

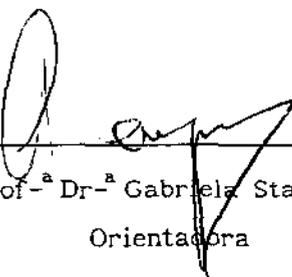
EPM/C

12

TÍTULO DA TESE : TRANSFORMAÇÕES E DIAGNÓSTICOS EM REGRESSÃO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sr^a Maria Cláudia Cabrini Grácio e aprovada pela Comissão Julgadora.

Campinas, 08 de fevereiro de 1991


Prof.ª Dr.ª Gabriela Stangenhau
Orientadora

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

3479

o cad.

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

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO
- I M E C C -

TRANSFORMAÇÕES E DIAGNÓSTICOS EM REGRESSÃO

MARIA CLÁUDIA CABRINI GRÁCIO

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

Stangenhau...

CAMPINAS - 1991

UNICAMP
BIBLIOTECA CENTRAL

Ao meu marido José Carlos

AGRADECIMENTOS

À Dr^a Gabriela Stangenhau, pela orientação, colaboração, sugestões, compreensão e amizade presentes durante todas as minhas atividades.

Ao meu marido, José Carlos, pela paciência, incentivo, confiança e carinho sempre dedicados.

Aos Professores do Departamento de Estatística, pelos ensinamentos e colaboração concedidos.

Aos funcionários do I.M.E.C.C., pela agradável convivência.

Aos colegas da Pós-Graduação, pelo apoio, amizade e pelos bons momentos que passamos juntos.

Aos meus cunhados, Marco e Eliane, pela hospitalidade acolhedora com que sempre me receberam em Campinas.

À CAPES e UNICAMP, pelo apoio financeiro.

À UNESP, pelas facilidades concedidas para conclusão do curso de pós-graduação.

Aos meus familiares, pela confiança transmitida em todos os momentos.

A DEUS, por me fornecer, sempre, força e coragem para superar os obstáculos que surgiram no meu caminho.

RESUMO

Este trabalho reúne informações a respeito do que tem sido pesquisado na área de transformações do tipo Box e Cox para a variável resposta de um modelo de regressão linear, sob a minimização dos erros quadráticos e a minimização dos erros absolutos.

Primeiramente, apresentamos métodos para a escolha de uma transformação do tipo Box e Cox adequada para a variável resposta de um modelo de regressão linear, sob a minimização dos erros quadráticos, algumas alternativas robustas e a minimização dos erros absolutos, usando critérios estabelecidos.

Em seguida, apresentamos métodos de diagnósticos de influência de um subconjunto de observações do conjunto de dados sobre a estimativa do parâmetro de transformação da variável resposta, sob a minimização dos erros quadráticos e a minimização dos erros absolutos.

Finalmente, a título de ilustração e comparação, utilizando alguns dos métodos catalogados, analisamos alguns conjuntos de dados simulados e um conjunto de dados real.

ÍNDICE

CAPÍTULO 1	INTRODUÇÃO	1
CAPÍTULO 2	TRANSFORMAÇÕES DO TIPO BOX E COX	7
2.1	INTRODUÇÃO	7
2.2	FAMÍLIA DE TRANSFORMAÇÕES DE BOX E COX	9
2.3	TESTES PARA TRANSFORMAÇÕES A FIM DE OBTER NORMALIDADE APROXIMADA.....	17
2.3.1	MÉTODO DE ANDREWS	18
2.3.2	MÉTODO DE ATKINSON	23
2.4	ALTERNATIVAS ROBUSTAS PARA A ESTIMAÇÃO DO PARÂMETRO DE TRANSFORMAÇÃO	26
2.4.1	MÉTODO DE CARROLL	26
2.4.2	ESTIMADORES DE PSEUDO-VEROSSIMILHANÇA	30
2.4.3	ESTIMADORES DE INFLUÊNCIA LIMITADA	35
2.5	TRANSFORMAÇÕES SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS	38
2.5.1	SELEÇÃO DE UMA TRANSFORMAÇÃO	41
2.6	ANÁLISE ROBUSTA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS	44
2.6.1	ESTIMADORES DE PSEUDO-VEROSSIMILHANÇA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS	44
2.6.2	ESTIMADORES DE INFLUÊNCIA LIMITADA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS	46
CAPÍTULO 3	DIAGNÓSTICOS DE INFLUÊNCIA PARA TRANSFORMAÇÕES DO TIPO BOX E COX	48
3.1	INTRODUÇÃO	48
3.2	DIAGNÓSTICOS DE INFLUÊNCIA PARA REGRESSÃO DE MÍNIMOS QUADRADOS	49
3.2.1	GRÁFICO DE VARIÁVEL ADICIONADA	50
3.2.2	MÉTODO GRÁFICO DE ATKINSON	53

3.2.3	MÉTODO DE COOK E WANG	55
3.2.4	MÉTODO DE ATKINSON	60
3.3	DIAGNÓSTICOS DE INFLUÊNCIA PARA REGRESSÃO L_1	62
3.3.1	MÉTODOS GRÁFICOS DE PARKER	63
CAPÍTULO 4 APLICAÇÕES		68
4.1	INTRODUÇÃO	68
4.2	PROBLEMA 1	70
4.3	PROBLEMA 2	84
4.4	PROBLEMA 3	98
4.5	PROBLEMA 4	110
4.6	COMENTÁRIOS	126
APÊNDICE 1		128
BIBLIOGRAFIA		147

CAPÍTULO 1

INTRODUÇÃO

Entre os métodos estatísticos disponíveis para análise de dados, um dos mais discutidos é a regressão. A teoria da regressão é utilizada para resolvermos problemas de aproximação de uma função a um conjunto de pontos dado. Na tentativa de ajustarmos uma curva a esse conjunto de pontos, procuramos a função que esteja mais próxima possível dos pontos. Usualmente, o tipo de função é sugerido por evidências empíricas ou argumentos teóricos. Após escolhermos um modelo, devemos analisá-lo para verificarmos se ele está adequado.

A técnica mais comumente utilizada para solucionarmos problemas de regressão é o método dos mínimos quadrados, proposto por Legendre em 1805. A popularidade do estimador de mínimos quadrados pode ser explicada, pelo menos parcialmente, pela simplicidade computacional e matemática do método e pelo fato de que quando os erros aleatórios são independentes e possuem variâncias constantes, o estimador de mínimos quadrados de um modelo de regressão linear é BLUE (melhor estimador linear não-tendencioso) assim como equivale ao estimador de máxima verossimilhança; conseqüentemente as inferências são realizadas facilmente.

Entretanto, sabemos que os estimadores de mínimos quadrados são sensíveis a valores aberrantes, assim como podem ser altamente ineficientes quando os erros não tem distribuição normal, ou possuem caudas alongadas, como por exemplo, as distribuições dupla-exponencial e Cauchy. Em uma variedade de aplicações tem surgido evidências teóricas e empíricas de que os resíduos revelam características distribucionais diferentes daquelas de distribuições normais.

Diversos estimadores alternativos tem sido desenvolvidos a fim de superarmos essas dificuldades da regressão de mínimos quadrados. Esses estimadores, que procuram acomodar a possibilidade da presença de valores aberrantes, são denominados estimadores robustos. A ocorrência de valores aberrantes pode ser estimulada pelo fato da distribuição de erros ter caudas alongadas.

A regressão L_1 , isto é, que minimiza a soma dos erros absolutos 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.

Embora proposta meio século antes da publicação do "Princípio de Mínimos Quadrados" por Legendre, o desenvolvimento da regressão de erros absolutos foi lenta. O problema computacional associado à regressão L_1 tornou seu uso praticamente impossível por muitos anos. Charnes, Cooper e Ferguson(1955) formularam o problema de regressão L_1 como um problema de programação linear e solucionaram-no usando o algoritmo simplex. Este trabalho foi de extrema importância, transformando a regressão L_1 em uma técnica computacionalmente tratável.

Além das dificuldades computacionais, uma grande restrição na ampla aplicação da regressão L_1 foi a falta de ferramental para inferência estatística, se comparada à regressão de mínimos quadrados. Entretanto, nos últimos dez anos diversos estudos de Monte Carlo foram desenvolvidos sobre propriedades dos estimadores para pequenas amostras, e surgiram resultados assintóticos sobre a distribuição dos estimadores e procedimentos para a inferência.

Quando a distribuição dos erros é dupla-exponencial (ou Laplace), os estimadores dos parâmetros da regressão L_1 são estimadores de máxima verossimilhança, portanto são assintoticamente não tendenciosos, eficientes e normalmente distribuídos.

A regressão L_1 é mais eficiente do que a regressão de mínimos quadrados quando os erros tem distribuição simétrica para a qual a mediana amostral é um estimador de posição mais eficiente do que a média amostral, por exemplo, quando os erros tem distribuição com caudas alongadas como as distribuições dupla-exponencial, Cauchy ou normais contaminadas. Quando a função de perda é baseada nos erros absolutos a regressão L_1 é mais apropriada do que a regressão de mínimos quadrados.

Muitos procedimentos em regressão de mínimos quadrados tem sido destinados a viabilizar a aplicação de modelos lineares a um conjunto de dados que não possua as suposições subjacentes à sua utilização e também a detectar discordâncias entre os dados e o modelo, então, assumido. Entre esses procedimentos, métodos comumente utilizados são as famílias de transformações da variável resposta, proposto por Box e Cox em 1964 e a utilização de diagnósticos de influência, que surgiu no final da década de 1970.

A metodologia, na área de transformações da variável resposta, cresceu consistentemente na década de 1970 e conjuntamente

com a metodologia desenvolvida para diagnósticos de influência na década de 1980. Entretanto, ambas metodologias encontram-se espalhadas em diversos artigos, relatórios e em poucos livros. Tomamos conhecimento, através da seção "Book Reviews" do volume 85 da revista JASA, publicado em junho de 1990, do lançamento de um livro que trata quase que exclusivamente do desenvolvimento de metodologias nessas áreas. Tal livro, publicado em 1988, é da autoria de Carroll e Ruppert e intitulado por "Transformation and Weighting in Regression".

Em regressão L_1 , devido ao recente desenvolvimento de metodologia, encontramos um único artigo específico relacionado à aplicação de transformações da variável resposta e métodos de diagnósticos de influência na escolha de tais transformações.

Nesse artigo, publicado por Parker em 1988, a obtenção de resultados interessantes relativos à aplicação da metodologia de transformações na variável resposta, sob minimização dos erros absolutos, em um conjunto de dados real motivou-nos a realizar esse trabalho. A aplicação, nesse artigo, foi realizada sobre um conjunto de dados bastante conhecido por "Stack-Loss Data", primeiramente analisado por Brownlee(1965), e que vem sendo analisado por diversos pesquisadores conhecidos, como C.Daniel e F.S.Wood(1980), D.F.Andrews(1974), A.C. Atkinson(1982), e R.J.Carroll e D.Ruppert(1985). Todos os pesquisadores mencionados acima tratam os dados como possuindo erros aleatórios independentes e identicamente distribuídos normais e aplicam, então, regressão de mínimos quadrados ou algum método de regressão robusta. Em todas as análises por eles desenvolvidas, chegou-se à conclusão que cerca de 20% das observações são aberrantes, devendo portanto serem desprezadas. Parker(1988), no entanto, trata os dados como possuindo erros aleatórios independentes e com distribuição dupla-exponencial e após aplicar a metodologia desenvolvida para transformação da variável resposta, sob minimização dos erros absolutos, consegue acomodar perfeitamente todas as

observações.

O fato é que, não podemos, na verdade, afirmar que tais dados pertencem a uma distribuição normal ou a uma distribuição dupla-exponencial, devido ao pequeno tamanho da amostra ($n=21$), mas quando trabalhamos sob a hipótese de normalidade desprezamos cerca de 20% de informações, que podem na realidade estar evidenciando a não adequação da suposição assumida, ao passo que, quando trabalhamos sob a hipótese de dupla-exponencialidade não jogamos fora informação. Cabe, portanto, nesse caso ao pesquisador a escolha da distribuição de erros subjacente ao problema, de acordo com a consciência de suas necessidades e das restrições de seus resultados.

A estimação L_1 vem ganhando uma grande preocupação por parte dos pesquisadores nos últimos anos e por esse motivo realizamos esse trabalho, cuja finalidade é reunir informações sobre o que tem sido pesquisado na área de transformações do tipo Box e Cox para a variável resposta e de diagnósticos de influência, sob minimização de erros quadráticos e sob minimização de erros absolutos, a fim de facilitar o estudo daqueles estatísticos ou pesquisadores que se interessam pela área. Nesse trabalho, desejamos, também, verificar a partir das linhas de pesquisas desenvolvidas nessas áreas para regressão de mínimos quadrados, o que pode ser desenvolvido por analogia para regressão L_1 .

No capítulo 2, apresentaremos a família de transformações do tipo Box e Cox e as metodologias desenvolvidas para selecionarmos a melhor transformação, baseada no critério de máxima verossimilhança, sob minimização dos erros quadráticos, sob minimização dos erros absolutos e sob outras alternativas robustas.

No capítulo 3, apresentaremos os métodos de diagnósticos desenvolvidos para avaliarmos a influência de observações

individuais na escolha de uma transformação em particular, sob minimização dos erros quadráticos e sob minimização dos erros absolutos.

No capítulo 4, utilizaremos um conjunto de dados retirado do livro de Gunst e Mason (1980) e aplicaremos os métodos acima citados em conjuntos de dados simulados.

CAPÍTULO 2

TRANSFORMAÇÕES DO TIPO BOX E COX

2.1 INTRODUÇÃO

Quando usamos as técnicas de análise de regressão assumimos certas suposições subjacentes ao conjunto de dados em estudo a fim de que as conclusões tiradas dessa análise possam ser confiáveis. As técnicas usuais de modelos lineares sob a minimização dos erros quadráticos supõem que o conjunto de dados analisado represente uma amostra aleatória de uma população que possua as seguintes características:

- (i) normalidade aproximada,
- (ii) variância dos erros constantes,
- (iii) independência das observações.

Muitas vezes , entretanto, o conjunto de dados original não possui as características assumidas acima. Transformações tem sido utilizadas a fim de tentarmos alcançar, pelo menos aproximadamente, esses requisitos para análise de modelos lineares, sob a minimização dos erros quadráticos.

Nesse capítulo, estudaremos alguns tipos de transformações convenientes quando não conhecemos precisamente a distribuição da população e a forma funcional do relacionamento entre $E(Y)$ e as variáveis explanatórias. Um método de procedimento nesta situação, quando desejamos minimizar a soma dos erros quadráticos, foi primeiramente dado por Box e Cox (1964), onde eles especificaram uma família de transformações $Y^{(\lambda)}$, indexadas por um vetor de parâmetros desconhecidos, $\underline{\lambda}$. Através de métodos padrões de inferência aplicados ao conjunto de dados, esse procedimento estima $\underline{\lambda}$ e os parâmetros do modelo tal que seja possível obtermos um modelo que atenda, pelo menos aproximadamente, as características (i) a (iii).

Na seção (2.2) apresentaremos a família de transformações de Box e Cox e o método de análise utilizado por eles, baseado em considerações de verossimilhança. Na seção (2.3) introduziremos dois métodos adicionais para avaliar a necessidade de uma transformação na variável resposta a fim de obtermos um modelo linear com distribuição de erros aproximadamente normal. Estes métodos, propostos por Andrews(1971) e Atkinson(1973), são baseados em testes de hipóteses para a hipótese nula $\underline{\lambda} = \underline{\lambda}_0$, onde $\underline{\lambda}_0$ é um valor dado para $\underline{\lambda}$.

Em regressão sabemos, entretanto, que os estimadores obtidos por máxima verossimilhança, sob normalidade, são muito sensíveis a observações discrepantes, podendo, assim, serem altamente ineficientes se a distribuição dos erros tiver caudas mais "pesadas" que a distribuição normal. A seguir (seções (2.4) e (2.5)) apresentaremos, então, alternativas robustas para a estimação do parâmetro de transformação, propostas por Carroll(1980), Carroll e Ruppert(1985) e Parker(1988).

A alternativa proposta por Parker(1988) na seção (2.5) baseia-se no uso da família de transformações do tipo Box e Cox para

modelos lineares cujos erros aleatórios são independentes com distribuição comum dupla-exponencial, utilizando-se a minimização da soma dos erros absolutos. Na seção (2.6) apresentaremos uma proposta de união dos métodos desenvolvidos em (2.4) e (2.5) a fim de acomodarmos os valores aberrantes da variável resposta em um modelo linear e fornecermos robustez a valores aberrantes nos pontos de delineamento, sob minimização dos erros absolutos.

2.2 FAMÍLIA DE TRANSFORMAÇÕES DE BOX E COX

Suponha que observamos um vetor $(n \times 1)$ de resposta Y , $\underline{Y}' = (y_1, y_2, \dots, y_n)$, e que para algum $\underline{\lambda}$ desconhecido, o vetor de respostas transformadas $\underline{Y}^{(\lambda)} = (y_1^{(\lambda)})$ pode ser escrito como:

$$\underline{Y}^{(\lambda)} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad , \quad (2.2.1)$$

onde: \underline{X} : matriz $(n \times p)$ conhecida com i -ésima linha denotada por \underline{x}_i ;

$\underline{\beta}$: vetor $(p \times 1)$ de parâmetros desconhecidos;

$\underline{\varepsilon}$: vetor $(n \times 1)$ de erros aleatórios cuja f.d.p. é $N(0, \sigma^2 \mathbf{1})$;

$\underline{\lambda}$: vetor $(v \times 1)$ de parâmetros da transformação.

(quando $v = 1$ $\underline{\lambda}$ será denotado por λ)

O objetivo agora é, portanto, determinarmos qual valor $\underline{\lambda}$ deve assumir tal que as respostas transformadas satisfaçam, pelo menos aproximadamente, as suposições de modelos lineares sob teoria normal, isto é, satisfaçam (2.2.1).

Box e Cox(1964) sugeriram utilizarmos uma generalização da família de transformações de potência monótona, usada anteriormente por Tukey(1957), como a família de transformações a partir da qual obteremos o valor de λ . Exemplos da família de transformações de Box e Cox são dados a seguir.

Se Y for uma variável resposta positiva ($Y > 0$), um exemplo importante dessa família de transformações é dado por:

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & ; \lambda \neq 0 \\ \log(Y) & ; \lambda = 0 \end{cases} \quad (2.2.2)$$

Chamaremos (2.2.2) por família de transformações simples. Esta família contém as transformações log, inversa e raiz quadrada como casos especiais e é muito útil para obtermos simetria na variável resposta. Um efeito da transformação log, por exemplo, é diminuir uma cauda da distribuição. Geralmente, a utilização de (2.2.2) será sensata em situações onde a origem ocorre naturalmente e a resposta é assimétrica e positiva.

Se respostas negativas ocorrem ou a origem é artificial, temos outro exemplo importante da família de transformações, obtido aumentando a flexibilidade da família de transformações simples para:

$$Y_{\sim}^{(\lambda)} = \begin{cases} \frac{(Y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & ; \lambda_1 \neq 0 \\ \log (Y + \lambda_2) & ; \lambda_1 = 0 \end{cases} \quad (2.2.3)$$

onde $Y \geq -\lambda_2$. Chamaremos (2.2.3) por família de transformações ampliada.

Em algumas situações pode ser satisfatório substituir λ_2 por algum valor adequado e então procedermos usando (2.2.2) em comunhão com a resposta corrigida $Y + \lambda_2$. Em geral, assumimos que para cada λ , $Y_{\sim}^{(\lambda)}$ é uma função monótona de Y sobre o seu domínio.

Box e Cox sugeriram dois procedimentos para obtermos a transformação que satisfaça (2.2.1), um baseado em considerações de verossimilhança e outro baseado em considerações bayesianas. Em nosso trabalho consideraremos somente estimadores sob o método de máxima verossimilhança.

A função de verossimilhança das respostas não transformadas Y_{\sim} , é dada por

$$l(\sigma, \beta, \lambda) = (2\pi\sigma^2)^{-n/2} \exp \left\{ - \frac{(Y_{\sim}^{(\lambda)} - X\beta)' (Y_{\sim}^{(\lambda)} - X\beta)}{2\sigma^2} \right\} J(\lambda, Y_{\sim}) \quad (2.2.4)$$

onde

$$J(\underline{\lambda}, \underline{Y}) = \prod_{i=1}^n \left| \frac{dy_i^{(\lambda)}}{dy_i} \right| \quad (2.2.4.a)$$

é o jacobiano da transformação em relação a \underline{Y} .

Para $\underline{\lambda}$ fixo, o suporte da distribuição não depende de seus parâmetros, logo podemos definir o log da função de verossimilhança por

$$L(\sigma, \underline{\beta}, \underline{Y}) = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{(\underline{Y}^{(\lambda)} - \underline{X}\underline{\beta})'(\underline{Y}^{(\lambda)} - \underline{X}\underline{\beta})}{2\sigma^2} + \log J(\underline{\lambda}, \underline{Y}) \quad (2.2.5)$$

Os estimadores de máxima verossimilhança para $\underline{\beta}$ e σ são obtidos pela solução de $(dL/d\underline{\beta})(\sigma, \underline{\beta}, \underline{Y}) = 0$ e $(dL/d\sigma)(\sigma, \underline{\beta}, \underline{Y}) = 0$.

Denotando, para $\underline{\lambda}$ fixo, por $\hat{\underline{\beta}}(\underline{\lambda})$ e $\hat{\sigma}^2(\underline{\lambda})$ os estimadores de máxima verossimilhança de $\underline{\beta}$ e σ^2 , respectivamente, quando \underline{X} é de posto completo, temos

$$\hat{\underline{\beta}}(\underline{\lambda}) = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}^{(\lambda)}, \quad (2.2.6)$$

$$\hat{\sigma}^2(\underline{\lambda}) = \frac{\underline{Y}^{(\lambda)'}(I - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y}^{(\lambda)}}{n} = \frac{S(\underline{\lambda})}{n}$$

onde $S(\lambda)$ é a soma de quadrados de resíduos obtida quando ajustamos o modelo usando a resposta transformada $Y^{(\lambda)}$.

Os estimadores $\hat{\beta}(\lambda)$ e $\hat{\sigma}^2(\lambda)$ são também os estimadores de mínimos quadrados para a variável resposta $Y^{(\lambda)}$, já que para λ fixo maximizar (2.2.5) equivale a minimizar $(Y^{(\lambda)} - X\beta)'(Y^{(\lambda)} - X\beta)$.

Para λ fixo, temos então que o máximo de (2.2.5), a menos de uma constante, é dado por

$$L_{\max}(\lambda) = -0.5 n \log \hat{\sigma}^2(\lambda) + \log J(\lambda, Y) \quad (2.2.7)$$

Quando utilizamos o caso especial $Y^{(\lambda)} = (Y^\lambda - 1)/\lambda$ da transformação de potência simples (2.2.2), temos que o termo $\log J(\lambda, Y)$ em (2.2.7) é dado por

$$(\lambda - 1) \sum_{i=1}^n \log y_i \quad (2.2.8)$$

e no caso $Y^{(\lambda)} = ((Y + \lambda_2)^{\lambda_1} - 1)/\lambda_1$ da transformação de potência ampliada (2.2.3), esse termo é dado por

$$(\lambda_1 - 1) \sum_{i=1}^n \log (y_i + \lambda_2) \quad (2.2.9)$$

Quando λ é um escalar, Box e Cox sugerem fazermos um gráfico de $L_{\max}(\lambda)$ versus λ para alguns valores selecionados de λ , e a partir deste gráfico obtermos o valor de λ que maximiza L_{\max} , ou seja, um estimador aproximado de máxima verossimilhança para λ . Denotaremos tal valor de λ por $\hat{\lambda}$.

Podemos obter, aproximadamente, uma região de $(1-\alpha)100\%$ de confiança para $\hat{\lambda}$ a partir do conjunto de todos $\tilde{\lambda}$ satisfazendo a desigualdade

$$L_{\max}(\hat{\lambda}) - L_{\max}(\tilde{\lambda}) \leq 0.5 \chi_{(\alpha, v)}^2, \quad (2.2.10)$$

onde $\chi_{(\alpha, v)}^2$ representa o α -ésimo percentil de uma distribuição qui-quadrado com v graus de liberdade.

Para λ um escalar, esta região de confiança será facilmente construída a partir do gráfico de $L_{\max}(\lambda)$ versus λ .

Quando o objetivo final do estudo é a escolha do valor de $\tilde{\lambda}$, desejamos estimar $\tilde{\lambda}$ mais precisamente; neste caso determinamos o valor de $\hat{\tilde{\lambda}}$ maximizando (2.2.7), ou seja, achando, através de um método numérico, a solução para $(dL_{\max}(\tilde{\lambda}))/d\tilde{\lambda} = 0$.

Podemos simplificar os resultados obtidos acima através da padronização da transformação, isto é, trabalhando com a seguinte resposta transformada

$$\underline{Z}^{(\lambda)} = \frac{Y^{(\lambda)}}{J^{1/n}} \quad (2.2.11)$$

onde $J = J(\lambda, Y)$. Neste caso, o estimador de máxima verossimilhança de σ^2 , para λ fixo, quando a matriz X é de posto completo, é dado por

$$\hat{\sigma}^2(\lambda, Z) = \frac{\underline{Z}^{(\lambda)'} (I - X(X'X)^{-1}X') \underline{Z}^{(\lambda)}}{n} = \frac{S(\lambda, Z)}{n} \quad (2.2.12)$$

onde $S(\lambda, Z)$ é a soma de quadrados dos resíduos quando ajustamos o modelo usando a resposta transformada padronizada $\underline{Z}^{(\lambda)}$ e teremos, portanto,

$$L_{\max}(\lambda) = -0.5 n \log \hat{\sigma}^2(\lambda, Z). \quad (2.2.13)$$

O estimador de máxima verossimilhança $\hat{\lambda}$ é obtido minimizando $S(\lambda, Z)$ com respeito a λ .

A transformação padronizada, por apresentar a soma de quadrados dos resíduos na escala original para todos os valores de λ , torna-se bastante conveniente quando desejamos comparar vários modelos, pois ela proporciona uma base comum de comparação para esses modelos.

Para a família de potência simples (2.2.2), a transformação padronizada é dada por

$$Z^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda (\text{mg}(\underline{Y}))^{\lambda-1}} & , \lambda \neq 0 \\ \text{mg}(\underline{Y}) \log(Y) & , \lambda = 0 \end{cases} \quad (2.2.14)$$

onde $\text{mg}(\underline{Y})$ é a média geométrica do vetor de respostas observadas \underline{Y} .

A transformação padronizada para a família de potência ampliada (2.2.3) é dada por

$$Z^{(\lambda)} = \begin{cases} \frac{(\underline{Y} + \lambda_2 \underline{1})^{\lambda_1} - 1}{\lambda_1 (\text{mg}(\underline{Y} + \lambda_2 \underline{1}))^{\lambda_1-1}} & , \lambda_1 \neq 0 \\ \text{mg}(\underline{Y} + \lambda_2 \underline{1}) \log(\underline{Y} + \lambda_2 \underline{1}) & , \lambda_1 = 0 \end{cases} \quad (2.2.15)$$

onde $\underline{1}$ é um vetor $(n \times 1)$ de 1's e $\text{mg}(\underline{Y} + \lambda_2 \underline{1})$ é a média geométrica do vetor de observações corrigidas $\underline{Y} + \lambda_2 \underline{1}$.

2.3 TESTES PARA TRANSFORMAÇÕES A FIM DE OBTER NORMALIDADE APROXIMADA

O método de verossimilhança desenvolvido por Box e Cox fornece um teste para a hipótese nula $\underline{\lambda} = \underline{\lambda}_0$ sob o modelo (2.2.1), onde utilizamos a distribuição χ^2 assintótica com v graus de liberdade, do log da razão de verossimilhança. Para facilitar a comparação com outros métodos a serem apresentados, utilizaremos a raiz quadrada da estatística, obtendo assim

$$T_v(\lambda_0) = \left[-2 \left\{ L_{\max}(\lambda_0) - L_{\max}(\hat{\lambda}) \right\} \right]^{1/2} \quad (2.3.1)$$

Assintoticamente a distribuição nula de T_v será normal padrão.

Observamos, entretanto, que as regiões de confiança obtidas por (2.2.10) e os testes para $H_0: \underline{\lambda} = \underline{\lambda}_0$ dados pela estatística (2.3.1) possuem somente validade assintótica, onde o número de parâmetros deve ser pequeno comparado com o número de observações.

Andrews(1971) mostrou através de um exemplo que o método de verossimilhança para a escolha de uma transformação possui também a desvantagem de ser sensível a observações aberrantes, confirmando assim o fato que o método de máxima verossimilhança sob o modelo normal não é robusto.

Nesta seção, apresentaremos dois métodos adicionais para avaliar a necessidade de uma transformação na variável resposta. Estes métodos, pesquisados por Andrews(1971) e Atkinson(1973), são um

suporte necessário à análise de verossimilhança devido à facilidade dos seus cálculos e poderes em diagnosticar a necessidade de uma transformação.

2.3.1 MÉTODO DE ANDREWS

O método de Andrews é baseado em um teste exato para a hipótese nula $\underline{\lambda} = \underline{\lambda}_0$, onde $\underline{\lambda}_0$ é um valor dado para $\underline{\lambda}$, para o modelo (2.2.1).

Assumindo que a transformação $Y^{(\underline{\lambda})}$ seja suficientemente regular, podemos aproximá-la através de uma expansão linear em torno de $\underline{\lambda}_0$ e escrever

$$\underline{Y}^{(\underline{\lambda})} = \underline{Y}^{(\underline{\lambda}_0)} + V^{(\underline{\lambda}_0)} (\underline{\lambda} - \underline{\lambda}_0),$$

$$\text{tal que } y_1^{(\underline{\lambda})} = y_1^{(\underline{\lambda}_0)} + v_1^{(\underline{\lambda}_0)} (\underline{\lambda} - \underline{\lambda}_0),$$

(2.3.1.1)

$$\text{onde } v_{1j}^{(\underline{\lambda}_0)} = \left[\frac{\partial (y_1^{(\underline{\lambda})})}{\partial \lambda_j} \right]_{\underline{\lambda} = \underline{\lambda}_0}, \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, v \end{matrix}$$

Então,

$$\underline{Y}^{(\underline{\lambda}_0)} = X\underline{\beta} + V^{(\underline{\lambda}_0)}\underline{\phi} + \underline{\varepsilon} \quad (2.3.1.2)$$

onde $\underline{\phi} = -(\underline{\lambda} - \underline{\lambda}_0)$.

Como podemos notar a matriz $V_{\sim 0}^{(\underline{\lambda})}$ depende de \underline{Y} . O modelo (2.2.1) sob a hipótese $H_0: \underline{\lambda} = \underline{\lambda}_0$ proporciona valores ajustados para \underline{Y} , logo a matriz $V_{\sim 0}^{(\underline{\lambda})}$ pode ser aproximada para $\hat{V}_{\sim 0}^{(\underline{\lambda})}$ usando os valores ajustados, tal que

$$\hat{v}_{ij} = \left[\frac{\partial \{y_i^{(\underline{\lambda})}\}}{\partial \lambda_j} \right]_{\underline{\lambda} = \underline{\lambda}_0, \underline{Y} = \hat{\underline{Y}}} \quad , \quad (2.3.1.3)$$

a fim de obtermos um teste simples.

Podemos, agora, construir um teste para a hipótese nula $H_0: \underline{\lambda} = \underline{\lambda}_0$ a partir do modelo modificado

$$\underline{Y}_{\sim 0}^{(\underline{\lambda})} = X\underline{\beta} + \hat{V}_{\sim 0}^{(\underline{\lambda})}\underline{\phi} + \underline{\varepsilon} \quad . \quad (2.3.1.4)$$

Observamos que testar a hipótese $\underline{\lambda} = \underline{\lambda}_0$ é idêntico a testar $H_0: \underline{\phi} = \underline{0}$.

O modelo (2.3.1.4) pode ser apresentado da seguinte maneira:

$$\underline{Y}_{\sim 0}^{(\underline{\lambda})} = X^* \underline{\beta}^* + \underline{\varepsilon}$$

onde $X^* = \begin{pmatrix} X & | & \hat{V}_{-0}^{(\lambda)} \end{pmatrix}$

e $\underline{\beta}^* = \begin{pmatrix} \underline{\beta} \\ \underline{\phi} \end{pmatrix}$

O estimador de mínimos quadrados de $\underline{\beta}^*$ é dado por

$$\begin{aligned} \hat{\underline{\beta}}^* &= \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{\phi}} \end{pmatrix} = \left(X^{*'} X^* \right)^{-1} X^{*'} \underline{Y}_{-0}^{(\lambda)} \\ &= \begin{pmatrix} X'X & X' \hat{V}_{-0}^{(\lambda)} \\ \hat{V}_{-0}^{(\lambda)'} X & \hat{V}_{-0}^{(\lambda)'} \hat{V}_{-0}^{(\lambda)} \end{pmatrix}^{-1} \begin{pmatrix} X' \\ \hat{V}_{-0}^{(\lambda)'} \end{pmatrix} \cdot \underline{Y}_{-0}^{(\lambda)} \\ &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} X' \underline{Y}_{-0}^{(\lambda)} \\ \hat{V}_{-0}^{(\lambda)'} \underline{Y}_{-0}^{(\lambda)} \end{pmatrix} \\ &= \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix} \begin{pmatrix} X' \underline{Y}_{-0}^{(\lambda)} \\ \hat{V}_{-0}^{(\lambda)'} \underline{Y}_{-0}^{(\lambda)} \end{pmatrix} \end{aligned}$$

onde (Mardia(1979) p. 459) :

$$A^{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$$

$$A^{12} = - A^{11} A_{12} A_{22}^{-1}$$

$$A^{22} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$$

$$A^{21} = - A^{22} A_{21} A_{11}^{-1}$$

Portanto

$$\hat{\underline{\phi}} = \left(\hat{V}_{-0}^{(\lambda)} \underline{X}_r \hat{V}_{-0}^{(\lambda)} \right)^{-1} \cdot \hat{V}_{-0}^{(\lambda)} \underline{X}_r \underline{Y}_{-0}^{(\lambda)} \quad (2.3.15)$$

onde $\underline{X}_r = I - X(X'X)^{-1}X'$

$\hat{\underline{\phi}}$ pode ser expresso por uma constante vezes $\underline{Y}_{-0}^{(\lambda)}$,
portanto

$$\hat{\underline{\phi}} \sim N_v \left(\underline{\phi}, \sigma^2 (\hat{V}_{-0}^{(\lambda)} \underline{X}_r \hat{V}_{-0}^{(\lambda)})^{-1} \right)$$

Através da generalização da distribuição t-Student univariada (Mardia(1979) p. 43), temos que a estatística de teste para a hipótese $\underline{\phi} = \underline{0}$ é dada por

$$T_{E \sim 0}(\lambda) = - \hat{\sigma}^2 (\hat{V}_{\sim 0}^{(\lambda)'} X_r \hat{V}_{\sim 0}^{(\lambda)})^{-1/2} \hat{V}_{\sim 0}^{(\lambda)'} X_r \tilde{Y}_{\sim 0}^{(\lambda)} \quad (2.3.1.6)$$

onde $\hat{\sigma}^2$ é um estimador da variância σ^2 , dado pelo quadrado médio dos resíduos do modelo (2.3.1.4). Sob a hipótese nula, $(T_{E \sim 0}(\lambda))'(T_{E \sim 0}(\lambda))$ possui distribuição F com v e $n-p-v$ graus de liberdade.

