RT-MAE 2000-23

COMPARAÇÃO DE ESTIMADORES DO PARÂMETRO DE ESCALA DA DISTRIBUIÇÃO DOS ERROS NA REGRESSÃO L_1

by

Carmen Diva Saldiva de Andre
and
Silvia Nagib Elian

Palavras-Chave: Regressão MSAE, regressão linear múltipla, seleção de variáveis, distribuições com caudas pesadas.
Classificação AMS: 62J05, 62F12.
(AMS Classification)

- Dezembro de 2000 -

Comparação de estimadores do parâmetro de escala da distribuição dos erros na regressão L_1

Resumo

O método de estimação L_1, ou de mínima soma dos erros absolutos, dos parâmetros de modelos de regressão linear é uma alternativa robusta ao método de mínimos quadrados, quando são observados valores aberrantes na variável resposta, ou os erros do modelo seguem uma distribuição com caudas pesadas, ou a função de perda apropriada é proporcional à soma dos erros absolutos. O cálculo dos erros padrão dos estimadores L_1, a construção de intervalos de confiança e testes de hipótese sobre os parâmetros do modelo , bem como o cálculo de um coeficiente de determinação robusto, requerem a estimação do parâmetro de escala τ, definido de tal forma que τ^2/n é a variância da mediana amostral em uma amostra de tamanho n da distribuição dos erros. O estimador consistente $\hat{\tau}$ recomendado por McKean e Schrader (1987) apresenta comportamento instável em pequenas amostras, podendo apresentar maior valor quando novas variáveis independentes são adicionadas ao modelo. Quando os erros seguem distribuição de Laplace, o estimador de máxima verossimilhança de τ, $\hat{\tau}^*$, é dado pela média dos valores absolutos dos resíduos, que denominaremos erro médio absoluto. Neste trabalho, pretendemos estudar propriedades estatísticas de $\hat{\tau}^*$ assumindo diferentes distribuições para os erros do modelo, e estabelecer , através de um estudo de simulação, situações definidas pelo tamanho da amostra e distribuição dos erros nas quais $\hat{\tau}^*$ pode ser utilizado como uma alternativa a $\hat{\tau}$.

Palavras chave : *regressão MSAE, regressão linear múltipla, seleção de variáveis, distribuições com caudas pesadas.*

1- Introdução

Consideremos o modelo de regressão linear múltipla:

$$y = X\beta + \varepsilon,$$

onde

y é um vetor $n \times 1$ de valores da variável resposta correspondentes a X, uma matriz $n \times k$ de variáveis preditoras que pode incluir uma coluna de " 1's " para o intercepto,

β é um vetor $k \times 1$ de parâmetros e

ε é um vetor $n \times 1$ de erros aleatórios não observáveis.

Os componentes de ε são variáveis aleatórias independentes e identicamente distribuídas com função distribuição acumulada F. Vamos supor que F tenha uma única mediana igual a zero, e derivada contínua f em uma vizinhança de zero, tal que f(0)>0. Seja τ o parâmetro de escala definido como :

$$\tau = (2\, f(v))^{-1}, \qquad (1.1)$$

de modo que τ^2/n é a variância da mediana amostral em uma amostra de tamanho n da distribuição dos erros.

O estimador L_1 dos parâmetros do modelo, $\hat{\beta}$, é o valor de β que minimiza $\sum_{i=1}^{n} | y_i - x_i \beta |$, onde y_i é o i-ésimo elemento do vetor y, e x_i é a i-ésima linha da matriz X. O critério L_1 é uma alternativa robusta ao método de mínimos quadrados sempre que os dados contêm valores aberrantes, ou quando os erros seguem distribuição com caudas pesadas, como, por exemplo as distribuições de Cauchy e Laplace, ou quando a função de perda é proporcional ao valor absoluto dos erros.

O método de estimação L_1 tem uma longa história em aplicações estatísticas, desde a sua formulação inicial feita por Roger Boscovich em 1760. Após a proposição do método de mínimos quadrados por Legendre em 1805, esse método caiu no esquecimento, devido à dificuldade computacional requerida para o cálculo dos estimadores.

Entretanto, após 1955, o método L_1 voltou a atrair a atenção de pesquisadores, devido à sua formulação como um problema de programação linear feita por Charnes, Cooper e Ferguson (1955). A partir de então, surgiram vários trabalhos propondo algoritmos para a obtenção de estimativas L_1 em modelos de regressão linear simples e múltipla e estudando suas propriedades estatísticas . Um histórico sobre a regressão L_1 pode ser encontrado em Narula et al (1999) e em Portnoy e Koenker(1997).

Sabe-se que os estimadores L_1 são também estimadores de máxima verossimilhança, e portanto assintoticamente não viesados e eficientes, quando os erros seguem distribuição de Laplace. Basset e Koenker(1978) provaram que os estimadores L_1 dos parâmetros do modelo de regressão são assintoticamente não viesados, consistentes e seguem assintoticamente distribuição normal multivariada com matriz de variâncias e covariâncias dada por $\tau^2 (X' X)^{-1}$. Uma conseqüência importante desse resultado é que os estimadores L_1 de β possuem menor elipsóide de confiança do que os estimadores de mínimos quadrados, para qualquer distribuição dos erros na qual a mediana é uma medida de posição mais eficiente do que a média.

Com base na distribuição assintótica dos estimadores L_1 foram deduzidas fórmulas para a construção de intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo (Dielman e Pfaffenberger, 1982 e Narula, 1987).Para o cálculo dos erros padrão dos estimadores de β e também na construção dos referidos intervalos de confiança e testes de hipótese , é necessário termos uma estimativa do parâmetro τ. Vários estimadores para τ foram propostos (McKean e Schrader, 1984). Um estimador consistente recomendado por esses autores é dado por:

$$\hat{\tau} = \sqrt{n^*}\,(e_{(n^*-m+1)} - e_{(m)})/4,$$

onde

$m = (n^* + 1)/2 - \sqrt{n^*}$,

n^* é o número de resíduos não nulos,

$e_{(1)}, \ldots, e_{(n)}$ são os resíduos não nulos ordenados em ordem crescente.

Observemos que $\hat{\tau}$ é uma medida de variabilidade dos resíduos, sendo influenciado por todos os resíduos não nulos, mas determinado por apenas dois deles.

