y_i: i = 1, 2, \dots, n\}$ é uma amostra aleatória da variável aleatória Y , com função da distribuição F . O caso da me-

diana ($q = \frac{1}{2}$) é bem conhecido, mas o mesmo não se dá com o caso geral (ver, por ex., Ferguson (1967)).

Koenker & Basset (1978) propuseram uma generalização direta do problema de minimização apresentado em (2.2) para obtenção do estimador L no modelo de regressão (1.1). O q -ésimo quantil de regressão (ou percentil de regressão, $\hat{\beta}(q)$) é definido como qualquer solução para o problema de minimização

$$\min_{\beta \in R^p} \left[\sum_{i \in \{i: y_i \geq x_i^T \beta\}} q |y_i - x_i^T \beta| + \sum_{i \in \{i: y_i < x_i^T \beta\}} (1-q) |y_i - x_i^T \beta| \right] \quad (2.3)$$

No modelo de locação ($p = 1$, $x_i = 1$, $\forall i$), (2.2) e (2.3) são equivalentes.

Um caso especial importante é a mediana de regressão ("regression median"), isto é, o quantil de regressão para $q = \frac{1}{2}$, que corresponde ao estimador que minimiza a soma dos resíduos absolutos. Este estimador foi objeto de estudo em muitos contextos e sob uma variedade de nomes tais como: MAD - "minimum absolute deviation", LAE - "least absolute errors", LAV - "least absolute values", L_1 - norm (ou, simplesmente, L_1 como será de notado ao longo deste texto) e "median regression". Com o desenvolvimento de vários algoritmos eficientes e alguns procedimentos de inferência, esse estimador é, segundo Narula & Wellington (1982), uma alternativa viável e desejável para o estimador de mínimos quadrados, particularmente, como valor inicial para solução iterativa de outros procedimentos robustos.

Dielman (1984) apresentou uma revisão bastante completa dos trabalhos publicados sobre o estimador L_1 . Koenker & Basset (1978) propuseram, também, uma generalização dos estimadores L de locação mais comuns, tais como tri-média, estimador de Gastwirth e média α -aparada, para o modelo linear baseando-se nos quantis de regressão.

2.3. ESTIMADORES R

Estimadores R foram introduzidos por Hodges & Lehmann (1963) e seu nome se deve ao fato de sua construção ser feita a partir de testes baseados em postos ("rank"). Estes estimadores foram originalmente construídos a partir de testes para uma amostra, porém, vamos apresentar aqui a abordagem mais usual que usa testes para duas amostras (Huber, 1981).

Sejam $X_1, \dots, X_n$ e $Y_1, \dots, Y_m$ duas amostras independentes das distribuições $F(x)$ e $G(x) = F(x + \Delta)$, respectivamente, onde Δ é um deslocamento desconhecido na locação. Denotemos por R_i o posto de X_i na amostra de tamanho $N = n + m$ obtida pela combinação das duas amostras. O teste de postos para $H: \Delta = 0$ contra $A: \Delta > 0$ é baseado na estatística

$$S_N = \frac{1}{n} \sum_{i=1}^n a_N(R_i), \quad (2.4)$$

onde $a_N(i)$, $i = 1, 2, \dots, n + m$ são escores gerados, por exemplo, por

$$a_N(i) = N \int_{(i-1)/N}^{i/N} J(s) ds, \quad (2.5)$$

onde J é uma função conveniente.

Um estimador R de locação para uma amostra é definido por $T_n = T(X_1, \dots, X_n)$, onde T_n é escolhido de tal forma que (2.4) torne-se tão próximo de zero quanto possível quando calculado para as amostras $(X_1, \dots, X_n)$ e $(2T_n - X_1, \dots, 2T_n - X_n)$, isto é, $S_N \approx 0$. Notemos que nem sempre é possível encontrar S_N exatamente igual a zero devido a descontinuidades da função S_N .

Diversos estimadores R podem ser obtidos através de diferentes funções J . Essas funções podem ser definidas através da função de distribuição hipotética dos dados; segundo Huber (1981, cap. 3) é complicado construir estimadores R, assim como estimadores L, com propriedades de robustez especificadas a priori.

Citaremos, a seguir alguns exemplos de estimadores R definidos a partir de diferentes funções J :

- i. estimador baseado em "escores normais" ($J(t) = \Phi^{-1}(t)$, onde $\Phi(t)$ é a distribuição normal padrão). Esse estimador é qualitativamente robusto, mas não é B-robusto e seu ponto de ruptura é $\epsilon^* \approx 0,239$ (Hampel et alii, 1986, cap. 2);
- ii. estimador de Hodges-Lehmann, derivado do teste de Wilcoxon ($J(t) = t - \frac{1}{2}$). Esse estimador é dado pela mediana de $\{(X_i + X_j)/2; i = 1, \dots, n \text{ e } j = 1, \dots, n\}$ e seu ponto de ruptura é $\epsilon^* \approx 0,293$ (Hampel et alii, 1986, cap. 2).

As estimativas R para o modelo de regressão (1.1) são obtidas minimizando

$$\sum_{i=1}^n a_n(R_i) r_i, \quad (2.6)$$

onde os r_i , $i = 1, \dots, n$ são os resíduos amostrais

$(y_i - \tilde{x}_{i\sim}^T \beta)$, R_i é o posto de r_i em $(r_1, \dots, r_n)$ e $a_n(i)$

é uma função escore monótona satisfazendo $\sum_{i=1}^n a_n(i) = 0$.

Jaeckel (1972) provou que, sob condições bastante gerais, este problema de minimização é equivalente a solucionar as p equações

$$\sum_{i=1}^n x_{ik} a_n(R_i) \approx 0, \quad k = 1, \dots, p \quad (2.7)$$

que devem ser resolvidas aproximadamente devido a descontinuidades em $a_n(i)$ e R_i .

Estas equações aproximadas podem ser transformadas, por exemplo, no problema de minimização de

$$\sum_{k=1}^p \left| \sum_{i=1}^n a_n(R_i) x_{ik} \right| \quad (2.8)$$

A equação (2.8) foi proposta por Jurečková (1971) e Jaeckel (1972) provou a equivalência assintótica entre (2.7) e (2.8). A solução das equações (2.7) ou (2.8) é obtida por técnicas de programação linear, porém, segundo Huber (1981, cap. 7) pode tornar-se muito trabalhosa, a menos que n e p sejam bastante pequenos.

2.4. ESTIMADORES M

Consideremos o problema de estimação do parâmetro θ de uma distribuição descrita pela densidade $f_{\theta}(x)$, através de uma amostra aleatória $X_1, X_2, \dots, X_n$. O estimador de máxima verossimilhança para θ é definido como o valor $T_n = T(X_1, X_2, \dots, X_n)$ que maximiza $\prod_{i=1}^n f_{\theta}(X_i)$, ou que minimiza

$$\sum_{i=1}^n [-\ln f_{\theta}(X_i)]$$

Huber (1964) propôs uma generalização desse método considerando o problema de minimização de

$$\sum_{i=1}^n \rho(X_i, \theta) \quad (2.9)$$

onde ρ é uma função arbitrária não-constante. Supondo que ρ tem derivada $\psi(x, t) = \frac{\partial}{\partial t} \rho(x, t)$, a estimativa T_n satisfaz a equação implícita

$$\sum_{i=1}^n \psi(X_i, T_n) = 0 \quad (2.10)$$

Qualquer estimador definido por (2.9) ou (2.10) é chamado estimador M (tipo máxima verossimilhança). No caso de modelos de locação é natural usarmos funções escore da forma $\psi(x - \theta)$, para as quais assumimos que $\int \psi dF = 0$; o estimador M é definido então por

$$\sum_{i=1}^n \psi(x_i - T_n) = 0 \quad (2.11)$$

Sua função de influência é dada por

$$I(x; \psi, F) = \frac{\psi(x - T(F))}{-\int \frac{\partial}{\partial \theta} [\psi(y - \theta)]_{T(F)} dF(y)} \quad (2.12)$$

Notemos que a função de influência dos estimadores M é proporcional à função escore ψ ; esse fato nos habilita a construir estimadores M robustos através da escolha conveniente da função ψ . Por exemplo, funções ψ limitadas geram estimadores M que são insensíveis à presença de dados discrepantes; funções ψ que são proporcionais a $f'_0(x)/f_0(x)$ geram estimadores M que são altamente eficientes sob o modelo com densidade f_0 ; funções ψ contínuas geram estimadores M que são insensíveis a erros de agrupamentos e arredondamento.

Huber (1964) mostrou que os estimadores M possuem, assintoticamente, distribuição normal com variância dada por

$$V(T, F) = \frac{\int \psi^2(x, T(F)) dF(x)}{\left\{ \int \frac{\partial}{\partial \theta} [\psi(y, \theta)]_{T(F)} dF(y) \right\}^2} \quad (2.13)$$

No caso de a distribuição F ser simétrica, é natural escolher uma função ψ satisfazendo $\psi(-x) = -\psi(x)$. Nessa situação, todos os estimadores M com função ψ estritamente monótona têm as seguintes propriedades (Hampel et alii (1986), Cap. 2):

- i. se ψ é limitada, então os estimadores são B-robustos, qualitativamente robustos e têm ponto de ruptura $\epsilon^* = \frac{1}{2}$;
- ii. se ψ é não-limitada então os estimadores não são nem B-robustos nem qualitativamente robustos e $\epsilon^* = 0$.

No caso de ψ não ser estritamente monótona, a solução de (2.11)

pode não ser única. Por exemplo, para os estimadores M com ponto de rejeição finito

$$\sum_{i=1}^n \psi(X_i - T_n) = 0$$

para todo valor absoluto grande de T_n . Hampel et alii (1986, Cap. 2) sugeriram, neste caso, três maneiras de definir a sequência $\{T_n; n \geq 1\}$ de soluções que é única e consistente:

- i. tomar o mínimo global de (2.9);
- ii. tomar a solução de (2.11) mais próxima da mediana amostral; ou
- iii. usar o método de Newton (apresentado no Cap. 3) com a mediana como valor inicial.

Os métodos (ii) e (iii) são fáceis de executar e mantêm o ponto de ruptura $\epsilon^* = \frac{1}{2}$ da mediana.

Entre muitas funções ψ propostas na literatura para obtenção de estimadores M robustos, as mais usadas e estatisticamente convenientes são aquelas sugeridas por Huber (1964) através da expressão (1.9), por Hampel (1974)

$$\psi_{a, b, c} = \begin{cases} x & \text{se } 0 \leq |x| \leq a \\ a \operatorname{sign}(x) & \text{se } a \leq |x| \leq b \\ a \frac{c - |x|}{c - b} \operatorname{sign}(x) & \text{se } b \leq |x| \leq c \\ 0 & \text{se } |x| \geq c \end{cases} \quad (2.14)$$

onde $a \leq b < c < \infty$ são constantes positivas ou por Tukey que sugeriu a função ψ bi-ponderada (Beaton & Tukey, 1974)

$$\psi_B = \begin{cases} x(1 - (x/B)^2)^2 & \text{se } |x| \leq B \\ 0 & \text{se } |x| > B \end{cases} \quad (2.15)$$

onde B é uma constante positiva. Estimadores obtidos pelo uso de (1.9) têm a propriedade de limitar, sem eliminação total, a influência de dados discrepantes (função ψ monótona) enquanto que aqueles obtidos usando (2.14) ou (2.15) têm a propriedade de rejeitá-los completamente (funções ψ redescendentes).

As funções escore (2.14) e (2.15) são esboçados na Figura 2.1.

Figura 2.1. Função escore ψ : (a) Hampel e (b) bi-ponderada de Tukey.

As constantes de ajuste k , a , b , c e B dos estimadores M propostos por Huber, Hampel e Tukey determinam as propriedades dos respectivos estimadores, tais como eficiência, função de influência e sensibilidade a erros grosseiros. Elas podem ser escolhidas para ajustar, por exemplo, a eficiência do estimador a uma distribuição particular, tal como a distribuição normal.

Uma grande parte da literatura sobre estimadores M é dedicada ao estudo de problemas de locação. Os estimadores M obtidos de (2.11) não são invariantes em relação à escala; conseqüentemente, na prática temos que levar a escala em consideração pelo menos como um parâmetro de interesse secundário ("nuisance parameter"). A estimação simultânea de locação e escala pode trazer algumas dificuldades teóricas quanto à convergência dos algoritmos de cálculo, mas uma alternativa mais simples que produz estimadores M de locação invariantes em relação à escala é disponível. Ela é conhecida na literatura por "estimadores M

de locação com estimativa preliminar da escala" e consiste em tomar uma estimativa robusta da escala (S_n) e definir o estimador M de locação T_n como uma solução para

$$\sum_{i=1}^n \psi\left(\frac{x_i - T_n}{S_n}\right) = 0 \quad (2.16)$$

Neste caso S_n é considerada fixa e é conveniente tomar

$$S_n = \frac{1}{0,6745} \text{MAD}(x_i) = 1,483 \text{ med}_i \{|x_i - \text{med}_j(x_j)|\} \quad (2.17)$$

que tem ponto de ruptura $\epsilon^* = 50\%$, (med denota mediana e MAD denota "median absolute deviation"). O fator 1,483 ($1/\phi^{-1}(3/4)$) torna S_n uma estimativa consistente da escala quando a distribuição subjacente é normal.

Estimativas M em um passo (também chamadas estimativas m) são definidas por

$$T_n = T_n^{(0)} + S_n^{(0)} \frac{\sum_{i=1}^n \psi\left(\frac{x_i - T_n^{(0)}}{S_n^{(0)}}\right)}{\sum_{i=1}^n \psi'\left(\frac{x_i - T_n^{(0)}}{S_n^{(0)}}\right)} \quad (2.18)$$

onde $T_n^{(0)}$ e $S_n^{(0)}$ são estimativas iniciais de locação e escala (Bickel, 1975). As estimativas T_n resultantes são o primeiro passo do algoritmo iterativo de Newton para estimadores M, que produzem estimadores que têm a mesma função de influência e, essencialmente, o mesmo comportamento assintótico que suas versões completamente iteradas, pelo menos quando Ψ é ímpar e a distribuição F subjacente é simétrica. Segundo Hampel et alii (1986, Cap. 2) no caso em que ψ não é monótona, a versão (2.18) é mesmo mais segura porque evita o problema de multiplicidade de soluções. Entretanto, estimadores M em um passo devem ser usados somente quando os valores iniciais são muito robustos; caso contrário, a estimativa resultante pode não ser robusta. A recomendação usual é

$$T_n^{(0)} = \text{med}_i (x_i)$$

$$S_n^{(0)} = 1,483 \text{ MAD } (x_i)$$

Andrews et alii (1972) mostraram que o uso da média aritmética como valor inicial leva a resultados desastrosos e, também, que a amplitude interquartílica, que possui ponto de ruptura de a penas 25%, não é tão apropriada quanto MAD para a estimativa i nicial da escala.

