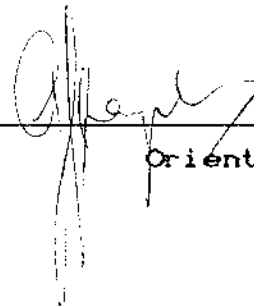


MÉTODOS PARA ANÁLISE DE REGRESSÃO L₁

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Manoel Ivanildo Silvestre Bezerra e aprovada pela Comissão Julgadora.

Campinas, 31 de agosto de 1990

Prof.^a Dr.^a



Orientadora

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística

UNIVERSIDADE ESTADUAL DE CAMPINAS
- U N I C A M P -

INSTITUTO DE MATEMATICA, ESTATISTICA E CIENCIA DA COMPUTAÇÃO
- I M E C C -

MÉTODOS PARA ANÁLISE DE REGRESSÃO L_1

MANOEL IVANILDO SILVESTRE BEZERRA

ORIENTADORA: PROF.^a DR.^a GABRIELA STANGENHAUS

CAMPINAS - 1990

A G R A D E C I M E N T O S

À Profa. Dra. Gabriela Stangenhans pela orientação do trabalho.

Ao amigo José Ramos pelo auxílio na parte computacional.

Aos colegas do IMECC e alguns professores pela amizade durante este tempo.

À UNICAMP, CAPES e CNPQ pelo apoio financeiro.

Dedico este trabalho aos meus Pais

I N D I C E

CAPITULO 1 - INTRODUÇÃO	1
CAPITULO 2 - REGRESSÃO QUANTILICA	5
2.1 - INTRODUÇÃO	5
2.2 - PROPRIEDADES	7
2.3 - DISTRIBUIÇÃO ASSINTÓTICA DOS ESTIMADORES DA REGRES SÃO QUANTILICA	13
2.4 - REGRESSÃO DA MSEA (MÍNIMA SOMA DOS ERROS ABSOLUTOS) OU REGRESSÃO L_1	16
CAPITULO 3 - INFERÊNCIA	19
3.1 - INTRODUÇÃO	19
3.2 - INTERVALO DE CONFIANÇA	20
3.3 - TESTE DE HIPÓTESE	25
3.4 - ANÁLISE DE VARIÇÃO	26
3.4.1 - TESTE PARA A HIPÓTESE LINEAR GERAL	27
3.4.2 - EXEMPLO PARA CONSTRUÇÃO DA TABELA DE ANÁLISE DE VA RIAÇÃO	30
CAPITULO 4 - SELEÇÃO DE VARIÁVEIS	32
4.1 - INTRODUÇÃO	32
4.2 - ALGORITMO DO NARULA E WELLINGTON (1979) PARA ENCON TRAR O MELHOR SUBCONJUNTO DE VARIÁVEIS PREDITORAS EM REGRESSÃO L_1	33
4.2.1 - SISTEMA DE ROTULAÇÃO	36
4.3 - SUGESTÕES PARA ACELERAR O ALGORITMO DE ENUMERAÇÃO IMPLÍCITA	36
4.3.1 - MODELOS "QUASE-ÓTIMOS" DE TAMANHO m	36
4.3.2 - MODELOS "IGUALMENTE-BONS"	37
4.3.3 - LIMITE PARA O NÚMERO DE VARIÁVEIS PREDITORAS	38

CAPITULO 5 - ROBUSTEZ	39
5.1 - INTRODUÇÃO	39
5.2 - ANÁLISE DE SENSIBILIDADE	40
5.2.1 - REGRESSÃO LINEAR SIMPLES	41
5.2.2 - ALGORITMO DE KARST	42
5.2.3 - INTERVALO PARA VARIÁVEL RESPOSTA Y	43
5.2.4 - INTERVALO PARA UM VALOR DA VARIÁVEL PREDITORA X ..	44
5.3 - ANÁLISE DE SENSIBILIDADE PARA O MODELO DE REGRESSÃO	
MÚLTIPLA	46
5.4 - ANÁLISE DE OBSERVAÇÕES	50
 CAPITULO 6 - ANÁLISE DE RESÍDUOS	 52
6.1 - INTRODUÇÃO	52
6.2 - HETEROSCEDASTICIDADE	53
6.3 - VALORES ABERRANTES E VALORES INFLUENTES	54
6.3.1 - METODO PARA DETECTAR VALORES ABERRANTES ATRAVÉS DE	
REGRESSÕES QUANTILICAS	55
6.3.2 - ALGORITMO	56
6.4 - VERIFICAÇÃO DO MODELO	57
6.4.1 - USO DO GRÁFICO DE PROBABILIDADE NORMAL PARA DETEC	
TAR DISTRIBUIÇÕES DE ERROS COM CAUDAS ALONGADAS ..	60
 CAPITULO 7 - APLICAÇÃO	 64
7.1 - INTRODUÇÃO	64
7.2 - CONJUNTO DE DADOS	64
7.3 - ANÁLISE	66
7.4 - CONCLUSÕES	74
 BIBLIOGRAFIA	 96

CAPÍTULO 1

INTRODUÇÃO

1 - INTRODUÇÃO

Na vasta literatura estatística, podemos encontrar diversos livros e artigos sobre regressão linear usando o método dos mínimos quadrados, e diversos programas para computadores estão disponíveis.

A regressão de mínimos quadrados será ótima, e oferecerá estimadores de máxima verossimilhança para os parâmetros do modelo, se os erros são independentes e possuem distribuição normal, porém para diversos casos em que os erros não tem distribuição normal, ou possuem caudas alongadas, como por exemplo as distribuições de Cauchy e Laplace, a regressão de mínimos quadrados não é adequada.

Um outro problema é o caso de valores aberrantes, quando usamos a regressão de mínimos quadrados, mesmo que a distribuição dos erros seja normal, será preciso usar algum procedimento robusto, ou seja, que acomoda a possibilidade desses valores, ou técnicas de diagnósticos que identificam esses valores.

A regressão da Mínima Soma dos Erros Absolutos (MSEA), ou regressão L_1 , é considerada adequada para casos em que a distribuição dos erros tem caudas alongadas. Vários autores, consideram-na uma boa alternativa robusta para a regressão de mínimos quadrados. A robustez ocorre pelo fato da regressão L_1 ser menos sensível a valores aberrantes do que a regressão de mínimos quadrados.

A regressão L_1 tem inúmeras aplicações, como por exemplo: (1) estimação de funções e previsão de investimento; (2) detecção de erros em conjunto de dados, a fim de tentar descobrir formações na superfície, onde recursos naturais podem existir; (3) processamento de dados orbitais em objetos espaciais; (4) astronomia; (5) modelagem de diversos dados geofísicos; (6) estimação de funções de custos; (7) análise de dados sísmicos; (8) estimação de parâmetros farmacocinéticos e (9) obtenção de estimativas de estados em sistemas de potência.

Se os erros tem distribuição de Laplace, os estimadores dos parâmetros da regressão L_1 são estimadores de máxima verossimilhança, portanto são assintoticamente não-tendenciosos e eficientes.

A regressão L_1 foi introduzida no século XVIII por Boscovich (1757), ou seja, meio século antes da regressão de mínimos quadrados proposta por Legendre, mas seu desenvolvimento foi muito lento, devido ao fato que a regressão de mínimos quadrados é computacionalmente mais simples. Sendo assim, o seu uso tornou-se praticamente impossível por muitos anos.

Boscovich (1760) também desenvolveu um algoritmo geométrico para resolver o problema. A formulação algébrica do algoritmo foi feita por Laplace (1772).

A regressão L_1 reapareceu na década de 1880 devido ao trabalho de Edgeworth, que em 1887 reformulou o problema de regressão L_1 proposto por Boscovich (1757), e em 1888 discutiu sobre a não unicidade da solução da regressão L_1 , como também contribuiu na demonstração que os estimadores L_1 são de máxima verossimilhança, quando os erros tem distribuição de Laplace, e em um algoritmo para determinar a regressão L_1 .

Charnes, Cooper e Ferguson (1955) formularam o problema de regressão L_1 , como um problema de programação linear, e solucionaram-no usando o método simplex. Daí então foram desenvolvidos vários algoritmos eficientes. O problema de regressão L_1 pode ser resolvido no computador usando o sistema de programas estatísticos SAS e o conjunto de rotinas IMSL. Ambos

procedimentos para resolver o problema de regressão L_1 .

A falta de ferramenta para inferência estatística foi também outro problema para a regressão L_1 , se comparado com a regressão de mínimos quadrados. Porém nos últimos dez anos diversos estudos de Monte Carlo foram desenvolvidos sobre propriedades dos estimadores para pequenas amostras. Resultados assintóticos sobre a distribuição dos estimadores e procedimentos para inferência também foram apresentados.

A literatura sobre regressão L_1 encontra-se espalhada em diversos artigos, relatórios, monografias e em poucos livros. Dentre os livros poderíamos citar: Bloomfield e Steiger (1983), que trata sobre métodos de regressão L_1 em vários aspectos, porém de difícil compreensão e com linguagem e notação inadequadas. Como monografia poderíamos citar: Análise de Regressão L_1 (VIII SINAPE), mas não tem demonstrações e foi escrita no nível de graduação.

O trabalho aqui apresentado consiste em um levantamento bibliográfico de diversos artigos, com vários métodos para análise de regressão L_1 , dispersos na literatura, com intuito de facilitar o estudo daqueles estatísticos ou pesquisadores que se interessem pela área.

No capítulo 2 apresentamos a classe das regressões quantílicas e alguns teoremas em relação a esta classe, que são baseadas nos quantis amostrais e tem como caso particular a regressão L_1 .

No capítulo 3 apresentamos uma seção sobre inferência estatística, mostrando como calcular intervalos de confiança, testes de hipóteses e construir a tabela de análise de "variação" para regressão L_1 e alguns resultados assintóticos.

No capítulo 4 apresentamos um método para seleção de variáveis baseado em um algoritmo e algumas sugestões para eficiência deste algoritmo.

No capítulo 5 apresentamos uma seção sobre robustez, usando análise de sensibilidade para variável resposta e para uma variável preditora e análise de observações.

No capítulo 6 tratamos da análise de resíduos, observando o caso de heteroscedasticidade, ou seja, quando variância dos resíduos não é constante, e apresentamos um algoritmo para detectar valores aberrantes através de regressões quantílicas.

Finalmente no capítulo 7 fizemos uma aplicação em um conjunto de dados, usando os métodos acima citados.

CAPÍTULO 2

REGRESSÕES QUANTÍLICAS

2.1 - INTRODUÇÃO

A regressão linear tem sido extensivamente explorada na literatura estatística, utilizando-se a teoria dos mínimos quadrados. Porém, esse método nem sempre fornece estimativas adequadas como, por exemplo, nos casos em que a distribuição dos erros possui caudas pesadas ou existem valores aberrantes no conjunto de dados.

Pensando em resolver estes problemas, muitos autores começaram a estudar estimadores alternativos à média amostral para o modelo de posição, que fossem adequados para aqueles casos. As primeiras propostas de estimadores robustos de posição surgiram no século XVIII. A mediana, médias aparadas, combinações lineares de estatísticas de ordem e outras estatísticas foram usadas em cálculos astronômicos no século XIX.

Mosteller (1946), propôs estatísticas, chamadas "ineficientes" baseadas em quantis amostrais. Mostrou-se então que estimadores desse tipo podem ser construídos de forma a serem quase tão eficientes quanto os estimadores de máxima verossimilhança. Gastwirth (1966) e outros autores estabeleceram que existem vários estimadores deste tipo com uma boa eficiência para uma variedade de distribuições. Como por exemplo, a média ponderada dos quantis $1/3, 1/2$ e $2/3$ com pesos 0.3, 0.4 e 0.3, respectivamente tem eficiência assintótica de aproximadamente 80%

para a distribuição normal, Laplace, Logística e Cauchy. Por outro lado, a média amostral é o estimador mais eficiente para a distribuição normal, mas para a distribuição de Laplace a sua eficiência com relação à mediana é $1/2$ e para a distribuição de Cauchy sua eficiência é zero. Trabalhos recentes foram desenvolvidos sobre estimadores baseados em combinações lineares de estatísticas de ordem, como por exemplo: os L-estimadores. O L-estimador de posição mais comum é a α -média aparada, que é simplesmente a média amostral com uma proporção α removida das extremidades.

Andrews, et al.(1972) desenvolveram um estudo de Monte Carlo, que tornou-se clássico, no qual utilizou-se várias alternativas de estimadores de posição, e comparou-se suas eficiências usando várias distribuições.

Com relação aos modelos lineares, Boscovich (1757) propôs o ajuste de uma reta minimizando a soma dos desvios absolutos, como também desenvolveu um algoritmo geométrico em 1760 para resolver o problema. Em 1818, Laplace provou através de um modelo de regressão linear passando pela origem, que o estimador L_1 tem menor variância assintótica do que o estimador de mínimos quadrados. Muitos anos depois, mostrou-se que a regressão L_1 é menos sensível a valores aberrantes do que a regressão de mínimos quadrados.

Bassett e Koenker (1978) provaram que os estimadores da regressão L_1 são assintoticamente não-viciados com distribuição normal.

O trabalho apresentado por Bassett e Koenker em 1978, sobre regressões quantílicas, pode ser visto como uma tentativa de estender para o modelo linear a idéia de quantil amostral, tendo como caso particular a regressão L_1 . Os teoremas aqui apresentados já tem alguns dos seus resultados conhecidos em regressão L_1 . Como a teoria da regressão quantílica engloba a regressão L_1 , apresentaremos alguns teoremas para essa nova classe de regressões.

2.2 - PROPRIEDADES

Seja $\{y_i: i = 1, \dots, n\}$ uma amostra aleatória da variável aleatória Y , com distribuição F . O θ -ésimo quantil amostral pode ser definido como solução do problema de minimização

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

onde $0 < \theta < 1$ (2.2.1)

Seja $Y = (Y_1, \dots, Y_n)$ um vetor de variáveis aleatórias do modelo linear

$$Y_i = X_i \beta + \varepsilon_i, \quad i = 1, \dots, n \quad (2.2.2)$$

onde β é um vetor de parâmetros desconhecidos e X_i é a i -ésima linha da matriz X ($n \times k$) e os ε_i 's são os erros aleatórios. Para a amostra aleatória $\{Y_i: i = 1, \dots, n\}$, a θ -ésima regressão quantílica é qualquer solução do problema de minimização

$$\min_{\tilde{\beta} \in \mathbb{R}^k} \left[\sum_{i \in \{i: y_i \geq x_i \tilde{\beta}\}} \theta |y_i - x_i \tilde{\beta}| + \sum_{i \in \{i: y_i < x_i \tilde{\beta}\}} (1 - \theta) |y_i - x_i \tilde{\beta}| \right]$$

onde $0 < \theta < 1$ (2.2.3)

O θ é um parâmetro que pondera a regressão quantílica, ou seja, localizando-a no conjunto de dados de forma que, pelo menos $100\theta\%$ das observações encontra-se acima (ou sobre: resíduos positivos ou nulos) da regressão quantílica e pelo menos $100(1-\theta)\%$

das observações encontra-se abaixo (ou sobre: resíduos negativos ou nulos) a regressão quantílica. Quando $\theta = 1/2$ temos a regressão L_1 , que é um caso importante da regressão quantílica, tendo como medida de posição a mediana amostral. No final deste capítulo trataremos da regressão L_1 , como caso particular da regressão quantílica.

Introduziremos a seguir, as principais notações e propriedades dos elementos $\beta^*(\theta) \in B^*(\theta)$, onde $B^*(\theta)$ é o conjunto de todas as soluções possíveis do problema de minimização (2.2.3) da θ -ésima regressão quantílica.

Seja $\Gamma = \{1, \dots, n\}$ e H um conjunto de subconjuntos de dimensão k de Γ . Os elementos $h \in H$ e $\bar{h} = \Gamma - h$, ambos servem para particionar y e X . Se tivermos um vetor $y(h)$ de k elementos $\{y_i : i \in h\}$, então $X(\bar{h})$ será a matriz de dimensão $(n-k) \times k$ com linhas definidas por $\{x_i : i \in \bar{h}\}$. O vetor composto de k elementos iguais a um será representado por 1_k . Seja

$$H_1 = \{h \in H \mid \text{Posto} X(h) = k\} \quad (2.2.4)$$

ou seja, o posto $X(h \in H)$ será igual ao número de colunas da matriz X .

TEOREMA 2.1 - Se X tem posto k então o conjunto de regressões quantílicas, $B^*(\theta)$, tem no mínimo um elemento da forma

$$\beta^*(\theta) = X(h)^{-1}y(h) \quad \text{para algum } h \in H \quad (2.2.5)$$

onde $X(h)$ é uma matriz $(k \times k)$ e $\beta^*(\theta) = \check{\beta}$. Além disso, $B^*(\theta)$, é o fecho convexo de todas as soluções possíveis de (2.2.5).

PROVA: O problema de minimização (2.2.3) pode ser formulado como o seguinte problema de programação linear:

$$\min [\theta 1_n' e^+ + (1 - \theta) 1_n' e^-] \quad (P)$$

Sujeito a

$$Y = X\tilde{\beta} + \varepsilon^+ - \varepsilon^-$$
$$(\tilde{\beta}, \varepsilon^+, \varepsilon^-) \in \mathbb{R}^k \times \mathbb{R}_+^{2n},$$

onde $\mathbf{1}'_n = (1, \dots, 1)$, um vetor de n uns, e a solução deste problema é o estimador L_1 de β e $\varepsilon^+, \varepsilon^-$ são os erros positivos e negativos respectivamente. O dual de (P) é um problema de variáveis limitadas, e é definido como

$$\max [Y'd]$$

(D)

Sujeito a

$$X'd = 0$$
$$d \in [\theta - 1, \theta]^n$$

onde $[\theta - 1, \theta]^n$ representa os n -produtos cruzados do intervalo fechado $[\theta - 1, \theta]$.

Pela solução do problema (P) ou (D) através do método simplex temos que pelo menos uma solução é dada por (2.2.5) em que $X(h)^{-1}$ é a inversa da base obtida na última etapa do método simplex. Se a solução do problema (P) não é única então qualquer combinação convexa dessas soluções também é solução ótima. ■

Este teorema mostra que pelo menos uma regressão quantílica passa pelo menos por k das observações. Por exemplo os quantis amostrais coincidem com estatísticas de ordem da amostra ou, podem assumir qualquer combinação convexa de duas estatísticas de ordem adjacentes. O teorema acima generaliza essa característica para regressão quantílica, onde normais aos hiperplanos definidos por subconjuntos das k observações fazem o papel de estatísticas de ordem.

TEOREMA 2.2 - Se $\beta^*(\theta, y, X) \in B^*(\theta, y, X)$ então as expressões abaixo serão soluções dos respectivos problemas transformados:

$$(i) \beta^*(\theta, \lambda y, X) = \lambda \beta^*(\theta, y, X), \lambda \in [0, \infty)$$

$$(ii) \beta^*(1-\theta, \lambda y, X) = \lambda \beta^*(\theta, y, X), \lambda \in (-\infty, 0]$$

$$(iii) \beta^*(\theta, y + X\gamma, X) = \beta^*(\theta, y, X) + \gamma, \gamma \in \mathbb{R}^k$$

$$(iv) \beta^*(\theta, y, XA) = A^{-1} \beta^*(\theta, y, X), A_{k \times k} \text{ não singular.}$$

PROVA : Seja

$$\psi(\tilde{\beta}; \theta, y, X) = \sum_{\{i: y_i \geq x_i \tilde{\beta}\}} \theta |y_i - x_i \tilde{\beta}| + \sum_{\{i: y_i < x_i \tilde{\beta}\}} (1 - \theta) |y_i - x_i \tilde{\beta}|$$

$$\psi(\tilde{\beta}; \theta, y, X) = \sum_{i=1}^n \theta [y_i - x_i \tilde{\beta}] - 1/2 [y_i - x_i \tilde{\beta}] + 1/2 \operatorname{sgn}(y_i - x_i \tilde{\beta}) [y_i - x_i \tilde{\beta}]$$

$$\text{onde } \operatorname{sgn}(u) = \begin{cases} 1, & u > 0 \\ 0, & u = 0 \\ -1, & u < 0 \end{cases} \quad \text{e } \tilde{\beta} = \beta^*(\theta)$$

então nós temos que:

$$(i) \lambda \psi(\tilde{\beta}; \theta, y, X) = \psi(\lambda \tilde{\beta}; \theta, \lambda y, X), \lambda \in [0, \infty)$$

$$(ii) -\lambda \psi(\tilde{\beta}; \theta, y, X) = \psi(\lambda \tilde{\beta}; 1 - \theta, \lambda y, X), \lambda \in (-\infty, 0]$$

$$(iii) \psi(\tilde{\beta}; \theta, y, X) = \psi(\tilde{\beta} + \gamma; \theta, y + X\gamma, X), \gamma \in \mathbb{R}^k$$

$$(iv) \psi(\tilde{\beta}; \theta, y, X) = \psi(A^{-1}\tilde{\beta}; \theta, y, XA), |A_{k \times k}| \neq 0. \blacksquare$$

Este teorema apresenta diversas propriedades sobre características equivariantes da regressão quantílica. Os itens (i) e (ii) implicam que a solução da regressão L_1 é equivariante em escala, (iii) constata que $\beta^*(\theta)$ é equivariante de posição e (iv) afirma que $\beta^*(\theta)$ é equivariante à reparametrização da matriz de planejamento.

Bassett e Koenker (1978) mostraram que $\beta^*(\theta) = X(h)^{-1}y(h)$ é a solução única para o problema de regressão quantílica, se somente se

$$(\theta - 1)1'_k < \sum_{i \in h} \left[1/2 - 1/2 \operatorname{sgn}(y_i - x_i \tilde{\beta}) - \theta \right] x_i X(h)^{-1} < \theta 1'_k$$

Se na expressão acima considerarmos os quantis amostrais, em que $x_i = 1$, $i = 1, \dots, k$, $H = \Gamma$ e F é contínua teremos que $\beta^*(\theta) = y(h)$ é o único θ -ésimo quantil amostral, se e somente se $(\theta - 1) < \sum_{i \in h} \left[1/2 - 1/2 \operatorname{sgn}(y_i - y(h)) - \theta \right] < \theta$. Esta desigualdade implica que o número de y_i menor que $y(h)$ esteja estritamente entre $k\theta - 1$ e $k\theta$, portanto $k\theta$ não pode ser inteiro. A continuidade de F elimina o problema de empates no modelo de posição e também no modelo linear. Na ausência dessa condição, a solução única ocorrerá dependendo da matriz de planejamento. No modelo de posição isso implica que $k\theta$ seja não inteiro. Para obtermos uma sequência de soluções únicas, pode-se retirar de uma sequência de matrizes de planejamento $\{X_k\}$ uma subsequência (ou no pior caso uma subsequência perturbada) adequada, ou adotar uma regra arbitrária para escolher um único elemento de um conjunto de soluções.

TEOREMA 2.3 - Seja $P(u^*(\theta))$, $N(u^*(\theta))$ e $Z(u^*(\theta))$ o número de

elementos positivos, negativos e iguais a zero no vetor $u^*(\theta) = y - X\beta^*(\theta)$. Então se X contém uma coluna de uns,

$$N(u^*(\theta)) \leq n\theta \leq n - P(u^*(\theta)) = N(u^*(\theta)) + Z(u^*(\theta)) \quad (2.2.7)$$

para todo $\beta^*(\theta) \in B^*(\theta)$. Se $\beta^*(\theta)$ é único, isto é, $\beta^*(\theta) = B^*(\theta)$, então as desigualdades acima são estritas.

PROVA : Particionamos a matriz X com $X = [1 : \tilde{X}]$. Pelo argumento usado no teorema (2.3), $\beta^*(\theta) \in B^*(\theta)$ se e somente se

$$\psi'(\tilde{\beta}; w) = \sum_{i=1}^n \left[1/2 - 1/2 \operatorname{sgn}^*(y_i - x_i \tilde{\beta}; -x_i w) - \theta \right] x_i w \geq 0 \quad (2.2.8)$$

onde $\psi'(\tilde{\beta}; w)$ é a derivada de $\psi(\tilde{\beta}; \theta, y, X)$ na direção w , para todo $w \neq 0$. Em particular para $w^+ = (1, 0, \dots, 0)$ e $w^- = (-1, 0, \dots, 0)$ em \mathbb{R}^k . Então $x_i w^+ = 1$ e $x_i w^- = -1$ para todo x_i , dos resultados mostrados por Bassett e Koenker (1978) implica que $\sum_{i=1}^n \pm [1/2 - 1/2 \operatorname{sgn}^*(y_i - x_i \tilde{\beta}; \pm 1) - \theta] > 0$ que é equivalente às seguintes condições

$$(i) -\theta P + (1 - \theta)N + (1 - \theta)Z > 0$$

$$(ii) -\theta P + (1 - \theta)N - Z < 0$$

que são equivalentes à relação (2.2.12). Se $\beta^*(\theta)$ é único então todas as desigualdades são estritas ■

Este teorema é importante para sabermos como vai ser dividido o número de observações, acima e abaixo do hiperplano de regressão quantílica, de acordo com o parâmetro θ .

Se F for uma função contínua então $Z(u^*(\theta)) = k$ com probabilidade um, então existirá no mínimo $n\theta$ observações abaixo ou sobre o θ -ésimo hiperplano de regressão quantílica e no máximo

$(n\theta+k)$ observações acima do hiperplano de regressão quantílica.

TEOREMA 2.4 - Se $\beta^*(\theta) \in B^*(\theta, y, X)$, então $\beta^*(\theta) \in B^*(\theta, X\tilde{\beta} + Du^*(\theta), X)$, onde $u^*(\theta) = y - X\beta^*(\theta)$ e D é uma matriz diagonal $(n \times n)$ com elementos não-negativos.

PROVA : $\tilde{\beta} = \beta^*(\theta) \in B^*(\theta, y, X)$ implica que

$$\sum_{i=1}^n \left[\frac{1}{2} - \frac{1}{2} \operatorname{sgn}^*(y_i - x_i \tilde{\beta}; -x_i w) - \theta \right] x_i w \geq 0 \quad \text{para todo } w \neq 0.$$

Podemos notar que

$$\left[\frac{1}{2} - \frac{1}{2} \operatorname{sgn}^*(x_i \tilde{\beta} + d_i(y_i - x_i \tilde{\beta}) - x_i \tilde{\beta}; -x_i w) - \theta \right] x_i w = \left(\frac{1}{2} - \theta \right) x_i w - \frac{1}{2} \operatorname{sgn}^*(d_i(y_i - x_i \tilde{\beta}); -x_i w) x_i w \geq \left(\frac{1}{2} - \theta \right) x_i w - \frac{1}{2}$$

$$\operatorname{sgn}^*(y_i - x_i \tilde{\beta}; -x_i w) x_i w$$

de onde segue o resultado do teorema. ■

Este teorema tem uma interpretação geométrica importante que se baseia no seguinte: se tivermos várias observações dispersas no \mathbb{R}^k , com a θ -ésima regressão quantílica passando através desses pontos, e se considerarmos o efeito do movimento das observações para cima ou para baixo do hiperplano, de tal forma que cada observação continue do mesmo lado do hiperplano (o resíduo não troca de sinal), então a θ -ésima regressão quantílica não se altera.

2.3 - DISTRIBUIÇÃO ASSINTÓTICA DOS ESTIMADORES DA REGRESSÃO QUANTÍLICA

Neste item estudaremos a teoria assintótica da regressão

quantílica, isto é, a distribuição assintótica dos estimadores da θ -ésima regressão quantílica.

O seguinte teorema, devido a Mosteller (1946), proporciona o fundamento para a teoria das grandes amostras das estatísticas sistemáticas, ou seja, estimadores que são combinações lineares de quantis amostrais. A mediana é o caso particular mais importante com variância $[2f(\xi(1/2))]^{-2}$. Esse valor é menor do que a variância da média, para uma ampla classe de distribuições com caudas alongadas.

TEOREMA 2.5 - Seja $\{\xi_n^*(\theta_1), \dots, \xi_n^*(\theta_M)\}$ com $0 < \theta_1 < \theta_2 < \dots < \theta_M < 1$ uma sequência de quantis amostrais únicos de amostras aleatórias de tamanho n de uma população cuja inversa da função de distribuição $\xi(\theta) = F^{-1}(\theta)$. Se F é contínua e tem f.d.p. f positiva em $\xi(\theta_i)$, $i = 1, \dots, M$, então $\sqrt{n} [\xi_n^*(\theta_1) - \xi(\theta_1), \dots, \xi_n^*(\theta_M) - \xi(\theta_M)]$ converge em distribuição para um vetor de variáveis aleatórias normais M -dimensional, com média 0, e matriz de covariância $\Omega(\theta_1, \dots, \theta_M; F)$ com

$$\omega_{ij} = \frac{\theta_i(1 - \theta_j)}{f(\xi(\theta_i))f(\xi(\theta_j))}, \quad i \leq j \quad (2.2.9)$$

A analogia entre quantis amostrais e regressão quantílica no modelo linear torna-se ainda mais evidente com a forte semelhança entre os comportamentos assintóticos, apresentada no seguinte teorema:

TEOREMA 2.6 - (Bassett e Koenker (1978)) Seja $\{\beta_n^*(\theta_1), \dots, \beta_n^*(\theta_M)\}$ com $0 < \theta_1 < \dots < \theta_M < 1$ uma sequência de soluções únicas das regressões quantílicas do modelo $P(Y_i < y) = F(y - x_i\beta)$. Seja $\xi(\theta) = F^{-1}(\theta)$, $\xi(\theta) = (\xi(\theta), 0, \dots, 0) \in \mathbb{R}^k$ e $\xi_n^*(\theta) = \beta_n^*(\theta) - \beta$. Suponhamos que:

(i) F é contínua e tem densidade positiva, f , em $\xi(\theta_i)$,
 $i = 1, \dots, M$ e

(ii) $x_{1i} = 1 : i = 1, 2, \dots$ e $\lim_{n \rightarrow \infty} n^{-1} X'X = Q$, uma matriz positiva definida. Então, $\sqrt{n}[\xi_n^*(\theta_1) - \xi(\theta_1), \dots, \xi_n^*(\theta_M) - \xi(\theta_M)]$ converge em distribuição para um vetor de variáveis aleatórias normais Mk -dimensional com média zero e matriz de covariância $\Omega(\theta_1, \dots, \theta_M) \otimes Q^{-1}$, onde Ω é a matriz de covariância dos M quantis amostrais de amostras aleatórias da distribuição F .

O paralelismo existente entre o comportamento assintótico dos quantis amostrais no modelo de posição e regressão quantílica no modelo linear, sugere a extensão da teoria de grandes amostras das estatísticas sistemáticas do modelo de posição para o modelo linear. O próximo teorema explicita esta idéia.

TEOREMA 2.7 - Seja $\pi(\theta) = (\pi(\theta_1), \dots, \pi(\theta_M))'$ uma distribuição de probabilidade discreta e simétrica assumindo valores em $\langle \theta_i : i = 1, \dots, M \mid 0 < \theta_1 < \dots < \theta_M < 1 \rangle$. Suponha que $F(0) = 1/2$, então $\xi(1/2) \equiv F^{-1}(1/2) \equiv 0$, e as condições (i) e (ii) do teorema 2.6 valem, então $\tilde{\beta}_n(\pi(\theta)) = \sum \pi(\theta) \beta_n^*(\theta)$ é invariante de posição, escala e reparametrização do modelo, de acordo com o teorema 2.2 e $\sqrt{n}(\tilde{\beta}_n(\pi(\theta)) - \beta)$ converge em distribuição para um vetor k -dimensional com distribuição normal de média 0 e matriz de covariância $\pi' \Omega \pi Q^{-1}$.

PROVA: Do teorema 2.2 nós temos que:

$$(i) \sum \pi(\theta) \beta_n^*(\theta, \lambda y, X) = \lambda \sum \pi(\theta) \beta_n^*(\theta, y, X) \quad \lambda \in [0, \infty),$$

$$(ii) \sum \pi(1 - \theta) \beta_n^*(1 - \theta, \lambda y, X) = \lambda \sum \pi(\theta) \beta_n^*(\theta, y, X) \quad \lambda \in (-\infty, 0],$$

$$(iii) \sum \pi(\theta) \beta_n^*(\theta, y + X\gamma, X) = \sum \pi(\theta) \beta_n^*(\theta, y, X) + \gamma, \gamma \in \mathbb{R}^k,$$

$$(iv) \sum \pi(\theta) \beta_n^*(\theta, y, XA) = A^{-1} \sum \pi(\theta) \beta_n^*(\theta, y, X) A_{k \times k}, \text{ não-singular.}$$

Do resultado do teorema 2.5 nós temos que se o teorema 2.7 satisfaz as condições (i) e (ii) então $\sqrt{n}(\tilde{\beta}_n(\pi(\theta)) - \beta)$ converge em distribuição para um vetor de variáveis aleatórias $N(0, \Pi' \Omega \Pi Q^{-1})$ de dimensão k . ■

Neste teorema nós temos que $\tilde{\beta}_n(\pi(\theta))$ é uma ponderação de $\beta_n^*(\theta)$, cujos pesos são os $\pi(\theta_i)$. Diversos estimadores da forma $\tilde{\beta}_n(\pi(\theta))$ são eficientes para uma classe de distribuições, assim como os estimadores de Gastwirth e a trimédia. A eficiência da média e do estimador $\hat{\beta}$ de mínimos quadrados para o modelo linear, ao contrário dos estimadores anteriores, diminui a medida que a distribuição dos erros distancia-se da distribuição normal em direção à distribuições com caudas mais alongadas.

Tukey (1975) sugeriu que deveriam ser desenvolvidos esforços para encontrar estimadores que modificassem o método de mínimos quadrados reduzindo significativamente sua sensibilidade a valores aberrantes, mas conservando as suas boas qualidades.

2.4 - REGRESSÃO DA MSEA (MÍNIMA SOMA DOS ERROS ABSOLUTOS) OU REGRESSÃO L_1

Estudaremos aqui o caso em que $\theta = 1/2$, ou seja, a regressão L_1 .

O problema de estimação da regressão L_1 pode ser escrito como um problema de programação linear (P.P.L.), como segue:

$$\min [1'e^+ + 1'e^-] \quad (P1)$$

sujeito à

$$Y = X\tilde{\beta} + \epsilon^+ - \epsilon^-,$$

$$\tilde{\beta}, \epsilon^+, \epsilon^- \geq 0$$

onde a solução deste problema é o estimador L_1 de β e ϵ^+ , ϵ^- são os erros positivos e negativos respectivamente e $1' = (1, \dots, 1)$ com dimensão $(1 \times n)$.

Podemos associar este problema (P) ao seu dual que é definido como:

$$\max \{ Y'd \} \tag{D1}$$

sujeito à

$$X'd = 0, \quad -1 < d_i < 1, \quad i = 1, \dots, n$$

Quando a matriz X é de posto completo, podemos destacar algumas propriedades da regressão L_1 , que são casos particulares das propriedades da regressão quantílica.

Propriedade P_1 - Existe pelo menos um hiperplano ótimo da regressão L_1 que passa através de k pontos observados. Essa propriedade é consequência do teorema 2.1.

Propriedade P_2 - Um solução para o problema (P1) é um hiperplano tal que $|n^+ - n^-| \leq k$ em que n^+ é igual ao número de observações acima e n^- é o número de observações abaixo do hiperplano. Essa propriedade é derivada do teorema 2.3.

Propriedade P_3 - Sem perda de generalidade β pode ser localizada de modo que $F(0) = 1/2$, então $\xi(1/2) = 0$. Temos então que a distribuição assintótica da variável aleatória $\sqrt{n}(\hat{\beta}(1/2) - \beta)$ é normal k -dimensional com média zero e matriz de variância $[2f(0)]^{-2}Q^{-1}$.

Vê-se portanto que a regressão L_1 é mais eficiente do que a regressão de mínimos quadrados, no caso linear, para qualquer distribuição em que a mediana é mais eficiente do que a média no modelo de posição.

CAPÍTULO 3

INFERÊNCIA

Neste capítulo desenvolveremos os seguintes aspectos da Inferência Estatística em Regressão L_1 : Intervalo de Confiança, teste de Hipótese e Análise de Variação.

3.1 - INTRODUÇÃO

Vários autores analisaram, através de estudos de simulação, a eficiência relativa entre o estimador de regressão L_1 e o estimador de mínimos quadrados, utilizando várias distribuições. Dentre eles podemos citar Rice e Whitte (1964), Pfaffemberger e Dinkel (1978), Rosemberg e Carlson (1977). Mostrou-se que o estimador L_1 é mais eficiente do que o estimador de mínimos quadrados, considerando inclusive, pequenas amostras, quando os erros tem distribuição Cauchy, Laplace e Pareto Estável com expoente menor ou igual a 1,5 ou certas distribuições normais contaminadas. Ashar e Wallace (1963) e Wilson (1978) mostraram que quando as suposições de mínimos quadrados são satisfeitas a eficiência do estimador L_1 é de aproximadamente 80 %.

Um dos primeiros estudos sobre propriedades dos estimadores L_1 para pequenas amostras, através da simulação, usando regressões múltiplas e utilizando erros com distribuição normal e normais contaminadas simétricas foi feito por Rosemberg e Carlson (1977), usando 100.000 regressões L_1 , e obtendo os seguintes resultados:

- (1) O estimador L_1 tem menor desvio padrão do que o estimador de mínimos quadrados para regressões com curtose grande.
- (2) Os estimadores L_1 são quase normais quando a distribuição dos erros tem curtose grande.
- (3) A diferença $(\tilde{\beta} - \beta)$ em regressão L_1 , é aproximadamente normal com média 0 e matriz de covariância $\lambda^2(X'X)^{-1}$, em que λ^2/n é a variância da mediana de uma amostra de tamanho n .

Outros autores fizeram também estudos sobre as propriedades para grandes amostras, como por exemplo Taylor (1974), foi o primeiro a provar algumas propriedades para os estimadores L_1 e apresentou resultados para estimadores L_1 não-viciados cujo resíduos tem distribuição simétrica em torno de zero.

Rosemberg e Carlson publicaram seus resultados sobre a teoria da distribuição assintótica dos estimadores L_1 em 1977, que foi mais tarde desenvolvido analiticamente por Bassett e Koenker (1978). Este resultado apresentado por Bassett e Koenker é o mais importante para distribuição assintótica dos estimadores L_1 . Eles provaram que os estimadores L_1 , $\tilde{\beta}$, são assintoticamente normais e consistentes, com média β e matriz de covariância $\lambda^2(X'X)^{-1}$, onde X é uma matriz $(n \times k)$ e λ^2/n é a variância assintótica da mediana de uma amostra aleatória de tamanho n . Esses resultados mostraram que para as distribuições de erros, cuja mediana é assintoticamente mais eficiente do que a média, o estimador L_1 tem uma elipsóide de concentração estritamente menor do que a dos mínimos quadrados. Através da teoria assintótica dos estimadores L_1 , podemos estudar Intervalos de Confiança, testes de Hipótese e Análise de Variação dentro da regressão L_1 .

3.2 - INTERVALO DE CONFIANÇA

Denotemos por $\tilde{\beta}$ o estimador de regressão L_1 de β no

modelo (2.2.2), definido no capítulo 2. Estamos interessados em um intervalo de confiança para $h'\beta$, onde h é um vetor de constantes e β é o vetor de parâmetros desconhecidos, então o estimador L_1 de $h'\beta$ é $h'\tilde{\beta}$.

Baseado nos resultados assintóticos nós podemos escrever um intervalo de $(1 - \alpha)100\%$ de confiança para $h'\beta$ como:

$$h'\tilde{\beta} \pm z_{\alpha/2} \lambda \left[h'(X'X)^{-1}h \right]^{1/2} \quad (3.1)$$

onde $z_{\alpha/2}$ é o percentil da distribuição normal padronizada, e λ é o desvio-padrão amostral da mediana dos resíduos.

Para uma componente β_i de β , o intervalo de $(1 - \alpha)100\%$ de confiança será:

$$\tilde{\beta}_i \pm z_{\alpha/2} \lambda (X'X)^{-1/2}_{ii} \quad (3.2)$$

onde $(X'X)^{-1/2}_{ii}$ é a raiz quadrada do i -ésimo elemento da diagonal da matriz $(X'X)^{-1}$.

Se o valor de λ não for conhecido, teremos que estimá-lo. Temos que $\lambda = \theta/2$, onde $\theta = 1/f(\eta)$ e $f(\eta)$ é o valor da função de densidade dos resíduos calculado na mediana.

Para estimarmos λ , será necessário estimarmos a f.d.p. $f(\eta)$. Uma estimativa consistente de $[f(\eta)]^{-1}$, dada por Cox e Hinkley (1974) é:

$$\hat{\theta} = \frac{e_{(i)} - e_{(r)}}{(i - r)/n} \quad \begin{array}{l} i = 1, \dots, n \\ r = 1, \dots, n \end{array} \quad (3.3)$$

em que denotamos os resíduos observados por \hat{e}_i , $i = 1, \dots, n$, onde $\hat{e}_i = y_i - x_{ij}\tilde{\beta}_j$, $j = 1, \dots, k$ e $e_{(i)}$ e $e_{(r)}$ são o i -ésimo e o

r-ésimo resíduos ordenados. Os valores i e r em (3.5) devem ser simétricos em torno da mediana amostral dos resíduos. Além disso

$$i = [n/2] + v \quad e \quad r = [n/2] - v \quad (3.4)$$

onde $[.]$ indica o maior inteiro contido, e v é escolhido de tal forma que i e r fiquem simétricos em torno da mediana amostral dos resíduos. Recomenda-se que v seja pequeno e $e_{(i)}^*, e_{(r)}^* \neq 0$. $\hat{\theta}$ é um estimador consistente de 2λ , logo uma estimativa para λ é dada por $\hat{\lambda} = \hat{\theta}/2$.

Outras propostas foram apresentadas por Maritz e Jarret (1978) e Efron (1979), ou seja;

a) O estimador proposto por Maritz e Jarret é obtido através do método Bootstrap. Seja $n^* = n - k$ e sejam e_1^*, \dots, e_n^* os resíduos diferentes de zero. Para n^* ímpar o estimador de λ é dado por:

$$\hat{\lambda}_{BS}^2 = \sum_{i=1}^{n^*} (e_{(i)}^* - \tilde{e}^*)^2 \left[B(m-1; n^*, (i-1)/n^*) - B(m-1; n^*, i/n^*) \right] \quad (3.5)$$

onde $n^* = 2m - 1$ é o número de observações sem os resíduos iguais a zero, \tilde{e}^* é a mediana dos resíduos que permanecem. $B(k; n^*, p)$ é a distribuição binomial. Em Maritz e Jarret (1978) encontra-se a expressão para n^* par.

b) O estimador proposto por Efron é baseado em um intervalo da forma $(e_{(k)}^*, e_{(n-k+1)}^*)$, onde o valor aproximado de k é $((n^* + 1)/2) - z_{\alpha/2} (n^*/4)^{1/2}$ e $z_{\alpha/2}$ é o percentil superior da $N(0,1)$. O estimador de λ é,

$$\hat{\lambda}_{1-\alpha} = \sqrt{n^*} (e_{(n-k+1)}^* - e_{(k)}^*) / (2z_{\alpha/2}) \quad (3.6)$$

Schrader e McKean (1982) fizeram um estudo de Monte Carlo comparando os dois estimadores, e concluíram que considerando o modelo de posição, estes estimadores tem aproximadamente o mesmo desempenho. No estudo feito por McKean e Schrader (1987) $\alpha = 0.05$ parece ser o melhor valor para o modelo (3.6).

O intervalo de confiança, com $(1-\alpha)100\%$ de $h'\beta$ pode ser agora calculado como:

$$h'\beta \pm z_{\alpha/2} \hat{\lambda} \left[h'(X'X)^{-1}h \right]^{1/2} \quad (3.7)$$

Podemos também construir um intervalo de confiança para a resposta média Y_0 . Seja $X'_0 = (X_{10}, \dots, X_{k0})$ um vetor de observações de X , a estimativa de Y_0 correspondente a X_0 , é dada por:

$$\hat{Y}_0 = X'_0 \tilde{\beta} \quad (3.8)$$

e o intervalo $(1 - \alpha)100\%$ de confiança será dado por:

$$\hat{Y}_0 \pm z_{\alpha/2} S(\hat{Y}_0) \quad (3.9a)$$

onde $S^2(\hat{Y}_0) = \hat{\lambda}^2 (X'_0 (X'X)^{-1} X_0)$ (3.9b)

Ainda podemos construir um intervalo de confiança para uma nova observação, Y_{nova} , em relação ao vetor X_0 , que é dado por:

$$\hat{Y}_{nova} \pm z_{\alpha/2} S(\hat{Y}_{nova}) \quad (3.10)$$

Os intervalos de confiança acima apresentados são baseados na normalidade assintótica de Y , suposição esta que depende do tamanho da amostra e da distribuição dos erros.

Dielman e Pfaffenberger (1988) fizeram um estudo, através de simulação Monte Carlo para determinar o tamanho de amostra necessário para obter normalidade, e chegaram a conclusão que se os erros são normalmente distribuídos, o tamanho de amostra $n = 20$ parece ser suficiente para utilizar-se os resultados assintóticos. Se os erros não são normalmente distribuídos, a convergência para a distribuição normal depende do número de variáveis preditoras dentro do modelo e deve-se tomar uma amostra suficientemente grande para garantir a normalidade assintótica. Se a distribuição dos erros tem caudas muito pesadas, tamanhos de amostra $n \geq 200$ foram recomendados para utilizar aproximação normal. A multicolinearidade parece ter um efeito na velocidade de convergência para normalidade dos estimadores L_1 . Os resultados sugerem que deve ser feita uma análise de resíduos do modelo ajustado para depois usar a inferência em MSEA.

Stangenhans e Narula (1987) em estudo de Monte Carlo, mostraram que mesmo com tamanhos de amostra relativamente pequenos ($10 \leq n \leq 100$) a aproximação normal já é satisfatória e proporciona bons resultados para fins de construção de intervalos de confiança e testes de hipótese.

O método Bootstrap tem sido sugerido na determinação de intervalo de confiança e construção de testes de hipótese para pequenas amostras, como por exemplo: Stangenhans e Narula (1987); Schrader e McKean (1987); Filho, Pedro F. (1988) e Dielman e Pfaffenberger (1988).

3.3 - TESTES DE HIPÓTESE

Construiremos testes para as seguintes hipóteses:

a) Para testar a hipótese nula, $H_0: h'\beta = c$ contra $H_A: h'\beta \neq c$ com um nível de significância α , teremos as seguintes regiões:

Rejeitar H_0 se

$$Z^* = \frac{h'\tilde{\beta} - c}{\hat{\lambda}(h'(X'X)^{-1}h)^{1/2}} > Z_{\alpha/2} \text{ ou } Z^* < -Z_{\alpha/2} \quad (3.11)$$

b) Para testar $H_0: \beta_l = 0$ contra $H_A: \beta_l \neq 0$, temos:

Rejeitar H_0 se

$$Z^* = \frac{\tilde{\beta}_l}{\hat{\lambda}(X'X)^{-1/2}_{ll}} > Z_{\alpha/2} \text{ ou } Z^* < -Z_{\alpha/2} \quad (3.12)$$

c) Para testar $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ contra $H_A: \beta_l \neq 0$ para $l = 1, \dots, k$

$$W = \frac{(\tilde{\beta}^* - \beta^*)' X^* X^* (\tilde{\beta}^* - \beta^*)}{\hat{\lambda}^2} \quad (3.13)$$

em que (*) significa que a coluna de 1's da matriz X foi retirada. A expressão acima tem assintoticamente uma distribuição

Qui-Quadrada com $k-1$ graus de liberdade.

Rejeitar H_0 se

$$W > \chi^2_{(1-\alpha, k-1)} \quad (3.14)$$

onde $\chi^2_{(1-\alpha, k-1)}$ é o percentil da Qui-Quadrado $(1 - \alpha)$ e $k-1$ graus de liberdade.

Os testes de hipótese desenvolvidos aqui são assintoticamente equivalentes ao de Neyman-Pearson e ao da razão de verossimilhança, como foi mostrado por Bassett e Koenker (1978).

3.4 - ANÁLISE DE VARIACÃO

A Análise de Variação foi proposta por McKean e Schrader (1987). A Análise de Variação em regressão L_1 não é propriamente como a análise de variância em Mínimos Quadrados, pois aqui não temos projeções. O procedimento é desenvolvido no mesmo espírito utilizado no método dos mínimos quadrados para testar uma hipótese linear geral, sendo que trocamos a soma de quadrados dos resíduos pela soma dos resíduos absolutos, isto é, utilizando a redução na soma de desvios absolutos. O objetivo é criar uma estatística que tenha uma distribuição assintoticamente conhecida, e usando estas estatísticas construiremos os testes de hipótese.

3.4.1 - TESTE PARA A HIPÓTESE LINEAR GERAL

Na parte de teste de hipótese vimos vários casos. Aqui construiremos um teste de hipótese para o modelo linear geral da forma,

$$H_0: H\beta = 0 \text{ contra } H_A: H\beta \neq 0 \quad (3.15)$$

onde H é uma matriz $q \times p$ tal que $H\beta$ é um conjunto de q funções estimáveis, linearmente independentes, em que a definição de estimabilidade aqui é análoga à definição utilizada em mínimos quadrados.

Seja $Y = X\beta + \varepsilon$ o modelo linear completo. Sob a hipótese $H_0: H\beta = 0$, temos o subespaço $\omega = \langle X\beta \mid H\beta = 0 \rangle$ de Ω de dimensão $k-q$ do modelo reduzido, onde k é a dimensão do espaço das colunas de X .

Para o caso dos mínimos quadrados o teste é feito através da $SQ(\omega)$ (Soma dos erros ao quadrado), que é a distância Euclidiana entre Y e ω , e a estatística Q será baseada na redução da soma dos quadrados quando passamos do modelo reduzido para o modelo completo, que é definida como:

$$SQ = SQ(\omega) - SQ(\Omega) \quad (3.16)$$

No teste para os parâmetros da regressão L_1 apenas substituiremos a distância Euclidiana pela L_1 , ou seja, $SEAC(\omega)$ (Soma dos erros absolutos) que é a distância L_1 entre Y e ω , e a estatística do teste será baseada na redução da SEA (Soma dos erros absolutos), passando do modelo reduzido para o modelo completo:

$$SA = SEAC(\omega) - SEAC(\Omega) \quad (3.17)$$

Rejeitaremos H_0 para valores grandes de (3.17).

Se considerarmos o teste da razão de verossimilhança, denotado como Λ , então $SQ = -2\ln\Lambda$ se os erros tem distribuição normal, e $SA = -2\ln\Lambda$ quando os erros tem distribuição de Laplace. Em geral a forma da distribuição não é conhecida e devemos encontrar uma padronização adequada para a estatística do teste. O teste para os mínimos quadrados é dado por

$$Q_{MQ} = \frac{SQ(\omega) - SQ(\Omega)}{\hat{\sigma}^2} \quad (3.18a)$$

onde $\hat{\sigma}^2 = \frac{SQ(\Omega)}{(n - k)}$ (3.18b)

Para a estatística L_1 , teremos o seguinte teste:

$$Q_{L_1} = \frac{SAC(\omega) - SAC(\Omega)}{q(\hat{\lambda}/2)} \quad (3.19)$$

onde $\hat{\lambda}$ é a estimativa do desvio-padrão amostral da mediana dos resíduos, já apresentada anteriormente na seção 3.2.

Se considerarmos o teste de hipótese da forma:

$$H_n: H\beta_n = n^{-1/2}\theta \quad ||\theta|| \neq 0 \quad \text{com } n \rightarrow \infty,$$

Schrader e Hettmansperger (1976) provaram que sob as seguintes condições de regularidade:

(i) $\lim \text{Sup}(\text{diag } P_{\Omega_n}) = 0$ quando $n \rightarrow \infty$

onde Ω_n contém Ω , que é o espaço das colunas de X , e P_{Ω_n} é a projeção sobre Ω_n .

(ii) $F'(\eta) = f(\eta) > 0$

onde η é a mediana dos e_i 's (erros aleatórios), $i = 1, \dots, n$.

tem-se os seguintes resultados:

(a) Sob H_0 , qQ_{L_1} tem assintoticamente uma distribuição Qui-Quadrado com q graus de liberdade.

(b) Sob H_n , qQ_{L_1} tem assintoticamente uma distribuição Qui-Quadrado com q graus de liberdade e parâmetro de não-centralidade,

$$\delta_{L_1} = \frac{(H\beta_n)' \left[H(X'X)^{-1}H' \right]^{-1} (H\beta_n)}{2\lambda^2} \quad (3.20)$$

Se assumirmos que o erro da distribuição tem variância finita então qQ_{MQ} é assintoticamente Qui-Quadrado, e sob H_n tem parâmetro assintótico de não-centralidade

$$\delta_{MQ} = \frac{(H\beta_n)' \left[H(X'X)^{-1}H' \right]^{-1} (H\beta_n)}{2\sigma^2} \quad (3.21).$$

Portanto a ERA (Eficiência Relativa Assintótica) de Q_{L_1} com relação a Q_{MQ} é definida como a razão dos parâmetros de não-centralidade,

$$ERA = \frac{\sigma^2}{\lambda^2} \quad (3.22).$$

Se os erros tem distribuição normal então ERA = 0.63, entretanto se eles tiverem distribuição de Laplace, ERA = 2.0. Podemos assim verificar que a análise de regressão L_1 é mais eficiente do que a análise de mínimos quadrados quando a distribuição dos erros tem caudas muito pesadas.

3.4.2 - EXEMPLO PARA CONSTRUÇÃO DA TABELA DE ANÁLISE DE VARIÇÃO

Vamos supor que estamos interessados em testar o paralelismo entre duas retas, para isso consideramos o seguinte modelo:

$$Y_{ij} = \mu + \alpha_i + \beta_i X_j + \varepsilon_{ij} \quad (3.23)$$

onde $i = 1, 2$, e $j = 1, 2, \dots, n$.

Desejamos testar a seguinte hipótese:

$$H_0: \begin{aligned} \alpha_1 &= \alpha_2 \\ \beta_1 &= \beta_2 \end{aligned}$$

versus

$$H_A: \begin{aligned} \alpha_1 &\neq \alpha_2 \\ \beta_1 &\neq \beta_2 \end{aligned}$$

A tabela terá a seguinte forma:

TABELA DE ANÁLISE DE VARIAÇÃO PARA O MODELO ACIMA

FONTE	SA	GRAUS DE LIBERDADE	SMA	Q_{L_1}
A	$SAC(\omega) - SAC(\Omega)$	4 - 2	$\frac{SAC(\omega) - SAC(\Omega)}{2}$	$\frac{SMA}{\hat{\lambda}/2}$
ERRO	-	$2(n - 1)$	$\hat{\lambda}/2$	
TOTAL	-	$2n$		

onde

- A = FONTE REFERENTE A HIPÓTESE H_0
- SA = SOMA ABSOLUTA
- SMA = SOMA ABSOLUTA/GRAUS DE LIBERDADE DE A
- Q_{L_1} = ESTATÍSTICA QUI-QUADRADO
- $\hat{\lambda}$ = ESTIMATIVA DO DESVIO PADRÃO AMOSTRAL DA MEDIANA DOS RESÍDUOS

CAPÍTULO 4

SELEÇÃO DE VARIÁVEIS

4.1 - INTRODUÇÃO

Frequentemente em problemas de regressão, podemos selecionar um modelo com $m < k$ variáveis preditoras sem perdermos essencialmente muita informação sobre a variável resposta. Um modelo simplificado pode ajudar a interpretar melhor o problema em estudo, diminuir o trabalho computacional e as análises estatísticas, o que pode ser desejável no ponto de vista econômico. Escolher um modelo com menos do que k variáveis implica na busca dentre 2^{k-1} possíveis modelos.

Existem vários procedimentos para a seleção de variáveis, como o passo a passo, o passo atrás e o passo a frente, mas nem sempre estes métodos garantem encontrar o "melhor" subconjunto de variáveis preditoras. Por outro lado, a enumeração explícita de todos os modelos possíveis não é conveniente pois alguns modelos podem ser inadequados. Algoritmos de enumeração implícita fornecem melhores soluções para o problema. Além de serem eficientes proporcionam o "melhor" modelo de cada dimensão.

Um modelo de tamanho $m (=1, \dots, k-1)$, é dito ser o melhor se, dentre os $\binom{k}{m}$ possíveis modelos ele tem o menor valor para o critério que for usado. Narula e Wellington (1979), propuseram um algoritmo para selecionar o melhor modelo de $m (=1, \dots, k-1)$ variáveis preditoras que será apresentado na próxima seção.

4.2 - ALGORITMO DE NARULA E WELLINGTON (1979) PARA ENCONTRAR O MELHOR SUBCONJUNTO DE VARIÁVEIS PREDITORAS EM REGRESSÃO L_1

O algoritmo proposto aqui reduzirá bastante o esforço computacional e não será preciso examinar todos os subconjuntos do modelo completo. A sua eficiência está associada às seguintes observações:

1) Seja $Z(m)$ o valor da MSEA (Mínima Soma dos Erros Absolutos) referente a menor soma dos desvios entre os modelos de tamanho m ($= 1, \dots, k-1$) variáveis preditoras, ou seja melhor modelo, então

$$Z(k) \leq Z(k-1) \leq \dots \leq Z(2) \leq Z(1) \quad (4.1)$$

2) O valor da MSEA para um modelo com m variáveis preditoras é menor ou igual ao valor da MSEA de um modelo com r ($= 1, \dots, m-1$) variáveis preditoras.

O algoritmo tem como base o teste de sondagem e o teste de pré-otimalidade, partindo do princípio que, quando o problema (P1) (definido no capítulo 2 seção 2.4) for primal ou dual factível, então aplicaremos o algoritmo primal ou dual, usando o teste de sondagem, mas se este não tiver sucesso, usaremos o teste de pré-otimalidade que está associado com o dual do problema.

O teste de sondagem é baseado nas observações citadas acima, e será definido como:

Se o valor da SEA (Soma dos Erros Absolutos), associado a um modelo com m variáveis preditoras é maior ou igual a Z_r para algum $r < m$, os modelos com $r, r+1, \dots, m-1$ variáveis preditoras não precisa ser examinados.

O teste de pré-otimalidade que está associado com o dual é definido da seguinte maneira:

Se para alguma iteração a SEA (Soma dos Erros Absolutos) para um subconjunto de $m < k$ variáveis preditoras é maior ou igual ao valor da SEA associado a qualquer modelo com $m < k$ variáveis preditoras já encontrado, não será preciso examinar os subconjuntos restantes.

Para facilitar o uso do algoritmo, foi usado um arranjo, para todos os possíveis subconjuntos, ao qual deu-se o nome de árvore. Esta árvore fará com que o esforço computacional seja menor. Assim sendo, para conseguir este objetivo nós devemos observar que:

(i) A factibilidade dual de (P1) é mantida ou perdida quando uma ou mais variáveis são retiradas (ou adicionadas) de um subconjunto, e quando o problema dual for factível o teste de pré-otimalidade e de sondagem pode ser aplicado antes de cada iteração.

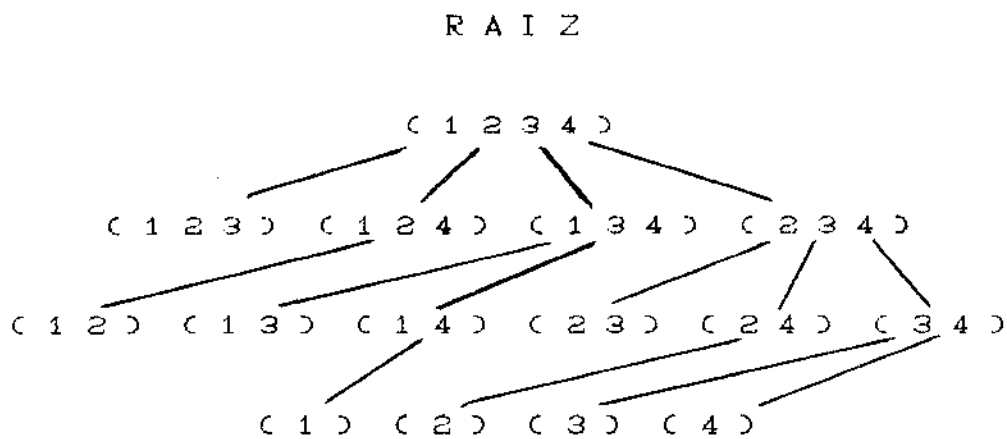
(ii) Devido a observação (2) se o valor da MSEA para um subconjunto m é maior do que a MSEA para um subconjunto $r < m$, então os $r, r+1, \dots, m-1$ subconjuntos que podem ser obtidos das m variáveis não precisam ser examinados. É desejável relacionar os subconjuntos entre si de tal forma que diversos conjuntos sejam implicitamente examinados, se o teste de sondagem tiver sucesso.

A árvore foi construída da seguinte maneira: partimos do modelo completo que é chamado a raiz da árvore com suas ramificações e seus respectivos nós, que são os subconjuntos das variáveis preditoras.

As seguintes relações entre os subconjuntos são feitas:

se o subconjunto de tamanho $m-1$ é obtido de um subconjunto de tamanho m , então o subconjunto de tamanho m é chamado de pai, e o de tamanho $m-1$ é chamado de filho ou sobrinho, e todos os subconjuntos de tamanho $m-1$ do mesmo subconjunto de tamanho m são chamados de irmãos ou primos. A árvore é construída de tal maneira que um subconjunto de variáveis de mesmo tamanho não seja repetido.

Como exemplo vamos construir a árvore para um modelo com 4 variáveis preditoras:



onde nós temos que:

(i) A raiz (1 2 3 4) é pai dos nós: (1 2 3), (1 2 4), (1 3 4) e (2 3 4), que são irmãos.

(ii) Os nós (1 3 4), (2 3 4) e (1 2 4) são pais dos nós: [(1 3), (1 4)]; [(2 3), (2 4), (3 4)] e (1 2) respectivamente, onde [(1 3) e (1 4)] são irmãos e primos de [(2 3), (2 4) e (3 4)] e (1 2).

(iii) E os nós (1 4), (2 4), (3 4) são pais de: (1), (2) e [(3), (4)] respectivamente, que são irmãos e primos análogo a (ii). Os nós (1), (2), (3) e (4) são chamados de nós terminais.

4.2.1 - SISTEMA DE ROTULAÇÃO

Para podermos aplicar algum dos métodos de seleção de variáveis proposto aqui, foi desenvolvido um sistema de rotulação de tal maneira que não examinaremos um rótulo duas vezes. A rotulação é feita em cada nó da árvore com os inteiros $1, 2, 3, \dots, k$ e um ponto a esquerda dos números, ou seja, $.1 \ 2 \ \dots \ k$. As variáveis que estão no modelo são: aquelas que ficam à direita do ponto (não sublinhadas) e as que ficam à esquerda do ponto, mas que estão sublinhadas.

Se aplicarmos o método passo a frente (de pai para filho) para um nó, teremos o seguinte procedimento: move-se o ponto um espaço para direita, um nó terminal terá sido pesquisado se todos os números estão a esquerda do ponto e sendo assim nenhum passo a frente pode ser mais executado.

Para o método passo atrás (de filho para pai) o procedimento será: escolhido um número a direita de um número sublinhado e(ou) à esquerda do ponto teremos um passo atrás. Se todos os inteiros estão a esquerda do ponto e estão sublinhados, a raiz da árvore foi pesquisada, e nenhum passo atrás pode ser executado.

4.3 - SUGESTÕES PARA ACELERAR O ALGORITMO DE ENUMERAÇÃO IMPLÍCITA

Apresentaremos aqui algumas propostas para acelerar o algoritmo de enumeração implícita, descrito em 4.2, que ajudará a selecionar um modelo mais criteriosamente.

4.3.1 - MODELOS "QUASE-ÓTIMOS" DE TAMANHO m

São modelos com $m(=1, \dots, k-1)$ variáveis preditoras que aproximam-se do melhor modelo com m variáveis preditoras, estes modelos foram chamados $(1 - \delta)$ - "quase ótimos" que são definidos

COMO:

Um modelo será $(1-\delta)$ - "quase ótimo" com $m(=1, \dots, k-1)$ variáveis preditoras dentre os $\binom{k}{m}$ possíveis modelos, se o valor da MSEA difere do melhor modelo com m variáveis preditoras não mais do que δ por cento, com $\delta \geq 0$.

Em outras palavras podemos dizer que um modelo com m variáveis preditoras é $(1-\delta)$ - "quase ótimo" se o valor da MSEA, Z_m , é menor ou igual a $(1+\delta)Z(m)$.

Um algoritmo para encontrar o modelo $(1-\delta)$ - "quase ótimo" com $m(=1, \dots, k-1)$ variáveis preditoras terá o seguinte critério:

Se o valor da MSEA, Z_m , é maior do que $Z_r/(1+\delta)$ para algum $r < m$, os modelos $r, r+1, \dots, m-1$ variáveis preditoras não precisam ser examinados.

Esta última regra acelera o algoritmo, tornando o processo de seleção mais rápido.

4.3.2 - MODELOS "IGUALMENTE-BONS".

Para alguns casos onde o número de variáveis preditoras é muito grande, pode surgir o problema da multicolinearidade entre as variáveis, fazendo com que a contribuição de algumas variáveis na regressão L_1 seja muito pequena, e o valor da MSEA para um modelo com poucas variáveis não vai diferir muito do modelo completo $\{Z(k)\}$. Um modelo que tem o valor da MSEA próximo de $Z(k)$ será chamado de modelo $(1-\delta)$ - "igualmente bom".

Um modelo será $(1-\delta)$ - "igualmente bom" se o valor da

MSEA for menor ou igual a $(1+\delta)Z(k)$, onde $\delta \geq 0$.

O algoritmo para este caso será baseado na regra de sondagem com uma pequena modificação, que é:

Se o valor da MSEA, Z_m , for maior do que $Z_r = (1+\delta)Z(k)$ para algum $r < m$, os modelos com $r, r+1, \dots, m-1$ variáveis preditoras não precisam ser examinados.

A modificação no algoritmo acelera o processo computacional diminuindo o número de modelos a serem examinados.

O δ será escolhido de acordo com problema que se está trabalhando à critério do pesquisador.

4.3.3 - LIMITE PARA O NÚMERO DE VARIÁVEIS PREDITORAS

Quando trabalhamos com um modelo cujo número de variáveis preditoras é muito grande o ideal é determinar um limite inferior e superior, ou seja, um intervalo tal que o número m ($=1, \dots, k-1$) esteja entre $L < m < U$, onde L = limite superior e U = limite inferior. Trabalhando com este intervalo será mais fácil analisar o modelo e não haverá perda de informações importantes. A idéia do limite superior (U) para o número de variáveis preditoras foi mencionada por Lamotte e Hocking (1970) e Furnival e Wilson (1976).

O limite para o número de variáveis preditoras pode partir do investigador. Uma sugestão inicial seria os métodos de seleção e eliminação de variáveis "passo a frente" e "passo atrás".

Quando reduzimos o número de variáveis preditoras nós não só diminuimos o esforço computacional como também o número de modelos a serem escolhidos.

CAPÍTULO 5

ROBUSTEZ

5.1 - INTRODUÇÃO

Métodos robustos em regressão tem sido objeto de estudo há longo tempo e desenvolvidos de tal forma que não são muito afetados por desvios com relação às hipóteses sobre a distribuição dos erros ou pela presença de valores aberrantes, isto é, acomodam a possibilidade de presença desses valores.

A regressão L_1 é considerada por vários autores uma alternativa robusta para o caso em que valores aberrantes estão presentes. Mostrou-se que o estimador L_1 é menos sensível a valores aberrantes, ou seja, produz estimativas mais resistentes do que o estimador de mínimos quadrados. Por exemplo, Huber (1974) afirmou que com relação às estimativas da regressão L_p , "p = 1" proporciona robustez do ponto de vista técnico, isto é, resistência com relação a valores aberrantes arbitrários.

Sposito, Smith e McCornick (1978) mostraram que a estimativa L_1 passa pelo menos por k observações quando é única e quando existem infinitas soluções então existe pelo menos uma solução que passa por k observações. Essas observações são denominadas explicativas, enquanto que as outras com resíduo diferente de zero serão chamadas de não-explicativas. Se existirem exatamente k observações com resíduos iguais a zero, então o estimador L_1 é completamente determinado por estas observações, não sendo afetado pelo restante das observações.

Cabe aqui observar que, a relação que existe entre a regressão L_1 e a regressão de mínimos quadrados é a mesma que existe entre a mediana e a média amostrais. Por exemplo nós vimos que tanto a média amostral como o estimador de mínimos quadrados é influenciado por todas as observações, enquanto que a mediana amostral e o estimador L_1 são determinados somente por um subconjunto de observações. Sendo assim, a mediana amostral não será afetada pelo restante das observações mesmo que modifiquemos uma delas, desde que elas permaneçam com o mesmo sinal de resíduo com relação a mediana amostral. Um resultado análogo ocorre na regressão L_1 , ou seja, o valor do estimador L_1 não é alterado quando mudamos os valores da variável resposta associados com as observações não-explicativas e estas permanecem com o mesmo sinal dos resíduos.

Com relação a estes fatos faremos um estudo, neste capítulo, baseado nos trabalhos de Narula e Wellington (1985), que mostraram a existência de um intervalo para o valor da variável resposta y , em relação a i -ésima observação não-explicativa, tal que se mudarmos y_i dentro deste intervalo a regressão L_1 não se altera para esse conjunto de dados. Um procedimento similar para as observações explicativas também foi desenvolvido, ou seja, um intervalo para y_i , tal que o conjunto de variáveis explicativas não se altere.

Nesta mesma linha, Ha e Narula (1986) desenvolveram um algoritmo para determinar a máxima perturbação, ρ^* , que o vetor Y pode sofrer independentemente e simultaneamente sem alterar o conjunto das observações explicativas. Por exemplo, se a análise revela $\rho^* = 0.5$, isto é, o valor da variável Y pode mudar $\pm 50\%$ sem alterar o vetor de parâmetros $\tilde{\beta}$. Esse estudo foi baseado no trabalho de Wendell (1984) que introduziu uma certa tolerância para o estudo de análise de sensibilidade dentro do problema de programação linear. No nosso caso usaremos a expressão análise de perturbação no lugar de tolerância.

5.2 - ANÁLISE DE SENSIBILIDADE

5.2.1 - REGRESSÃO LINEAR SIMPLES

Seja Y_i a variável resposta e X_i a variável preditora, $i=1, \dots, n$, no modelo de regressão simples

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i=1, \dots, n \quad (5.1)$$

onde β_0 e β_1 são parâmetros desconhecidos e os ε_i 's são os erros aleatórios.

Para o modelo de regressão linear simples nós vimos anteriormente que a regressão L_1 passa, no mínimo, através de duas das n observações e a diferença entre o número de observações acima (n^+) e o número de observações abaixo (n^-) é menor ou igual a dois (pequenas variações nas respostas dessas observações podem alterar substancialmente a estimativa de β).

Assumindo que a regressão L_1 passa exatamente por dois pontos: (x_p, y_p) e (x_q, y_q) , então, os estimadores para β_0 e β_1 são dados por

$$\tilde{\beta}_0 = y_p - \beta_1 x_p = y_q - \beta_1 x_q \quad (5.2a)$$

o

$$\tilde{\beta}_1 = \frac{(y_p - y_q)}{(x_p - x_q)} \quad (5.2b)$$

Podemos também encontrar o estimador da regressão L_1 , em relação ao modelo 5.1, através do algoritmo do Karst (1958), que será apresentado a seguir e no qual usaremos os seus resultados para fazer a análise de sensibilidade.

5.2.2 - ALGORITMO DO KARST

Para encontrar β_0 e β_1 definimos:

$$Z_i(t) = \begin{cases} \frac{y_i(t)}{x_i(t)} = \frac{(y_i - y_t)}{(x_i - x_t)}, & t = p, q; i = 1, \dots, n, i \neq t \\ \text{n\~{a}o definido,} & \text{para } i = t \end{cases} \quad (5.3)$$

ordenando, ent\~{a}o, os valores de $Z_i(t)$ em ordem n\~{a}o-decrescente:

$$Z_{i_1}(t) \leq Z_{i_2}(t) \leq \dots \leq Z_{i_n}(t), \quad t = p, q. \quad (5.4)$$

Definimos:

$$S_r(t) = -\sum_{i=1}^r |x_{i_1}(t)| + 2\sum_{k=1}^{n-r} |x_{i_k}(t)|, \quad t = p, q \quad (5.5)$$

a soma $S_r(t)$ determina o valor de $1 \leq r \leq n$, para todo t , tal que

$$S_{r-1}(t) < 0 \text{ e } S_r(t) > 0 \quad (5.6)$$

e o valor de r determina o ponto (x, y) onde $\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i|$ \u00e9 m\u00ednimo.

Seja u o valor de r para o qual as desigualdades acima s\~{a}o satisfeitas para $t = p$; e v o valor de r para $t = q$. Ent\~{a}o a observa\u00e7\~{a}o que corresponde ao posto u , para $t = p$, \u00e9 a observa\u00e7\~{a}o q e a observa\u00e7\~{a}o que corresponde ao posto v , para $t = q$, \u00e9 a

observação p. Então a estimativa $\tilde{\beta}_1$ é dada por:

$$\tilde{\beta}_1 = Z_{iu}(p) = Z_q(p)$$

ou

(5.7)

$$\tilde{\beta}_1 = Z_{iv}(q) = Z_p(q) = \frac{(y_p - y_q)}{(x_p - x_q)}$$

Baseado no desenvolvimento acima apresentado, foi construído um algoritmo para encontrar intervalos para a variável resposta y e para a variável preditora x, tal que a regressão L_1 não mude e as observações explicativas permaneçam as mesmas.

5.2.3 - INTERVALO PARA VARIÁVEL RESPOSTA Y

Um intervalo para a variável resposta será construído da seguinte maneira: para cada observação i não-explicativa teremos um intervalo para y_i , sem alterar a regressão L_1 , usando $t=p$ ou $t=q$. Se para alguma observação i, o valor de $Z_i(p)$ é menor (ou maior) do que $\tilde{\beta}_1$, então as estimativas $\tilde{\beta}_0$ e $\tilde{\beta}_1$ não mudam desde que o valor de y_i mude tal que o novo $Z_i^*(p)$ continue menor (ou maior) do que $\tilde{\beta}_1$, ou seja, o intervalo é obtido resolvendo-se as desigualdades:

$$(y_i^* - y_p) / (x_i - x_p) \leq \tilde{\beta}_1 \quad (5.8a)$$

ou

$$(y_i^* - y_p) / (x_i - x_p) \geq \tilde{\beta}_1 \quad (5.8b)$$

onde y_i^* é o novo valor de y_i .

Entretanto, se considerarmos as observações explicativas p ou q, os valores de $\tilde{\beta}_0$ e $\tilde{\beta}_1$ mudam para qualquer alteração em y_p ou y_q . Para que as observações p e q continuem sendo explicativas, nós mudaremos os valores de y_p (ou y_q) tal que a observação p (ou q) continue a corresponder ao ponto u ou v satisfazendo as desigualdades (5.6). Para a observação p isto ocorre se o novo valor y_p^* é tal que $Z_p^*(q)$ esteja no intervalo $[Z_{i(v-1)}(q); Z_{i(v+1)}(q)]$. O valor de $\tilde{\beta}_1^* = Z_p^*(q)$ e $\tilde{\beta}_0^* = y_q^* - \tilde{\beta}_1^* x_q$. O mesmo raciocínio segue para a observação q.

5.2.4 - INTERVALO PARA UM VALOR DA VARIÁVEL PREDITORA X

Construiremos um intervalo para a variável preditora X com relação a cada observação i, tal que a regressão L_1 não mude ou as observações explicativas permaneçam as mesmas. Qualquer mudança em um valor da variável X altera os valores de $Z_i(t)$ e $S_i(t)$.

Se considerarmos uma observação não-explicativa i, $Z_i(t) \leq \tilde{\beta}_1$, então x_i^* , o novo valor de x_i será tal que

$$-\sum_{t=1}^n |x_i(t)| - |x_i - x_i^*| + |x_i^* - x_i| + 2 \sum_{k=1}^{r-1} |x_{i_k}(t)| < 0 \quad (5.9a)$$

$$-\sum_{t=1}^n |x_i(t)| - |x_i - x_i^*| + |x_i^* - x_i| + 2 \sum_{k=1}^r |x_{i_k}(t)| \geq 0 \quad (5.9b)$$

$$e \quad y_i(t)/(x_i^* - x_i) \leq \tilde{\beta}_1 \quad (5.9c)$$

Para $Z_i(t) \geq \tilde{\beta}_1$, x_i^* será tal que

$$-\sum_{k=1}^n |x_i(t)| + |x_i - x_i| - |x_i^* - x_i| + 2\sum_{k=1}^{i-1} |x_{i_k}(t)| < 0 \quad (5.10a)$$

$$-\sum_{k=1}^n |x_i(t)| + |x_i - x_i| - |x_i^* - x_i| + 2\sum_{k=1}^i |x_{i_k}(t)| \geq 0 \quad (5.10b)$$

$$e \quad y_i(t)/(x_i^* - x_i) \geq \tilde{\beta}_1 \quad (5.10c)$$

Os valores x_i^* que satisfazem as desigualdades acima são tais que a regressão L_1 não muda.

Considerando agora uma observação explicativa, as estimativas $\tilde{\beta}_0$ e $\tilde{\beta}_1$ mudam, entretanto para que a observação p continue como explicativa, x_p^* , o novo valor de x_p , deve satisfazer as seguintes desigualdades:

$$-\sum_{k=1}^n |x_i(q)| - |x_p - x_q| + |x_p^* - x_q| + 2\sum_{k=1}^{i-1} |x_{i_k}(q)| < 0 \quad (5.11a)$$

$$-\sum_{k=1}^n |x_i(q)| - |x_p - x_q| + |x_p^* - x_q| + 2\sum_{k=1}^i |x_{i_k}(q)| \geq 0 \quad (5.11b)$$

$$e \quad y_i(q)/(x_p^* - x_q) \in [Z_{i(v-1)}(q); Z_{i(v+1)}(q)] \quad (5.11c)$$

Para que a observação q permaneça explicativa o procedimento será similar, ou seja;

$$y_i(p)/(x_q^* - x_p) \in [Z_{i(u-1)}(p); Z_{i(u+1)}(p)] \quad (5.11c)$$

5.3 - ANÁLISE DE SENSIBILIDADE PARA O MODELO DE REGRESSÃO MÚLTIPLA.

Estudaremos agora a análise de sensibilidade para a regressão múltipla, observando o quanto o vetor Y pode mudar sem modificar o hiperplano de regressão L_1 ou então sem mudar o conjunto de variáveis explicativas.

Seja $Y = X\beta + e$ um modelo linear múltiplo já apresentado anteriormente no capítulo 2. A regressão L_1 pode ser escrita para este modelo da seguinte maneira:

$$\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \\ Y_{(3)} \end{bmatrix} = \begin{bmatrix} X_{(1)} & 0 & 0 \\ X_{(2)} & I_{(2)} & 0 \\ X_{(3)} & 0 & -I_{(3)} \end{bmatrix} \cdot \begin{bmatrix} \tilde{\beta} \\ e_{(2)}^+ \\ e_{(3)}^- \end{bmatrix} \quad (5.12)$$

onde os subscritos (1), (2) e (3) referem-se às observações com resíduos zero, positivos e negativos, respectivamente. $I_{(2)}$ e $I_{(3)}$ são matrizes identidades com dimensões que são iguais ao número de resíduos positivos e negativos, respectivamente. De (5.12) nós temos que:

$$\tilde{\beta} = X_{(1)}^{-1} Y_{(1)}$$

$$e_{(2)}^+ = Y_{(2)} - X_{(2)} X_{(1)}^{-1} Y_{(1)} \quad (5.13)$$

$$e_{(3)}^- = -Y_{(3)} + X_{(3)} X_{(1)}^{-1} Y_{(1)}$$

Do sistema (5.13) nós podemos escrever:

$$\begin{bmatrix} \tilde{Y}^{(1)} \\ \tilde{Y}^{(2)} \\ \tilde{Y}^{(3)} \end{bmatrix} = \begin{bmatrix} I \\ X_{(2)} X_{(1)}^{-1} \\ X_{(3)} X_{(1)}^{-1} \end{bmatrix} \cdot Y_{(1)} = \begin{bmatrix} G_{(1)} \\ G_{(2)} \\ G_{(3)} \end{bmatrix} \cdot Y_{(1)} \quad (5.14)$$

onde I é uma matriz identidade ($k \times k$) e a matriz G particionada em $G_{(1)} = I$, $G_{(2)}$ e $G_{(3)}$, se assemelha em alguns aspectos à matriz $H = [h_{ij}]$ na regressão de mínimos quadrados, onde $H = X(X'X)^{-1}X'$ e representa a projeção ortogonal de Y sobre Y -estimado no espaço de X . Os valores grandes de h_{ii} representam pontos influentes. Podemos observar que $G_{(1)} = I$, isto é, para observações explicativas $g_{ii} = 1$, logo, na terminologia da regressão de mínimos quadrados esses valores tem ponto de ruptura grande.

Dentro desse estudo foi usado o estimador $\tilde{\beta}$ de β para determinar a quantidade máxima que a variável Y pode ser modificada sem afetar a regressão L_1 .

Para estudar o efeito da perturbação na mudança da variável Y , vamos considerar o seguinte Problema de Programação Linear Perturbado (PPLP):

$$\min 1'e^+ + 1'e^-$$

Sujeito a

$$X\tilde{\beta} + e^+ - e^- = Y + Y^0\delta \quad (PLP)$$

$$e^+, e^- \geq 0, \tilde{\beta} \text{ não-restrito no sinal.}$$

onde Y^0 é uma matriz diagonal ($n \times n$), com elementos na diagonal Y_i^0 e δ é um vetor ($n \times 1$), usado para determinar a maior perturbação que o vetor Y pode sofrer simultaneamente e independentemente

dentro do (PLP). A matriz Y^0 servirá como peso para o vetor de perturbação δ . Usaremos aqui $\|\delta\| = \max\langle |\delta_i|, i=1, \dots, n \rangle = \rho^*$. Logo o lado direito da igualdade em (PLP) pode ser escrito como $y_i + y_i^0 \rho^*$.

O intervalo para a variável Y pode ser construído, para cada observação i , tal que as observações explicativas não se alterem, quando tomarmos em $y_i + y_i^0 \rho^*$, $y_i^0 = 1$ para todo i . Se tomarmos $Y_{(1)}^0 = 0$, $Y_{(2)}^0 = 1$ e $Y_{(3)}^0 = 1$ em (PLP) poderemos encontrar o quanto cada componente $y_{(i)}$, $i=2,3$ pode variar sem alterar a regressão L_1 . Considerando-se o problema (PLP), pode-se resolvê-lo da seguinte maneira:

$$\tilde{\beta}^0 = X_{(1)}^{-1} (Y_{(1)} + Y_{(1)}^0 \delta_{(1)}) \quad (5.15a)$$

$$v_{(2)}^+ = Y_{(2)} + Y_{(2)}^0 \delta_{(2)} - X_{(2)} X_{(1)}^{-1} (Y_{(1)} + Y_{(1)}^0 \delta_{(1)}) \quad (5.15b)$$

$$v_{(3)}^- = -Y_{(3)} - Y_{(3)}^0 \delta_{(3)} + X_{(3)} X_{(1)}^{-1} (Y_{(1)} + Y_{(1)}^0 \delta_{(1)}) \quad (5.15c)$$

onde $\tilde{\beta}^0$ é o estimador L_1 de β para o problema perturbado. Os vetores $v_{(2)}^+$ e $v_{(3)}^-$ serão os novos resíduos para o problema perturbado correspondendo às n observações nos conjuntos (2) e (3), ou seja, relativo as observações que estão associadas com os resíduos positivos e negativos respectivamente. Como a regressão não se altera quando os valores de y_i são modificados tal que os resíduos permanecem com mesmo sinal, enquanto $v_{(2)}^+$ e $v_{(3)}^-$ forem maior ou igual a zero em (5.15a) e (5.15b), as observações explicativas continuam as mesmas.

O estimador L_1 é determinado pelo mesmo conjunto de observações para o problema perturbado. As condições $v_{(2)}^+ \geq 0$ e $v_{(3)}^- \geq 0$ podem ser estabelecidas como

$$X_{(2)} X_{(1)}^{-1} Y_{(1)}^{\circ} \delta_{(1)} - Y_{(2)}^{\circ} \delta_{(2)} \leq Y_{(2)} - X_{(2)} X_{(1)}^{-1} Y_{(1)} \quad (5.16a)$$

$$-X_{(3)} X_{(1)}^{-1} Y_{(1)}^{\circ} \delta_{(1)} + Y_{(3)}^{\circ} \delta_{(3)} \leq -Y_{(3)} + X_{(3)} X_{(1)}^{-1} Y_{(1)} \quad (5.16b)$$

As desigualdades acima podem ser escritas da seguinte forma:

$$B\delta \leq b \quad (5.17)$$

onde B é uma matriz $(n-k) \times n$ dada por:

$$B = \begin{bmatrix} X_{(2)} X_{(1)}^{-1} Y_{(1)}^{\circ} & -Y_{(2)}^{\circ} & 0 \\ -X_{(3)} X_{(1)}^{-1} Y_{(1)}^{\circ} & 0 & Y_{(3)}^{\circ} \end{bmatrix} \quad (5.18a)$$

e

$$b = \begin{bmatrix} Y_{(2)} - X_{(2)} X_{(1)}^{-1} Y_{(1)} \\ -Y_{(3)} + X_{(3)} X_{(1)}^{-1} Y_{(1)} \end{bmatrix} = \begin{bmatrix} \varepsilon_{(2)}^+ \\ \varepsilon_{(3)}^- \end{bmatrix} \quad (5.18b).$$

Como $\varepsilon_{(2)}^+, \varepsilon_{(3)}^- \geq 0$, podemos supor que $b \geq 0$. Como Y representa as observações numa variável resposta, nós podemos ter uma informação a priori que os valores do vetor Y são não-negativos ou que eles estão dentro de um intervalo. Além disso, considerando um caso geral, sem nenhuma perda de generalidade, é preciso que δ pertença a um subconjunto Δ de \mathbb{R}^n . Porém, nos problemas práticos, nós podemos definir $\Delta = \Delta(L, U)$ como sendo um intervalo, onde L e U representam o limite superior e inferior respectivamente para o valor de δ . Estes limites podem ser baseados no conhecimento da

posição dos valores de Y.

Os resultados obtidos por Narula e Wellington (1985) podem ser vistos como um caso particular. Por exemplo: seja $\Delta = \mathbb{R}^n$, $Y_{(1)}^o = 0$ e $Y_{(3)}^o = 0$, $Y_{(2)}^o = 1$ para um certo i dentro do conjunto (2) e todos os outros valores são iguais a zero, então a condição para que a MSEA e MSEA-PERTURBADA tenham a mesma regressão L_1 é:

$$(\delta_{(2)})_i \geq -(\epsilon_{(2)}^+)_i \quad (5.19)$$

Ha e Narula (1986) e Wendell (1984) mostraram que o valor de ρ_i^* ($i=1, \dots, n-k$) é finito se e somente se existe um $\delta \in \Delta$ tal que a i-ésima linha de $(B\delta)_i$ seja maior do que b_i . Do mesmo modo, se $\rho_i^* < \infty$, também satisfaz a desigualdade $(B\delta)_i \leq b_i$, onde $\|\delta\| \leq \rho$. Sendo assim o conjunto de todos $\rho_i \geq 0$ está no intervalo fechado $[0, \rho_i^*]$. Similarmente, o conjunto de todos $\rho \geq 0$ que satisfaz a desigualdade $B\delta \leq b$ está no intervalo fechado $[0, \rho^*]$. Então,

$$[0, \rho^*] = \bigcap_{i=1}^{n-k} [0, \rho_i^*] \quad (5.20a)$$

em que

$$\rho^* = \min_{(i=1, \dots, n-k)} \rho_i^* \quad (5.20b)$$

5.4 - ANÁLISE DE OBSERVAÇÕES

A análise de observações pode ser útil para analisar a influência de cada observação. Ela é feita tirando-se cada observação do conjunto de dados e analisando a sua influência sobre as estimativas da regressão L_1 . Para sabermos o quanto a observação está influenciando a regressão observamos as diferenças

$\tilde{\beta}_j - \tilde{\beta}_j(i)$, para cada observação i , onde $\tilde{\beta}_j$ e $\tilde{\beta}_j(i)$ denotam as estimativas do j -ésimo coeficiente com todas as observações e sem a i -ésima observação, respectivamente. Se o principal objetivo da análise de regressão é seu uso para predição, as diferenças $e_i - e_{(i)}$, $i = 1, \dots, n$, onde e_i e $e_{(i)}$ representam os resíduos com e sem a i -ésima observação, respectivamente, podem ser úteis para analisar essa influência.

Se as diferenças $\tilde{\beta}_j - \tilde{\beta}_j(i)$ ou $e_i - e_{(i)}$, forem muito grandes então podemos dizer que a i -ésima observação influencia a regressão.

Gráficos dos resíduos e_i e $e_{(i)}$ e das diferenças $\tilde{\beta}_j - \tilde{\beta}_j(i)$ e $e_i - e_{(i)}$ versus a observação i podem auxiliar na análise das observações influentes.

Dentro do contexto: análise de observações Daniel e Wood (1971) propuseram algumas sugestões que podem ajudar muito na análise de observações. Como por exemplo, para analisar: (i) o efeito de valores aberrantes, dentro de um conjunto de dados, que podem ou não ser influentes; (ii) o efeito da diferença do erro estimado de duas observações que estão próximas.

Estas análises são feitas usando algumas estatísticas que mostram a influência que cada observação pode ter nos dados. É um processo que exige vários passos, já que vai ser preciso fazer comparações dos parâmetros do modelo, com e sem aquelas observações que são consideradas influentes.

CAPÍTULO 6

ANÁLISE DE RESÍDUOS

6.1 - INTRODUÇÃO

Neste capítulo será feito um estudo sobre análise de resíduos, que é útil para analisar a validade do modelo e pressuposições estabelecidas sobre o modelo.

Este capítulo foi dividido em três seções:

- (i) heteroscedasticidade;
- (ii) valores influentes e aberrantes e
- (iii) verificação do modelo.

A análise de resíduos é muito importante dentro da análise de regressão. O que o modelo de regressão não pode explicar vai para o resíduo. Se o modelo de regressão é correto, o resíduo mostrará essa tendência confirmando as suposições feitas sobre o modelo, ou pelo menos não mostrará uma suposição incorreta.

Na análise de resíduo devemos fazer uma cuidadosa inspeção de $y_i - \hat{y}_i$, depois de ter ajustado o modelo ao conjunto de dados, que ajudará a descobrir alguma anomalia nos dados, evitando assim uma má estimação dos parâmetros do modelo em estudo.

Existem várias técnicas para examinarmos os resíduos. As

mais comuns são através de gráficos, ou através do cálculo de estatísticas, que são utilizadas para detectar valores influentes ou aberrantes em um conjunto de dados.

Os gráficos de resíduos são muito úteis, principalmente para observar que tendências tem os dados, como por exemplo: linear, quadrática ou exponencial.

Os gráficos de resíduos também são úteis para detectar valores aberrantes e variância dos erros não-constante (heteroscedasticidade), verificar suposições sobre o modelo (distribuição dos erros) e determinar a importância das variáveis preditoras. Porém, a interpretação de gráficos de resíduos é algo individual que requer uma cuidadosa análise do investigador, para não tirar conclusões erradas sobre o modelo de regressão em estudo.

Os gráficos usados para a análise dos resíduos são: resíduos (\hat{e}_i) versus valores preditos (\hat{y}_i), resíduos (\hat{e}_i) versus variáveis preditoras (x_i) e o gráfico probabilístico. Nunca devemos fazer o gráfico de resíduos versus valores observados (y_i), por serem eles, geralmente correlacionados, levando-nos a conclusões erradas sobre o modelo de regressão em estudo. Através dos resíduos poderemos concluir se a análise sobre um conjunto de dados foi satisfatória ou não.

6.2 - HETEROSCEDASTICIDADE

Neste item falaremos do caso em que os elementos da diagonal principal da matriz de covariâncias não são iguais, e os elementos fora da diagonal são zeros.

Existem várias técnicas estatísticas para sabermos se os erros de um modelo de regressão tem variâncias desiguais. Uma das maneiras mais simples é construindo o gráfico de resíduos (\hat{e}_i) versus valores preditos (\hat{y}_i), este gráfico pode mostrar se existe heteroscedasticidade. Se existe homocedasticidade espera-se que haja dispersão aleatória dos pontos em torno de uma linha paralela

a abscissa do gráfico. Para detectar modelos heteroscedásticos através de gráficos de resíduos precisa-se de um número razoável de observações.