Um estimador consistente de τ também é necessário para o cálculo do coeficiente de determinação robusto proposto por McKean e Sievers(1987). Esse coeficiente é uma medida informal do ajuste do modelo. Sua expressão é:

$$R_2 = RSEA/(RSEA + (n-p-1)(\hat{\tau}/2)),$$

onde :

$$RSEA = \sum_{i=1}^{n} \left| y_i - \hat{\beta}_0 \right| - \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|,$$

$\hat{\beta}_0$ = mediana $(y_1, \ldots, y_n)$,

$\hat{y}_i$ é o valor previsto da i-ésima observação , ou seja, $\hat{y}_i = x_i \hat{\beta}$,

$\hat{\beta}$ é o estimador L_1 dos coeficientes de regressão do modelo.

Uma propriedade desejável para um coeficiente de determinação é que seu valor aumente quando novas variáveis preditoras são adicionadas ao modelo (Kvalseth, 1985). Para o coeficiente de determinação R_2 , isto ocorre somente se a estimativa de τ diminui à medida que novas variáveis são adicionadas ao modelo. Isto nem sempre acontece, como mostramos no exemplo a seguir.

Exemplo 1.1- Os resultados apresentados neste exemplo foram obtidos a partir de dados relativos a imóveis, retirados de Narula e Wellington (1977), contendo as variáveis Imposto, em centenas de dólares (x_1), Área do Terreno, em milhares de pés quadrados (x_2), Área Construída, em milhares de pés quadrados (x_3), Idade da Residência, em anos (x_4), e a variável resposta, Preço de Venda da casa (y), em milhares de dólares.

Na Tabela 1.1 apresentamos todos os possíveis modelos que podem ser obtidos a partir das variáveis explicativas consideradas , o número de parâmetros (k), e os valores observados de $\hat{\tau}$ e do coeficiente de determinação R_2.

Tabela 1.1. Número de parâmetros (k) e valores observados de $\hat{\tau}$ e de R_2 para todos os modelos de regressão possíveis, utilizando os dados de imóveis.

Variáveis no Modelo	k	$\hat{\tau}$	R_2
Nenhuma	1	10,9629	0,0000
x_1	2	4,0079	0,7105
x_2	2	8,0285	0,3830
x_3	2	9,8973	0,4681
x_4	2	11,8750	0,1164
x_1, x_2	3	3,9590	0,7215
x_1, x_3	3	5,4301	0,6860
x_1, x_4	3	3,9636	0,7212
x_2, x_3	3	8,0913	0,5470
x_2, x_4	3	6,5918	0,4576
x_3, x_4	3	6,6577	0,6098
x_1, x_2, x_3	4	6,1331	0,6701
x_1, x_2, x_4	4	7,3275	0,5937
x_1, x_3, x_4	4	5,4875	0,7011
x_2, x_3, x_4	4	4,0022	0,7410
x_1, x_2, x_3, x_4	5	4,0246	0,7704

Observamos que o valor de $\hat{\tau}$ no modelo com a variável x_1 (4,0079) é menor que no modelo com x_1 e x_3 (5,4301), fazendo com que o R_2 do modelo com x_1 (0,7105) seja maior que o R_2 no modelo com x_1 e x_3 (0,6860). Entretanto, a contribuição adicional da variável x_3, dado que x_1 está no modelo, é significante (nível descritivo menor que 0,01).

Portanto, devido à instabilidade das estimativas de τ em pequenas amostras, que decorre provavelmente do fato do valor de $\hat{\tau}$ ser determinado por apenas dois resíduos, pode ocorrer que, mesmo sendo introduzida uma variável que traga contribuição na explicação da variável resposta, a estimativa de τ aumente quando se passa do modelo reduzido para o completo. Este fato diminui o valor de R_2 , fazendo com que esse modelo não seja selecionado, se o coeficiente de determinação for o critério utilizado para escolha do modelo final entre todas as possíveis regressões.

Sabe-se que, quando os erros seguem uma distribuição de Laplace, o estimador de máxima verossimilhança de τ é o Erro Médio Absoluto (ver, por exemplo, Elian et al, 2000), dado por:

$$\hat{\tau}^* = SEA/n \tag{1.2}$$

onde

$$SEA = \sum_{i=1}^{n} |y_i - \hat{y}_i|.$$

Apesar das condições de regularidade usuais não estarem satisfeitas, $\hat{\tau}^*$ é um estimador consistente de τ (Engelhardt e Bain, 1973). Esse estimador é uma medida de variabilidade dos resíduos e possui a propriedade de diminuir quando novas variáveis preditoras são adicionadas ao modelo. Sua utilização permitiu a construção de um coeficiente de determinação robusto que satisfaz às condições estabelecidas por Kvalseth (1985) e um coeficiente de determinação ajustado pelo número de variáveis preditoras no modelo (André et al, 2000).

Nosso objetivo neste trabalho é estudar a possibilidade da utilização do estimador $\hat{\tau}^*$ como uma alternativa a $\hat{\tau}$ quando os erros seguem outras distribuições além da Laplace. A importância do método de estimação L_1 foi recentemente apontada por Portnoy e Koenker(1997) , que apresentaram várias situações práticas nas quais a aplicação do método é recomendada . Portanto, a busca de procedimentos que tornem a aplicação do método mais eficiente traz contribuição importante para a teoria estatística.

Na seção 2, deduzimos distribuições aproximadas para $\hat{\tau}^*$, considerando erros com distribuição Normal, Normal Contaminada Laplace e Logística. Foram também feitas algumas considerações sobre o viés assintótico desse estimador.

Na seção 3, realizamos um estudo de simulação no qual distribuições empíricas de $\hat{\tau}^*$ e $\hat{\tau}$ foram geradas a partir do ajuste de modelos com uma variável preditora e erros com as distribuições consideradas na seção 2 . Médias e desvios padrão das distribuições empíricas foram comparados com os correspondentes parâmetros das distribuições aproximadas obtidas na seção 2, para diferentes tamanhos de amostra.

Considerando adicionalmente a distribuição de Cauchy, comparamos, $\hat{\tau}^*$ e $\hat{\tau}$ quanto ao viés e erro quadrático médio.

Procuramos, na seção 4, associar os resultados obtidos nas seções 2 e 3 ao peso das caudas das distribuições de erros consideradas .

As conclusões do trabalho são apresentadas na seção 5.