Huber (1973) estendeu os resultados anteriores para o caso da estimação de β no modelo de regressão linear (1.1). Vamos con siderar, inicialmente, o parâmetro de escala fixo e conhecido; sem perda de generalidade tomemos $\sigma = 1$. O estimador M de β é a solução $\hat{\beta}$ para o problema de minimização de

$$\sum_{i=1}^n \rho(y_i - \tilde{x}_i^T \beta) \quad (2.19)$$

Se ρ tem derivada $(\partial/\partial r) \rho(r) = \psi(r)$, $\hat{\beta}$ satisfaz o sistema de equações

$$\sum_{i=1}^n \psi(y_i - \tilde{x}_i^T \beta) \tilde{x}_i = 0 \quad (2.20)$$

Notemos que os estimadores de mínimos quadrados e máxima veros similhança são casos particulares de estimadores M com $\rho(x) = x^2$ e $\rho(x) = -\log f(x, \beta)$, respectivamente, onde f é a função den sidade dos erros.

Huber (1973) mostrou que o estimador M é único, consistente e as sintoticamente normal sob certas condições de regularidade.

Em particular, esses resultados são válidos quando a função es core ψ é limitada e não-decrescente e $hp^2 \rightarrow 0$ com $n \rightarrow \infty$, on- de p é a dimensão de β e $h = \max_{1 \leq i \leq n} (\tilde{x}_i^T (\tilde{X}^T \tilde{X})^{-1} \tilde{x}_i)$, isto é, o

maior valor da diagonal da matriz de projeção $\tilde{X} (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T$.

Essa condição exige, pois, um certo controle das váriáveis ex-

plicativas para que h seja pequeno, garantindo a ausência de pontos alavanca ("leverage points"). Huber também mostrou que alguns resultados mais fracos valem no caso de $h_p \rightarrow 0$ com $n \rightarrow \infty$. Como $h_p \rightarrow 0$ implica $p^2/n \rightarrow 0$, ainda neste caso é necessário um número muito grande de observações mesmo quando a dimensão de β é moderada. Yohai & Maronna (1979) melhoraram os resultados de Huber mostrando que $h_p^{3/2} \rightarrow 0$ com $n \rightarrow \infty$ é suficiente para consistência e normalidade assintótica dos estimadores. Se a hipótese de monotonicidade de ψ for relaxada, esses resultados ainda valem no sentido local.

Também neste caso os estimadores M não são invariantes em relação a escala. Para contornar esse problema uma solução é introduzir em (2.20) uma estimativa robusta ($\hat{\sigma}$) do parâmetro de escala σ que seja invariante em relação à locação, passando a considerar o sistema

$$\sum_{i=1}^n \psi \left[(y_i - \tilde{x}_i^T \hat{\beta}) / \hat{\sigma} \right] \tilde{x}_i = 0 \quad (2.21)$$

onde $\hat{\sigma}$ é considerado fixo. Nos Capítulos 3 e 4 retomaremos este problema e apresentaremos algumas soluções alternativas.

Para estudar algumas propriedades estatísticas dos estimadores M apresentamos sua função de influência, ignorando o parâmetro de escala, como em Hampel et alii (1986)

$$I(y, \tilde{x}; F, \hat{\beta}) = \psi(y - \tilde{x}^T \hat{\beta}) \tilde{M}^{-1} \tilde{x} \quad (2.22)$$

onde

$$\tilde{M} = \int \psi' (y - \tilde{x}^T \hat{\beta}) \tilde{x} \tilde{x}^T dF(y)$$

Este resultado mostra que a influência do resíduo é limitada, o que caracteriza uma melhoria em relação aos estimadores de mínimos quadrados do ponto de vista de robustez. No entanto, a influência da posição dos valores observados das variáveis explicativas não é limitada. Assim, uma única observação dis

crepante no conjunto das variáveis explicativas pode determinar o ajuste quase que completamente porque a função de influência (2.22) é diretamente proporcional a $\underline{x}$. Entretanto, se as variáveis explicativas estiverem sob controle, estimadores M robustos de $\underline{\beta}$ podem ser obtidos através de uma escolha apropriada da função escore ψ .

Vários autores consideram estimadores alternativos, denominados estimadores de influência limitada, obtidos pela ponderação simultânea das variáveis respostas e explicativas. Consideremos $\underline{x}_1, \dots, \underline{x}_n$ vetores aleatórios; denotaremos por $f_{\underline{\beta}}(\underline{x}, y)$ a densidade conjunta de $(\underline{x}_i, y_i)$ e por $F_{\underline{\beta}}(\underline{x}, y)$ a função distribuição correspondente. Notemos que usualmente consideramos o modelo (1.1) com $\underline{x}_1, \dots, \underline{x}_n$ fixos. Entretanto, segundo Hampel et alii (1986, Cap. 6), é conveniente utilizar um modelo com $\underline{x}_1, \dots, \underline{x}_n$ aleatórios para lidar com valores discrepantes nas variáveis explicativas. Os resultados obtidos para esse caso podem ser interpretados e aplicados num modelo onde $\underline{x}_1, \dots, \underline{x}_n$ são fixos.

Um estimador M generalizado ($\hat{\underline{\beta}}_n$) para $\underline{\beta}$ em modelos da forma (1.1) é definido implicitamente pela equação

$$\sum_{i=1}^n \eta(\underline{x}_i, (y_i - \underline{x}_i^T \hat{\underline{\beta}}_n)) \underline{x}_i = \underline{0}, \quad (2.23)$$

onde a função $\eta: R^D \times R \rightarrow R$ deve satisfazer algumas condições (ver Hampel et alii (1986), Cap. 6).

Existem várias propostas para a função η , porém, todas elas podem ser escritas na forma

$$\eta(\underline{x}, r) = \omega(\underline{x}) \cdot \psi(r \cdot v(\underline{x})) \quad (2.24)$$

onde $\psi: R \rightarrow R$, $\omega: R^D \rightarrow R^+$ e $v: R^D \rightarrow R^+$ são funções convenientes e $r = y - \underline{x}^T \hat{\underline{\beta}}_n$. Huber (1973) considerou

$v(\underline{x}) = w(\underline{x}) = 1$ que corresponde à definição usual de estimadores M (2.20). Segundo Hampel et alii (1986, Cap. 6), Mallows considerou $v(\underline{x}) = 1$ e Schweppe $v(\underline{x}) = 1/w(\underline{x})$, conhecidos como estimadores de influência limitada tipo Mallows e tipo Schweppe, respectivamente. A Figura 2.2, obtida de Hampel et alii (1986 pág. 322), mostra esquematicamente a maneira pela qual as observações têm sua influência limitada através de estimadores M tipo Huber (definidos por (2.20)), tipo Mallows e estimador Hampel-Krasker, caso particular do estimador tipo Schweppe.

Huber

$$\sum_{i=1}^n \psi_C(r_i) \underline{x}_i = 0$$

Huber não limita a influência da posição.

Mallows

$$\sum_{i=1}^n \psi_C(r_i) w_i \underline{x}_i = 0$$

Mallows limita a influência de pontos de alavanca separadamente dos resíduos.

Hampel-Krasker

$$\sum_{i=1}^n \psi_C\left(\frac{r_i}{\tilde{w}_i}\right) \tilde{w}_i \underline{x}_i = \sum_{i=1}^n \psi_{C\tilde{w}_i}(r_i) \underline{x}_i = 0$$

Hampel-Krasker limita a influência de pontos de alavanca somente se os resíduos correspondentes são grandes.

Figura 2.2. Limitação da influência do resíduo e da posição das variáveis explicativas em alguns estimadores M.

Krasker & Welsch (1982) consideraram uma classe de estimadores de influência limitada eficientes com função η tipo Schweppe. Essencialmente esses estimadores minimizam a matriz de covariância assintótica de $\hat{\beta}$, sujeito a um limite na sensibilidade a erros grosseiros; a escolha desse limite deve ser tal que a eficiência assintótica no modelo normal seja apenas ligeiramente prejudicada. Uma discussão interessante sobre esses estimadores é apresentada em Huber (1983) e Ruppert (1985).

Considerando a ampla revisão de estimadores M generalizados (definidos em (2.23)) desenvolvida em Hampel et alii (1986, cap. 6), nos abstermos de transcrever as propriedades desses estimadores exceto quanto ao ponto de ruptura, que apresentamos a seguir.

No caso de estimadores de mínimos quadrados o ponto de ruptura é $\epsilon^* = 0$. A classe de estimadores M , definida em (2.20) é o primeiro passo na direção de robustificação de estimadores de mínimos quadrados. O ponto de ruptura desses estimadores, no entanto, nunca é maior que 25% e depende da matriz de especificação X ; em particular ϵ^* aproxima-se arbitrariamente de zero se a distribuição de x tem caudas pesadas.

No caso dos estimadores M generalizados definidos em (2.23), o ponto de ruptura não excede $1/p$, onde p é a dimensão do espaço de parâmetros. Assim, o ponto de ruptura torna-se muito pequeno quando o número de parâmetros é grande.

2.5. ALGUNS ESTIMADORES ALTERNATIVOS

O desafio de construir ajustes robustos de modelos de regressão tem estimulado muitos pesquisadores, com conseqüente produção de uma ampla variedade de abordagens. Apresentamos nesta seção uma breve seleção de estimadores recentemente propostos; alguns deles são modificações de estimadores M .

2.5.1. Estimadores M Tipo Huber Adaptativos e Invariantes em Relação à Escala

Jurečková & Sen (1984) propuseram uma versão adaptativa dos estimadores M para regressão invariantes em escala baseando-se no conceito de quantis de regressão devido a Koenker & Bassett (1978). Esse estimador de β no modelo (1.1) é um estimador M como definido em (2.20) com a função escore ψ proposta por Huber (1.9); os autores definem a constante de ajuste de forma adaptativa, isto é, utilizam os resultados da amostra na sua determinação. Nesse caso a função escore é dada por:

$$\psi_n(x) = \begin{cases} x, & \text{se } |x| \leq k_n(\alpha) \\ k_n \operatorname{sign}(x), & \text{se } |x| > k_n(\alpha) \end{cases} \quad (2.25)$$

onde

$$k_n(\alpha) = \frac{1}{2} (\hat{\beta}_1(1 - \alpha) - \hat{\beta}_1(\alpha)), \quad (2.26)$$

$\hat{\beta}_1(\alpha)$ e $\hat{\beta}_1(1 - \alpha)$ são as primeiras componentes dos vetores constituídos pelo α -ésimo e $(1 - \alpha)$ -ésimo quantis de regressão, respectivamente.

Jurečková & Sen (1984) mostraram que esse estimador é invariante em relação à escala e tem a mesma distribuição assintótica que o estimador M de Huber usual desde que as condições relacionadas a seguir sejam satisfeitas.

(A) Condições sobre a sequência de matrizes de especificação X_n ($n \times p$):

$$(A.1) \quad x_{i1} = 1, \quad i = 1, \dots, n; \quad \sum_{i=1}^n x_{ij} = 0; \\ j = 2, \dots, p.$$

$$(A.2) \quad \left(\max_{1 \leq i \leq n} |x_{ij}| \right) \left(\sum_{i=1}^n x_{ij}^2 \right) \rightarrow 0, \quad \text{com } n \rightarrow \infty; \quad j = 2, \dots, p.$$

(A.3) Existe uma matriz $\underline{Q}$ ($p \times p$) definida positiva tal que

$$\lim_{n \rightarrow \infty} n^{-1} \underline{X}' \underline{X} = \underline{Q}$$

(A.4) $\max_{1 \leq j \leq p} (n^{-1} \sum_{i=1}^n |x_{ij}|^3) = o(1)$, com $n \rightarrow \infty$ e

$$\max_{1 \leq j \leq p} (n^{-1} \sum_{i=1}^n x_{ij}^4) = o(1), \text{ com } n \rightarrow \infty.$$

(B). Condições sobre a função de distribuição F dos erros aleatórios:

(B.1) F tem uma densidade absolutamente contínua e informação de Fisher finita.

(B.2) $f(-x) = f(x)$, $x \in \mathbb{R}$

(B.3) $0 < f(x) < \infty$ e $f'(x)$ é limitada na vizinhança de $F^{-1}(\alpha_0)$ e $F^{-1}(1 - \alpha_0)$, onde

$$\alpha_0 = (1 - \epsilon) (1 - \Phi(k)) + \frac{\epsilon}{2} \quad (2.27)$$

e k e ϵ são definidos através do modelo de contaminação (1.5) e (1.8).

Sob os conjuntos de condições (A) e (B) os autores mostraram ainda que o estimador proposto tem, assintoticamente, distribuição normal com média $\underline{\beta}$ e matriz de covariância

$$n^{-1} \sigma^2(\alpha_0, F) \cdot \underline{Q}^{-1} \quad (2.28)$$

onde $\sigma^2(\alpha_0, F)$ é a variância assintótica da média α_0 -aparada, isto é,

$$\begin{aligned} \sigma^2(\alpha_0, F) = (1 - 2\alpha_0)^{-2} \left\{ \int_{\alpha_0}^{1-\alpha_0} (F^{-1}(t))^2 dt + \right. \\ \left. + 2\alpha_0 (F^{-1}(1 - \alpha_0))^2 \right\} \end{aligned} \quad (2.29)$$

e é assintoticamente minimax sob o modelo normal contaminado(1.5).

Esse estimador será retomado no Capítulo 4 como uma possível alternativa para a solução do problema de escala dos estimadores M.