A detecção de heteroscedasticidade através da regressão baseada em quantis amostrais ou regressão quantílica, apresentada no capítulo 2, é uma sugestão de alguns autores, como por exemplo Hogg (1975), foi um dos primeiros a apresentar um exemplo, onde são usadas as regressões de percentis: 25%, 50% e 75%, para detectar a presença de heteroscedasticidade.

O problema de heteroscedasticidade pode ser resolvido fazendo transformações na variável resposta Y , como por exemplo: $1/y$, $\ln y$ e \sqrt{y} , ou usar MSEAP (Mínima Soma dos Erros Absolutos ponderados), ou seja, regressão L_1 ponderada, regressão esta que não foi citada neste trabalho. Estas transformações acima citadas podem estabilizar a variância dos resíduos, deixando-a constante.

6.3 - VALORES ABERRANTES E VALORES INFLUENTES

Um ponto é considerado valor aberrante entre os resíduos se ele está bastante afastado dos demais em valor absoluto. O valor aberrante é uma particularidade e indica um ponto não comum aos demais pontos. Ele deve ser submetido a um exame cuidadoso para determinar a sua importância em relação as outras observações, podendo ser um ponto influente, porém, o fato de uma observação ser um valor aberrante não quer dizer que ela seja influente no conjunto de dados.

Existem várias regras para a remoção de valores aberrantes, sendo depois os dados reanalisados sem estes valores aberrantes. Nem sempre será aconselhável remover valores aberrantes ou influentes sem antes fazer um séria análise nos dados, algumas vezes o valor aberrante ou influente é de vital importância para o conjunto de dados em estudo.

6.3.1 - MÉTODO PARA DETECTAR VALORES ABERRANTES ATRAVÉS DE REGRESSÕES QUANTÍLICAS

Uma maneira de identificar valores aberrantes é através das regressões quantílicas propostas por Bassett e Koenker (1978, 1982).

Duas sugestões são feitas por Portnoy (1987) para detectar valores aberrantes, baseadas em regressões quantílicas. Primeiramente espera-se que valores aberrantes sejam ajustados pelos extremos das regressões quantílicas $\tilde{\beta}(0)$ e $\tilde{\beta}(1)$. Em uma segunda sugestão podemos ver que os valores aberrantes são provavelmente ajustados exatamente por um grande número de soluções $\tilde{\beta}(\theta)$ ($0 < \theta < 1$), ou seja, para cada observação (x_i, y_i) nós temos

$$T_i = \{\theta: e_i(\tilde{\beta}(\theta)) = 0, i=1, \dots, n\} \quad (6.1)$$

onde T_i é o conjunto de valores para o qual (x_i, y_i) é exatamente ajustado por $\tilde{\beta}(\theta)$. Podemos ver que T_i é uma união de intervalos da forma: (θ_j, θ_{j+1}) , $j = 1, \dots, p-1$ ($p < n$), e $L_i = \text{comprimento}(T_i)$. Portanto, grandes valores de L_i indicam que (x_i, y_i) é um valor aberrante.

Em relação à primeira sugestão existem dificuldades, ou seja, através dela é possível construir exemplos onde um valor aberrante pode não estar posicionado entre as regressões quantílicas $\tilde{\beta}(0)$ e $\tilde{\beta}(1)$. Uma solução para este problema é perturbar aleatoriamente todos os dados com uma pequena quantia. Portanto, como os valores aberrantes estão distantes do centro dos dados, ele se posicionará na regressão quantílica $\tilde{\beta}(0)$ ou $\tilde{\beta}(1)$ perturbada (com probabilidade um). Se a perturbação é menor do que a precisão que está sendo usada para os dados, então a aleatorização não terá efeito para identificar valores aberrantes.

6.3.2 - ALGORITMO

1- Somar uma perturbação aleatória para todos os dados.

2- Identificar como potenciais valores aberrantes às observações dentro do conjunto

$$I = \{ (x_i, y_i) : e_i(\hat{\beta}(0)) = 0 \text{ ou } e_i(\hat{\beta}(1)) = 0 \} \quad (6.2)$$

3- Escolher um particular $(x^*, y^*)_1 \in I$ como o "mais afastado" e removê-lo (formando um conjunto de dados de tamanho $n-1$).

Foram sugeridas duas alternativas para escolher o ponto mais afastado, ou seja,

3.1 - O ponto $(x, y)_{j+1} \in I_j$ que está mais afastado da mediana de todos os dados, usando como medida de distância a Euclidiana.

3.2- O ponto $(x, y)_{j+1} \in I_j$ que está mais afastado do hiperplano da regressão L_1 de todos os dados, usando como medida de distância a Euclidiana.

OBS.: Na prática 3.1 tem-se mostrado melhor critério para detectar valores aberrantes. Além disso é computacionalmente mais rápido.

4 - Iterativamente repetir os passos (2) e (3) p vezes (utilizando o conjunto de dados sem as observações consideradas mais afastadas até aquela etapa) determinando os valores aberrantes $(x^*, y^*)_1, (x^*, y^*)_2, \dots, (x^*, y^*)_p$.

5- Depois de retirar as p observações, determina-se o estimador $L_1(\hat{\beta}(1/2))$, baseado nas observações que permanecem, o qual chamaremos de $\tilde{\beta}^*$.

6- Fazer o gráfico de todos os resíduos (n), em relação ao modelo de regressão do passo 5 e identificar os valores aberrantes como os resíduos que tem os maiores valores.

OBS.: O número de pontos retirados fica a critério do analista.

6.4 - VERIFICAÇÃO DO MODELO

Como tínhamos falado anteriormente, os gráficos de resíduos são muito importantes e indispensáveis na análise de regressão. Os gráficos que estudaremos são:

- (1) resíduos (\hat{e}_i) versus valores preditos (\hat{y}_i) e
- (2) resíduos (\hat{e}_i) versus variáveis preditoras (x_i).

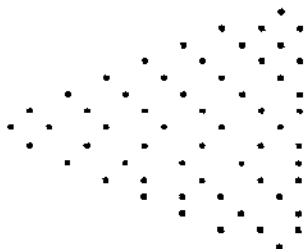
1- O gráfico de resíduos (e_i) versus valores preditos (\hat{y}_i), pode ter os seguintes comportamentos:

(1.a)

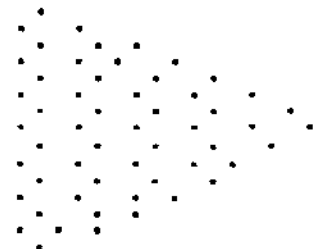


Uma faixa aleatória que não indica nenhuma tendência.

(1.b)

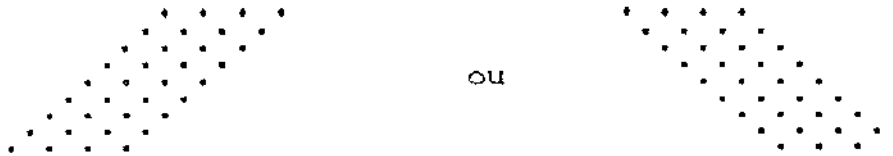


ou



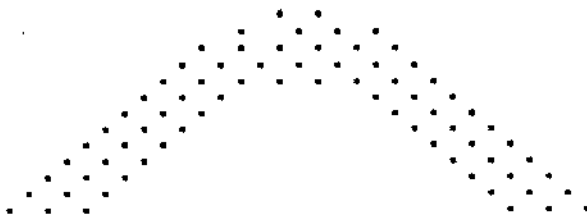
Indica variância dos erros não-constante (heteroscedasticidade), pode ser necessário ponderar os dados ou fazer transformações na variável resposta y .

(1.c)



Tendência linear no modelo (positiva ou negativa) não removida.

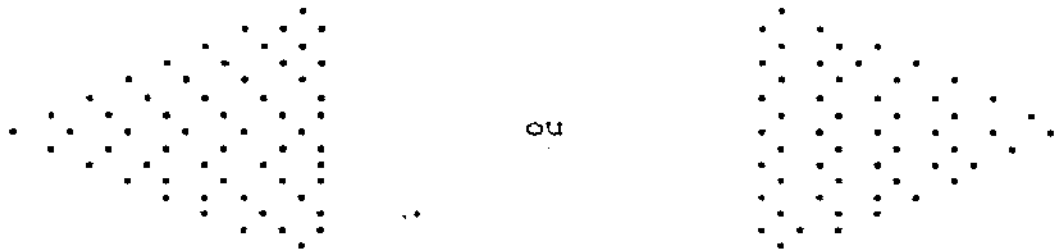
(1.d)



O Modelo necessita de um termo extra, isto é, um termo quadrático ou produto cruzado, ou uma transformação na variável resposta y .

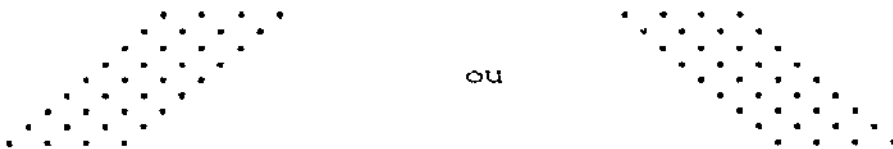
2- O gráfico de resíduos (\hat{e}_i) versus variáveis preditoras (x_{ij} , $i=1, \dots, n$; $j=1, \dots, k$), pode também ter o mesmo comportamento que o gráfico do \hat{e}_i versus \hat{y}_i , ou seja,

(2.a)



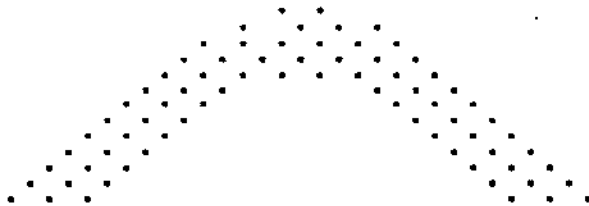
Indica variância dos erros não-constante (heteroscedasticidade); é necessário ponderar os dados ou fazer transformação na variável preditora x_j .

(2.b)



Efeito linear na variável preditora x_j (positivo ou negativo) não removido.

(2.c)



Necessita de termos extras; isto é, um termo quadrático na variável preditora x ou transformação na variável resposta y .

6.4.1 - USO DO GRÁFICO DE PROBABILIDADE NORMAL PARA DETECTAR DISTRIBUIÇÕES DE ERROS COM CAUDAS ALONGADAS

Aqui o gráfico de probabilidade normal será usado para detectar distribuições de erros com caudas mais alongadas do que a normal, caso em que a regressão L_1 é mais adequada. A fim de visualizarmos o uso do gráfico de probabilidade normal faremos a simulação de três amostras aleatórias de tamanho 200 baseada em uma distribuição que supostamente não é normal. Como exemplo utilizaremos a distribuição da Normal Contaminada e a distribuição de Laplace.

(i) Gráfico de Probabilidade Normal de uma amostra aleatória de 200 observações geradas segundo uma Normal Contaminada $0.7.N(0,1) + 0.3.N(0,25)$. Ver figura 6.1.

(ii) Gráfico de Probabilidade Normal de uma amostra aleatória de 200 observações geradas segundo uma distribuição Normal Contaminada $0.9.N(0,1) + 0.1.N(0,25)$. Ver figura 6.2.

(iii) Gráfico de Probabilidade Normal de uma amostra aleatória de 200 observações geradas segundo a distribuição de Laplace. Ver figura 6.3.

Normal Probability Plot

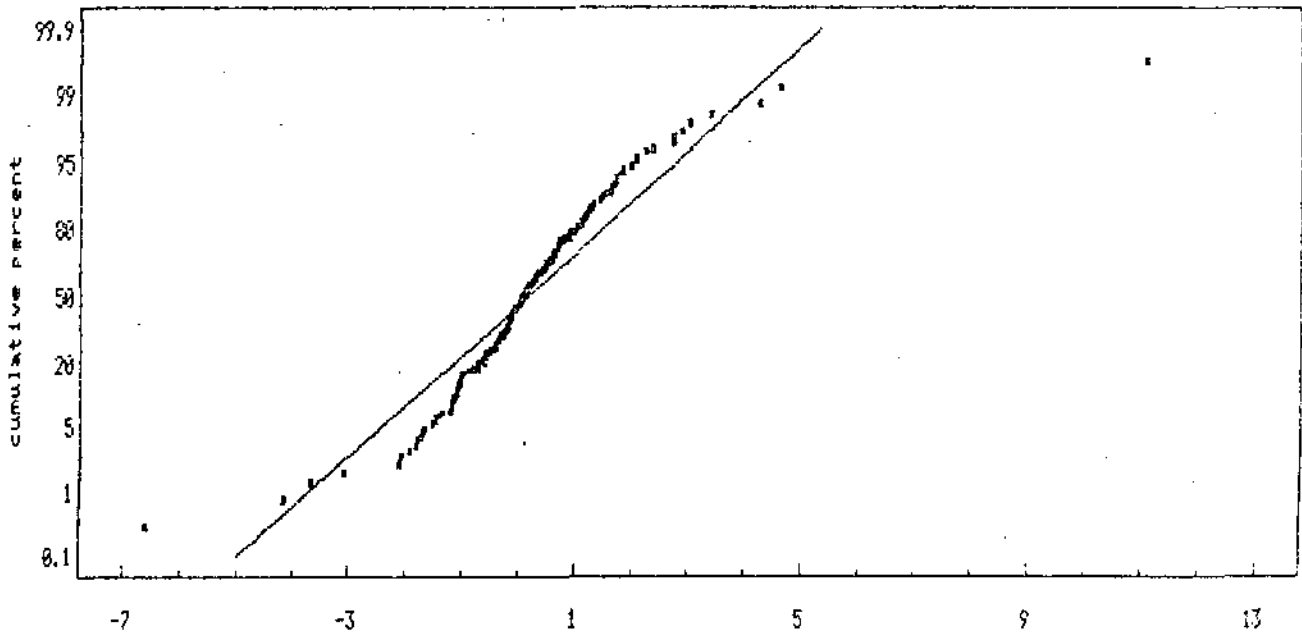


FIGURA - 8.1

Grafico Probabilístico Normal

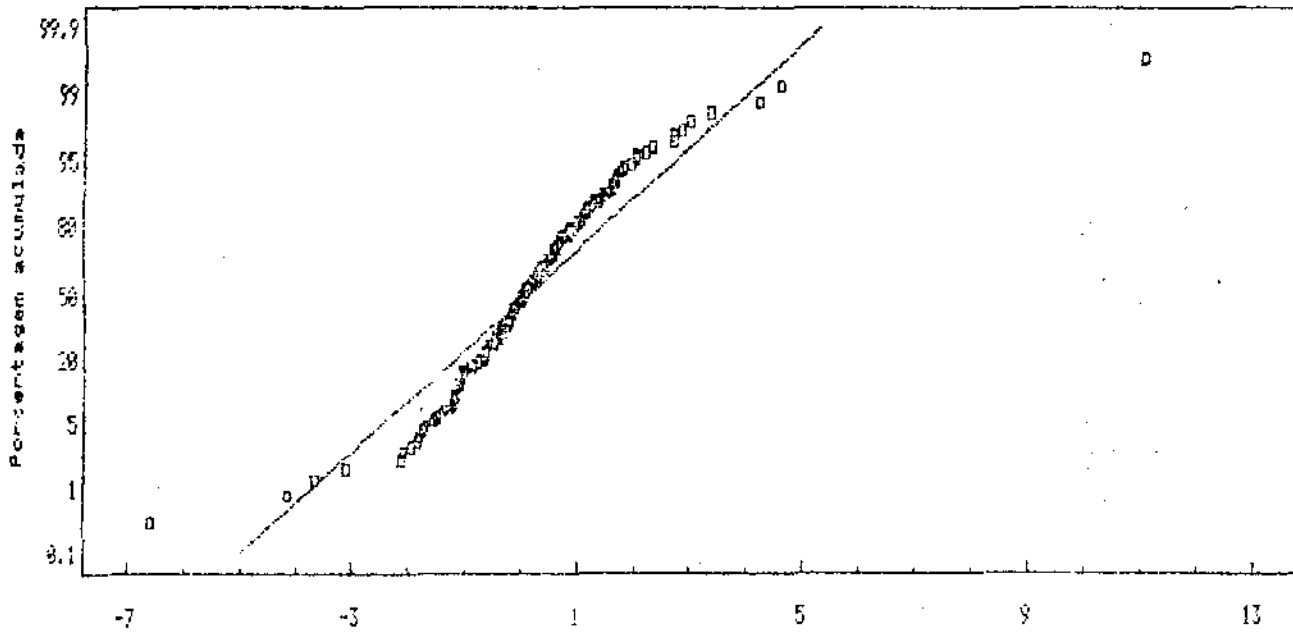


FIGURA - 6.2

Normal Probability Plot

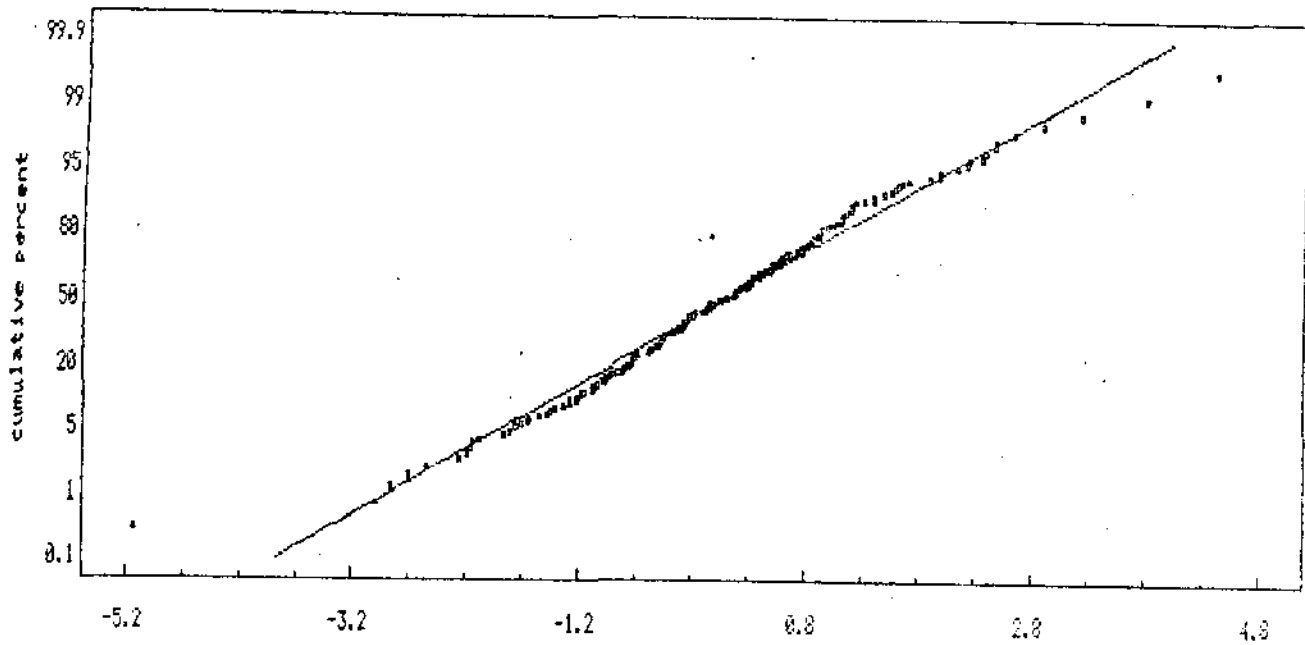


FIGURA - 6.3

Os gráficos das figuras: 6.1; 6.2 e 6.3, mostram claramente o comportamento das distribuições de erros com caudas alongadas, ou seja, estas distribuições apresentam os gráficos em forma de uma curva com as extremidades achatadas.

CAPÍTULO 7

APLICAÇÃO

7.1 - INTRODUÇÃO

Neste capítulo apresentaremos a análise de um conjunto de dados, utilizando os métodos apresentados nos capítulos anteriores, com a finalidade de ilustrar uma aplicação dos métodos de análise de regressão L_1 .

7.2 - CONJUNTO DE DADOS

Os dados aqui apresentados são referentes à várias características demográficas de 49 países, e foram retirados do livro de Gunst e Mason [(1980) - pag. 358], que estão descritos abaixo na tabela - 1:

TABELA 1 - CARACTERÍSTICAS DEMOGRÁFICAS DE VÁRIOS PAÍSES

PAÍSES	Y	X1	X2	X3	X4	X5	X6
Austrália	1316	19.5	960	1	21	98.5	856
Áustria	670	37.5	695	84	1720	98.5	546
Barbados	200	60.4	3000	548	7121	91.1	24
Bélgica	1196	35.4	819	301	5257	96.7	536
Guiana Britânica	235	67.1	3900	3	192	74.0	27

Bulgaria	365	45.1	740	72	1380	85.0	456
Canadá	1947	27.3	900	2	257	97.5	645
Chile	379	127.9	1700	11	1164	80.1	257
Costa Rica	357	78.9	2600	24	948	79.4	326
Chipre	467	29.9	1400	62	1042	60.5	78
Tchecoslováquia	680	31.0	620	108	1821	97.5	398
Dinamarca	1057	23.7	830	107	1434	98.5	570
El Salvador	219	76.3	5400	127	1497	39.4	89
Finlândia	794	21.0	1600	13	1512	98.5	529
França	943	27.4	1014	83	1288	96.4	667
Guatemala	189	91.9	6400	36	1365	29.4	135
Hong Kong	272	41.5	3300	3082	98143	57.5	176
Hungria	490	47.6	650	108	1370	97.5	258
Islândia	572	22.4	840	2	79	98.5	446
Índia	73	225.0	5200	138	2279	19.3	220
Irlanda	550	30.5	1000	40	598	98.5	362
Itália	516	48.7	746	164	2323	87.5	362
Jamaica	316	58.7	4300	143	3410	77.0	42
Japão	306	37.7	930	254	7563	98.0	750
Luxemburgo	1388	31.5	910	123	2286	96.5	36
Malásia	356	68.9	6400	54	2980	38.4	475
Malta	377	38.3	980	1041	8050	57.6	142
Maurícios	225	69.5	4500	352	4711	51.8	14
México	262	77.7	1700	18	296	50.0	258
Holanda	836	16.5	900	346	4855	98.5	923
Nova Zelândia	1310	22.8	700	9	170	98.5	839
Nicarágua	160	71.7	2800	10	824	38.4	110
Noruega	1130	20.2	946	11	3420	98.5	258
Panamá	329	54.8	3200	15	838	67.5	371
Polónia	475	74.7	1100	96	1411	95.0	351
Portugal	224	77.5	1394	100	1087	55.9	272
Porto Rico	563	52.4	2200	271	4030	81.0	1192
Roménia	360	75.7	788	78	1248	89.0	226
Singapura	400	32.3	2400	2904	108214	50.0	437
Espanha	293	43.5	1000	61	1347	87.0	258

Suécia	1380	16.6	1089	17	1705	98.5	401
Suíça	1428	21.1	765	133	2320	98.5	398
Tailândia	161	30.5	1500	305	10446	54.0	329
Trindade	423	45.4	2300	168	4383	73.8	61
Reino Unido	1189	24.1	935	217	2677	98.5	460
Estados Unidos	2577	26.4	780	20	399	98.0	1983
USSR	600	35.0	578	10	339	95.0	539
Alem. Ocidental	927	33.8	798	217	3631	98.5	528
Iugoslávia	265	100.0	1637	73	1215	77.0	524

onde Y é a variável resposta e X_i , $i = 1, \dots, 6$, são as variáveis preditoras, definidas abaixo:

- $Y \equiv$ Produto Interno Bruto (calculado em dólares - 1957)
- $X_1 \equiv$ Mortalidade Infantil por 1000 crianças nascidas vivas
- $X_2 \equiv$ Número de Habitantes para cada médico
- $X_3 \equiv$ População por km^2
- $X_4 \equiv$ População por 1000 hectares de agricultura no país
- $X_5 \equiv$ Percentagem de alfabetizados maiores de 15 anos
- $X_6 \equiv$ Estudantes matriculados no nível superior para cada 100.000 habitantes

7.3 - ANÁLISE

Inicialmente procuramos encontrar um conjunto de dados que fosse adequado para utilizar a regressão L_1 , ou seja, que tivessem distribuições dos erros não Normais. Para isso, testamos vários conjuntos de dados, nos quais foram verificados se a suposição de normalidade na distribuição dos erros não era satisfeita. Isso foi possível utilizando o pacote de procedimentos estatísticos SAS [28].