2. Distribuição aproximada de $\hat{\tau}^*$

Nesta seção, determinaremos a distribuição assintótica de $\hat{\tau}^*$, considerando erros com distribuição Normal (0, σ^2), Normal Contaminada consistindo de variáveis aleatórias selecionadas da Normal (0,1) com probabilidade p, e de uma N(0, σ^2) com probabilidade (1-p), Logística com média zero e variância $\gamma^2\pi^2/3$ e Laplace com média zero e variância $2\sigma^2$.

Notemos inicialmente que $\hat{\tau}^*$ pode ser escrito como:

$$\hat{\tau}^* = \frac{1}{n}\sum_{i=1}^{n}| y_i - x_i\hat{\beta} | = \frac{1}{n}\sum_{i=1}^{n}| y_i - x_i\beta - (x_i\hat{\beta} - x_i\beta)|, \qquad (2.1)$$

onde $\hat{\beta}$ é o estimador L_1 do vetor de parâmetros β.

ε_i

Como $x_i\hat{\beta}$ é um estimador consistente de $x_i\beta$ (Basset e Koenker, 1978), a distribuição assintótica de $\hat{\tau}^*$ é a mesma de $\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta|$, que, por sua vez é igual a $\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|$, onde ε_i é o i-ésimo elemento do vetor de erros do modelo.

Passaremos a estudar a distribuição assintótica dessa variável aleatória para diferentes distribuições dos erros do modelo.

Erros com distribuição N(0,σ^2)

Utilizando o fato de que $\frac{\varepsilon_i}{\sigma} \sim$ N (0,1), verificamos (ver Apêndice A) que

$$E\left(\frac{|\varepsilon_i|}{\sigma}\right) = \frac{\sqrt{2}}{\sqrt{\pi}} \quad \text{e} \quad Var\left(\frac{|\varepsilon_i|}{\sigma}\right) = \frac{\pi - 2}{\pi} \quad .$$

Além disso, as variáveis aleatórias $\frac{|\varepsilon_i|}{\sigma}$, i = 1,..., n, são independentes e identicamente distribuídas e assim, devido ao Teorema Central do Limite,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi-2}}{\sqrt{\pi}\sqrt{n}}} \xrightarrow{D} N\,(0,1).$$

Como conseqüência,

$$W_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - x_i\hat{\beta}\right|}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi-2}}{\sqrt{\pi}\sqrt{n}}} = \frac{\frac{\hat{\tau}^*}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi-2}}{\sqrt{\pi}\sqrt{n}}} \xrightarrow{D} \text{N }(0,1).$$

Como $\varepsilon_i \sim$ Normal (0, σ^2), temos que $\tau = \frac{\sqrt{2\pi}}{2}\sigma$, e então

$$W_n = \left(\frac{\pi}{\sqrt{2(\pi-2)}}\sqrt{n}\,\frac{\hat{\tau}^*}{\tau} - \frac{\sqrt{2n}}{\sqrt{2(\pi-2)}}\right) \xrightarrow{D} \text{N }(0,1),$$

ou ainda

$$W_n = \frac{\sqrt{n}\pi}{\sqrt{2(\pi-2)}}\left(\frac{\hat{\tau}^*}{\tau} - \frac{2}{\pi}\right) \xrightarrow{D} \text{N}(0,1)$$

Erros com distribuição Normal Contaminada

No caso em que os erros têm distribuição Normal Contaminada, isto é, consistem em variáveis aleatórias selecionadas da Normal (0,1) com probabilidade p, e de uma N(0, σ^2) com probabilidade (1-p), a função densidade de probabilidades de ε_i é da forma

$$f(\varepsilon_i) = \frac{p}{\sqrt{2\pi}}\exp(-\frac{\varepsilon_i^2}{2}) + \frac{1-p}{\sqrt{2\pi}\sigma}\exp(-\frac{\varepsilon_i^2}{2\sigma^2}), \quad -\infty < \varepsilon_i < \infty.$$

Verifica-se que, nestas condições,

E(ε_i) = 0, Var(ε_i) = p+(1-p)σ^2 e que o parâmetro τ é dado por

$$\tau = \frac{\sqrt{2\pi}\sigma}{2(p\sigma + 1 - p)}.$$

Adicionalmente, $|\varepsilon_i| = |y_i - x_i\beta|$, i=1,...,n são variáveis aleatórias independentes e identicamente distribuídas com média e variância (ver Apêndice A) dadas por

$$E(|\varepsilon_i|)=\frac{2p}{\sqrt{2\pi}}+\frac{2(1-p)\sigma}{\sqrt{2\pi}} \text{ e}$$

$$\operatorname{Var}(|\varepsilon_i|)=\bar{\sigma}^2=\frac{(\pi(1-p)-2(1-p)^2)\sigma^2-4p(1-p)\sigma+p\pi-2p^2}{\pi}$$

Assim, pelos mesmos argumentos utilizados no caso de erros com distribuição Normal,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i-x_i\beta|-\frac{2p+2(1-p)\sigma}{\sqrt{2\pi}}}{\bar{\sigma}/\sqrt{n}}\xrightarrow{D} N(0,1),$$

e com isso,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i-x_i\hat{\beta}|-\frac{2p+2(1-p)\sigma}{\sqrt{2\pi}}}{\bar{\sigma}/\sqrt{n}}\xrightarrow{D} N(0,1),$$

ou seja,

$$U_n=\frac{\sqrt{n}\,\frac{\hat{\tau}^*}{\tau}-\frac{2(p+(1-p)\sigma)\sqrt{n}}{\sqrt{2\pi}\,\tau}}{\bar{\sigma}/\tau}\xrightarrow{D} N(0,1).$$

Erros com distribuição Logística

Admitindo que a distribuição dos erros é Logística com função densidade

$$f(\varepsilon_i)=\frac{\exp(-\varepsilon_i/\gamma)}{\gamma[1+\exp(-\varepsilon_i/\gamma)]^2}, \quad -\infty<\varepsilon_i<\infty,$$

de modo que $E(\varepsilon_i) = 0$, $\operatorname{Var}(\varepsilon_i) = \gamma^2\pi^2/3$, verificamos que o parâmetro τ é 2γ.

Por outro lado, $|\varepsilon_i|$, i = 1, ..., n são variáveis aleatórias independentes e identicamente distribuídas e verifica-se (ver Apêndice A) que a média e a variância dessas variáveis são respectivamente $1,386\gamma$ e $1,37\ \gamma^2$.