2.5.2. Regressão de Mínima Mediana de Quadrados ("least median of squares")

Rousseeuw (1984) introduziu a idéia de minimização da mediana dos quadrados (MMQ) dos resíduos como uma proposta de obtenção de estimadores com alto ponto de ruptura. Esse estimador é definido como o valor de $\tilde{\beta}$ que minimiza

$$\text{med}_i \{ (y_i - \tilde{x}_i^T \tilde{\beta})^2 \} \quad (2.30)$$

Notemos que a minimização de $\text{med} \{ r_i^2 \}$, onde $r_i = y_i - \tilde{x}_i^T \tilde{\beta}$, quando n é ímpar é equivalente à minimização de $\text{med} \{ |r_i| \}$ que é, essencialmente, o estimador de escala MAD. O ponto de ruptura para amostras finitas é $[(n/2) - p + 2]/n$. Portanto, assintoticamente, $\epsilon^* = 50\%$ que é o melhor valor possível.

A aplicação da abordagem MMQ a problemas de regressão múltipla envolve um grande esforço computacional; no entanto são disponíveis programas de computador. (Leroy & Rousseeuw, 1984).

Uma desvantagem do estimador MMQ é sua baixa eficiência quando os erros têm distribuição normal. Alguns autores (Hampel et alii (1986), por exemplo) sugerem sua utilização como valor inicial para estimadores M devido a sua interessante propriedade de alto ponto de ruptura; porém, deve-se levar em consideração o elevado esforço computacional exigido.

Duas variantes mais eficientes de estimadores MMQ são:

- i. o estimador de mínimos quadrados aparados ("least trimmed squares") sugerido por Rousseeuw (1984), e

- ii. o estimador S que se baseia na minimização de um estimador M da escala dos resíduos (Rousseeuw & Yohai, 1984).

Hampel et alii (1986) sugerem o uso de uma estimativa de alto ponto de ruptura seguido de um estimador M em um passo com uma função ψ redescendente; na presença de pontos de alavanca esses autores sugerem o uso de um estimador de influência limitada. Yohai (1987) comenta que esse tipo de procedimento provavelmente conserva o alto ponto de ruptura do estimador inicial e melhora sua eficiência através do uso do estimador M; no entanto, seu ponto de ruptura exato não é conhecido.

2.5.3. Estimadores MM

Yohai (1987) introduziu uma classe de estimadores robustos para o modelo de regressão (estimadores MM) que são altamente eficientes quando os erros têm distribuição normal e têm ponto de ruptura $\epsilon^* = 50\%$. Esses estimadores são definidos por um procedimento em três estágios:

- i. cálculo de uma estimativa inicial de regressão que seja robusta consistente e tenha alto ponto de ruptura mas não seja necessariamente eficiente;
- ii. cálculo de uma estimativa M da escala dos erros usando os resíduos obtidos através da estimativa inicial, e
- iii. cálculo de uma estimativa M dos parâmetros de regressão baseada em uma função ψ redescendente apropriada.

Yohai (1987) prova a consistência e normalidade assintótica das estimativas MM assumindo $\underline{x}_1, \dots, \underline{x}_n$ aleatórios.

Mais recentemente, Yohai & Zamar (1988) apresentaram uma classe alternativa de estimadores (estimadores τ) que possuem também as propriedades de alto ponto de ruptura e alta eficiência

sob erros normais. Essas estimativas são definidas pela minimização de uma nova estimativa da escala dos erros introduzida pelos autores. Uma vantagem adicional dessa abordagem é que ela fornece simultaneamente com as estimativas dos coeficientes de regressão, uma estimativa da escala dos erros com alto ponto de ruptura e alta eficiência.

3. ESTIMADORES M

3.1. INTRODUÇÃO

Em geral, estimadores M não são invariantes em relação à escala; mais especificamente, quando definidos por (2.19) ou (2.20), os estimadores M dependem da unidade de y. Os estimadores de mínimos quadrados e L_1 são as exceções. Então, mesmo que haja interesse em estimar apenas β é preciso levar em consideração o parâmetro de escala. Apresentamos na Seção 3.2 duas abordagens que produzem estimadores M invariantes em relação à escala e alguns algoritmos de cálculo propostos na literatura. Na Seção 3.3 relacionamos alguns pacotes de computador disponíveis para o cálculo dos estimadores M e na Seção 3.4 estudamos estimadores para a matriz do covariância de $\hat{\beta}$.

3.2. ESTIMADORES M INVARIANTES EM RELAÇÃO À ESCALA

3.2.1. Introdução

Vamos considerar duas estratégias que permitem a obtenção de estimadores M invariantes em relação à escala. A primeira consiste na obtenção de uma estimativa preliminar robusta da escala $\hat{\sigma}$ e no cálculo de uma estimativa M para β através do sistema de equações (2.21) considerando $\hat{\sigma}$ fixo.

A segunda consiste na obtenção de estimativas simultâneas de β e σ ; nesse sentido Huber (1981, Cap. 7) propôs a estimação de β e σ através da solução das equações simultâneas

$$\sum_{i=1}^n \psi \left(\frac{y_i - x_i^T \beta}{\sigma} \right) x_i = 0 \quad (3.1)$$

$$\sum_{i=1}^n \chi \left(\frac{y_i - \tilde{x}_i^T \underline{\beta}}{\sigma} \right) = na \quad (3.2)$$

onde $\chi(x)$ é uma função par e $\underline{a}$ é uma constante positiva. Huber sugere que ψ e χ sejam relacionadas pois, caso contrário, há problemas nas provas de existência dos estimadores M simultâneos de $\underline{\beta}$ e σ e convergência dos algoritmos utilizados para obtenção dos estimadores. A escolha recomendada por Huber(1981) é

$$\chi(x) = x \psi(x) - \rho(x) \quad (3.3)$$

Huber sugere também tomar a constante $\underline{a}$ como

$$a = \frac{(n - p)}{n} E_{\Phi}(\chi) \quad (3.4)$$

de forma que a estimativa da escala sob o modelo normal seja consistente e produza as estimativas clássicas quando $\rho(x) = x^2/2$.

3.2.2. Algoritmos para Cálculo de Estimativas M Invariantes em Relação à Escala

$\hat{\sigma}$ fixo

O estimador M para os coeficientes de regressão é definido neste caso pelo sistema de equações (2.21). Holland & Welsch (1977) consideram os seguintes três métodos iterativos para solução dessas equações:

1. Método de Newton

O princípio desse método é a utilização de uma aproximação linear em $\underline{\beta}$ para a função $\sum \psi[(y_i - \tilde{x}_i^T \underline{\beta}) / \hat{\sigma}] \tilde{x}_i$. A aplicação do método de Newton (ou Newton - Raphson) (ver Carnahan et alii, 1969, por exemplo), sugere a utilização iterativa da expressão

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \hat{\sigma} \left[\sum_{i=1}^n \tilde{x}_i \psi' \left(r_i^{(m)} / \hat{\sigma} \right) \tilde{x}_i^T \right]^{-1} \tilde{x}^T \psi \left(\frac{\tilde{y} - \tilde{X} \hat{\beta}^{(m)}}{\hat{\sigma}} \right) \quad (3.5)$$

onde

$$\psi'(t) = d\psi(t)/dt,$$

$$r_i^{(m)} = y_i - \tilde{x}_i^T \hat{\beta}^{(m)}$$

$$\psi \left(\frac{\tilde{y} - \tilde{X} \hat{\beta}^{(m)}}{\hat{\sigma}} \right) = \left[\psi \left(\frac{r_1^{(m)}}{\hat{\sigma}} \right), \dots, \psi \left(\frac{r_n^{(m)}}{\hat{\sigma}} \right) \right]^T,$$

$m = 0, 1, \dots$ e $\hat{\beta}^{(0)}$ é um valor inicial arbitrário.

2. Método de Huber ou de resíduos modificados

Huber (1977) sugeriu que a solução de (2.21) fosse obtida através da resolução iterativa de:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \hat{\sigma} (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \psi \left(\frac{\tilde{y} - \tilde{X} \hat{\beta}^{(m)}}{\hat{\sigma}} \right) \quad (3.6)$$

onde

$m = 0, 1, \dots$ e $\hat{\beta}^{(0)}$ é um valor inicial arbitrário.

A motivação do algoritmo está na substituição do vetor de resíduos $\tilde{r}$ na expressão alternativa do estimador de mínimos quadrados

$$\hat{\beta} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{y} = \hat{\beta} + (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{r}$$

pelo vetor de resíduos modificados $\psi(\tilde{r} / \hat{\sigma})$, onde $\tilde{r} = \tilde{y} - \tilde{X} \hat{\beta}$.

3. Método dos mínimos quadrados reponderados iterativamente (MQRI) ou método dos pesos modificados.

Beaton & Tukey (1974) descreveram outro método para a solução aproximada da equação (2.21) a partir da relação

$$\begin{aligned} \sum_{i=1}^n \psi[(y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma}] x_{ij} &= \\ &= \sum_{i=1}^n x_{ij} \frac{\psi[(y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma}]}{(y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma}} (y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma} = 0 \end{aligned}$$

$$j = 1, 2, \dots, p$$

que pode ser reescrita como

$$\sum_{i=1}^n x_{ij} w_i (y_i - \tilde{x}_i^T \hat{\beta}) = 0, \quad j = 1, 2, \dots, p$$

onde

$$w_i = \frac{\psi[(y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma}]}{(y_i - \tilde{x}_i^T \hat{\beta})/\hat{\sigma}}$$

Beaton & Tukey (1974) sugeriram a solução iterativa de

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\tilde{X}^T \tilde{W}^{(m)} \tilde{X})^{-1} \tilde{X}^T \tilde{W}^{(m)} (\tilde{y} - \tilde{X} \hat{\beta}^{(m)}) \quad (3.7)$$

onde $\tilde{W}^{(m)}$ é uma matriz diagonal com os elementos

$$w_i^{(m)} = \frac{\psi[(y_i - \tilde{x}_i^T \hat{\beta}^{(m)})/\hat{\sigma}]}{(y_i - \tilde{x}_i^T \hat{\beta}^{(m)})/\hat{\sigma}},$$

$m = 0, 1, \dots$ e $\hat{\beta}^{(0)}$ é um valor inicial arbitrário.

Uma forma equivalente para (3.7) é:

$$\hat{\beta}^{(m+1)} = (\tilde{X}^T \tilde{W}^{(m)} \tilde{X})^{-1} \tilde{X}^T \tilde{W}^{(m)} \tilde{y}, \quad (3.8)$$

que deu origem ao termo "mínimos quadrados reponderados iterativamente".

Comparando as equações (3.5), (3.6) e (3.7) vemos que o

método de Newton requer o uso de ψ' , o que pode torná-lo mais difícil de implementar. O método de Huber tem propriedades computacionais convenientes uma vez que $(\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T$ é calculada apenas uma vez. O método MQRI permite utilizar, a cada iteração, pacotes estatísticos já disponíveis para cálculo de estimadores de mínimos quadrados ponderados ou, ainda, um programa de mínimos quadrados após multiplicação de y_i e $\tilde{x}_i^T$ por $\sqrt{w_i^{(m)}}$.

Se a função escore ψ for monótona todos os três métodos produzem seqüências convergentes de $\hat{\beta}^{(m)}$, sendo que o método de Newton, segundo Holland & Welsch (1977), converge mais rapidamente; o método MQRI converge um pouco mais rapidamente do que o método de Huber embora a evidência empírica mostre somente uma pequena diferença (ver Huber (1981), cap. 7).

Com funções ψ não monótonas (por exemplo, funções ψ redescendentes), o sistema de equações (2.21) pode ter raízes múltiplas, incluindo tanto máximos quanto mínimos locais. Tanto o método de Newton quanto o método MQRI, sob condições suaves sobre ψ (geralmente satisfeitas pelas funções escores usuais) e com escala fixa, sempre convergem para alguma raiz. Nesse caso, é essencial começar o processo iterativo com uma boa estimativa para os parâmetros. Por exemplo com uma estimativa baseada num estimator consistente; nesse caso, pelo menos para grandes amostras, teremos grande probabilidade de estarmos "perto" do valor do parâmetro.

Estimação Simultânea de β e σ

A segunda abordagem para o problema da invariância dos estimadores M em relação à escala consiste em estimar β e σ simultaneamente através das equações (3.1) e (3.2). Apresentamos a seguir proposta de Huber (1981) para os processos iterativos de solução.

i. Escala

Sejam $\hat{\beta}^{(m)}$ e $\hat{\sigma}^{(m)}$ os valores de β e σ na iteração m e $r_i = y_i - \tilde{x}_i^T \hat{\beta}^{(m)}$. Então

$$(\hat{\sigma}^{(m+1)})^2 = \frac{1}{na} \sum_{i=1}^n \chi(r_i / \hat{\sigma}^{(m)}) (\hat{\sigma}^{(m)})^2 \quad (3.9)$$

ii. Coeficientes de regressão

Existem duas alternativas sugeridas por Huber (1981).

1. o método de Huber ou de resíduos modificados, dado por (3.6) ou
2. o método MQRI ou de pesos modificados, dado por (3.7).

Huber (1981) mostrou a convergência desses algoritmos, independentemente do valor inicial, quando ψ é monótona; se esse não é o caso, então pode-se obter somente convergência local.

Valores iniciais para os parâmetros

i. Coeficientes de regressão

Estimadores M são particularmente sensíveis à escolha dos valores iniciais quando as funções ψ são redescendentes; um estimador inicial ruim pode levar a uma solução da equação (2.21) que esteja "longe" do mínimo global em termos de β . No problema de locação, a escolha usual é a mediana (Andrews et alii, 1972), o que sugere o estimador L_1 , como generalização natural para os modelos de regressão. Por outro lado, o estimador L_1 pode consumir mais tempo computacional do que o desejável para obtenção de um valor inicial. Uma estratégia alternativa recomendada por vários autores como Hogg (1979) por exemplo, é utilizar estimativas de mínimos quadrados como valores iniciais e uma função monótona com estimativa simultânea de β e σ até obter convergência e, então, realizar poucas iterações (talvez apenas uma) com uma função ψ redescendente e $\hat{\sigma}$ mantido fixo.

Se a função ψ for monótona, o estimador de mínimos quadrados é uma escolha possível; não há dificuldades com convergência, é fácil de calcular e os usuários geralmente desejam conhecer os resultados de mínimos quadrados e compará-los com as estimativas robustas.

ii. Escala

A sugestão usual (Holland & Welsch, 1977) para a estimativa inicial do parâmetro de escala nos algoritmos onde é mantido fixo é

$$\hat{\sigma} = \text{MAD}/0,6745 = 1,483 \text{ med}_i \{ |y_i - \tilde{x}_i^T \hat{\beta}^{(0)}| \} - \text{med}_j \{ |y_j - \tilde{x}_j^T \hat{\beta}^{(0)}| \} \quad (3.10)$$

Holland & Welsch (1977) examinaram, através de um estudo de Monte Carlo, os efeitos em vários estimadores M do número de iterações e da estimação do parâmetro de escala quando comparada à situação onde a escala é conhecida. Seus resultados indicaram que, em geral, os estimadores M são mais sensíveis à estimação do parâmetro de escala do que ao número de iterações utilizado. Esses resultados motivaram a realização de um estudo comparativo de algumas soluções disponíveis para obtenção de estimadores M invariantes em relação à escala, que apresentamos no Capítulo 4.