Quando $\tilde{Y}_{\sim 0}^{(\lambda)}$ corresponde a qualquer família de transformação uni-paramétrica, ou seja $v = 1$, temos que para σ^2 conhecido, $T_{E \sim 0}$ possui exatamente uma distribuição normal padrão, e para σ^2 desconhecido, $T_{E \sim 0}$ possui distribuição t-Student univariada com $n-p-1$ graus de liberdade, sob a hipótese nula.

Este teste exato proposto por Andrews, pode também ser obtido, como mencionado por Atkinson(1973), desenvolvendo-o a partir da soma de quadrados dos resíduos de $\tilde{Y}_{\sim 0}^{(\lambda)}$. Por este procedimento, inicialmente diferenciamos a soma de quadrados dos resíduos para o modelo (2.2.1) com respeito a $\tilde{\lambda}$ e para formar o teste exato substituímos, então, $V_{\sim 0}^{(\lambda)}$ por $\hat{V}_{\sim 0}^{(\lambda)}$ dado em (2.3.1.3).

O método de Andrews possui três vantagens sobre o método de verossimilhança : (i) fornece um teste de significância exato, não necessitando assim que o número de parâmetros seja pequeno comparado com o número de observações; (ii) reduz a quantidade de cálculos em relação ao teste da razão de verossimilhança; (iii) através de exemplos, Andrews mostrou que seu método é mais robusto a valores aberrantes que o método de Box e Cox.

Estas propriedades vantajosas, entretanto, são obtidas ignorando o jacobiano da transformação, um ponto que Andrews não torna explícito.

2.3.2 MÉTODO DE ATKINSON

Assim como o método de Andrews, o método de Atkinson(1973) baseia-se em um teste para a hipótese $\lambda = \lambda_0$, para o modelo (2.2.1). Para o desenvolvimento deste método, Atkinson considerou somente famílias de transformações padronizadas uni-paramétricas, onde λ é um escalar.

A estatística de teste pode ser construída expandindo $\underline{Z}^{(\lambda)}$ em torno de λ_0 ,

$$\underline{Z}^{(\lambda)} = \underline{Z}^{(\lambda_0)} + (\lambda - \lambda_0) \underline{D}^{(\lambda_0)}, \quad (2.3.2.1)$$

onde

$$\underline{D}^{(\lambda_0)} = \left[\frac{\partial \underline{Z}^{(\lambda)}}{\partial \lambda} \right]_{\lambda = \lambda_0} \quad (2.3.2.1.a)$$

Como $\underline{Z}^{(\lambda)} = \underline{X}\underline{\beta} + \underline{\varepsilon}$,

$$\underline{Z}^{(\lambda_0)} = \underline{X}\underline{\beta} + \phi \underline{D}^{(\lambda_0)} + \underline{\varepsilon}, \quad (2.3.2.2)$$

onde $\phi = -(\lambda - \lambda_0)$.

Este modelo é similar ao modelo (2.3.1.2) usado na construção da estatística $T_E(\lambda_0)$. A estatística de teste para a

hipótese $\lambda = \lambda_0$ é, fora uma mudança de sinal, igual a estatística de teste para a hipótese $\phi = 0$ no modelo (2.3.2.2), dada por

$$T_A(\lambda_0) = - \frac{\tilde{D}^{(\lambda_0)'} \tilde{X}_r \tilde{Z}^{(\lambda_0)}}{\hat{\sigma}_z^2 \{\tilde{D}^{(\lambda_0)'} \tilde{X}_r \tilde{D}^{(\lambda_0)}\}^{1/2}}, \quad (2.3.2.3)$$

onde $\hat{\sigma}_z^2$ é uma estimativa da variância de $\tilde{Z}^{(\lambda)}$, dada pelo quadrado médio dos resíduos para o modelo (2.3.2.2). O sinal negativo em (2.3.2.3) ocorre porque $\phi = -(\lambda - \lambda_0)$.

Cabe aqui ressaltarmos que as estatísticas de teste propostas nessa seção e na seção anterior diferem pela utilização ou não de uma aproximação para a derivada da variável resposta transformada em relação a λ . Andrews(1971) ao utilizar essa aproximação para a derivada possibilitou a construção de um teste exato, enquanto que Atkinson(1973) não utilizou tal aproximação e trabalhou com distribuição e teste assintóticos.

Assintoticamente, a distribuição nula de $T_A(\lambda_0)$ é normal padrão, mas sua distribuição em pequenas amostras é intratável, pois ambos $\tilde{Z}^{(\lambda_0)}$ e $\tilde{D}^{(\lambda_0)}$ são variáveis aleatórias com distribuições não padronizadas.

O estimador de mínimos quadrados de ϕ dado por:

$$\hat{\phi} = \frac{\tilde{Z}^{(\lambda_0)'} \tilde{X}_r \tilde{D}^{(\lambda_0)}}{\tilde{D}^{(\lambda_0)'} \tilde{X}_r \tilde{D}^{(\lambda_0)}} \quad (2.3.2.4)$$

pode ser usado para obtermos um estimador de λ , dado por

$$\tilde{\lambda} = \lambda_0 - \hat{\phi} \quad (2.3.2.5)$$

Este método proposto por Atkinson pode também ser obtido através de uma aproximação para o teste usado por Neyman(1959) que é desenvolvido a partir da inclinação da função de verossimilhança.

Atkinson(1973) mostrou através de estudos de Monte Carlo que, sob normalidade, $T_V(\lambda_0)$ e $T_A(\lambda_0)$ fornecem testes com poderes semelhantes. Nenhum ganho é obtido no poder do teste por calcularmos a estimativa de máxima verossimilhança $\hat{\lambda}$, observando que todas as informações parecem estar contidas na inclinação da verossimilhança sob a hipótese nula.

A vantagem do teste da razão de verossimilhança é que a distribuição de T_V sob a hipótese nula, está mais próxima da distribuição normal padrão que a distribuição de T_A . Atkinson mostrou também que embora o método de Andrews forneça um teste exato, sob a distribuição normal, os testes $T_A(\lambda_0)$ e $T_V(\lambda_0)$ são uniformemente mais poderosos que $T_E(\lambda_0)$. O método de Andrews é, entretanto, mais robusto a valores aberrantes que os outros dois métodos considerados.

Assim como o método de Andrews, este método possui a vantagem sobre o método da razão de verossimilhança, dado em (2.3.1), de não precisar calcular o estimador de máxima verossimilhança $\hat{\lambda}$.

2.4 ALTERNATIVAS ROBUSTAS PARA A ESTIMAÇÃO DO PARÂMETRO DE TRANSFORMAÇÃO

O estimador de máxima verossimilhança utilizado por Box e Cox, como já mencionado anteriormente, é sensível a valores aberrantes, podendo assim ser altamente ineficiente se a distribuição dos erros tiver caudas mais "pesadas" que a distribuição normal. Como as transformações de potência são, entretanto, destinadas a obter somente normalidade aproximada, Carroll(1980) deu o primeiro passo na utilização de estimadores robustos para o parâmetro de transformação, ao desenvolver um método que considera situações onde os erros estão próximos à normalidade, possuindo todavia caudas possivelmente mais "pesadas". Com o propósito de fornecer robustez também a pontos de alta ruptura, ou seja, a pontos do delineamento que estão bastante afastados do centro de massa(média) dos outros valores \underline{x} , Carroll e Ruppert(1985) desenvolveram dois métodos, a partir de idéias de pseudo-verossimilhança e de função de influência limitada.

2.4.1 MÉTODO DE CARROLL

Carroll(1980), baseando-se na teoria de Huber(1964), sugere utilizarmos em lugar do modelo (2.2.1), um modelo cujos erros aleatórios estejam próximos a normalidade, possuindo todavia caudas mais "pesadas". O modelo utilizado é dado por

$$\underline{Y}^{(\lambda)} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad , \quad (2.4.1.1)$$

onde \underline{X} e $\underline{\beta}$ são definidos como em (2.2.1), λ é um escalar, ou seja, $Y^{(\lambda)}$ corresponde a famílias de transformações uniparamétricas e $\underline{\varepsilon}$ é o vetor $(n \times 1)$ de erros aleatórios independentes e identicamente distribuídos cuja densidade é proporcional a $\exp\{-\rho((y_1^{(\lambda)} - \underline{x}_1 \underline{\beta})/\sigma)\}$, $i = 1, 2, \dots, n$, sendo que para algum k ,

$$\rho(x) = \begin{cases} \frac{x^2}{2} & ; |x| \leq k \\ k \left(|x| - \frac{k}{2} \right) & ; |x| > k \end{cases} \quad (2.4.1.2)$$

O valor de k , geralmente, é escolhido pelo estatístico, com $k = 1.5$ ou 2.0 . A função densidade dos erros aleatórios do modelo (2.4.1.1) é, portanto, normal no centro e exponencial nas caudas.

Os estimadores de máxima verossimilhança obtidos para o modelo (2.4.1.1) são os estimadores mais robustos em situações onde os erros são de fato independentes com função de distribuição comum F , onde F pertence à classe $\mathfrak{F} = \{G : G = (1-\alpha)\Phi + \alpha H, H \text{ simétrica}\}$, $0 < \alpha < 1$ e Φ é a distribuição normal padrão, ou seja, F pertence à classe de distribuições normais contaminadas.

A função de verossimilhança para o modelo (2.4.1.1) é proporcional a

$$l(\underline{\beta}, \sigma, \lambda) = \sigma^{-n} \exp \left\{ \sum_{i=1}^n -\rho \left(\frac{y_1^{(\lambda)} - \underline{x}_1 \underline{\beta}}{\sigma} \right) \right\} J(\lambda, \underline{Y}) \quad (2.4.1.3)$$

Para λ fixo, a fim de obtermos os estimadores de máxima verossimilhança para $\underline{\beta}$ e σ procedemos como no problema de locação dado por Huber, onde tomamos um valor inicial para σ e maximizamos (2.4.1.3) com respeito a $\underline{\beta}$ resolvendo a equação

$$\sum_{i=1}^n \psi \left\{ \frac{y_i^{(\lambda)} - \underline{x}_i \underline{\beta}}{\sigma} \right\} \cdot \underline{x}_i = 0 \quad (2.4.1.4)$$

onde $\psi (\cdot)$ é a derivada de $\rho (\cdot)$.

A seguir, estimamos σ resolvendo a equação adicional

$$(n-p)^{-1} \sum_{i=1}^n \psi^2 \left\{ \frac{y_i^{(\lambda)} - \underline{x}_i \underline{\beta}}{\sigma} \right\} = E_{\Phi} \psi^2(Z) \quad (2.4.1.5)$$

onde calculamos a esperança do lado direito de (2.4.1.5) sob a distribuição normal padrão. Tal procedimento é realizado iterativamente até convergir ao estimador de σ .

Para um dado λ , denominamos as estimativas obtidas por $\hat{\underline{\beta}}_r$ e $\hat{\sigma}_r$. O valor estimado de λ é obtido então maximizando $l(\hat{\underline{\beta}}_r, \hat{\sigma}_r, \lambda)$ com respeito a λ , onde a solução é denotada por $\hat{\lambda}_r$. É interessante notarmos que quando $k = \infty$, este método produz o mesmo resultado obtido por Box e Cox.

Por este método podemos também construir um teste da razão de verossimilhança para a hipótese $\lambda = \lambda_0$, obtendo assim a

estatística

$$\Lambda_c = -2 \log \left\{ \frac{l(\hat{\beta}_r(\lambda_0), \hat{\sigma}_r(\lambda_0), \lambda_0)}{l(\hat{\beta}_r(\hat{\lambda}_r), \hat{\sigma}_r(\hat{\lambda}_r), \hat{\lambda}_r)} \right\} \quad (2.4.1.6)$$

que, para $\lambda = \lambda_0$ e n grande, é assintoticamente χ^2 com um grau de liberdade.

Através de um estudo de Monte Carlo, Carroll(1980) mostrou que este método é preferível ao método de Box e Cox em termos de inferência sobre λ e β , por oferecer poder de teste compatível ao poder deste último e pelo fato de seu erro tipo I estar mais próximo à significância nominal, quando consideramos a possibilidade dos erros possuírem outra distribuição que não seja a normal. Este método parece, também, preferível em relação ao método de Andrews, pois ele é consideravelmente mais poderoso que este último, embora o método de Carroll pareça ligeiramente menos robusto a valores aberrantes que o método de Andrews.

Este estimador proposto por Carroll é, entretanto, destinado a fornecer robustez somente a valores aberrantes nos erros(ϵ_i), mas não fornece robustez a valores aberrantes nas variáveis explanatórias(X). É importante, todavia, termos estimadores que possam fornecer robustez, também, a valores aberrantes nos pontos de delineamento, uma vez que os modelos de regressão são aproximações adequadas para a realidade somente se a variação dos valores \underline{x} for moderada, não estando aptos, portanto, a trabalhar com valores aberrantes em \underline{x} , que podem ser erros grosseiros.

Nas subseções a seguir, apresentaremos os estimadores

propostos por Carroll e Ruppert(1985) que são robustos a valores aberrantes em ambos $\underline{\varepsilon}$ e \underline{x} .

2.4.2 ESTIMADORES DE PSEUDO-VEROSSIMILHANÇA

Nessa seção obteremos estimadores para os parâmetros de um modelo de regressão e para λ a partir da idéia de pseudo-verossimilhança. Basicamente, esta idéia envolve uma modificação da definição de log da verossimilhança, ponderando-se adequadamente cada observação, a fim de obtermos certas propriedades desejáveis, como a robustez a valores aberrantes e pontos de alta ruptura.

Suponha que observamos uma amostra (y_i, \underline{x}_i) , $i=1,2,\dots,n$ que é distribuída de acordo com o modelo

$$Y^{(\lambda)} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (2.4.2.1)$$

onde λ , \underline{X} e $\underline{\beta}$ são definidos em (2.2.1) e $\underline{\varepsilon}$ é o vetor de erros aleatórios independentes e identicamente distribuídos cuja densidade é proporcional a $\exp\{-\rho((y_i^{(\lambda)} - \underline{x}_i\underline{\beta})/\sigma)\}$, onde $\rho(\cdot)$ é alguma função com derivada $\psi(\cdot)$.

Carroll e Ruppert(1985) trabalharam somente com $Y_1^{(\lambda)}$, $i = 1,2,\dots,n$, pertencentes à família das transformações de potência simples dada em (2.2.2).

Seja $\{(s_i, t_i)\}$ conjuntos de pesos entre zero e um. O logaritmo da pseudo-verossimilhança para o modelo (2.4.2.1) é, então, definido como

$$-\sum_{i=1}^n s_i t_i \left\{ \log \sigma^* + \rho \left(\frac{y_i^{(\lambda)} - \underline{x}_i \beta}{\sigma^* t_i C(\lambda)} \right) \right\}, \quad (2.4.2.2)$$

onde $C(\lambda) = \left\{ J(\lambda, \underline{Y}) \right\}^{1/n}$ e $\sigma^* = \sigma/C(\lambda)$.

Os estimadores de pseudo-verossimilhança para os parâmetros do modelo de regressão (2.4.2.1) são aqueles que maximizam (2.4.2.2). Se $s_i = t_i = 1$ e $\rho(x) = x^2/2$, então (2.4.2.2) se reduz ao log de verossimilhança usual dado por (2.2.5); observamos assim que os estimadores de pseudo-verossimilhança incluem os estimadores de Box e Cox.

Se

$$\rho(x) = \begin{cases} \frac{x^2}{2} & ; |x| \leq k \\ k \left(|x| - \frac{k}{2} \right) & ; |x| > k \end{cases} \quad (2.4.2.3)$$

e $s_i = t_i = 1$, obteremos os estimadores propostos por Carroll(1980).

Para uma maior proteção a grandes valores aberrantes na variável resposta, Carroll(1982) sugere o uso da função

$$\begin{aligned} \rho'(x) &= \psi(x) \\ &= -\psi(-x) = x && ; 0 \leq x < a \\ &= a && ; a \leq x < b \\ &= a(c-x)/(c-b) && ; b \leq x < c \\ &= 0 && ; x \geq c \end{aligned}$$

com $s_i = t_i = 1$.

O propósito dos termos $\{(s_i, t_i)\}$ em (2.4.2.2) é fornecer alguma proteção à influência conjunta de valores aberrantes na resposta e pontos de alta ruptura. A escolha de $\{(s_i, t_i)\}$ é portanto muito importante.

Carroll e Ruppert(1985) utilizaram em seus exemplos $s_i = t_i = w_{x,i}$, onde $w_{x,i}$ é definido como os pesos de delineamento para a matriz de 2º momento de $\{x\}$ ponderada, dada por

$$A = n^{-1} \sum_{i=1}^n w_{x,i} \begin{matrix} x_i' \\ x_i \end{matrix} \quad (2.4.2.4)$$

A matriz A ($p \times p$) atribui pesos baixos a valores aberrantes e é obtida através de iterações, com $w_{x,i}$ satisfazendo

$$w_{x,i} = \min \left\{ 1, p \lambda_x (x_i' A^{-1} x_i)^{-1} \right\} \quad (2.4.2.5)$$

Na expressão acima, a_x deve ser uma constante maior que 1. Carroll e Ruppert experimentaram valores para a_x entre 1.1 e 1.7. A determinação do valor exato para a_x não pareceu ser crucial. Eles estabeleceram, então, em seus exemplos o valor 1.4 para a_x por acomodar razoavelmente valores aberrantes. Os valores atribuídos a $\{(s_i, t_i)\}$ e a_x por Carroll e Ruppert fornecem eficientemente pouco peso a respostas aberrantes que ocorrem em pontos de delineamento discrepantes.

Consequentemente, para λ fixo, os estimadores de (β, σ) satisfazem

$$\sum_{i=1}^n x_{i,1} w_{x,i} w_{r,i}(\lambda) r_i(\lambda) = 0 \quad (2.4.2.6)$$

e

$$\sum_{i=1}^n w_{x,i}^2 w_{r,i}(\lambda) \left(r_i^2(\lambda) - 1 \right) = 0 \quad (2.4.2.7)$$

onde

$$r_i(\lambda) = \frac{y_i^{(\lambda)} - x_{i,1} \beta}{\sigma^* w_{x,i} C(\lambda)} \quad (2.4.2.8)$$

e

$$w_{r,i}(\lambda) = \frac{\psi(r_i(\lambda))}{r_i(\lambda)} \quad (2.4.2.9)$$

Para testarmos $H_0: \lambda = \lambda_0$, obtemos um teste robusto,

substituindo o log de verossimilhança (2.2.5) pelo pseudo log de verossimilhança (2.4.2.2). Especificamente, seja $L_{\max}^*(\lambda)$ o máximo de (2.4.2.2) com relação a β e σ . A estatística de teste robusta é

$$T_p(\lambda_0) = 2 D \cdot \left[L_{\max}^*(\hat{\lambda}) - L_{\max}^*(\lambda_0) \right] \quad (2.4.2.10)$$

onde D, o fator de correção é dado por

$$D = \frac{\sum_{i=1}^n \left(\frac{\psi(r_i / \sigma t_i) \xi_i^2 w_{x,i}}{t_i} \right)}{\sum_{i=1}^n \psi_k^2(r_i / \sigma t_i) \xi_i^2 w_{x,i}} \quad (2.4.2.11)$$

$$e \quad \xi_i^2 = \left[\frac{\partial (x_i \beta)^{(\lambda)}}{\partial \lambda} \right]_{\lambda = \hat{\lambda}, \beta = \hat{\beta}(\hat{\lambda})}$$

O fator de correção é introduzido para que $T_p(\lambda_0)$ seja assintoticamente chi-quadrado com um grau de liberdade, sob a hipótese $H_0: \lambda = \lambda_0$.

2.4.3 ESTIMADORES DE INFLUÊNCIA LIMITADA (EIL)

Este método visa, também, fornecer um estimador para o parâmetro de transformação da família (2.2.2) que seja robusto a valores aberrantes em \underline{y} e \underline{X} . Baseia-se na função de influência limitada.

$$\text{Seja : } \underline{\theta} = (\underline{\beta}, \sigma, \lambda); \quad (2.4.3.1)$$

$f(y_1, \underline{x}_1, \underline{\theta})$ = densidade de y_1 dado \underline{x}_1 e $\underline{\theta}$;

$l(y_1, \underline{x}_1, \underline{\theta}) = \nabla_{\underline{\theta}} \log f(y_1, \underline{x}_1, \underline{\theta})$, ou seja ,
o gradiente do log de $f(y_1, \underline{x}_1, \underline{\theta})$.

O estimador de influência limitada $\hat{\underline{\theta}}$ é dado pelo valor de $\underline{\theta}$ que resolve

$$\sum_{i=1}^n w(y_i, \underline{x}_i, \underline{\theta}) \left[l(y_i, \underline{x}_i, \underline{\theta}) \right] = 0 \quad , \quad (2.4.3.2)$$

onde $w(y_i, \underline{x}_i, \underline{\theta})$ são os pesos da matriz \mathbb{B} ,

$$\mathbb{B} = n^{-1} \sum_{i=1}^n E_Y \left\{ w^2(y_i, \underline{x}_i, \underline{\theta}) l(y_i, \underline{x}_i, \underline{\theta}) (l(y_i, \underline{x}_i, \underline{\theta}))' \right\} , \quad (2.4.3.3)$$

com $w(y, \underline{x}, \underline{\theta})$ satisfazendo

$$w(y, \underline{x}, \underline{\theta}) = \min \left\{ 1, a(p+2)^{1/2} [(l(y, \underline{x}, \underline{\theta}))' B^{-1} l(y, \underline{x}, \underline{\theta})]^{-1/2} \right\}$$

onde a é uma constante maior que um. É interessante notarmos que, a medida que $a \rightarrow \infty$, $w(y, \underline{x}, \underline{\theta}) \rightarrow 1$, para $i=1,2,\dots,n$; conseqüentemente $\hat{\underline{\theta}}$ converge para o estimador de máxima verossimilhança de $\underline{\theta}$.

Os pesos $w(y, \underline{x}, \underline{\theta})$ associados às observações possuem valor entre zero e um, sendo que pesos pequenos são atribuídos a (i) pontos de alta ruptura, (ii) valores aberrantes, (iii) grandes valores absolutos de $(\partial/\partial\lambda)\log f(y, \underline{x}_i, \hat{\underline{\theta}})$ ou alguma combinação de (i), (ii) e (iii). O critério de atribuição de pesos às observações é baseado em uma função do inverso da distância tipo Mahalanobis (Mardia(1979)-p.17) de $l(y, \underline{x}_i, \hat{\underline{\theta}})$ em relação ao centro de $\{ l(y, \underline{x}_i, \hat{\underline{\theta}}), i=1,2,\dots,n \}$. A distância tipo Mahalanobis possui valor alto quando a observação é discrepante.

A matriz B é uma versão robusta da matriz de segundo momento de $l(y, \underline{x}, \underline{\theta})$. A equação (2.4.3.2) mostra que o estimador de influência limitada da transformação é um estimador de máxima verossimilhança ponderado, enquanto que o estimador de pseudo-verossimilhança é obtido modificando-se o log de verossimilhança e então maximizando-o.

Tanto B quanto $\underline{\theta}$ são estimados através de um processo iterativo (Carroll e Ruppert(1985)), onde utilizamos como estimativa preliminar para $\underline{\theta}$, os estimadores de máxima verossimilhança do modelo (2.2.1) e como estimativa preliminar de B , o resultado de (2.4.3.3) após substituirmos $\underline{\theta}$ por seu valor preliminar e $w(y, \underline{x}_i, \underline{\theta}) = 1$, $i=1,2,\dots,n$. Calculamos, então, iterativamente, os valores para os pesos da matriz B , as estimativas de $\underline{\theta}$ e B , nessa ordem apresentada,

até convergirmos ao estimador de influência limitada, $\hat{\underline{\theta}}$, que deve satisfazer o seguinte critério

$$\| \hat{\underline{\theta}}^{(j)} - \hat{\underline{\theta}}^{(j-1)} \|_1 / \left(\| \hat{\underline{\theta}}^{(j-1)} \|_1 + 10^{-5} \right) \leq 0.02$$

onde $\hat{\underline{\theta}}^{(j)}$ representa o estimador de influência limitada, $\hat{\underline{\theta}}^{(j-1)}$ representa a estimativa de $\underline{\theta}$ no passo anterior a $\hat{\underline{\theta}}^{(j)}$ e $\| \underline{x} \|_1$ é a soma das coordenadas absolutas de \underline{x} .

Carroll e Ruppert(1985) mostraram que $\hat{\underline{\theta}}$ possui influência limitada usando a função de influência (FI) calculada em (y_1, \underline{x}_1) ,

$$FI(y_1, \underline{x}_1, \hat{\underline{\theta}}) = C^{-1} w(y_1, \underline{x}_1, \hat{\underline{\theta}}) l(y_1, \underline{x}_1, \hat{\underline{\theta}})$$

onde

$$C = n^{-1} \sum_{i=1}^n E_y \left\{ w^2(y_i, \underline{x}_i, \hat{\underline{\theta}}) l(y_i, \underline{x}_i, \hat{\underline{\theta}}) (l(y_i, \underline{x}_i, \hat{\underline{\theta}}))' \right\}$$

e uma medida escalar de influência (Krasker e Weisch(1982)) dada por

$$\begin{aligned} \gamma_2 &= \max_i \left\{ (FI(y_i, \underline{x}_i, \hat{\underline{\theta}}))' W^{-1} FI(y_i, \underline{x}_i, \hat{\underline{\theta}}) \right\}^{1/2} \\ &= \max_i \left\{ (l(y_i, \underline{x}_i, \hat{\underline{\theta}}))' B^{-1} l(y_i, \underline{x}_i, \hat{\underline{\theta}}) \right\}^{1/2} \end{aligned}$$

onde V é a matriz de covariância assintótica de $\hat{\theta}$, dada por $C^{-1}BC^{-1}$, e verificando que $\gamma_2 \leq a$.

Desde que $\partial/\partial\lambda \log f(y_i, \underline{x}_i, \underline{\theta})$ é uma função de $\log(y_i)$; se y_i está próximo de zero ou é grande, a i -ésima observação pode ter um valor absoluto grande para $\partial/\partial\lambda \log f(y_i, \underline{x}_i, \hat{\theta})$ e ser influente na determinação de $\hat{\lambda}$, mesmo se $|\varepsilon_i|$ for pequeno e \underline{x}_i for somente um ponto de baixa ruptura. Observamos, portanto, que entre os estimadores apresentados para λ , somente o estimador de influência limitada (EIL) consegue detectar e limitar esse tipo de influência na determinação do estimador de λ .

Não é possível construirmos um teste do tipo razão de verossimilhança para EIL; pois ele não é baseado na otimização de um critério.

Ambas as técnicas de estimação para a transformação de Box e Cox propostas em (2.4.2) e (2.4.3) são, ao contrário do estimador de máxima verossimilhança proposto na seção (2.2), relativamente insensíveis a valores aberrantes em ambos pontos de delineamento e variável resposta, como mostrado por Carroll e Ruppert(1985) em seus exemplos.

2.5 TRANSFORMAÇÕES SOB MINIMIZAÇÃO DE ERROS ABSOLUTOS

Vimos nas seções anteriores que transformações da variável resposta podem ser utilizadas como uma ferramenta poderosa na obtenção de modelos lineares cujos erros sejam independentes com variâncias constantes e distribuições idênticas aproximadamente

normais. Entretanto, quando observamos desvios maiores do que seria esperado se os erros fossem distribuídos normalmente, pode ser mais apropriado assumirmos uma distribuição de erros com caudas longas, como as distribuições dupla-exponencial e Cauchy. Neste caso, porém, é amplamente sabido que a regressão de mínimos quadrados está longe de ser ótima; além do fato de distribuições de caudas longas estimularem a ocorrência de valores aberrantes, sendo que a regressão de mínimos quadrados é muito sensível a valores aberrantes.

Nos últimos anos com o reconhecimento de que muitos conjuntos de dados de diversas áreas, como economia e finanças, possuem distribuições com caudas longas e que valores aberrantes podem na verdade carregarem informações relevantes quanto à estrutura subjacente ao problema, tem havido um crescente interesse em procedimentos de estimação robusta aplicáveis para modelos de regressão linear. Em situações onde os dados se acomodam melhor a tais tipos de distribuições de erros ou valores aberrantes sejam prováveis ocorrer, um procedimento robusto que se destaca pela sua resistência a valores aberrantes é a regressão L_1 , que minimiza a soma dos erros absolutos.

A regressão L_1 fornece estimadores de máxima verossimilhança para os parâmetros do modelo quando os erros são independentes com distribuições dupla-exponenciais idênticas, além de ser resistente a certos tipos de valores aberrantes. O fato da regressão de erros absolutos ser menos sensível a valores aberrantes do que a regressão de mínimos quadrados pode ser parcialmente explicado pelo seguinte: observações inconsistentes são tratadas de forma diferente pelos dois métodos. Se uma observação está afastada do restante das observações, o método de mínimos quadrados é mais influenciado por esse dado do que pelos demais, o que é indesejável. A regressão de erros absolutos, por outro lado, assegura um esforço igual com relação aos dados aberrantes e o restante do conjunto de

dados. Isto é, os estimadores de erros absolutos são influenciados por todas as observações, mas determinados somente por um subconjunto de observações. No caso do modelo linear de locação ($\underline{Y} = \underline{\mu} + \underline{\varepsilon}$), por exemplo, o estimador L_1 é dado pela mediana que conhecidamente sofre a influência de todas as observações, mas é determinada apenas por um subconjunto delas. Portanto, podemos esperar que dados aberrantes irão destacar-se mais quando usamos a regressão de erros absolutos.

A regressão de erros absolutos, também, tem melhor desempenho do que a regressão de mínimos quadrados quando os erros tem distribuição simétrica para a qual a mediana amostral é um estimador de posição mais eficiente do que a média amostral. Este procedimento de estimação tem a vantagem sobre os outros métodos robustos de não necessitar de uma "constante de ajuste" .

Em situações comuns em que as técnicas de análise de regressão L_1 são utilizadas, assume-se que o conjunto de dados analisado pertença a uma população que possua independência das observações e uma distribuição de erros aproximadamente dupla-exponencial com homogeneidade de variância. Muitas vezes, entretanto, mesmo o conjunto de dados se assemelhando mais às suposições associadas à regressão L_1 que à regressão de mínimos quadrados, o modelo não é linear ou alguma pressuposição do modelo não é satisfeita. Neste caso, um procedimento apropriado para aproximarmos os dados, ainda mais, da suposição de erros aleatórios independentes com distribuição comum dupla-exponencial é a utilização de alguma transformação adequada.

Nessa seção, apresentaremos o procedimento proposto por Parker(1988) que sugere o uso da família de transformações de Box e Cox a fim de obtermos modelos lineares cujos erros são independentes e possuem distribuições dupla-exponenciais idênticas, sob a minimização da soma de erros absolutos.

2.5.1 SELEÇÃO DE UMA TRANSFORMAÇÃO

Considere uma variável resposta positiva, Y , de n observações e que, para algum λ desconhecido, o vetor de respostas transformadas simples padronizadas, dadas em (2.2.14), possa ser escrito como

$$\underline{z}^{(\lambda)} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (2.5.1.1)$$

onde \underline{X} e $\underline{\beta}$ estão definidos em (2.2.1) e $\underline{\varepsilon}$ é o vetor ($n \times 1$) de erros aleatórios independentes com funções densidade de probabilidade idênticas, dadas por $f(\varepsilon_i) = \exp\{-|\varepsilon_i|/\theta\}/2\theta$; $\theta > 0$, $i=1, \dots, n$, isto é, distribuição dupla-exponencial.

Para estimarmos λ tal que as respostas transformadas padronizadas satisfaçam, pelo menos aproximadamente, (2.5.1.1), utilizaremos o procedimento baseado em considerações de verossimilhança, análogo ao desenvolvido na seção (2.2).

A função de verossimilhança das respostas originais é dada por

$$\frac{1}{(2\theta)^n} \exp \left\{ - \sum_{i=1}^n \frac{|z_i^{(\lambda)} - \underline{x}_i \underline{\beta}|}{\theta} \right\} \quad (2.5.1.2)$$

O suporte desta distribuição, para λ fixo, também não

depende de seus parâmetros, portanto o log da função de verossimilhança é dado por

$$-n \log(2\theta) - \sum_{i=1}^n \frac{|z_i^{(\lambda)} - x_i \beta|}{\theta} \quad (2.5.1.3)$$

Para λ fixo, sob a minimização da soma de erros absolutos, existe um subconjunto de p observações que interpolam o hiperplano ótimo (veja Bloomfield e Steiger(1983)). Portanto, para λ fixo, os estimadores de máxima verossimilhança de β e θ são os estimadores L_1

$$\hat{\beta}^\circ(\lambda) = X_{(1)}^{-1} Z_{(1)}^{(\lambda)} \quad (2.5.1.4)$$

e

$$\hat{\theta}^\circ(\lambda) = \sum_{i=1}^n |z_i^{(\lambda)} - x_i \hat{\beta}^\circ(\lambda)| / n \quad (2.5.1.5)$$

onde o índice (1) representa as observações básicas, ou seja, as observações cujos resíduos são nulos. A expressão (2.5.1.4) mostra como a regressão L_1 é influenciada por todas as observações, mas determinada somente por um subconjunto das observações.

Substituindo (2.5.1.4) e (2.5.1.5) em (2.5.1.3), obtemos o log da verossimilhança como uma função de λ , onde omitindo o termo constante, será dado por

$$L_{\max}(\lambda) = -n \log \left\{ \sum_{i=1}^n | z_i^{(\lambda)} - x_i \beta^{\circ}(\lambda) | \right\} \quad (2.5.1.6)$$

Como λ é um escalar, um modo fácil de obtermos, aproximadamente, o estimador de máxima verossimilhança para λ , λ° , é através da construção de um gráfico de $L_{\max}(\lambda)$ (dado em (2.5.1.6)) versus λ , para valores selecionados de λ , e a partir deste observarmos que valor λ° deve assumir.

Podemos obter um intervalo de confiança de aproximadamente $(1-\alpha)100\%$ para λ a partir de todos valores de λ que satisfaçam a desigualdade

$$L_{\max}(\lambda^{\circ}) - L_{\max}(\lambda) \leq 0.5 \chi_{1,\alpha}^2, \quad (2.5.1.7)$$

onde $\chi_{1,\alpha}$ representa o α -ésimo percentil de uma distribuição qui-quadrado com um grau de liberdade. É importante notarmos que (2.5.1.7) supõe que σ e β assumam os valores de seus estimadores de máxima verossimilhança.

2.6 ANÁLISE ROBUSTA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS

O procedimento de regressão L_1 , embora, como se sabe, forneça os estimadores de máxima verossimilhança para os parâmetros do modelo linear quando os erros aleatórios são independentes com distribuições dupla-exponenciais idênticas, é robusto somente a valores aberrantes na variável resposta. Esse método não é robusto, portanto, a valores aberrantes nas variáveis explanatórias, tal que mesmo pequenos erros nos valores dos pontos de delineamento podem nos levar a estimadores bastante distintos. Um modo de amenizarmos tal problema seria considerarmos a comunhão dos métodos descritos em (2.4.2) e (2.4.3) com o método dado em (2.5), ou seja, utilizarmos os métodos robustos propostos por Carroll e Ruppert(1985) sob modelos lineares com erros aleatórios independentes e distribuição comum dupla-exponencial, obtendo assim alternativas robustas para a regressão, sob minimização da soma dos erros absolutos, onde teremos, portanto, robustez a valores aberrantes em \underline{Y} e valores aberrantes em \underline{X} .