Desta forma, pelo Teorema Central do Limite,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta| - 1,386\gamma}{\gamma\sqrt{1,37}\Big/\sqrt{n}} \xrightarrow{D} N(0,1)\ .$$

Devido aos argumentos utilizados anteriormente, segue como conseqüência que

$$\frac{\hat{\tau}^* - 1,386\gamma}{\gamma\sqrt{1,37}\Big/\sqrt{n}} = \frac{\frac{\hat{\tau}^*}{\tau} - \frac{1,386\gamma}{\tau}}{\gamma\sqrt{1,37}\Big/\tau\sqrt{n}} \xrightarrow{D} N\ (0,1).$$

Como neste caso $\tau = 2\ \gamma$, obtemos

$$\frac{2\sqrt{n}}{\sqrt{1,37}}\left(\frac{\hat{\tau}^*}{\tau} - 0,693\right) \xrightarrow{D} N\ (0,1).$$

Erros com distribuição de Laplace

Para erros com distribuição de Laplace com média zero e variância $2\sigma^2$, temos que $\tau = \sigma$, $E(|\varepsilon_i|) = \sigma$ e $Var(|\varepsilon_i|) = \sigma^2$.

Portanto, pelo Teorema Central do Limite,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta| - \sigma}{\sigma\Big/\sqrt{n}} \xrightarrow{D} N\ (0,1),$$

e, com isso,

$$\frac{\hat{\tau}^* - \tau}{\tau / \sqrt{n}} = \sqrt{n}\left(\frac{\hat{\tau}^*}{\tau} - 1\right) \xrightarrow{D} \mathrm{N}\,(0,1).$$

Comentários

Pelo fato da média amostral dos valores absolutos dos resíduos ser uma função contínua e limitada, temos, pelo Lema de Helly-Bray (ver, por exemplo, Sen and Singer, 1993), que a esperança desse estimador converge para a média de sua distribuição assintótica, e desta forma a expressão do viés assintótico de $\hat{\tau}^*$ pode ser obtida.

Os resultados obtidos sobre a distribuição assintótica de $\hat{\tau}^*$, mostram que o viés assintótico desse estimador é não nulo para todas as distribuições de erros consideradas, a menos da distribuição de Laplace. No caso de erros com distribuição Normal (0, σ^2), o viés assintótico é dado por

$$\frac{\tau(2-\pi)}{\pi} < 0,$$

ou seja , $\hat{\tau}^*$, em média, subestima τ. Se os erros seguem distribuição Normal Contaminada, a expressão do viés assintótico é

$$\frac{p(1-p)(\sigma^2-1)^2 + (1-\pi)\sigma}{\sqrt{2\pi}\,(p\sigma + (1-p))},$$

que assume valor negativo se $p(1-p)(\sigma^2-1)^2 < (\pi-1)\sigma$ e positivo, caso contrário. Quando a distribuição dos erros é Logística, o viés assintótico é dado por

$$-0{,}614\,\gamma = -0{,}307\,\tau$$

sendo, portanto, sempre negativo.

Com base nas distribuições assintóticas apresentadas , intervalos de confiança assintóticos para τ podem ser construídos. Por exemplo, no caso de erros normais, um intervalo de confiança assintótico para τ é dado por

$$\left(k_1\,\hat{\tau}^*, k_2\hat{\tau}^*\right), \qquad (2.1)$$

onde

$$k_1 = \frac{\sqrt{n}\,\pi}{2\sqrt{n} + z\sqrt{2(\pi-2)}},$$

$$k_2 = \frac{\sqrt{n}\,\pi}{2\sqrt{n} - z\sqrt{2(\pi-2)}} \text{ e}$$

z é o quantil de ordem (1+ γ)/2 da distribuição Normal padrão, sendo γ o coeficiente de confiança do intervalo.

O intervalo de confiança acima permite testar hipóteses do tipo H: $\tau = \tau_0$ ao nível de significância α = (1- γ).

3. Simulação

Um estudo de simulação foi realizado com o objetivo de gerar distribuições empíricas dos estimadores $\hat{\tau}^*$e $\hat{\tau}$. Considerando as seguintes distribuições para os erros ε_i do modelo:

- Normal (0,1) $\Rightarrow$ τ = 1,253
- Logística com média zero e variância $\pi^2/3 \Rightarrow \tau$ =2,00,
- Laplace com média zero e variância 2 $\Rightarrow$ τ = 1,00,
- Normal contaminada (NC85-15) que consiste de variáveis aleatórias selecionadas da Normal(0,1) com probabilidade 0.85, e da N(0,49) com probabilidade 0.15 $\Rightarrow$ τ = 1,439,
- Normal contaminada (NC80-20) que consiste de variáveis aleatórias selecionadas da Normal(0,1) com probabilidade 0.80, e da N(0,49) com probabilidade 0.20 $\Rightarrow$ τ = 1,513 e
- Cauchy com mediana igual a zero e parâmetro de escala igual a um $\Rightarrow$ τ = 1,571,

o estudo foi conduzido da seguinte forma :

- Foram considerados modelos de regressão contendo 1 variável independente gerada a partir da distribuição Normal(0,1), independentemente dos erros. Sem perda de generalidade, aos coeficientes do modelo foi atribuído valor unitário.

- Foram considerados tamanhos de amostra (n) iguais a 10, 20, 30, 50, 100 e 200.

- Em cada uma das situações definidas por n ,e distribuição dos erros foram gerados 1000 conjuntos de dados.

- A cada conjunto de dados gerado foi ajustado o modelo de regressão pelo método L_1 e obtidos os valores de $\hat{\tau}^*$ e $\hat{\tau}$. Para cada tamanho de amostra e distribuição dos erros, esse procedimento gerou 1000 valores de $\hat{\tau}^*$ e $\hat{\tau}$, a partir dos quais foram calculados os resultados apresentados a seguir.

Para a realização da simulação, foi construído um programa no aplicativo S-Plus.

As distribuições de $\hat{\tau}^*$ e $\hat{\tau}$ geradas na simulação foram utilizadas com dois propósitos. Considerando inicialmente as distribuições geradas para $\hat{\tau}^*$, utilizamos os resultados da simulação para comparar as distribuições aproximadas de $\hat{\tau}^*$ obtidas a partir da distribuição assintótica desenvolvida na seção 2 com as geradas no estudo, para amostras de tamanho n = 10, 30, 100 e 200. A comparação foi feita somente através do cálculo das médias e desvios padrão das duas distribuições, que são apresentadas nas Tabelas 3.1 a 3.5. Uma comparação mais completa será feita futuramente estimando-se probabilidades de cobertura de intervalos de confiança para τ construídos com base na distribuição assintótica.