Utilizando os procedimentos acima citados no conjunto de dados da tabela - 1, foi possível então: testar a normalidade;

calcular a curtose; contruir os gráficos: ramos e folha, box-plot e de probabilidade normal, tanto para regressão de Mínimos Quadrados quanto para regressão L_1 . Os resultados referentes à regressão de Mínimos Quadrados pode ser visto nas figuras 7.1, 7.2, 7.3, 7.4 e 7.5, enquanto que os resultados referentes a regressão L_1 podem ser vistos nas figuras 7.6, 7.7 e 7.8. Observamos então que os dados eram adequados para aplicar regressão L_1 .

Logo em seguida utilizamos um programa em FORTRAN - 77, para calcular regressões quantílicas, desenvolvido por Narula e Wellington. Calculamos algumas regressões quantílicas para o conjunto de dados acima citado, ou seja, para $\theta = 0.2; 0.4; 0.5; 0.6$ e 0.8 . Através destas regressões, usando gráfico dos resíduos contra preditos, detectamos heteroscedasticidade no conjunto de dados. Mostraremos a seguir as regressões quantílicas para $\theta = 0.5; 0.2$ e 0.8 com os seus respectivos gráficos dos resíduos contra os preditos. Os resultados são mostrados abaixo:

1.a - Regressão quantílica para $\theta = 0.5$ ou regressão L_1 ;

$$\hat{Y} = 134.8374 - 3.9619X_1 + 0.0333X_2 + 0.0939X_3 - 0.0046X_4 + 3.0632X_5 + 0.8714X_6$$

MSEA = 11152.950 - Gráfico - ver figura 7.9

1.b - Regressão quantílica para $\theta = 0.2$;

$$\hat{Y} = 287,1313 - 3,2762X_1 - 0.0100X_2 - 0.1108X_3 + 0.0031X_4 + 2,7063X_5 + 0.2248X_6$$

MSEA = 6430.875 - Gráfico - ver figura 7.10

1.c - Regressão quantílica para $\theta = 0.8$;

$$\hat{Y} = 5.0999 - 0.0071X_1 + 0X_2 - 0.0001X_3 + 0X_4 + 0.0192X_5 + 0.0005X_6$$

MSEA = 9.508 - Gráfico - ver figura 7.11

Olhando os gráficos das figuras 7.9; 7.10 e 7.11 podemos notar a presença de heteroscedasticidade no conjunto de dados. Para resolver este problema usamos as transformações $1/y$, $\ln y$ e \sqrt{y} , e observamos que a melhor transformação foi $z = \ln y$, ou seja, aquela que melhor estabilizou a variância dos dados. Os resultados abaixo mostrarão as regressões quantílicas para $\theta = 0.5$; 0.2 e 0.8, com os seus respectivos gráficos de resíduos contra predito:

2.a - Regressão quantílica para $\theta = 0.5$ ou regressão L_1 , utilizando a transformação $z = \ln y$;

$$\hat{Z} = 5.8421 - 0.0099X_1 + 0X_2 - 0.0004X_3 + 0X_4 + 0.0072X_5 + 0.0008X_6$$

MSEA = 13.842 - Gráfico - ver figura 7.12

2.b - Regressão quantílica para $\theta = 0.2$, utilizando a transformação $z = \ln y$;

$$\hat{Z} = 4.8214 - 0.0075X_1 + 0.0001X_2 + 0X_3 + 0X_4 + 0.015X_5 + 0.0006X_6$$

MSEA = 10.086 - Gráfico - ver figura 7.13

2.c - Regressão quantílica para $\theta = 0.8$, utilizando a transformação $z = \ln y$;

$$\hat{Z} = 5.0999 - 0.0071X_1 + 0X_2 - 0.0001X_3 + 0X_4 + 0.0192X_5 + 0.0005X_6$$

MSEA = 9.508 - Gráfico - ver figura 7.14

A seguir fizemos uma seleção de variáveis para o modelo 2.a utilizando um programa em FORTRAN - 77, proposto por Narula e Wellington (1979) baseado no algoritmo apresentado no capítulo 4. Encontramos os melhores modelos de tamanho $p = 1, \dots, 5$ variáveis preditoras. Fizemos testes de hipótese (capítulo 3) nos parâmetros de cada modelo de tamanho $p = 1, \dots, 6$, testando a significância dos respectivos parâmetros, e comparamos cada um dos modelos um a um para verificar qual modelo seria mais adequado. Partindo daí chegamos a conclusão que o modelo mais adequado seria aquele com as variáveis X_1 , X_5 e X_6 , cujo modelo e testes de hipótese mostraremos a seguir:

3a. - Modelo escolhido para o conjunto de dados tabela - 1;

$$\hat{Z}^* = 5.2669 - 0.0061X_1 + 0.0111X_5 + 0.0008X_6$$

MSEA = 14.0425

4 - TESTES DE HIPÓTESE

4.a - Teste para hipótese $H_0: \beta_1 = 0$;

$$Z^* = | -4.8129 | = 4.8129$$

4.b - Teste para hipótese $H_0: \beta_5 = 0$;

$$Z^* = | 5.3287 | = 5.3287$$

4.c - Teste para hipótese $H_0: \beta_6 = 0$;

$$Z^* = | 7.1311 | = 7.1311$$

A probabilidade de significância para cada um dos testes acima apresentados mostrou-se inferior a 0.1%, portanto rejeitamos a hipótese $H_0: \beta_i = 0$, $i = 1, 5, 6$.

Logo em seguida utilizamos o algoritmo proposto por Portnoy (1987) apresentado no capítulo 6, para detectar possíveis valores aberrantes no conjunto de dados, através das regressões quantílicas. Fizemos vinte interações e encontramos 4 observações que consideramos valores aberrantes, ou seja, as observações: 24, 25, 37 e 46, que podem ser vistas no gráfico da figura 7.15, de resíduos contra observações ($i = 1, \dots, 49$), que foi baseado no seguinte modelo:

$$5.a - \hat{Z}^{**} = 4.8404 - 0.0058X_1 + 0.0127X_5 + 0.0016X_6$$

$$MSEA = 7.299$$

6.a - Modelo sem as observações: 24, 25, 37 e 46;

$$\hat{Z}^{***} = 5.2516 - 0.0063X_1 + 0.0097X_5 + 0.0013X_6$$

$$MSEA = 11.179$$

Fazendo uma análise no modelo 6.a, ou melhor, observando os parâmetros desse modelo e o gráfico de resíduos contra preditos (ver figura 7.16), podemos ver que são semelhantes ao do modelo 3.a (ver figura 7.17), não mostrando diferenças significativas em

relação a este modelo (3.a), resolvemos então ficar com o modelo contendo todas as observações.

Os gráficos de resíduos contra X1, X5 e X6, para o modelo 3.a são mostrados nas figuras 7.18; 7.19 e 7.20.

Escolhido o modelo 3.a construímos intervalos de confiança para os parâmetros do modelo, os quais são mostrados a seguir:

7 - INTERVALOS DE CONFIANÇA

7.a - Intervalo $(1 - \alpha)100\%$ de confiança para $\tilde{\beta}_1 = -0.0061$, com $\alpha=0.05$;

$$\left[-0.0086; -0.0036 \right]$$

7.b - Intervalo $(1 - \alpha)100\%$ de confiança para $\tilde{\beta}_5 = 0.0111$, com $\alpha=0.05$;

$$\left[0.0070; 0.0152 \right]$$

7.c - Intervalo $(1 - \alpha)100\%$ de confiança para $\tilde{\beta}_6 = 0.0008$, com $\alpha=0.05$;

$$\left[0.0006; 0.0010 \right]$$

8 - ANÁLISE DE VARIAÇÃO

Para testar se os parâmetros são conjuntamente iguais a zero, fizemos a análise de "variação" para o modelo 3.a.

Seja SAC(X1, X5, X6) a MSEA com relação ao modelo 3.a e SAC(O) a MSEA com relação ao modelo completo 2.a. A seguir testaremos a seguinte hipótese:

$$H_0: \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$Q_{L_1} = \frac{SAC(X1, X5, X6) - SAC(\Omega)}{q \times \hat{\lambda}/2} = \frac{14.026 - 13.842}{.3 \times 0.4846} = 0.1258$$

onde $\hat{\lambda}/2 = 0.4846$ é a estimativa do desvio padrão da mediana amostral dos resíduos dividido por dois.

TABELA DE ANÁLISE DE VARIACÃO PARA O MODELO 3.A

FONTE	GRAUS DE LIBERDADE	SAC(X1, X5, X6) - - SAC(Ω)	Q_{L_1}
H_0	3	0.0610	0.1258
ERRO	49-3 = 46	$\hat{\lambda}/2 = 0.4846$	
TOTAL	49		

A probabilidade de significância é de aproximadamente 2%, portanto rejeitamos H_0 .

9 - ANÁLISE DE SENSIBILIDADE PARA A VARIÁVEL RESPOSTA Z^*

Finalmente fizemos uma análise de sensibilidade para a variável resposta $z^* = \ln y$ do modelo (3.a), com base no exemplo apresentado por Ha e Narula (1987) e encontramos o valor $\rho^* = \min_{i=1, \dots, n-k} \rho_i^* = 1.7 \times 10^{-4}$, ou seja, $\rho^* \approx 0.0002$.

Considerando somente as observações não explicativas, temos os seguintes resultados:

TABELA - 2 - INTERVALOS PARA A VARIÁVEL RSPOSTA Z_i^*

i	Z_i^*	VALOR DO LIMITE INFERIOR	VALOR DO LIMITE SUPERIOR
1	6.9600	6.9598	∞
2	6.5908	$-\infty$	6.5910
3	5.9321	$-\infty$	5.9322
4	6.5753	6.5751	∞
5	5.7033	$-\infty$	5.7035
6	6.3188	$-\infty$	6.3190
7	6.7248	6.7246	∞
8	6.5919	6.5917	∞
9	5.9412	$-\infty$	5.9414
10	5.8230	5.8228	∞
11	6.4957	6.4955	∞
12	6.6953	6.6951	∞
13	5.3136	5.3134	∞
15	6.6775	$-\infty$	6.6777
16	6.7303	6.7301	∞
17	5.1454	5.1452	∞
18	5.8006	$-\infty$	5.8008
19	6.2772	$-\infty$	6.2774
21	6.5987	$-\infty$	6.5989
23	6.4798	$-\infty$	6.4800
24	5.8006	$-\infty$	5.8008
25	6.7546	$-\infty$	6.7548
26	6.1789	6.1787	∞
27	5.6704	5.6702	∞
28	5.7928	5.7926	∞
29	5.4306	$-\infty$	5.4308
30	5.5644	5.5642	∞
31	7.0344	$-\infty$	7.0346
32	6.9256	6.9254	∞
33	5.3482	$-\infty$	5.3484
34	6.4558	6.4556	∞

35	5.9738	-∞	5.9739
36	6.1615	6.1613	∞
37	5.6430	-∞	5.6432
38	6.8450	-∞	5.8452
40	5.9841	-∞	5.9843
41	6.1854	-∞	6.1856
42	6.5974	6.5972	∞
43	5.9568	-∞	5.9570
44	5.8621	5.8619	∞
45	6.6009	6.6007	∞
47	6.5613	-∞	6.5615
48	6.5984	6.5982	∞
49	5.9511	-∞	5.9513

OBS.: As observações explicativas são: 14, 20, 22, 39 e 46.

7.4 - CONCLUSÕES

Verificamos através do modelo 3.a que as variáveis X1 (Mortalidade Infantil por 1000 nascidos vivos; X5 (Percentagem de Alfabetizados Maiores de 15 anos) e X6 (Estudantes Matriculados no Nível Superior para cada 100.000 Habitantes) explicam adequadamente o modelo com relação ao Produto Interno Bruto dos países apresentados na tabela - 1, referente ao ano de 1957.

Observamos que o gráfico de resíduos contra preditos, modelo 3.a, figura - 7.17, não mostra nenhuma tendência no conjunto de dados, ou seja, o modelo está adequado.

O gráfico de resíduos contra a variável preditora X1, modelo 3.a, figura - 7.18, parece razoavelmente bom, a não ser uma observação que se destacou das outras, ou melhor, a observação 20, referente a Índia que tem o maior índice de mortalidade infantil para cada 1000 crianças nascidas vivas no ano de 1957.

O gráfico de resíduos contra a variável preditora X5

(Porcentagem de Alfabetizados Maiores de 15 Anos), modelo 3.a, figura - 7.19, apresenta uma pequena variabilidade, assim como destacamos as observações: 13,16,20,26 e 32 que tem os menores valores, ou seja, El Salvador, Guatemala, Índia, Malásia e Nicarágua respectivamente. Uma transformação logarítmica da variável X5 parece que seria adequada.

E finalmente o gráfico de resíduos contra a variável preditora X6, figura 7.20 não mostra nenhuma tendência nos dados, a não ser a observação 46, que se destacou das demais, ou melhor, esta observação é referente aos Estados Unidos, que tem o maior número de estudantes matriculados no nível superior para cada 100.000 habitantes no ano de 1957. Observamos também que esta observação 46 foi considerada um valor aberrante e é uma observação explicativa, ou seja, resíduo igual a zero.

Model: MODEL1
 Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	6	7441315.3636	1240219.2273	9.823	0.0001
Error	42	5302611.1262	126252.64586		
C Total	48	12743926.490			
Root MSE		355.32048	R-square	0.5839	
Dep Mean		647.89726	Adj R-sq	0.5245	
C.V.		54.84204			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t for H0: Parameter=0	Prob > t
INTERCEPT	1	31.190563	462.18438280	0.067	0.9465
X1	1	-3.440571	1.98619004	-1.732	0.0906
X2	1	0.014666	0.05377705	0.273	0.7864
X3	1	-0.172978	0.37440540	-0.462	0.6463
X4	1	0.003102	0.01096337	0.283	0.7786
X5	1	6.311843	4.24266734	1.488	0.1443
X6	1	0.690100	0.16639380	4.147	0.0002

FIGURA - 7.1

UNIVARIATE PROCEDURE

Variable=RS

Residual

Moments

N	49	Sum Wgts	49
Mean	0	Sum	0
Std Dev	332.3719	Variance	110471.1
Skewness	0.743572	Kurtosis	1.053851
USS	5302611	CS	5302611
CV	.	Std Mean	47.4817
T:Mean=0	0	Prob> T	1.0000
Sgn Rank	-74.5	Prob> S	0.4643
Num ^= 0	49		

Quantiles(Def=5)

100% Max	935.567	99%	935.567
75% Q3	103.395	95%	640.3938
50% Med	-67.1875	90%	489.1558
25% Q1	-186.251	10%	-334.785
0% Min	-724.784	5%	-365.512
		1%	-724.784
Range	1660.35		
Q3-Q1	289.6456		
Mode	-724.784		

FIGURA - 7.2

Extremes

Lowest	Obs	Highest	Obs
-724,784(24)	489,1558(41)
-619,655(37)	577,6179(42)
-365,512(30)	640,3938(46)
-364,893(6)	832,0888(25)
-334,788(43)	935,567(7)

Stem	Leaf	%	Boxplot
9	4	1	0
8	3	1	0
7			
6	4	1	0
5	8	1	0
4	89	2	
3	235	3	
2			
1	0356	4	+-----+
0	224888899	9	+
-0	98721	3	*-----*
-1	9877443310	10	+-----+
-2	98633	5	
-3	76322	5	
-4			
-5			
-6	2	1	
-7	2	1	0

-----+-----+-----+-----+
 Multiply Stem,Leaf by 10**+2

FIGURA - 7.3

SAS

UNIVARIATE PROCEDURE

variable=RS.

Residual

Normal Probability Plot

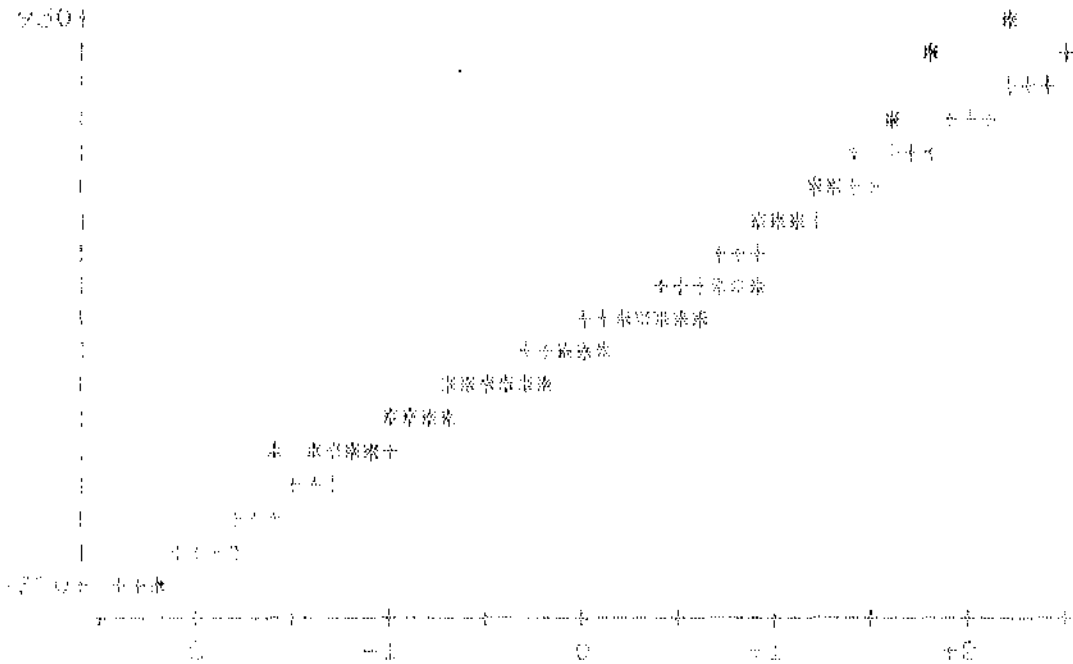


FIGURA - 7.4

Plot of RS*PREDITO. Legend: A = 1 obs, B = 2 obs, etc.

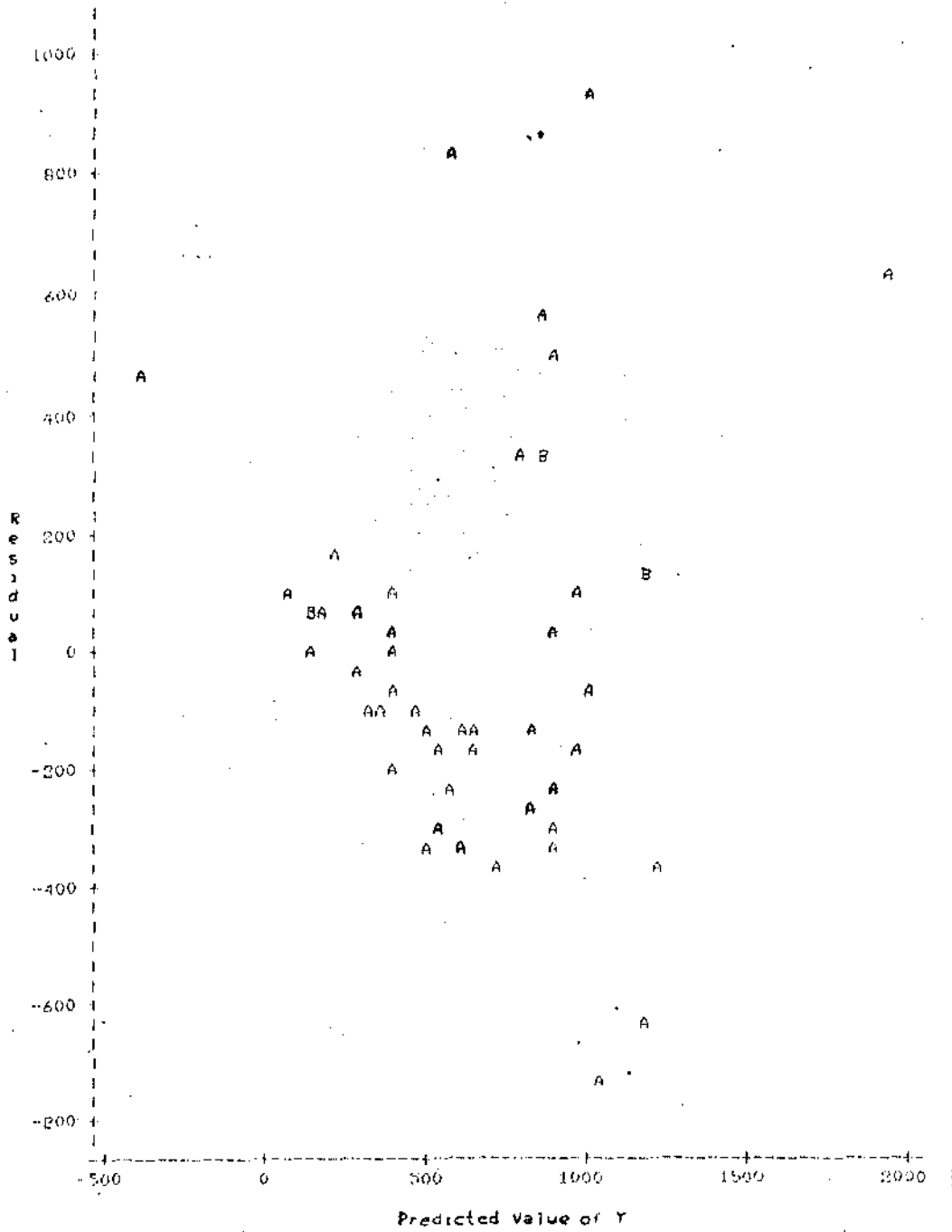


FIGURA - 7.5

CAS

UNIVARIATE

VARIABLE=RS

MOMENTS

N	49	SUM WGTs	49
MEAN	54.9298	SUM	2691.56
STD DEV	344.36	VARIANCE	118584
SKEWNESS	0.781077	KURTOSIS	1.7207
USS	5839863	CBS	5692016
CV	626.909	STD MEAN	49.1943
T:MEAN=0	1.11659	PROB>ITI	0.269728
SGN RANK	42	PROB>ISI	0.670361
NUM ^= 0	48		
W:NORMAL	0.921829	PROB<W	<.01

QUANTILES(DEF=4)

0% MAX	1030.59	99%	1030.59
5% Q3	184.29	95%	860.166
0% MED	-1.000E-04	90%	531.674
5% Q1	-116.26	10%	-274.02
0% MIN	-731.47	5%	-516.7
		1%	-731.47
ANGE	1762.06		
3-Q1	300.551		
ODE	-0.0004		

EXTREMES

LOWEST	HIGHEST
-731.47	531.674
-653.57	629.649
-379.84	700.806
-336.11	1019.53
-274.02	1030.59

FIGURA - 7.6

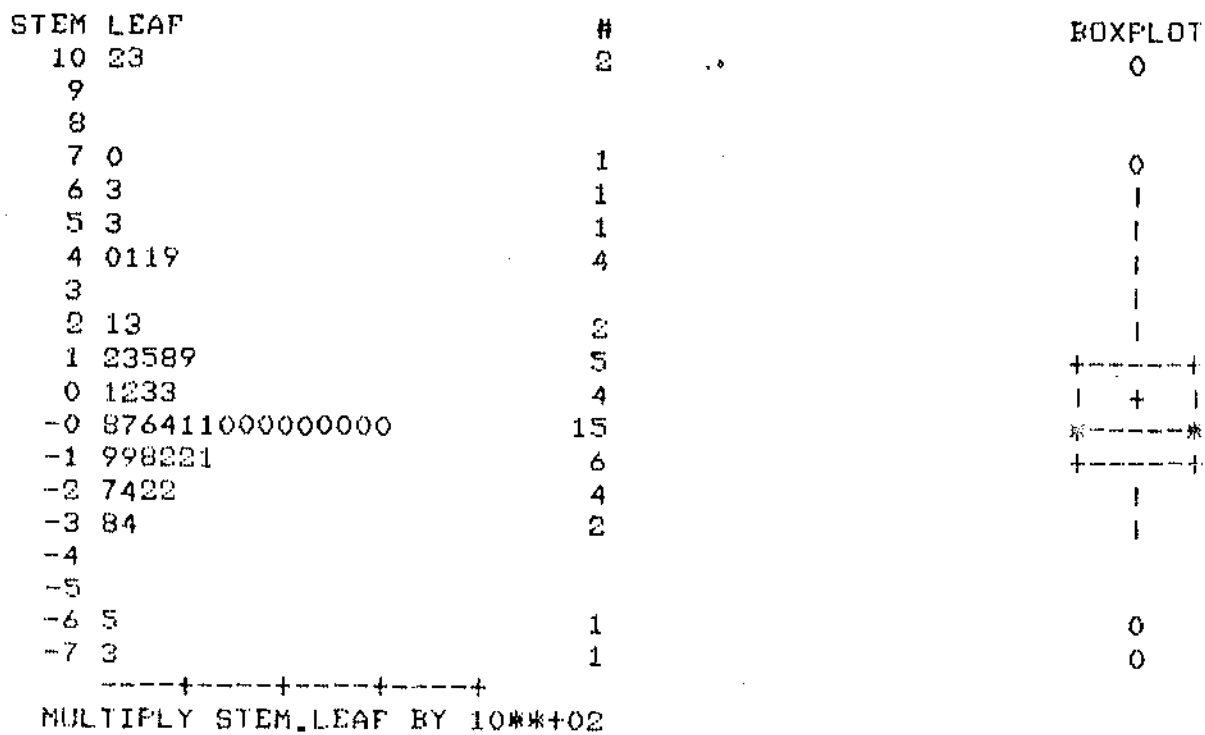


FIGURA - 7.7

SAS

UNIVARIATE

VARIABLE=RS

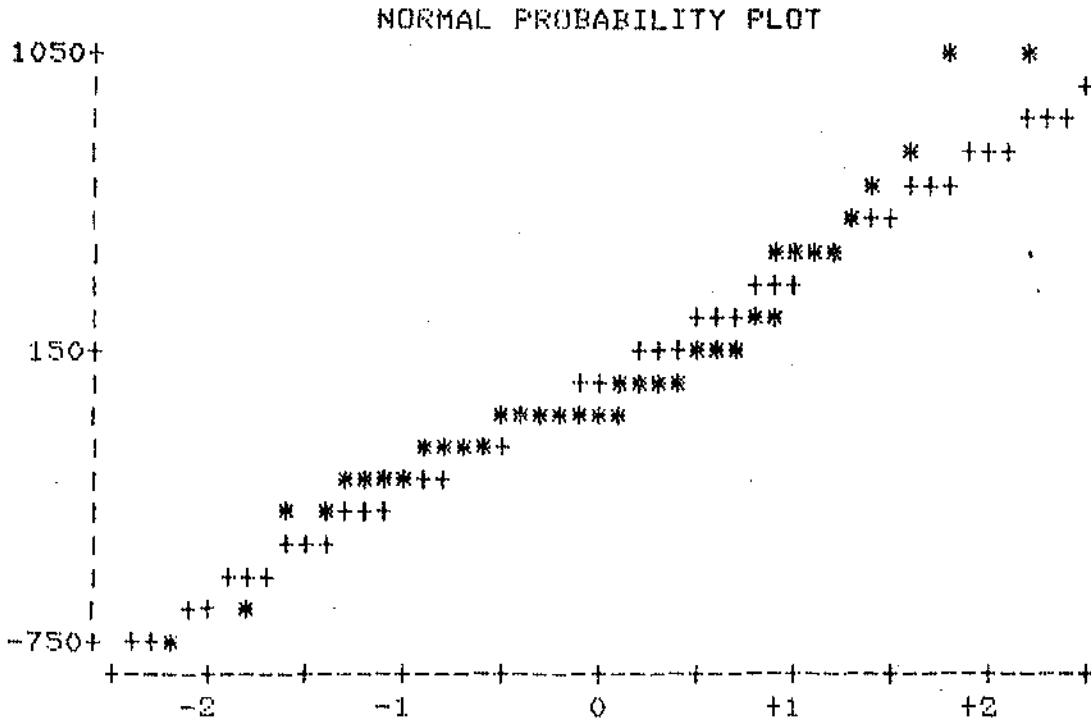


FIGURA - 7.8

DADOS ORIGINAIS - RESÍDUOS CONTRA
PREDITOS - TETA=0,5

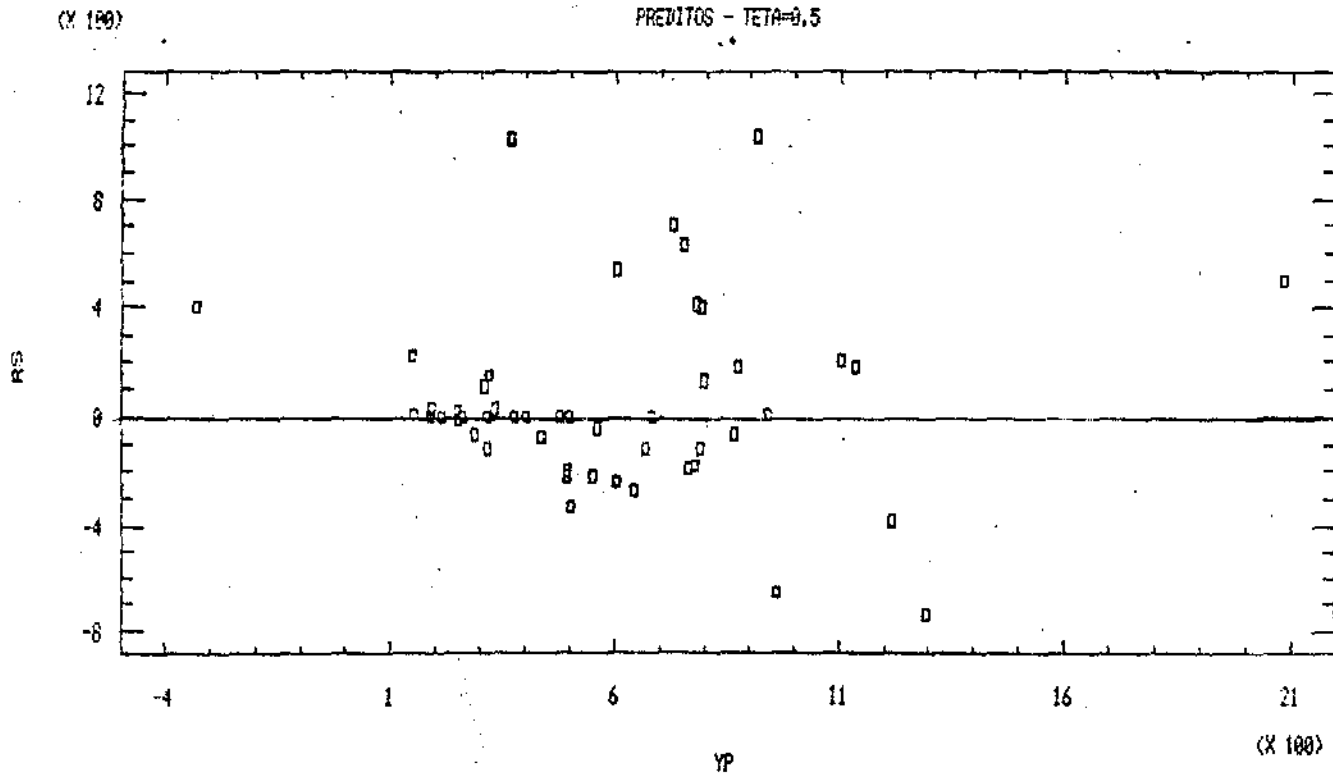


FIGURA - 7.9

DADOS ORIGINAIS - RESIDUOS CONTRA
PRELITOS - TETA=0.2

(X 100)

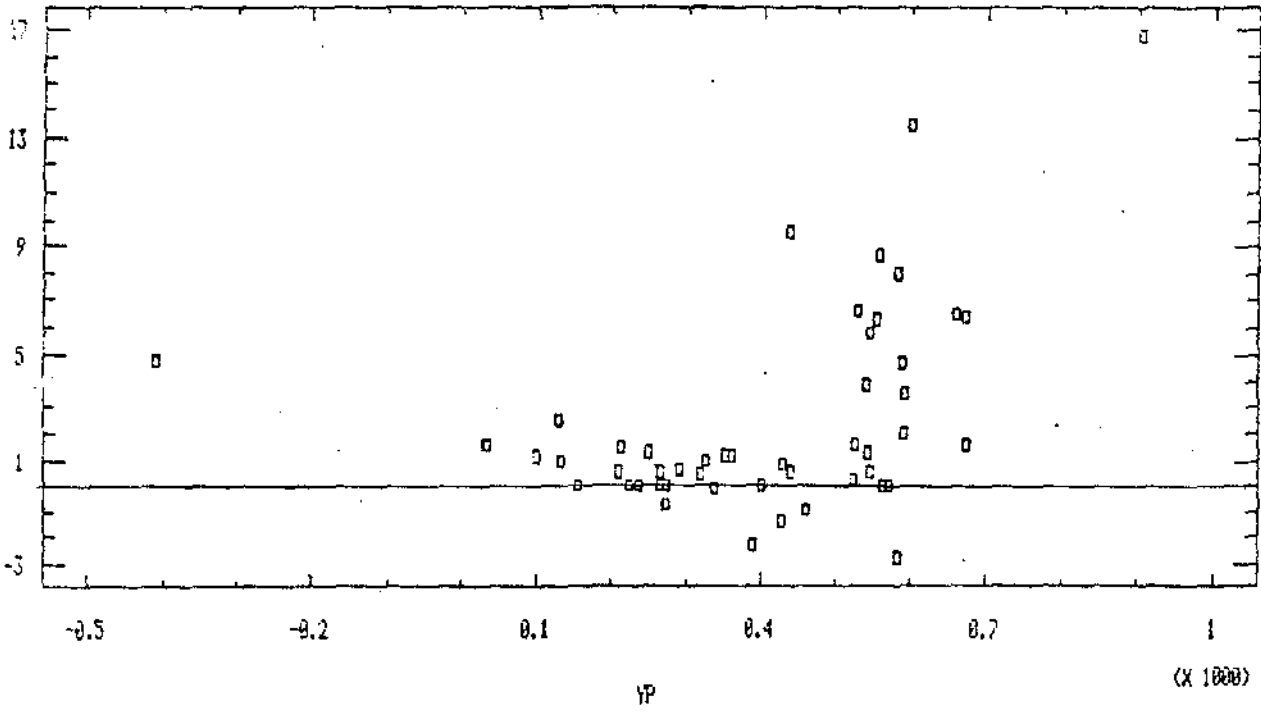


FIGURA - 7.10

DADOS ORIGINAIS - RESIDUOS CONTRA
PREBITOS - TETA=0.8

(X 100)

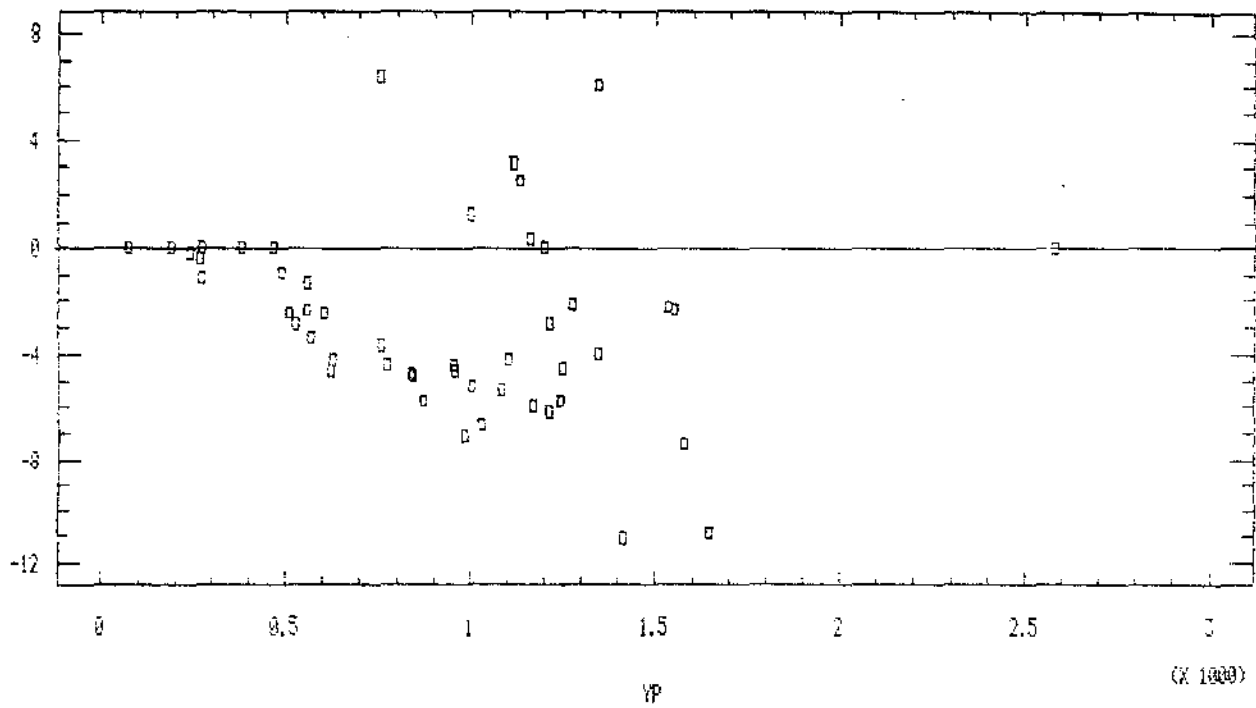


FIGURA - 7.11

DADOS TRANSFORMADOS - RESÍDUOS CONTRA
PREDITOS - TETA = 0.5

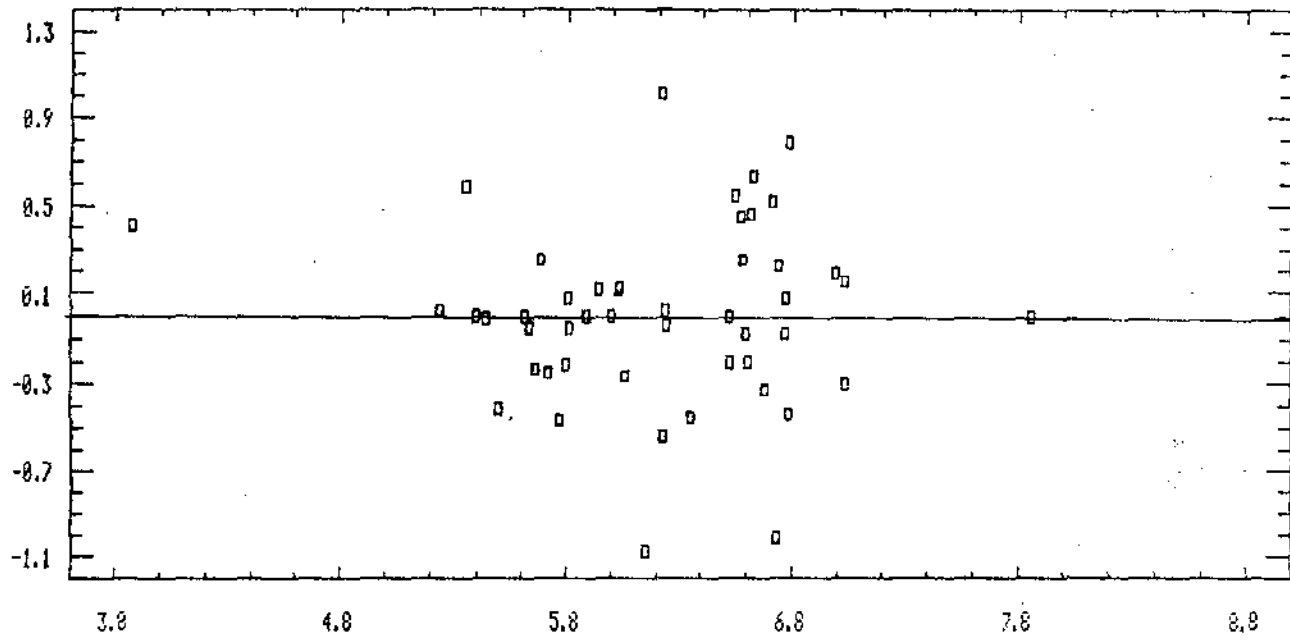


FIGURA - 7.12

DADOS TRANSFORMADOS - RESÍDUOS CONTRA
PREDITOS - TETA = 0.20

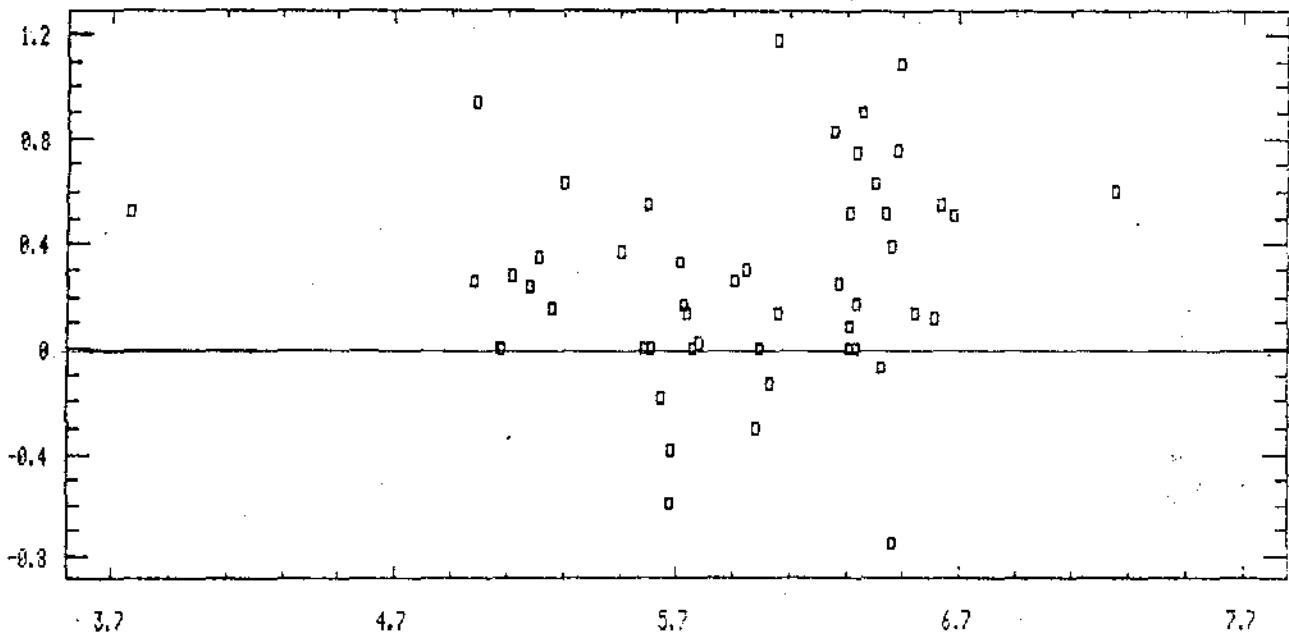


FIGURA - 7.13

DADOS TRANSFORMADOS - RESÍDUOS CONTRA
PREDITOS - TETA = 0.80

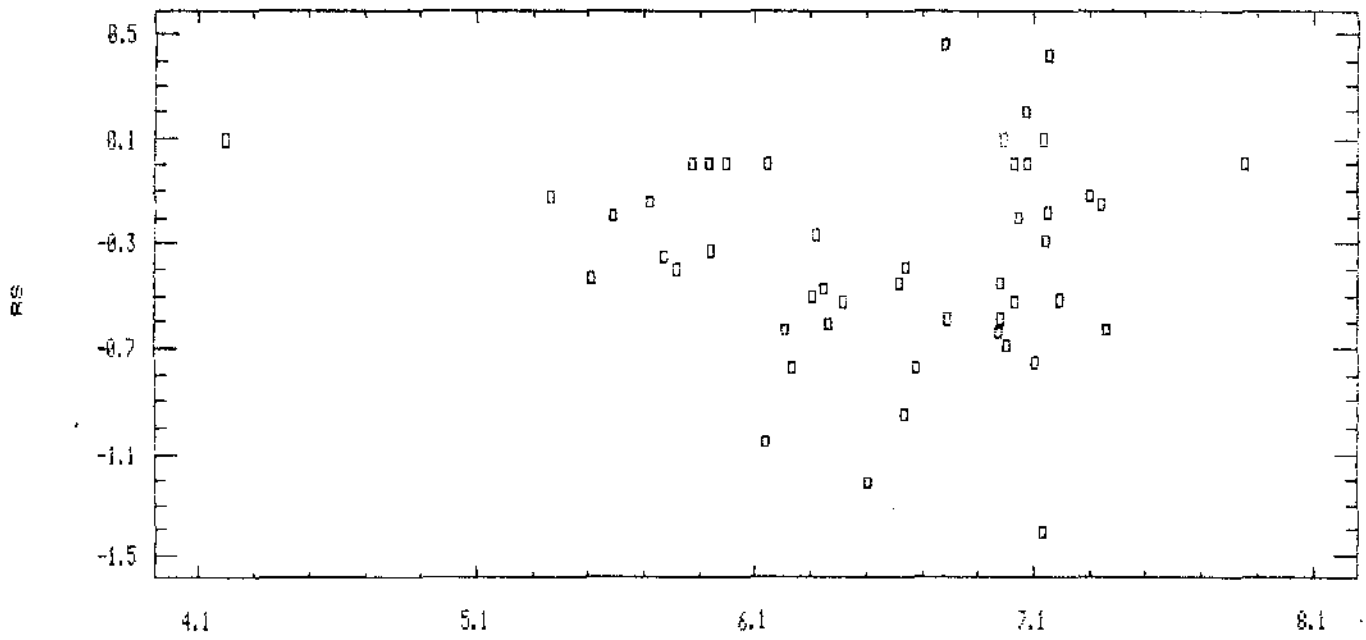


FIGURA - 7.14

RESIDUOS EM VALORES ABSOLUTOS VERSUS OBSERVACOES

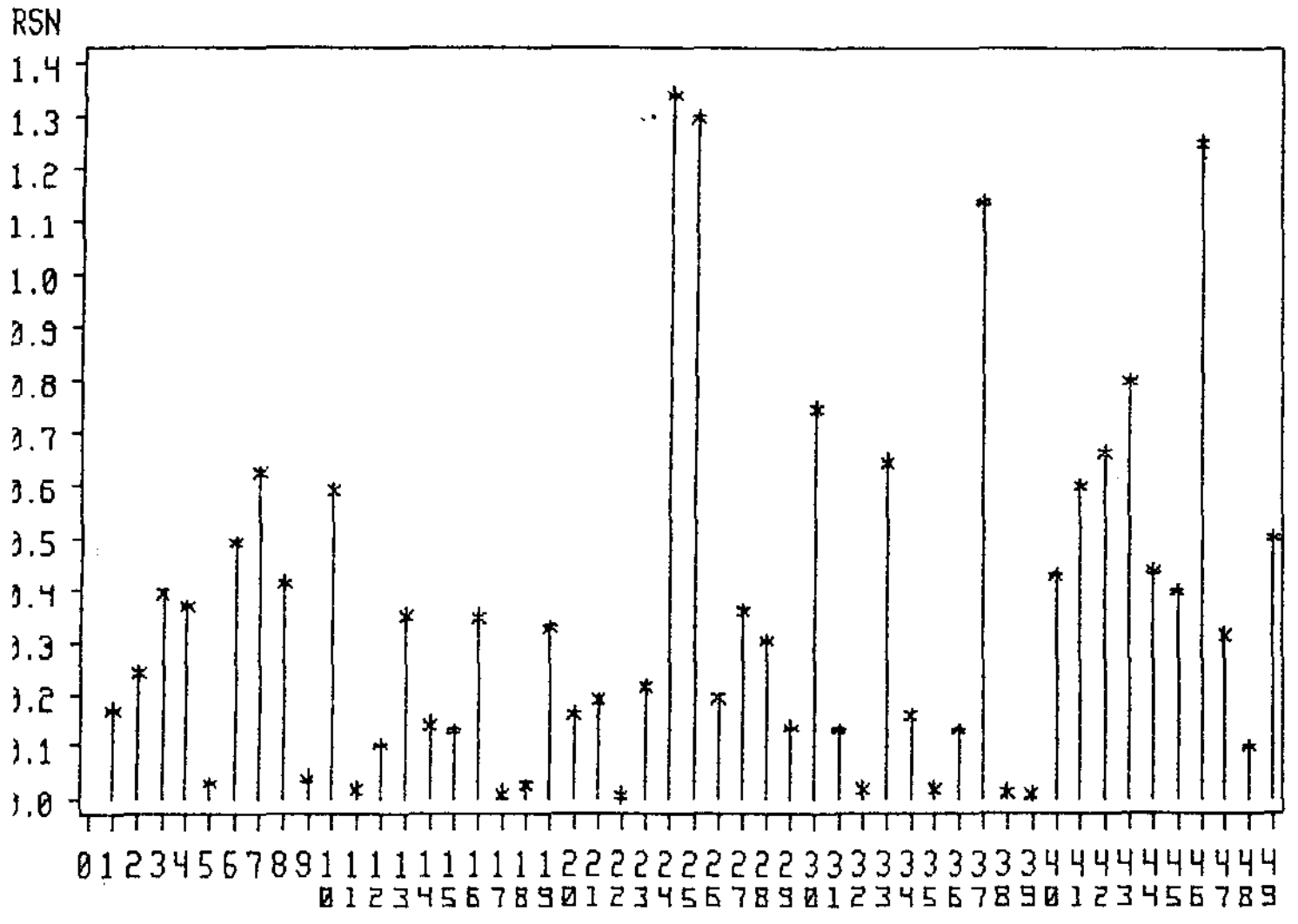


FIGURA - 7.15

RESIDUOS CONTRA FRENTOS

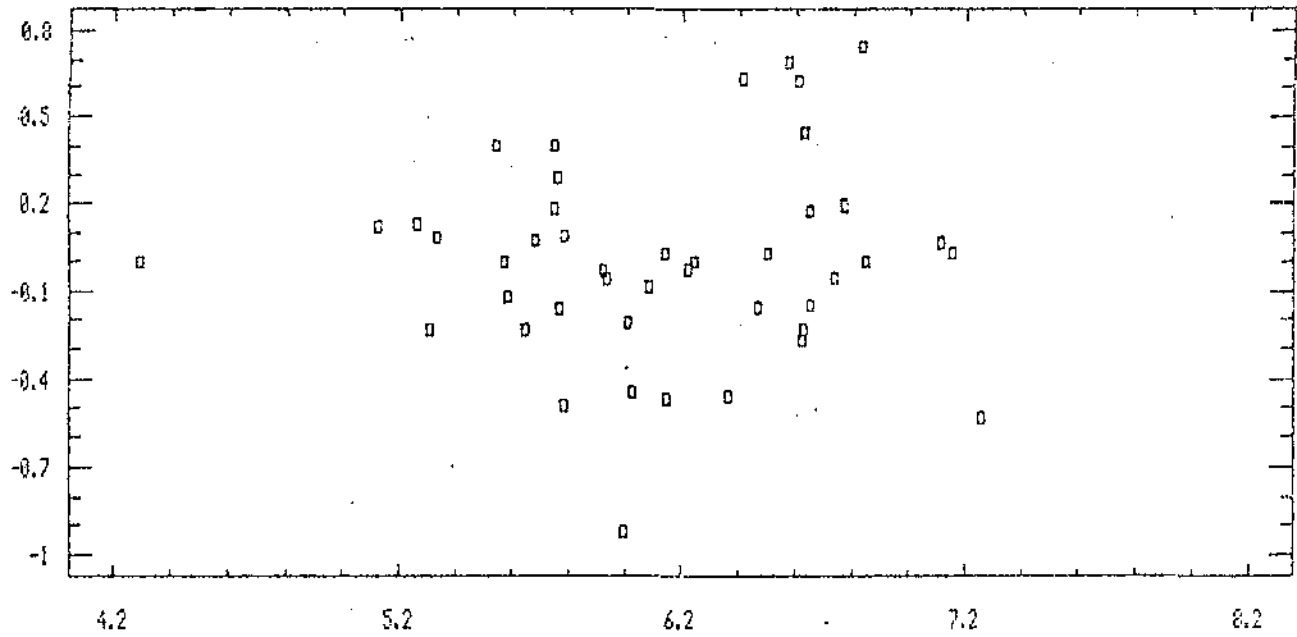


FIGURA - 7.16

residuos contra preditos

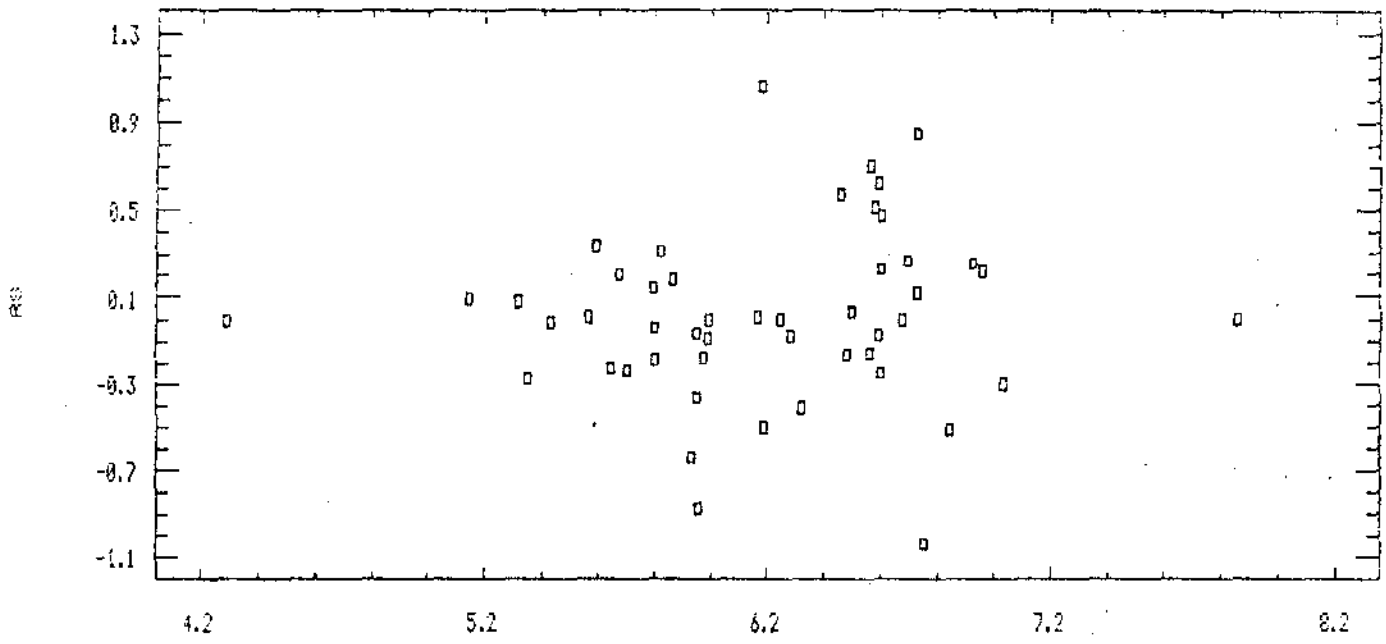


FIGURA - 7.17

RESIDUOS CONTRA X1

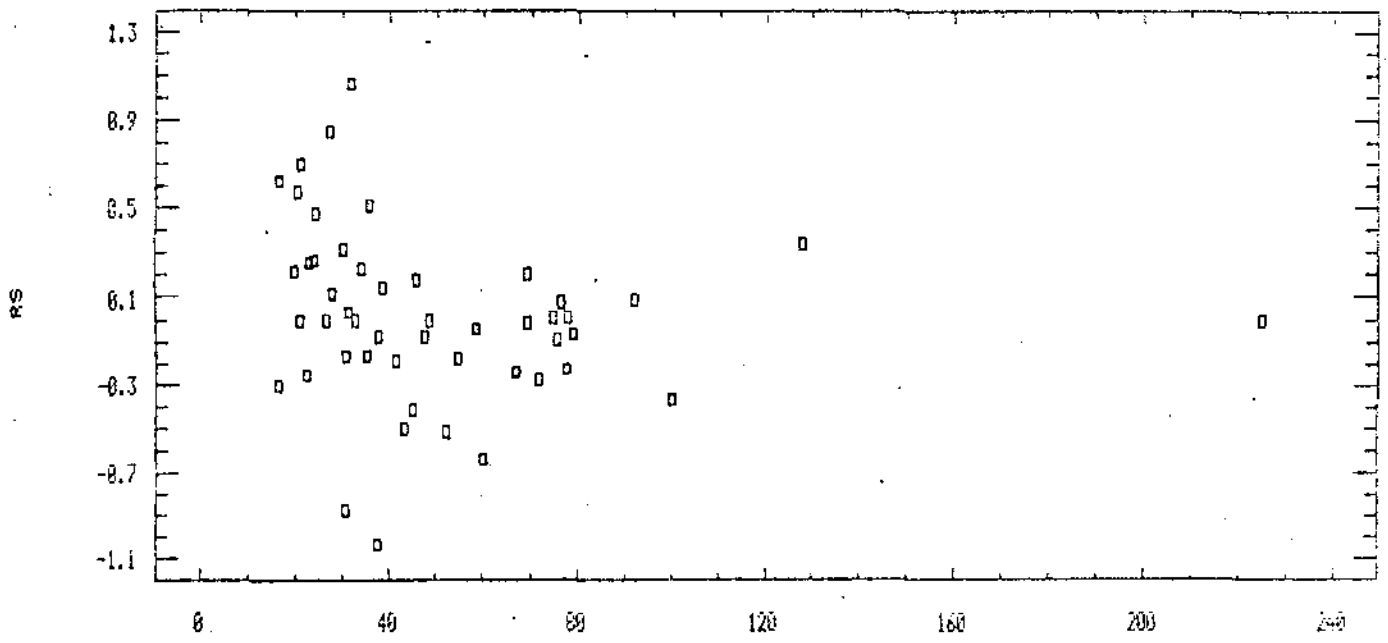


FIGURA - 7.18

RESÍDUOS CONTRA X₆

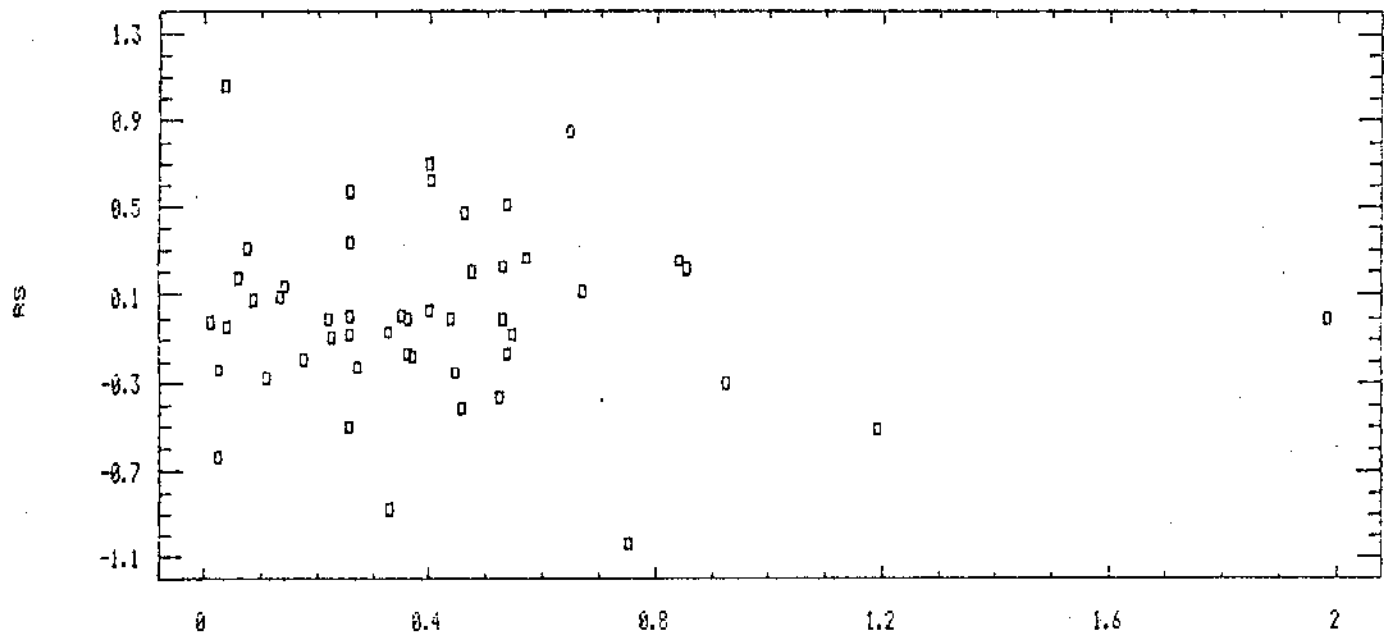


FIGURA - 7.20

- B I B L I O G R A F I A -

- [1] ARMSTRONG, R. D., FROME, E.L., AND KUNG, D.S. A revised simplex algorithm for the absolute deviation curve fitting problem. *Communications in Statistics - Simulation and Computation*. B8: 179-190, 1979.
- [2] BARRODALE, I. AND ROBERTS, F.D.K. An improved algorithm for discrete L_1 linear approximation. *SIAM - Journal on Numerical Analysis*. 10: 839-848, 1973.
- [3] BASSETT, G. AND KOENKER, R. Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*. 73: 618-622, 1978.
- [4] BASSETT, G. AND KOENKER, R. . An Empirical Quantile Function For Linear Models With iid Errors. *Journal of the American Statistical Association*. 77: 407-415, 1982.
- [5] BASSETT, G. AND KOENKER, R. Regression Quantiles. *Econometrica*. 46: 33-50, 1978.
- [6] BLOOMFIELD, P. AND STEIGER, W.L. *Least Absolute Deviations - Theory, Applications and Algorithms*. Boston, Basel, 1983.
- [7] BOSCOVICH, R.J. De litteraria expeditione per pontificamditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiesi Scientiarum et Artum Instituto atque Academia Commentarii* 4: 353-96, 1757.
- [8] COX, R.D. & WEISBERG, S. *Residuals and Influence in Regression*. New York, Chapman and Hall Ltd, 1982.

- [9] COX, R.D. & LINKLEY, D.V. *Theoretical statistics*. London, Chapman and hall Ltd, 1974.
- [10] CHARLES, A., COPPER, W.W., AND FERGUNSON, R.O. Optimal estimation of executive compensation by linear programming. *Management Science*, 1: 138-151, 1955.
- [11] DACHS, J.N. E CARVALHO, J.F. *Diagnóstico em Regressão*. VI SINAPE, Rio de Janeiro, 1984.
- [12] DANIEL, C. & WOOD, F.S. *Fitting Equations to Data*. New York, John Wiley & Sons, Inc, 1971.
- [13] DIELMAN, T. AND PFAFFENBERGER, R. LAV (Least Absolute Value) estimation in the regression model: A review. *Time Studies in the Management Sciences*. 10: 31-52, 1982.
- [14] DIELMAN, T. AND PFAFFENBERGER, R. Least Absolute Value Regression: Necessary Sample Sizes to use Normal Theory Inference Procedures. *Decision Sciences*. 19: 734-743, 1988.
- [15] DIELMAN, T. AND PFAFFENBERGER, R. Booststrapping in Least Absolute Value Regression: An Aplication to Hypotesis Testing. *Communications in Statistics - Simulation and Computation*. 17(3): 843-856, 1988.
- [16] DRAPER, N.R. & SIMTH, H. *Applied Regression Analysis*. New York, John Wiley & Sons, Inc, 1981.
- [17] EFRON, B. Booststrapping Methods: Another Look at the Jackknife. *Ann. Statist.* 7: 1-26, 1979.
- [18] FILHO, P.F. *Intervalos de Confiança para os Parâmetros da Regressão L_1 : Pequenas Amostras*. Tese (mestrado). Campinas, UNICAMP, 1988.

- [19] GUNST & MASON. *Regression Analysis and its Application - A Data Approach*. New York, 43, 1980.
- [20] HA, C.D. AND NARULA, S.C. An Optimization Problem Arising from the Tolerance Approach to Sensitivity Analysis. *Research Report, Virginia Commonwealth University, Richmond, Virginia, U.S.A., 1986.*
- [21] HA, C.D. AND NARULA, S.C. Perturbation Analysis for the Minimum Sum of Absolute Errors Regression. *Research Report, Virginia Commonwealth University, Richmond, Virginia, U.S.A., 1987.*
- [22] HOGG, R.V. Estimates of Percentile Regression Lines Using Salary Data. *JASA*, 70: 56-59, 1975.
- [23] HUBER, P.J. Comment on "Adaptive robust procedures: A partial review and some suggestions for future applications and theory", by R.V. Hogg, *JASA*, 69: 926-927, 1974.
- [24] KARST, O.J. Linear curve fitting using least deviations. *JASA*, 53: 118-132, 1958.
- [25] LAPLACE, P.S. Sur quelques points du systems du monde. *Memoires de l'Academic Royale des Sciences de Paris Ann ee 1789*, 1-87. Reprinted in *Oeuvres Completes de Laplace*, Vol. II. Paris: Gauthier-Villars (1895), 477-558, 1793.
- [26] MARITZ, J.S. and JARRETT, R.G. A Note on Estimating the Variance of the Sample Median. *JASA*, 73(361): 194-196, 1978.
- [27] McKEAN, J.W. AND SCHRADER, R.M. *Least Absolute Erros Analysis*

of Variance. Elsevier Science Publisher B.V., 297-305, 1987.

- [28] NARULA, S.C. AND WELLINGTON, J.F. Prediction, linear regression and minimum sum of relative errors. *Technometrics*, 19: 185-190, 1977.
- [29] NARULA, S.C. AND WELLINGTON, J.F. An Algorithm for the Minimum Sum of Weighted Absolute Errors Regression. *Communications in Statistics - Simulation and Computation*, B6(4): 341-352, 1977.
- [30] NARULA, S.C. AND WELLINGTON, J.F. Selection of variables in linear regression using the minimum sum of weighted absolute errors criterion. *Technometrics*, 21: 299-306, 1979.
- [31] NARULA, S.C. AND WELLINGTON, J.F. Selection of variables in linear regression: A pragmatic approach. *Journal of Statistical Computation and Simulation*, 17: 159-172, 1983.
- [32] NARULA, S.C. and WELLINGTON, J.F. Interior analysis for the minimum sum of absolute error regression. *Technometrics*, 27: 181-188. 1985.
- [33] NARULA, S.C. and WELLINGTON, J.F. A Branch and Bound Procedure for Selection of Variables in Minimax Regression. *SIAM J. Appl. Math.* 6, 573- 581, 1985.
- [34] NARULA, S.C. The minimum sum of absolute errors regression. *Journal of Quality Technology*, 19: 37-45, 1987.
- [35] PORTNOY, S. Using regression fractiles to identify outliers. Elsevier Science Publishers. B.V., p. 345-356, 1987.

- [36] ROSENBERG, B. AND CARLSON, D. A simplex approximation of the sampling distribution of least absolute residual regression estimates. *Communications in Statistics - Simulation and Computation*, B6: 421-427, 1977.
- [37] SAS Institute, Inc. *SUGI Supplemental Library User's Guide*, 1983 Edition. SAS Institute, Inc., Cary, N.C., 1983.
- [38] STANGENHAUS, G. and NARULA S.C. *Análise de Regressão L_1* . 8^o SINAPE - IMPA, Rio de Janeiro, 1988.
- [39] STANGENHAUS, G. AND NARULA S.C. Inference procedures for the L_1 - Regression. *Relatório Técnico No. 49/87*. Universidade de Campinas, Campinas, S.P., 1987.
- [40] SPOSITO, V.A. and McCORMICK, G.F. Using the L_2 - estimator in L_1 - estimation: *SIAM J. Numer. Anal.* 13, 337-43, 1976.
- [41] TALARICO, R.M.P.M. *Regressão Linear Múltipla através da Mínima Soma dos Desvios Absolutos*. Tese (mestrado). São Carlos, USP, 1984.
- [42] WELLINGTON, J.F. and NARULA S.C. Variable selection in multiple linear regression using the minimum sum of weighted absolute errors criterion. *Communications in Statistics - Simulation and Computation*, B10: 641-648, 1981.
- [43] WELLINGTON, J.F. and NARULA S.C. An Algorithm for Regression Quantiles. *Communications in Statistics - Simulation and Computations*. 13(5): 683-704, 1984.