2.6.2 ESTIMADORES DE PSEUDO-VEROSSIMILHANÇA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS

Um modo bastante simples de efetuarmos a comunhão entre os estimadores de pseudo-verossimilhança dados em (2.4.2) e a transformação para a variável resposta de modelos lineares com distribuição de erros dupla-exponencial tratados em (2.5) seria considerarmos em (2.4.2),

$$\rho(x) = |x| \tag{2.6.1.1}$$

e

$$\rho'(x) = \psi(x) = \text{sign}(x)$$

e procedermos então os cálculos dos estimadores de pseudo-verossimilhança utilizando $a_x = 1.4$ e $s_1 = t_1 = w_{x,i}$, onde $w_{x,i}$ é determinado por (2.4.2.4) e (2.4.2.5).

A partir dessas substituições, o log da função de pseudo-verossimilhança (2.4.2.2) é dado por

$$-\sum_{i=1}^n w_{x,i}^2 \left\{ \log \sigma^* + \left| \frac{y_1^{(\lambda)} - x_1 \beta}{\sigma w_{x,i}} \right| \right\} \tag{2.6.1.2}$$

E para λ fixo, os estimadores de (β e σ) satisfazem:

$$\sum_{i=1}^n x_1 w_{x,i} \text{sign}(e_1^{(\lambda)}) = 0 \tag{2.6.1.3}$$

e

$$\sum_{i=1}^n w_{x,i}^2 \frac{\text{sign}(e_1^{(\lambda)})}{e_1^{(\lambda)}} - w_{x,i}^2 = 0$$

onde

$$e_1^{(\lambda)} = \left[\frac{y_1^{(\lambda)} - x_1 \beta}{\sigma w_{x,i}} \right]$$

Esta comunhão dos procedimentos desenvolvidos nas seções (2.4.2) e (2.5) deve nos fornecer um método robusto a valores aberrantes em \underline{Y} e \underline{X} , através de estimadores que atribuem pouco peso a respostas aberrantes que ocorram em pontos de delineamento discrepantes.

2.6.2 ESTIMADORES DE INFLUÊNCIA LIMITADA SOB MINIMIZAÇÃO DOS ERROS ABSOLUTOS

A união dos métodos de influência limitada e das transformações sob minimização dos erros absolutos dados nas seções (2.4.3) e (2.5), respectivamente, deve nos fornecer um procedimento robusto a valores aberrantes em \underline{Y} e \underline{X} , através de estimadores que limitam a função de influência. A fim de efetuarmos tal comunhão consideramos $f(y, \underline{x}_1, \underline{\theta})$ igual a função densidade de uma distribuição dupla-exponencial em (2.4.3) e procedemos então os cálculos dos estimadores dados em (2.4.3.2) e (2.4.3.3).

$$\text{Assumindo } f(y_1, \underline{x}_1, \underline{\theta}) = \frac{1}{2\sigma} \exp\left\{-\frac{|y_1^{(\lambda)} - \underline{x}_1 \underline{\beta}|}{\sigma}\right\},$$

teremos então:

$$\underline{\theta} = (\underline{\beta}, \sigma, \lambda) \tag{2.6.2.2}$$

$$l(y_1, \underline{x}_1, \underline{\theta}) = \nabla_{\underline{\theta}} \left\{ -\log 2\sigma - \frac{|y_1^{(\lambda)} - \underline{x}_1 \underline{\beta}|}{\sigma} \right\}$$

e o estimador de influência limitada para $\underline{\theta}$ será dado pelo valor de $\underline{\theta}$ que resolva

$$\sum_{i=1}^n w(y_i, \underline{x}_i, \underline{\theta}) \nabla_{\underline{\theta}} \left\{ -\log 2\sigma - \frac{|y_i^{(\lambda)} - \underline{x}_i \underline{\beta}|}{\sigma} \right\} = 0 \quad (2.6.2.3)$$

onde $w(y_i, \underline{x}_i, \underline{\theta})$ é dado por (2.4.3.3) com $l(y_i, \underline{x}_i, \underline{\theta})$ dado por (2.6.2.2).

CAPÍTULO 3

DIAGNÓSTICOS DE INFLUÊNCIA PARA TRANSFORMAÇÕES DO TIPO BOX E COX

3.1 INTRODUÇÃO

A seleção de uma transformação pode ser visualizada como um método auxiliar na escolha de um modelo que represente adequadamente um conjunto de dados. Com a extensa disponibilidade de softwares computacionais sofisticados tornou-se bastante simples ajustarmos um modelo de regressão qualquer ao conjunto de dados. Todavia, não devemos permitir que essa facilidade nos torne desatentos com respeito à verificação da existência de alguma observação que, por alguma razão, seja aberrante em relação ao restante das observações do conjunto de dados sendo analisado. A escolha de uma transformação em particular pode estar sendo demasiadamente dirigida pela presença de casos influentes ou aberrantes e ter efeitos permanentes que são difíceis de descobrir na análise subsequente, uma vez que a análise de dados transformados é geralmente condicionada ao valor estimado, $\hat{\lambda}$, de λ . Um valor aberrante ou caso influente na escala original, por exemplo, pode adaptar-se à escala transformada. Isto pode ser devido ao fato da escala transformada ser mais apropriada que a escala original, ou porque a seleção da transformação foi dirigida pelo caso

em questão. Portanto, sem dúvida, é importante investigarmos se a evidência para a seleção de uma transformação está uniformemente espalhada através do conjunto de dados ou se depende basicamente de alguns casos em particular. A fim de avaliarmos essa adequabilidade do modelo escolhido em relação às tendências gerais do conjunto de dados utilizaremos métodos de diagnósticos de influência.

Na primeira seção desse capítulo descreveremos métodos gráficos (Atkinson(1981) e Atkinson(1982)) e não gráficos (Cook e Wang(1983) e Atkinson(1986)) para avaliarmos o grau de influência de um subconjunto de observações na estimativa do parâmetro de transformação da variável resposta num modelo de regressão linear sob mínimos quadrados. Na seção seguinte, apresentaremos métodos de diagnósticos de influência de observações individuais, disponíveis para a estimativa do parâmetro de transformação da variável resposta, λ , num modelo de regressão linear sob a minimização da soma dos erros absolutos.

3.2 DIAGNÓSTICOS DE INFLUÊNCIA PARA REGRESSÃO DE MÍNIMOS QUADRADOS

Sob o método dos mínimos quadrados, ao examinarmos um modelo de regressão linear ajustado para um conjunto de dados, uma observação pode ser considerada influente se características importantes da análise, neste caso $\hat{\lambda}$, são alteradas substancialmente quando a observação é eliminada. Observações que possuem um resíduo grande após o modelo escolhido ser ajustado, são denominadas observações aberrantes. O fato de uma observação ser aberrante, não necessariamente significa que a observação seja influente com respeito ao modelo ajustado.

Na seção (3.2.1) apresentaremos um método gráfico baseado em considerações de variáveis adicionadas e nas seções (3.2.2), (3.2.3) e (3.2.4) diagnósticos de influência de observações sobre a estimativa do parâmetro de transformação, λ , num modelo de regressão linear de mínimos quadrados, baseados em métodos de eliminação de observações do conjunto de dados.

3.2.1 GRÁFICO DE VARIÁVEL ADICIONADA

O gráfico de variável adicionada tem como finalidade avaliarmos a necessidade de uma nova variável explanatória W , em termos de poder de predição da variável resposta, na presença de todas as outras variáveis explanatórias. Esse gráfico é aqui utilizado para analisarmos a influência de observações individuais sobre a transformação. Consideramos o seguinte modelo

$$\underline{Y} = \underline{X}\underline{\beta} + \delta \underline{W} + \underline{\varepsilon} \quad (3.2.1.1)$$

A equação (3.2.1.1) é construída de modo a ressaltarmos, no modelo, a presença da variável W , que pode ser uma variável explanatória do conjunto de dados ainda não incluída na matriz de delineamento X ou uma função de uma variável explanatória já presente em X . Um teste conveniente para avaliarmos a necessidade da variável W no modelo (3.2.1.1) é o teste F , para a hipótese $\delta = 0$. Entretanto, há sempre a possibilidade de que as conclusões tiradas a partir desse teste em relação à inclusão ou exclusão da nova variável explanatória W sejam indevidamente influenciadas por uma ou apenas

algumas observações. Um gráfico complementar à análise do teste F que visa fornecer tal tipo de informação pode ser obtido a partir do procedimento dado a seguir.

Seja $H = X(X'X)^{-1}X'$. Após multiplicarmos ambos os lados de (3.2.1.1) por $(I - H)$, temos

$$(I - H)\underline{Y} = (I - H)\underline{X}\underline{\beta} + \delta(I - H)\underline{W} + (I - H)\underline{\varepsilon} \quad (3.2.1.2)$$

O lado esquerdo de (3.2.1.2) é o mesmo vetor residual, \underline{e} , obtido ao ajustarmos o modelo (2.2.1). O primeiro termo no lado direito de (3.2.1.2) é exatamente igual a zero. Tomando esperanças em (3.2.1.2) obtemos,

$$E(\underline{e}) = \delta (I - H)\underline{W} \quad (3.2.1.3)$$

que sugere que um gráfico de \underline{e} versus $(I - H)\underline{W}$ será linear, passando pela origem, se o modelo (3.2.1.1) é adequado. Chamaremos o gráfico de \underline{e} versus $(I - H)\underline{W}$ um gráfico de variável adicionada, já que ele é destinado a medir o efeito de adicionar uma variável ao modelo.

Por analogia, Atkinson(1982) propôs, então, que o teste $T_A(\lambda_0)$, dado em (2.3.2.3), que usamos para avaliar a necessidade de uma transformação na variável resposta seja proveitosamente completado por um gráfico de variável adicionada. A variável resposta será

representada por $\underline{Z}^{(\lambda_0)}$ e a nova variável explanatória, W , será representada por $\underline{D}^{(\lambda_0)}$, dado em (2.3.2.1.a). Deste modo, o teste $T_A(\lambda_0)$ pode ser ampliado fornecendo, também, uma análise gráfica do efeito de observações individuais sobre a transformação estimada.

Idealmente, este gráfico, ou seja, o gráfico de \underline{e} versus $(I - H)\underline{D}^{(\lambda_0)}$, deve mostrar uma tendência linear constante, indicando que a evidência para uma transformação está uniformemente espalhada através do conjunto de dados. Valores aberrantes em um gráfico de variável adicionada podem corresponder a casos que estão deturpando a evidência para uma transformação e então necessitam de atenção especial. Uma curvatura substancial pode ser uma indicação de que uma modificação da família de transformações permitiria uma melhor representação do modelo verdadeiro.

Na regressão de \underline{e} versus $(I - H)\underline{D}^{(\lambda_0)}$, o intercepto é zero e a inclinação é estimada por:

$$\hat{\phi} = \frac{\underline{D}^{(\lambda_0)'} (I - H)\underline{Z}^{(\lambda_0)}}{\underline{D}^{(\lambda_0)'} (I - H)\underline{D}^{(\lambda_0)}} \quad (3.2.1.4)$$

Devemos ressaltar que $\hat{\phi}$ é idêntico a estimativa de mínimos quadrados de ϕ obtida ao ajustarmos o modelo (2.3.2.2). A partir disto, temos imediatamente que os resíduos no gráfico de variável adicionada são os mesmos resíduos obtidos da regressão de $\underline{Z}^{(\lambda_0)}$ sobre \mathcal{X} e $\underline{D}^{(\lambda_0)}$.

3.2.2 MÉTODO GRÁFICO DE ATKINSON

Atkinson(1981) descreveu dois métodos gráficos para avaliar a presença de valores aberrantes ou influentes tanto na variável resposta quanto nas variáveis explanatórias.

A fim de detectar valores aberrantes na variável resposta, Atkinson(1981) sugeriu a construção de um gráfico semi-normal dos resíduos "studentizados", dados por

$$t_i = \frac{y_i - \hat{y}_i}{s_{(i)} (1 - h_{ii})^{1/2}}, \quad i=1,2,\dots,n \quad (3.2.2.1)$$

onde $s_{(i)}^2$ é a estimativa de σ^2 dada pelo quadrado médio dos resíduos após a i -ésima observação ter sido eliminada e h_{ii} é o i -ésimo elemento da diagonal da matriz de projeção H .

Para detectarmos observações influentes nas variáveis explanatórias, Cook(1977) propôs a estatística:

$$D_i = \left[\frac{y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}}{s (1 - h_{ii})^{1/2}} \right]^2 \cdot \frac{h_{ii}}{p (1 - h_{ii})}, \quad i = 1,2,\dots,n \quad (3.2.2.2)$$

Atkinson(1981) propôs, então, modificarmos a estatística dada por Cook a fim de obtermos um gráfico de diagnóstico para os pontos de delineamento influentes. Após substituir s^2 por $s_{(i)}^2$

em (3.2.2.2) a fim de enfatizar valores extremos, padronizou $(D_1)^{1/2}$, obtendo assim a seguinte estatística modificada:

$$C_i = \left\{ \frac{n - p}{p} \left(\frac{h_{ii}}{1 - h_{ii}} \right) \right\}^{1/2} \cdot |t_i|, \quad i=1,2,\dots,n \quad (3.2.2.3)$$

onde t_i é dado por (3.2.2.1).

Em um delineamento ótimo, todas as observações têm a mesma influência e $h_{ii} = p/n$. Devido à padronização de $(D_1)^{1/2}$, teremos, portanto, que os gráficos de C_i e $|t_i|$ são idênticos, quando o delineamento for ótimo.

O gráfico obtido a partir de (3.2.2.3) nos indicará o grau de influência do i -ésimo ponto do conjunto de dados sobre as estimativas paramétricas do modelo assumido. Grandes valores de C_i indicam uma observação que tem uma influência apreciável sobre os parâmetros estimados. Se esta observação não for usada no ajuste, há uma grande mudança na estimativas paramétricas.

Atkinson(1981) propôs um estudo em conjunto dos gráficos de t_i e C_i a fim de detectarmos valores aberrantes e influentes em ambos conjuntos de variáveis, \underline{Y} e \underline{X} . Observamos, entretanto, que os $2n$ valores de t_i e C_i proporcionam um campo potencialmente abundante para interpretações excessivas dos dados. Para evitar isso, ele propôs que as estatísticas sejam apresentadas como gráficos semi-normais que parecem mostrar com mais eficiência que os gráficos normais, grandes valores aberrantes.

Na ausência de valores aberrantes e observações

indevidamente influentes, os gráficos de t_1 e C_1 devem ser aproximadamente uma linha reta com inclinação de cerca de 45° . Se um valor aberrante ou influente está somente nos pontos de delineamento, este valor é detectado somente pelo gráfico de C_1 .

Atkinson(1982) a fim de melhor avaliar a contribuição de observações individuais para a escolha de uma transformação em particular, propõe a análise conjunta do gráfico de variável adicionada dado em (3.2.1) e do gráfico semi-normal da estatística de Cook modificada dada em (3.2.2.3) para o modelo univariado dos resíduos $(I - H)Y_0^{(\lambda)}$ sobre os resíduos da variável adicionada $(I - H)D_0^{(\lambda)}$. Ele propôs a união desses dois gráficos devido ao fato de que em alguns casos o gráfico proposto na seção anterior não mostra com clareza certas observações altamente influentes para a transformação, enquanto que o gráfico semi-normal para a estatística de Cook modificada aponta estas observações.

3.2.3 MÉTODO DE COOK E WANG

Cook e Wang(1983) expuseram que o método de variável adicionada proposto por Atkinson é sem dúvida útil e facilmente executado para as famílias de transformações simples uni-paramétricas e que o método frequentemente identificará casos influentes, mas que esta habilidade de identificação não pode ser garantida pelo fato do método não tomar os cuidados adequados com os pontos de alta ruptura do problema, ou seja, com os pontos do delineamento que estão bastante afastados do centro de massa(média) dos outros valores \tilde{x} . Para essas observações os resíduos de ambos $Z_0^{(\lambda)}$ e $D_0^{(\lambda)}$ tenderão a ser pequenos, com variâncias proporcionais a $(1-h_{ii})$, onde h_{ii} é o i -ésimo

elemento da diagonal da matriz H, fazendo com que estas observações estejam próximas da origem no gráfico de variável adicionada. Conseqüentemente, qualquer informação contida nessas observações sobre a transformação é aparentemente sem importância.

Eles propuseram, então, uma aproximação para a distância de verossimilhança

$$LD_I = 2 \left[L_{\max}(\hat{\lambda}) - L_{\max}(\hat{\lambda}_I) \right] \quad (3.2.3.1)$$

onde $L_{\max}(\hat{\lambda})$ é definido em (2.2.13) e $\hat{\lambda}_I$ é o estimador de máxima verossimilhança de λ obtido após eliminarmos os casos indexados por $I = (i_1, \dots, i_k)$. A estatística dada em (3.2.3.1) pode ser comparada aos percentis de uma distribuição χ^2 com um grau de liberdade.

O estimador $\hat{\lambda}_I$ é obtido a partir do modelo de eliminação

$$\underline{z}_I^{(\lambda)} = \underline{x}_I \underline{\beta} + \underline{\varepsilon}_I \quad (3.2.3.2)$$

onde o subscrito I significa "com os casos indexados por I eliminados". Observamos, entretanto, que para cada valor de I o jacobiano J dado em (2.2.4.a) é diferente, o que torna a quantidade de cálculos muito grande e a estatística intratável.

Com a finalidade de viabilizar a utilização de

(3.2.3.1), Cook e Wang(1983) propuseram a substituição do modelo (3.2.3.2) pelo modelo de valores aberrantes dado por

$$\underline{Z}^{(\lambda)} = \underline{X}\underline{\beta} + \underline{E}_I \underline{\omega} + \underline{\varepsilon} \quad (3.2.3.3)$$

onde $\underline{\omega}$ é um vetor ($k \times 1$) e \underline{E}_I é uma matriz ($n \times k$), onde a j -ésima coluna de \underline{E}_I contém um 1 na i_j -ésima posição conforme indexado por $I = (i_1, \dots, i_k)$ e zeros nas outras posições. Para $\underline{\lambda}$ fixo, as estimativas de máxima verossimilhança de $\underline{\beta}$ são as mesmas sob os modelos de eliminação e de valores aberrantes.

Sob o modelo de valores aberrantes, o estimador de máxima verossimilhança de λ , $\hat{\lambda}_I$, minimiza

$$\underline{Z}^{(\lambda)'} (I - H_E) \underline{Z}^{(\lambda)} \quad (3.2.3.4)$$

onde $H_E = \underline{X}_E (\underline{X}_E' \underline{X}_E)^{-1} \underline{X}_E'$ e $\underline{X}_E = (\underline{X} : \underline{E}_I)$

Substituindo $\underline{Z}^{(\lambda)}$ em (3.2.3.4) pela sua expansão linear em série de Taylor sobre $\hat{\lambda}$,

$$\underline{Z}^{(\lambda)} \cong \underline{Z}^{(\hat{\lambda})} + (\lambda - \hat{\lambda}) \underline{D}^{(\hat{\lambda})} \quad (3.2.3.5)$$

e minimizando a expressão resultante, temos o correspondente estimador

$\hat{\lambda}_I^1$ dado por

$$\hat{\lambda}_I^1 = \hat{\lambda} + \frac{\hat{e}'_{\sim Z, I} (I - H_I)^{-1} \hat{e}_{\sim D, I}}{S_D^2(n-p) - \hat{e}'_{\sim D, I} (I - H_I)^{-1} \hat{e}_{\sim D, I}} \quad (3.2.3.6)$$

onde H_I é a submatriz ($k \times k$) de H correspondendo às linhas e colunas indexadas por I ; $\hat{e}_{\sim Z, I}$ e $\hat{e}_{\sim D, I}$ são, respectivamente, os subvetores ($k \times 1$) de

$$\hat{e}_{\sim Z} = (I - H)Z^{(\hat{\lambda})} \quad \text{e} \quad \hat{e}_{\sim D} = (I - H)D^{(\hat{\lambda})}$$

e $S_D^2 = \hat{e}'_{\sim D} \hat{e}_{\sim D} / (n-p)$.

Usando $\hat{\lambda}_I^1$ como uma aproximação para $\hat{\lambda}_I$ na distância de verossimilhança temos que

$$LD_I \cong 2 \left[L_{\max}(\hat{\lambda}) - L_{\max}(\hat{\lambda}_I^1) \right] \quad (3.2.3.7)$$

Se usarmos a aproximação dada em (3.2.3.5) para novamente simplificar (3.2.3.7) teremos

$$LD_I \cong \hat{LD}_I = n \log \left\{ 1 + \left[\frac{\begin{matrix} \hat{r}'_{\sim Z, I} & \hat{r}_{\sim D, I} \\ \hat{r}_{\sim D, I} & \hat{r}_{\sim D, I} \end{matrix}}{n - p - \begin{matrix} \hat{r}'_{\sim D, I} & \hat{r}_{\sim D, I} \\ \hat{r}_{\sim D, I} & \hat{r}_{\sim D, I} \end{matrix}} \right]^2 \right\} \quad (3.2.3.8)$$

onde

$$\begin{aligned} \dot{r}_{z,I} &= \frac{\hat{e}_{z,I} (I - H_1)^{-1/2}}{S_z} , \\ S_z^2 &= \frac{\hat{e}'_{z,I} \hat{e}_{z,I}}{n - p} \end{aligned} \quad (3.2.3.9)$$

e

$$\dot{r}_{D,I} = \frac{\hat{e}_{D,I} (I - H_1)^{-1/2}}{S_D} .$$

Cook e Wang recomendaram esta forma final para uso na prática. Eles expuseram que uma diferença importante entre este procedimento e o procedimento de Atkinson é que este último calcula os resíduos $\hat{e}_{z,I} = (I - H)Z^{(\lambda)}$ e $\hat{e}_{D,I} = (I - H)D^{(\lambda)}$ sob o valor da hipótese nula $\lambda = \lambda_0$ em lugar da estimativa de máxima verossimilhança $\hat{\lambda}$. Além disso, a estrutura de ruptura do problema é levada em consideração em \hat{LD}_I através do uso de resíduos padronizados, onde pontos influentes tenderão a ser aqueles onde ambos resíduos padronizados $\dot{r}_{D,I}$ e $\dot{r}_{z,I}$ são grandes simultaneamente.

Cook e Wang(1983) expuseram exemplos demonstrando como o método de variável adicionada proposto por Atkinson falha na identificação de valores aberrantes e pontos influentes, enquanto que seu método identifica tais pontos com facilidade. A fim de melhor visualizar as informações obtidas a partir da estatística (3.2.3.8), Cook e Wang expuseram em seus exemplos essa estatística sob a forma gráfica, ou seja, utilizaram um gráfico de \hat{LD}_I versus i , considerando a eliminação de uma única observação por vez.

3.2.4 MÉTODO DE ATKINSON

A partir do estimador $\tilde{\lambda}$ dado em (2.3.2.5) e do teste para a hipótese $\lambda = \lambda_0$ dado em (2.3.2.3) por

$$T_A(\lambda_0) = - \frac{\underline{D}^{(\lambda_0)'} \underline{X}_r \underline{Z}^{(\lambda_0)}}{\hat{\sigma}_z \left(\underline{D}^{(\lambda_0)'} \underline{X}_r \underline{D}^{(\lambda_0)} \right)^{1/2}}$$

Atkinson(1986) propôs um método não gráfico para julgar a influência de um subconjunto de observações sobre a estimativa do parâmetro de transformação da variável resposta num modelo de regressão linear sob mínimos quadrados. Ele propôs, então, eliminarmos tais observações do conjunto de dados e analisarmos os efeitos sobre $\tilde{\lambda}$ e $T_A(\lambda_0)$.

A estatística $T_A(\lambda_0)$ e o estimador $\tilde{\lambda}$ são funções da soma de produtos de resíduos da forma:

$$S(\underline{a}, \underline{b}) = \underline{a}'(I - H)\underline{b} \quad (3.2.4.1)$$

onde \underline{a} e \underline{b} são vetores associados às observações. O efeito de eliminarmos k observações indexadas por $I = (i_1, \dots, i_k)$ sobre (3.2.4.1) é dado por (Belsley, Kuh e Welsch(1980) - p.64)

$$S_I(\underline{a}, \underline{b}) = S(\underline{a}, \underline{b}) - \underline{a}_I' (I - H_I)^{-1} \underline{b}_I \quad (3.2.4.2)$$

onde H_I é a submatriz ($k \times k$) de H correspondente às linhas e colunas indexadas por I e \underline{a}_I° e \underline{b}_I° são os k membros relevantes dos vetores de resíduos $(I - H)\underline{a}$ e $(I - H)\underline{b}$, respectivamente. Essas somas de produtos com observações eliminadas podem então serem usadas para calcularmos estatísticas que meçam o efeito de eliminarmos determinados subconjuntos de observações.

Para o modelo de regressão (2.3.2.2), a estimativa do coeficiente de regressão para $D^{(\lambda_0)}$ após os k casos indexados por I terem sido eliminados é, portanto, baseando-nos em (3.2.4.2), dado por

$$\hat{\phi}_I = \frac{D^{(\lambda_0)'} (I - H) \underline{Z}_I^{(\lambda_0)} - D_I^\circ (I - H_I)^{-1} \underline{Z}_I^\circ}{D^{(\lambda_0)'} (I - H) D_I^{(\lambda_0)} - D_I^\circ (I - H_I)^{-1} D_I^\circ} \quad (3.2.4.3)$$

onde D_I° e \underline{Z}_I° são os k membros, indexados por I , pertencentes aos vetores residuais $(I - H)D^{(\lambda_0)}$ e $(I - H)\underline{Z}^{(\lambda_0)}$, respectivamente.

Logo, o estimador $\tilde{\lambda}$, após eliminarmos as k observações, é dado por

$$\lambda_I = \lambda_0 - \frac{D^{(\lambda_0)'} (I - H) \underline{Z}_I^{(\lambda_0)} - D_I^\circ (I - H_I)^{-1} \underline{Z}_I^\circ}{D^{(\lambda_0)'} (I - H) D_I^{(\lambda_0)} - D_I^\circ (I - H_I)^{-1} D_I^\circ} \quad (3.2.4.4)$$

O efeito de eliminarmos k observações sobre a estatística $T_A(\lambda_0)$, será dado por

$$\tilde{T}_{A(I)}(\lambda_0) = - \frac{\underline{D}^{(\lambda_0)'} (I - H) \underline{Z}^{(\lambda_0)} - \underline{D}_I^{\circ'} (I - H_I)^{-1} \underline{Z}_I^{\circ}}{\hat{\sigma}_{z(I)}^2 (\underline{D}^{(\lambda_0)'} (I - H) \underline{D}^{(\lambda_0)} - \underline{D}_I^{\circ'} (I - H_I)^{-1} \underline{D}_I^{\circ})^{1/2}} \quad (3.2.4.5)$$

onde $\hat{\sigma}_{z(I)}^2$ é o quadrado médio dos resíduos obtido para o modelo (2.3.2.2) após as k observações terem sido eliminadas.

Atkinson(1985) mostrou através de exemplos que este método desenvolvido não apresenta nenhuma desvantagem em relação ao método desenvolvido por Cook e Wang(1983), pois ambas as estatísticas desenvolvidas $\tilde{\lambda}_I$ e $\hat{\lambda}_I^1$ apresentam valores bastante próximos nos problemas exemplificados. Uma desvantagem do método de Cook e Wang, entretanto, é o fato do método deles necessitar do valor da estimativa de máxima verossimilhança do parâmetro de transformação, $\hat{\lambda}$. Ambas as estimativas $\tilde{\lambda}_I$ e $\tilde{T}_{A(I)}$ mostraram uma boa aproximação com os respectivos valores exatos $\hat{\lambda}_I$ e $T_{A(I)}$.

3.3 DIAGNÓSTICOS DE INFLUÊNCIA PARA REGRESSÃO L_1

Com o intuito de verificarmos se a evidência para uma escolha de transformação em particular está uniformemente espalhada através do conjunto de dados ou se está sendo indevidamente influenciada por uma ou mais observações, desenvolvemos na seção anterior alguns métodos de diagnósticos sob regressão de mínimos quadrados. Para regressão L_1 , sabemos, entretanto, que seus parâmetros são completamente determinados por um subconjunto das observações;

consequentemente diagnósticos de influência similares àqueles empregados em regressão de mínimos quadrados não são prontamente válidos. As observações básicas tem resíduos iguais a zero e são influentes, mas não no contexto de mínimos quadrados. O conceito de um subconjunto de observações básicas é único para regressão L_1 e nenhuma comparação pode ser feita com regressão de mínimos quadrados. As observações básicas são um conjunto à parte com suas próprias marcas especiais de influência. Apresentaremos, então, medidas de influência das observações remanescentes sobre a escolha do valor do parâmetro de transformação, apresentadas por Parker(1988).

3.3.1 MÉTODOS GRÁFICOS DE PARKER

O sistema de equações correspondentes ao ajuste de regressão L_1 para o modelo (2.5.1.1) pode ser escrito como:

$$\begin{bmatrix} Z_{(1)}^{(\lambda)} \\ Z_{(2)}^{(\lambda)} \end{bmatrix} = \begin{bmatrix} X_{(1)} & 0 \\ X_{(2)} & I^\circ \end{bmatrix} \cdot \begin{bmatrix} \beta^\circ \\ e_{(2)} \end{bmatrix} \quad (3.3.1.1)$$

onde os índices (1) e (2) referem-se às observações com resíduos iguais a zero (observações básicas) e resíduos diferentes de zero, respectivamente. β° é o vetor $(p \times 1)$ definido em (2.5.1.4), o vetor $e_{(2)}$ é composto pelos valores absolutos dos resíduos referentes ao conjunto (2) e I° é uma matriz diagonal $(n-k) \times (n-k)$ com elementos da diagonal que são +1 ou -1 conforme o resíduo e_i correspondente àquela

observação seja negativo ou positivo. As observações indexadas por (1), caem sobre o hiperplano de regressão L_1 .

Deste modo, degenerescências à parte, somente as observações básicas possuem resíduo igual a zero e não contribuem para o log da verossimilhança (2.5.1.6), que pode, então, ser reexpressado como:

$$-n \log \left\{ \sum_{i \in (2)}^{n-p} \left| Z_i^{(\lambda)} - \underline{x}_i \underline{X}_{(1)}^{-1} Z_{(1)}^{(\lambda)} \right| \right\} \quad (3.3.1.2)$$

A inclinação do log da verossimilhança acima, dada por

$$S(\lambda) = \frac{\sum_{i \in (2)} \text{sign}(e_i^{(\lambda)}) (D_i^{(\lambda)} - \underline{x}_i \underline{X}_{(1)}^{-1} D_{(1)}^{(\lambda)})}{\sum_{i \in (2)} |e_i^{(\lambda)}|} \quad (3.3.1.3)$$

onde $\underline{e}^{(\lambda)}$ é o vetor residual para o modelo transformado e $D_{(1)}^{(\lambda)} = \partial Z_{(1)}^{(\lambda)} / \partial \lambda$, fornece um modo de avaliarmos quão próximo um dado valor λ_0 está do valor ótimo de λ , λ° . Para o valor ótimo do parâmetro de transformação, λ° ,

$$S(\lambda^\circ) = 0,$$

logo quanto mais próximo $S(\lambda_0)$ estiver de zero, mais próximo λ_0 está de λ° .

Para um determinado valor do parâmetro de transformação, λ_0 , denominamos a estatística $S(\lambda_0)$ por escore de eficiência (Cox e Hinkley (1978)). O escore de eficiência $S(1)$ é uma medida da necessidade de uma transformação. Quando $S(1)$ é menor (maior) que zero temos que o valor ótimo do parâmetro de transformação, λ^0 , é menor (maior) que um; para $S(1)$ igual a zero observamos que não há necessidade de uma transformação da variável resposta.

A partir de (3.3.1.3) podemos, então, obter uma medida da contribuição da i -ésima observação para a inclinação do log da verossimilhança em λ_0 , dada pela quantidade

$$\text{sign}(e_i^{(\lambda_0)}) \left(D_i^{(\lambda_0)} - \mathbf{x}_i \mathbf{X}^{-1} D_{(1)}^{(\lambda_0)} \right) \quad (3.3.1.4)$$

Esta medida proporciona, conseqüentemente, uma análise da influência de observações individuais sobre a transformação. Quanto mais a medida (3.3.1.4) se afasta do zero no sentido negativo (positivo), mais temos evidências de que esta observação pertence a um modelo cujo valor ótimo do parâmetro de transformação, λ^0 , é menor (maior) que λ_0 .

Para a hipótese de nenhuma transformação, que é $\lambda_0 = 1$, podemos visualizar a influência de observações individuais na determinação da necessidade de uma transformação sob regressão L_1 a partir do gráfico de índices da seguinte medida de influência:

$$T_i = \text{sign}(e_i^{(1)}) \left(D_i^{(1)} - \mathbf{x}_i \mathbf{X}^{-1} D_{(1)}^{(1)} \right) \quad (3.3.1.5)$$

Tal gráfico é de bastante interesse, pois algumas vezes necessitamos escolher entre uma resposta não-transformada e uma resposta transformada, como por exemplo a transformação log, raiz quadrada ou inversa, que ocorrem tão frequentemente. Observações que possuem grande influência na escolha de transformações do tipo log, raiz quadrada ou inversa, por exemplo, apresentam grandes valores positivos de T_1 , ao passo que, grandes valores negativos de T_1 estão associados a observações que rejeitam a necessidade de transformações de tais tipos.

Um outro modo de avaliarmos o efeito de observações individuais sobre a transformação pode ser obtido através da eliminação de uma observação do conjunto de dados por vez e posterior verificação na mudança ocorrida na estimativa do parâmetro de transformação λ° devido à ausência desta observação. Denominando por λ_1° a estimativa de máxima verossimilhança do parâmetro de transformação, após a i -ésima observação ter sido eliminada, podemos construir, então, um gráfico de índices de λ_1° .

Parker(1988) aplicou tais gráficos de influência sobre conjuntos de dados conhecidos a fim de comparar as observações consideradas influentes sob análise de regressão L_1 e observações influentes detectadas por outros métodos sob estimação de mínimos quadrados e sob estimação robusta. Para um conjunto de dados frequentemente analisado e conhecido por "Stack-Loss Data", Parker observou que não há observações individuais ou um subconjunto de observações direcionando demasiadamente a escolha da transformação sob minimização dos erros absolutos; ao passo que pelos outros métodos mencionados(diagnósticos para regressão de mínimos quadrados (Atkinson(1982)) e regressão robusta(Carroll e Ruppert(1985)) existem subconjuntos de observações consideradas potencialmente influentes. Ele concluiu, portanto, que os erros aleatórios se acomodavam melhor, neste caso, a uma distribuição de caudas longas e conseqüentemente a

uma análise de mínimos desvios absolutos.

Um outro método de diagnóstico de influência, não considerado por Parker, mas que poderia ser perfeitamente utilizado para regressão L_1 , seria a construção de um gráfico de LD_1 dado por (3.2.3.1) versus i , considerando a eliminação de uma única observação por vez e os erros aleatórios independentes e identicamente distribuídos dupla-exponenciais. Essa utilização em regressão L_1 de um método desenvolvido para regressão de mínimos quadrados é bastante razoável uma vez que podemos obter tanto $L_{\max}(\lambda^\circ)$ quanto $L_{\max}(\lambda^\circ)_1$ através de (2.5.1.6). LD_1 , sob minimização dos erros absolutos, também possui uma distribuição χ^2 com v graus de liberdade.