Tabela 3.1- Médias e Desvios padrão das distribuições teórica e empírica de $\hat{\tau}^*$, para n=10, 30, 100 e 200, quando os erros têm distribuição N(0,1), (τ = 1,253)

n	distribuição aproximada		distribuição empírica	
	média	desvio padrão	média	desvio padrão
10	0,798	0,191	0,675	0,179
30	0,798	0,110	0,758	0,130
100	0,798	0,060	0,785	0,061
200	0,798	0,043	0,788	0,044

Tabela 3.2- Médias e Desvios padrão das distribuições teórica e empírica de $\hat{\tau}^*$, para n=10, 30, 100 e 200, quando os erros têm distribuição Logística com média zero e variância $\pi^2/3$, (τ=2,00).

n	distribuição aproximada		distribuição empírica	
	média	desvio padrão	média	desvio padrão
10	1,386	0,370	1,189	0,355
30	1,386	0,214	1,311	0,212
100	1,386	0,117	1,363	0,120
200	1,386	0,083	1,380	0,082

Tabela 3.3- Médias e Desvios padrão das distribuições teórica e empírica de $\hat{\tau}^*$, para n=10, 30, 100 e 200, quando os erros têm distribuição NC85-15 (τ = 1,439).

n	distribuição aproximada		distribuição empírica	
	média	desvio padrão	média	desvio padrão
10	1,516	0,768	1,340	0,747
30	1,516	0,444	1,498	0,450
100	1,516	0,243	1,490	0,236
200	1,516	0.172	1,502	0,165

Tabela 3.4- Médias e Desvios padrão das distribuições teórica e empírica de $\hat{\tau}^*$, para n=10, 30, 100 e 200, quando os erros têm distribuição NC80-20, (τ = 1,513)

n	distribuição aproximada		distribuição empírica	
	média	desvio padrão	média	desvio padrão
10	1,755	0,867	1,624	0,876
30	1,755	0,501	1,498	0,450
100	1,755	0,274	1,739	0,277
200	1,755	0,194	1,745	0,190

Tabela 3.5- Médias e Desvios padrão das distribuições teórica e empírica de $\hat{\tau}^*$, para n=10, 30, 100 e 200, quando os erros têm distribuição Laplace com média zero e variância 2, ($\tau = 1$)

n	distribuição aproximada		distribuição empírica	
	média	desvio padrão	média	desvio padrão
10	1	0,316	0,872	0,310
30	1	0,183	0,975	0,182
100	1	0,100	0,996	0,097
200	1	0,071	0,993	0,070

Os resultados apresentados nas Tabelas 3.1 a 3.5 indicam que as médias e desvios padrão das distribuições assintóticas e empíricas de $\hat{\tau}^*$ são próximas , sendo uma indicação preliminar de que a distribuição assintótica é uma boa aproximação para a distribuição desse estimador mesmo para amostras pequenas, qualquer que seja a distribuição dos erros.

Os resultados da simulação foram também utilizados com o propósito de comparar as distribuições dos estimadores $\hat{\tau}^*$ e $\hat{\tau}$.

Nas Tabelas B1 a B7 do Apêndice B apresentamos estatísticas descritivas das distribuições desses estimadores. Os resultados obtidos sugerem que :

- $\hat{\tau}^*$ pode ser adotado como uma alternativa a $\hat{\tau}$, quando os erros seguem distribuição de Laplace, NC85-15, ou NC80-20, pois neste caso o estimador proposto apresenta viés e erro quadrático médio menor ou da mesma ordem que $\hat{\tau}$, principalmente em amostras de tamanho inferior a 50.
- Quando os erros seguem distribuição Normal ou Logística , $\hat{\tau}^*$ tende a subestimar τ. As médias das distribuições do estimador $\hat{\tau}$ geradas no estudo mostraram-se mais próximas ao verdadeiro valor do parâmetro, sendo seus erros quadráticos médios

praticamente uniformemente menores que os de $\hat{\tau}^*$ para todos os tamanhos de amostra.

- Para erros com distribuição de Cauchy, o estimador em estudo tende a superestimar τ. Este resultado pode ser atribuído ao fato de todos os resíduos serem considerados no cálculo de $\hat{\tau}^*$. Embora com menor viés e erro quadrático médio menor que o de $\hat{\tau}^*$, $\hat{\tau}$ não apresentou bom comportamento para amostras de tamanho menor ou igual a 30.

4. Algumas características sobre as distribuições consideradas no estudo

As distribuições consideradas nas seções 2 e 3 são todas simétricas em torno de zero e podem ser ordenadas segundo o peso relativo de suas caudas (Groeneveld e Meeden, 1984). Um coeficiente que pode ser utilizado com essa finalidade é dado por:

$$b_2(\alpha) = [F^{*-1}(1-\alpha)+F^{*-1}(\alpha) - 2v^*]/ [F^{*-1}(1-\alpha) + F^{*-1}(\alpha)],$$

onde :

$F^*(x) = 2F(x)-1$,

$F(x)$ é a função distribuição dos erros,

v^* é a mediana da função densidade associada a F^*.

Esse coeficiente é uma extensão do coeficiente de assimetria de Bowley (Bowley, 1920) e apresenta as seguintes características:

- $-1 \leq b_2(\alpha) \leq 1$,
- $b_2(\alpha)$ não requer a existência de qualquer momento da distribuição dos erros, e seu valor é independente dos parâmetros de locação e escala.

Apresentamos na Tabela 4.1 os valores de $b_2(\alpha)$ para as distribuições consideradas no estudo, considerando α=0,10 e α=0,05. Temos , para os dois valores de α considerados, a seguinte ordenação das distribuições segundo o peso de suas caudas:

Normal < Logística < Laplace < NC85-15 < NC80-20 < Cauchy.