3.3. ALGUNS PACOTES COMPUTACIONAIS

Hampel et alii (1986) descrevem três pacotes de computador para o cálculo de estimadores M :

1. **ROBETH.** É um pacote escrito em FORTRAN 77 que contém sub-rotinas para análise de regressão linear robusta e obtenção de estimativas M de matrizes de covariância. Foi escrito na ETH, Zurique e vem sendo continuamente ampliado na Universidade de Lausanne (ver Marazzi, 1980).

a, b). O pacote ROBETH incorpora os principais procedimentos de regressão robusta, incluindo estimadores de influência limitada e testes robustos para parâmetros de modelos lineares. Ele apresenta três opções para o cálculo dos estimadores M em regressão, incluindo o algoritmo MQRI e o método de Huber, que foram adaptados para o cálculo de estimadores de influência limitada. Do ponto de vista numérico, a principal característica dessas sub-rotinas é o uso de transformações Householder (ver Lawson & Hanson, 1974), o que permite obter considerável estabilidade numérica.

2. LINWDR. Considera estimadores M de regressão e é uma versão modificada do programa de mínimos quadrados LINWOOD (Daniel & Wood, 1971, 1980). Foi desenvolvido por Dutter (1976, 1983) e utiliza o método de Huber ou resíduos modificados no cálculo dos estimadores.
3. TROLL. É um grande sistema interativo de modelagem e análise de dados desenvolvido no MIT. O sistema contém um pacote chamado BIF para o cálculo de estimadores de influência limitada cujo algoritmo é basicamente o mesmo de ROBETH (ver Peters et alii, 1982).

O estimador L_1 de regressão pode, ainda, ser calculado pelo pacote estatístico SAS através do procedimento LAV desenvolvido por Lee & Gentle (1986) e disponível na biblioteca suplementar. O procedimento usa uma versão modificada do algoritmo de Barrodale & Roberts (1974).

Os algoritmos propostos na Seção 3.2 podem, também, ser implementados em pacotes estatísticos não específicos para regressão robusta. Entre as diversas possibilidades, as estimativas M podem ser calculadas através dos recursos de álgebra matricial disponíveis nos pacotes; adaptação do procedimento de mínimos quadrados através do MQRI ou, ainda, adaptação de procedimentos de cálculo de estimadores de regressão não-linear, por exemplo, o procedimento NLIN do SAS.

3.4. ESTIMADORES PARA A MATRIZ DE COVARIÂNCIA DE $\hat{\beta}$

Em problemas práticos usualmente há interesse em conhecer a precisão das estimativas dos coeficientes de regressão para construção de intervalos de confiança ou testes de hipóteses envolvendo os componentes de $\hat{\beta}$.

Sob algumas condições de regularidade, a matriz de covariância assintótica dos estimadores dos coeficientes de regressão é dada por

$$\text{COV}(\hat{\beta}) = \sigma^2 \frac{E \psi^2 \left(\frac{\epsilon}{\sigma} \right)}{[E \psi' \left(\frac{\epsilon}{\sigma} \right)]^2} (\tilde{X}^T \tilde{X})^{-1} \quad (3.11)$$

É natural estimar $E \psi^2 \left(\frac{\epsilon}{\sigma} \right)$ por $\sum_{i=1}^n \psi^2 \left(\frac{r_i}{\hat{\sigma}} \right) / n$ e $E \psi' \left(\frac{\epsilon}{\sigma} \right)$ por $\sum_{i=1}^n \psi' \left(\frac{r_i}{\hat{\sigma}} \right) / n$; para obter consistência de resultados no caso do estimador de mínimos quadrados, Huber (1981) propôs o seguinte estimador para $\text{COV}(\hat{\beta})$

$$\hat{\text{COV}}(\hat{\beta}) = \frac{\sum_{i=1}^n \psi \left(\frac{r_i}{\hat{\sigma}} \right)^2 / (n - p)}{[\sum_{i=1}^n \psi' \left(\frac{r_i}{\hat{\sigma}} \right) / n]^2} (\tilde{X}^T \tilde{X})^{-1} \hat{\sigma}^2 \quad (3.12)$$

Huber (1981, cap. 7) apresenta outras sugestões que sob algumas condições são estimativas não-viciadas de $\text{COV}(\hat{\beta})$; entre elas temos (3.12) multiplicada por um fator de correção

$$k = \left(1 + \frac{P}{n} \frac{\text{Var}(\psi')}{(E \psi')^2} \right)^2 \quad (3.13)$$

Na prática $E \psi'$ e $\text{Var} \psi'$ são desconhecidos e podem ser estimados por

$$\begin{aligned} E \psi' &\approx \frac{1}{n} \sum_{i=1}^n \psi' (r_i / \hat{\sigma}) = m \\ \text{Var} \psi' &\approx \frac{1}{n} \sum_{i=1}^n (\psi' (r_i / \hat{\sigma}) - m)^2 \end{aligned} \quad (3.14)$$

A expressão (3.12) é sensível em relação a valores discrepantes de $\tilde{x}$; existem várias outras sugestões para estimação da matriz de covariância de $\tilde{\beta}$ (Hill (1979), por exemplo), mas nenhuma delas é completamente satisfatória.

4. COMPARAÇÃO DE ALGUNS ESTIMADORES ROBUSTOS EM MODELOS DE REGRESSÃO

4.1. INTRODUÇÃO

Neste capítulo vamos considerar alguns estimadores M invariantes em relação à escala, com base nos princípios de:

- i. estimação simultânea de β e σ (sistema de equações (3.1) - (3.4));
- ii. estimação preliminar da escala (sistema de equações (2.21)), e
- iii. estimação adaptativa (descrita na Seção 2.5.1).

Na Seção 4.2 apresentamos detalhadamente esses estimadores; na Seção 4.3 desenvolvemos um estudo de Monte Carlo objetivando comparar o desempenho desses estimadores em situações envolvendo diversos desvios do modelo normal e diferentes tamanhos de amostra. Na Seção 4.4 ilustramos a utilização desses métodos de estimação através de problemas práticos de regressão.

4.2. DEFINIÇÃO DOS ESTIMADORES E ASPECTOS COMPUTACIONAIS

Descrevemos a seguir quatro estimadores escolhidos para o estudo.

1. Estimador M baseado em estimativas simultâneas de β e σ

Escolhemos a função escore proposta por Huber (equação (1.9)) com constante de ajuste $k = 1,345$ que produz 95% de eficiência assintótica sob o modelo de locação quando a distribuição subjacente é normal (ver Holland & Welsch (1977), por exemplo). As estimativas de mínimos quadrados foram escolhidas como valores iniciais para β e σ no processo iterativo e os cálculos foram feitos utilizando

o método de pesos modificados implementado em ROBETH. O critério de convergência escolhido foi

$$\left| d_j \right| < 0,001 \hat{\sigma}^{(m)} \sqrt{\text{COV}_{jj}}, \quad j = 1, 2, \dots, p \quad (4.1)$$

e

$$\frac{\left| \hat{\sigma}^{(m+1)} - \hat{\sigma}^{(m)} \right|}{\max(1, \hat{\sigma}^{(m)})} < 0,001 \quad (4.2)$$

com a limitação de que o número de iterações (m) fosse menor ou igual a 100; onde d_j representa o $j^{\text{ésimo}}$ componente do 2º termo da equação (3.7)

$$\underline{d} = (\underline{X}^T \underline{W}^{(m)} \underline{X})^{-1} \underline{X}^T \underline{W}^{(m)} (\underline{y} - \underline{X} \hat{\underline{\beta}}^{(m)}) \quad (4.3)$$

e

$$\text{COV} = \frac{E_{\Phi} \psi^2}{[E_{\Phi} \psi']^2} (\underline{X}^T \underline{X})^{-1} \quad (4.4)$$

Denotamos este estimador por HUSI (função de Huber, escala iterada).

2. Estimador M baseado numa estimativa preliminar da escala

Selecionamos dois estimadores como representantes dessa abordagem; um com função ψ monótona (estimador HUSC) e o outro com função ψ redescendente (estimador TUSC).

Estimador HUSC: função ψ proposta por Huber com $k = 1,345$, estimativa de mínimos quadrados como valor inicial de $\underline{\beta}$ ($\hat{\underline{\beta}}^{(0)}$) e estimativa robusta de escala obtida através da equação (3.10) a partir da estimativa de mínimos quadrados.

Estimador TUSC: função ψ bi-ponderada de Tukey (equação 2.15), estimativa L_1 como valor inicial de $\underline{\beta}$ e es-

timativa preliminar da escala baseada nos resíduos da estimação $L_1(r_i^{(0)})$

$$\hat{\sigma} = \text{med}_i (|r_i^{(0)}|) / 0,6745 \quad (4.5)$$

Para que os estimadores sejam comparáveis, a constante de ajuste da função ψ bi-ponderada foi escolhida sob o mesmo critério da função ψ de Huber, ou seja, 95% de eficiência no modelo de locação sob distribuição normal. Essa constante é $B = 4,685$ (ver Holland & Welsch(1977), por exemplo).

Para obtenção das estimativas, utilizamos o algoritmo de pesos modificados de ROBETH com critério de convergência

$$|d_j| < 0,001 \hat{\sigma} \sqrt{\text{COV}_{jj}}, \quad j = 1, 2, \dots, p \quad (4.6)$$

e número de iterações menor ou igual a 100, onde d_j e COV são definidos em (4.3) e (4.4), respectivamente.

A estimativa L_1 também foi calculada com o pacote ROBETH que utiliza o algoritmo de Barrodale & Roberts (1974). Em ambas as abordagens, a matriz de covariância dos estimadores dos coeficientes de regressão é estimada pela equação (3.12), multiplicada pelo fator de correção (3.13).

3. Estimador M adaptativo de Jurečková & Sen

Jurečková & Sen (1984) propuseram um estimador M de $\tilde{\beta}$ invariante em relação à escala através da definição da constante de ajuste da função de Huber de forma adaptativa, conforme as equações (2.25) e (2.26).

A determinação do quantil α foi feita através da equação (2.27) com ϵ e k definidos pela relação (1.8). Para tornar esse estimador comparável aos demais adotamos o seguinte procedimento: com o valor da constante de ajuste da função de Huber escolhido para os estimadores HUSI e HUSC ($k = 1,345$), determinamos a proporção de con

taminação através de (1.8), obtendo $\epsilon \approx 0,058$. Com o par $(k = 1,345; \epsilon = 0,058)$ em (2.27), determinamos o valor do quantil α a ser considerado na determinação da constante de ajuste adaptativa, $k_n(\alpha)$, obtendo $\alpha = 0,11$.

Com a matriz de especificação $\tilde{X}$, transformada de forma a satisfazer o conjunto de condições (A) descrito na Seção 2.5.1, obtivemos os primeiros componentes dos quantis de regressão $\hat{\beta}_1(0,11)$ e $\hat{\beta}_1(0,89)$ utilizando o algoritmo de Wellington & Narula (1984) e, finalmente, determinamos a constante de ajuste adaptativa através de (2.26).

Para o cálculo de $\hat{\beta}$ utilizamos o método de pesos modificados implementado em ROBETH com estimativa inicial dos coeficientes de regressão obtida pelo método dos mínimos quadrados. O critério de convergência utilizado foi

$$|d_j| < 0,001 \hat{\sigma} \sqrt{(\tilde{X}^T \tilde{X})_{jj}^{-1}}, \quad j = 1, 2, \dots, p \quad (4.7)$$

com a limitação de que o número de iterações fosse menor ou igual a 100 onde:

$$\hat{\sigma} = \frac{1}{0,6745} \text{MAD}(r_i^{(0)}),$$

$r_i^{(0)}$: i -ésima componente do vetor de resíduos da estimativa de mínimos quadrados;

d_j : j -ésima componente do vetor $\underline{d}$ da fórmula de iteração do algoritmo de pesos modificados, dada abaixo:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \underline{d}, \quad \text{onde} \quad (4.8)$$

$$\underline{d} = (\tilde{X}^T \tilde{W}^{(m)} \tilde{X})^{-1} \tilde{X}^T \tilde{W}^{(m)} (\underline{y} - \tilde{X} \hat{\beta}^{(m)})$$

$\tilde{W}^{(m)}$: matriz diagonal com os elementos não nulos dados por

$$w_i^{(m)} = \Psi_n(r_i^{(m)})/r_i^{(m)}, \quad i = 1, 2, \dots, n$$

Este estimador será denotado por J&S.

A matriz de covariância assintótica é estimadora por:

$$\text{COV}(\hat{\beta}) = \hat{\sigma}^2(\alpha) (\tilde{X}'\tilde{X})^{-1}, \quad (4.9)$$

onde $\hat{\sigma}^2$ é uma estimativa consistente de σ^2 (Bickel, 1973) dada por

$$\hat{\sigma}^2(\alpha) = \frac{1}{(1 - 2\alpha)} \left\{ \frac{1}{n-p} \sum_{i=r+1}^{n-r} r_{(i)}^2 + \alpha(r_{(r+1)}^2 + r_{(n-r)}^2) \right\}, \quad (4.10)$$

onde

$r = [n\alpha]$, o maior inteiro contido em $n\alpha$ e

$r_{(1)} < r_{(2)} < \dots < r_{(n)}$ são os resíduos ordenados.

4.3. ESTUDO DE MONTE CARLO

Esse estudo tem por objetivo comparar o desempenho dos estimadores definidos na Seção 4.2 avaliados, principalmente, em termos do erro quadrático médio em situações com desvios do modelo normal representadas por cinco distribuições de probabilidade.