CAPÍTULO 4

APLICAÇÕES

4.1 INTRODUÇÃO

A título de ilustração e comparação, utilizando alguns métodos catalogados nos capítulos anteriores, sob minimização da soma de erros quadráticos e minimização da soma de erros absolutos, analisamos alguns conjuntos de dados gerados com o auxílio do "Statistical Analysis System (SAS)" e um conjunto de dados real retirado do livro de Gunst e Mason [(1980) - p. 358].

Em todos os conjuntos de dados gerados, os valores da variável resposta, Y, foram determinados através do modelo

$$Y = \exp\left\{ 1 + X + \varepsilon \right\} ,$$

onde X é um valor fixo obtido a partir do gerador uniforme(0,1) e ε

representa o erro gerado a partir de distribuição normal, dupla-exponencial ou normal contaminada. Cada conjunto de dados gerado contém 30 observações.

Foram utilizadas as funções RANUNI, RANNOR e RANEXP do sistema de programas SAS para gerar amostras de distribuição uniforme, normal e exponencial, respectivamente. O método de geração de observações da distribuição dupla-exponencial foi:

- (i) Gerar uma observação da distribuição uniforme(0,1), usando a função RANUNI, u .
- (ii) Gerar uma observação da distribuição exponencial com $\theta = 3$, usando a função RANEXP, e .
- (iii) Se $u < 0.5$, então e é o valor gerado da distribuição dupla-exponencial, caso contrário o valor gerado é $-e$.

Todos os cálculos necessários para obtermos os resultados dos métodos aplicados na análise dos conjuntos de dados foram desenvolvidos através do procedimento IML do sistema de programas SAS e encontram-se no apêndice 1. Para calcularmos a regressão de mínimos desvios absolutos, assim como suas aplicações, sob minimização dos desvios absolutos, usamos um programa que utiliza a subrotina de programação linear contida no procedimento IML do sistema de programas SAS. Esse programa consta no manual do SAS/IML versão 6.03 (pg.180). Escolhemos trabalhar com este algoritmo para a resolução das regressões de mínimos desvios absolutos, pelo fato dele ser desenvolvido no procedimento IML do sistema de programas SAS, onde desenvolvemos o restante dos métodos analisados.

Após gerarmos os problemas (conjuntos de dados), aplicamos alguns métodos estudados para verificarmos se conseguem indicar a transformação correta (transformação pré-estabelecida com a

simulação) e apontar os possíveis valores influentes presentes no conjunto de dados.

4.2 PROBLEMA 1

No primeiro conjunto de dados gerado, mostrado na tabela 4.2.1, ϵ é uma amostra de uma distribuição normal com média zero e desvio padrão igual a três.

OBS	X	Y
1	0.57663	0.250
2	0.44536	0.274
3	0.65664	541.518
4	0.51567	0.109
5	0.30051	0.709
6	0.01247	0.097
7	0.34488	3.392
8	0.58768	0.030
9	0.67852	145.804
10	0.64308	1.091
11	0.53687	1.800
12	0.49618	4.249
13	0.71133	488.587
14	0.25794	13.336
15	0.16794	7.469
16	0.42766	0.392
17	0.42048	7.767

18	0.09227	0.596
19	0.69213	41.588
20	0.35331	66.789
21	0.92884	0.982
22	0.40392	0.017
23	0.24245	0.838
24	0.25156	0.923
25	0.06455	68.669
26	0.13990	6.247
27	0.67155	40.294
28	0.83292	2.495
29	0.90244	0.017
30	0.14356	10.478

Tabela 4.2.1 Lista de dados gerados sob distribuição de erros normal com média zero e desvio padrão igual a três.

Como um estudo preliminar, para avaliarmos a necessidade de uma transformação, ajustamos o modelo de primeira ordem completo, $\hat{Y} = \hat{a} + \hat{b}X$, sob mínimos quadrados e mínimos desvios absolutos. Os modelos obtidos são dados abaixo:

Regressão de mínimos quadrados:

$$\hat{Y} = -12.5 + 135.8 X$$

Soma dos Quadrados dos Resíduos = 461564.95

Regressão de mínimos desvios absolutos:

$$\hat{Y} = 4.8 - 4.1 X$$

Soma dos Resíduos Absolutos = 1439.59

Através dos gráficos probabilísticos para os resíduos dos modelos ajustados, dados nas figuras 4.2.1 e 4.2.2, temos fortes evidências de que uma transformação da variável resposta, tanto sob análise de mínimos quadrados quanto sob análise de mínimos desvios absolutos, podem valer a pena.

Aplicando a família de transformações de Box e Cox, dada em (2.2.2), na variável resposta, sob análise de mínimos quadrados, obtemos que o estimador de máxima verossimilhança de λ é $\hat{\lambda} = -0.05$. O gráfico do log de verossimilhança $L_{\max}(\lambda)$ versus λ , sob mínimos quadrados, mostrado na figura 4.2.3, fornece o intervalo de confiança de 95% para λ , definido por (2.2.10), que abrange os valores entre -0.10 e 0.10. Desse modo, temos que a transformação log pré-estabelecida é apropriada.

O gráfico do log de verossimilhança $L_{\max}(\lambda)$ versus λ , sob a análise de mínimos desvios absolutos é mostrado na figura 4.2.4. O log de verossimilhança assume o máximo em $\hat{\lambda} = -0.05$. O intervalo de confiança de 95% para λ , definido em (2.5.1.7), abrange os valores entre -0.15 e 0.15. Vemos, portanto, que sob análise de mínimos desvios absolutos, a transformação log também é identificada.

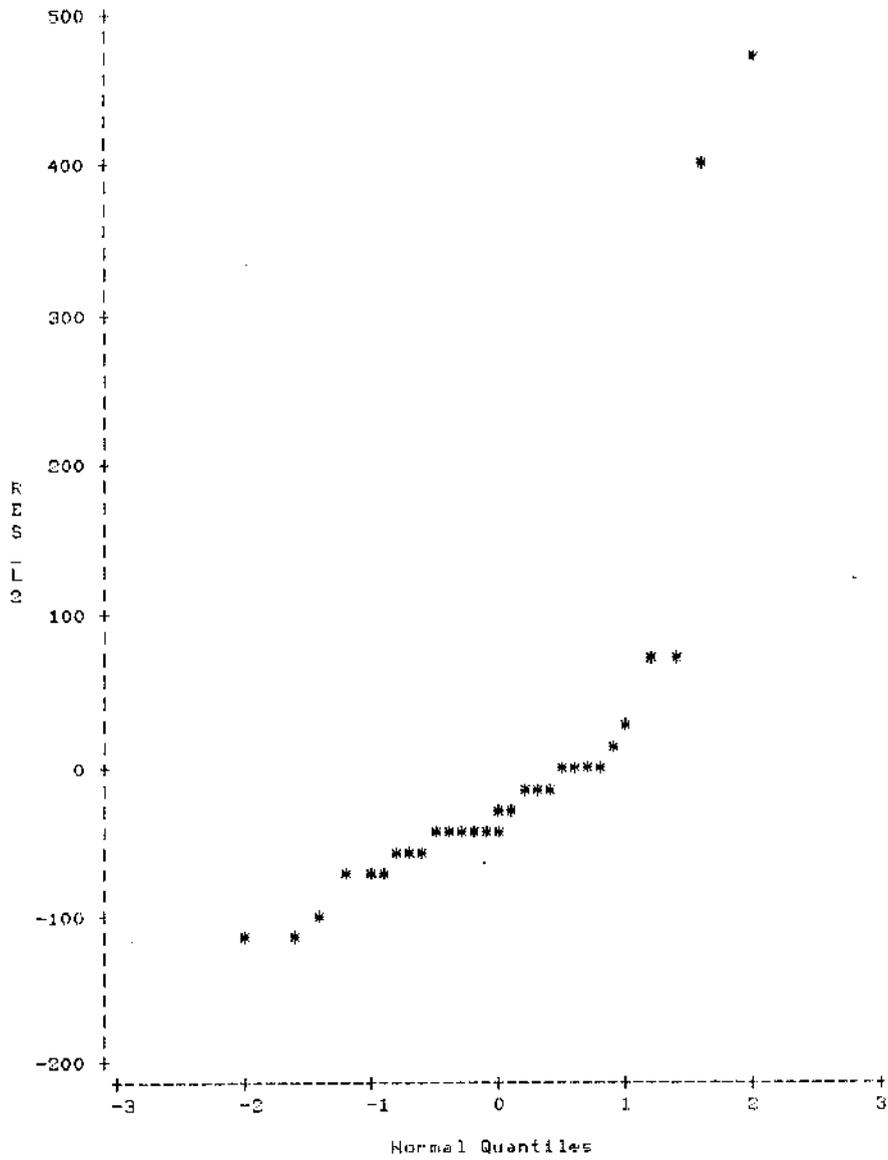


Figura 4.2.1 Gráfico probabilístico normal para os resíduos do modelo não transformado ajustado por mínimos quadrados.

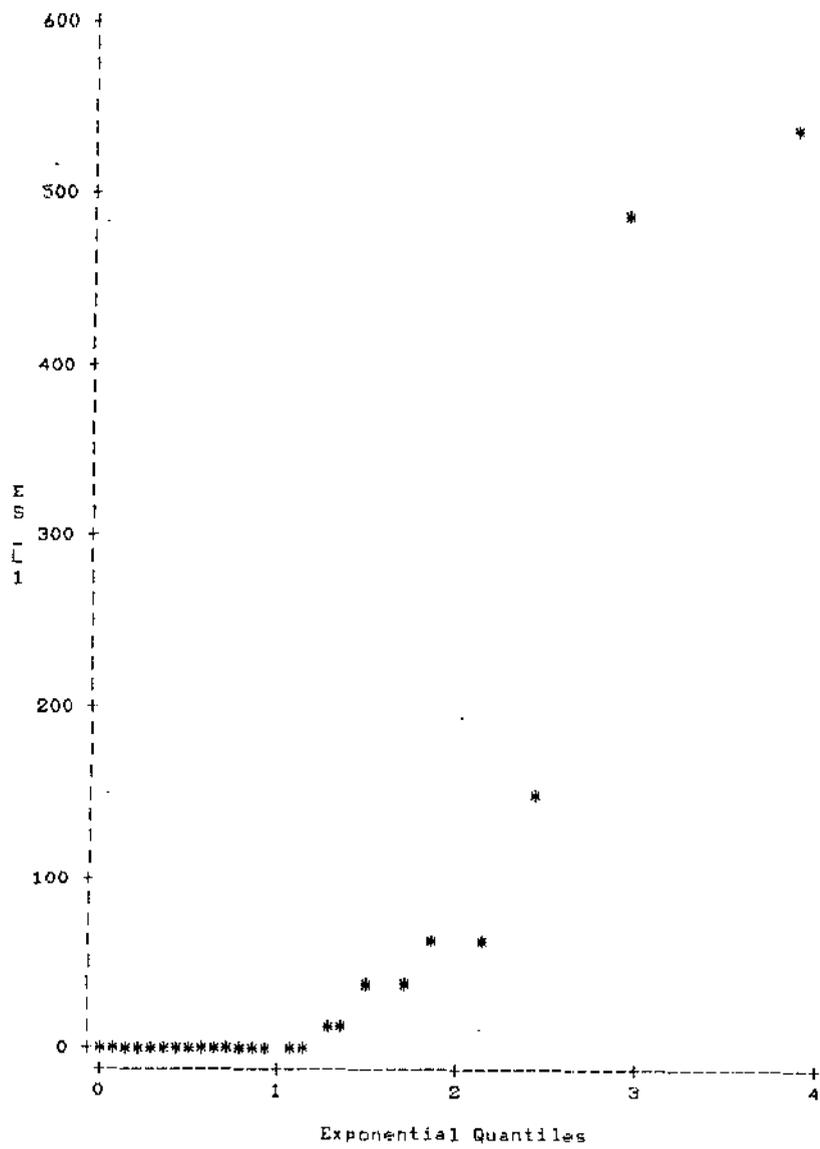


Figura 4.2.2 Gráfico probabilístico exponencial para os resíduos absolutos do modelo não transformado ajustado por mínimos desvios absolutos.

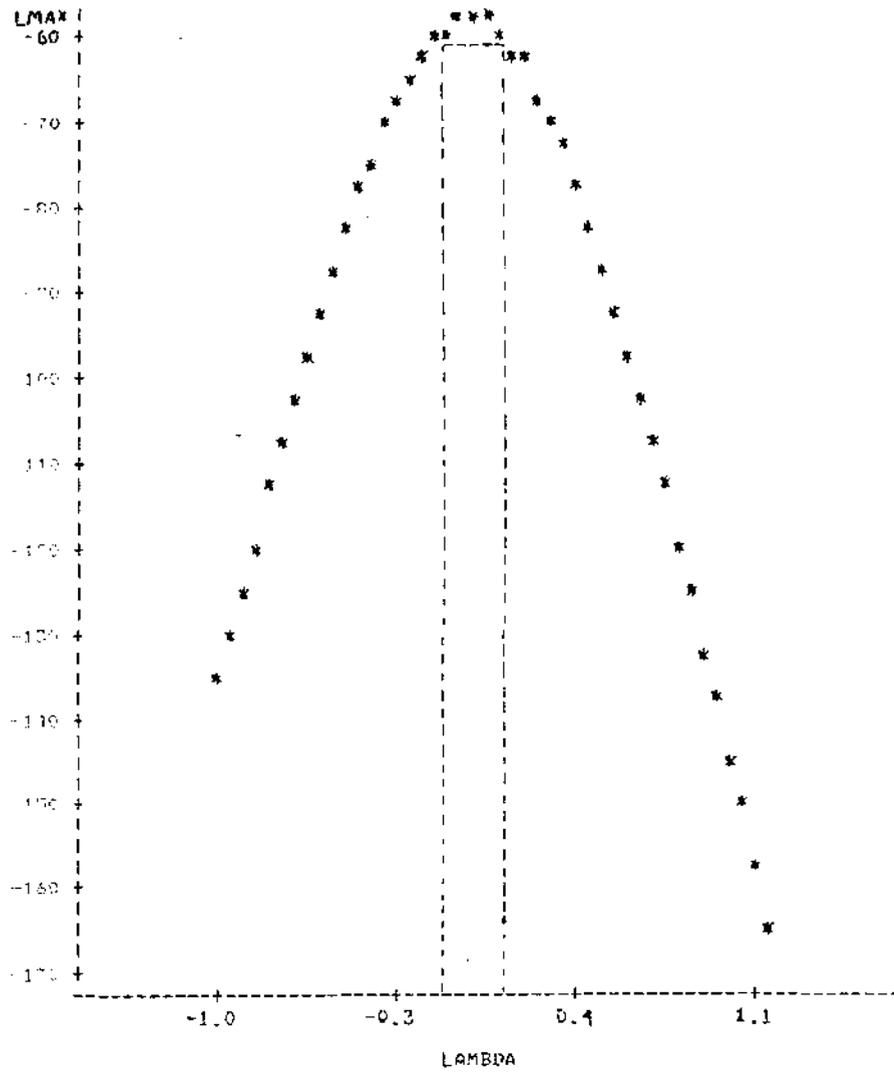


Figura 4.2.3 Gráfico do log de verossimilhança L_{\max} versus λ , sob mínimos quadrados.

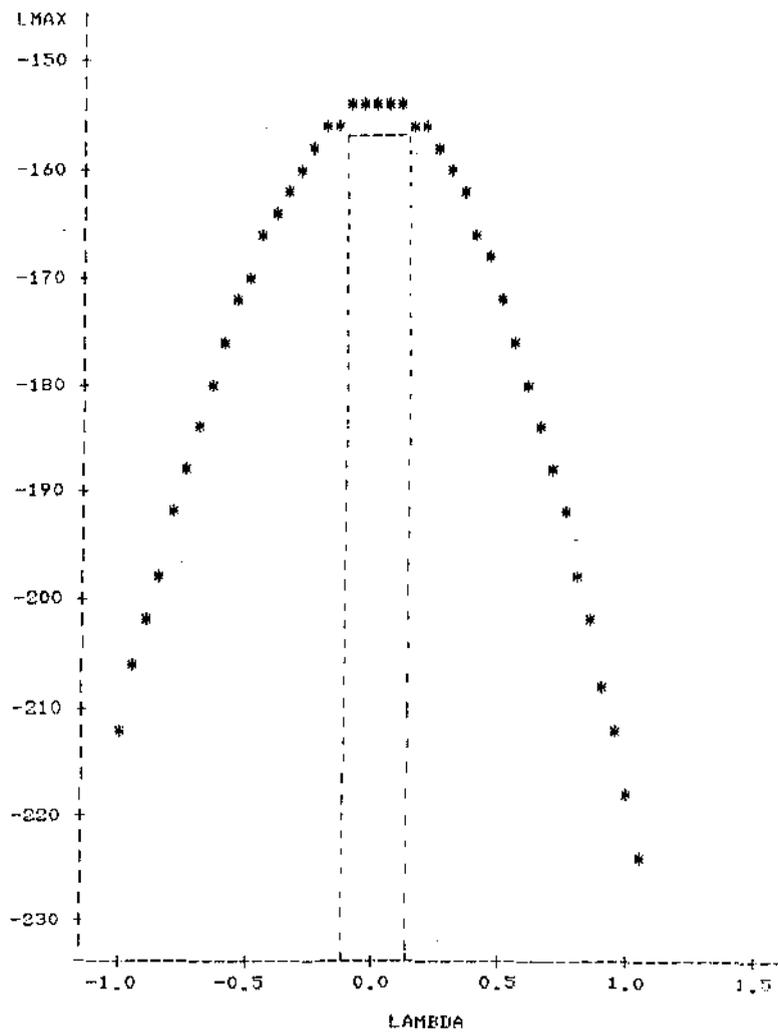


Figura 4.2.4 Gráfico do log de verossimilhança L_{\max} versus λ , sob mínimos desvios absolutos.

Após estimarmos os parâmetros dos modelos transformados ($\lambda=0$), por análise de mínimos quadrados e análise de mínimos desvios absolutos, obtemos os seguintes modelos ajustados:

Regressão de mínimos quadrados:

$$\widehat{\log(Y)} = 0.78 + 0.3 X$$

Soma dos Quadrados dos Resíduos = 229.94

Regressão de mínimos desvios absolutos:

$$\widehat{\log(Y)} = 1.95 - 2.12 X$$

Soma dos Resíduos Absolutos = 67.52

O gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0$), sob a análise de mínimos quadrados, é mostrado na figura 4.2.5. Observa-se uma melhora decisiva em relação à figura 4.2.1, sendo que, agora, todas as observações parecem estar adequadamente ajustadas.

A figura 4.2.6 mostra o gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado, sob análise de mínimos desvios absolutos. Observando este gráfico, verificamos que a hipótese de erros dupla-exponenciais não é aceita, pois os valores não estão próximos a nenhuma reta.

O gráfico de diagnóstico de Parker, mostrado na figura 4.2.7, indica que as observações 3 e 13 são as mais influentes na determinação de necessidade de uma transformação. A exclusão de tais observações do conjunto de dados, entretanto, não altera a permanência da transformação log no intervalo de confiança de 95% para λ , sob análise de mínimos desvios absolutos, assim como não proporciona mudanças significativas nas estimativas paramétricas e não melhora a disposição dos resíduos absolutos no gráfico probabilístico exponencial (tais análises não foram apresentadas por não conterem

mudanças significativas).

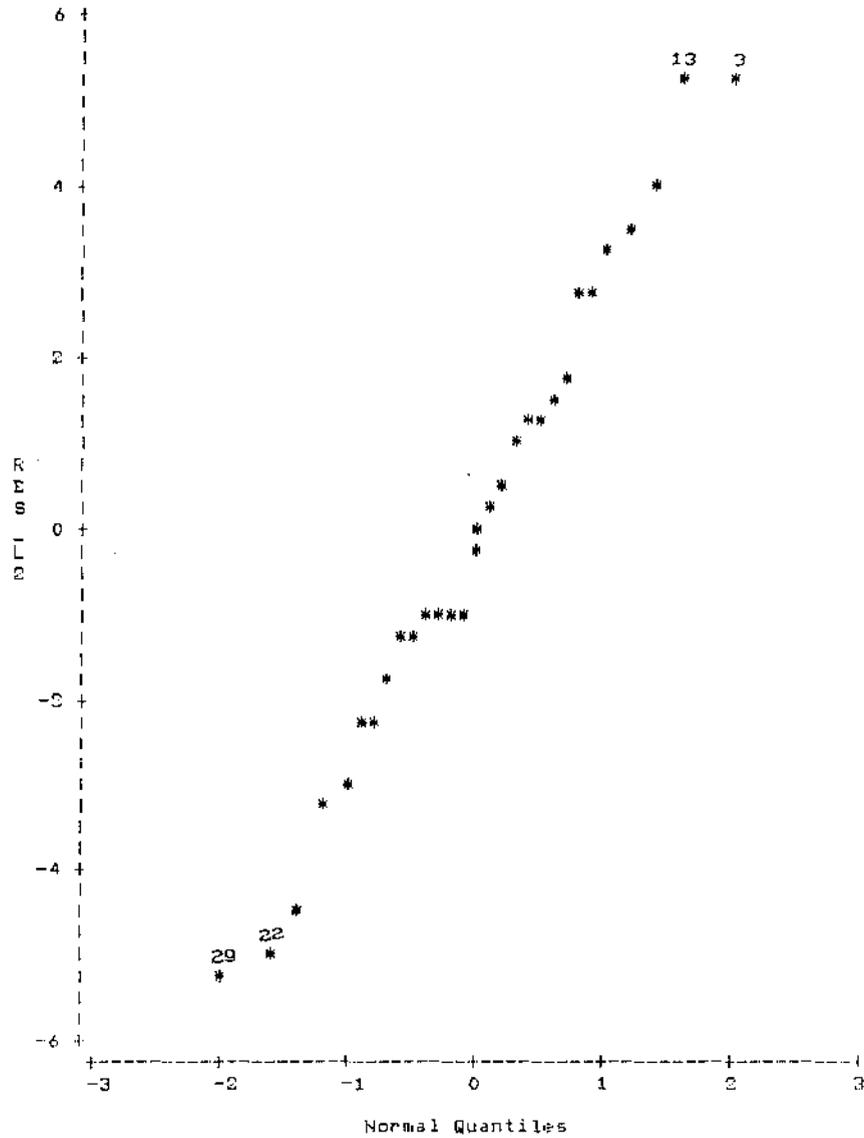


Figura 4.2.5 Gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0$) ajustado por mínimos quadrados.

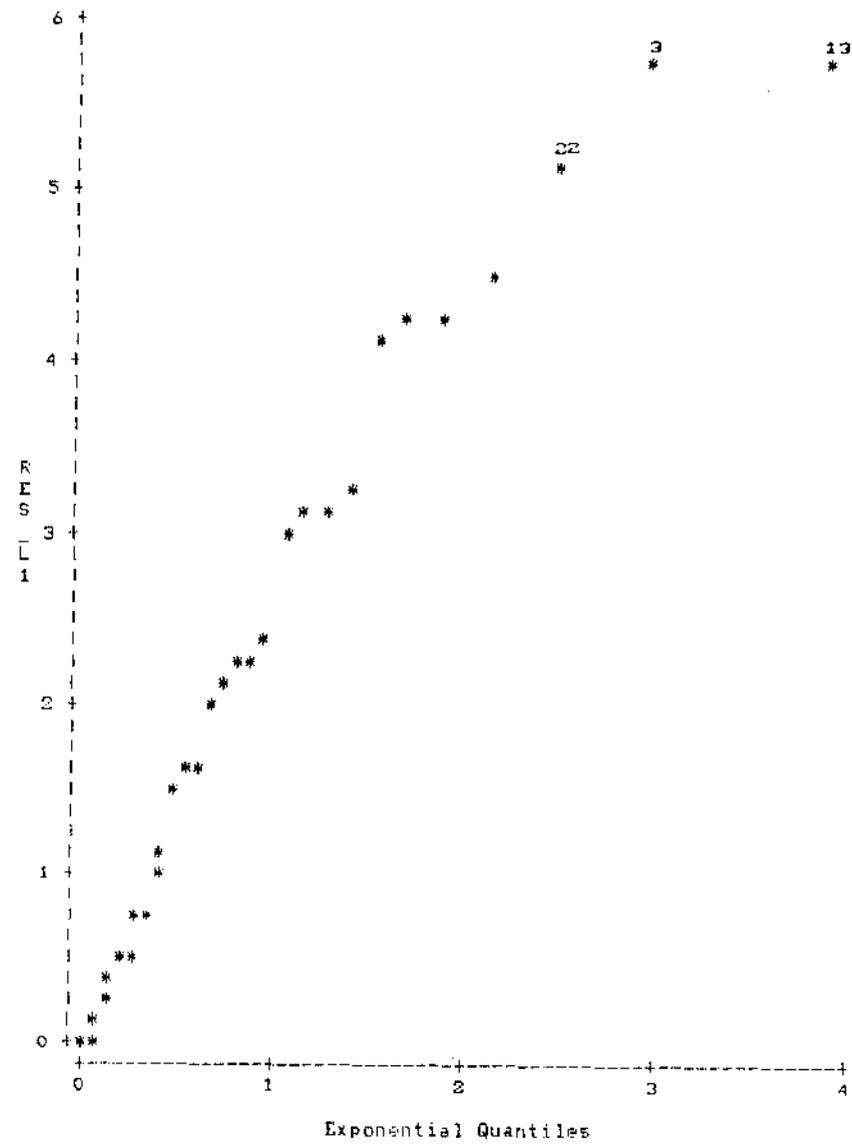


Figura 4.2.6 Gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) ajustado por mínimos desvios absolutos.

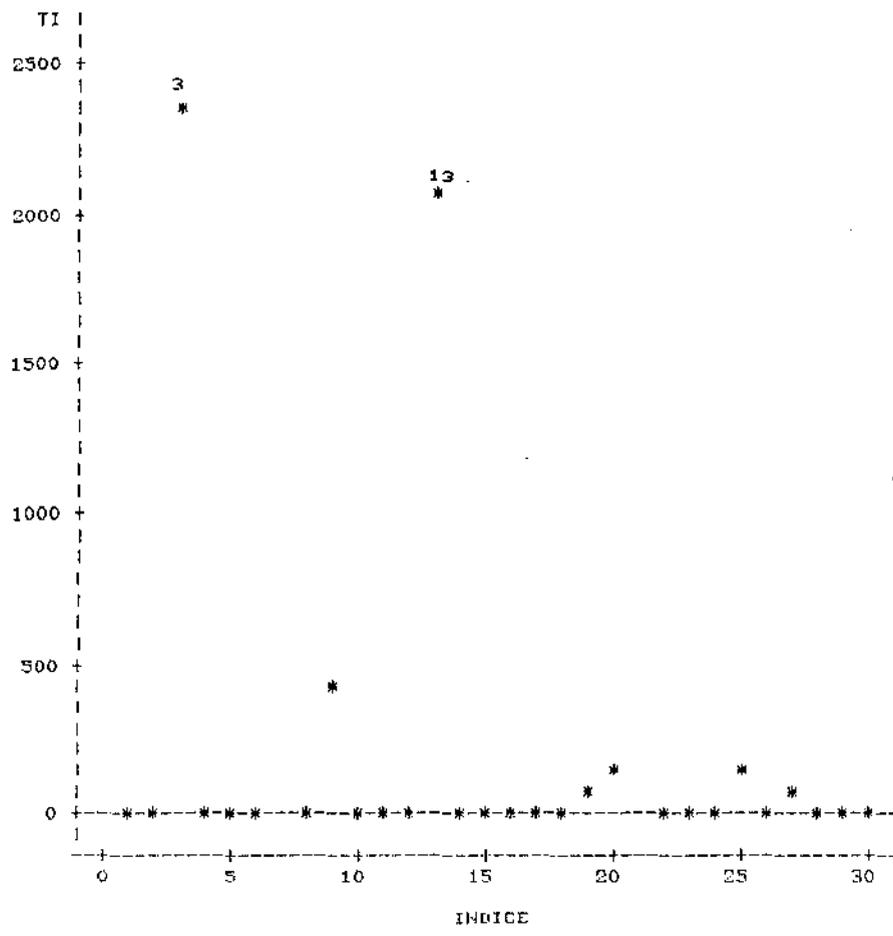


Figura 4.2.7 Gráfico de diagnóstico de Parker para observações não básicas.

Os diagnósticos de influência, sob minimização dos erros quadráticos, dados nas seções 3.2.1 e 3.2.2, mostrados nas figuras 4.2.8 e 4.2.9, identificam as observações 3 e 13 como as mais influentes na escolha da transformação. No gráfico de variável adicionada, figura 4.2.8, observamos que, embora, esta duas

observações estejam distantes das demais, elas estão de acordo com a tendência indicada pelas demais observações, ou seja, evidenciam a necessidade da mesma transformação indicada pelas outras observações. A ausência dessas observações, sob análise de mínimos quadrados, também não proporciona, entretanto, mudanças significativas nas análises subsequentes.

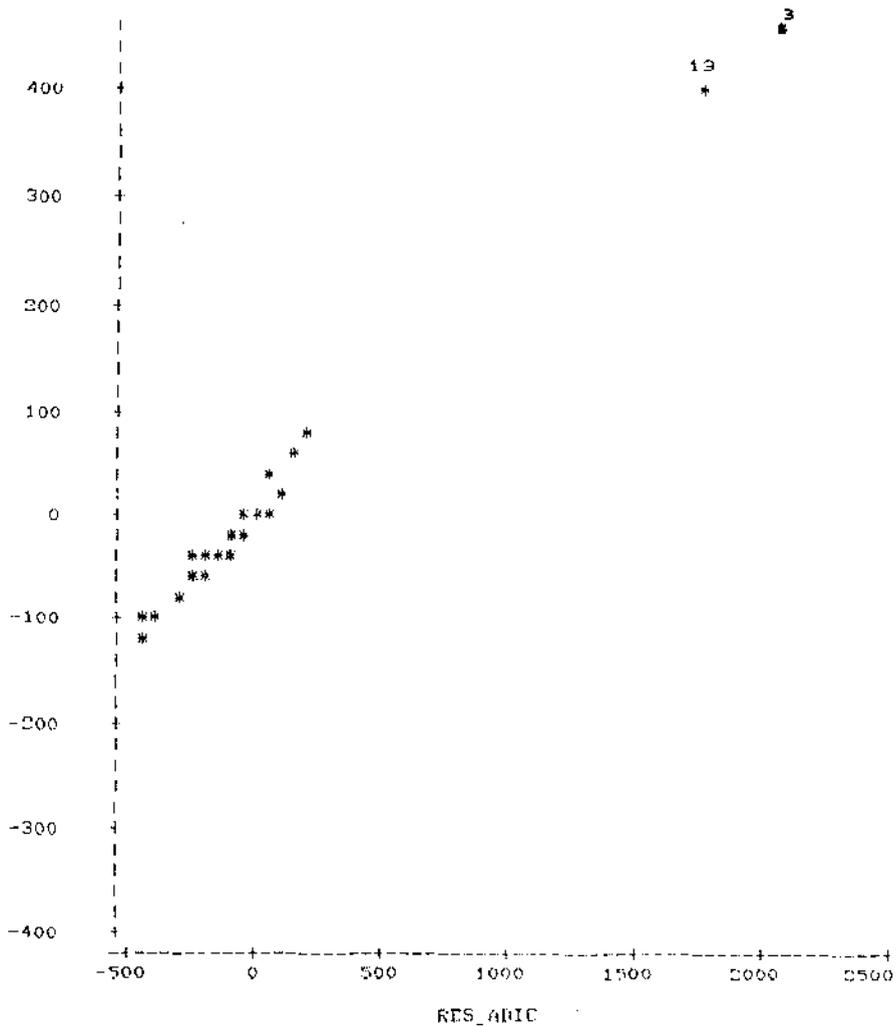


Figura 4.2.8 Gráfico de variável adicionada

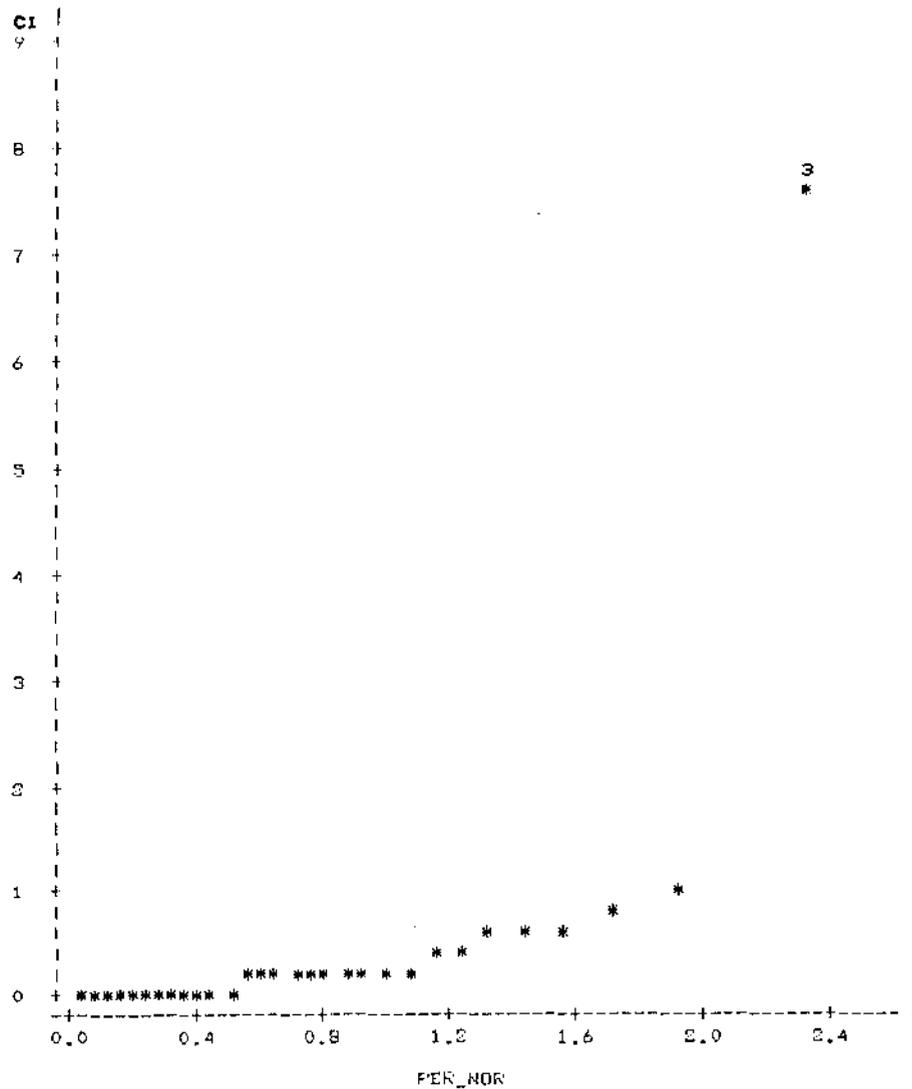


Figura 4.2.9 Gráfico semi-normal para estatística de Cook modificada proposta por Atkinson.