Tabela 4.1 - Valores de $b_2(\alpha)$ para as distribuições consideradas no estudo

Distribuição	b2(0,10)	b2(0,05)
Normal	0,2775	0,3551
Logística	0,3455	0,4396
Laplace	0,4650	0,5641
NC85-15	0,5618	0,7848
NC80-20	0,6972	0,8076
Cauchy	0,7265	0,8541

5. Considerações Finais

Neste trabalho foram estudados o comportamento do estimador $\hat{\tau}^*$ com o objetivo de utilizar esse estimador como uma alternativa a $\hat{\tau}$. Determinamos sua distribuição assintótica considerando diferentes distribuições para os erros do modelo. Verificamos que esse estimador é, de uma forma geral, viesado assintoticamente, possuindo viés assintótico nulo quando os erros seguem distribuição de Laplace. Nos casos de erros com distribuição Normal ou Logística o viés assintótico é negativo. Através de um estudo de simulação comparamos os dois estimadores quanto ao viés e erro quadrático médio, considerando distribuições com diferentes pesos nas caudas e diferentes tamanhos de amostra. Observamos que $\hat{\tau}^*$é uma boa alternativa a $\hat{\tau}$ quando os erros do modelo seguem distribuição de Laplace ou Normal Contaminada com os parâmetros considerados no estudo. Quando os erros possuem distribuição Normal, Logística (caudas mais leves) ou Cauchy (caudas mais pesadas), $\hat{\tau}$ apresentou melhor comportamento, para todos os tamanhos de amostra. Entretanto, no caso da distribuição de Cauchy, embora $\hat{\tau}$ tenha apresentado melhor comportamento que $\hat{\tau}^*$, sua utilização não é recomendada em amostras de tamanho menor ou igual a 30, devido ao viés apresentado por esse estimador.

Os resultados do estudo sugerem que $\hat{\tau}^*$ deve ser utilizado quando a distribuição dos erros for próxima da Laplace. Devido às propriedades desse estimador citadas na seção 1, sugerimos, como estratégia de análise, verificar se os erros apresentam desvios da distribuição de Laplace através da construção de um *Q-Q plot* dos resíduos. No caso de não serem observados sérios desvios, esse estimador pode ser utilizado. Caso contrário, transformações de tipo Box-Cox podem ser aplicadas aos dados, segundo metodologia apresentada em Parker (1988). Na análise dos dados transformados $\hat{\tau}^*$ pode então ser ou ainda utilizado na construção de intervalos de confiança, e testes de hipótese sobre os parâmetros do modelo, e no cálculo de coeficientes de determinação robustos com e sem a correção pelo número de variáveis preditoras no modelo.

Bibliografia

André, C.D.S., Elian, S.N., Narula, S.C. and Tavares, R.A. (2000). Coefficients of Determination for variable selection in the MSAE regression. *Communication in Statistics – Theory and Methods*, **29(3)** – 623-642.

Basset, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**, 618-622.

Charnes, A., Cooper, W. W., and Ferguson, R. D. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, **1**, 138 -151.

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.

Elian, S. N., André, C.D.S. and Narula, S.C. (2000). Influence Measure for the L_1 regression. *Communication in Statistics – Theory and Methods*, **29(4)** – 837-850

Engelhardt, M. and Bain L.J. (1973). Interval Estimation for the Two - parameter Double Exponential Ditribution. *Technometrics*, **15(4)**, 875 - 887.

Kvalseth,T. O .(1985). Cautionary note about R^2.*American Statistician* , **39**, 279-285

McKean , J. W. and Schrader, R.M. (1987). Least absolute errors analysis of variance. *Statistical Data Analysis Based on the L_1-Norm and Related Methods*. (Y. Dodge, editor), Elsevier Science Publishers B. V., 297 - 305.

McKean , J. W. and Sievers, G. L. (1987). Coefficient of Determination for least absolute deviation analysis. *Statistics and Probability Letters*, **5**, 49 - 54.

Narula, S.C. and Wellington, J. F. (1977). Prediction, linear regression and minimum sum of absolute errors. *Technometrics*, **19**, 185-190.

Narula, S.C. (1987). The minimum sum of absolute errors regression. *Journal of Quality Technology*, 19, 37-45.

Narula, S.C., Saldiva, P.H.N., André, C.D.S., Elian, S.N., Ferreira, A .F. and Capelozzi,V. (1999). *Statistics in Medicine*, 18, 1401-1417.

Parker, I. (1988). Transformations and influential observations in minimum sum of absolute errors regression. *Technometrics*, 30, 215-220.

Portnoy, S. and Koenker, R.(1997). The Gaussian hare and the Laplacian Tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4), 4, 279-300.

Sen, P.K. and Singer, J.M. (1993). Large sample methods is Statistics - An introduction with applications. Chapman and Hall, New York.

Apêndice A – Resultados utilizados na seção 2

Na seção 2, quando obtivemos as distribuições assintóticas de $\hat{\tau}^* / \tau$, as distribuições consideradas para os erros eram simétricas em torno de zero. Verifica-se facilmente que se X é uma variável aleatória com valores no intervalo] -∞, ∞ [, simétrica em torno do zero e com densidade *f(x)*, então a densidade de Y=|X| é

$$g(y) = \begin{cases} 2 f(y), & \text{para } y \geq 0 \\ 0, & \text{para } y < 0. \end{cases}$$

Utilizando esse fato, obtivemos E($|\varepsilon_i|$), quando ε_i tem distribuição Normal, Normal Contaminada , Logística e Laplace.

Para $\varepsilon_i \sim N(0, \sigma^2)$, segue que $W_i = |\varepsilon_i / \sigma|$ tem densidade

$$g(w_i) = \begin{cases} \dfrac{2}{\sqrt{2\pi}} exp(-\dfrac{1}{2} w_i^2), & w_i \geq 0 \\ 0, & w_i < 0 \end{cases}$$

e assim,

$$E(W_i) = \int_0^\infty \frac{2w_i}{2\pi} exp(-\frac{1}{2} w_i^2) dw_i = \frac{2}{\sqrt{2\pi}} \left[- exp\left(- \frac{1}{2} w_i^2 \right) \right]_0^\infty = \frac{2}{\sqrt{2\pi}} .$$

Por outro lado,

$$E(W_i^2) = E\left(\frac{\varepsilon_i^2}{\sigma^2} \right) = \frac{1}{\sigma^2} E(\varepsilon_i^2) = Var(\varepsilon_i)/\sigma^2 = 1,$$

e, portanto,

$$Var(W_i) = (\pi-2)/\pi.$$

Quando os erros ε_i têm distribuição normal contaminada, a densidade de $U_i = |\varepsilon_i|$ é dada por

$$f(u_i) = \begin{cases} \frac{2p}{\sqrt{2\pi}} exp(-\frac{1}{2}u_i^2 + \frac{2(1-p)}{\sqrt{2\pi\sigma}} exp(-\frac{1}{2}u_i^2), & u_i \geq 0 \\ 0, & u_i < 0 \end{cases}$$