Os fatores envolvidos nesse estudo são:

- (1) **estimadores:** HUSI, HUSC, TUSC, J&S, descritos na Seção 4.2 além dos estimadores de mínimos quadrados (MQ) e L_1 .
- (2) **distribuições dos erros:** i. normal; ii. distribuições simétricas com desvios em relação à normal que classificamos como:
 - . desvios suaves a moderados (caudas moderadas): distribuições normal padrão ($N(0,1)$) contaminada com 5, 10 e 20% com distribuição $N(0,9)$;
 - . desvios moderados a grandes (caudas pesadas): distribuição $N(0,1)$ com 5% de contaminação com distribuição $N(0,100)$;
 - . desvios acentuados (caudas muito pesadas): distribuição "Slash" ($N(0,1)/U(0,1)$) onde $U(0,1)$ representa a distribuição uniforme no intervalo $[0,1]$, que possui aproximadamente caudas semelhantes às da

Cauchy e região central como a da distribuição normal.

Esses desvios podem ser visualizados nas Figuras 4.1 a 4.3 onde apresentamos os histogramas e gráficos de probabilidade normal correspondentes a uma amostra aleatória de tamanho 1000 gerada por cada distribuição utilizada nesse estudo; as distribuições normais contaminadas foram padronizadas para variância unitária.

(3) tamanho das amostras:

- i. $n=20$ (mínimo encontrado em muitos trabalhos práticos;
- ii. $n=40$ (moderado) e
- iii. $n=100$ (grande), para verificar se com este tamanho de amostra os métodos robustos são ainda necessários.

(4) matriz de especificação $\tilde{X}$: consideramos o caso $p=2$ com

$$x_{i1} = 1, \quad x_{i2}^* = i, \quad \text{para } i = 1, 2, \dots, n.$$

Os $\{x_{i2}^*\}$ foram transformados para $\{x_{i2}\}$ de maneira que apresentassem média zero e soma de quadrados igual a n ; nesse caso, $\tilde{X}^T \tilde{X} = nI$. Essa transformação satisfaz as condições do estimador J&S e torna os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ não-correlacionados e de mesma variância.

A variável resposta y_i foi obtida através do modelo

$$y_i = 1 + x_{i2} + \text{Erro}_i; \quad i = 1, 2, \dots, n$$

(5) número de repetições escolhemos 1500 repetições para amostras de tamanho 20 e 1000 repetições para amostras de tamanho 40 e 100.

Os programas foram escritos em FORTRAN 77 e compilados pelo compilador IBM FORTRAN VS em um computador IBM 4341 do Laboratório de Computação Científica da UFMG. Os erros aleatórios foram obtidos através de transformação Box-Muller (ver Box & Muller (1958)) a partir de números aleatórios gerados através da subrotina GGUW do IMSL.

Figura 4.1. Histogramas baseados em amostras de tamanho 1000.

Figura 4.2. Gráficos de probabilidade normal baseados em amostras de tamanho 1000.

(a)

(b)

(c)

Figura 4.3. Distribuição "Slash": (a) histograma; (b) ampliação das caudas do histograma (a) e (c) gráfico de probabilidade normal, baseados em uma amostra de tamanho 1000.

Resultados

Um resumo dos resultados numéricos das simulações está apresentado no Apêndice A. O erro quadrático médio (EQM) total foi calculado através da soma dos erros quadráticos médios obtidos para cada parâmetro individualmente e a eficiência foi definida por

$$\frac{\text{EQM total do estimador de referência}}{\text{EQM total do estimador}} \times 100 \quad (4.11)$$

Analizamos a seguir os principais aspectos desses resultados nas seguintes situações:

i. distribuição dos erros da forma $\epsilon N(0,9) + (1 - \epsilon) N(0,1)$

A Figura 4.4 apresenta a eficiência dos diversos estimadores em relação ao estimador de mínimos quadrados para porcentagens de contaminação (ϵ) iguais a 0, 5, 10 e 20.

Para a distribuição normal ($\epsilon = 0$), apesar de selecionarmos as constantes de ajuste das funções escore com um mesmo critério (95% de eficiência assintótica), o estimador TUSC apresenta eficiência um pouco mais baixa que os demais (89%). Porém, o aspecto mais notável dos resultados é que, mesmo para desvios ainda suaves ($\epsilon = 5\%$) o estimador de MQ já perde bastante em eficiência quando comparado aos demais estimadores, com exceção do estimador L_1 que é mais eficiente que o estimador de MQ apenas para $\epsilon = 20\%$. Os demais estimadores têm comportamento bastante semelhante, com HUSI e J&S bastante próximos em todas as situações; HUSC só é ligeiramente superior a HUSI e J&S para $\epsilon = 20\%$. Os estimadores HUSC e TUSC apresentam um desempenho um pouco superior aos demais para $\epsilon = 20\%$ em todos os tamanhos de amostra.

Concluimos que os estimadores pesquisados apresentam

Figura 4.4. Eficiência em relação ao estimador de Mínimos Quadrados para diferentes valores de porcentagem de contaminação (ϵ) no modelo $\epsilon N(0,9) + (1 - \epsilon) N(0,1)$.

desempenho similar na presença de desvios suaves em relação à normal, com ligeira perda de eficiência sob o modelo normal ao se usar o estimador TUSC.

A Tabela 4.1 mostra a eficiência de cada estimador em relação ao que possui menor EQM total para a mesma combinação distribuição-tamanho de amostra, o que permite discriminar melhor os estimadores que apresentam comportamento mais próximo.

ii. distribuição dos erros 5% $N(0,100)$ + 95% $N(0,1)$

A Figura 4.5 apresenta o erro quadrático médio total dos estimadores para os tamanhos de amostra investigados. Nessa situação o estimador MQ apresenta comparativamente uma grande queda de eficiência; o estimador L_1 é mais eficiente que MQ para todos os tamanhos de amostra, porém, bem menos eficiente que os demais (ver também Tabela 4.1). O estimador TUSC é, sistematicamente, o de menor EQM total mas para amostras de tamanho 40 e 100 os estimadores HUSI, HUSC, TUSC e J&S praticamente se equivalem. Para amostras menores há alguma discriminação entre os estimadores; com ordem decrescente de desempenho temos: TUSC, HUSI, HUSC e J&S.

iii. distribuição dos erros da forma "Slash"

Os resultados envolvendo o erro quadrático médio total dos estimadores em função do tamanho da amostra são apresentados na Figura 4.6. Nessa situação de cauda muito mais pesada que a normal, os estimadores apresentam desempenho bastante desigual. Os estimadores MQ e HUSC perdem eficiência drasticamente; o estimador J&S mostra um comportamento ruim para amostras pequenas a moderadas e o estimador HUSI apresenta-se um pouco melhor. O estimador TUSC é sistematicamente o de maior eficiência seguido de perto pelo estimador L_1 .

Tabela 4.1. Eficiência (%) em relação no estimador de menor erro quadrático médio total na linha.

DISTRIBUIÇÃO	n	E S T I M A D O R					
		HUSI	HUSC	TUSC	J&S	MQ	L_1
NORMAL	20	97	94	89	95	100	65
	40	93	92	89	92	100	61
	100	95	94	92	93	100	64
5%N(0,9)+95%N(0,1)	20	100	99	96	99	84	71
	40	100	99	99	100	85	70
	100	98	98	100	98	79	69
10%N(0,9)+90%N(0,1)	20	100	100	98	99	77	78
	40	98	99	100	98	73	72
	100	99	98	100	98	73	70
20%N(0,9)+80%N(0,1)	20	97	100	99	97	71	78
	40	98	100	100	98	67	78
	100	98	99	100	98	62	76
5%N(0,100)+95%N(0,1)	20	92	87	100	83	22	67
	40	95	93	100	93	19	66
	100	96	96	100	95	21	69
"SLASH"	20	65	0	100	28	0	89
	40	68	4	100	47	0	90
	100	75	1	100	66	0	91

Figura 4.5. Erro quadrático médio total em relação ao tamanho de amostra para erros com distribuição 5% $N(0,100)$ + 95% $N(0,1)$.

DADOS FORA DO GRAFICO :

60

Figura 4.6. Erro quadrático médio em relação ao tamanho de amostra para erros com distribuição "Slash".

Um aspecto que deve ser levado em consideração quando se comparam estimadores é o esforço computacional envolvido. Para isto esboçamos na Figura 4.7 o número de iterações (mínimo, mediano e máximo) e o tempo computacional/repetição gasto no cálculo dos diversos estimadores, em cada situação de tamanho de amostra e distribuição dos erros. Os aspectos mais evidentes são:

- . o estimador HUSI apresenta maior número mediano de iterações particularmente à medida que a distribuição se afasta da normal assim como maior dispersão no número de iterações;
- . o estimador J&S apresenta pequeno número de iterações (não computadas as iterações utilizadas para o cálculo da constante de ajuste), porém, apresenta um tempo computacional bem maior do que os demais principalmente para amostras maiores. Isto se deve ao cálculo da constante adaptativa de ajuste que, devemos lembrar, envolve o cálculo de dois quantis de regressão;
- . o estimador HUSC é o que exige menor tempo computacional.

Avaliamos também as estimativas da matriz de covariância dos parâmetros que são da forma

$$\text{CÔV } \hat{\beta} = f_c \hat{\sigma}^2 (\tilde{X}^T \tilde{X})^{-1} \quad (4.12)$$

com

$f_c = 1$, para o estimador J&S,

$$f_c = k \frac{\sum_{i=1}^n \psi \left(\frac{r_i}{\hat{\sigma}} \right)^2 / (n - p)}{\left[\sum_{i=1}^n \psi' \left(\frac{r_i}{\hat{\sigma}} \right) / n \right]^2}, \text{ para HUSI, HUSC e TUSC}$$

com k definido em (3.13) e (3.14).

Com a matriz de especificação escolhida, $(\tilde{X}^T \tilde{X})^{-1} = \frac{1}{n} \cdot I$, temos

Figura 4.7. (a) Nº de iterações (min; med; max) e (b) Tempo/repetição (seg) em relação ao tamanho de amostra.

$$\text{Var } \hat{\beta}_0 = \text{Var } \hat{\beta}_1 = \frac{f_c \sigma^2}{n} \quad (4.13)$$

e

$$\text{COV } (\hat{\beta}_0, \hat{\beta}_1) = 0$$

Como não conhecemos os valores $\text{Var } \hat{\beta}_0$ ou $\text{Var } \hat{\beta}_1$ vamos avaliar os estimadores das variâncias através de uma aproximação do EQM total

$$\text{"EQM" total} = \sum_{i=0}^1 (n \hat{\text{Var}} \hat{\beta}_i - f_c \hat{\sigma}^2) + \hat{\text{Var}} (f_c \hat{\sigma}^2) \quad (4.14)$$

Apresentamos na Figura 4.8 os resultados obtidos com as distribuições normal e normais contaminadas que foram geradas com mesma variância ($\sigma^2 = 1$).

Notamos que para situações com distribuição normal ou com desvios moderados da mesma, os estimadores apresentam desempenho bem diversificado para amostras de tamanho 20 mas os estimadores HUSI e TUSC apresentam um perfil similar ao longo das distribuições pesquisadas, o que se repete para amostras de tamanho 40. Para amostras grandes todos apresentam desempenho similar.

Os resultados obtidos com a distribuição "Slash" são apresentados na Tabela 4.2. Verificamos a "explosão" da variabilidade da estimativa da variância para o estimador HUSC o que não recomenda sua utilização; o resultado é, também, muito ruim para o estimador J&S com amostras moderadas e HUSI com amostra pequena. Esses resultados são compatíveis com o que já havíamos verificado na avaliação das estimativas dos parâmetros β_0 e β_1 (ver Figura 4.6).

Figura 4.8. "EQM" total da estimativa de (n variância) dos parâmetros.

Tabela 4.2. "EQM" total da estimativa ($f_c \hat{\sigma}^2$) com distribuição SLASH.

TAMANHO DA AMOSTRA	ESTIMADOR			
	HUSI	HUSC	TUSC	J&S
20	101,28	2×10^{10}	34,19	6326,75
40	26,96	$1,5 \times 10^6$	10,89	159,90
100	6,39	$3,0 \times 10^7$	3,20	10,81

4.4. EXEMPLOS

Nesta seção aplicamos os estimadores definidos na Seção 4.2 em conjuntos de dados reais divulgados na literatura. Abordamos, inicialmente, um problema de regressão linear simples onde é fácil identificar dados discrepantes; em seguida apresentamos um modelo de regressão múltipla. Nosso objetivo aqui é adquirir mais conhecimento sobre os estimadores robustos utilizando-os em situações menos controladas, inerentes aos conjuntos de dados reais.

Exemplo 1: Pesos do cérebro e corpo

A Tabela 4.3 apresenta os pesos médios do cérebro (em gramas) e corpo (em kilogramas) de 28 espécies de animais. Esses dados foram obtidos de Rousseeuw & Leroy (1987, pág. 57) e são uma amostra retirada de conjuntos de dados maiores (Weisberg, 1980 e Jerison, 1973). Vamos considerar o problema de investigar e modelar a relação entre peso do cérebro e peso do corpo.

Os dados apresentam uma elevada amplitude em ambas as variáveis sugerindo, pois, transformações logarítmicas como candidatas a expressar de maneira mais clara a relação entre as variáveis. O diagrama de dispersão na escala logarítmica (base 10) é esboçado na Figura 4.9, indicando que há forte relação linear nes-

Tabela 4.3. Pesos médios do cérebro e corpo de 28 espécies.

Índice	Espécies	Peso do Corpo (kg)	Peso do Cérebro (g)
1	Camundongo	0,023	0,400
2	"Hamster"	0,120	1,000
3	Toupeira	0,122	3,000
4	Rato	0,280	1,900
5	Porquinho-da-Índia	1,040	5,500
6	Castor da montanha	1,350	8,100
7	Coelho	2,500	12,100
8	Gato	3,300	25,600
9	Macaco "Rhesus"	6,800	179,000
10	Macaco "Potar"	10,000	115,000
11	Cabra	27,700	115,000
12	Canguru	35,000	56,000
13	Lobo cinzento	36,300	119,500
14	Chimpanzé	52,200	440,000
15	Ovelha	55,500	175,000
16	Homem	62,000	1.320,000
17	Jaguar	100,000	157,000
18	Burro	187,100	419,000
19	Porco	192,000	180,000
20	Gorila	207,000	406,000
21	Vaca	465,000	423,000
22	Cavalo	521,000	655,000
23	Girafa	529,000	680,000
24	Elefante asiático	2.547,000	4.603,000
25	Elefante africano	6.654,000	5.712,000
26	Triceratops	9.400,000	70,000
27	Diplodocus	11.700,000	50,000
28	Brachiosaurus	87.000,000	154,500

sa escala. O ajuste linear aos dados transformados ($\log \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \log x$) é equivalente ao modelo da forma

$$\hat{y} = \hat{\beta}_0' \times \hat{\beta}_1$$

Figura 4.9. Log (peso do cérebro) em relação a Log (peso do corpo) de 28 espécies com retas ajustadas.

entre as variáveis peso do cérebro (y) e peso do corpo (x), com $\hat{\beta}'_0 = 10^{\hat{\beta}_0}$.

As estimativas de MQ, HUSI, HUSC, TUSC e J&S para os parâmetros do modelo linear com seus respectivos desvios padrão são apresentadas na Tabela 4.4; as retas associadas a essas estimativas podem ser visualizadas na Figura 4.9.

Os estimadores robustos em estudo, comparados ao estimador MQ, apresentam uma inclinação mais íngreme e maior precisão nas estimativas sendo que os estimadores HUSI, HUSC e J&S apresentam resultados bem semelhantes com estimativas intermediárias entre MQ e TUSC.

Os pesos finais (ver equação 3.8) atribuídos às observações constituem uma ferramenta útil para detecção de valores discrepantes; as observações que apresentaram baixo peso na estimativa e, portanto, devem ser consideradas com atenção são apresentadas na Tabela 4.5 com seus respectivos resíduos (e_i) e pesos (w_i). As observações com pesos mais baixos (peso nulo para TUSC) apresentam resíduos negativos e são responsáveis pela baixa inclinação no ajuste de MQ; elas se referem a três dinossauros que possuíam cérebro pequeno relativamente ao cor

Tabela 4.4. Estimativas dos parâmetros do modelo $\log y = \beta_0 + \beta_1 \log x + \epsilon$ baseados nos dados referentes a 28 espécies.

Estimador	β_0		β_1	
	Estimativa	Desvio Padrão	Estimativa	Desvio Padrão
MQ	1,110	(0,179)	0,496	(0,078)
HUSI	0,982	(0,119)	0,647	(0,052)
HUSC	1,005	(0,128)	0,627	(0,056)
TUSC	0,867	(0,076)	0,752	(0,033)
J & S	1,035	(0,127)	0,594	(0,055)

Tabela 4.5. Resíduos (e_i) e pesos (w_i) referentes às estimativas da Tabela 4.4.

Índice	Espécie	HUSI		HUSC		TUSC		J & S	
		e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i
9	Macaco "Rhesus"	0,73	0,74	0,73	0,89	0,76	0,22	0,72	1,00
16	Homem	0,98	0,55	0,99	0,65	0,91	0,06	1,02	0,80
26	Triceratops	-1,70	0,32	-1,65	0,39	-2,01	0	-1,55	0,53
27	Diplodocus	-1,91	0,28	-1,85	0,35	-2,23	0	-1,75	0,47
28	Brachiosaurus	-1,99	0,27	-1,91	0,34	-2,39	0	-1,78	0,46

po pesado. Nesse aspecto eles contrastam com o restante do conjunto de dados que se referem a mamíferos. O estimador TUSC apresenta peso baixo também para as observações 9 e 16, isto é, o macaco "Rhesus" que é muito utilizado em pesquisas científicas e o homem. Para eles o peso real do cérebro é maior do que aquele predito pelo modelo linear (resíduos positivos). Esses resultados levam à conclusão que dinossauros, homens e macacos "Rhesus" não obedecem à mesma tendência seguida pela maioria das espécies estudadas (ver Figura 4.9).

Consideremos, agora, apenas os animais mamíferos. As novas estimativas de MQ e dos estimadores em estudo são apresenta

Tabela 4.6. Estimativas dos parâmetros do modelo $y = \beta_0 + \beta_1 \log x + \epsilon$ baseadas nos dados referentes aos mamíferos.

Estimador	β_0		β_1	
	Estimativa	Desvio Padrão	Estimativa	Desvio Padrão
MQ	0,934	(0,087)	0,752	(0,046)
HUSI	0,893	(0,074)	0,750	(0,039)
HUSC	0,885	(0,072)	0,751	(0,038)
TUSC	0,780	(0,067)	0,779	(0,035)
J&S	0,891	(0,091)	0,750	(0,048)

Tabela 4.7. Resíduos (e_i) e pesos (w_i) referentes às estimativas da Tabela 4.6 para observações com $w_i < 0,50$.

Índice	Espécie	HUSI		HUSC		TUSC		J&S	
		e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i
9	Macaco "Rhesus"	0,73	0,48	0,74	0,41	0,82	0,13	0,74	0,46
16	Homem	0,88	0,40	0,84	0,34	0,94	0,02	0,88	0,38

das na Tabela 4.6; os resíduos e pesos referentes às observações que obtiveram peso menor do que 0,50 estão na Tabela 4.7.

Uma análise desses resultados indica que nesse caso os estimadores em estudo apresentam desempenho similar detectando as duas observações discrepantes (macaco "Rhesus" e homem); porém não produzem resultados muito diferentes do estimador de MQ. O estimador TUSC, ao contrário dos demais, apresenta resultados similares nas duas situações devido ao uso do estimador inicial L_1 , mais robusto que MQ, e da função ψ bi-ponderada de Tukey que tem a propriedade de limitar completamente a influência de observações com resíduos elevados.

Exemplo 2: Perda de amônia ("Stack-loss") na produção de ácido nítrico.

O conjunto de dados de Brownlee (1965, pág. 454), reproduzido na Tabela 4.8, representa a perda de amônia (y) registrada em 21 dias consecutivos de operação de uma planta que oxida amônia para produzir ácido nítrico. Registraram-se, também, os valores de três variáveis explicativas que descrevem as condições sob as quais a planta está operando: fluxo de ar (x_1), temperatura de entrada da água de refrigeração (x_2) e concentração do ácido (x_3).

Tabela 4.8. Dados de perda de amônia e variáveis operacionais.

Índice	Fluxo de ar (x_1)	Temperatura (x_2)	Concentração do ácido (x_3)	Perda de amônia (y)
1	80	27	89	42
2	80	27	88	37
3	75	25	90	37
4	62	24	87	28
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

Este conjunto de dados foi analisado por um grande número de estatísticos (Daniel & Wood (1971, 1980), Andrews (1974) e Atkinson (1981), entre vários outros) através de diversos métodos. Daniel & Wood (1971, 1980) apresentaram argumentos justificando que as observações 1, 3, 4 e 21 se referem a es dados transientes e, possivelmente, não podem ser representadas pelo mesmo modelo que as demais que se encontram em regime permanente. Andrews (1974) encontrou as mesmas quatro observações discrepantes quando ajustou um modelo de 1ª ordem por um método robusto; alguns autores consideram, também, a observação 2 como dado discrepante. Atkinson (1981) forneceu evidências da conveniência do uso de uma transformação logarítmica.

Seguindo Andrews (1974), focalizaremos o modelo de 1ª ordem

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

e não consideraremos os modelos alternativos discutidos em algumas das referências citadas.

A Tabela 4.9 apresenta um sumário do ajuste de mínimos quadrados em duas situações: considerando todos os pontos e após retirada das observações 1, 3, 4 e 21 consideradas anômalas por Daniel & Wood (1971, 1980).

Tabela 4.9. Estimativas de mínimos quadrados.

Variável	Todas as 21 observações			Retiradas as observações 1,3,4 e 21		
	Coeficiente	Desvio Padrão	Intervalo 95% confiança	Coeficiente	Desvio Padrão	Intervalo 95% confiança
Intercepto	-39,9	11,90	[-65,0;-14,8]	-37,7	4,73	[-47,9;-27,5]
x_1	0,716	0,135	[-0,43; 1,001]	0,798	0,0674	[0,652;0,944]
x_2	1,30	0,368	[0,52; 2,08]	0,577	0,166	[0,218;0,936]
x_3	-0,152	0,156	[-0,481;0,177]	-0,0671	0,0616	[-0,200;0,710]
	$R^2 = 91,4\%$		$\hat{\sigma} = 3,243$	$R^2 = 97,9\%$		$\hat{\sigma} = 1,253$

Observemos que ao retirar as observações 1, 3, 4 e 21 há considerável aumento na precisão das estimativas e alteração significativa na estimativa de β_2 ; em ambos os casos a variável x_3 não contribui para explicar a perda na amônia.

A Tabela 4.10 apresenta os resultados referentes aos estimadores robustos em estudo e a Tabela 4.11 mostra os resíduos(e_i)

Tabela 4.10. Estimativa dos parâmetros do modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$ com todas as 21 observações.

Variável	HUSI		HUSC		TUSC		J & S	
	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão
Intercepto	-41,15	(10,66)	-41,11	(10,47)	-41,03	(9,07)	-41,15	(9,97)
x_1	0,816	(0,121)	0,819	(0,119)	0,939	(0,103)	0,806	(0,113)
x_2	0,986	(0,330)	0,973	(0,323)	0,549	(0,281)	1,026	(0,309)
x_3	-0,132	(0,140)	-0,131	(0,138)	-0,112	(0,119)	-0,114	(0,131)

Figura 4.11. Resíduos (e_i) e pesos (w_i) referentes aos ajustes com todas as 21 observações.

Nº Obs.	MQ (sem obs. 1,3,4 e 21)		HUSI		HUSC		TUSC		J & S	
	e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i
1	6,22	0	2,93	1,00	2,96	1,00	3,06	0,74	2,90	1,00
3	6,43	0	4,12	0,94	4,13	0,90	3,97	0,59	4,11	1,00
4	8,17	0	6,32	0,61	6,36	0,59	7,39	0,04	6,21	0,68
21	-8,63	0	-8,74	0,44	-8,78	0,43	-10,48	0	-8,60	0,49

e pesos (w_i) correspondentes às observações consideradas anômalas por Daniel & Wood (1971, 1980). Uma análise desses resultados indica que apenas o estimador TUSC apresenta todas as es

timativas dos coeficientes dentro do intervalo de 95% de confiança construído a partir das estimativas de MQ sem as observações anômalas; os demais estimadores apresentam $\hat{\beta}_2$ fora do intervalo correspondente. Os resultados da Tabela 4.11 mostram que os estimadores em estudo limitaram a influência das observações 4 e 21, com peso praticamente nulo para o estimador TUSC. As observações 1 e 3 não foram consideradas discrepantes (peso elevado), a não ser no caso do estimador TUSC que lhes atribuiu peso moderado. Esses resultados são justificados pela observação cuidadosa dos valores das variáveis explicativas (ver Tabela 4.8 e Figura 4.10(a)) e variável resposta correspondente, que pode ser visualizada na Figura 4.10(b) em relação às variáveis explicativas significativas (x_1 e x_2). As observações 1 e 2 estão afastadas das demais quanto aos valores das variáveis x_1 e x_2 e são potencialmente influentes; o mesmo ocorre com a observação 3, porém, em menor extensão. A Figura 4.10(b) mostra que a presença das observações 1 e 3 "puxam" o plano ajustado em sua direção se não for limitada a influência das variáveis explicativas discrepantes; a observação 2, apesar de sua posição, não influencia o ajuste devido ao fato de o valor de y correspondente se adaptar ao plano ajustado pelos demais dados. As observações 4 e 21 apresentam valores de y alto e baixo, respectivamente, em relação às demais observações. Esses argumentos explicam o baixo peso que os estimadores em estudo atribuíram às observações 4 e 21 e a incapacidade de detectar as observações 1 e 3 como observações discrepantes, já que eles estão aptos, por construção, a limitar apenas a influência do resíduo; a limitação da influência da posição das variáveis explicativas pode ser obtida pelo uso de uma estimativa inicial com essa propriedade como, por exemplo, em Andrews (1974).

Para comparação de resultados apresentamos nas Tabelas 4.12 e 4.13 as estimativas obtidas ao se retirar a observação 1 que está distante das demais variáveis quanto aos valores x_1 e x_2 e a resposta registrada corresponde a uma operação da planta fora do equilíbrio (conforme Daniel & Wood (1971, 1980)). Esses resultados estão mais próximos daqueles obtidos com o ajuste de MQ sem as quatro observações anômalas, particularmente no que se refere ao estimador TUSC; o estimador J&S foi o menos capaz de detectar as observações discrepantes e apresenta,

(a)

(b)

Figura 4.10. (a) Valores das variáveis explicativas onde foram feitas as observações.

(b) Perda de amônia em relação às variáveis x_1 e x_2 .

ainda, estimativa de β_2 fora do intervalo de 95% de confiança referente ao ajuste de MQ sem as 4 observações supostamente anômalas. Esse resultado se deve, provavelmente, à posição desfavorável da observação 3 como já mencionado.

Tabela 4.12. Estimativa dos parâmetros do modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \quad \text{sem a observação 1.}$$

Variável	HUSI		HUSC		TUSC		J & S	
	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão	Coeficiente	Desvio Padrão
Intercepto	-38,7	(8,06)	-38,8	(8,74)	-38,1	(5,78)	-39,1	(7,17)
x_1	0,749	(0,096)	0,742	(0,104)	0,828	(0,069)	0,730	(0,092)
x_2	0,877	(0,250)	0,917	(0,271)	0,545	(0,179)	0,996	(0,239)
x_3	-0,092	(0,107)	-0,095	(0,116)	-0,074	(0,077)	-0,102	(0,103)

Tabela 4.13. Resíduos (e_i) e pesos (w_i) referentes aos ajustes sem a observação 1.

Nº Obs.	MQ (sem obs. 1,3,4 e 21)		HUSI		HUSC		TUSC		J & S	
	e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i	e_i	w_i
1	6,22	0	5,40	0	5,14	0	5,73	0	4,85	0
3	6,43	0	5,83	0,49	5,75	0,57	6,01	0,14	5,60	0,72
4	8,17	0	7,17	0,40	7,04	0,46	8,09	0	6,77	0,59
21	-8,63	0	-7,95	0,36	-7,85	0,42	-9,05	0	-7,68	0,52

5. CONCLUSÕES E SUGESTÕES PARA FUTUROS TRABALHOS

Os resultados do estudo de Monte Carlo e exemplo sugerem o uso de métodos M em aplicações práticas onde a matriz de especificação X está sob controle. Em particular indicamos sua utilização em situações onde nenhum método de detecção de dados discrepantes é empregado ou, ainda, em grandes conjuntos de dados onde se tem informação da existência de considerável proporção de erros grosseiros.

Recomendamos a seguinte estratégia para escolha dos estimadores robustos em aplicações práticas de análise de regressão: na presença de pequena proporção de erros grosseiros onde há suspeita de que a distribuição dos erros difere apenas moderadamente da normal pode-se utilizar qualquer dos estimadores HUSI, HUSC, TUSC e J&S; em situações onde se sabe a priori que há grande proporção de erros grosseiros indicamos a utilização do estimador TUSC. Sugerimos que se proceda também, sempre que possível, à estimação pelo método dos mínimos quadrados para comparação dos resultados e obtenção de familiaridade com o métodos robustos.

Acreditamos que a reduzida utilização de métodos robustos em análise de regressão se deve em grande parte à escassa oferta de pacotes de computador à disposição dos usuários. Nesse sentido, sugerimos o desenvolvimento de procedimentos que capacitem os pacotes estatísticos de ampla utilização (tais como SOC e SAS) a calcular os estimadores M.

Sugerimos também que sejam desenvolvidos trabalhos de divulgação e avaliação em situações práticas de métodos M de seleção de variáveis, estimadores de influência limitada, estimadores MM e τ .

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDREWS, D. F. (1974). A Robust Method for Multiple Linear Regression. Technometrics 16: 523-531.
2. ANDREWS, D. F. et alii (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton, N. J.
3. ATKINSON, A. C. (1981). Two Graphical Displays for Outlying and Influential Observations in Regression. Biometrika 68: 13-20.
4. BARRODALE, I. & ROBERTS, F. D. K. (1974). Algorithm 478: Solution of an Overdetermined System of Equations in the L_1 Norm. Comm. of the ACM 17: 319-320.
5. BEATON, A. E. & TUKEY, J. W. (1974). The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data. Technometrics 16: 147 - 185.
6. BELSLEY, D. A. et alii (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley, New York.
7. BICKEL, P. J. (1973). On Some Analogues to Linear Combinations of Order Statistics in the Linear Model. Ann Statist. 1: 597 - 616.
8. BICKEL, P. J. (1975). One-step Huber Estimates in the Linear Model. J. Am. Statist. Assoc. 70: 428 - 434.
9. BOX, G. E. P. & MULLER, M. E. (1958). A Note on the Generation of Normal Deviates. Ann Math. Statist. 29: 610-611.
10. BROWNLIE, K. A. (1965). Statistical Theory and Methodology in Science and Engineering. 2. ed. John Wiley, New York.

11. CARNAHAN, B. et alii (1969). Applied Numerical Methods. John Wiley, New York.
12. DANIEL, C. & WOOD, F. S. ((1971) 1980). Fitting Equations to Data. John Wiley, New York.
13. DIELMAN, T. E. (1984). Least Absolute Value Estimation in Regression Models: an Annotated Bibliography. Commun Statist. - Theor. Meth. 13: 513-541.
14. DRAPER, N. R. & SMITH, H. (1981). Applied Regression Analysis. 2. ed. John Wiley, New York.
15. DUTTER, R. (1976). Computer Linear Robust Curve Fitting Program LINWDR. Research Report 10. Fachgruppe für Statistik, ETH, Zurich.
16. DUTTER, R. (1983). Computer Program BLINWDR for Robust and Bounded Influence Regression. Research Report 8. Institut für Statistik, Technische Universität, Graz, Austria.
17. FERGUSON, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.
18. HAMPEL, F. R. (1968). Contribution to the Theory of Robust Estimation. Ph. D. Thesis. University of California, Berkeley.
19. HAMPEL, F. R. (1974). The Influence Curve and its Role in Robust Estimation. J. Am. Statist. Assoc. 69: 383-393.
20. HAMPEL et alii (1986). Robust Statistics: The Approach Based on Influence Functions. John Wiley, New York.
21. HILL, R. W. (1979). On Estimating the Covariance Matrix of Robust Regression M-Estimates. Commun. Statist. Theor. Meth. 8: 1183-1196.

22. HODGES, J.L. Jr. & LEHMANN, E.L. (1963). Estimates of Location Based on Rank Tests. Ann. Math. Statist. 34: 598-611.
23. HOGG, R. V. (1979). Statistical Robustness: One View of its Use in Applications Today. Am. Statist. 33: 108-115.
24. HOLLAND, P. W. & WELSCH, R. E. (1977). Robust Regression Using Iteratively Reweighted Least-Squares. Commun. Statist. Theor. Meth. 6: 813-827.
25. HUBER, P. J. (1964). Robust Estimation of a Location Parameter. Ann. Math. Statist. 35: 73-101.
26. HUBER, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. Ann. Statist 1: 799-821.
27. HUBER, P. J. (1977). Robust Methods of Estimation of Regression Coefficients. Math. Operationsforschung Statist. Ser. Statis.. 8: 41-53.
28. HUBER, P. J. (1981). Robust Statistics. John Wiley, New York.
29. HUBER, P. J. (1983). Minimax Aspects of Bounded-influence Regression (with discussion). J. Am. Statist. Assoc. 78: 66-80.
30. IMSL Library, Edition 7 (1979), IMSL, Inc., Houston.
31. JAECKEL, L. A. (1972). Estimating Regression Coefficients by Minimizing the Dispersion of Residuals. Ann. Math. Statist. 43: 1449-1458.
32. JERISON, H. J. (1973). Evolution of the Brain and Intelligence. Academic Press, New York.
33. JUREČKOVÁ, J. (1971). Nonparametric Estimate of Regression Coefficients. Ann. Math. Statist. 42: 1328-1338.
34. JUREČKOVÁ, J. & SEN, P. K. (1984). On Adaptive Scale-equivariant M-Estimators in Linear Models. Statistics & Decisions, Supplement Issue 1: 31-46.

35. KOENKER, R. & BASSETT, G. Jr. (1978). Regression Quantiles. Econometrica 46: 33-50.
36. KRASKER, W. S. & WELSCH, R. E. (1982). Efficient Bounded-Influence Regression Estimation. J. Am. Statis. Assoc. 77: 595-604.
37. LAWSON, C. L. & HANSON, R.J. (1974). Solving Least Squares Problems. Prentice Hall, Englewood Cliffs, N.J.
38. LEE, W. & GENTLE, J. (1986). The LAV Procedure. In: SUGI Supplementary Libray User's Guide, Version 5 Edition. SAS Institute Inc., Cary, NC.
39. LEROY, A. & ROUSSEEUW, P. J. (1984). PROGRESS: A Program for Robust Regression. Technical Report 201. Free University, Brussels, Bélgica.
40. MARAZZI, A. (1980a). ROBETH, A Subroutine Library for Robust Statistical Procedures. COMPSTAT 1980, Proceedings in Computational Statistics. Physica, Viena.
41. MARAZZI, A. (1980b). Robust Linear Regression Programs in ROBETH. ROBETH Document Nº 2. Research Report 23. Fachgruppe für Statistik, ETH, Zurich.
42. NARULA, S. C. & WELLINGTON, J. F. (1982). The Minimum Sum of Absolute Errors Regression: A State of the Art Survey. International Statistical Review 50: 317-326.
43. PETERS, S. C. et alii (1982). Computational Procedures for Bounded-Influence and Robust Regression (TROLL: BIF and BIFMOD). Technical Report 30. Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, Mass.
44. ROUSSEEUW, P. J. (1984). Least Median of Squares Regression. J. Am. Statist. Assoc. 79: 871-880.

45. ROUSSEEUW, P. J. & LEROY, A. M. (1987). Robust Regression and Outlier Detection. Joh Wiley, New York.
46. ROUSSEEUW, P. J. & YOHAI, V. (1984). Robust Regression by Means of S-estimators. In: Robust and Nonlinear Time Series Analysis, J. Franke et alii (eds.), Lecture Notes in Statistics 26, Springer, New York.
47. RUPPERT, D. (1985). On the Bounded-Influence Regression Estimator of Krasker and Welsch. J. Am. Statist. Assoc. 80: 205-208.
48. SAS User's Guide, 1985 Version Edition 5. SAS Institute Inc., Cary, NC.
49. STUDENT (1927). Errors of Routine Analysis. Biometrika 19: 151-164.
50. WEISBERG, S. (1980). Applied Linear Regression. John Wiley, New York.
51. WELLINGTON, J. F. & NARULA, S. C. (1984). An Algorithm for Regression Quantiles. Commun. Statist. - Simula. Computa. 13: 683-704.
52. YOHAI, V. J. (1987). High Breakdown-point and High Efficiency Robust Estimates for Regression. Ann. Statist. 15: 642-656.
53. YOHAI, V. J. & MARONNA, R. A. (1979). Asymptotic Behavior of M-Estimators for the Linear Model. Ann. Statist. 7: 258-268.
54. YOHAI, V. J. & ZAMAR, R. H. (1988). High Breakdown-point estimates of Regression by Means of the Minimization of an Efficient Scale. J. Am. Statist. Assoc. 83: 406-413.

APÊNDICE A

Tabela A.1. Resultados da simulação com erros seguindo distribuição normal e tamanho de amostra = 20, semente = 91675

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{Var}(\hat{\beta}_0)$	$\widehat{Var}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\widehat{Var}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,005	1,001	0,0507	0,0501	1,0101	0,1414	5 [2, 17]	128	0,1008	97	97
HUSC	1,004	1,001	0,0524	0,0516	1,0477	0,1887	5 [1, 12]	112	0,1040	94	94
TUSC	1,005	1,001	0,0545	0,0552	1,0683	0,1661	5 [3, 21]	169	0,1097	89	89
J&S	1,005	1,001	0,0517	0,0508	1,0547	0,1750	5 [2, 16]	328	0,1025	95	95
MQ	1,005	1,001	0,0491	0,0485	0,9746	0,0262	-		0,0976	100	100
L ₁	1,004	0,993	0,0731	0,0772	0,8840	0,0575	4 [2, 7]		0,1503	65	65

Tabela A.2. Resultados da simulação com erros seguindo distribuição normal contaminada (5%N(0,9) + 95% N(0,1)) e tamanho de amostra = 20, semente = 13689

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\overline{f_c \sigma^2}$	$\widehat{Var(f_c \sigma^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,002	1,000	0,0438	0,0419	0,8615	0,1265	6 [3, 26]	140	0,0857	119	100
HUSC	1,003	1,000	0,0446	0,0421	0,8723	0,1451	5 [1, 16]	119	0,0867	117	99
TUSC	1,003	0,999	0,0453	0,0436	0,8825	0,1330	5 [3, 21]	171	0,0889	114	96
J&S	1,002	1,001	0,0445	0,0421	0,8886	0,1742	5 [2, 11]	334	0,0866	117	99
MQ	1,002	1,001	0,0534	0,0483	0,9733	0,0655	-	-	0,1017	100	84
L ₁	1,000	0,996	0,0598	0,0609	0,7984	0,0504	4 [2, 7]		0,1207	84	71

Tabela A.3. Resultados da simulação com erros seguindo distribuição normal contaminada (10%N(0,9) + 90%N(0,1)) e tamanho de amostra = 20, semente = 479193.

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{\text{Var}}(\hat{\beta}_0)$	$\widehat{\text{Var}}(\hat{\beta}_1)$	$\overline{f_c} \hat{\sigma}^2$	$\widehat{\text{Var}}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,999	1,000	0,0368	0,0399	0,7630	0,1250	6 [3, 25]	143	0,0767	130	100
HUSC	0,999	0,999	0,0367	0,0400	0,7586	0,1168	5 [1, 14]	119	0,0767	130	100
TUSC	1,000	0,997	0,0374	0,0405	0,7682	0,1258	6 [2, 21]	170	0,0779	128	98
J&S	0,999	0,999	0,0369	0,0403	0,7742	0,1264	5 [2, 15]	330	0,0772	129	99
MQ	0,999	1,004	0,0484	0,0515	0,9564	0,0757	-	-	0,0999	100	77
L ₁	1,003	0,989	0,0464	0,0520	0,7339	0,0412	4 [2, 8]		0,0984	102	78

Tabela A.4. Resultados da simulação com erros seguindo distribuição normal contaminada (20%N(0,9) + 80%N(0,9)) e tamanho de amostra = 20, semente = 1500.

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\text{Var}(\hat{\beta}_0)$	$\text{Var}(\hat{\beta}_1)$	$\overline{f_c} \hat{\sigma}^2$	$\text{Var}(f_c \hat{\sigma}^2)$	NÚMERO LTERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,000	1,004	0,0350	0,0372	0,6937	0,1336	7 [3, 28]	153	0,0722	137	97
HUSC	1,000	1,004	0,0337	0,0366	0,6802	0,1223	5 [1, 13]	119	0,0703	141	100
TUSC	0,998	1,002	0,0345	0,0362	0,6780	0,1356	6 [3, 25]	169	0,0707	140	99
J&S	1,000	1,005	0,0353	0,0372	0,7049	0,1646	5 [2, 13]	327	0,0725	137	97
MO	1,001	1,003	0,0480	0,0511	0,9543	0,0773	-	-	0,0991	100	71
L_1	0,999	1,000	0,0423	0,0480	0,6718	0,0394	4 [2, 8]	-	0,0903	110	78

Tabela A.5. Resultados da simulação com erros seguindo distribuição normal contaminada (5%N(0,100) + 95%N(0,1)) e tamanho de amostra = 20
semente = 31116

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\widehat{f_c \sigma^2}$	$\widehat{Var(f_c \sigma^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,006	0,992	0,0111	0,0115	0,2140	0,0102	7 [2, 23]	152	0,0226	428	92
HUSC	0,996	0,998	0,0115	0,0124	0,2220	0,0128	5 [1, 16]	117	0,0239	405	87
TUSC	0,996	1,000	0,0103	0,0106	0,1993	0,0076	5 [3, 27]	164	0,0209	463	100
J&S	0,995	0,998	0,0125	0,0128	0,2303	0,0332	5 [2, 11]	330	0,0253	382	83
MQ	0,989	0,990	0,0478	0,0487	0,7929	0,2890	-		0,0967	100	22
L ₁	0,996	0,995	0,0149	0,0162	0,3945	0,0122	5 [2, 8]		0,0311	311	67

Tabela A.6. Resultados da simulação com erros seguindo distribuição SLASH e tamanho de amostra = 20, semente = 934.517.

Nº DE REPETIÇÕES = 1500

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\text{Var}(\hat{\beta}_0)$	$\text{Var}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\text{Var}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,992	1,002	0,5026	0,5432	10,88	100,60	12 [3, 60]	183	1,0459	$3,5 \times 10^6$	65
HUSC	1,323	1,359	186,76	311,66	3972,25	20.197.425.152	5 [1, 15]	123	498,65	$7,4 \times 10^3$	0
TUSC	0,985	1,011	0,3319	0,3490	7,35	33,56	6 [3, 36]	166	0,6812	$5,4 \times 10^6$	100
J&S	0,970	1,007	1,4700	0,9365	17,41	6181,14	5 [3, 13]	331	2,4074	$1,5 \times 10^6$	28
MQ	4,50	3,69	14.706,41	22.198,02	43,2	283130,41	-	-	36923,92	100	0
L ₁	0,995	1,013	0,3634	0,3985	2,18	0,5912	3 [2, 6]	-	0,7621	$4,8 \times 10^6$	89

Figura A.7. Resultados da simulação com erros seguindo distribuição normal e tamanho de amostra = 40; semente = 31.517.322.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\overline{f_c} \hat{\sigma}^2$	$\widehat{Var(f_c \hat{\sigma}^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,008	1,001	0,0259	0,0257	1,0347	0,0730	5 [3, 14]	144	0,0516	93	93
HUSC	1,008	1,001	0,0264	0,0260	1,0525	0,0854	4 [1, 9]	121	0,0524	92	92
TUSC	1,009	1,001	0,0271	0,0268	1,0635	0,0709	5 [3, 12]	173	0,0539	89	89
J&S	1,008	1,001	0,0262	0,0261	1,1099	0,0902	5 [3, 9]	419	0,0523	92	92
MQ	1,007	1,001	0,0238	0,0242	0,9890	0,0134	-	-	0,0480	100	100
L ₁	1,017	0,995	0,0372	0,0412	1,0638	0,0709	4 [2, 9]	-	0,0784	61	61

Tabela A.8. Resultados da simulação com erros seguindo distribuição normal contaminada (5%N(0,9) + 95%N(0,1)) e tamanho de amostra = 40, semente = 1.215.201.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\hat{Var}(\hat{\beta}_0)$	$\hat{Var}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\hat{Var}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,002	1,005	0,0250	0,0214	0,8377	0,0531	6 [3, 16]	157	0,0464	117	100
HUSC	1,003	1,005	0,0252	0,0216	0,8470	0,0567	5 [2, 10]	128	0,0468	116	99
TUSC	1,003	1,007	0,0252	0,0219	0,8461	0,0514	5 [2, 12]	176	0,0471	115	99
J&S	1,003	1,005	0,0251	0,0214	0,8945	0,0625	5 [3, 9]	431	0,0465	117	100
MQ	1,001	1,001	0,0292	0,0252	0,9875	0,0361	-	-	0,0544	100	85
L ₁	1,007	1,002	0,0347	0,0315	0,8426	0,0266	4 [2, 9]	-	0,0662	82	70

Tabela A.9. Resultados da simulação com erros seguindo distribuição normal contaminada (10%N(0,9) + 90%N(0,1)) e tamanho de amostra = 40, semente = 17.953.190

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{\text{Var}}(\hat{\beta}_0)$	$\widehat{\text{Var}}(\hat{\beta}_1)$	$\overline{f_c} \hat{\sigma}^2$	$\widehat{\text{Var}}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,000	1,000	0,0179	0,0178	0,7398	0,0517	7 [3, 14]	169	0,0357	136	98
HUSC	1,000	1,000	0,0177	0,0179	0,7420	0,0513	5 [1, 9]	130	0,0356	136	99
TUSC	1,000	0,999	0,0171	0,0180	0,7358	0,0522	5 [3, 11]	179	0,0351	138	100
J&S	1,000	1,000	0,0179	0,0179	0,7836	0,0614	5 [3, 8]	435	0,0358	135	98
MQ	0,999	1,001	0,0254	0,0230	0,9739	0,0399	-	-	0,0484	100	73
L ₁	1,001	1,004	0,0234	0,0252	0,7666	0,0224	4 [2, 9]	-	0,0486	100	72

Tabela A.10. Resultados da simulação com erros seguindo distribuição normal contaminada (20%N(0,9) + 80%N(0,1)) e tamanho de amostra = 40, semente = 311261

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\widehat{f_c \hat{\sigma}^2}$	$\widehat{Var(f_c \hat{\sigma}^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,005	1,001	0,0179	0,0153	0,6612	0,0514	8 [3, 17]	187	0,0332	148	98
HUSC	1,005	1,001	0,0175	0,0152	0,06493	0,0476	5 [3, 9]	134	0,0327	150	100
HUST	1,004	1,003	0,0174	0,0153	0,6434	0,0527	6 [3, 14]	181	0,0327	150	100
J&S	1,005	1,001	0,0180	0,0153	0,7032	0,0691	5 [3, 8]	438	0,0333	147	98
MQ	1,005	0,999	0,0262	0,0228	0,9826	0,0403	-	-	0,0490	100	67
L ₁	1,000	0,996	0,0222	0,0197	0,6959	0,0188	4 [2, 8]	-	0,0419	117	78

Tabela A.11. Resultados da simulação com erros seguindo distribuição normal contaminada (5%N(0,100) + 95%N(0,1)) e tamanho de amostra = 40, semente = 519.171.710

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\overline{f_c \hat{\sigma}^2}$	$\widehat{Var(f_c \hat{\sigma}^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
JUSI	1,001	0,998	0,0045	0,0052	0,2111	0,0041	8 [3, 20]	187	0,0097	504	95
HUSC	1,001	0,998	0,0046	0,0053	0,2138	0,0045	5 [3, 9]	137	0,0099	494	93
HUST	1,001	0,997	0,0043	0,0049	0,1945	0,0030	5 [3, 19]	179	0,0092	532	100
J&S	1,001	0,998	0,0046	0,0053	0,2251	0,0052	5 [3, 11]	464	0,0099	494	93
MQ	1,004	0,997	0,0238	0,0251	0,9068	0,2115	-	-	0,0489	100	19
L ₁	0,999	0,996	0,0065	0,0075	0,4149	0,0065	4 [2, 9]	-	0,0140	349	66

Tabela A.12. Resultados da simulação com erros seguindo distribuição SLASH e tamanho de amostra = 40, semente = 81.297.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\widehat{E(\beta_0)}$	$\widehat{E(\beta_1)}$	$\widehat{Var(\beta_0)}$	$\widehat{Var(\beta_1)}$	$\overline{f_c} \hat{\sigma}^2$	$\widehat{Var(f_c \hat{\sigma}^2)}$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,994	1,013	0,2407	0,2431	9,1581	26,4197	14 [4, 31]	294	0,4840	$6,5 \times 10^5$	68
HUSC	1,049	1,086	3,2631	4,3557	130,724	1.511.195	5 [3, 12]	137	7,6286	$4,1 \times 10^4$	4
TUST	0,978	1,011	0,1679	0,1603	6,5848	10,8447	6 [3, 13]	177	0,3288	$9,5 \times 10^5$	100
J&S	1,014	1,016	0,3629	0,3417	12,3631	153,570	5 [3, 9]	447	0,7051	$4,4 \times 10^5$	47
MQ	2,29	1,88	1921,07	1208,26	56,88	79045,32	-	-	3.131,77	100	0
L ₁	0,998	1,0130	0,1777	0,1865	2,1915	0,2555	4 [2, 7]	-	0,3651	$8,6 \times 10^5$	90

Tabela A.13. Resultados da simulação com erros seguindo distribuição normal e tamanho de amostra = 100, semente = 9.056.713.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\text{Var}(\hat{\beta}_0)$	$\text{Var}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\text{Var}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,996	1,001	0,0104	0,0104	1,0577	0,0312	5 [3, 10]	303	0,0208	95	95
HUSC	0,996	1,001	0,0105	0,0104	1,0619	0,0324	4 [2, 7]	258	0,0209	94	94
TUSC	0,996	1,001	0,0106	0,0107	1,0686	0,0277	5 [3, 8]	371	0,0213	92	92
J&S	0,996	1,001	0,0106	0,0105	1,0569	0,0353	5 [3, 6]	1267	0,0211	93	93
MQ	0,997	1,001	0,0098	0,0099	1,0008	0,0051	-	-	0,0197	100	100
L_1	0,992	1,003	0,0153	0,0154	0,9841	0,0150	5 [2, 9]	-	0,0307	64	64

Tabela A.14. Resultados da simulação com erros seguindo distribuição normal contaminada (5%N(0,9) + 95%N(0,1)) e tamanho de amostra = 100, semente = 497.321.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{\text{Var}}(\hat{\beta}_0)$	$\widehat{\text{Var}}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\widehat{\text{Var}}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,001	0,997	0,008250	0,008314	0,8367	0,0215	6 [3, 12]	363	0,016574	125	98
HUSC	1,001	0,997	0,008265	0,008354	0,8394	0,0221	5 [3, 7]	282	0,016629	124	98
TUSC	1,002	0,997	0,007978	0,008264	0,8328	0,0197	5 [3, 8]	397	0,016251	127	100
J&S	1,001	0,997	0,008278	0,008324	0,8312	0,0227	5 [3, 7]	1299	0,016612	124	98
MQ	1,001	0,998	0,01026	0,01041	0,9912	0,0165	-	-	0,02067	100	79
L_1	1,001	0,995	0,01153	0,01187	0,8657	0,0108	5 [2, 9]	-	0,02343	88	69

Tabela A.15. Resultados da simulação com erros seguindo distribuição normal contaminada (10%N(0,9) + 90%N(0,1)) e tamanho da amostra = 100, Semente = 1.975.493.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\hat{Var}(\hat{\beta}_0)$	$\hat{Var}(\hat{\beta}_1)$	$\hat{f}_c \hat{\sigma}^2$	$\hat{Var}(\hat{f}_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,996	0,999	0,007325	0,007209	0,7262	0,0182	7 [3, 13]	395	0,014551	136	99
HUSC	0,996	0,999	0,007351	0,007248	0,7274	0,0185	5 [3, 7]	296	0,014616	136	98
TUSC	0,996	0,999	0,007203	0,007170	0,7161	0,0172	5 [3, 9]	401	0,014390	138	100
J&S	0,996	0,999	0,007363	0,007241	0,7177	0,0179	5 [3, 6]	1293	0,014621	136	98
MQ	0,998	0,999	0,01002	0,00979	0,9868	0,0174	-	-	0,019815	100	73
L_1	0,997	0,996	0,01046	0,01019	0,7919	0,0088	5 [2, 10]	-	0,02068	96	70

Tabela A.16. Resultados da simulação com erros seguindo distribuição normal contaminada (20%N(0,9) + 80%N(0,1)) e tamanho da amostra = 100, semente = 839.510.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{\text{Var}}(\hat{\beta}_0)$	$\widehat{\text{Var}}(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$\widehat{\text{Var}}(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	0,999	0,997	0,006198	0,006158	0,6410	0,0177	8 [4, 13]	438	0,012366	158	98
HUSC	0,999	0,997	0,006108	0,006113	0,6365	0,0170	5 [3, 7]	297	0,012231	160	99
HUST	0,999	0,997	0,006084	0,006040	0,6266	0,0187	5 [3, 8]	404	0,012134	161	100
J&S	0,998	0,997	0,006189	0,006149	0,6247	0,0169	5 [3, 8]	1291	0,012351	158	98
MQ	0,998	0,996	0,00970	0,00983	0,9915	0,01554	-	-	0,01955	100	62
L_1	0,996	0,997	0,008093	0,007903	0,7195	0,0084	5 [2, 10]	-	0,016021	122	76

Tabela A.17. Resultados da simulação com erros seguindo distribuição normal contaminada ($5\%N(0,100) + 95\%N(0,1)$) e tamanho de amostra = 100, semente = 519.443.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$\hat{E}(\hat{\beta}_0)$	$\hat{E}(\hat{\beta}_1)$	$\widehat{\text{Var}}(\hat{\beta}_0)$	$\widehat{\text{Var}}(\hat{\beta}_1)$	$\overline{f_c} \hat{\sigma}^2$	$\widehat{\text{Var}}(\overline{f_c} \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,002	1,000	0,002073	0,001997	0,2086	0,0014	8 [4, 13]	431	0,004074	456	96
HUSC	1,002	1,000	0,002078	0,002004	0,2097	0,0015	5 [3, 8]	300	0,004086	455	96
TUSC	1,001	1,000	0,001989	0,001922	0,1915	0,0010	5 [3, 7]	389	0,003912	475	100
J&S	1,002	1,000	0,002090	0,002021	0,2053	0,0014	5 [3, 8]	1391	0,004115	451	95
MQ	1,001	0,998	0,009545	0,009027	0,9602	0,0898	-	-	0,018577	100	21
L _I	1,000	1,000	0,003014	0,002692	0,1915	0,0010	5 [2, 9]	-	0,005706	326	69

Tabela A.18. Resultados da simulação com erros seguindo distribuição SLASH e tamanho de amostra = 100, semente = 1.715.111.

Nº DE REPETIÇÕES = 1000

ESTIMADOR	$E(\hat{\beta}_0)$	$E(\hat{\beta}_1)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$\overline{f_c \hat{\sigma}^2}$	$Var(f_c \hat{\sigma}^2)$	NÚMERO ITERAÇÕES MEDIANO [min; máx]	TEMPO TOTAL (SEG)	EQM TOTAL	EFICIÊNCIA (%) EM RELAÇÃO A	
										M Q	MENOR EQM TOTAL
HUSI	1,002	0,993	0,072397	0,082341	7,9475	5,8096	14 [9, 24]		0,154791	$8,9 \times 10^6$	75
HUSC	1,022	0,914	2,8895	5,3986	286,51	29.945.552	5 [3, 9]		8,2960	$1,7 \times 10^5$	1
TUSC	0,998	0,995	0,056388	0,059852	5,9973	3,0682	6 [3, 10]		0,116269	$1,2 \times 10^7$	100
J&S	1,002	0,992	0,082907	0,093955	8,4104	9,8277	6 [4, 9]		0,176930	$7,8 \times 10^6$	66
MQ	1,11	-1,91	5181,12	8626,69	78,5	508.226,41	-		13.816,29	100	0
L_1	0,992	0,995	0,060639	0,067439	21,820	0,097095	4 [2, 9]		0,128167	$1,1 \times 10^7$	91