O método de Cook e Wang, como mostrado na figura 4.2.10, identifica as observações 22 e 29 como as mais influentes na escolha

da transformação, sob mínimos quadrados. Essas observações, assim como as observações 3 e 13, situadas nas extremidades da linha na figura 4.2.5, não apontam, entretanto, desvios da suposição de normalidade dos resíduos.

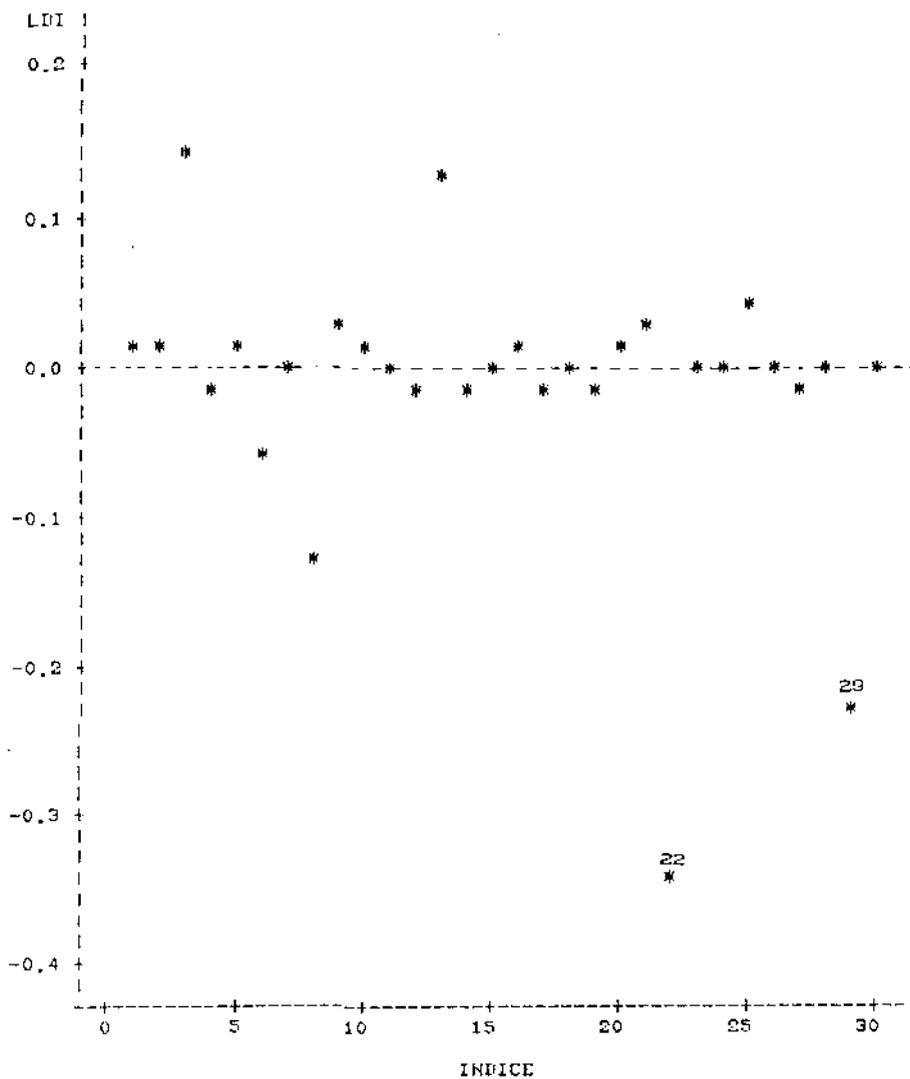


Figura 4.2.10 Gráfico para estatística de Cook e Wang.

Comparativamente, vemos que ambas análises de mínimos quadrados e mínimos desvios absolutos conseguem identificar a transformação log pré-estabelecida. Observando, entretanto, os gráficos probabilísticos, mostrados nas figuras 4.2.5 e 4.2.6, verificamos que os resíduos se acomodam adequadamente à suposição de normalidade sob mínimos quadrados e que na análise de mínimos desvios absolutos, os resíduos obtidos não apresentam distribuição dupla-exponencial.

4.1 PROBLEMA 2

O segundo conjunto de dados gerado, mostrado na tabela 4.3.1, possui erros provenientes de uma distribuição normal contaminada. A distribuição normal contaminada foi gerada através da distribuição normal, fazendo $f(x) = 0.85N_1 + 0.15N_2$, onde N_1 e N_2 representam normais com médias zero e desvios padrões iguais a um e cinco, respectivamente.

OBS	X	Y
1	0.24035	2.9072
2	0.20973	3.4433
3	0.85830	3.9076
4	0.69654	17.0814
5	0.86712	2.4000
6	0.58542	17.4154
7	0.50374	3.6733
8	0.43785	4.7912

9	0.60673	17.0603
10	0.75400	3.0896
11	0.40864	4.9165
12	0.28936	5.0222
13	0.49779	3.8464
14	0.79941	2.4303
15	0.66488	22.8810
16	0.55157	8.1044
17	0.01076	1.4590
18	0.53938	19.6928
19	0.82150	14.4348
20	0.17168	0.2578
21	0.76741	2.1292
22	0.32791	0.1729
23	0.87664	0.5295
24	0.85447	4.4532
25	0.83493	14.3818
26	0.53027	3.2940
27	0.65950	22.1659
28	0.60267	1.3337
29	0.28028	22.5794
30	0.27313	3.9508

Tabela 4.3.1 Lista de dados gerados sob distribuição de erros normal contaminada.

Inicialmente, ajustamos o modelo de primeira ordem, $\hat{Y} = \hat{a} + \hat{b} X$, por análise de mínimos quadrados e mínimos desvios absolutos. Os modelos obtidos são:

Regressão de mínimos quadrados:

$$\hat{Y} = 4.9 + 5.2 X$$

Soma dos Quadrados dos Resíduos = 1623.46

Regressão de mínimos desvios absolutos:

$$\hat{Y} = 3.1 + 1.6 X$$

Soma dos Resíduos Absolutos = 163.30

As análises dos gráficos probabilísticos para os resíduos dos modelos não transformados, sob mínimos quadrados e mínimos desvios absolutos, conforme as figuras 4.3.1 e 4.3.2, evidenciam a necessidade de uma transformação da variável resposta para ambas análises de mínimos quadrados e mínimos desvios absolutos.

Utilizando o procedimento de Box e Cox, sob mínimos desvios absolutos, obtemos que a transformação ótima é dada por $\hat{\lambda} = 0.1$ e que o intervalo de confiança de 95% para λ abrange os valores entre -0.2 e 0.45. A análise de mínimos desvios absolutos, como no problema anterior, identifica, portanto, a transformação log pré-estabelecida como uma transformação apropriada para a variável resposta.

Sob mínimos quadrados, através do procedimento de Box e Cox, obtemos que o log de verossimilhança assume o máximo em $\hat{\lambda} = 0.2$. O intervalo de confiança de 95% para λ abrange os valores entre 0.05 e 0.45. Portanto, temos que o procedimento de Box e Cox, sob mínimos quadrados, não consegue identificar a transformação log

pré-estabelecida como uma transformação apropriada para esse conjunto de dados.

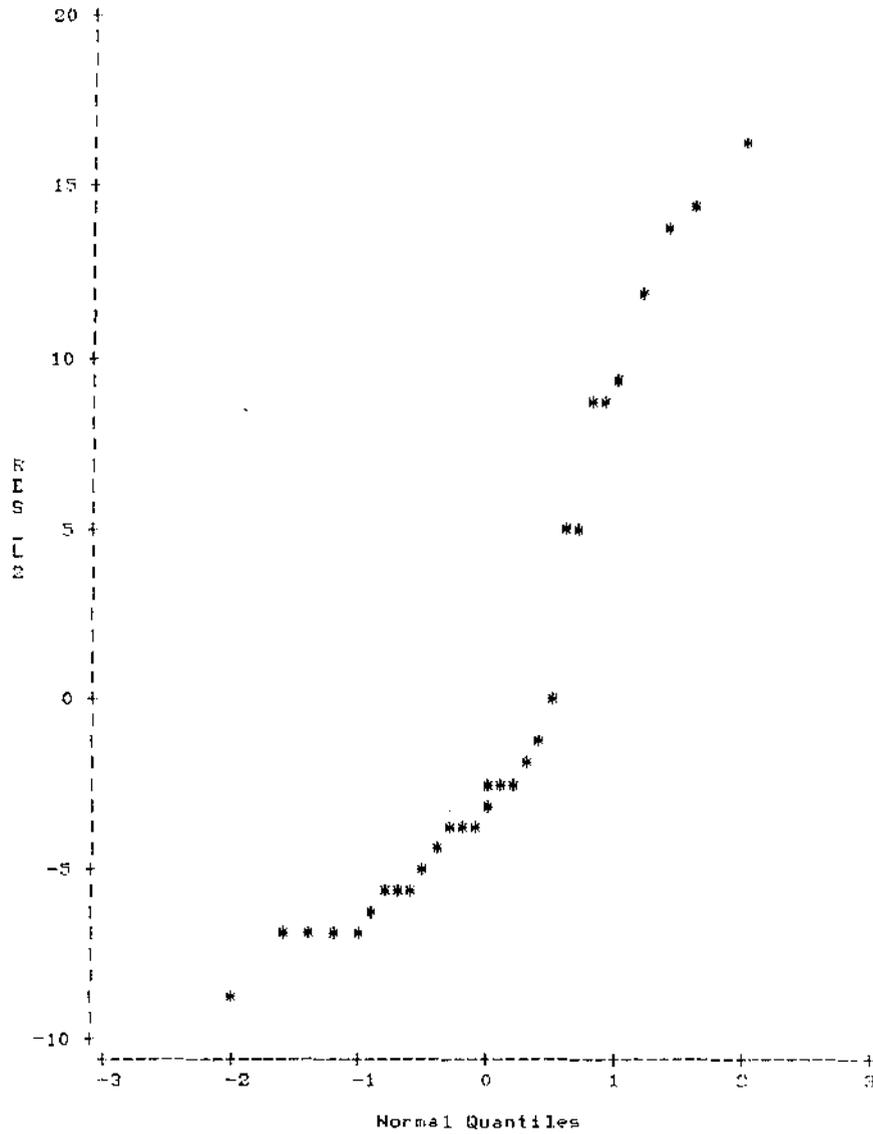


Figura 4.3.1 Gráfico probabilístico normal para os resíduos do modelo não transformado ajustado por mínimos quadrados.

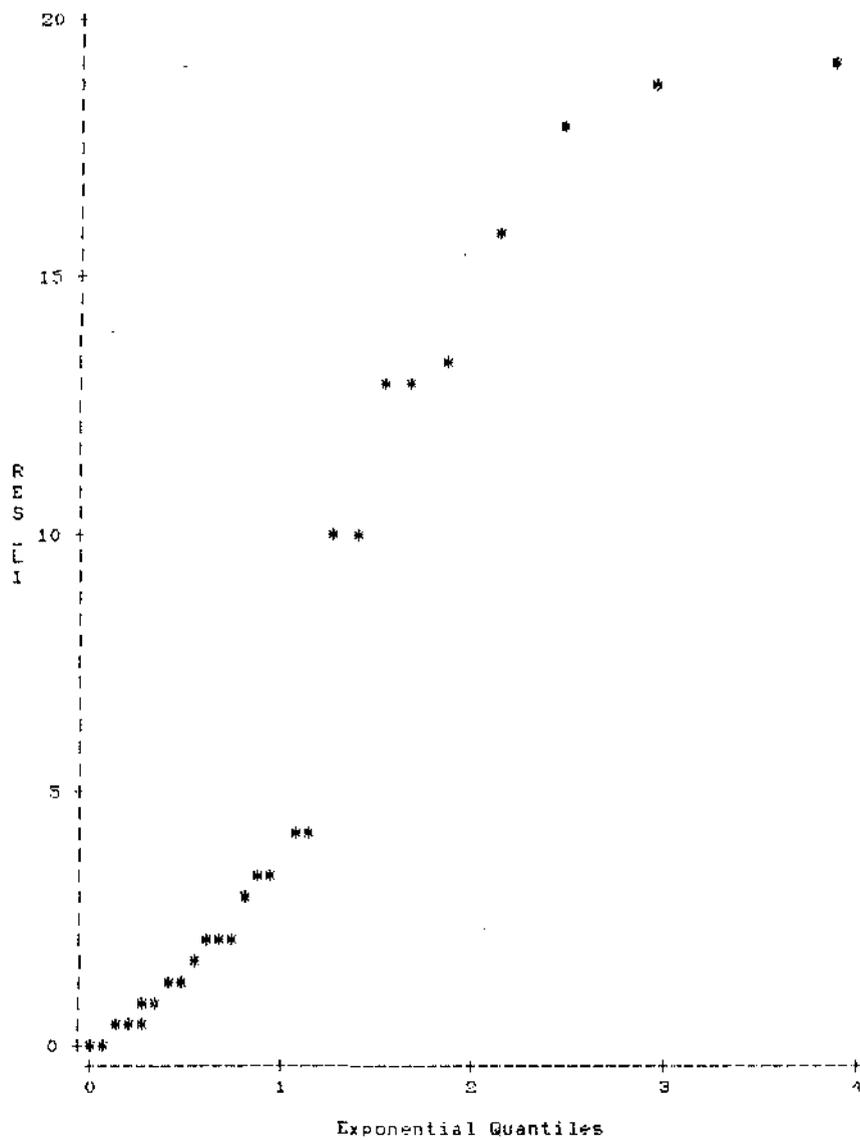


Figura 4.3.2 Gráfico probabilístico exponencial para os resíduos absolutos do modelo não transformado ajustado por mínimos desvios absolutos.

Após estimarmos os parâmetros dos modelos transformados, por análise de mínimos quadrados e análise de mínimos desvios absolutos, obtemos as seguintes regressões:

Regressão de mínimos quadrados:

$$\hat{Y}^{(0.2)} = 1.1 + 1.5 X$$

Soma dos Quadrados dos Resíduos = 72.45

Regressão de mínimos desvios absolutos:

$$\hat{\log(Y)} = 1.2 + 0.4 X$$

Soma dos Resíduos Absolutos = 27.66

O gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) por análise de mínimos desvios absolutos, mostrado na figura 4.3.3, evidencia que, agora, todas as observações parecem estar razoavelmente modeladas, com as observações 22, 20 e 23 possuindo os maiores resíduos e fora do alinhamento das demais observações. A exclusão dessas observações não alteraram, entretanto, os resultados obtidos por análise de mínimos desvios absolutos.

A figura 4.3.4 mostra o gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0.2$), sob a análise de mínimos quadrados. Observa-se, que o gráfico probabilístico indica claramente que a hipótese de normalidade não é atendida.

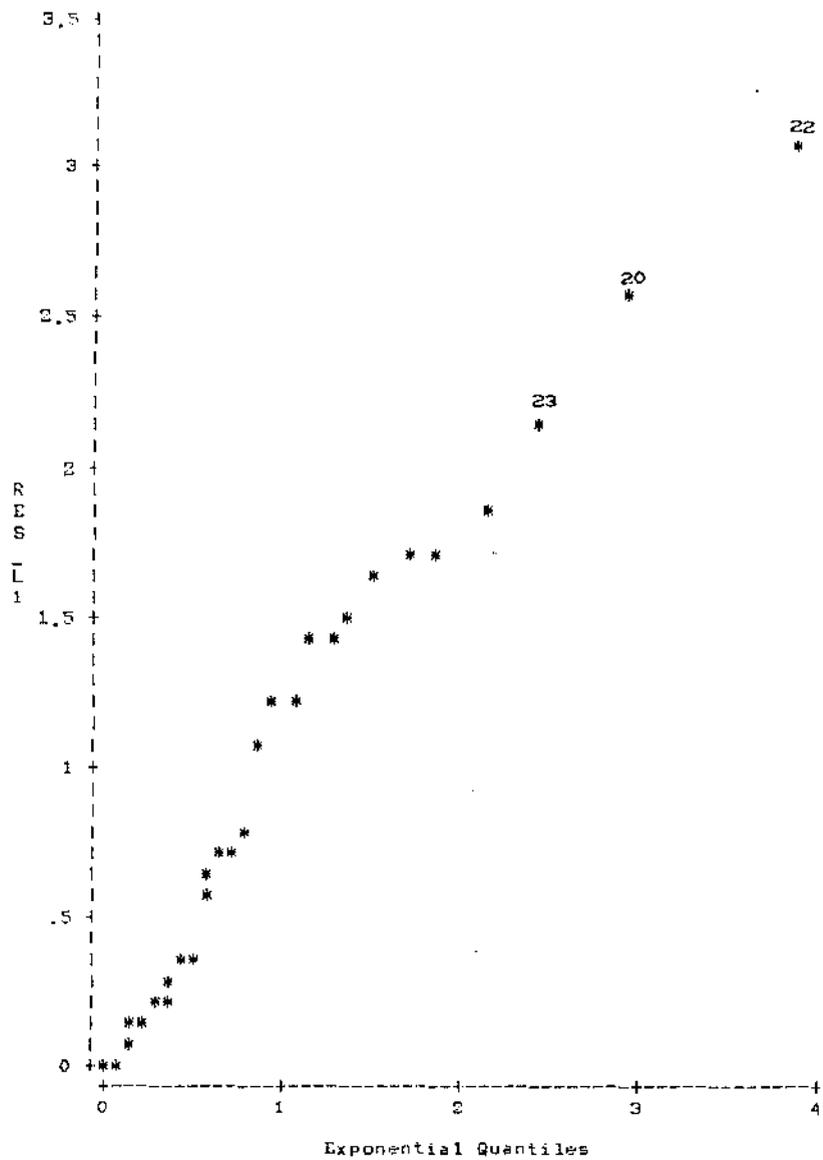


Figura 4.3.3 Gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) ajustado por mínimos desvios absolutos.

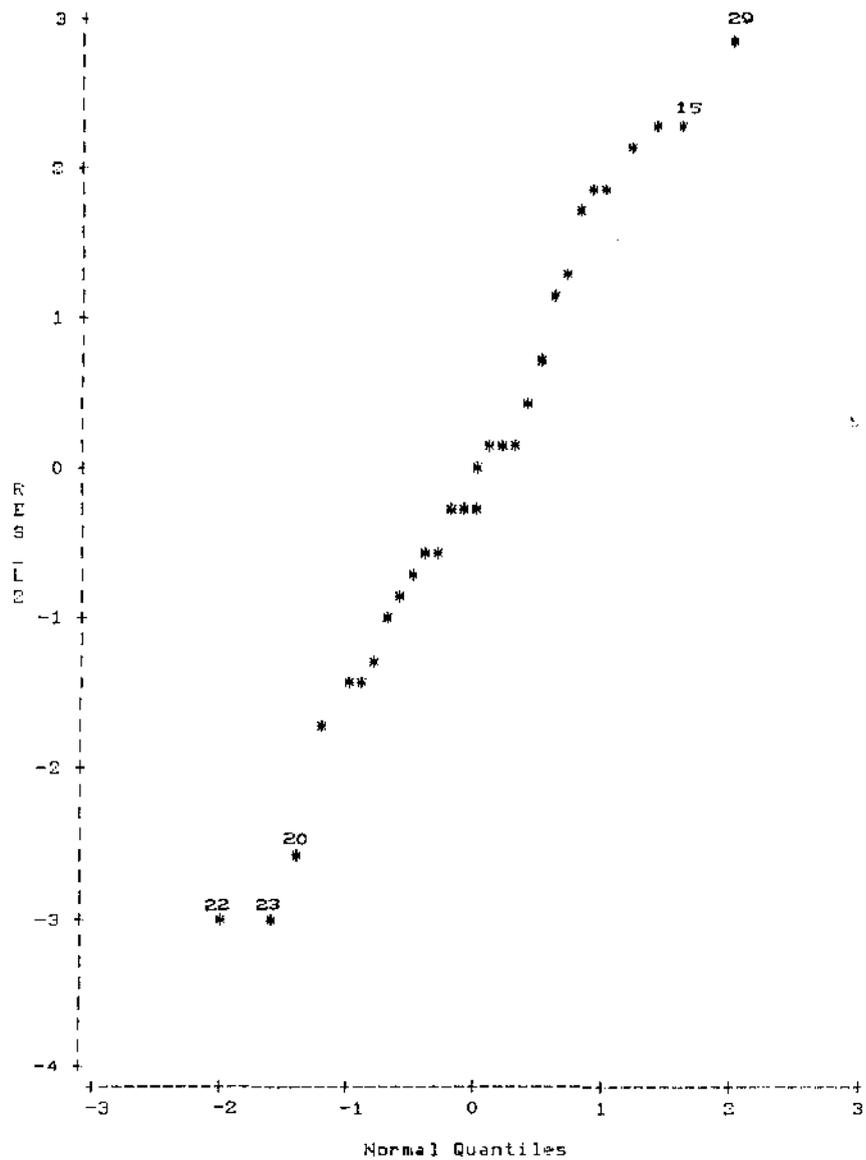


Figura 4.3.4 Gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0.2$) ajustado por mínimos quadrados.

Através do diagnóstico de influência de Cook e Wang, ilustrado na figura 4.3.5, observamos que este identifica as observações 20, 22 e 23 como observações aberrantes e/ou pontos influentes. O gráfico de variável adicionada, mostrado na figura 4.3.6, não identifica nenhuma observação como observação aberrante, enquanto que o gráfico para a estatística de Cook modificada, apresentado na figura 4.3.7, aponta a observação 15 como possível valor influente.

Após a exclusão das observações 20, 22 e 23 do conjunto de dados, o procedimento de Box e Cox, sob mínimos quadrados, seleciona como ótima a transformação $\hat{\lambda} = -0.2$ e fornece um intervalo de confiança de 95% para λ que abrange os valores entre -0.065 e 0.30 . Portanto, após a retirada das observações identificadas por Cook e Wang como aberrantes e/ou influentes, temos que a análise de mínimos quadrados identifica a transformação log pré-estabelecida. O modelo transformado ($\lambda=0$) ajustado por mínimos quadrados sem as observações 20, 22 e 23 é dado por:

Regressão de mínimos quadrados :

$$\hat{\log(Y)} = 1.39 + 0.67 X$$

Soma dos Quadrados dos Resíduos = 20.60

O gráfico probabilístico normal para os resíduos deste modelo transformado, mostrado na figura 4.3.8, continua mostrando a não adequabilidade da suposição de normalidade dos resíduos, ao evidenciar uma distribuição dos resíduos típica de distribuições de caudas alongadas.

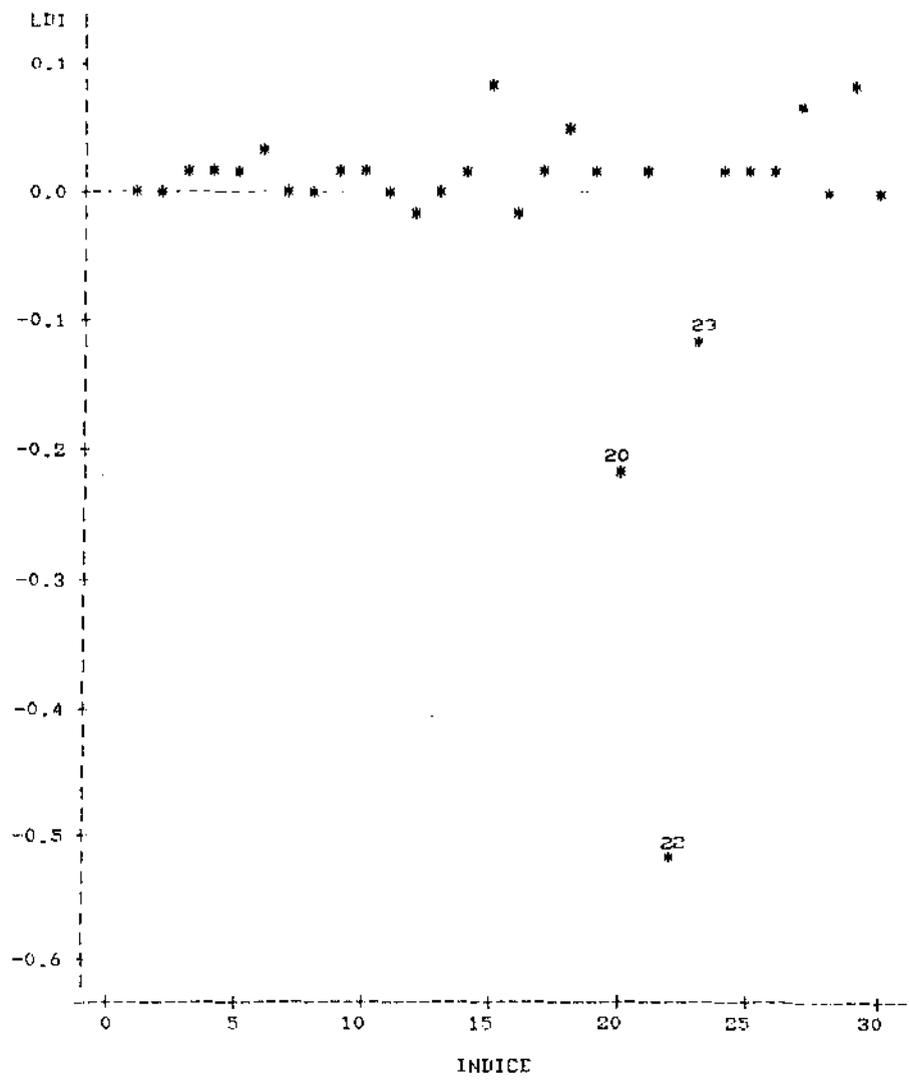


Figura 4.3.5 Gráfico para estatística de Cook e Wang.

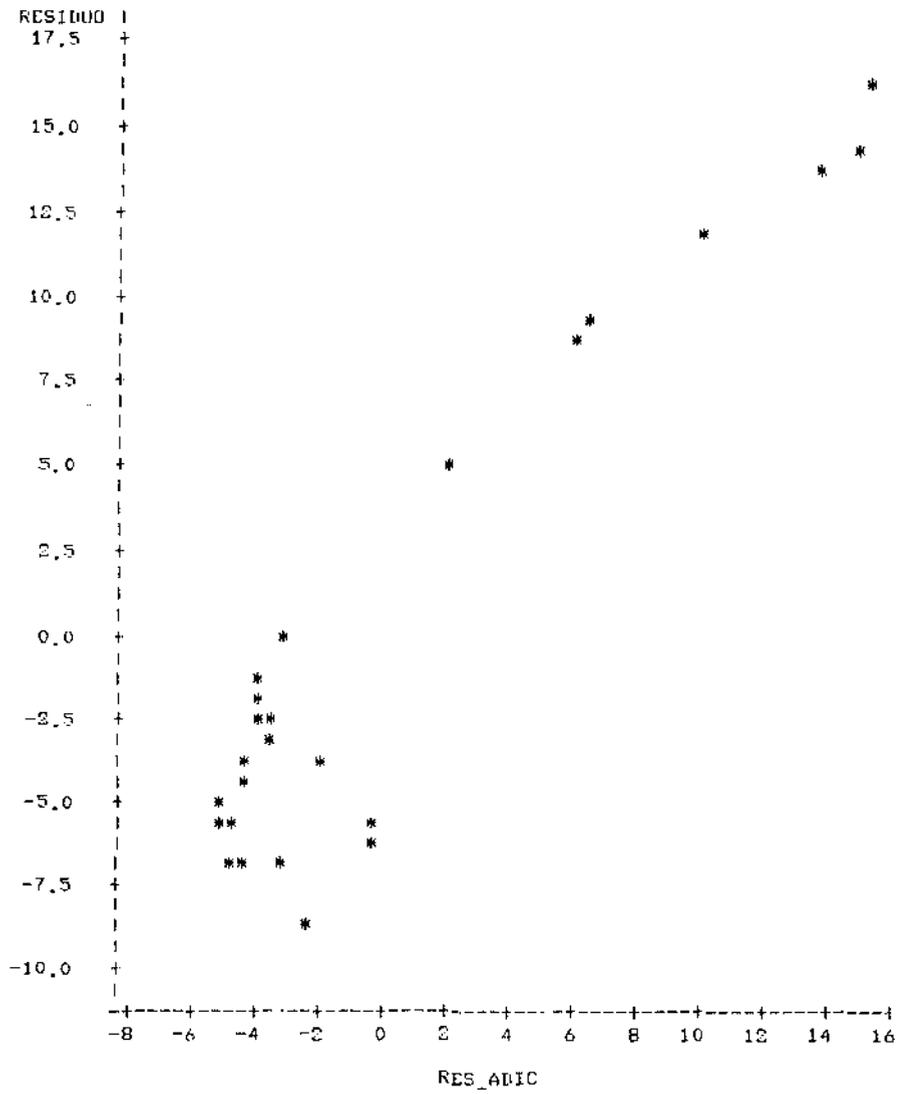


Figura 4.3.6 Gráfico de variável adicionada.

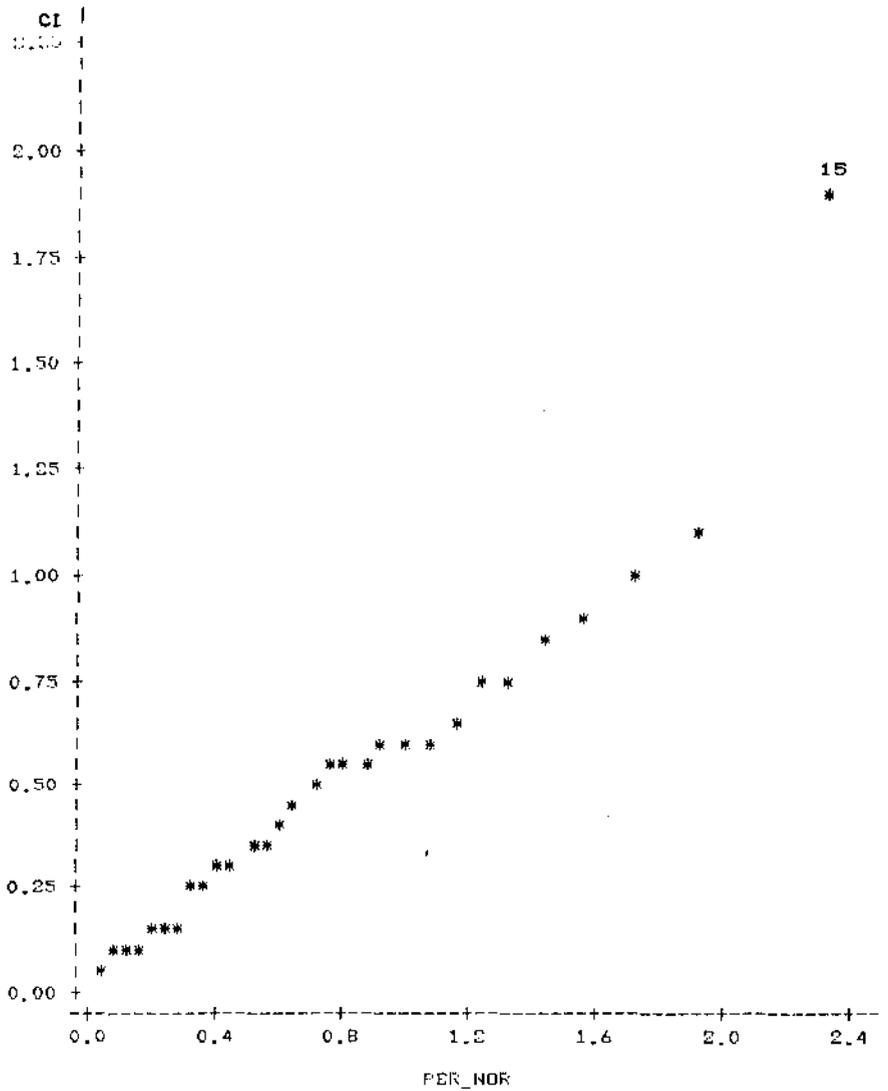


Figura 4.3.7 Gráfico semi-normal para estatística de Cook modificada proposta por Atkinson.

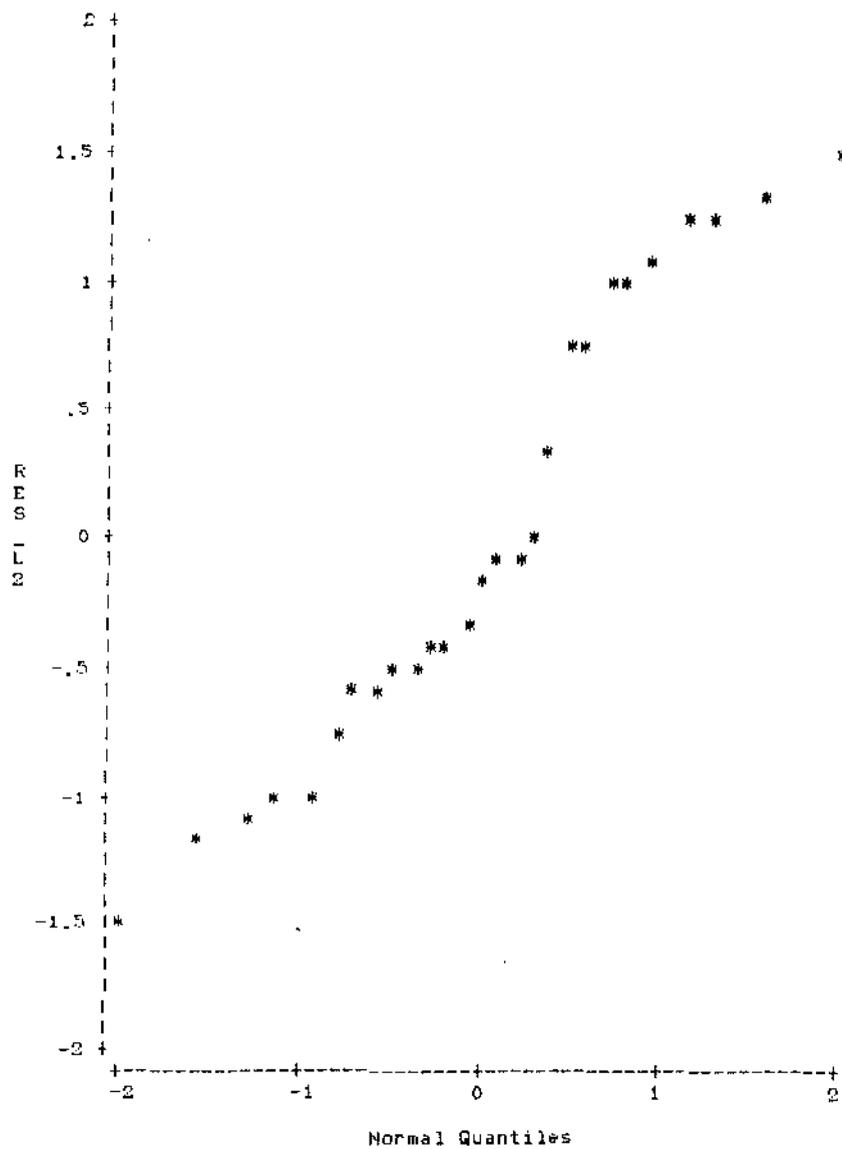


Figura 4.3.8 Gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0$) ajustado por mínimos quadrados, após a exclusão das observações 20, 22 e 23.

Para investigarmos a influência de observações individuais na determinação da necessidade de uma transformação da variável resposta, sob análise de mínimos desvios absolutos, construímos o gráfico de diagnóstico de Parker, mostrado na figura 4.3.9. Este método de diagnóstico indica as observações 15, 27 e 29 como as mais influentes em determinar a necessidade de uma transformação; entretanto, vemos que outras observações também possuem influência significativa em determinar essa necessidade. Desse modo, temos que a necessidade de uma transformação está amplamente espalhada através do conjunto de dados.

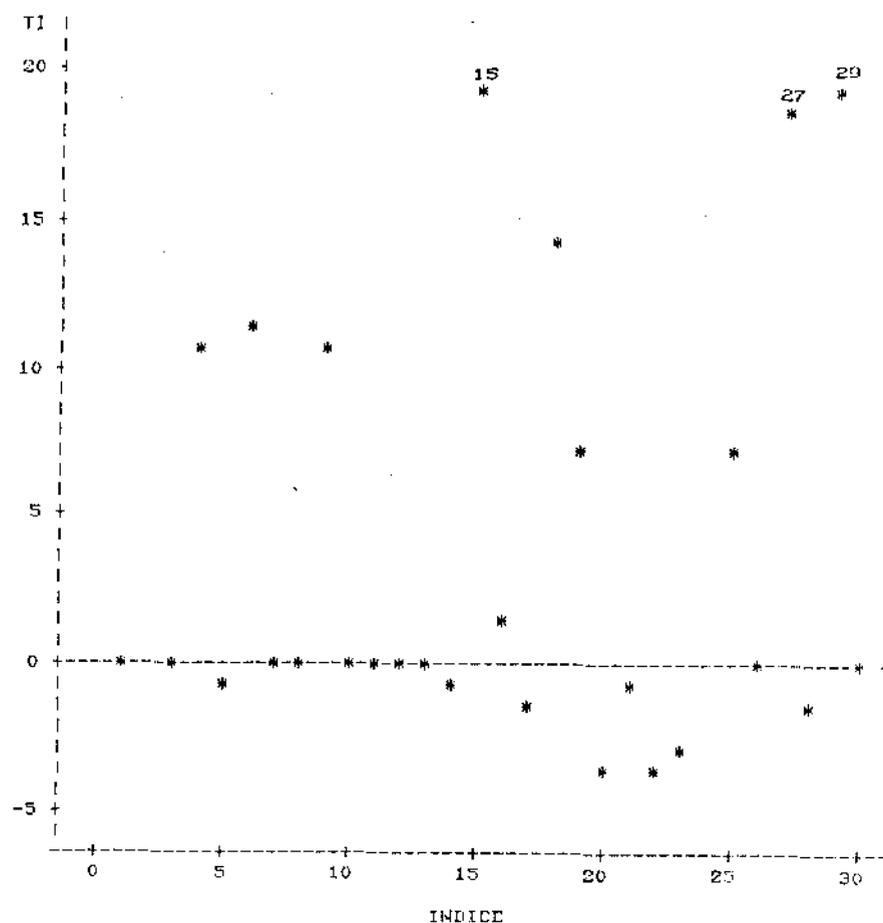


Figura 4.3.9 Gráfico de diagnóstico de Parker para observações não básicas.

Comparando os resultados obtidos por análise de mínimos quadrados e mínimos desvios absolutos, temos que parece mais razoável ajustarmos ao conjunto de dados, o modelo transformado por análise de mínimos desvios absolutos ($\lambda=0$).

4.4 PROBLEMA 3

Para o terceiro conjunto de dados gerado, mostrado na tabela 4.4.1, ϵ é uma amostra de uma distribuição dupla-exponencial com $\theta = 3$.

OBS	X	Y
1	0.63459	20.06
2	0.94395	10.25
3	0.20833	2.82
4	0.31270	5.79
5	0.73129	0.05
6	0.68019	0.00
7	0.82735	203.35
8	0.79867	37.52
9	0.39280	1.57
10	0.20163	0.00
11	0.94028	2.19
12	0.62943	0.02
13	0.35380	6.84

14	0.22326	7.61
15	0.54077	9.16
16	0.37388	12.06
17	0.37473	0.24
18	0.65878	1432.16
19	0.24526	423.76
20	0.85314	432.36
21	0.16818	3.87
22	0.32020	0.28
23	0.70115	21.98
24	0.60474	0.30
25	0.88693	149.00
26	0.18298	2.31
27	0.46658	389.51
28	0.82426	39.33
29	0.96678	45.20
30	0.33925	1.50

Tabela 4.4.1 Lista de dados gerados sob distribuição de erros dupla-exponencial $\theta = 3$.

Como nos problemas anteriores, ajustamos, inicialmente, o modelo $\hat{Y} = \hat{a} + \hat{b} X$ por análise de mínimos quadrados e mínimos desvios absolutos. As regressões obtidas são dadas por:

Regressão de mínimos quadrados:

$$\hat{Y} = 31.9 + 140.5 X$$

Soma dos Quadrados dos Resíduos = 2245164.6

Regressão de mínimos desvios absolutos:

$$\hat{Y} = -8.1 + 44.3 X$$

Soma dos Resíduos Absolutos = 3129.55

A não adequabilidade desses modelos ajustados é apontada pelos gráficos probabilísticos para os resíduos, mostrados nas figuras 4.4.1 e 4.4.2. Através desses gráficos podemos observar fortíssimas evidências da necessidade de uma transformação da variável resposta, sob análise de mínimos quadrados e análise de mínimos desvios absolutos.

Através do procedimento de Box e Cox, sob mínimos desvios absolutos, obtemos que a transformação ótima é dada por $\hat{\lambda} = 0.05$ e que o intervalo de confiança de 95% para λ abrange os valores entre -0.05 e 0.2. Portanto, a análise de mínimos desvios absolutos, novamente identifica a transformação log pré-estabelecida como uma transformação adequada para a variável resposta.

Sob a análise de mínimos quadrados, através do procedimento de Box e Cox, obtemos que o log de verossimilhança assume o máximo em $\hat{\lambda} = 0.10$. O intervalo de confiança de 95% para λ , sob mínimos quadrados, abrange os valores entre 0.05 e 0.15. Como no problema anterior, observamos, então, que o procedimento de Box e Cox, sob mínimos quadrados, não consegue selecionar a transformação log pré-estabelecida.

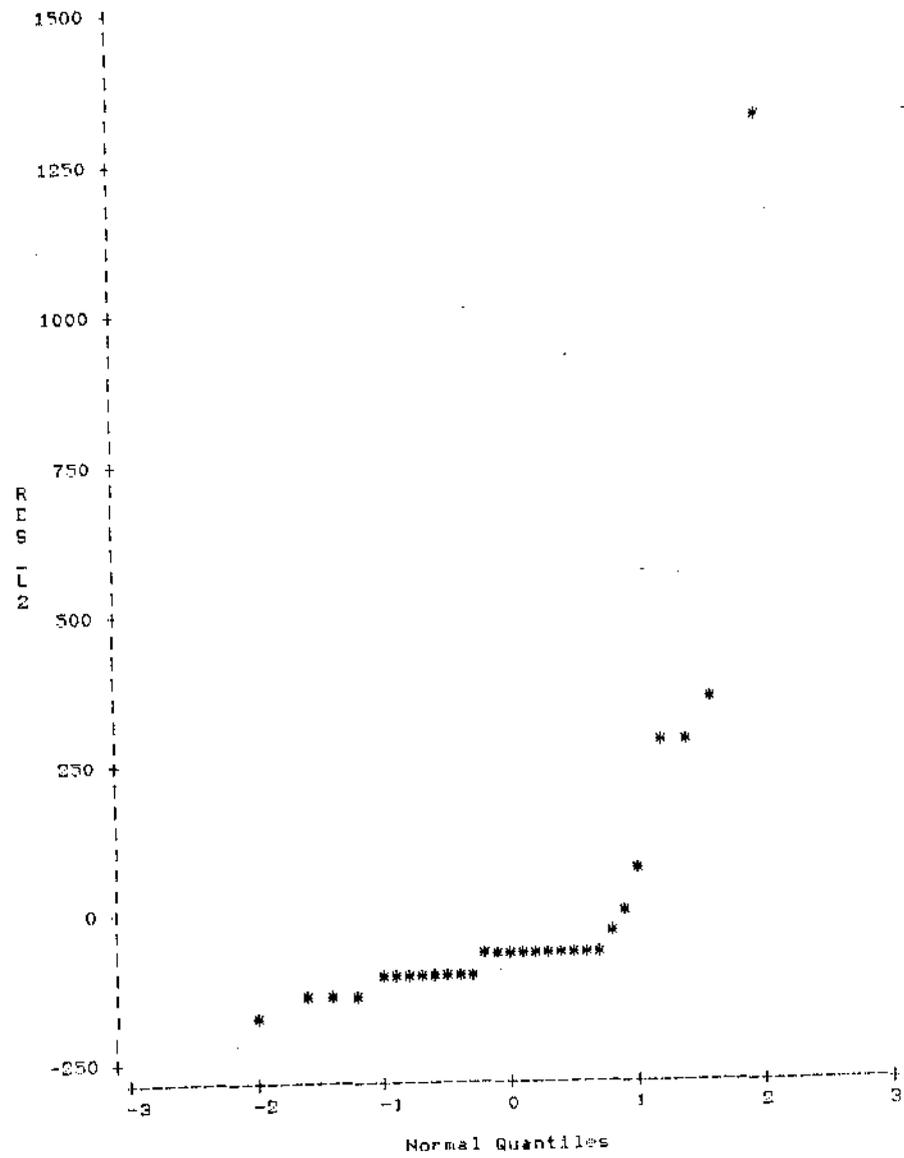


Figura 4.4.1 Gráfico probabilístico normal para os resíduos do modelo não transformado ajustado por mínimos quadrados.

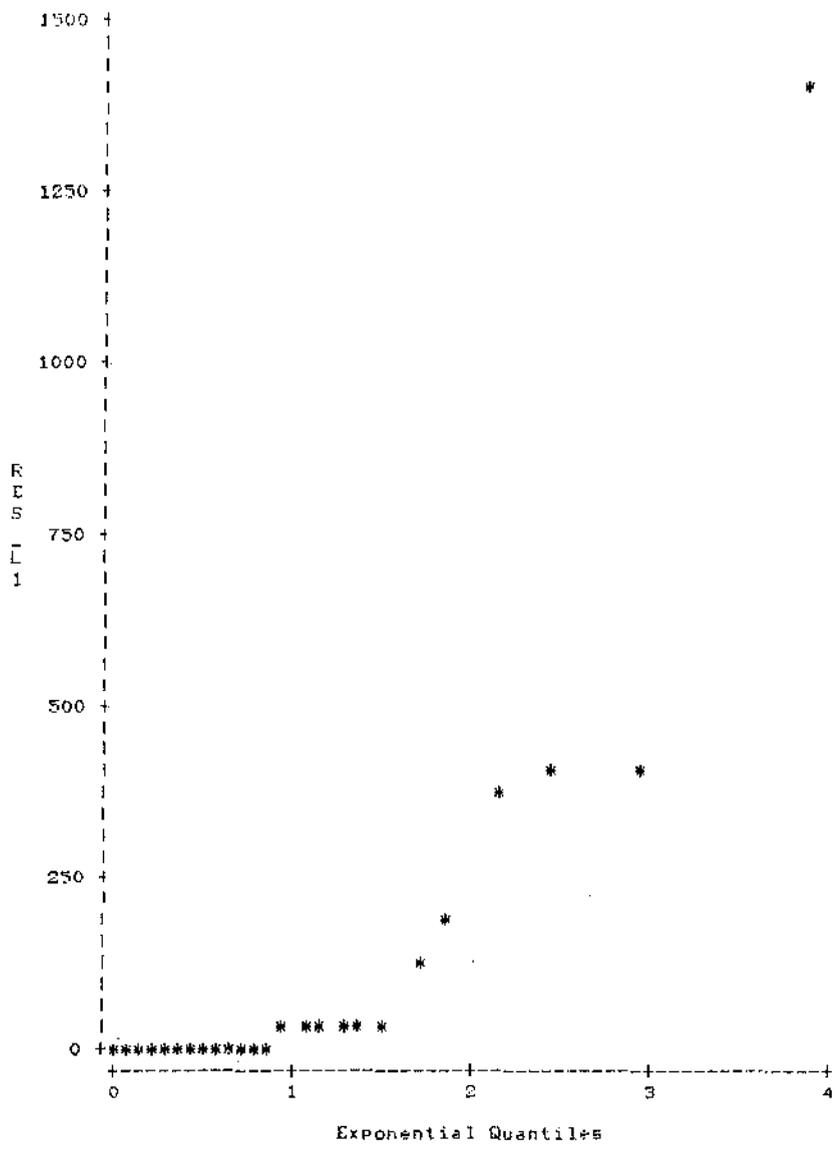


Figura 4.4.2 Gráfico probabilístico exponencial para os resíduos absolutos do modelo não transformado ajustado por mínimos desvios absolutos.

Os modelos transformados ajustados por análise de

mínimos desvios absolutos ($\lambda=0$) e por análise de mínimos quadrados ($\lambda=0.10$) são dados por:

Regressão de mínimos quadrados:

$$\hat{Y}^{(0.1)} = -0.03 + 4.30 X$$

Soma dos Quadrados dos Resíduos = 405.86

Regressão de mínimos desvios absolutos:

$$\hat{\log(Y)} = 0.17 + 4.17 X$$

Soma dos Resíduos Absolutos = 69.74

A figura 4.4.3 mostra o gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) sob análise de mínimos desvios absolutos. Ele ilustra uma melhora decisiva sobre o gráfico probabilístico exponencial para os resíduos absolutos do modelo não transformado (figura 4.4.2). A figura 4.4.3 evidencia também que, sob o modelo transformado, todas as observações se acomodam adequadamente à suposição de dupla-exponencialidade, com exceção da observação 6 que está ligeiramente distanciada das demais observações.

A exclusão da observação 6, sob análise de mínimos desvios absolutos, não altera, entretanto, o intervalo de confiança de 95% para λ , assim como as estimativas paramétricas continuam as mesmas obtidas para o modelo transformado ($\lambda=0$) com todas as observações; os demais resultados subsequentes, também, não sofrem mudanças significativas.

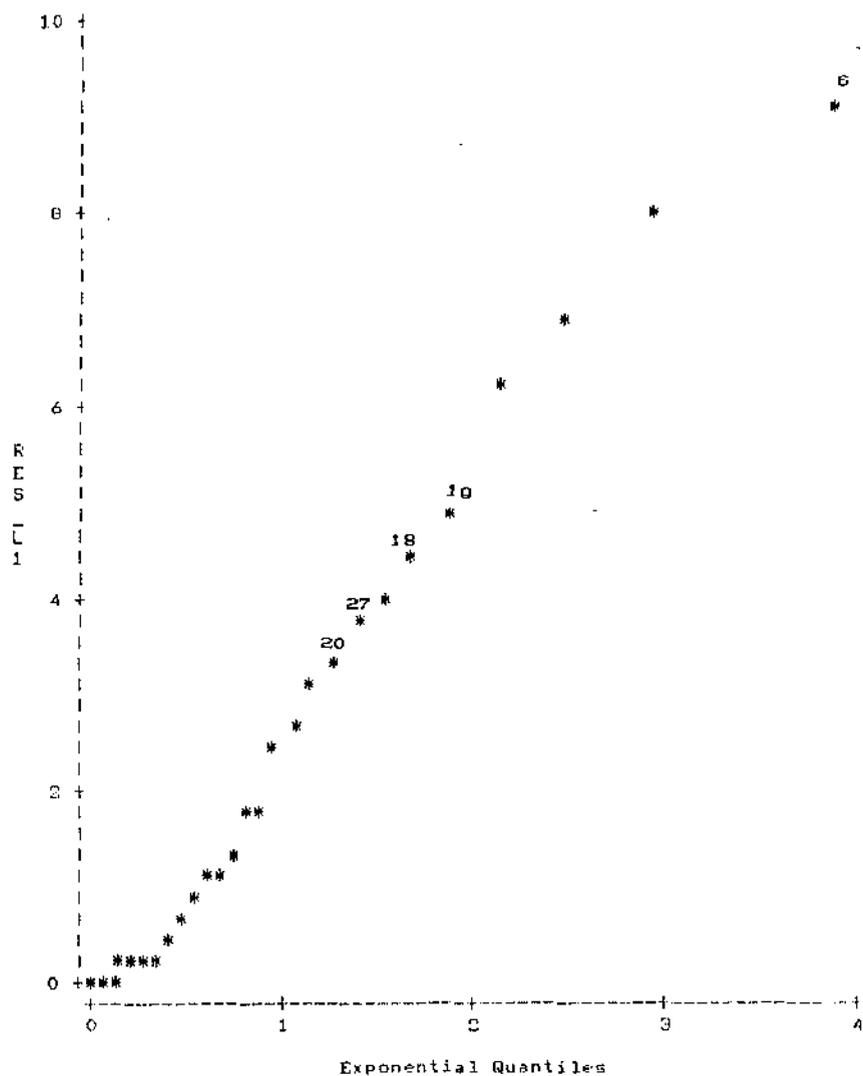


Figura 4.4.3 Gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) ajustado por mínimos desvios absolutos.

O gráfico probabilístico normal para os resíduos do

modelo transformado ($\lambda=0.10$) sob análise de mínimos quadrados é mostrado na figura 4.4.4. A variação em relação a uma tendência linear é muito grande para ser explicada somente por flutuações aleatórias, levando-nos a concluir que os resíduos não são normalmente distribuídos.

Para avaliarmos a influência de observações individuais na determinação da necessidade de uma transformação, sob mínimos quadrados, utilizamos as figuras 4.4.5, 4.4.6 e 4.4.7, que ilustram os gráficos de variável adicionada, estatística de Cook modificada e estatística de Cook e Wang, respectivamente. Através dos gráficos de variável adicionada e estatística de Cook modificada, obtemos que a observação 18 é uma observação aberrante. Após retirarmos essa observação do conjunto de dados, identificamos através desses mesmos gráficos as observações 19, 20 e 27 como valores aberrantes. A figura que ilustra a estatística de Cook e Wang aponta como observações influentes as seguintes observações: 18 e 10.

O procedimento de Box e Cox, sob mínimos quadrados, aplicado ao conjunto de dados após a exclusão de cada um dos seguintes subconjuntos de observações: {18}, {18, 19, 20}, {18, 19, 20, 27}, não fornece, em nenhum dos conjuntos resultantes, um intervalo de confiança de 95% para λ que contenha a transformação log pré-estabelecida. Para os conjuntos de dados resultantes da exclusão dos subconjuntos {18}, {18, 19, 20} o procedimento continua apontando como ótima a transformação $\hat{\lambda} = 0.10$, enquanto que para o conjunto resultante da exclusão do subconjunto {18, 19, 20, 27}, o procedimento aponta como ótima a transformação $\hat{\lambda} = 0.15$. Os gráficos probabilísticos normais para os resíduos dos modelos transformados, após a exclusão desses subconjuntos de observações evidenciam, também, as mesmas tendências observadas na figura 4.4.4.

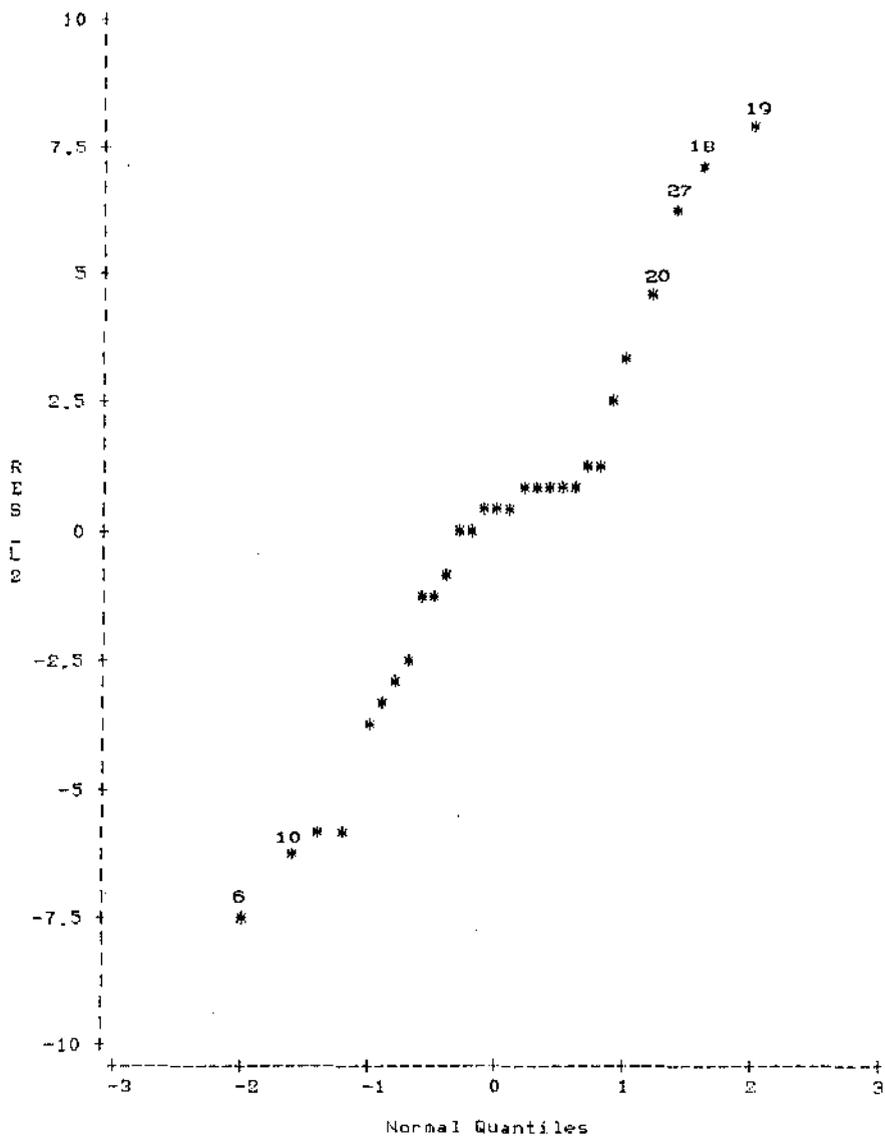


Figura 4.4.4 Gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0.1$) ajustado por mínimos quadrados.

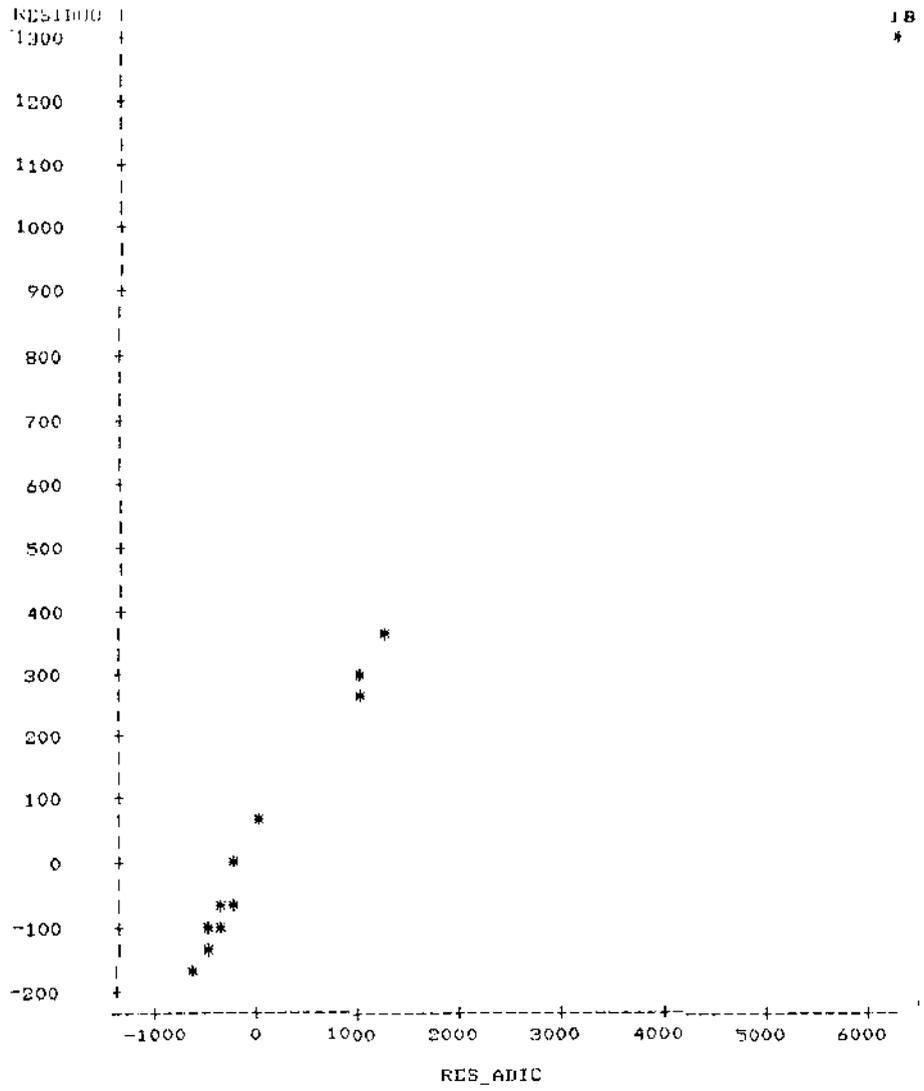


Figura 4.4.5 Gráfico de variável adicionada.

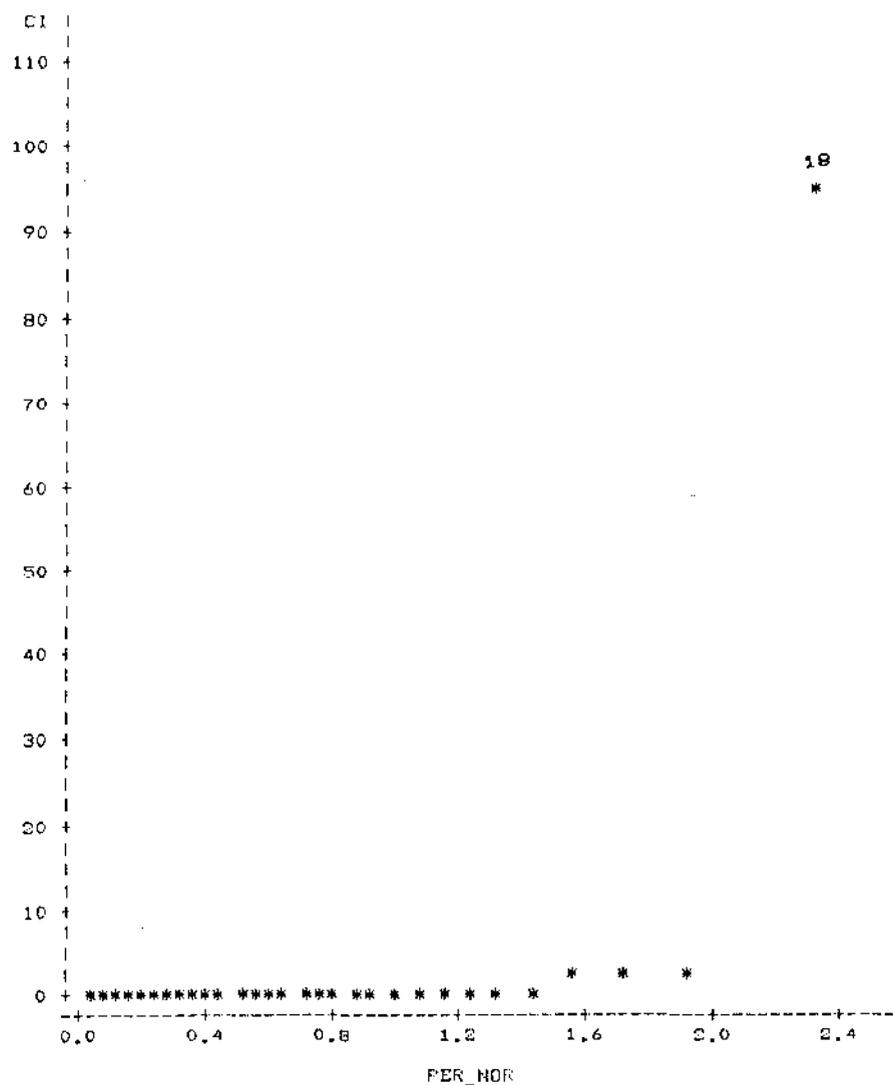


Figura 4.4.6 Gráfico semi-normal para estatística de Cook modificada proposta por Atkinson.

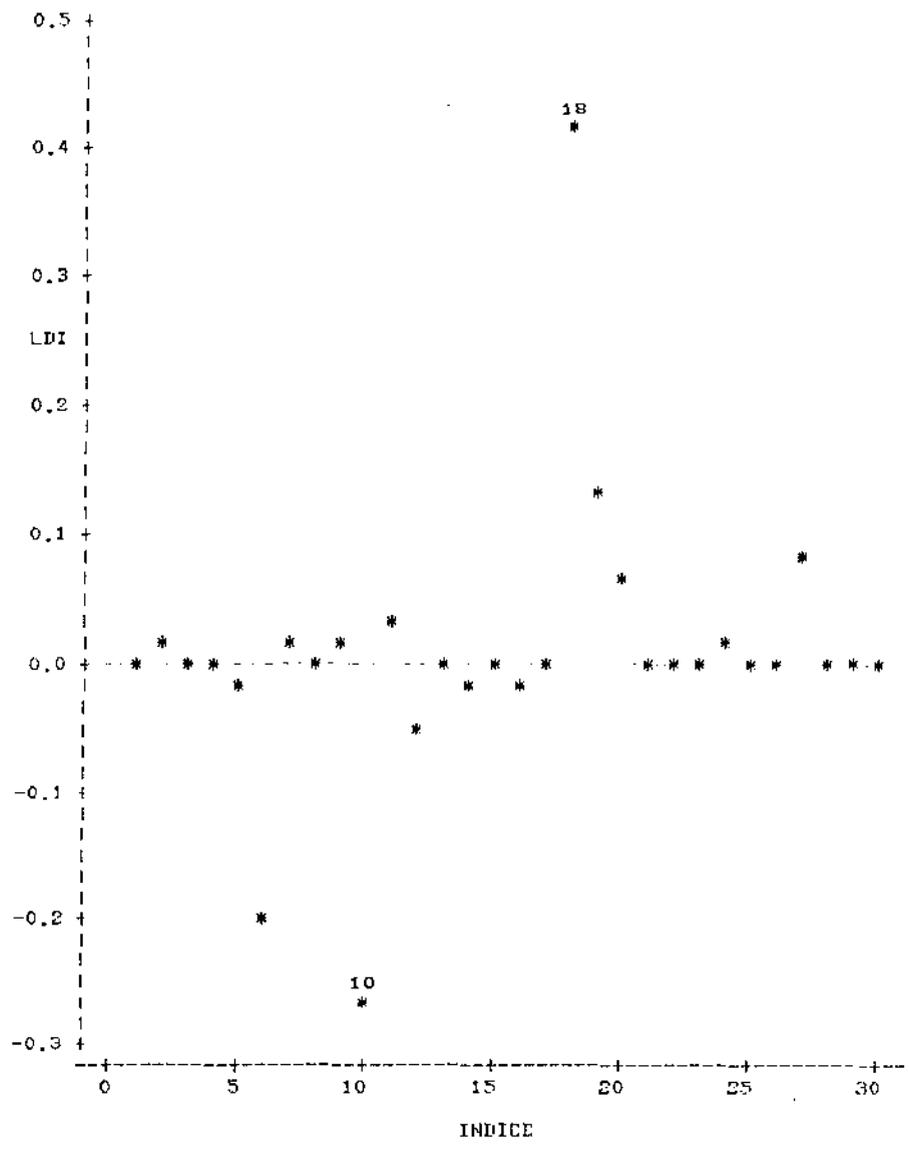


Figura 4.4.7 Gráfico para estatística de Cook e Wang.

Para investigarmos a influência de observações

individuais na determinação da necessidade de uma transformação da variável resposta, sob análise de mínimos desvios absolutos, construímos o gráfico de Parker, mostrado na figura 4.4.8. Este gráfico ilustra a observação 18 como a mais influente em determinar a necessidade de uma transformação, seguida pelas observações 19, 20 e 27. Este gráfico não indica nenhuma observação influente em rejeitar a necessidade de uma transformação. Através do gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=0$) ajustado por mínimos desvios absolutos, figura 4.4.3, observamos que as observações identificadas pelo método de Parker se acomodam perfeitamente à escala transformada.

Nesse problema, os resultados obtidos usando a análise de mínimos desvios absolutos são mais satisfatórios que aqueles obtidos usando análise de mínimos quadrados. Através da análise de mínimos desvios absolutos e uma distribuição de erros dupla-exponencial, todas as 30 observações são adequadamente modeladas por um modelo transformado ($\lambda=0$) de primeira ordem. Usando regressão de mínimos quadrados e uma distribuição de erros normal, as observações não são razoavelmente modeladas, como ilustra a figura 4.4.4. Além disso, sob mínimos quadrados, as observações 18, 19, 20 e 27 são consideradas aberrantes.

4.5 PROBLEMA 4

Neste problema utilizamos um conjunto de dados, retirado do livro de Gunst e Mason [(1980) - p.358], que representa algumas características demográficas de 49 países. Os dados, mostrados na tabela 4.5.1, representam as seguintes características:

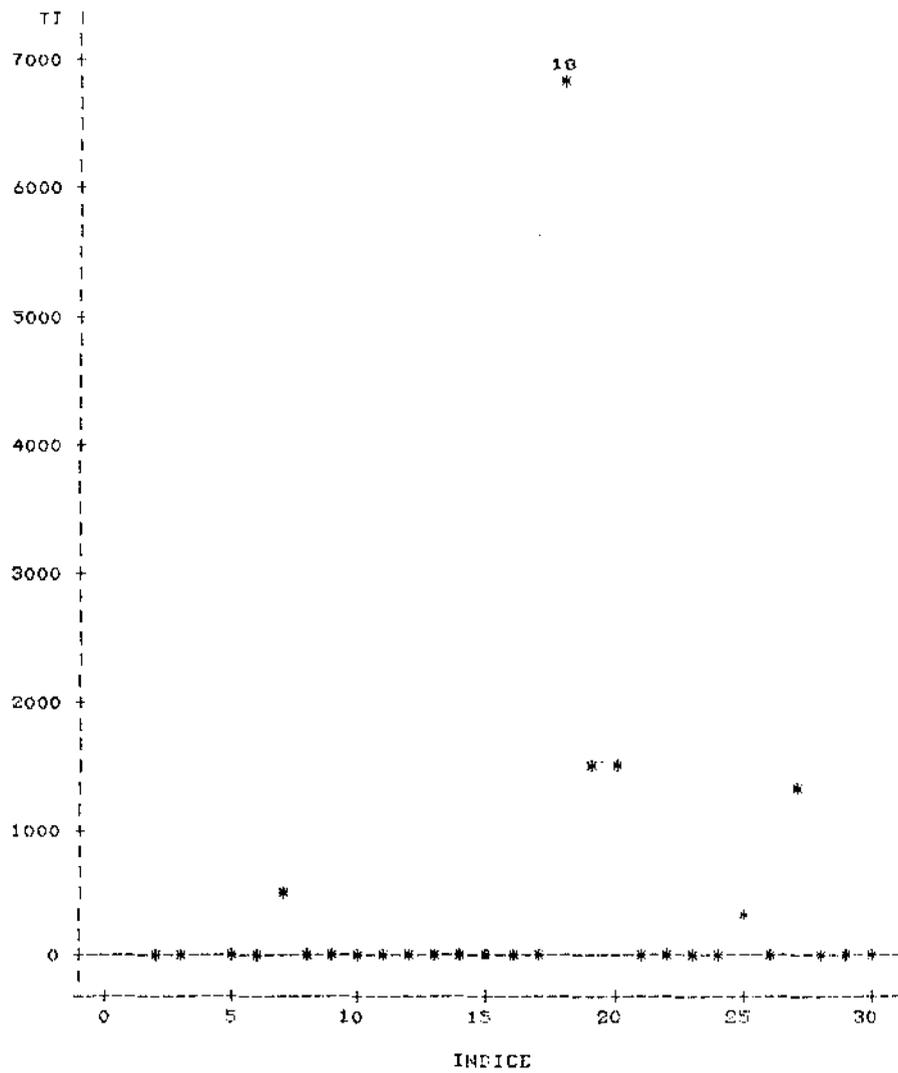


Figura 4.4.8 Gráfico de diagnóstico de Parker para observações não básicas.

Y = Produto Interno Bruto (calculado em dólares - 1957)

X₁ = Mortalidade Infantil por 1000 crianças nascidas vivas

X₂ = Número de Habitantes para cada médico

X₃ = População por km²

X₄ = População por 1000 hectares de agricultura no país

X₅ = Percentagem de alfabetizados maiores de 15 anos

X₆ = Estudantes matriculados no nível superior para cada 100000 habitantes

PAÍSES	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1 Austrália	1316	19.5	860	1	21	98.5	856
2 Áustria	670	37.5	695	84	1720	98.5	546
3 Barbados	200	60.4	3000	548	7121	91.1	24
4 Bélgica	1196	35.4	819	301	5257	96.7	536
5 Guiana Britânica	235	67.1	3900	3	192	74.0	27
6 Bulgária	365	45.1	740	72	1380	85.0	456
7 Canadá	1947	27.3	900	2	257	97.5	645
8 Chile	379	127.9	1700	11	1164	80.1	257
9 Costa Rica	357	78.9	2600	24	948	79.4	326
10 Chipre	467	29.9	1400	62	1042	60.5	78
11 Tchecoslováquia	680	31.0	620	108	1821	97.5	398
12 Dinamarca	1057	23.7	830	107	1434	98.5	570
13 El Salvador	219	76.3	5400	127	1497	39.4	89
14 Finlândia	794	21.0	1600	13	1512	98.5	529

15 França	943	27.4	1014	83	1288	96.4	667
16 Guatemala	189	91.9	6400	36	1365	29.4	135
17 Hong Kong	272	41.5	3300	3092	98143	57.5	176
18 Hungria	490	47.6	650	108	1370	97.5	258
19 Islândia	572	22.4	840	2	79	98.5	445
20 Índia	73	225.0	5200	138	2279	19.3	220
21 Irlanda	550	30.5	1000	40	598	98.5	362
22 Itália	516	48.7	746	164	2323	87.5	362
23 Jamaica	316	58.7	4300	143	3410	77.0	42
24 Japão	306	37.7	930	254	7563	98.0	750
25 Luxemburgo	1388	31.5	910	123	2286	96.5	36
26 Malásia	356	68.9	6400	54	2980	38.4	475
27 Malta	377	38.3	980	1041	8050	57.6	142
28 Maurícios	225	69.5	4500	352	4711	51.8	14
29 México	262	77.7	1700	18	296	50.0	258
30 Holanda	836	16.5	900	346	4855	98.5	923
31 Nova Zelândia	1310	22.8	700	9	170	98.5	839
32 Nicarágua	160	71.7	2800	10	824	38.4	110
33 Noruega	1130	20.2	946	11	3420	98.5	258
34 Panamá	329	54.8	3200	15	838	67.5	371
35 Polónia	475	74.7	1100	96	1411	95.0	351
36 Portugal	224	77.5	1394	100	1087	55.9	272
37 Porto Rico	563	52.4	2200	271	4030	81.0	1192
38 Romênia	360	75.7	788	78	1248	89.0	226
39 Singapura	400	32.3	2400	2904	108214	50.0	437
40 Espanha	293	43.5	1000	61	1347	87.0	258
41 Suécia	1380	16.6	1089	17	1705	98.5	401
42 Suíça	1428	21.1	765	133	2320	98.5	398
43 Tailândia	161	30.5	1500	305	10446	54.0	329
44 Trindade	423	45.4	2300	168	4383	73.8	61
45 Reino Unido	1189	24.1	935	217	2677	98.5	460
46 Estados Unidos	2577	26.4	780	20	399	98.0	1983

47 USSR	600	35.0	578	10	339	95.0	539
48 Alem. Ocidental	927	33.8	798	217	3631	98.5	528
49 Iugoslávia	265	100.0	1637	73	1215	77.0	524

Tabela 4.5.1 Lista de algumas características demográficas de 49 países.

Gunst e Mason(1980) optaram por retirar as observações 17 e 39, referentes a Hong Kong e Singapura, antes de ajustar qualquer modelo, pelo fato de terem-nas identificado, através de análise exploratória de dados, como observações aberrantes e por elas serem cidades e não países. Em seu trabalho, eles realizaram todas as análises sob minimização da soma dos erros quadráticos.

Gunst e Mason, através de suas análises sobre as 47 observações restantes, observaram a necessidade de uma transformação na variável resposta e resolveram, então, utilizar as escalas log e inversa para transformar a variável resposta. Para ambos os modelos transformados ($\lambda=0$ e $\lambda=-1$) ajustados, concluíram, após observar os gráficos de resíduos versus variáveis explanatórias e gráficos probabilísticos normais para os resíduos, que as transformações melhoraram substancialmente o ajuste. O gráfico probabilístico normal para os resíduos do modelo ajustado sob escala inversa ($\lambda=-1$) é ilustrado na figura 4.5.1.

Eles escolheram, então, o modelo transformado pela escala inversa , cujo ajuste é dado por:

$$\begin{aligned} \hat{Y}^{-1} = & 0.004103 + 0.0000380 X_1 - 0.0000001 X_2 \\ & - 0.0000021 X_3 + 0.0000003 X_4 - 0.0000420 X_5 \\ & - 0.0000004 X_6 \end{aligned}$$

Nas análises seguintes realizadas sobre os resíduos deste modelo transformado, eles identificaram as observações 8 e 20, referentes ao Chile e a Índia, como observações aberrantes. Desse modo, dos 49 países iniciais do conjunto de dados, eles decidiram excluir quatro: Chile, Hong Kong, Índia e Singapura.

Através do procedimento de Box e Cox, sob mínimos quadrados, verificamos que o intervalo de confiança de 95% para λ , para o conjunto de dados completo, abrange os valores entre -0.5 e 0. Observamos, portanto, que dentre as duas transformações consideradas por Gunst e Mason(1980), somente a transformação log pertence ao intervalo de confiança para λ .

O modelo ajustado sob escala log, para o conjunto de dados completo, é dado por :

$$\begin{aligned} \hat{\log(Y)} = & 5.16032 - 0.007595 X_1 + 0.000016 X_2 \\ & - 0.000264 X_3 + 0.000000 X_4 + 0.014680 X_5 \\ & + 0.00061 X_6 \end{aligned}$$

O gráfico probabilístico normal para os resíduos deste modelo transformado ($\lambda=0$) é mostrado na figura 4.5.2. Comparando as figuras 4.5.1 e 4.5.2, observamos que, os resíduos do modelo transformado pela escala log evidenciam uma melhor acomodação à suposição de normalidade que os resíduos resultantes do modelo transformado pela escala inversa escolhido por Gunst e Mason(1980).

O gráfico para estatística de Cook modificada, mostrado na figura 4.5.3, indica a observação 46 como a observação mais influente na estimação do parâmetro de transformação. Após a exclusão desta observação, a estatística de Cook modificada aponta as observações 7, 25 e 24 como as mais influentes. Análises realizadas sem estas observações não alteraram, entretanto, a permanência da

transformação log no conjunto de dados, assim como não proporcionaram mudanças significativas nas estimativas paramétricas e nos gráficos probabilísticos normais.

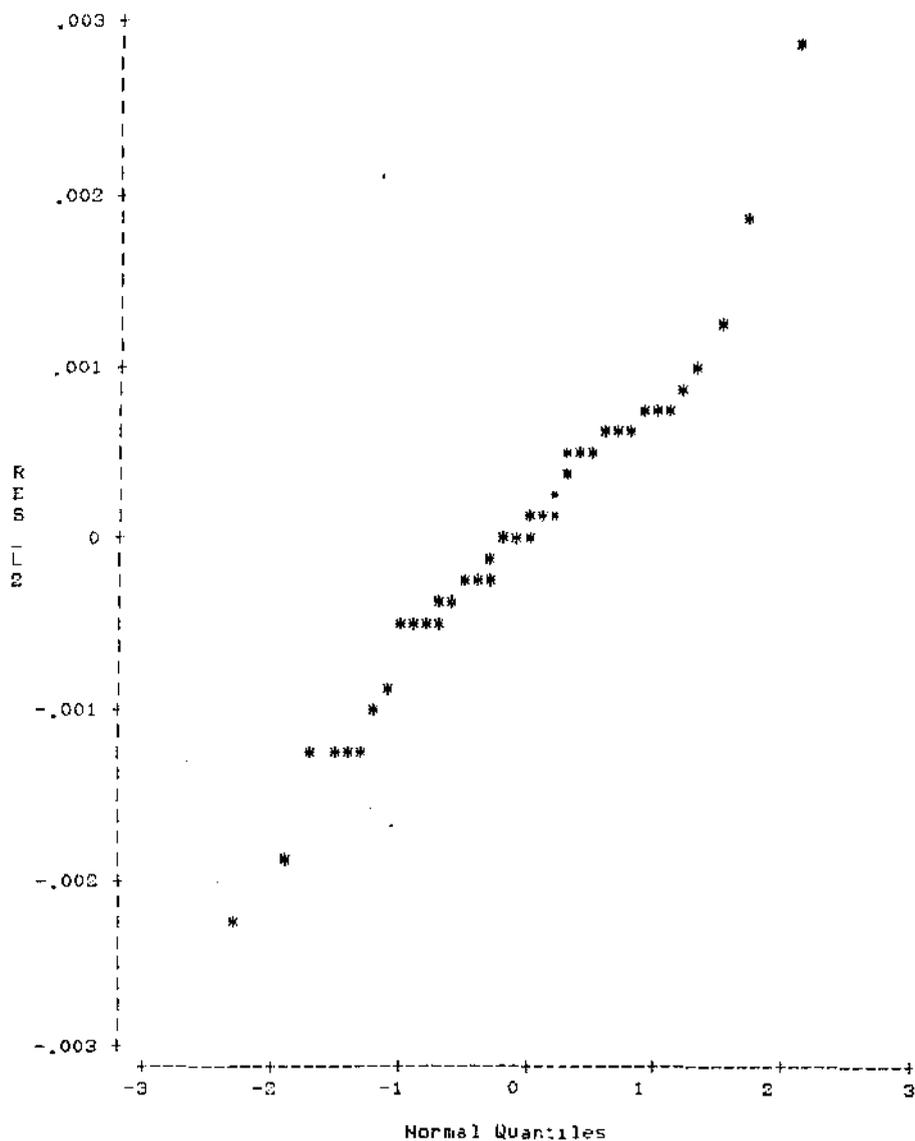


Figura 4.5.1 Gráfico probabilístico normal para os resíduos do modelo escolhido por Gunst e Mason(1980).

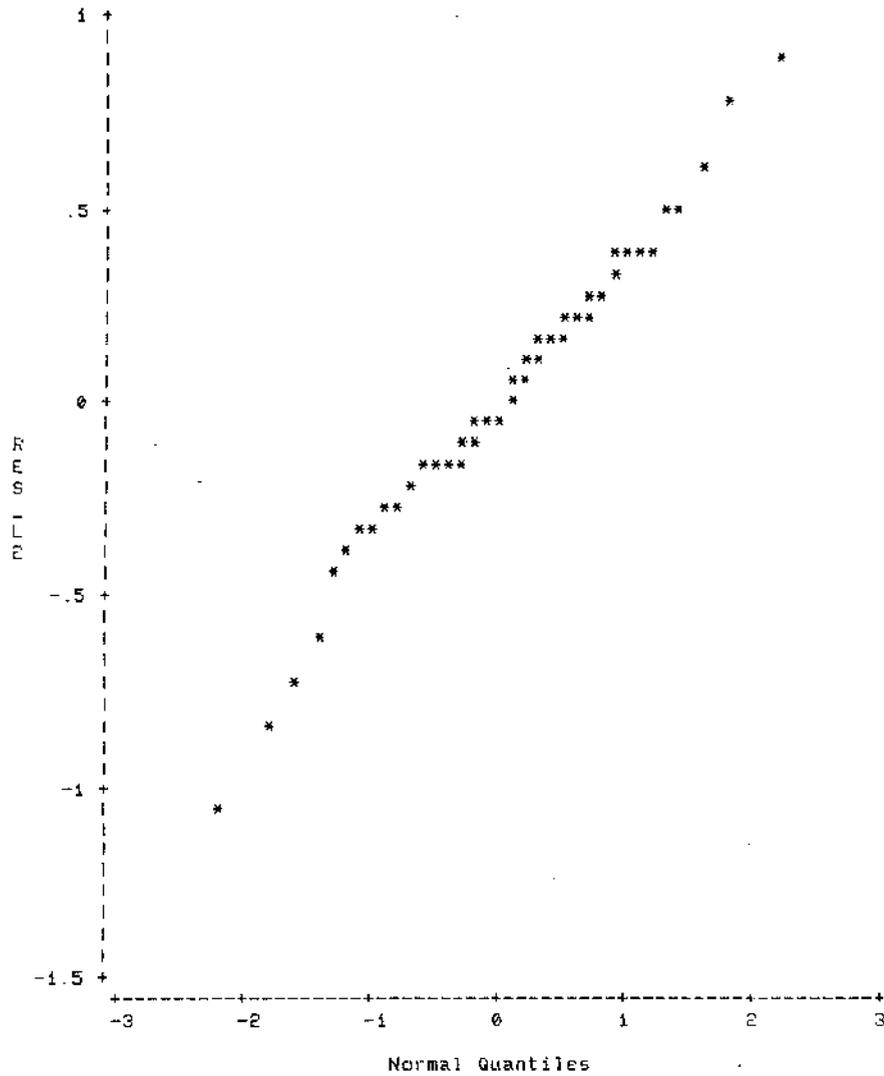


Figura 4.5.2 Gráfico probabilístico normal para os resíduos do modelo transformado ($\lambda=0$) ajustado por mínimos quadrados.

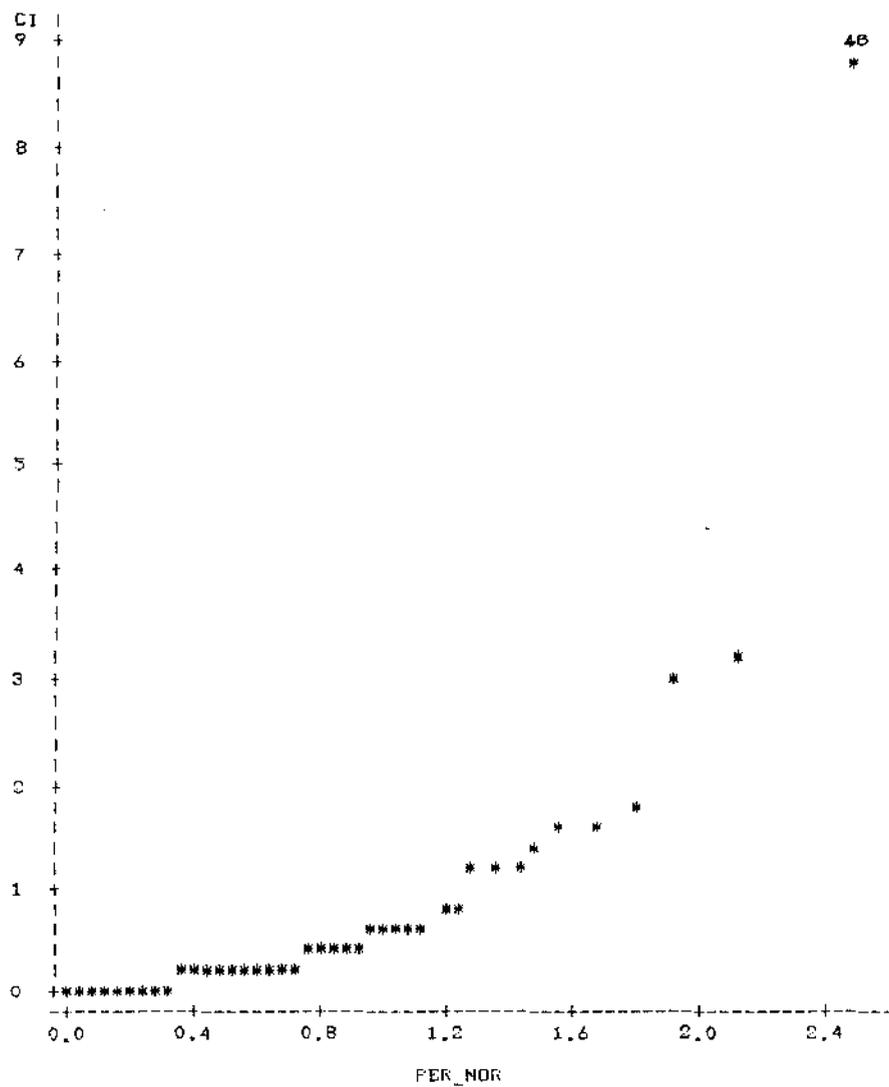


Figura 4.5.3 Gráfico semi-normal para estatística de Cook modificada proposta por Atkinson.

O gráfico de variável adicionada, ilustrado na figura 4.5.4, não indica nenhuma observação como aberrante, além de

evidenciar que a necessidade de uma transformação está uniformemente espalhada através do conjunto de dados. A estatística de Cook e Wang, ilustrada na figura 4.5.5, aponta as observações 43 e 8 como as observações mais influentes na determinação do valor de $\hat{\lambda}$.

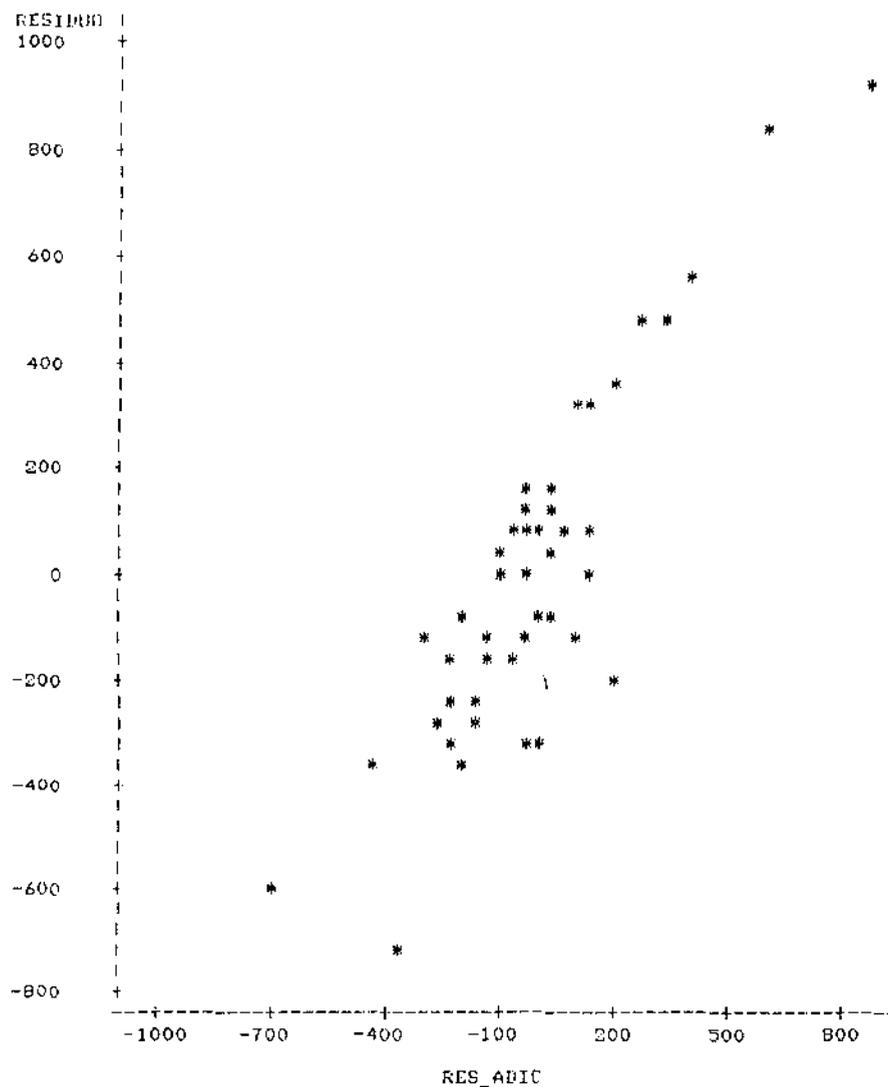


Figura 4.5.4 Gráfico de variável adicionada.

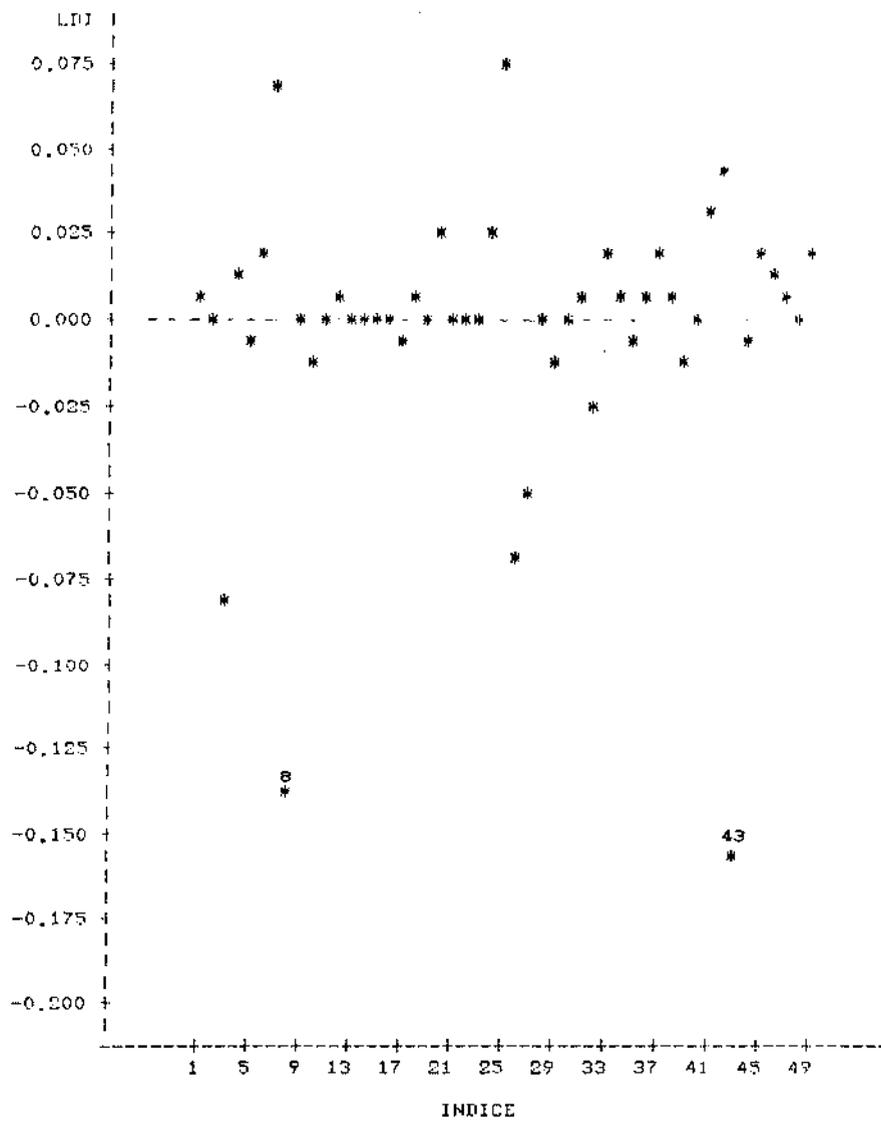


Figura 4.5.5 Gráfico para estatística de Cook e Wang.

Para análise de mínimos desvios absolutos, ajustamos, inicialmente, o modelo de primeira ordem completo. A regressão resultante é dada por:

$$\hat{Y} = 134.837 - 3.961928 X_1 + 0.0333461 X_2 + 0.0939167 X_3 - 0.004561 X_4 + 3.0631897 X_5 + 0.8713822 X_6$$

Soma dos Resíduos Absolutos = 11152.953

O gráfico probabilístico exponencial para os resíduos absolutos, mostrado na figura 4.5.6, demonstra que embora os resíduos estejam bastante próximos de uma linha reta, eles podem ainda ser considerados grandes e uma transformação da variável resposta pode, então, valer a pena.

O procedimento de Box e Cox, sob mínimos desvios absolutos, identifica como ótima a transformação $\hat{\lambda} = -0.5$ e fornece o intervalo de confiança de 95% para λ , que abrange os valores entre -0.6 e -0.3. Escolhemos, então, ajustar o modelo sob a escala inversa da raiz quadrada ($\lambda = -0.5$).

O modelo transformado ajustado por mínimos desvios absolutos é, portanto, dado por:

$$\hat{Y}^{(-0.5)} = 1.89879 - 0.000613 X_1 - 0.000000 X_2 - 0.000032 X_3 + 0.000000 X_4 + 0.000456 X_5 + 0.000017 X_6$$

Soma dos Resíduos Absolutos = 0.55

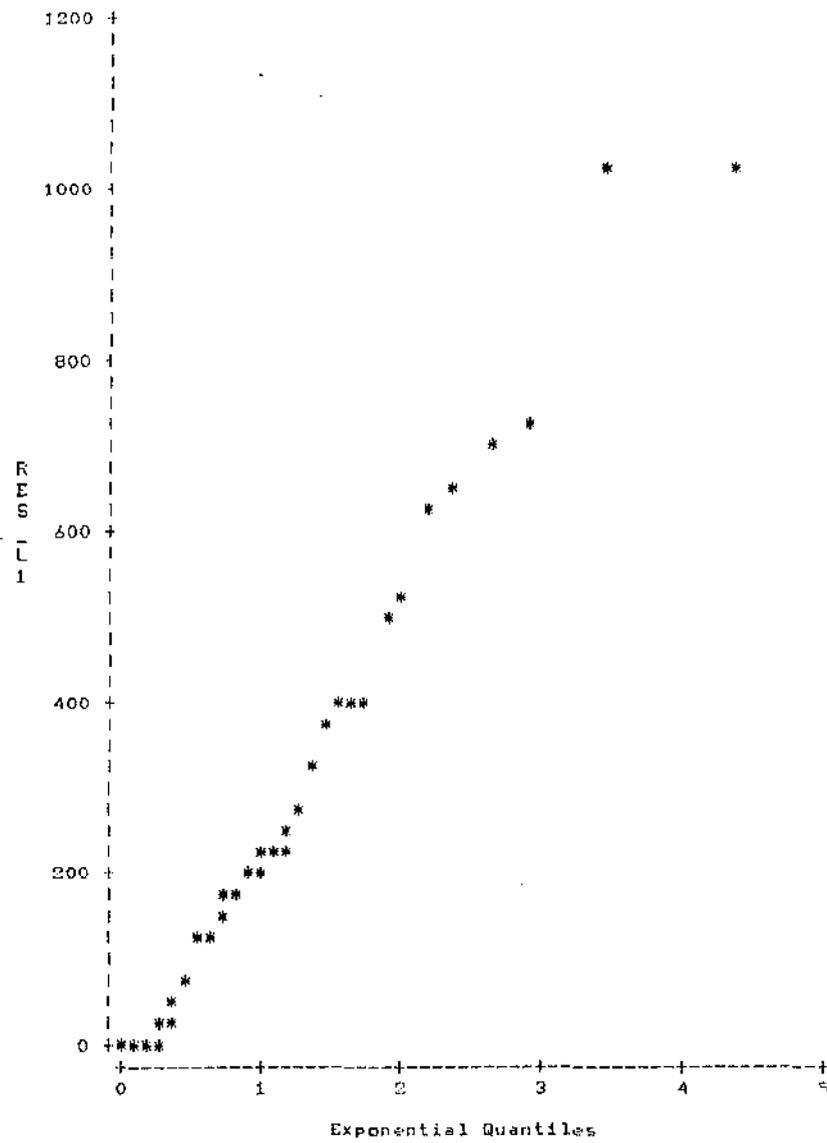


Figura 4.5.6 Gráfico probabilístico exponencial para os resíduos absolutos do modelo não transformado ajustado por mínimos desvios absolutos.

A figura 4.5.7 ilustra o gráfico probabilístico exponencial para os resíduos absolutos deste modelo transformado ajustado por mínimos desvios absolutos. Através desta figura, podemos observar uma tendência linear caracterizando todos os pontos, com exceção da observação 43 situada na extremidade superior do gráfico que destaca-se ligeiramente das demais observações. Deste modo, a figura 4.5.7, leva-nos a concluir que os resíduos do modelo transformado ($\lambda=-0.5$) possuem distribuição dupla-exponencial.

Para avaliarmos a influência de observações individuais na determinação da necessidade de uma transformação, utilizamos o gráfico de diagnóstico de Parker, ilustrado na figura 4.5.8. Este gráfico aponta as observações 46 e 7 como as mais influentes em determinar a necessidade de uma transformação. Esta mesma influência é observada, também, em quase todas as outras observações, entretanto, com menos significância.

Se excluirmos as observações 46 e 7 do conjunto de dados, continuaremos, entretanto, com o mesmo valor ótimo e intervalo de confiança de 95% para λ , assim como não obteremos mudanças significativas nas estimativas paramétricas ou no gráfico probabilístico exponencial para os resíduos absolutos.

Comparando os gráficos probabilísticos para os modelos transformados por análise de mínimos quadrados e por análise de mínimos desvios absolutos, verificamos que as observações parecem ser mais adequadamente modeladas pelo modelo linear de primeira ordem na escala -0.5 ajustado por regressão de mínimos desvios absolutos, sob uma distribuição de erros dupla-exponencial.

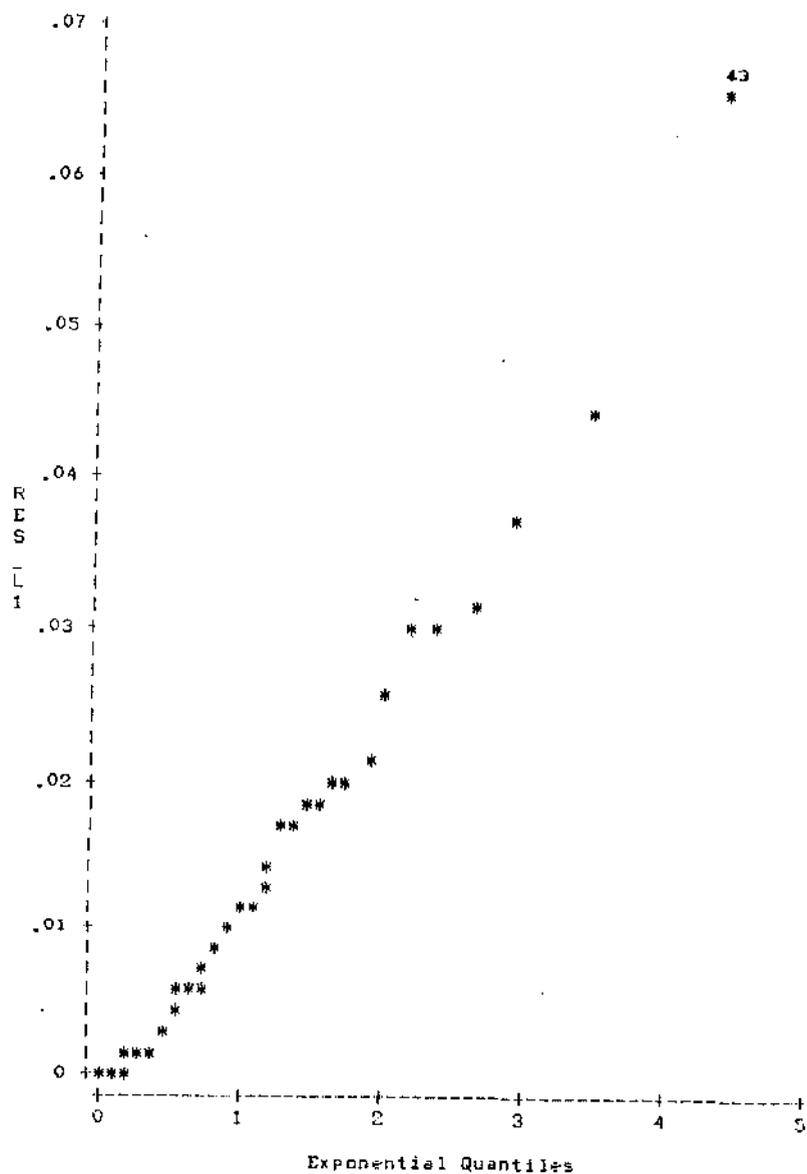


Figura 4.5.7 Gráfico probabilístico exponencial para os resíduos absolutos do modelo transformado ($\lambda=-0.5$) ajustado por mínimos desvios absolutos.

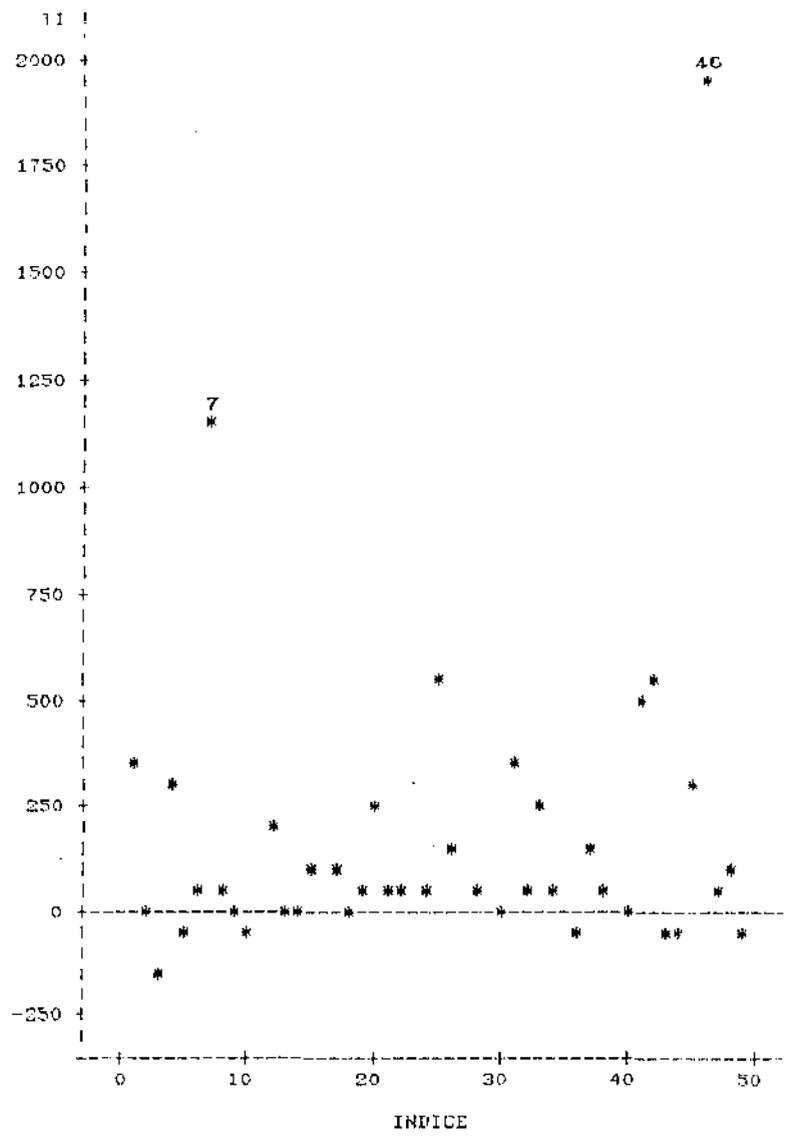


Figura 4.5.8 Gráfico de diagnóstico de Parker para observações não básicas.