Desta forma,

$$E(U_i) = \int_0^\infty \frac{2p}{\sqrt{2\pi}} u_i\, exp(-\frac{1}{2}u_i^2)\, du_i + \int_0^\infty \frac{2p(1-p)}{\sqrt{2\pi\sigma}} u_i\, exp(-\frac{1}{2\sigma^2}u_i^2) du_i$$

$$= \frac{2p}{\sqrt{2\pi}}\left[-exp(-\frac{1}{2}u_i^2\right]_0^\infty + \frac{2(1-p)}{\sqrt{2\pi\sigma}}\left[-\sigma^2\, exp(-\frac{1}{2\sigma^2}u_i^2)\right]_0^\infty$$

$$= \frac{2p}{\sqrt{2\pi}} + \frac{2(1-p)\sigma}{\sqrt{2\pi}} \quad \text{e}$$

$E(U_i^2) = E(\varepsilon_i^2) = Var(\varepsilon_i) = p + (1\text{-}p)\ \sigma^2$

que é a variância de uma variável aleatória cuja distribuição é uma mistura de normais, com parâmetros p, (1-p), médias iguais a zero e variâncias iguais a 1 e σ^2.

Como consequência,

$$Var\ (U_i) = \frac{\left[\pi(1-p) - 2(1-p)^2\right]\sigma^2 - 4p(1-p)\sigma + p\pi - 2p^2}{\pi}.$$

No caso em que ε_i tem distribuição **Logística** com média zero e variância $\gamma^2\pi^2/3$,

$$E(|\varepsilon_i|) = \int_0^\infty \frac{2x\, exp\left(-x/_\gamma\right)}{\lambda\left[1+exp\left(-x/_\gamma\right)\right]^2} dx$$

$$= 2\gamma\int_0^\infty \frac{z\, exp(-z)}{[1+exp(-z)]^2} dz = 1{,}386\ \gamma\ ,$$

pois,

$$\int_0^{\infty} \frac{z\,exp(-z)}{[1+exp(-z)]^2}\,dz = 0{,}693.$$

Por outro lado,

$E(Z_i^2) = E(\varepsilon_i^2) = \gamma^2\pi^2/3$ e, como conseqüência,

$Var(Z_i) = Var(|\varepsilon_i|) = 0{,}3289\,\gamma^2 - 1{,}386^2\,\gamma^2 = 1{,}37\,\gamma^2$.

Para ε_i com distribuição de Laplace com média zero e variância $2\sigma^2$, $|\varepsilon_i|$ tem distribuição exponencial com média σ e variância σ^2. Portanto,

$E(|\varepsilon_i|) = \sigma = \tau$ e

$Var(|\varepsilon_i|) = \sigma^2 = \tau^2$.

Apêndice B - Tabelas

Tabela B1- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^*$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição Normal (0,1) (τ = 1,253)e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	1,233	0,675	1,193	0,661	0,198	0,366	0,276	0,001	2,893	1,073
20	1,316	0,735	1,289	0,730	0,149	0,286	0,175	0,351	2,681	1,210
30	1,252	0,758	1,216	0,753	0,109	0,257	0,368	0,477	2,404	1,117
50	1,138	0,776	1,123	0,773	0,097	0,051	0,447	0,501	2,250	1,030
100	1,164	0,785	1,158	0,783	0,069	0,223	0,415	0,611	2,014	0,978
150	1,244	0,790	1,229	0,790	0,058	0,217	0,626	0,646	2,071	0,951
200	1,246	0,788	1,221	0,786	0,052	0,218	0,680	0,635	2,425	0,933

Tabela B2- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^*$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição **Logística** (τ = 2,00)e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	2,120	1,189	2,019	1,165	0,722	0,784	0,253	0,386	5,459	2,675
20	2,206	1,281	2,146	1,263	0,500	0,580	0,624	0,597	4,950	2,146
30	2,039	1,311	1,990	1,297	0,290	0,519	0,256	0,746	4,144	2,195
50	1,814	1,352	1,797	1,340	0,270	0,446	0,619	0,900	3,462	1,910
100	1,866	1,363	1,831	1,357	0,183	0,420	0,779	0,986	3,383	1,742
150	2,004	1,372	1,987	1,379	0,141	0,403	1,030	1,078	3,190	1,654
200	2,035	1,380	2,017	1,376	0,128	0,391	1,070	1,155	3,264	1,668

Tabela B3- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^{*}$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição de **Laplace** (τ = 1)e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$
10	1,461	0,872	1,368	0,843	0,705	0,112	0,162	0,238	4,413	2,207
20	1,442	0,936	1,401	0,917	0,443	0,052	0,373	0,482	3,479	1,942
30	1,314	0,975	1,263	0,960	0,261	0,034	0,462	0,481	3,105	1,621
50	1,072	0,979	1,042	0,974	0,100	0,021	0,382	0,616	2,351	1,453
100	1,076	0,996	1,058	0,992	0,068	0,009	0,403	0,711	2,025	1,393
150	1,118	0,991	1,106	0,989	0,065	0,007	0,555	0,767	1,993	1,263
200	1,099	0,993	1,093	0,992	0,053	0,005	0,557	0,788	1,942	1,223

Tabela B4- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^{*}$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição **0,85 Normal(0,1)+0,15 Normal(0,49)** (τ= 1,439) e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$	$\hat{\tau}$	$\hat{\tau}^{*}$
10	1,871	1,340	1,550	1,152	1,884	0,568	0,232	0,293	9,958	6,001
20	1,640	1,468	1,560	1,387	0,393	0,293	0,425	0,351	8,533	3,838
30	1,518	1,498	1,501	1,443	0,192	0,206	0,530	0,475	3,465	3,394
50	1,325	1,484	1,282	1,455	0,144	0,121	0,314	0,636	2,696	2,935
100	1,338	1,490	1,313	1,479	0,104	0,058	0,561	0,883	2,501	2,334
150	1,450	1,500	1,435	1,493	0,072	0,043	0,721	0,942	2,558	2,120
200	1,451	1,502	1,441	1,500	0,067	0,031	0,706	0,975	2,432	2,053