4.6 COMENTÁRIOS

Para os três problemas gerados o procedimento de Box e Cox, sob mínimos desvios absolutos usando distribuição de erros dupla-exponencial, identifica a transformação log pré-estabelecida como uma transformação apropriada. Usando o método dos mínimos quadrados e uma distribuição de erros normal, o referido método consegue identificar a transformação log pré-estabelecida somente no primeiro problema, onde os erros aleatórios possuem distribuição normal. Desse modo, verificamos que parece que a transformação de Box e Cox é sensível à não normalidade (caudas mais longas) quando usamos mínimos quadrados, sob a hipótese de normalidade; entretanto, quando usamos mínimos desvios absolutos, sob a hipótese de erros dupla-exponenciais, ela é resistente a outras distribuições, podendo modelá-las adequadamente sob esta última hipótese.

Os métodos de diagnósticos de influência, sob mínimos quadrados, evidenciaram com sucesso observações aberrantes e pontos influentes. O método de Cook e Wang e a estatística de Cook modificada proposta por Atkinson(1982) parecem apontar com mais clareza os pontos influentes que o gráfico de variável adicionada, pelo fato de levarem em consideração a estrutura de ruptura do problema, como já argumentado por Cook e Wang(1983) (seção 3.2.3), ao passo que o gráfico de variável adicionada consegue apontar com precisão as observações aberrantes. O método de diagnóstico de influência, sob mínimos desvios absolutos, detecta os valores influentes em determinar a necessidade de uma transformação, entretanto, verificamos que essas observações não são valores aberrantes.

Para o quarto problema, as conclusões obtidas usando as análises de mínimos quadrados e mínimos desvios absolutos, são bastante diferentes. Usando regressão de mínimos desvios absolutos,

todas as observações parecem estar adequadamente modeladas por um modelo de primeira ordem na escala -0.5 . Para análise de mínimos quadrados, comparativamente ao modelo acima descrito, o modelo transformado obtido não parece proporcionar tão bom ajuste, com observações 7, 24, 25 e 46 identificadas como aberrantes.

APÊNDICE 1

Todos os programas aqui catalogados aplicam-se a conjuntos de dados com uma variável explanatória (X_1) e uma variável resposta (Y) e será denominado DADOS.DAT. Podemos, entretanto, utilizar tais programas para conjuntos de dados maiores, através de ligeiras modificações.

Os programas catalogados são:

PROGRAMA 1

Programa para ajustar um modelo linear de primeira ordem, sob mínimos quadrados e mínimos desvios absolutos. Ele constrói um arquivo SAS permanente para os resíduos dos modelos ajustados, cujo nome é REG.SSD. Este arquivo será utilizado como o conjunto de dados de entrada no programa (PROGRAMA 7) que constrói os gráficos probabilísticos normal e exponencial para os resíduos do modelo ajustado por mínimos quadrados e para os resíduos absolutos ajustados por mínimos desvios absolutos, respectivamente.

```
LIBNAME RESIDUO 'A:';
```

```
/* PROGRAMA P/ AJUSTAR MODELO POR NORMAS :L1 E L2 */
```

```
OPTIONS LS = 70;
```

```
OPTIONS PS = 60;
```

```

DATA LOSS;
  INFILE 'A:DADOS.DAT';
  INPUT X1 Y;
RUN;
PROC PRINT;
RUN;
proc iml;
USE LOSS;
  READ ALL;
X = X1;
NOBS = NROW(X);
UNIT = J(NOBS,1,1);
X = UNIT || X;
START L1;
  bound=1.0e10;
  coef = X';
  m = nrow(coef);
  n = ncol(coef);
  r = repeat(0,m+2,1);
  l = repeat(-0.5,1,n) || repeat(0,1,m) || -bound || -bound ;
  u = repeat( 0.5,1,n) || repeat(.,1,m) || { . . } ;
  a = (          y'          || repeat(0,1,m) || { -1 0 }          ) //
      ( repeat(0,1,n) || repeat(-1,1,m) || { 0 -1 }          ) //
      (          coef          || I(m)          || repeat(0,m,2)          ) ;
  basis = n+m+2 - (0:n+m+1);
  call lp(rc,p,d,a,r,,u,l,basis);
  l = repeat(-0.5,1,n) || repeat(0,1,m) || -bound || {0} ;
  u = repeat( 0.5,1,n) || repeat(0,1,m) || { . 0 } ;
  call lp(rc,p,d,a,r,n+m+1,u,l,basis);
  b=d[3:m+2];
  PREDL1 = X*b;
  RL1 = y - PREDL1;
  wsum = 2*sum( choose(RL1<0 , -0.5*RL1 , 0.5*RL1) );
  finish;
RUN L1;

```

```

BO = B(:1, :);
B1 = B(:2, :);
RESET NONAME;
PRINT,, 'ANALISE DE REGRESSAO L1';
PRINT,, 'MODELO DE REGRESSAO: Y = BO + B1X1 ';
PRINT,, 'NUMERO DE OBSERVACOES: ' NOBS;
PRINT,, 'ESTIMADORES DE MAX. VEROS. DOS COEFICIENTES: ';
PRINT, ' BO =' BO ' B1 =' B1 ;
PRINT,, 'SOMA DOS DESVIOS ABSOLUTOS = ' WSUM;
BETA = INV(X**X)*X**Y;
BOL2 = BETA(:1, :);
BIL2 = BETA(:2, :);
PREDL2 = X*BETA;
RL2 = Y - PREDL2;
SQ = RL2**RL2;
PRINT,, 'ANALISE DE REGRESSAO L2';
PRINT,, 'MODELO DE REGRESSAO: Y = BO + B1X1 ';
PRINT,, 'NUMERO DE OBSERVACOES: ' NOBS;
PRINT,, 'ESTIMADORES DE MAX. VEROS. DOS COEFICIENTES: ';
PRINT, ' BO =' BOL2;
PRINT, ' B1 =' BIL2;
PRINT,, 'SOMA DE QUADRADOS DOS RESIDUOS = ' SQ;
SAIDA = PREDL1 ||RL1||PREDL2||RL2||X1;
NOME = {'YPRED_L1' 'RES_L1' 'YPRED_L2' 'RES_L2' 'X1'};
CREATE AJUSTE FROM SAIDA (:COLNAME = NOME!);
APPEND FROM SAIDA;
QUIT;
PROC PRINT DATA = AJUSTE;
VAR YPRED_L1 RES_L1;
TITLE'VALORES PREDITOS E RESIDUOS POR REG. L1';
RUN;
PROC PLOT ;
PLOT RES_L1*YPRED_L1/VREF=0;
PLOT RES_L1*X1/VREF=0;
RUN;

```

```

PROC PRINT DATA = AJUSTE;
  VAR YPRED_L2 RES_L2;
TITLE 'VALORES PREDITOS E RESIDUOS POR REG. L2';
  RUN;
PROC PLOT ;
  PLOT RES_L2*YPRED_L2/VREF=0;
  PLOT RES_L2*X1/VREF=0;
RUN;
PROC UNIVARIATE PLOT;
  VAR RES_L1 ;
TITLE 'ANALISE DOS RESIDUOS SOB REGRESSAO L1';
RUN;
PROC UNIVARIATE PLOT;
  VAR RES_L2;
TITLE 'ANALISE DOS RESIDUOS SOB REGRESSAO L2';
RUN;
DATA RESIDUO.REG;
  SET AJUSTE;
  RES_L1 = ABS(RES_L1);
RUN;

```

PROGRAMA 2

Programa para calcular o estimador de máxima verossimilhança do parâmetro de transformação, λ , um I.C. _{λ} (95%) e o gráfico de $L_{\max}(\lambda)$ versus λ , sob mínimos desvios absolutos. Ele oferece, também, as estimativas paramétricas do modelo linear de primeira ordem sob a escala transformada ($\hat{\lambda}$) para a variável resposta, assim como os valores ajustados para a variável resposta transformada e os resíduos. Os resíduos desse modelo transformado

ajustado são guardados em um arquivo SAS permanente cujo nome é RTR11.SSD. Este programa serve também para outras expressões da verossimilhança se mudarmos a linha correspondente no programa.

```
LIBNAME RESIDUO 'A:';
OPTIONS LS=70;
OPTIONS PS=60;
DATA LOSS;
  INFILE 'A:DADOS.DAT';
  INPUT X1 Y ;
RUN;
PROC IML;
USE LOSS;
  READ ALL;
  X=X1;          /* VARIAVEIS INDEPENDENTES */
NOBS=NROW(X);
UNIT=J(NOBS,1,1);
X=UNIT::X;
EXPO=1/NOBS;
A=Y##EXPO;
MG=A(:#,1);    /* MEDIA GEOMETRICA DE Y */
LY=LOG(Y);
FREE UNIT A ;
START L1(ZK,X,B1,LMAXK,Q);
  BOUND = 1.0E10;
  COEF = X';
  M = NROW(COEF);
  N = NCOL(COEF);
  R = REPEAT(0,M+2,1);
  L = REPEAT(Q-1,1,N)::REPEAT(0,1,M)::-BOUND::-BOUND;
  U = REPEAT(Q,1,N)::REPEAT(.,1,M)::{ . . };
  A = ( ZK'          :: REPEAT(0,1,M) :: { -1 0 }      ) //
      ( REPEAT(0,1,N):: REPEAT(-1,1,M):: { 0 -1 }      ) //
      ( COEF        :: I(M)          :: REPEAT(0,M,2) );
  BASIS = N+M+2- (0:N+M+1);
  CALL LP(RC,P,D,A,R,,U,L,BASIS);
  L = REPEAT(Q-1,1,N) :: REPEAT(0,1,M) :: -BOUND :: { 0 };
  U = REPEAT(Q,1,N)   :: REPEAT(0,1,M) :: { . 0 };
  CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
  B1 = D[3:M+2];
  ERROR = ZK - X*B1;
  WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
  LMAXK = -N*LOG(2*WSUM);
FINISH;
START EMV1;
DO L =-1 TO 0 BY 0.05;
  IF L < 0 THEN ZK= (Y##L-1)/(L*MG##(L-1));
  ELSE ZK=LY*MG;
```

```

    RUN LI(ZK,X,B,LMAX,0.5);
    SAI = SAI/(L::LMAX);
END;
FINISH;
RUN EMV1;
START EMV2;
DO K= 0 TO 1.10 BY 0.05;
    IF K > 0 THEN ZK=(Y##K-1)/(K*MG##(K-1));
    RUN LI(ZK,X,B,LMAX,0.5);
    SAI = SAI/(K::LMAX);
END;
FINISH;
RUN EMV2;
FREE L U A ERROR R;
QQIG=3.84;
A=QQIG/2;
T= SAI(:<:,2:);
LCHAPEU= SAI(:T,2:);
LAMMAX= SAI(:T,1:);
RESET NONAME;
PRINT,, 'ESTIMADOR DE MAX. VEROSSIM. DE LAMBDA',, LAMMAX;
PRINT,, 'VALOR DO LOG DA VEROSSIM. P/ EMV DE LAMBDA',, LCHAPEU;
LIMITE = LCHAPEU -A;
DIF = LIMITE-SAI(: ,2:);
POS=LOC(DIF<0);
IC = SAI(:POS,1:);
PRINT 'INTERV. DE CONF. DE 95% P/ LAMBDA CONTEM OS VALORES:',,IC;
NUM = NROW(SAI);
ZERO = J(NUM,1,0);
SAI = SAI::ZERO;
SAI(:POS,3:) = LIMITE;
START FINAL;
    IF LAMMAX = 0 THEN YK = LY;
        ELSE YK = (Y##LAMMAX-1)/LAMMAX;
FINISH;
RUN FINAL;
RUN LI(YK,X,B,LMAX,0.5);
BOL1 = B(:1, :);
BIL1 = B(:2, :);
PRINT,, 'ESTIMADORES DE MAX. VEROSSIM. DOS COEFICIENTES SOB LAM MAX: ';
PRINT ' BO =' BOL1 ' BI =' BIL1 ;
PRED = X*B;
RES = YK - PRED;
VET = YK::PRED::RES;
NOME=('LAMBDA' 'LMAX' 'INTER');
CREATE TRANSF FROM SAI (:COLNAME=NOME);
APPEND FROM SAI;
FREE SAI;
NOM={ 'RESP_TRN' 'PREDITO' 'RES'};
CREATE NOVO FROM VET (:COLNAME=NOM);
APPEND FROM VET;
FREE VET ;

```

```

QUIT;
PROC PRINT DATA=TRANSF; VAR LAMBDA LMAX;
RUN;
DATA NULO;
  SET TRANSF;
  IF INTER = 0 THEN INTER = '.';
RUN;
PROC PLOT DATA=NULO ;
  PLOT LMAX*LAMBDA='* ' INTER*LAMBDA ='_ ' /OVERLAY;
  TITLE 'VALOR MAX. DE VEROSSIMILHANCA P/ VALORES FIXOS DE LAMBDA';
RUN;
PROC PRINT DATA = NOVO;
TITLE'VALORES DAS: RESP. TRANSF., PREDITO POR LI E RESIDUO';
RUN;
DATA RESIDUO.RTRL1;
  SET NOVO;
  RES = ABS(RES);
RUN;
PROC UNIVARIATE DATA = NOVO PLOT;
  VAR RES;
TITLE'ANALISE DOS RESIDUOS POR REG. LI APOS TRANSFORMACAO';
RUN;

```

PROGRAMA 3

Programa para calcular o diagnóstico de influência de Parker. Ele constrói um gráfico de índice para a estatística de Parker.

```

OPTIONS PS=60;
OPTIONS LS = 70;
DATA LOSS;
  INFILE 'A:DADOS.DAT';
  INPUT X1 Y;
RUN;
PROC IML;

```

```

USE LOSS;
  READ ALL;
  X=X1 ;          /* VARIAVEIS INDEPENDENTES */
NOBS=NROW(X);
UNIT=J(NOBS,1,1);
X=UNIT#X;
EXPO=1/NOBS;
A=Y##EXPO;
MG=A{#,#,1};          /* MEDIA GEOMETRICA DE Y */
LY=LOG(Y);
FREE UNIT A ;
Z1 = Y - 1;
LMG = LOG(MG);
POND = LOG(Y/MG);
YPOND = Y#POND;
D1 = YPOND + LMG - Z1;
START L1;
  BOUND = 1.0E10;
  COEF = X';
  M = NROW(COEF);
  N = NCOL(COEF);
  R = REPEAT(0,M+2,1);
  L = REPEAT(-0.5,1,N)::REPEAT(0,1,M)::-BOUND::-BOUND;
  U = REPEAT(0.5,1,N)::REPEAT(.,1,M)::( . . );
  A = ( Z1'          :: REPEAT(0,1,M) :: { -1 0 } ) //
      ( REPEAT(0,1,N):: REPEAT(-1,1,M):: { 0 -1 } ) //
      ( COEF        :: I(M)          :: REPEAT(0,M,2) );
BASIS = N+M+2- (0:N+M+1);
CALL LP(RC,P,D,A,R,,U,L,BASIS);
  L = REPEAT(-0.5,1,N) :: REPEAT(0,1,M) :: -BOUND :: { 0 };
  U = REPEAT(0.5,1,N)  :: REPEAT(0,1,M) :: { . 0 };
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
B1 = D{3:M+2};
RESIDUO = Z1 - X*B1;
FINISH;

```

```

RUN L1;
FREE L U A R;
POS1 = LOC(RESIDUO <-1.0E-10);
POS2 = LOC(RESIDUO > 1.0E-10);
I = 1:NOBS;
I(: ,POS1:) = 0;
I(: ,POS2:) = 0;
IBAS = LOC(D0);
DBAS = D1(:IBAS, :);
XB = X(:IBAS, :);
CONST = INV(XB)*DBAS;
TIN = - ( D1(:POS1, :) - X(:POS1, :)*CONST );
TIP =      D1(:POS2, :) - X(:POS2, :)*CONST ;
TI = ( POS1' :: TIN ) // ( POS2' :: TIP );
ROTULO = {'INDICE' 'Ti'};
CREATE SAIDA FROM TI (:COLNAME = ROTULO!);
  APPEND FROM TI;
RESULT = Z1 :: X1 :: RESIDUO;
NOME = { 'RESP_TRN' 'X1' 'RESIDUO'};
CREATE FINAL FROM RESULT (:COLNAME = NOME!);
  APPEND FROM RESULT;
QUIT;
PROC PRINT DATA = SAIDA;
TITLE 'INDICE E VALOR DE Ti P/ OBS. NAO BASICAS';RUN;
PROC PLOT DATA = SAIDA;
  PLOT Ti*INDICE = '**' /VREF=0;
TITLE 'GRAFICO DE INFLUENCIA DAS OBS.NAO BASICAS';
RUN;

```

PROGRAMA 4

Programa para calcular o estimador de máxima verossimilhança do parâmetro de transformação, λ , um I.C. _{λ} (95%) e o gráfico de $L_{\max}(\lambda)$ versus λ , sob mínimos quadrados. Ele oferece, também, as estimativas paramétricas do modelo linear de primeira ordem sob a escala transformada ($\hat{\lambda}$) para a variável resposta, assim como os valores ajustados para a variável resposta transformada, os resíduos desse modelo ajustado.

```
OPTIONS LS=70;
OPTIONS PS=60;
DATA SIMUL;
  INFILE 'A:DADOS.DAT';
  INPUT X1 Y;
PROC IML;
USE SIMUL;
  READ ALL;
  X=X1;                * VARIÁVEIS INDEPENDENTE;
NOBS=NROW(X);
UNIT=J(NOBS,1,1);
X=UNIT::X;
EXPO=1/NOBS;
A=Y##EXPO;
B=A(:#,:);            * MEDIA GEOMETRICA DE Y ;
C=I(NOBS);
XIN VX=INV(X**X)*X';
XR=C-X*XIN VX;
LY=LOG(Y);
START EMV1;
DO L =-1 TO 0 BY 0.05;  * VALORES NEG. E NULO DE LAMBDA;
  IF L < 0 THEN ZK= (Y##L-1)/(L*B##(L-1));
  ELSE ZK=LY*B;
```

```

BETA=XINVX*ZK;
SQRK=ZK'*XR*ZK;
LMAXK=-0.5*NOBS*LOG((SQRK/NOBS));
VET=VET/(L::LMAXK);
END;
FINISH;
RUN EMV1;                                *VALORES POSITIVOS DE LAMBDA;
START EMV2;
DO K= 0 TO 1.15 BY 0.05;
  IF K > 0 THEN ZK=(Y##K-1)/(K*B##(K-1));
  BETA=XINVX*ZK;
  SQRK=ZK'*XR*ZK;
  LMAXK=-0.5*NOBS*LOG((SQRK/NOBS));
  VET=VET/(K::LMAXK);
END;
FINISH;
RUN EMV2;
QQ1G=3.84;
A=QQ1G/2;
X= VET(:,2);
LCHAPEU= VET(X,2);
LAMMAX= VET(X,1);
LIMITE = LCHAPEU -A;
DIF = LIMITE-VET(:,2);
POS=LOC(DIF<0);
IC = VET(POS,1);
NUM = NROW(VET);
ZERO = J(NUM,1,0);
VET = VET::ZERO;
VET(POS,3) = LIMITE;
START FINAL;
  IF LAMMAX = 0 THEN YK = LY;
  ELSE YK = (Y##LAMMAX-1)/LAMMAX;
FINISH;
RUN FINAL;

```

```

B = XINVT*YK;
BO = B(1, :);
BI = B(2, :);
SAI = YK-XI;
FREE C XINVT XR;
NOME={'LAMBDA' 'LMAX' 'INTER'};
CREATE RESULT FROM VET (:COLNAME=NOME);
APPEND FROM VET;
FREE VET;
NOM={'RESP_TRN' 'XI' };
CREATE TRANSF FROM SAI (:COLNAME=NOM);
APPEND FROM SAI;
QUIT;
DATA NOVO;
  SET RESULT;
  IF INTER = 0 THEN INTER = '.';
PROC PLOT DATA=NOVO ;
  PLOT LMAX*LAMBDA='*' INTER*LAMBDA ='_ '/OVERLAY;
  TITLE 'VALOR MAX. DE VEROSSIMILHANCA P/ VALORES FIXOS DE LAMBDA';
PROC REG DATA = TRANSF;
  MODEL RESP_TRN = XI;
  OUTPUT OUT= AJUSTE R = RESIDUO P = PREDITO;
PROC PRINT DATA = AJUSTE;
  VAR PREDITO RESIDUO;

```

PROGRAMA 5

Programa para calcular o diagnóstico de influência de Cook e Wang e construir um gráfico de índice para essa estatística.

```

OPTIONS LS=70;
OPTIONS PS=60;
DATA LOSS;
  INFILE 'A:DADOS.DAT';
  INPUT X1 Y; RUN;
PROC IML;
USE LOSS;
  READ ALL;
  X= X1; /* VARIAVEIS INDEPENDENTES*/
NOBS=NROW(X);
UNIT=J(NOBS,1,1);
X=UNIT::X;
EXPO=1/NOBS;
A=Y##EXPO;
B=A(:,1); /* MEDIA GEOMETRICA DE Y */
FREE A UNIT;
C=I(NOBS);
XIN VX=INV(X'*X)*X';
XR=C-X*XIN VX;
FREE C;
LY=LOG(Y);
P = 2;
START EMV1;
DO L =-1 TO 0 BY 0.05;
  IF L < 0 THEN ZK= (Y##L-1)/(L*B##(L-1));
  ELSE ZK=LY*B;
  SQRK=ZK'*XR*ZK;
  LMAXK=-0.5*NOBS*LOG((SQRK/NOBS));
  VET=VET//{(L::LMAXK)};
END;
FINISH;
RUN EMV1;
POS = VET({<>,2});
VET = VET({POS, 1});
START EMV2;

```

```

DO K= 0 TO 1.00 BY 0.05;
  IF K > 0 THEN ZK=(Y##K-1)/(K*B##(K-1));
  SQRK=ZK*XR*ZK;
  LMAXK=-0.5*NOBS*LOG((SQRK/NOBS));
  VET=VET/(K+LMAXK);
END;
FINISH;
RUN EMV2;
T = VET(1,2);
LCHAPEU= VET(T,2);
LAMMAX= VET(T,1);
RESET NONAME;
PRINT,, 'ESTIMADOR DE MAX. VEROSSIM. DE LAMBDA',, LAMMAX;
PRINT,, 'VALOR DO LOG DA VEROSSIM. P/ EMV DE LAMBDA',, LCHAPEU;
FREE VET;
START FINAL;
IF LAMMAX =0 THEN ZK=LY;
      ELSE ZK = (Y##LAMMAX-1)/LAMMAX;
RES = XR*ZK;
FINISH;
RUN FINAL;
YL = Y##LAMMAX;
MGL = B##LAMMAX;
P1 = YL#LY;
P11 = B*LAMMAX*MGL*P1;
P2 = MGL + LAMMAX * MGL * LOG(B);
P21 = B * P2;
P22 = P21 * ( YL - 1 );
P3 = ( LAMMAX * MGL ) ** 2;
DL = ( P11 - P22 ) / P3;
RESADIC = XR * DL;
SD = (RESADIC' * RESADIC) / (NOBS - P);
SZ = (RES' * RES) / (NOBS - P);
H = X * XINVS;
START WANG;

```

```

DO I = 1 TO NOBS;
  CT = 1 - H(I,I);
  RZ = RES(I, ) / SQRT( CT * SZ );
  RD = RESADIC(I, ) / SQRT( CT * SD );
  CALC = (RZ * RD) / (NOBS - P - RD * RD)**2;
  VALOR = LOG(1 + CALC);
  LDI = NOBS * VALOR;
  SAI = SAI // ( I :: LDI ) ;
END;
FINISH;
RUN WANG;
FREE XINVX XR H;
NOME = { 'INDICE ' 'LDI' };
CREATE GRAFICO FROM SAI (:COLNAME = NOME!);
  APPEND FROM SAI;
QUIT;
PROC PRINT DATA = GRAFICO;RUN;
PROC PLOT DATA = GRAFICO;
PLOT LDI * INDICE = '* ';
TITLE 'GRAFICO DE DIAGNOSTICO P/ EST. DE COOK E WANG';
RUN;

```

PROGRAMA 6

Programa para calcular ao teste de hipótese de Atkinson e os gráficos de variável adicionada e da estatística de Cook modificada.

```
/* PROGRAMA P/ CONSTRUCAO DOS GRAFICOS :
```

```
VAR. ADICONADA E ESTATISTICA DE COOK MODIFICADA */
```

```
OPTIONS LS=70;
```

```
OPTIONS PS=60;
```

```
DATA LOSS;
```

```
INFILE 'A:DADOS.DAT';
```

```
INPUT X1 Y;
```

```
PROC IML;
```

```
USE LOSS;
```

```
READ ALL;
```

```
X=X1; * VARIAVEIS INDEPENDENTES ;
```

```
NOBS=NROW(X);
```

```
UNIT=J(NOBS,1,1);
```

```
X=UNIT::X1;
```

```
EXPO=1/NOBS;
```

```
A=Y##EXPO;
```

```
B=A(:,!); * MEDIA GEOMETRICA DE Y;
```

```
C=I(NOBS);
```

```
XIN VX=INV(X'*X)*X';
```

```
XR=C-X*XIN VX;
```

```
LY=LOG(Y);
```

```
LAM = 1;
```

```
* OBSERVACAO: LAMBDA ZERO DEVE SER DIFERENTE DE ZERO;
```

```
RESET NONAME;
```

```
PRINT,, 'VALOR DE LAMBDA SELECIONADO P/ TESTE E MET. DE DIAGNOSTICOS:',,LAM;
```

```
ZK = (Y##LAM-1)/(LAM*B##(LAM-1));
```

```
RES = XR*ZK;
```

```
YL = Y##LAM;
```

```
MGL = B##LAM;
```

```
P1 = YL#LY;
```

```

P1 = B*LAM*MGL*P1;
P2 = MGL + LAM* MGL * LOG(B);
P21 = B * P2;
P22 = P21 * (YL - 1);
P3 = (LAM * MGL) **2;
DL = (P11 - P22)/P3;
RESADIC = XR * DL;
SUMSQ = (ZK'*RES) - ((ZK' * RESADIC) **2)/(DL'*RESADIC);
SUM = SQRT(SUMSQ/ (NOBS - NCOL(X) - 1));
DENOM = DL'*RESADIC;
RDENOM = SQRT (DENOM);
NUMER = DL'*RES;
COEF = NUMER/DENOM;
ESTIMADO = COEF*RESADIC;
TALAM = - NUMER/(SUM*RDENOM);
PRINT,, 'TESTE DE ATKINSON P/ OBTER NORMALIDADE ,T(LAMBDA)= ',,TALAM;
APRLAM = LAM - COEF;
PRINT,, 'VALOR APROXIMADO P/ LAMBDA CHAPEU',,APRLAM;
H1 = XR*DL*DL'*XR;
H2 = DL'*XR*DL;
H = H1/H2;
RESB = (C - H)*RES;
SQ = RES'*RESB;
START COOK;
  DO I = 1 TO NOBS;
    DIF = RES(:I, :) - ESTIMADO(:I, :);
    DIAG = H(:I,I:);
    BAIXO = SQRT(1-DIAG);
    RDIAG = SQRT(DIAG);
    PART1 = SQRT(NOBS - 1);
    PART2 = RDIAG/BAIXO;
    PART3A = DIF/BAIXO;
    PART3 = ABS(PART3A);
    PART41 = INV(1 - H(:I,I:));
    PART42 = RESB(:I, :)*PART41*RESB(:I, :);
  
```

```

PART43 = (SQ - PART42)/(NOBS-2);
PART44 = SQRT( PART43 );
PART4 = INV(PART44);
CI = PART1*PART2*PART3* PART4;
K = (1 - 0.375)/(NOBS+0.25);
K1 = (K + 1)/2;
SEMINOR = PROBIT(K1);
MODULO = MODULO // CI;
RESULT = RESULT // (CI::SEMINOR);
END;
FINISH;
RUN COOK;
RESULT = RESULT :: RES ::RESADIC;
ROTULO = { 'COOK_MOD' 'PER_NOR' 'RESIDUO' 'RES_ADIC' };
CREATE GRAF FROM RESULT (:COLNAME = ROTULO);
APPEND FROM RESULT;
NOME = { 'CI' };
CREATE MODIFICA FROM MODULO (:COLNAME = NOME !);
APPEND FROM MODULO;
FREE RESADIC XR DL RES ESTIMADO H H1 CI MCI SEMINOR;
QUIT;
PROC SORT DATA = MODIFICA;
BY Ci ;
DATA FINAL ;
MERGE MODIFICA GRAF ;
PROC PRINT DATA=FINAL;
VAR RESIDUO RES_ADIC COOK_MOD;
PROC PLOT DATA=FINAL;
PLOT RESIDUO*RES_ADIC = '* ';
TITLE 'GRAFICO DE VAR. ADICIONADA P/ LAMBDA SELECCIONADO';
PROC PLOT DATA = FINAL;
PLOT Ci*PER_NOR = '* ';
TITLE'GRAFICO SEMI-NORMAL P/ ESTAT. DE COOK MODIFICADA';

```

PROGRAMA 7

Programa para construir gráficos probabilísticos normais e exponenciais para os resíduos obtidos após o ajuste de mínimos quadrados e para os resíduos absolutos obtidos após o ajuste de mínimos desvios absolutos.

```
LIBNAME RESIDUO 'A:';
LIBNAME BROWN 'A';
PROC UPLOAD DATA = RESIDUO.REG OUT = BROWN.REG;
RUN;
OPTIONS TLS = 70;
OPTIONS TPS = 60;
PROC CAPABILITY DATA = BROWN.REG NOPRINT;
  VAR RES_L2 ;
  QQ/ QQSYMBOL = '* ' NORMAL;
  TITLE'GRAFICO PROBABILISTICO NORMAL P/RES_L2';
RUN;
PROC CAPABILITY DATA = BROWN.REG NOPRINT;
  VAR RES_L1 ;
  QQ/ QQSYMBOL = '* ' EXP;
  TITLE'GRAFICO PROBABILISTICO EXPONENCIAL P/RES_L1 ABSOLUTO';
RUN;
```

Os programas 1, 2, 3 e 5 foram executados no SAS/micro, os programas 4 e 6 foram executados no SAS/VAX e o programa 7 foi executado no SAS/IBM, devido à disponibilidade de algumas funções utilizadas em cada programa estar presente somente na versão SAS em que foi executado

BIBLIOGRAFIA

- ANDREWS, D.F. (1971). A note on the selection of data transformations. *Biometrika* 58:249-254.
- (1974). A robust method for multiple linear regression. *Technometrics* 16:523-531.
- ATKINSON, A.C. (1973). Testing transformations to normality. *J. R. Statist. Soc. B.* 35:473-479.
- (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika* 68:13-20.
- (1982). Regression diagnostics, transformations and constructed variables. *J. R. Statist. Soc. B.* 44:1-36.
- (1986). Diagnostic tests for transformations. *Technometrics* 28:29-37.
- (1987). Directions in diagnostics regression analysis. *REBRAPE* 2:23-57.
- BELSLEY, D.A.; KUH, E. and WELSCH, R.E. (1980). *Regression Diagnostics*. New York, John Wiley.
- BLOOMFIELD, P. and STEIGER, W.L. (1983). *Least Absolute Deviations; Theory, applications and algoritms*. Boston, Birkhäser.
- BOX, G.E.P. and COX, D.R. (1964). An analysis of transformations. *J. R. Statist. Soc. B.* 26:211-252.

- BROWNLEE, K.A. (1965). *Statistical Theory and Methodology In Science and Engineering*. New York, John Wiley.
- CARROLL, R.J. (1980). A robust method for testing transformations to achieve approximate normality. *J. R. Statist. Soc. B.* 42:71-78.
- (1982). Two examples of transformations when there are possible outliers. *Applied Statistics* 31:149-152.
- CARROLL, R.J. and RUPPERT, D. (1985). Transformations in regression: a robust analysis. *Technometrics* 27:1-12.
- CHARNES, A.; COOPER, W.W. and FERGUSON, R.O. (1955). Optimal estimation of executive compensation by linear programming. *Management Science* 1:138-151.
- COOK, R.D. (1977). Detection of influential observations in linear regression. *Technometrics* 19:15-18.
- COOK, R.D. and WANG, P.C. (1983). Transformations and influential cases in regression. *Technometrics* 25:337-343.
- COOK, R.D. and WEISBERG, S. (1982). *Residuals and Influence in Regression*. New York, Chapman and Hall.
- COX, D.R. and HINKLEY, D.V. (1978). *Problems and Solutions in Theoretical Statistics*. London, Chapman and Hall.
- DANIEL, C. and WOOD, F.S. (1980). *Fitting Equations to Data*. 2^a edição. New York, John Wiley.

- DIELMAN, T. and PFAFFENBERGER, R. (1982). LAV (Least Absolute Value) estimation in the regression model: A review. *Tims Studies in the Management Sciences* 19:31-52.
- DRAPER, N. R. and SMITH, H. (1981). *Applied Regression Analysis*. New York, John Wiley.
- GUNST, R. F. and MASON, R. L. (1980). *Regression Analysis and its Application - A Data Oriented Approach*. New York, Marcel Dekker, Inc..
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35:73-101.
- _____ (1983). Minimax of bounded-influence regression. *JASA* 78:66-80.
- KARST, O. J. (1958). Linear curve fitting using least deviations. *JASA* 53:118-132.
- KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *JASA* 77:595-604.
- MARDIA, K. V.; KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. London, Academic Press.
- NARULA, S. C. and WELLINGTON, J. F. (1982). The minimum sum of absolute errors regression: A state of the art Survey. *International Statistical Review* 50:317-326.
- _____ (1985). Interior analysis for minimum sum of absolute errors regression. *Technometrics* 27:181-188.

- NARULA, S.C. and STANGENHAUS, G. (1988). *Análise de regressão L_1* . IN: 8^o SINAPE. Rio de Janeiro, IMPA.
- NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. IN: *Probability and Statistics*. New York, John Wiley.
- PARKER, I. (1988). Transformations and influential observations in minimum sum of absolute errors regression. *Technometrics* 30:215-220.
- SAS INSTITUTE INC. (1985). *SAS/IML User's Guide*. Version 5 Edition. Cary, SAS Institute Inc..
- TUKEY, J.W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.* 28:602-632.