Tabela B5- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^*$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição 0,80 Normal(0,1)+0,20 Normal(0,49) (τ = 1,513) e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	2,182	1,624	1,716	1,414	2,933	0,638	0,259	0,249	12,5	5,204
20	1,808	1,69	1,677	1,626	0,684	0,427	0,393	0,541	9,292	4,419
30	1,610	1,498	1,558	1,498	0,221	0,270	0,606	0,475	4,200	3,394
50	1,410	1,727	1,399	1,697	0,153	0,193	0,480	0,803	2,883	3,244
100	1,405	1,739	1,367	1,719	0,111	0,126	0,671	1,069	2,738	2,830
150	1,519	1,737	1,499	1,729	0,083	0,098	0,759	1,146	2,488	2,617
200	1,523	1,745	1,511	1,740	0,075	0,092	0,784	1,212	2,383	2,480

Tabela B6- Valores de estatísticas descritivas observados nas distribuições dos estimadores $\hat{\tau}^*$ e $\hat{\tau}$ geradas nos estudos de simulação a partir de modelos com uma variável independente, erros com distribuição de **Cauchy** (τ = 1,571)e diferentes tamanhos de amostra.

n	média		mediana		erro quadrático médio		mínimo		máximo	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	3,480	7,39	2,458	2,070	20,15	2594,0	0,294	0,340	48,505	1203,7
20	2,375	22,8	2,147	2,800	1,794	182089	0,483	0,400	8,343	13019,1
30	1,931	14,74	1,863	3,010	0,587	51611	0,533	0,740	5,191	7064,1
50	1,578	6,902	1,543	3,608	0,168	348	0,533	0,941	3,920	285,1
100	1,530	41,7	1,509	3,700	0,117	1281098	0,684	1,500	2,790	35792,4
150	1,629	7,866	1,620	3,955	0,097	581	0,816	1,689	2,782	488,8
200	1,615	7,747	1,603	4,205	0,086	375	0,819	1,819	2,787	335,2

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2000-1 - **POPOV, S.Y., MACHADO, F.P.** One-dimensional branching random walk in a periodic random environment. 2000. 10p. (RT-MAE-2000-1)

2000-2 - **BORGES, W.S., HO L.L.; TURNES, O.** An analysis of taguchi's on-line quality monitoring procedure for attributes with diagnosis errors. 2000. 23p.(RT-MAE-2000-2)

2000-3 - **CORDEIRO, G.M., FERRARI, S.L.P., URIBE-OPAZO, M.A.** Bartlett-type corrections for two-parameter exponential family models. 2000. 21p. (RT-MAE-2000-3)

2000-4 - **SCARPA, O.** On coincidence of critical parameters in Poisson percolation model. 2000. 8p. (RT-MAE-2000-4)

2000-5 - **CANCHO, V.G., BOLFARINE, H.** Modelling the presence of immunes by using the exponentiated-weibull model. 2000. 18p. (RT-MAE-2000-5)

2000-6 - **BORGES, W., DIMITROV, B., KHALIL, Z., KOLEV, N.** System's performance under mixed minimal and imperfect repair maintenance strategies. 2000. 17p. (RT-MAE-2000-6)

2000-7 - **BARROSO, L.P., CORDEIRO, G.M., VASCONCELLOS, K.L.P.** Second-Order Asymptotics for score tests in heteroscedastic *t* regression models. 2000. 31p. (RT-MAE-2000-7)

2000-8 - **PEREIRA, C.A.B., NAKANO, F., STERN, J.M.** Actuarial Analysis via Branching Processes. 2000. 7p.(RT-MAE-2000-8)

2000-9 - **VIANA, M.A.G., PEREIRA, C.A.DE B.** Statistical Assessment of Jointly Observed Screening Tests. 2000 13p. (RT-MAE-200-9)

2000-10 - **MADRUGA, M.R., ESTEVES, L.G., WECHSLER, S.** On the Bayesianity of Pereira-Stern Tests. 2000. 9p. (RT-MAE-2000-10)

2000-11 - **PEREIRA, C.A., STERN, J.M.** *Full Bayesian Significance Test for Coefficients of Variation. 2000. 7p. (RT-MAE-2000-11)*

2000-12 - **BOLFARINE, H, CABRAL, C.R.B., PAULA, G.A.** Distance Tests Under Nonregular Conditions: Applications to the Comparative Calibration Model. 2000. 15p. (RT-MAE-2000-12)

2000-13 - **FERRARI, S.L.P., URIBE-OPAZO, M.** Corrected Likelihood Ratio Tests in a Class of Symmetric Linear Regression Models. 2000. 18p. (RT-MAE-2000-13)

2000-14 - **SVETLIZA, C.F., PAULA, G.A.** On Diagnostics in Log-Linear Negative Binominal Models. 2000. 17p. (RT-MAE-2000-14)

2000-15 - **FONTES, L.R.G., ISOPI, M., NEWMAN, C.M.** Random walks with strongly inhomogeneous rates and singular diffusions: convergence, localization and aging in one dimension. 2000. 22p. (RT-MAE-2000-15)

2000-16 - **SAHU, S.K., DEY, D.K., BRANCO, M.D.** A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models. 2000. 28p. (RT-MAE-2000-16)

2000-17 - **PEREIRA, C.A.B., STERN, J.M.** Model Selection: Full Bayesian Approach. 2000. 13p. (RT-MAE-2000-17)

2000-18 - **CORDEIRO, G.M., BOTTER, D.A.** Second-Order biases of maximum likelihood estimates in overdispersed generalized linear models. 2000. 16p. (RT-MAE-2000-18)

2000-19 - **GONZALEZ-LOPES, V.A., TANAKA, N.I.** Maximal Association, Fréchet Bounds and Homotopy in Prediction. 2000. 21p. (RT-MAE-2000-19)

2000-20 - **FERRARI, S.L.P., CRIBARI-NETO, F.** Corrected modified profile likelihood heteroskedasticity tests. 2000. 11p. (RT-MAE-2000-20)

2000-21 - **ESTEVES, L.G., WECHSLER, S., IGLESIAS, P.L.** De Finettian theorems for planar uniform models. 2000. 40p. (RT-MAE-2000-21)

2000-22 - **BUENO, V.C.** Modeling a coherent system at critical level. 2000. 9p. (RT-MAE-2000-22)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil