

INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS DE REGRESSÃO L_1 :
PEQUENAS AMOSTRAS

Este exemplar corresponde a reda
ção final da tese devidamente cor
rigida e defendida pelo Sr. Pedro
Ferreira Filho e aprovada pela Co
missão Julgadora.

Campinas, 04 de fevereiro de 1988


Profª Dra. Gabriela Stangenhuis
Orientadora

Dissertação apresentada ao Institu
to de Matemática, Estatística e
Ciência da Computação, UNICAMP, co
mo requisito parcial para obtenção
do Título de Mestre em Estatística.

A

Diva Lourdes M. Ferreira

(in memoriam)

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

À Profª Dra. Gabriela Stangenhauß pela orientação e incentivo durante o desenvolvimento deste trabalho.

Ao Prof. Dr. Subhash C. Narula pelas sugestões apresentadas.

À Estela pelo apoio, carinho e compreensão neste período e sempre.

Aos colegas pela amizade, incentivo e agradável convivência.

Aos companheiros do futebol pelas horas de lazer e diversão.

À M. Fatima Leticio pelo rápido e eficiente trabalho de datilografia.

Ao CNPq, UNICAMP e CAPES pelo apoio financeiro.

R E S U M O

Neste trabalho investigamos a viabilidade do uso do método bootstrap (Efron (1979)) como procedimento para obtenção do desvio padrão dos estimadores e construção de intervalos de confiança para os parâmetros da regressão L_1 quando o tamanho de amostra é pequeno.

Inicialmente apresentamos o problema de regressão L_1 como um problema de programação linear, suas propriedades e algoritmos propostos para obtenção das estimativas dos parâmetros.

Após, são apresentados os resultados inferenciais já conhecidos, o método bootstrap e sua aplicação ao caso de regressão L_1 .

A verificação da validade do método proposto é feita através de um estudo computacional que é descrito e cujos resultados são apresentados obtendo-se então uma conclusão final.

S U M M A R Y

In this work we investigate the possibility to use bootstrap method (Efrom (1979)) for to obtain estimates for standard deviation and construction of the confidence intervals on the parameters in the L_1 regression model when we have small sample size.

Firstly, we present, L_1 regression as one linear programming problem, properties and algorithms for to obtain the estimates.

After, inference results, bootstrap method and its application in the L_1 regression are presented.

Finally, we describe the computational study realized, the results obtained and final conclusion.

Í N D I C E

CAPÍTULO 1 - INTRODUÇÃO	1
CAPÍTULO 2 - REGRESSÃO L_1 COMO PROBLEMA DE PROGRAMAÇÃO <u>LI</u> NEAR	7
2.1 - Formulação do Problema	7
2.2 - Definições	9
2.3 - Propriedades dos Estimadores de Regressão L_1 ..	10
2.4 - Estimação dos Parâmetros da Regressão L_1	16
2.4.1 - Algoritmo de Barrodale e Roberts	18
2.4.2 - Algoritmo de Armstrong, Frome e Kung (AFK)	21
2.4.3 - Comparação Entre os Algoritmos	23
CAPÍTULO 3 - INFERÊNCIA EM REGRESSÃO L_1	27
3.1 - Introdução	27
3.2 - Teoria Assintótica	30
CAPÍTULO 4 - BOOTSTRAP E REGRESSÃO L_1	37
4.1 - Método Bootstrap	37
4.2 - Aplicação do Método Bootstrap à Regressão L_1 ..	40
CAPÍTULO 5 - METODOLOGIA E RESULTADOS	44
5.1 - Metodologia	44
5.1.1 - Estudo de Monte Carlo	46
5.1.2 - Método Bootstrap	47
5.2 - Resultados Para Reamostragem nos Pares	48
5.3 - Resultados Para Reamostragem nos Erros	62

CAPÍTULO 6 - CONCLUSÕES 71

BIBLIOGRAFIA 73

C A P Í T U L O 1

INTRODUÇÃO

A análise estatística de um conjunto de dados pode ser realizada através de uma variedade de métodos existentes. A escolha do método a ser utilizado deve estar de acordo com o tipo de informação disponível e os objetivos que desejamos alcançar.

Um dos métodos estatísticos utilizados com grande frequência em diferentes áreas e que conseqüentemente tem recebido especial atenção por parte dos pesquisadores em estatística é a análise de regressão. Técnica simples através da qual procuramos expressar o relacionamento entre um conjunto de variáveis. O seu principal objetivo é obter uma função que esteja o mais próximo possível dos pontos observados de acordo com um critério pré-estabelecido. O tipo de função a ser utilizada varia a cada caso, podendo ser sugerida por experiências anteriores com o mesmo tipo de informação, por evidências teóricas que estamos interessados em investigar, ou mesmo pela amostra observada. Num estudo deste tipo, torna-se de fundamental importância a análise a posteriori dos resultados obtidos a fim de se verificar a validade do modelo proposto.

O modelo usado com maior frequência é o linear:

$$Y = X\beta + \varepsilon \quad (1.1)$$

onde:

$Y = (y_i)$ $i=1,2,\dots,n$ é o vetor da variável resposta.

$X = (x_{ij})$ $i=1,2,\dots,n$ $j=1,2,\dots,m$; matriz das variáveis preditoras.

$\beta = (\beta_j)$ $j=1,2,\dots,m$ vetor de parâmetros desconhecidos.

$\epsilon = (\epsilon_i)$ $i=1,2,\dots,n$ vetor de erros aleatórios, independentes e identicamente distribuídos de acordo com uma determinada distribuição F .

Utilizando este modelo procuramos estimar o vetor de parâmetros desconhecidos β , de tal forma que o hiperplano ajustado esteja o mais próximo possível do vetor Y de observações. A medida de proximidade é, comumente estabelecida em termos da norma L_p , ou seja, obter um vetor $\hat{\beta}$ tal que:

$$\sum_{i=1}^n |y_i - x_i \hat{\beta}|^p \quad (1.2)$$

seja mínima, para $p \geq 1$, onde x_i é a i -ésima linha da matriz X .

Os valores de p mais usados são $p=1,2,\infty$. Para $p=1$ temos o critério da minimização dos erros absolutos; para $p=2$ o critério dos mínimos quadrados e $p=\infty$ o critério do erro máximo absoluto (estimador minimax ou Chebyshev). Dentre estes critérios o mais utilizado na análise de regressão é o método dos mínimos quadrados. Isso se deve em grande parte a simplicidade computacional e matemática do mesmo, e o que ainda é mais importante, está presente em praticamente todos pacotes de programas estatísticos disponíveis no mercado. Além disso o as

pecto inferencial é amplamente desenvolvido e divulgado. Entretanto, desvios da hipótese de que os erros são normalmente distribuídos e a presença de valores aberrantes podem fortemente afetar a análise através do método dos mínimos quadrados. Rice e White (1964) mostraram, através de um estudo de Monte Carlo, que se os erros se afastarem da hipótese de normalidade e se sua distribuição possuir cauda mais alongada do que a distribuição normal, o critério de minimização dos erros absolutos dá origem a estimadores com menor erro quadrático médio do que aqueles obtidos pelo método dos mínimos quadrados. Hogg (1979) sugeriu que se deveria dar maior atenção aos métodos robustos de estimação em análise de regressão. Ele recomenda ao ajustar um modelo do tipo (1.1) utilizar o método dos mínimos quadrados e também procedimentos robustos. Caso o conjunto de observações não satisfaça as hipóteses dos mínimos quadrados devemos investigá-los detalhadamente a fim de verificar valores aberrantes e dados que podem ter influência muito grande nas estimativas dos mínimos quadrados. Quando valores aberrantes ou valores influentes são encontrados, estes devem ser examinados cuidadosamente. Se são dados de má qualidade, ou erro de codificação, por exemplo, podemos corrigi-los ou abandoná-los. Se forem resultantes do processo de geração dos dados, eles devem ser mantidos na análise e então aplica-se um método robusto tal como o critério de minimização dos erros absolutos.

O critério de minimização dos erros absolutos (MSAE), ou mínimo valor absoluto (LAV), ou mínimos desvios absolutos (LAD), ou critério L_1, \dots , por ser resistente a pontos aberrantes é considerado uma boa alternativa robusta ao método dos mínimos quadrados. De acordo com Eisenhart (1961), este crité

rio foi sugerido por Boscovich (1757) antes mesmo de Legendre (1805) propor o método dos mínimos quadrados.

A partir da proposta de Boscovich alguns estudos foram realizados a cerca deste critério, entre eles, Fourier (1820), Edgeworth (1877) e Jackson (1921), porém, todos eles esbarraram nas dificuldades computacionais que se apresentavam. Este problema fez com que por várias décadas, este critério não fosse mais utilizado. Foi a partir do desenvolvimento da teoria da programação linear, especialmente do método simplex, que este aspecto começou a ser resolvido. Charnes, Cooper e Fergusson (1955) apresentaram o problema de análise de regressão sob critério de minimização dos erros absolutos formulado como um problema de programação linear. Este fato somado ao surgimento de computadores de grande porte, propiciou as condições necessárias para um melhor desenvolvimento da regressão sob o critério L_1 . Uma série de algoritmos foram propostos para estimar os parâmetros da regressão L_1 (Barrodale e Young (1966), Barrodale e Roberts (1973), Abdelmalek (1975), Armstrong, Frome e Kung (1979), Bloomfield e Steiger (1980), entre outros). Porém, um outro problema tão importante quanto o de estimação dos parâmetros da regressão L_1 , fez com que os pesquisadores relutassem em utilizar este critério: a falta de conhecimento da distribuição dos estimadores.

Vários estudos de Monte Carlo foram realizados abordando este problema (Ashar e Wallace (1963), Rice e White (1964), Rosemberg e Carlson (1970), Blattberg e Sargent (1971), Bourdon (1974)), porém o primeiro resultado indicando a forma desta distribuição somente foi apresentada por Rosemberg e

Carlson em 1977. Num estudo de Monte Carlo verificaram as propriedades para pequenas amostras dos estimadores da regressão L_1 no caso de erros simetricamente distribuídos. Concluíram que se a curtose da distribuição dos erros é grande, os estimadores L_1 têm distribuição muito próxima a normal e possui um erro médio quadrático significativamente menor que o estimador do método dos mínimos quadrados. Bassett e Koenker (1978) mostraram analiticamente que o estimador L_1 é consistente, não viciado e que assintoticamente segue a distribuição normal multivariada encontrada por Rosemberg e Carlson. A partir desse resultado foi possível desenvolver testes estatísticos e intervalos de confiança assintóticos para os parâmetros da regressão L_1 .

A convergência para esta distribuição assintótica foi estudada por Dielmann e Pfaffemberger (1982) sendo que a conclusão obtida foi de que para casos em que a distribuição dos erros se aproxima da distribuição normal, a convergência ocorre rapidamente. Porém, quando a distribuição dos erros apresenta caudas mais alongadas do que a distribuição normal, tal como as distribuições de Laplace e Cauchy, esta convergência somente começa a ocorrer a partir de tamanhos de amostras maiores do que 100. Portanto, o uso desta distribuição assintótica pode acarretar graves problemas quando utilizada em pequenas amostras.

Surge daí a necessidade do desenvolvimento de procedimentos estatísticos para pequenas amostras. Neste trabalho estudamos a viabilidade do uso do método bootstrap (Efrom (1979)) como procedimento para a determinação do desvio padrão dos estimadores e construção de intervalos de confiança para os parâmetros da regressão L_1 , quando o tamanho de amostra é pequeno.

Consideramos distribuições de erros simétricas como a normal padrão, normal contaminada, Laplace e Cauchy e diferentes tamanhos de amostra.

No Capítulo 2 abordamos o problema de regressão L_1 , visto através de um problema de programação linear, as propriedades dos estimadores bem como alguns algoritmos propostos para resolução do problema.

No Capítulo 3 apresentamos os estudos realizados com relação ao comportamento assintótico dos estimadores da regressão L_1 .

O método bootstrap e sua aplicação à regressão L_1 são expostos no Capítulo 4.

No Capítulo 5 descrevemos o estudo computacional utilizado para a verificação da nossa proposta e também apresentamos os resultados obtidos neste sentido.

As conclusões finais são apresentadas no Capítulo 6.

C A P Í T U L O 2

REGRESSÃO L_1 COMO PROBLEMA DE PROGRAMAÇÃO LINEAR

Neste capítulo apresentamos o problema de regressão L_1 sob o ponto de vista de um problema de programação linear, de definições, propriedades e algoritmos propostos para a obtenção das estimativas dos parâmetros.

2.1 - FORMULAÇÃO DO PROBLEMA

O problema de regressão L_1 pode ser visto como um problema de programação linear. Este aspecto tem sido exaustivamente apresentado na literatura (Charnes, Cooper e Fergusson (1955), Wagner (1959), Barrodale e Young (1966), Sposito (1976) entre outros). Esta abordagem do problema de regressão L_1 facilitou a obtenção de propriedades da regressão L_1 sendo também a base para os melhores algoritmos de estimação dos parâmetros.

Consideremos o modelo de regressão (1.1). O problema de regressão L_1 consiste em determinar o vetor de estimativas $\hat{\beta}$ para o vetor de parâmetros desconhecidos β , tal que:

$$\sum_{i=1}^n |Y_i - X_i \hat{\beta}| \quad (2.1)$$

seja mínima.

Sejam u_i e v_i os desvios absolutos positivos e negativos respectivamente em relação ao hiperplano ajustado, $\hat{Y} = X\hat{\beta}$. Para cada uma das observações, temos então que:

$$Y_i = \sum_{j=1}^m X_{ij} \hat{\beta}_j + u_i - v_i \quad i=1,2,\dots,n \quad (2.2)$$

Assim $(u_i + v_i)$ é o desvio absoluto com relação a i -ésima observação. Isto leva-nos à formulação do seguinte problema de programação linear:

$$\begin{aligned} \min \quad & 1'u + 1'v & u &= (u_1, u_2, \dots, u_n)' \\ & & v &= (v_1, v_2, \dots, v_n)' \end{aligned} \quad (2.3)$$

sujeito a

$$X\hat{\beta} + Iu - Iv = Y \quad u, v \geq 0$$

onde 1 é um vetor coluna unitário n -dimensional e I representa uma matriz identidade de ordem n .

O problema dual de (2.3) é dado por

$$\max \quad f'Y \quad (2.4)$$

sujeito a

$$X'f = 0$$

$$f = (f_1, \dots, f_n)', \quad f_i \in [-1, 1] \quad i=1,2,\dots,n.$$

Uma relação importante entre o problema primal e dual é dado pelo Teorema da Folga Complementar através do qual temos as condições onde a variável primal ótima não negativa assume valor zero ou o seu dual correspondente é satisfeito por

uma igualdade.

Para o problema (2.3) temos:

$$\begin{aligned} \text{i) } u_i &= 0 \quad \text{ou} \quad f_i = 1 \\ \text{ii) } v_i &= 0 \quad \text{ou} \quad f_i = -1 \end{aligned} \tag{2.5}$$

Conseqüentemente se o i -ésimo ponto está acima do h_i hiperplano ótimo, $u_i > 0$ e $f_i = 1$ e se está abaixo $v_i > 0$ e $f_i = -1$.

Este teorema e outros resultados fundamentais de programação linear são diretamente aplicados ao problema de regressão L_1 e, por serem amplamente encontrados na literatura não serão aqui apresentados.

2.2 - DEFINIÇÕES

DEFINIÇÃO 1:

Qualquer (β, u, v) que satisfaça $X\beta + Iu - Iv = Y$ é denominada uma solução para (2.3).

DEFINIÇÃO 2:

Qualquer solução $\bar{w} = (\beta, u, v)$ para (2.3) que também satisfaça $\bar{w}_j \geq 0$ para $j = m+1, \dots, m+1+2n$ é denominada uma solução possível para o problema (2.3).

DEFINIÇÃO 3:

Qualquer solução possível \bar{w} é chamada uma solução básica possível se as colunas de $A = (X \ I \ -I)$ correspondentes a componentes de \bar{w} diferentes de zero, formam um conjunto de vetores linearmente independentes no R^n .

DEFINIÇÃO 4:

Um ponto $\bar{w} \in F = \{\bar{w} \mid A\bar{w} = Y, \bar{w}_j \geq 0, j = m+1, \dots, m+1+2n\}$ é chamado um ponto extremo de F se não existem w_1 e $w_2 \in F$, $w_1 \neq w_2$, $0 < \lambda < 1$ tal que $\bar{w} = \lambda w_1 + (1-\lambda)w_2$.

DEFINIÇÃO 5:

Qualquer $\bar{w} \in F$, que minimiza $\sum_{i=1}^n (u_i + v_i)$ é chamado uma solução ótima do problema (2.3) e $X'\bar{w}$ é denominado hiperplano L_1 ótimo.

2.3 - PROPRIEDADES DOS ESTIMADORES DE REGRESSÃO L_1

Quando a matriz X das variáveis preditoras tem posto completo, temos as seguintes propriedades associadas às soluções do problema (2.3).

PROPRIEDADE 1:

Existe pelo menos um hiperplano L_1 ótimo que passa através de no mínimo m pontos observados.

Demonstração:

Em (2.3) existem n equações, $2n$ semi-espacos e $2n+m$ variáveis. Se um ponto não pertence ao hiperplano temos que u_i ou v_i é diferente de zero. Mas em um dos pontos extremos $n+m$ das $2n$ variáveis devem ser iguais a zero de tal forma que junto com as n equações fique definido um ponto extremo. Consequentemente, no máximo $n-m$ pontos estão fora do hiperplano correspondente a esse ponto extremo.

□

PROPRIEDADE 2:

Uma solução para o problema (2.3) é um hiperplano tal que a diferença absoluta entre o número de pontos com desvios positivos e negativos com relação a esse hiperplano nunca supera m .

Demonstração:

Sejam n_1 e n_2 o número de desvios absolutos positivos e negativos, respectivamente, em relação ao hiperplano L_1 ótimo. Pelo teorema da folga complementar temos:

$$\text{i) } u_i = 0 \quad \text{ou} \quad f_i = 1 \quad \text{e}$$

$$\text{ii) } v_i = 0 \quad \text{ou} \quad f_i = -1$$

Pela primeira restrição do problema dual (2.4),

$$\sum f_i = \sum_{i \in I_1} f_i + \sum_{i \in I_2} f_i + \sum_{i \in I_3} f_i = 0$$

onde I_1, I_2, I_3 representam respectivamente o conjunto de pontos com resíduos positivos, negativos e nulos em relação ao hiperplano L_1 ótimo.

Usando (i) e (ii) com $f_i \in [-1, 1]$ temos:

$$|n_1 - n_2| = \left| \sum_{i \in I_3} f_i \right| \leq m$$

□

Com estas duas propriedades podemos restringir a procura de um hiperplano ótimo a um subconjunto de $\binom{n}{m}$ hiperplanos.

No caso da matriz X não possuir posto completo, estas duas propriedades podem ser facilmente extendidas. Considere

mos que o posto da matriz X seja $K \leq m$. Na primeira propriedade de termos pelo menos um hiperplano L_1 ótimo que passa através de no mínimo K pontos observados. Na segunda propriedade, se o número de pontos no hiperplano L_1 ótimo for n_K , então a diferença absoluta entre o número de pontos acima e abaixo deste hiperplano não excederá a este valor.

PRÓPRIEDADE 3:

Se um conjunto de observações possuir múltiplos hiperplanos L_1 ótimos, então qualquer combinação linear convexa desses hiperplanos também é ótimo.

Demonstração:

Sejam b_1 e b_2 dois vetores L_1 ótimos e v a soma dos desvios absolutos.

Seja b^* uma combinação linear convexa de b_1 e b_2 , então:

$$b^* = \alpha b_1 + (1-\alpha)b_2 \quad \alpha \in [0,1]$$

e

$$\begin{aligned} \sum_{i=1}^n |Y_i - \sum_{j=1}^m X_{ij} b_j^*| &= \sum_{i=1}^n (u_i^* + v_i^*) \\ &= \sum_{i=1}^n [\alpha(u_{1i} + v_{1i}) + (1-\alpha)(u_{2i} + v_{2i})] \\ &= \alpha \sum_{i=1}^n (u_{1i} + v_{1i}) + (1-\alpha) \sum_{i=1}^n (u_{2i} + v_{2i}) \\ &= \alpha v + (1-\alpha)v \\ &= v \end{aligned}$$

Logo o hiperplano b^* também é L_1 ótimo.

□

PROPRIEDADE 4:

Se um hiperplano L_1 não passa por qualquer ponto observado então $|n_1 - n_2| = 0$.

Demonstração:

Suponhamos que $\hat{Y} = X'\hat{\beta}$ é um hiperplano que não passa por qualquer ponto observado e que $n_1 > n_2$.

$$\text{Seja } D = \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

Consideremos $\hat{\beta}^*$ um vetor cujos elementos são idênticos aos de $\hat{\beta}$ exceto para $\hat{\beta}_1^* = \hat{\beta}_1 + \epsilon$, onde ϵ é um número positivo menor ou igual ao mínimo desvio positivo, então:

$$\begin{aligned} D^* &= \sum_{i=1}^n |Y_i - Y_i^*| \\ &= \sum_{i=1}^n |Y_i - \hat{Y}_i - \epsilon| \\ &\leq \sum_{i=1}^n |Y_i - \hat{Y}_i| - (n_1 - n_2)\epsilon \\ &< D \end{aligned}$$

Portanto \hat{Y} não é um hiperplano L_1 ótimo.

Da mesma forma mostramos que se $n_1 < n_2$, então o hiperplano não é ótimo.

□

PROPRIEDADE 5:

Se o número de observações é ímpar, então qualquer hiperplano L_1 ótimo passa por pelo menos 1 ponto observado.

Demonstração:

Esta propriedade é consequência imediata do teorema anterior. Se o hiperplano não passa por nenhum ponto observado então $|n_1 - n_2| = 0$. Como n é ímpar isto não é possível, logo necessariamente o hiperplano passa por pelo menos um ponto observado.

□

PROPRIEDADE 6:

Se n é ímpar, todos hiperplanos L_1 ótimos passam através de um mesmo ponto observado.

Demonstração:

Suponhamos dois hiperplanos L_1 ótimos que não possuam um ponto em comum e que sejam paralelos. Logo, podemos encontrar um terceiro hiperplano L_1 ótimo que seja uma combinação linear desses dois e que não passa por nenhum ponto observado, contradizendo a Propriedade 5.

Se os pontos não são paralelos e a intersecção não é um ponto em comum, podemos encontrar um hiperplano L_1 ótimo, combinação linear dos hiperplanos considerados, que não passa por qualquer ponto observado, contradizendo novamente a Propriedade 5.

□

PROPRIEDADE 7:

Nenhum ponto observado poderá ter desvio positivo com relação a determinado hiperplano e um desvio negativo com relação a outro hiperplano.

Demonstração:

Sejam dois hiperplanos L_1 ótimos.

$$Y = X'b \quad (H_1)$$

$$Y = X'c \quad (H_2)$$

Suponhamos que exista um ponto observado (x_j, y_j) entre eles. Existe um hiperplano que passa por (x_j, y_j) que pode ser escrito como combinação linear convexa de H_1 e H_2 .

$$Y^* = \alpha X'b + (1-\alpha)X'c \quad (H_3)$$

O restante dos pontos tem desvios positivos em relação a H_1 ou negativos com relação a H_2 . Escrevendo a soma dos desvios absolutos com relação a H_3 temos:

$$\begin{aligned} & \sum_{i \neq j} |Y_i - (\alpha X'b + (1-\alpha)X'c)| \\ &= \sum_{i \neq j} |\alpha(Y_i - X'b) + (1-\alpha)(Y_i - X'c)| \\ & \stackrel{(A)}{=} \alpha \sum_{i \neq j} |Y_i - X'b|^+ + (1-\alpha) \sum_{i \neq j} |Y_i - X'c|^+ \\ & + \alpha \sum_{i \neq j} |Y_i - X'b|^- + (1-\alpha) \sum_{i \neq j} |Y_i - X'c|^- \\ &= \alpha \sum_{i \neq j} |Y_i - X'b| + (1-\alpha) \sum_{i \neq j} |Y_i - X'c| \end{aligned}$$

$$< \alpha \sum_{i=1}^n |Y_i - X' b| + (1-\alpha) \sum_{i=1}^n |Y_i - X' c|$$

$$(B) \sum_{i=1}^n |Y_i - X' c| = \sum_{i=1}^n |Y_i - X' b|$$

Temos (A) pelo fato dos desvios serem sempre tomados em módulo e serem positivos ou nulos com relação a H_1 ou negativos ou nulos em relação a H_2 , e (B) pelo fato dos dois hiperplanos H_1 e H_2 serem ótimos.

Portanto chegamos a um absurdo, pois toda combinação linear convexa de hiperplanos ótimos também é ótima.

□

2.4 - ESTIMAÇÃO DOS PARÂMETROS DA REGRESSÃO L_1

A maior dificuldade para a utilização do critério de minimização dos erros absolutos na análise de regressão foi, durante longo tempo, a falta de algoritmos rápidos e eficientes para obter a estimativa do vetor de parâmetros desconhecidos β .

Desde o século passado foram realizados estudos em busca de procedimentos para solução do problema de regressão L_1 . Trabalhos como o de Fourier (1820) e Edgeworth (1887) não puderam ser melhor desenvolvidos devido às dificuldades computacionais da época.

O desenvolvimento da programação linear e especialmente do método simplex possibilitaram o surgimento de um tratamento computacional mais adequado ao problema de regressão L_1 . O primeiro trabalho nesta linha foi apresentado por Singleton

em 1940, ainda sem o conhecimento do método simplex, mas com toda a estrutura do mesmo. Charnes, Cooper e Fergusson em 1955, apresentaram o primeiro trabalho utilizando as técnicas de programação linear para solução do problema de regressão L_1 .

A partir daí, durante os últimos 30 anos, vários métodos e algoritmos foram propostos em sua maioria baseados em técnicas de programação linear, alguns abordando o problema sob o ponto de vista primal e outros sob o ponto de vista dual. Dentre estes trabalhos temos:

- . Karst (1958): Método iterativo para estimação dos parâmetros da reta de melhor ajuste em duas situações:
 - 1) a reta deve passar num determinado ponto.
 - 2) Sem nenhuma restrição no ajuste.
- . Barrodale e Young (1966): Algoritmo usando uma modificação do método simplex.
- . Schlossmacher (1973): Método utilizando iterativamente o método dos mínimos quadrados ponderados.
- . Abdelmalek (1975): Método baseado no algoritmo simplex dual utilizando a decomposição LU para obter uma melhor estabilidade numérica.
- . Armstrong e Kung (1978): Modificação do algoritmo de Barrodale e Roberts (1973) para o caso específico de regressão linear near simples.
- . Bloomfield e Steiger (1980): Extensão do método proposto por Karst. Utiliza os mesmos procedimentos de pivotamento do método

todo simplex.

- . Wesolowsky (1981): Baseado no princípio de caminho descendente ao longo da superfície da função objetivo.
- . Josvanger e Sposito (1983): Modificação do algoritmo de Wesolowsky para o caso de regressão linear simples.

Entre todos os métodos até aqui apresentados os que aparecem com maior destaque são os propostos por Barrodale e Roberts (1973) e Armstrong, Frome e Kung (1979). Apresentaremos, desta forma, cada um desses métodos mais detalhadamente.

2.4.1 - ALGORÍTMO DE BARRODALE E ROBERTS

Este algoritmo, proposto em 1973, é baseado no método simplex e supera o algoritmo de Barrodale e Young (1966) pelo fato de em uma mesma iteração poder passar por vários vértices do conjunto de soluções possíveis, reduzindo significativamente o número de iterações necessárias para encontrar a solução ótima do problema.

Reescrevendo (2.3) obtemos:

$$\min 1'u + 1'v \quad (2.6)$$

sujeito a $X(b-c) + Iu - Iv = Y$

$$b, c, u, v \geq 0$$

onde $(b-c) = \beta$.

Temos então uma solução básica formada por u_1, u_2, \dots, u_n caso todos os Y_i sejam não negativos. Para o caso de algum valor negativo troca-se o sinal da linha correspon

dente e o elemento u_i desta linha é substituído pelo v_i correspondente na base.

Sabemos que $b_j = -c_j$, $u_i = -v_i$ e que a soma dos custos marginais de b_j e c_j é zero e de u_i e v_i é -2 . A partir destes resultados podemos construir um quadro inicial do método simplex condensado para o algoritmo, assumindo todos valores Y_i não negativos.

BASE	R	b_1	b_2	b_n
U_1	Y_1	x_{11}	x_{21}	x_{n1}
U_2	Y_2	x_{12}	x_{22}	x_{n2}
\vdots	\vdots	\vdots	\vdots	\vdots
U_m	Y_m	x_{1m}	x_{2m}	$x_{n,m}$
CUSTOS	$\sum_{i=1}^m Y_i$	$\sum_{i=1}^m x_{1i}$	$\sum_{i=1}^m x_{2i}$	$\sum_{i=1}^m x_{ni}$

O algoritmo é implementado em dois estágios. No estágio 1, a escolha da coluna pivô durante as iterações restringe-se aos vetores b_j e c_j . O vetor escolhido para entrar na base é aquele que apresenta maior custo marginal não negativo. O vetor escolhido para sair da base é aquele que proporciona maior redução na função objetivo.

Pode ocorrer que nenhum vetor seja encontrado para sair da base, isto é, a coluna pivotal escolhida, pode conter somente elementos não positivos; isto ocorre somente quando a matriz X não tem posto completo. Nesse caso ignora-se a coluna pivô b_j (ou c_j) e não realiza-se a transformação simplex.

Ao final do estágio 1 o posto da matriz X é determi

nado pelo número total de vetores b_j e c_j na base. Considerando que K dos vetores u_i (ou v_i) foram retirados da base, o quadro neste instante representa uma aproximação que interpola K pontos observados (Propriedade 1). Quando a aproximação interpolar mais do que K pontos, o quadro ficará degenerado, porém isto na prática não causa problemas.

No estágio 2 ocorre a troca dos vetores u_i e v_i não básicos com os vetores u_i e v_i básicos. Os vetores b_j ou c_j não podem deixar a base neste estágio. Novamente o vetor escolhido para entrar na base é aquele que apresenta maior custo marginal positivo e o vetor escolhido para sair da base é aquele que proporciona maior redução na função objetivo. O algoritmo termina quando todos os custos marginais são não positivos.

Neste segundo estágio, cada tabela do simplex corresponde a uma interpolação de K pontos observados. Em cada uma iteração $(K-1)$ pontos permanecem fixos, o vetor a ser retirado da base determina o ponto que será excluído da interpolação e conseqüentemente o vetor incluído na base determina o novo ponto de interpolação. Ao final deste estágio, ocorrendo valores negativos para alguns b_j ou c_j básicos, estes devem ser substituídos pelos correspondentes c_j ou b_j não básicos.

A principal modificação com relação ao método simplex padrão é a escolha do vetor u_i ou v_i a ser retirado da base. Em ambos os estágios, este vetor é escolhido como o que causa maior redução na função objetivo. Geometricamente a transformação proposta é equivalente a saltar vários vértices no método simplex padrão.

A regra para determinar o vetor para sair da base é

modificada da seguinte forma: primeiramente aplicamos a regra usual para determinar a linha pivô e subtraímos duas vezes o valor deste do custo marginal da coluna pivô. Se o resultado for não positivo fazemos a transformação simplex usando este pivô. Caso contrário, subtraímos duas vezes a linha pivô da linha de custos marginais, multiplicamos a linha pivô por -1 e trocamos o vetor u_i (ou v_i) da base, correspondente a linha pivô por v_i (ou u_i). A multiplicação pelo valor 2 é devida ao fato de que a soma dos custos marginais de cada par u_i ou v_i é -2 . Essa operação reduz o valor da função objetivo e troca o sinal do pivô. A regra normal para a escolha da linha pivô é novamente aplicada e um novo pivô é escolhido, repetindo-se este processo até que um pivô não seja rejeitado e a transformação simplex possa ser utilizada com este pivô.

2.4.2 - ALGORÍTMO DE ARMSTRONG, FROME E KUNG (AFK)

Este algoritmo apresentado em 1979 é uma modificação do algoritmo de Barrodale e Roberts. A principal diferença entre os dois algoritmos é que o de AFK utiliza o método simplex revisado com a decomposição LU na manutenção da base corrente. Porém, os critérios de entrada e saída de vetores da base é o mesmo.

O método simplex revisado é mais eficiente que o método simplex básico por não necessitar modificar toda a tabela simplex a cada iteração realizada, porém exige o cálculo de uma matriz inversa a cada entrada e saída de vetores de base.

Suponhamos o seguinte problema:

$$\max Z$$

(2.7)

sujeito a

$$Z - c'x = 0$$

$$Ax = b$$

$$x \geq 0$$

Existem $n+1$ variáveis, $m+1$ restrições e uma base de posto $m+1$.

Este problema pode ser reescrito como:

$$\max Z$$

sujeito a:

$$\left[\begin{array}{c|c} 1 & -c' \\ \hline 0 & A \end{array} \right] \begin{bmatrix} Z \\ X \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

Colocando-se as colunas de A de forma a obtermos uma base de posto m , podemos ter uma base de posto $m+1$ dada por:

$$B = \left[\begin{array}{c|c} 1 & -c'_B \\ \hline 0 & B_m \end{array} \right]$$

A inversa de B é dada por:

$$B^{-1} = \left[\begin{array}{c|c} 1 & c'_B B_m^{-1} \\ \hline 0 & B_m^{-1} \end{array} \right]$$

A partir de B^{-1} podemos determinar os valores da tabela seguinte do método simplex bem como a solução do problema neste instante. Portanto necessitamos atualizar esta matriz a cada iteração realizada.

Surge então a necessidade de um bom método de inver

são de matrizes para que o algoritmo tenha um bom desempenho. O algoritmo de AFK utiliza a decomposição LU para cálculo da matriz inversa inicial e também para sua atualização a cada iteração realizada. O resultado é um algoritmo rápido e eficiente com boa estabilidade numérica e que apresenta melhor desempenho que o algoritmo de Barrodale e Roberts.

2.4.3 - COMPARAÇÃO ENTRE OS ALGORÍTMOS

Diversos estudos foram realizados a fim de comparar os distintos algoritmos para a determinação da regressão L_1 .

Ao longo destes estudos verificamos que o desempenho dos algoritmos está relacionado a certas características do problema, entre as quais:

- i) número de observações;
- ii) número de parâmetros no modelo;
- iii) número de condição dos dados.

Os dois primeiros fatores acima podem ser facilmente controlados para efeito de comparação de algoritmos. Para o terceiro ponto, Kennedy, Gentle e Sposito (1977) e Hoffmann e Shier (1980) apresentaram métodos para geração aleatória de problemas testes em regressão L_1 com número de condição desejado e vetor de soluções conhecido.

Um estudo comparativo dos mais recentes algoritmos propostos para a solução do problema de regressão L_1 foi apresentado por Gentle, Narula e Sposito (1987). Neste trabalho foram estudados os algoritmos de Armstrong e Kung (1978) (AK), Josvanger e Sposito (1983) (JS), Abdelmalek (1980) (A), Armstrong,

Frome e Kung (1979) (AFK), Bloomfield e Steiger (1980) (BS). Destes algoritmos os de AK e JS são específicos para o caso de regressão linear simples. O algoritmo de Barrodale e Roberts não foi utilizado pelos autores devido a estudos anteriores (Armstrong, Frome e Kung (1979)) mostrando que sua performance é inferior ao algoritmo de AFK.

O desempenho dos algoritmos foi estudado sob diversas condições de número de observações (n) e número de parâmetros no modelo (m). Os valores da matriz de variáveis preditoras X foram geradas segundo uma distribuição uniforme $(0,1)$ e, para obtenção dos valores da variável dependente Y foram acrescentados erros segundo uma distribuição normal de média zero e desvio padrão σ assumindo diversos valores.

Os resultados obtidos são apresentados na Tabela 2.1 e 2.2. A comparação dos algoritmos é feita pelo tempo médio de CPU, em milisegundos, utilizados na solução de 100 problemas testes.

TABELA 2.1 - TEMPO MÉDIO CPU PARA MODELO LINEAR SIMPLES.

n	σ	AK	JS	A	AFK	BS
100	0.0	0.051	0.028	0.089	0.039	0.039
	3.0	0.021	0.023	0.094	0.034	0.023
500	0.0	0.702	0.560	1.287	0.345	0.229
	3.0	0.193	0.302	1.434	0.287	0.145
1000	0.0	2.420	2.070	4.309	0.881	0.565
	3.0	0.544	0.971	4.775	0.784	0.422
5000	0.0	17.068	10.148	198.21*	3.091	+
	3.0	1.262	2.837	211.23*	1.614	+

* Média de três execuções.

+ Não produz resposta correta.

TABELA 2.2 - TEMPO MÉDIO DE CPU PARA DIFERENTES VALORES DE m.

n	m	σ	A	AFK	BS
100	5	0.0	0.301	0.153	0.188
		3.0	0.331	0.149	0.114
	15	0.0	1.646	1.569	1.301
		3.0	1.976	1.313	0.933
500	5	0.0	3.483	1.156	1.182
		3.0	3.686	1.120	0.829
	15	0.0	15.978	8.349	9.714
		3.0	17.876	7.808	7.294
1000	5	0.0	11.253	2.876	+
		3.0	13.211	2.930	+
	15	0.0	43.547	18.096	+
		3.0	49.866	17.901	+
5000	5	0.0	118.81*	34.788	+
		3.0	248.91*	34.311	+
	15	0.0	706.74*	142.647	+
		3.0	687.31*	140.321	+

* Média de três execuções.

+ Não produz resposta correta.

Pelos resultados acima podemos verificar que os diferentes algoritmos utilizados não apresentam performance muito diferente para número de observações igual a 100. Porém, à medida que este número de observações cresce, o desempenho dos algoritmos difere bastante. Podemos notar que os algoritmos de AFK e BS apresentam uma estabilidade bem maior que os demais. O algoritmo de BS porém, apresenta problemas para grande número de observações principalmente quando o número de parâmetros no

modelo também é grande.

Portanto o algoritmo que se apresenta de maneira mais uniforme nas diferentes situações estudadas e que consequentemente é o mais recomendado numa situação geral é o algoritmo AFK.

C A P Í T U L O 3

INFERÊNCIA EM REGRESSÃO L_1

Neste capítulo abordamos aspectos da inferência em regressão L_1 , principalmente aqueles relacionados às propriedades assintóticas dos estimadores L_1 .

3.1 - INTRODUÇÃO

A inferência em regressão L_1 é ainda hoje uma área que necessita ser melhor estudada a fim de facilitar a utilização desta metodologia. Até pouco tempo atrás, por exemplo, não se conhecia qualquer tipo de informação com relação a distribuição dos estimadores dos parâmetros da regressão L_1 .

Os primeiros trabalhos nesta área foram baseados essencialmente em simulações e procuravam comparar a eficiência dos estimadores da regressão L_1 com relação aos tradicionais estimadores de regressão dos mínimos quadrados, sob as mais diversas distribuições de erros (Rice e White (1964), Blatteberg e Sargent (1971), Pffafemberger e Dinkel (1978), entre outros). Estes trabalhos indicaram que os estimadores de regressão L_1 são mais eficientes do que os estimadores obtidos pelo critério dos mínimos quadrados quando a distribuição dos erros apresenta caudas mais alongadas do que a distribuição normal. Ashar e Wallace (1963) e Wilson (1978) mostraram que os estimadores de

regressão L_1 têm uma eficiência relativa de 80% com relação aos estimadores de regressão L_2 quando as condições de mínimos quadrados são satisfeitas.

A consistência e a tendenciosidade dos estimadores de regressão L_1 foi estudada por Taylor (1974). Uma forma de consistência foi apresentada quando os resíduos são simetricamente distribuídos em torno de zero. A não tendenciosidade dos estimadores também foi apresentada para o caso de erros simetricamente distribuídos, porém, com a condição que o problema tenha solução única. Para o caso de múltiplas soluções Sielken e Hartley (1973) apresentaram um algoritmo que produz estimadores de regressão L_1 não tendenciosos.

O primeiro resultado com relação a distribuição dos estimadores dos parâmetros da regressão L_1 foi apresentado por Rosemberg e Carlson (1977). Neste trabalho eles investigaram as propriedades dos estimadores da regressão L_1 através de um estudo de Monte Carlo. Foram utilizados dois tamanhos amostrais $n=31$ e $n=39$, três tipos de variáveis preditoras (constante, normal padrão e normal contaminada com curtose alta) e três tipos de distribuições de erros, normal padrão e duas normais contaminadas ambas com mesmo percentual de contaminação, 15%, porém com variâncias da distribuição contaminante diferentes, 15 e 25. Para cada combinação destes três fatores foi realizado um experimento com 1000 repetições. Neste conjunto de repetições, os valores das variáveis preditoras foram mantidos fixos, variando a cada execução o valor da distribuição dos erros. Em cada uma das repetições foram obtidas as estimativas dos parâmetros de regressão usando os critérios de minimização dos erros absolutos e mínimos quadrados. Obtidas as 1000

estimativas para os parâmetros de regressão, foram então calculadas as variâncias das estimativas de ambos os métodos. Este procedimento foi feito para o caso de regressão linear simples, duas e seis variáveis preditoras e ao final as principais conclusões foram:

- i) Os estimadores de regressão L_1 possuem erro padrão significativamente menor que os estimadores de regressão L_2 quando a distribuição dos erros tem curtose elevada.
- ii) Os estimadores de regressão L_1 têm distribuição aproximadamente normal quando os erros têm curtose alta.
- iii) Os erros $(\hat{\beta} - \beta)$ nos estimadores de regressão L_1 são aproximadamente distribuídos segundo uma normal *m-variada* com vetor de médias 0 e matriz de variâncias e covariâncias $\lambda^2 (X'X)^{-1}$ onde X é a matriz das variáveis preditoras e λ^2/n é a variância assintótica da mediana de uma amostra de tamanho n na distribuição dos erros.

Este último resultado foi mais tarde obtido por Bassett e Koenker (1978) que mostraram, analiticamente, que a distribuição de amostragem do vetor de estimativas dos parâmetros de regressão L_1 , $\hat{\beta}$, é assintoticamente normal com média β e matriz de variâncias e covariâncias $\lambda^2 (X'X)^{-1}$. Este resultado possibilitou o desenvolvimento dos primeiros procedimentos inferenciais assim como a construção de intervalos de confiança e testes de hipóteses.

Os trabalhos recentes em regressão L_1 têm abordado pontos como o problema de seleção do melhor subconjunto de variáveis em regressão L_1 (Gentle e Hanson (1977), Narula e Wellington (1979)) e também a influência de cada observação na

regressão L_1 (Narula e Wellington (1985)).

3.2 - TEORIA ASSINTÓTICA

Consideremos o modelo (1.1) e suponhamos que as componentes do vetor ε_i são independentes e identicamente distribuídas com função de distribuição F cuja mediana é zero. Denotemos por $\hat{\beta}$, o estimador de β na regressão L_1 , isto é,

$$\min_{\beta} |Y - X\beta| = |Y - X\hat{\beta}| \quad (3.1)$$

Seja $\{\hat{\beta}_n\}$ a sequência de soluções únicas de problemas (3.1), $n \rightarrow \infty$.

Suponhamos que:

- i) F é contínua e tem densidade f positiva e contínua na mediana.
- ii) $\lim_{n \rightarrow \infty} n^{-1}(X'X)^{-1} = Q$

onde Q é uma matriz definida positiva, temos então que:

$\sqrt{n}(\hat{\beta} - \beta)$ converge em distribuição para uma distribuição normal m -variada com vetor de médias 0 e matriz de variâncias e covariâncias $\lambda^2 Q^{-1}$. λ^2/n é a variância assintótica da mediana de amostras de tamanho n da distribuição F . Este é o resultado de Bassett e Koenker e a consequência imediata do mesmo é que o critério de minimização dos erros absolutos é superior ao de mínimos quadrados no modelo linear geral, sempre que a distribuição dos erros é tal que a mediana é um estimador mais eficiente que a média como medida de posição.

A convergência dos estimadores da regressão L_1 para a distribuição normal foi estudada por Dielman e Pfaffemberger (1982). Concluíram que no caso de erros normalmente distribuídos é necessário que o tamanho de amostra seja 20 para que a hipótese de normalidade seja aceita. Para erros com distribuição normal contaminada é preciso uma amostra de tamanho 30 para que a convergência ocorra ao nível de significância de 10%. Porém, no caso de erros com distribuição de Cauchy a convergência ao nível de significância de 5% ocorre somente quando $n=100$ e se os erros são provenientes da distribuição de Laplace o tamanho de amostra necessário para convergência cresce para 150 ao nível de significância de apenas 1%.

Portanto, o uso da distribuição assintótica é adequado quando a distribuição dos erros é normal ou normal contaminada ou quando o tamanho de amostra é grande. A utilização da teoria assintótica em procedimentos estatísticos para pequenas amostras pode ocasionar resultados incorretos principalmente no caso dos erros provenientes de distribuições com caudas alongadas.

Com o resultado de Bassett e Koenker tornou-se simples a construção de intervalos de confiança e testes de hipóteses para os estimadores de regressão L_1 .

Supondo que estamos interessados em um intervalo de confiança para $r'\beta$, onde r é um vetor de constantes usado para especificar uma particular combinação linear dos parâmetros β . Um intervalo de confiança de $(1-\alpha)\%$ para $r'\beta$ pode ser escrito como:

$$r'\hat{\beta} \pm Z_{\alpha/2} \lambda [r'(X'X)^{-1}r]^{1/2} \quad (3.2)$$

onde Z denota o percentil apropriado da distribuição normal padrão.

Para a i -ésima componente do vetor β , um intervalo de confiança de $(1-\alpha)\%$ de confiança é:

$$\hat{\beta}_i \pm Z_{\alpha/2} \lambda(X'X)^{-1/2}_{ii} \quad (3.3)$$

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

Podemos também construir um teste para hipóteses do tipo:

$$H_0: r'\beta = h$$

$$H_1: r'\beta \neq h$$

Fixando um nível de significância α , as regiões de rejeição e aceitação são dados por:

$$\text{a) Rejeitar } H_0 \text{ se: } -Z_{\alpha/2} > Z^* \text{ ou } Z_{\alpha/2} < Z^*$$

$$\text{b) Aceitar } H_0 \text{ se: } -Z_{\alpha/2} \leq Z^* \leq Z_{\alpha/2}$$

onde

$$Z^* = \frac{r'\hat{\beta} - h}{\hat{\lambda}(r'(X'X)^{-1}r)^{1/2}} \quad (3.4)$$

Para uma componente específica do vetor β , onde desejamos testar a hipótese

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

temos:

a) Rejeitar H_0 se: $Z_{\alpha/2} < Z^*$ ou

$$-Z_{\alpha/2} > Z^*$$

b) Aceitar H_0 se: $-Z_{\alpha/2} \leq Z^* \leq Z_{\alpha/2}$

onde

$$Z^* = \frac{\hat{\beta}_i}{\hat{\lambda}(X'X)^{-1/2}_{ii}} \quad (3.5)$$

Uma alternativa para testar hipóteses sobre os parâmetros da regressão L_1 é através dos testes usuais de Wald, Multiplicadores de Lagrange e Razão de Verossimilhança. Estes três testes foram estudados no caso de regressão L_1 por Bassett e Koenker (1982), obtendo o seguinte resultado: Os testes de Wald, Razão de Verossimilhança e Multiplicadores de Lagrange são assintoticamente equivalentes e convergem em distribuição para uma qui-quadrado com K graus de liberdade e parâmetro de não centralidade η , onde

$$\eta = \lambda^{-2} \hat{\beta}'_i (D^{ij})^{-1} \hat{\beta}_i \quad (3.6)$$

K = posto da matriz X .

$\hat{\beta}_i$ é o vetor de estimativas do vetor β_i que se deseja testar e D^{ij} é o bloco correspondente na matriz $(X'X)^{-1}$.

O teste dos Multiplicadores de Lagrange é mais reco

mendado por ser mais simples do ponto de vista computacional.

Uma dificuldade para a utilização destes procedimentos é o fato de λ ser desconhecido. Assim uma estimativa de λ é necessária para que estes procedimentos sejam utilizados.

Uma estimativa para λ é sugerida por Cox e Hinkley (1974). Primeiro temos que assintoticamente $\lambda = \{2f_{(m)}\}^{-1}$ (Cramer, (1946)), onde $f_{(m)}$ é a ordenada da distribuição de erros avaliada na mediana. A sugestão de Cox e Hinkley é que $\{f_{(m)}\}^{-1}$ pode ser estimada por:

$$\hat{f} = \frac{e_{(t)} - e_{(s)}}{(t-s)/n} \quad (3.7)$$

onde $e_{(i)}$, $i=1,2,\dots,n$ é o vetor de resíduos ordenados e, $e_{(t)}$, $e_{(s)} \neq 0$.

O valor de t e s em (3.7) devem ser simétricos com relação a mediana amostral dos resíduos e a diferença entre eles deve ser pequena. Em particular:

$$\begin{aligned} t &= [n/2] + v \\ s &= [n/2] - v \end{aligned} \quad (3.8)$$

onde $[.]$ representa o maior inteiro contido e v é um valor inteiro positivo apropriado. (3.7) é um estimador consistente para 2λ , logo uma estimativa para λ é dada por:

$$\hat{\lambda} = \hat{f}/2 \quad (3.9)$$

Sposito (1987) verificou através de um estudo de Monte Carlo o comportamento de (3.9) como estimador de λ .

Neste estudo foram utilizadas as seguintes distribuições de erros.

1. Laplace:

$$f(x) = 1/2 \exp\{-|x|\} \quad -\infty < x < \infty$$

2. Cauchy:

$$f(x) = [\pi(1+x^2)]^{-1} \quad -\infty < x < \infty$$

3. Normal Padrão:

$$f(x) = (2\pi)^{-1/2} \exp\{-x^2/2\} \quad -\infty < x < \infty$$

4. Triangular:

$$f(x) = \begin{cases} 1-|x| & -\infty < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

5. Uniforme:

$$f(x) = \begin{cases} 1 & -1/2 \leq x \leq 1/2 \\ 0 & \text{caso contrário} \end{cases}$$

Tamanhos de amostra, 50, 100, 200, 300 e diferentes valores para ν foram testados utilizando o modelo $Y = \beta_0 + e$. Os resultados foram analisados separadamente para o caso de distribuições com amplitude finita e para as distribuições de amplitude infinita.

A partir dos resultados obtidos concluiu-se que:

1. Para distribuições com amplitude finita (3.9) aproxima λ rapidamente e para $n=100$, $\nu=3$ ou 4 apresenta uma boa aproximação. Se $n \geq 200$ valores de $\nu=2$ ou 3 mostraram-se adequados.
2. Para distribuições com amplitude infinita:

- i) No caso de distribuição normal uma boa aproximação pode ser obtida com $v=1$ independentemente do tamanho de amostra considerado.
- ii) Para a distribuição de Laplace uma boa aproximação ocorre com $n=100$ e $v=3$ e para $n \geq 200$, $v=1$ fornece bons resultados.
- iii) Se a distribuição dos erros for Cauchy é necessário um tamanho de amostra 300 e $v=1,2$, ou 3 para que uma boa aproximação ocorra.

Podemos então concluir que no caso de distribuições com amplitude infinita uma boa aproximação para λ utilizando-se (3.9) é obtida a partir de uma adequada escolha de v para cada tamanho de amostra.

C A P Í T U L O 4

BOOTSTRAP E REGRESSÃO L_1

Neste capítulo apresentamos o método bootstrap proposto por Efrom (1979), que será utilizado para estimar o desvio padrão e construir intervalos de confiança dos estimadores dos parâmetros da regressão L_1 .

4.1 - MÉTODO BOOTSTRAP

Um problema usual em estatística é aquele onde a partir de uma amostra aleatória $Z = (Z_1, Z_2, \dots, Z_n)$ proveniente de uma certa distribuição de probabilidade desconhecida F , desejamos fazer inferências sobre a distribuição de amostragem para uma determinada variável aleatória $R(Z, F)$, dependente de Z e F .

Os métodos clássicos de inferência estatística supõem que F é uma distribuição de probabilidade conhecida, que nem sempre é a verdadeira distribuição da variável aleatória em questão e, a partir dela fazem as inferências desejadas.

O método bootstrap não exige nenhuma suposição sobre a distribuição F e procura através dos próprios valores obtidos para a amostra $Z = (Z_1, Z_2, \dots, Z_n)$ buscar a solução do problema de inferência. A idéia geral do método é simples e consiste em retirar amostras aleatórias de tamanho n , com reposição, dos próprios valores obtidos para Z , de modo a obter uma distribui

ção bootstrap para os estimadores de interesse. Existem três propostas básicas para a obtenção da distribuição bootstrap:

- i) cálculo teórico;
- ii) aproximação Monte Carlo;
- iii) expansão em série de Taylor.

Neste trabalho utilizaremos a segunda proposta, a distribuição bootstrap será aproximada via método Monte Carlo.

Consideremos $Z = (Z_1, Z_2, \dots, Z_n)$ uma amostra aleatória cujas componentes são independentes e identicamente distribuídas (i.i.d.), provenientes de uma distribuição F da qual estamos interessados em estudar o parâmetro θ . Seja $\hat{\theta}(Z_1, Z_2, \dots, Z_n)$ a estatística de interesse cujo desvio padrão será escrito como $\sigma_\theta(F)$. Esta notação caracteriza o desvio padrão como uma função da distribuição de probabilidade desconhecida F . A estimativa bootstrap de $\sigma_\theta(F)$ será denotado por $\sigma_\theta(\hat{F}) = \sigma(\hat{F})$, onde \hat{F} é a estimativa de máxima verossimilhança não paramétrica de F .

O cálculo da estimativa bootstrap do desvio padrão, via método de Monte Carlo, segue o seguinte algoritmo.

1. Constrói-se \hat{F} como sendo:

$$\hat{F}: \text{massa } \frac{1}{n} \text{ sobre } Z_i, \quad i=1, 2, \dots, n \quad (4.1)$$

2. Retira-se uma amostra bootstrap de \hat{F} ,

$$Z_1^*, Z_2^*, \dots, Z_n^* \stackrel{i.i.d.}{\sim} \hat{F} \quad (4.2)$$

Calculamos então $\hat{\theta}^* = \hat{\theta}(Z_1^*, Z_2^*, \dots, Z_n^*)$.

3. Repete-se o passo 2 independentemente B vezes, obtendo em cada repetição bootstrap, $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$.

Calculamos então o desvio padrão dos $\hat{\theta}^{*i}$.

$$\hat{\sigma}_B = \left\{ \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}^{*i} - \hat{\theta}^*)^2 \right\}^{1/2} \quad (4.3)$$

onde

$$\hat{\theta}^* = \frac{1}{B} \sum_{i=1}^B \hat{\theta}^{*i} \quad (4.4)$$

Quando $B \rightarrow \infty$, $\hat{\sigma}_B$ aproxima $\sigma(\hat{F})$, o estimador bootstrap do desvio padrão.

A partir da estimativa $\sigma(\hat{F})$, podemos obter um intervalo de confiança aproximado para θ . Esta aproximação é dada por:

$$\hat{\theta} \pm \sigma(\hat{F}) \cdot Z_{\alpha} \quad (4.5)$$

onde Z_{α} é o α -ésimo percentil da distribuição normal padrão. Este intervalo é chamado "intervalo padrão" para θ e é baseado na aproximação normal para grandes amostras.

Outra forma de obtenção de um intervalo de confiança para θ é a partir da distribuição bootstrap de $\hat{\theta}^*$, é o chamado método dos percentis.

Seja

$$F(t) = Prob_*\{\hat{\theta} \leq t\} \quad (4.6)$$

a função de distribuição bootstrap de $\hat{\theta}^*$. Quando esta distribuição é obtida pelo método de Monte Carlo temos que $\hat{F}(t)$ é aproximada por

$$\frac{\#\{\hat{\theta}^{*i} \leq t\}}{B}$$

Para um dado α , entre 0 e 0.5, definimos:

$$\hat{\theta}_{INF}(\alpha) = \hat{F}^{-1}(\alpha) \quad (4.7)$$

$$\hat{\theta}_{SUP}(\alpha) = \hat{F}^{-1}(1-\alpha) \quad (4.8)$$

O método dos percentis fornece o intervalo dado por:

$$[\hat{\theta}_{INF}(\alpha), \hat{\theta}_{SUP}(\alpha)] \quad (4.9)$$

como um intervalo $(1-2\alpha)\%$ de confiança para θ .

Outras sugestões para a utilização da distribuição bootstrap na construção de intervalos de confiança foram apresentados por Efrom (1982) e Efrom e Tibshirani (1985), mas não serão consideradas neste trabalho.

4.2 - APLICAÇÃO DO MÉTODO BOOTSTRAP À REGRESSÃO L_1

Consideremos o modelo (1.1) com ε_i , *i.i.d.*, provenientes de uma distribuição desconhecida F . Suponhamos que F seja centrada em zero em algum sentido: $E_F(\varepsilon) = 0$ ou $P\{\varepsilon < 0\} = 0.5$. Para o caso de regressão L_1 , pela Propriedade 2 do Capítulo 2, temos que a mediana dos ε_i sempre se iguala a zero, isto é, $P\{\varepsilon < 0\} = 0.5$.

A estimativa bootstrap do desvio padrão $\hat{\sigma}_B = \hat{\sigma}(\hat{\beta}, F)$ pode ser utilizada para medir a precisão dos parâmetros da regressão L_1 . Neste caso \hat{F} é a distribuição empírica dos resíduos.

Consideremos $\hat{Y} = X\hat{\beta}$, o vetor estimado pela estimativa

L_1 dos parâmetros de regressão. Então, $\hat{\varepsilon}_i = Y_i - X_i \hat{\beta}$ é o vetor de erros estimado para o ajuste obtido através do critério L_1 .

O algoritmo bootstrap Monte Carlo pode ser então descrito da seguinte forma:

1. Constrói-se \hat{F} colocando-se massa $\frac{1}{n}$ a cada ponto do vetor $\hat{\varepsilon}_i$ observado.

$$\hat{F}: \text{massa } \frac{1}{n} \text{ sobre } \hat{\varepsilon}_i, \quad i=1,2,\dots,n \quad (4.10)$$

2. Retira-se uma amostra bootstrap Y^* , selecionando-se uma amostra i.i.d. de \hat{F} , $\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_n^*$, obtemos então:

$$Y_i^* = X_i \hat{\beta} + \varepsilon_i^*, \quad i=1,2,\dots,n \quad (4.11)$$

Calcula-se então uma estimativa L_1 do vetor β , $\hat{\beta}^*$, para esta amostra, mantendo-se X fixa.

3. O Passo 2 é repetido independentemente B vezes, obtendo-se as estimativas bootstrap, $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$.

Uma estimativa para o desvio padrão de cada parâmetro $\hat{\beta}_i^*$ é dada por:

$$\hat{\sigma}_{\beta_i} = \left\{ \frac{1}{B-1} \sum_{j=1}^B (\hat{\beta}_i^{*(j)} - \hat{\beta}_i^{*(.)}) \right\}^{1/2} \quad (4.12)$$

onde

$$\hat{\beta}_i^{*(.)} = \frac{1}{B} \sum_{j=1}^B \hat{\beta}_i^{*(j)} \quad (4.13)$$

A utilização deste algoritmo para o problema de regressão L_1 pode apresentar dificuldades devido a reamostragem

de resíduos iguais a zero levando a subestimação do desvio padrão. Pela Propriedade 1 do Capítulo 2, em regressão L_1 , existe pelo menos um hiperplano L_1 ótimo que passa por pelo menos m pontos observados, ou seja, existe pelo menos uma solução L_1 ótima que resulta em pelo menos m valores do vetor de resíduos $\hat{\varepsilon}_i$ iguais a zero, onde m é o posto da matriz das variáveis preditoras.

Por outro lado, estudos têm mostrado que em geral os problemas de regressão L_1 apresentam uma única solução ótima. Conjuntos de observações com múltiplas soluções ótimas precisam ser construídos artificialmente. Portanto, pela Propriedade 1, salvo raras exceções, as soluções ótimas de problemas de regressão L_1 passam por pelo menos m pontos observados.

Uma forma de resolver este problema é através do bootstrap alisado. Neste caso retira-se uma amostra bootstrap de $\hat{F} \otimes G$, onde \otimes significa convolução. Cada ponto ε_i da amostra bootstrap representa a soma de um ponto selecionado aleatoriamente dos dados originais com um ponto selecionado da distribuição G .

A reamostragem dos resíduos nulos pode também ser omitida retirando-se da amostra original de resíduos aqueles iguais a zero.

O método bootstrap pode ser também utilizado de uma forma mais simplificada em problemas de regressão. Considere mos cada observação $t_i = (Y_i, X_i)$ obtida através de amostragem aleatória simples de uma distribuição F . O método bootstrap é então aplicado ao conjunto de dados,

$$t_1, t_2, \dots, t_n \stackrel{i.i.d.}{\sim} F.$$

O algoritmo bootstrap Monte Carlo fica da seguinte forma:

1. Constrói-se \hat{F} , colocando-se massa $\frac{1}{n}$ sobre cada observação t_i .

$$\hat{F}: \text{massa } \frac{1}{n} \text{ sobre } t_i, \quad i=1,2,\dots,n \quad (4.14)$$

2. Retira-se uma amostra bootstrap

$$t_1^*, t_2^*, \dots, t_n^* \stackrel{i.i.d.}{\sim} \hat{F}$$

Calcula-se para esta amostra a estimativa L_1 bootstrap ($\hat{\beta}^{*(j)}$).

3. Repete-se 2, B vezes obtendo-se as estimativas bootstrap

$$\hat{\beta}^{*(j)}, \quad j=1,2,\dots,B$$

Com os valores $\hat{\beta}^{*(j)}$, calcula-se a estimativa bootstrap do desvio padrão através de (4.12) e (4.13).

C A P Í T U L O 5

METODOLOGIA E RESULTADOS

Neste capítulo apresentamos a metodologia utilizada para avaliar a aplicação do método bootstrap na obtenção de estimativas dos desvios padrões e intervalos de confiança dos parâmetros da regressão L_1 , bem como os resultados obtidos.

5.1 - METODOLOGIA

Para verificar a viabilidade do uso do método bootstrap como procedimento para obtenção do desvio padrão e intervalos de confiança para os estimadores da regressão L_1 quando o tamanho da amostra é pequeno, desenvolvemos um extenso estudo computacional utilizando o computador VAX11/785 do Centro de Computação da UNICAMP.

Este estudo foi realizado em duas partes. A primeira consistiu em um estudo de Monte Carlo para obter os valores conhecidos para o desvio padrão dos estimadores da regressão L_1 sob diversas condições de distribuições de erros e tamanhos de amostras. Estes valores foram então comparados aos resultados obtidos pelo método bootstrap, que é a segunda parte do estudo.

Com este procedimento desejamos mostrar que as estimativas obtidas através do bootstrap aproximam os valores obtidos através do estudo de Monte Carlo, indicando desta forma a viabi

lidade do uso do método bootstrap como procedimento para estimar o desvio padrão dos estimadores dos parâmetros da regressão L_1 em pequenas amostras.

O modelo que utilizamos foi o de regressão linear simples

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (5.1)$$

Foram utilizadas as seguintes distribuições simétricas de erros:

1. Normal Padrão:

$$f(x) = (2\pi)^{-1/2} \exp\{-x^2/2\} \quad -\infty < x < \infty$$

2. Normal Contaminada:

$$f(x) = 0.85 \cdot N(0,1) + 0.15 \cdot N(0,5) \quad -\infty < x < \infty$$

onde $N(0, \sigma)$ denota uma variável aleatória proveniente de uma distribuição normal com média 0 e desvio padrão σ , consequentemente $f(x)$ é uma mistura de distribuições.

3. Laplace:

$$f(x) = (1/2) \exp\{-|x|\} \quad -\infty < x < \infty$$

4. Cauchy:

$$f(x) = [\pi(1+x^2)]^{-1} \quad -\infty < x < \infty$$

Para a geração dos números aleatórios segundo essas distribuições, foram utilizados o Método Polar de Marsaglia para a distribuição normal e o Método da Transformação Inversa para

ra as variáveis de Cauchy e Laplace. Na geração de números aleatórios uniformes $[0,1]$ foi utilizado o gerador RAN da biblioteca FORTRAN/VAX.

Os tamanhos de amostra usados foram, 10, 20, 30, 50, 75, 100 para cada uma das quatro distribuições de erros. Para cada um destes tamanhos de amostra foram geradas 5 diferentes matrizes X . Foram escolhidos 5 valores para β_0 e β_1 , conforme a Tabela 5.1. Cada par desses valores foi atribuído a uma das matrizes para cada tamanho de amostra.

TABELA 5.1 - VALORES PARA β_0 E β_1 .

	β_0	β_1
1	1	1
2	3	7
3	2	6
4	3.5	2
5	3.1	2.4

Os valores de x_i foram gerados segundo a distribuição normal padrão, independentemente dos erros. Uma vez gerados os valores de x_i , para cada uma das cinco matrizes, estes valores foram então fixados durante todo o estudo. Os valores para a variável dependente, Y , em cada uma das matrizes X foram obtidos por

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i \quad i=1,2,\dots,n. \quad (5.2)$$

5.1.1 - ESTUDO DE MONTE CARLO

Na execução do estudo de Monte Carlo foram geradas 5.000 amostras, segundo cada uma das distribuições de erros. A

cada amostra dos erros que era gerada, um novo conjunto de valores para Y era obtido. Com os valores (Y, X) , calculávamos então as estimativas L_1 para β_0 e β_1 utilizando o algoritmo de Armstrong, Frome e Kung (1979).

Ao final do processo, para cada matriz, em cada tamanho de amostra, existiam 5.000 estimativas para β_0 e β_1 . A partir dessas estimativas foram então obtidos os desvios padrões destes estimadores pelo método de Monte Carlo; estes resultados são apresentados, para o caso de $\beta_0 = \beta_1 = 1$, na Tabela 5.2.

5.1.2 - MÉTODO BOOTSTRAP

Para cada matriz X gerada, o método bootstrap foi aplicado. Duas formas de reamostragem foram utilizadas;

- i) reamostragem nos pares t_i ;
- ii) reamostragem no vetor de resíduos ignorando-se aqueles iguais a zero.

A reamostragem utilizando a convolução $(F \otimes G)$, deixou de ser estudada por não apresentar resultados satisfatórios.

Para definirmos o número B de repetições bootstrap, realizamos um estudo comparando quatro possíveis alternativas: 100×100 , 100×200 , 200×100 e 200×200 , onde $A \times B$ significa que a estimativa bootstrap do desvio padrão é resultado da média de A execuções do método bootstrap com B repetições. Os resultados indicaram que dentre as 4 alternativas, aquela que melhor aproximou os valores do estudo de Monte Carlo com menor tempo de execução foi a de $A = 100$ e $B = 200$.

Para a construção de intervalos de confiança foi uti

lizado apenas o método padrão, pois o método dos percentis não apresentou bons resultados, bem como exigia um tempo computacional muito grande. Foi utilizado somente o caso de matrizes associadas a $\beta_0 = \beta_1 = 1$ e a performance dos intervalos construídos foi verificada através da probabilidade de cobertura obtida em 500 repetições do método bootstrap com 200 repetições.

5.2 - RESULTADOS PARA REAMOSTRAGEM NOS PARES

Para cada par de valores fixado, conforme a Tabela 5.1, para o vetor de parâmetros β , obtivemos uma tabela contendo o desvio padrão para as estimativas dos parâmetros da regressão L_1 pelo estudo de Monte Carlo (σ_{MC}), o resultado de Bassett e Koenker (σ_{BK}), a média das 100 estimativas bootstrap com reamostragem nos pares $t_i(\bar{\sigma})$ e o desvio padrão destas estimativas ($\hat{\sigma}_{EB}$), para cada uma das distribuições nos diversos tamanhos de amostra. No resultado de Bassett e Koenker foi usado o valor de λ assintótico de cada distribuição de erro. Na Tabela 5.2 apresentamos estes resultados para o caso de $\beta_0 = \beta_1 = 1$.

Estes resultados comprovam os problemas que podemos ter quando utilizamos o resultado de Bassett e Koenker. Quando a distribuição de erros é normal ou normal contaminada, o desvio padrão pelo resultado de Bassett e Koenker aproxima-se do resultado do estudo de Monte Carlo a partir de tamanhos de amostra 10 e 30 respectivamente. Porém, para distribuição de erros de Cauchy esta aproximação começa a ficar clara com tamanho de amostra 100, o que não ainda ocorre para erros de Laplace.

Outro resultado que se pode verificar é que os resultados obtidos pelo método bootstrap super-estimam os valores do

TABELA 5.2 - RESULTADOS PARA REAMOSTRAGEM NOS PARES $\beta_0 = \beta_1 = 1$.

DISTRIBUIÇÃO	TAMANHO DE AMOSTRA	n=10		n=20		n=30		n=50		n=75		n=100	
		β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
NORMAL	σ_{BK}	0.399	0.354	0.298	0.330	0.230	0.216	0.177	0.125	0.145	0.139	0.127	0.132
	σ_{MC}	0.378	0.364	0.289	0.326	0.228	0.218	0.176	0.168	0.146	0.139	0.129	0.132
	$\bar{\sigma}_B$	0.460	0.463	0.341	0.386	0.251	0.255	0.191	0.190	0.153	0.158	0.134	0.140
	σ_{EB}	0.137	0.143	0.129	0.121	0.060	0.107	0.045	0.045	0.032	0.041	0.027	0.033
NORMAL CONTAMINADA	σ_{BK}	0.453	0.402	0.443	0.390	0.261	0.246	0.201	0.191	0.165	0.161	0.145	0.151
	σ_{MC}	0.470	0.462	0.347	0.386	0.263	0.251	0.201	0.198	0.167	0.165	0.143	0.153
	$\bar{\sigma}_B$	0.733	0.782	0.405	0.490	0.301	0.311	0.216	0.209	0.178	0.189	0.152	0.166
	σ_{EB}	0.333	0.415	0.167	0.187	0.082	0.089	0.047	0.054	0.032	0.046	0.026	0.038
LAPLACE	σ_{BK}	0.318	0.282	0.211	0.233	0.183	0.173	0.141	0.134	0.115	0.113	0.102	0.106
	σ_{MC}	0.420	0.416	0.294	0.346	0.222	0.219	0.161	0.158	0.129	0.136	0.115	0.120
	$\bar{\sigma}_B$	0.578	0.617	0.379	0.446	0.260	0.271	0.193	0.194	0.141	0.156	0.120	0.135
	σ_{EB}	0.198	0.223	0.121	0.146	0.066	0.090	0.049	0.053	0.030	0.046	0.025	0.031
CAUCHY	σ_{BK}	0.500	0.443	0.374	0.413	0.288	0.271	0.222	0.211	0.182	0.178	0.160	0.167
	σ_{MC}	0.743	0.841	0.437	0.535	0.318	0.330	0.233	0.232	0.190	0.195	0.165	0.175
	$\bar{\sigma}_B$	13.412	21.051	0.755	0.902	0.422	0.448	0.278	0.271	0.213	0.230	0.178	0.203
	$\hat{\sigma}_{EB}$	86.916	148.175	0.528	0.565	0.167	0.187	0.061	0.065	0.050	0.075	0.042	0.058

estudo de Monte Carlo, sendo que à medida que o tamanho de amostra cresce, a diferença entre estes valores decresce, tanto para β_0 como β_1 .

Na Figura 5.1 apresentamos conjuntamente os valores obtidos nas diversas situações estudadas pelo método bootstrap e o estudo de Monte Carlo, para as estimativas de β_0 e β_1 . A partir deste gráfico verificamos uma diferença de comportamento das estimativas para β_0 e β_1 . Portanto os estudos a seguir serão feitos independentemente para β_0 e β_1 . Nas Figuras 5.2 e 5.3 representamos o gráfico anterior individualmente para β_0 e β_1 , respectivamente, sendo que o valor apresentado no gráfico representa o tamanho de amostra do qual o ponto é proveniente. Para que o método bootstrap seja uma boa aproximação do desvio padrão dos estimadores da regressão L_1 , devemos obter estes valores próximos a reta de identidade ($\hat{\sigma} = \sigma_{MC}$). Podemos notar que isto não ocorre porém, estes dois gráficos indicam a existência de alguma relação funcional entre estas variáveis distinta da identidade. Portanto, podemos então determinar uma relação funcional de modo a obter uma correção que forneça os valores do método bootstrap próximos aos do estudo de Monte Carlo.

Na busca de um ajuste adequado para esta função verificamos a existência de uma relação entre os valores observados e o tamanho de amostra do qual eram provenientes. Isto pode ser observado nas Figuras 5.2 e 5.3, quanto menor o tamanho de amostra, maior tem que ser a correção para que o bootstrap se aproxime do Monte Carlo, assim sendo, decidimos obter o ajuste destas correções em duas partes:

FIGURA 5.1

REAMOSTRAGEM NOS PARES

MONTE CARLO × BOOTSTRAP

* = BETA(0)
+ = BETA(1)

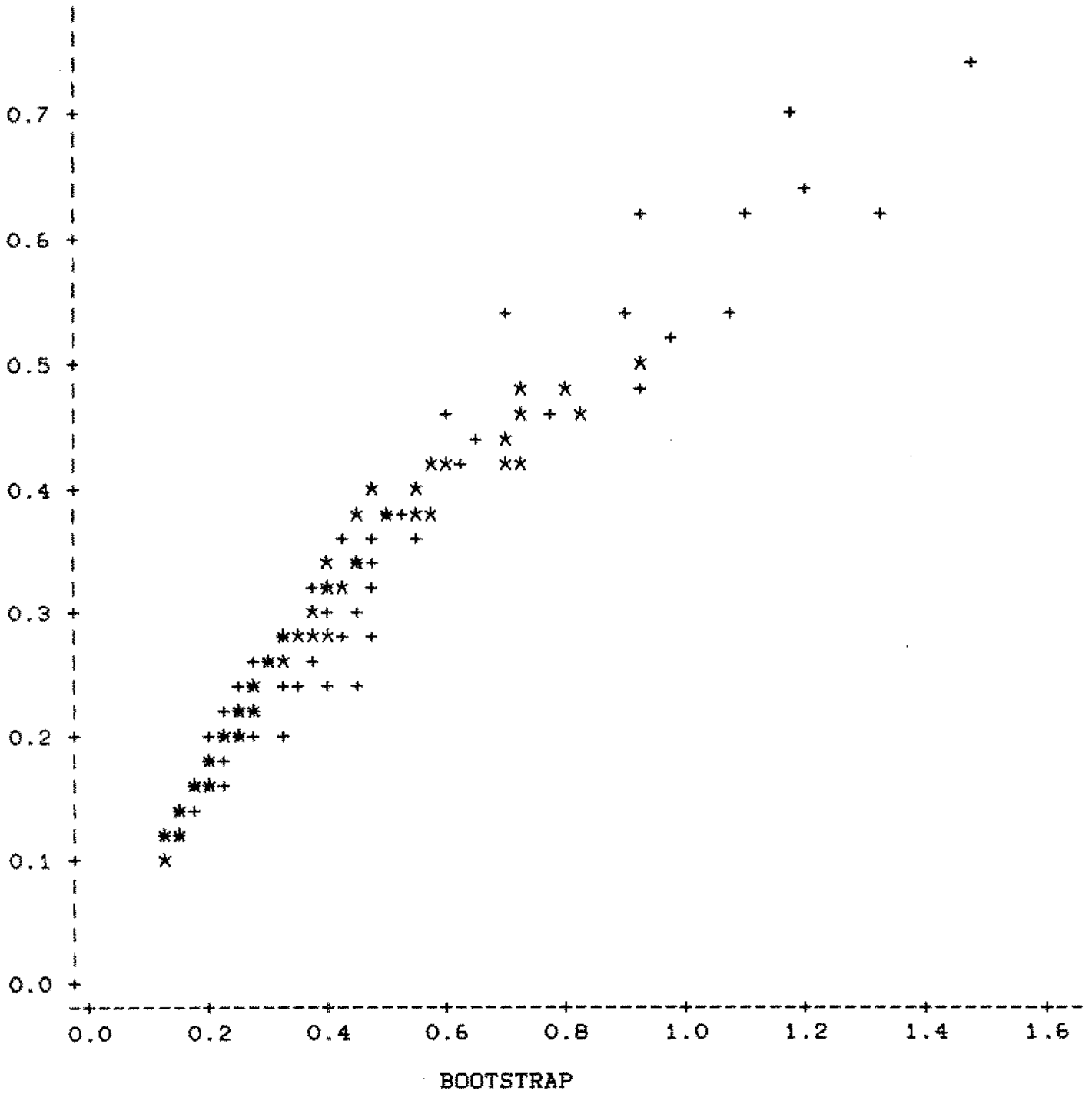


FIGURA 5.2

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(0) X BOOTSTRAP BETA(0)

- 1 = tamanho de amostra 10
- 2 = tamanho de amostra 20
- 3 = tamanho de amostra 30
- 5 = tamanho de amostra 50
- 7 = tamanho de amostra 75
- 0 = tamanho de amostra 100

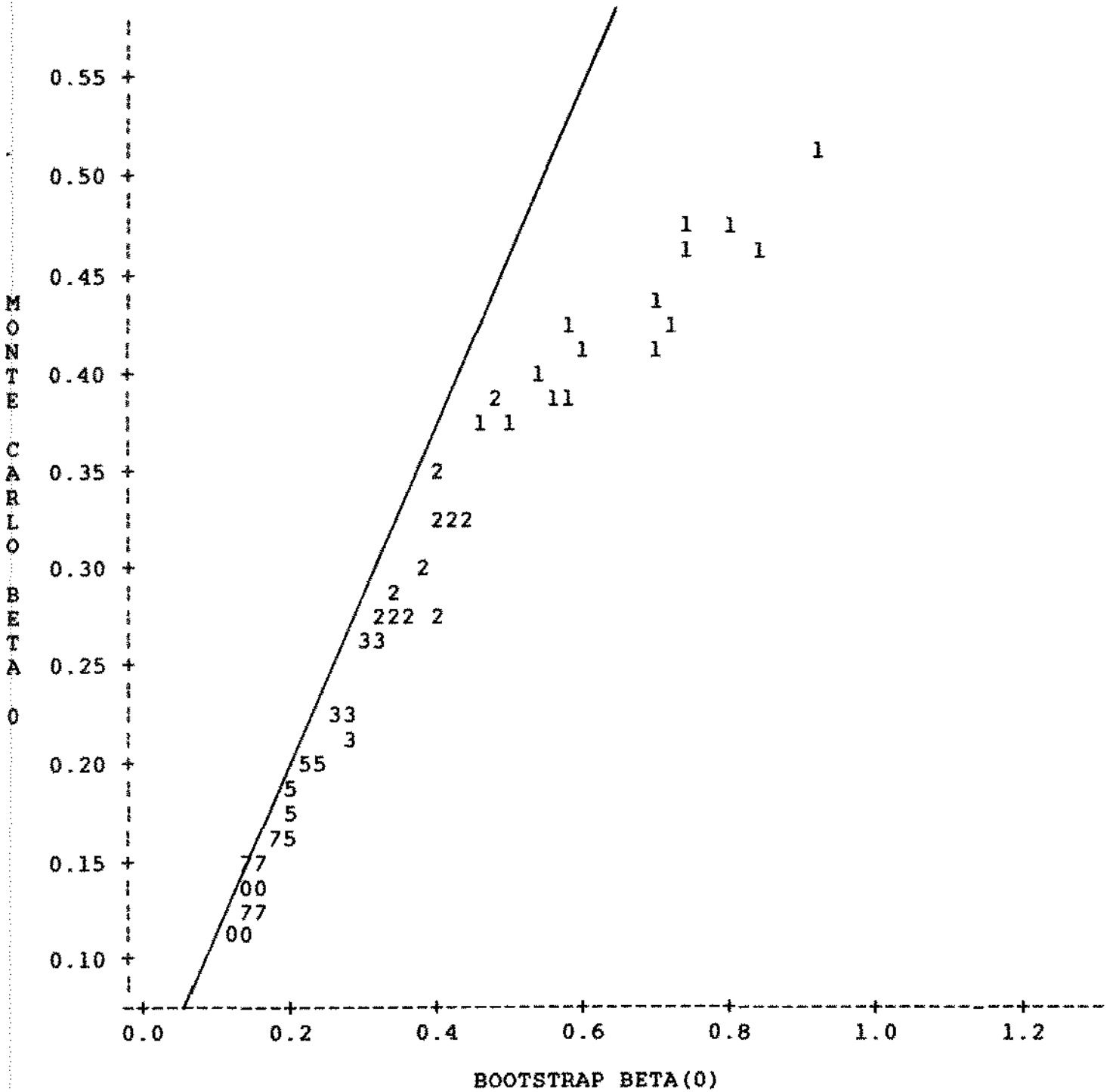
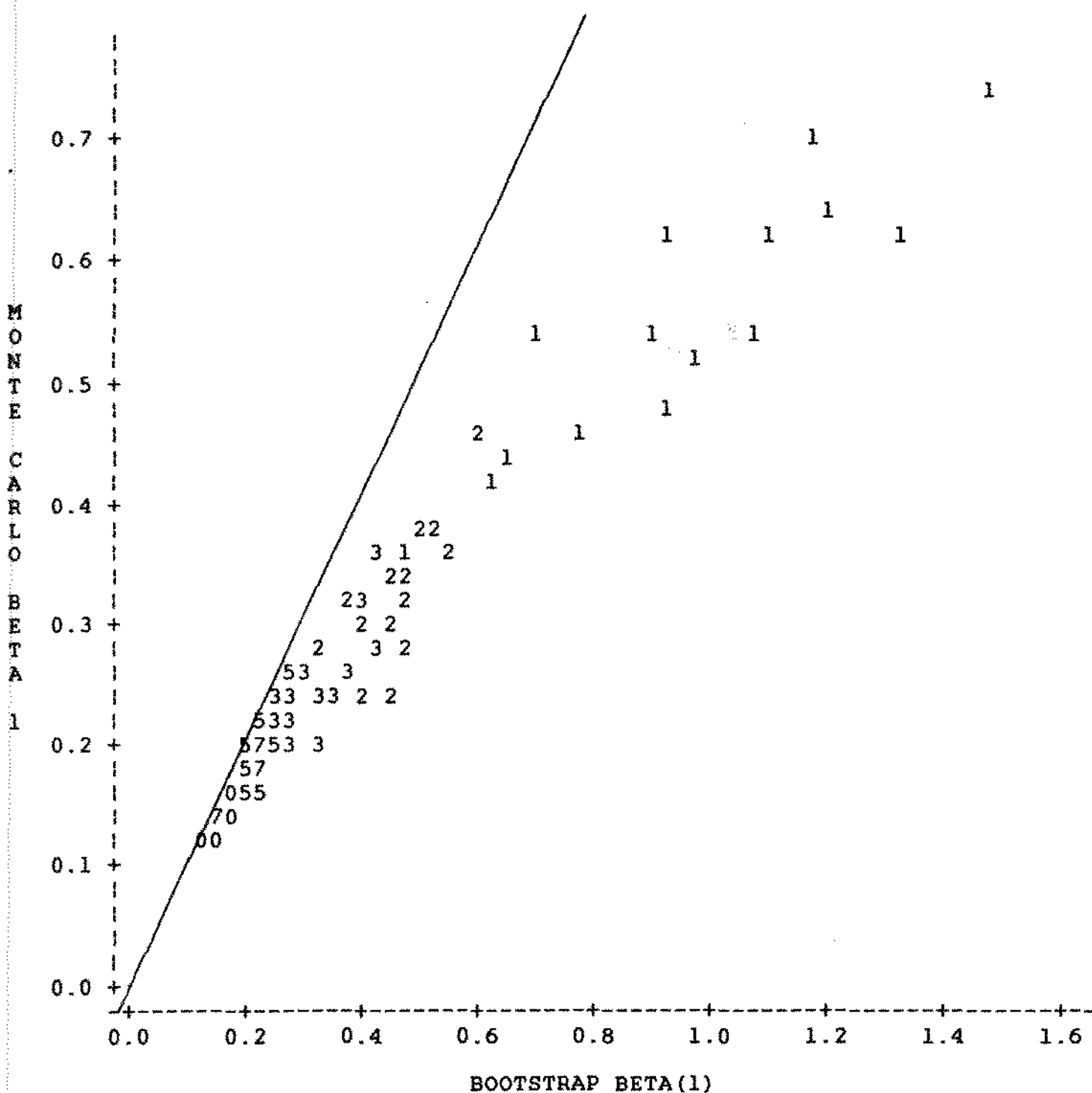


FIGURA 5.3

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(1) X BOOTSTRAP BETA(1)

- 1 = tamanho de amostra 10
- 2 = tamanho de amostra 20
- 3 = tamanho de amostra 30
- 5 = tamanho de amostra 50
- 7 = tamanho de amostra 75
- 0 = tamanho de amostra 100



- i) tamanhos de amostra menor ou igual a 30;
- ii) tamanho de amostra maior do que 30.

Obtivemos as seguintes correções em que BC denota bootstrap corrigido:

- i) $n \leq 30$

$$BC(\beta_0) = 0.13 + 0.47\bar{\sigma}_{\beta_0} - 0.08(\bar{\sigma}_{\beta_0})^2$$

$$BC(\beta_1) = 0.12 + 0.46\bar{\sigma}_{\beta_1} - 0.04(\bar{\sigma}_{\beta_1})^2$$

- ii) $n > 30$

$$BC(\beta_0) = 1.03\bar{\sigma}_{\beta_0} - 0.64(\bar{\sigma}_{\beta_0})^2$$

$$BC(\beta_1) = 0.046 + 0.64\bar{\sigma}_{\beta_1}$$

Nas Figuras 5.4 e 5.5 rerepresentamos as Figuras 5.2 e 5.3 aplicando estas correções onde podemos verificar que os pontos observados estão agora próximos à reta de identidade que desejávamos.

Para verificar a validade destas correções, novos conjuntos de valores foram gerados para diferentes tamanhos de amostra, diferentes matrizes X e diferentes valores dos parâmetros β_0 e β_1 . Para estes novos conjuntos de observações foram obtidos os resultados pelo método Monte Carlo e pelo método bootstrap nos quais foram aplicadas as correções acima propostas. Nas Figuras 5.6 e 5.7 estes valores são apresentados para β_0 e β_1 respectivamente. Podemos verificar que os pontos com as correções propostas, aproximam-se da identidade ($\bar{\sigma} = \sigma_{MC}$), in

FIGURA 5.4

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(0) X BOOTSTRAP CORRIGIDO BETA(0)

- 1 = tamanho de amostra 10
- 2 = tamanho de amostra = 20
- 3 = tamanho de amostra 30
- 5 = tamanho de amostra = 50
- 7 = tamanho de amostra 75
- 0 = tamanho de amostra = 100

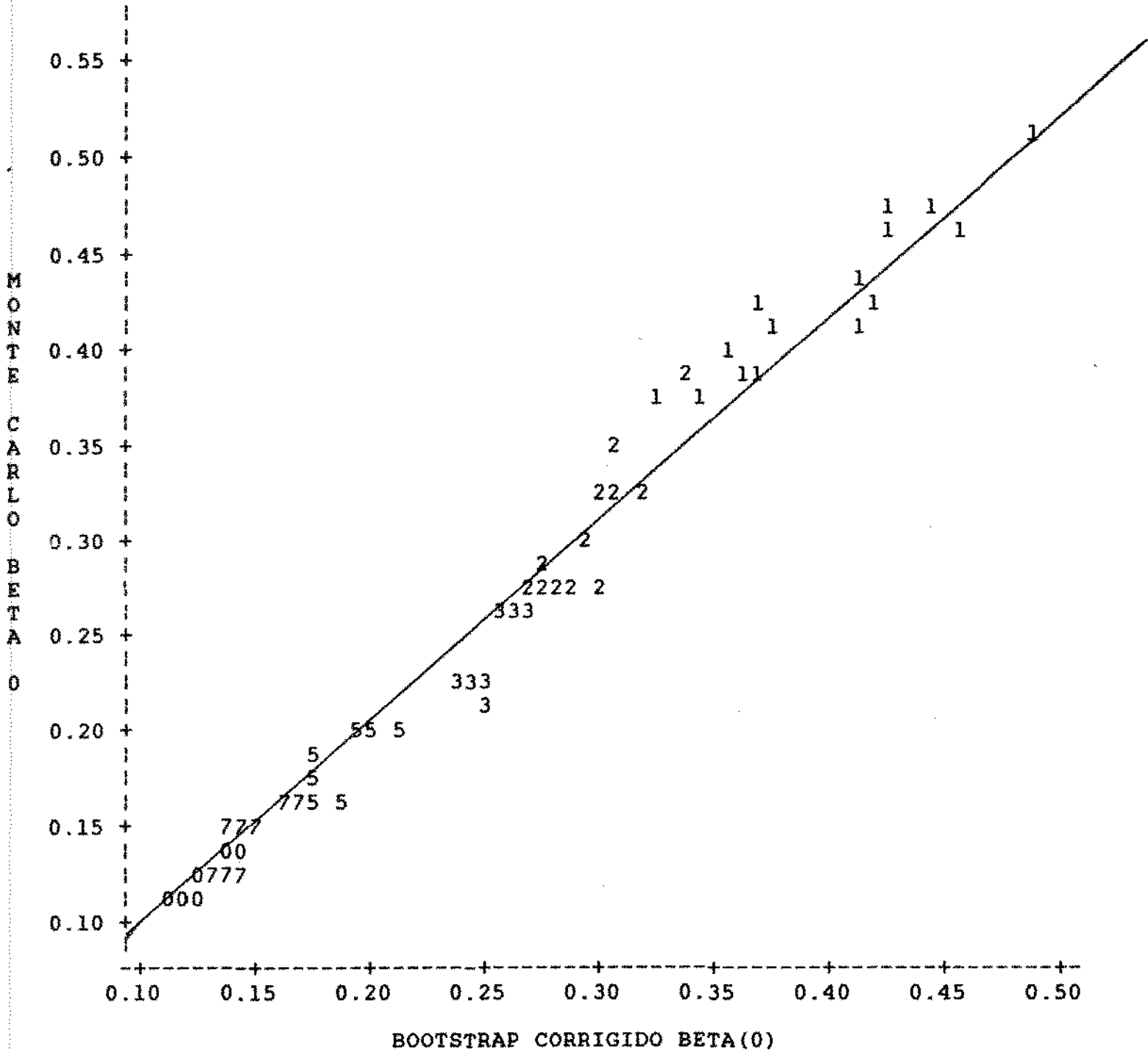


FIGURA 5.5

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(1) X BOOTSTRAP CORRIGIDO BETA(1)

- 1 = tamanho de amostra 10 2 = tamanho de amostra = 20
- 3 = tamanho de amostra 30 5 = tamanho de amostra = 50
- 7 = tamanho de amostra 75 0 = tamanho de amostra = 100

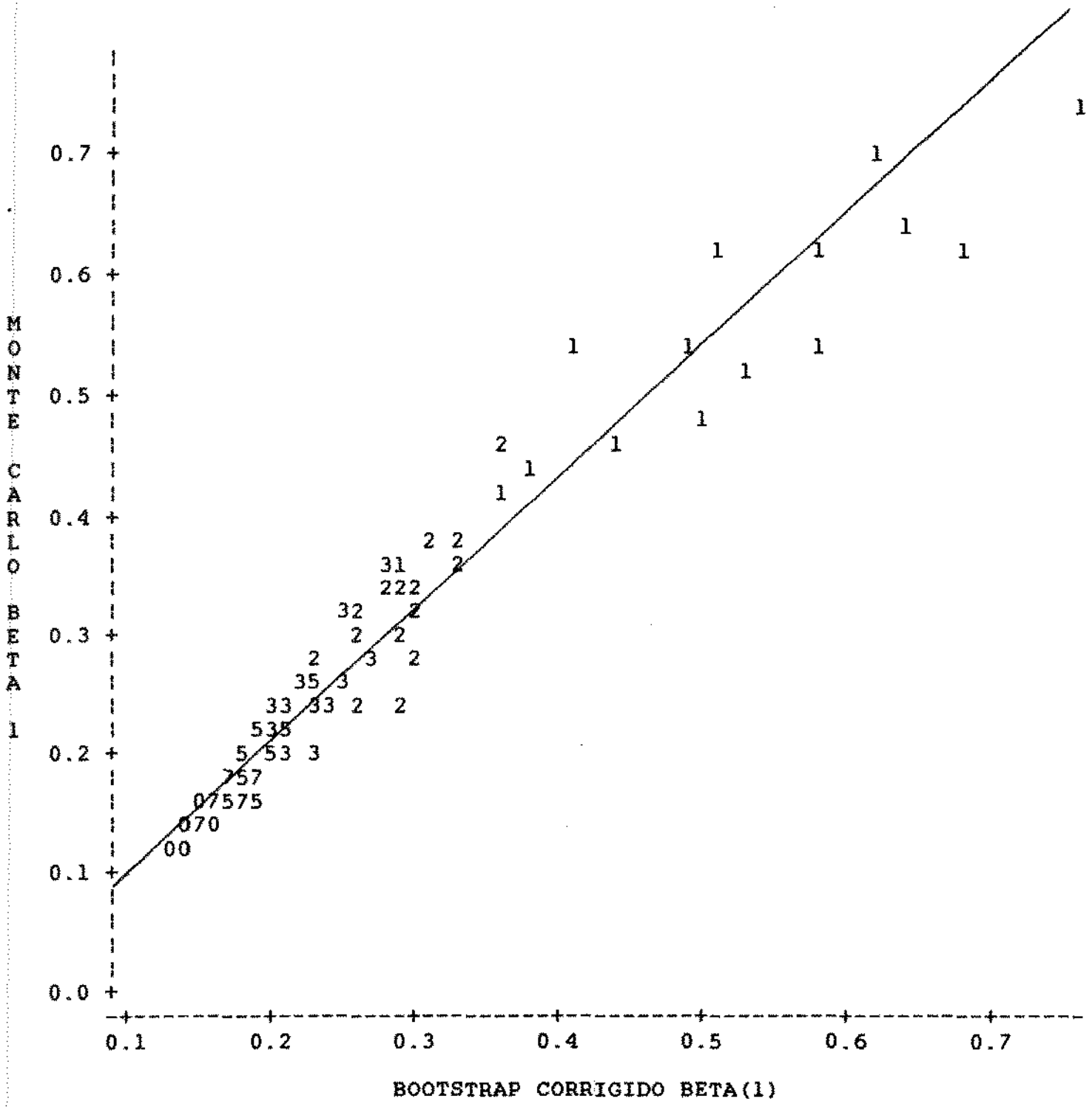


FIGURA 5.6

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(0) X BOOTSTRAP CORRIGIDO BETA(0)

- 1 = tamanho de amostra 15
- 2 = tamanho de amostra 25
- 3 = tamanho de amostra 35
- 6 = tamanho de amostra 60
- 7 = tamanho de amostra 70
- 8 = tamanho de amostra 85
- 0 = tamanho de amostra 120

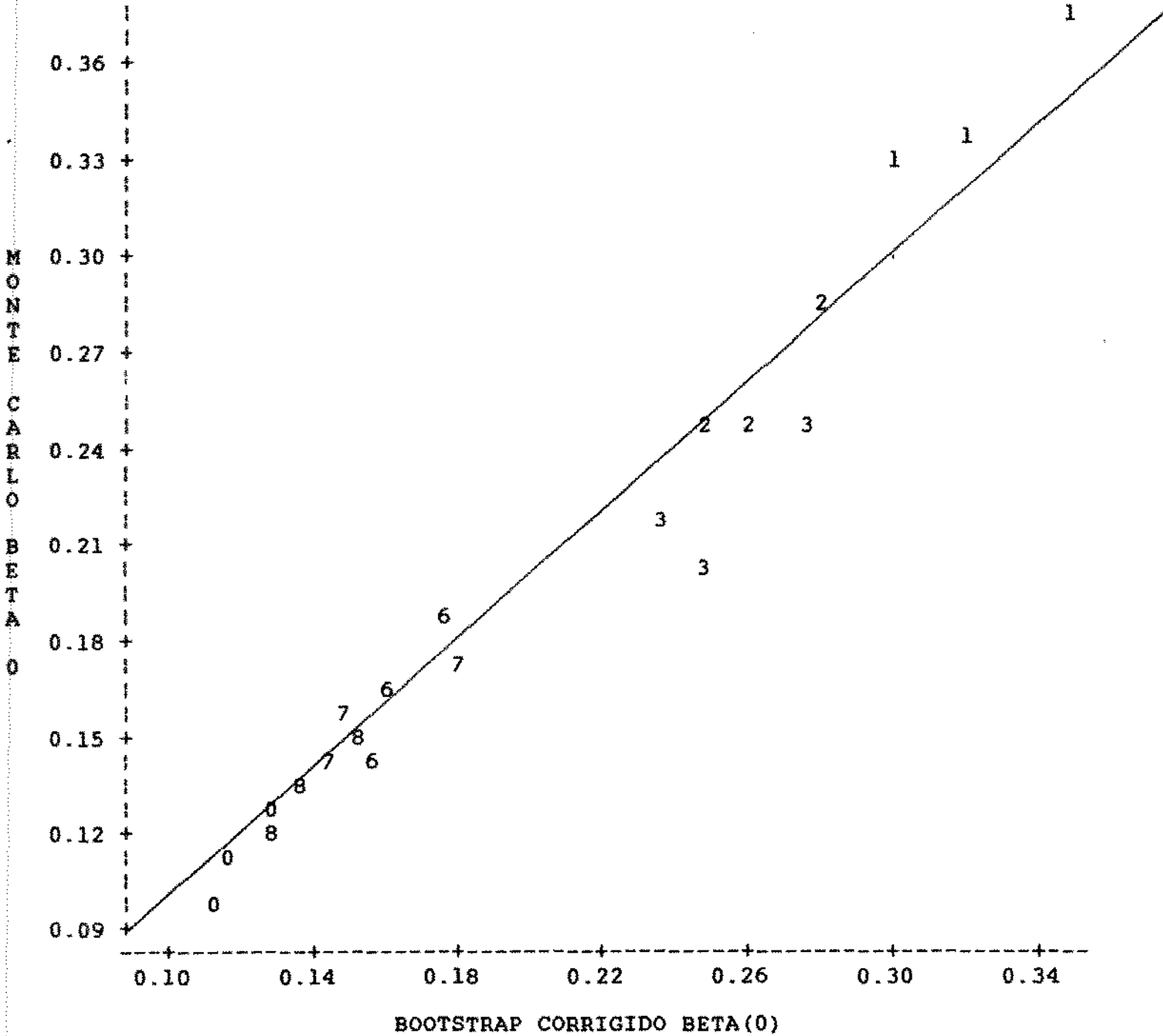
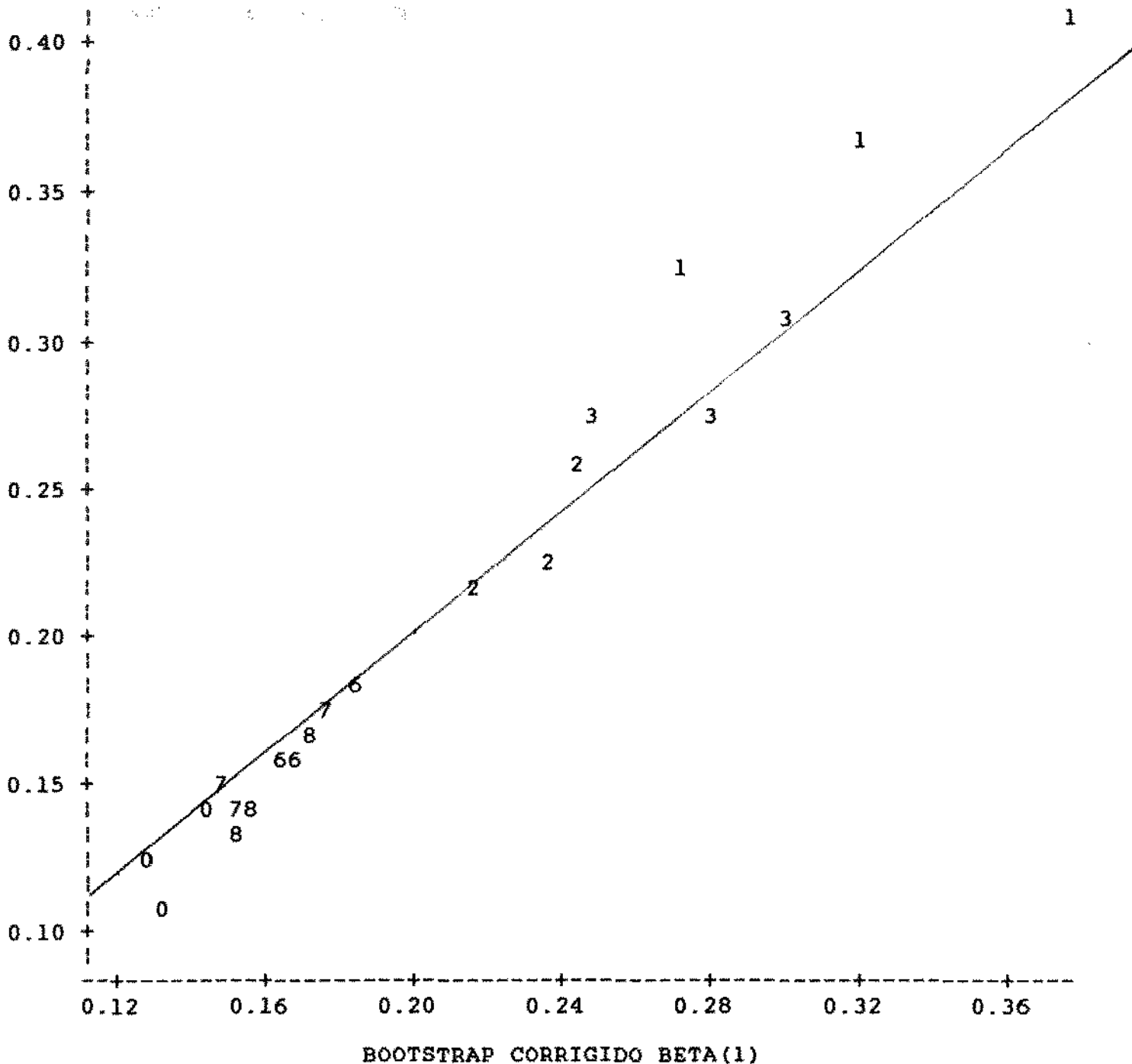


FIGURA 5.7

REAMOSTRAGEM NOS PARES

MONTE CARLO BETA(1) X BOOTSTRAP CORRIGIDO BETA(1)

- 1 = tamanho de amostra 15
- 2 = tamanho de amostra 25
- 3 = tamanho de amostra 35
- 6 = tamanho de amostra 60
- 7 = tamanho de amostra 70
- 8 = tamanho de amostra 85
- 0 = tamanho de amostra 120



dicando desta forma a validade das correções propostas às estimativas bootstrap.

Com as estimativas do desvio padrão, partimos então para a construção de intervalos de confiança para os parâmetros utilizando o método padrão (4.5) com $\sigma_{\beta_i}(\hat{F}) = BC(\beta_i)$. Calculamos então as probabilidades de cobertura destes intervalos, para cada uma das distribuições de erros nos diferentes tamanhos de amostra. Os resultados obtidos são apresentados nas Tabelas 5.3, 5.4, 5.5, 5.6 para as distribuições de erros normal, normal contaminada, Laplace e Cauchy, respectivamente.

TABELA 5.3 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS PA-
RES t_i . DISTRIBUIÇÃO DE ERROS NORMAL.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.896	.912	.932	.948	.980	.992
20	.924	.916	.962	.968	.992	.988
30	.902	.896	.944	.950	.986	.990
50	.908	.902	.948	.942	.988	.986
75	.906	.924	.930	.968	.980	.990
100	.908	.912	.946	.96	.982	.992

TABELA 5.4 - PROBABILIDADE DE COBERTURA. REAMOSTRAGEM NOS PA
RES. DISTRIBUIÇÃO DE ERROS NORMAL CONTAMINADA.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.900	.928	.940	.968	.980	.992
20	.922	.874	.962	.938	.990	.990
30	.898	.900	.942	.946	.984	.988
50	.906	.922	.952	.968	.988	.990
75	.906	.896	.944	.950	.990	.992
100	.916	.914	.964	.956	.982	.994

TABELA 5.5 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS PA
RES t_i . DISTRIBUIÇÃO DE ERROS LAPLACE.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.906	.938	.944	.972	.986	.990
20	.912	.902	.956	.954	.994	.992
30	.912	.92	.958	.962	.994	.986
50	.914	.904	.958	.960	.994	.996
75	.922	.938	.966	.968	.996	.990
100	.89	.904	.952	.950	.986	.986

TABELA 5.6 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS PA
RES t_i . DISTRIBUIÇÃO DE ERROS CAUCHY.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.632	.732	.696	.794	.758	.850
20	.872	.878	.930	.926	.978	.974
30	.884	.910	.940	.964	.994	.992
50	.912	.912	.958	.938	.992	.986
75	.890	.904	.952	.950	.986	.986
100	.918	.884	.962	.930	1	.982

A partir destes resultados podemos concluir que em ge
ral o intervalo de confiança obtido pelo método padrão apresent
a bons resultados. As probabilidades de cobertura aproximams
e das probabilidades fixadas. Como poderíamos prever existe
uma variabilidade nos resultados em torno do valor fixado, por
ém esta variabilidade, com algumas exceções, não ultrapassa a
2%. Podemos verificar também que esta variabilidade aparece
com maior evidência para probabilidade fixada de 90%. Para val
ores fixados de 95% e 99% a estabilidade do método padrão é
bem maior.

Na utilização desta reamostragem encontramos dois probl
emas:

- i) No cálculo da estimativa do desvio padrão as correções entr
eadas foram para $n \leq 30$ e $n > 30$. Quando obtivemos as
probabilidades de cobertura verificamos que para $n = 30$ os
valores encontrados estavam um pouco distantes dos valores

fixados. Utilizamos então a correção para $n > 30$ e recalculamos estas probabilidades e verificamos que os novos resultados estavam bem mais próximos dos valores fixados do que os valores obtidos com a correção para $n \leq 30$. Portanto os valores apresentados para $n = 30$ nas Tabelas 5.3, 5.4, 5.5 e 5.6 utilizam a correção para $n > 30$.

ii) Quando temos a distribuição de erros Cauchy, as probabilidades de cobertura para $n = 10$ estão longe dos valores fixados, com $n = 20$ esta diferença é menor e apenas com $n = 30$ os resultados se apresentam de acordo com o comportamento global do método. Portanto isto nos indica que devemos ter cuidado em usar o método padrão quando a variância da distribuição de erros for infinita, como a Cauchy.

5.3 - RESULTADOS PARA REAMOSTRAGEM NOS ERROS

Os resultados obtidos para reamostragem nos erros $\hat{\epsilon}_i$ diferentes de zero são apresentados na Figura 5.8 que foi construída a partir de tabelas análogas a Tabela 5.2. Verificamos que o comportamento das estimativas do desvio padrão para β_0 e β_1 são semelhantes para tamanhos de amostra maior do que 30. Também podemos verificar que os pontos observados estão próximos da identidade ($\hat{\sigma} = \sigma_{MC}$). Estes dois fatos ficam mais evidentes nas Figuras 5.9 e 5.10 onde são apresentados individualmente os resultados obtidos para β_0 e β_1 , respectivamente.

Podemos verificar que para tamanhos de amostra maior ou igual a 30, em ambos os casos, os valores observados estão em torno da reta de identidade, alguns acima outros abaixo da mesma. Porém para $n < 30$, os valores observados encontram-se um

FIGURA 5.8

REAMOSTRAGEM NOS ERROS

MONTE CARLO x BOOTSTRAP

* = BETA(0)
+ = BETA(1)

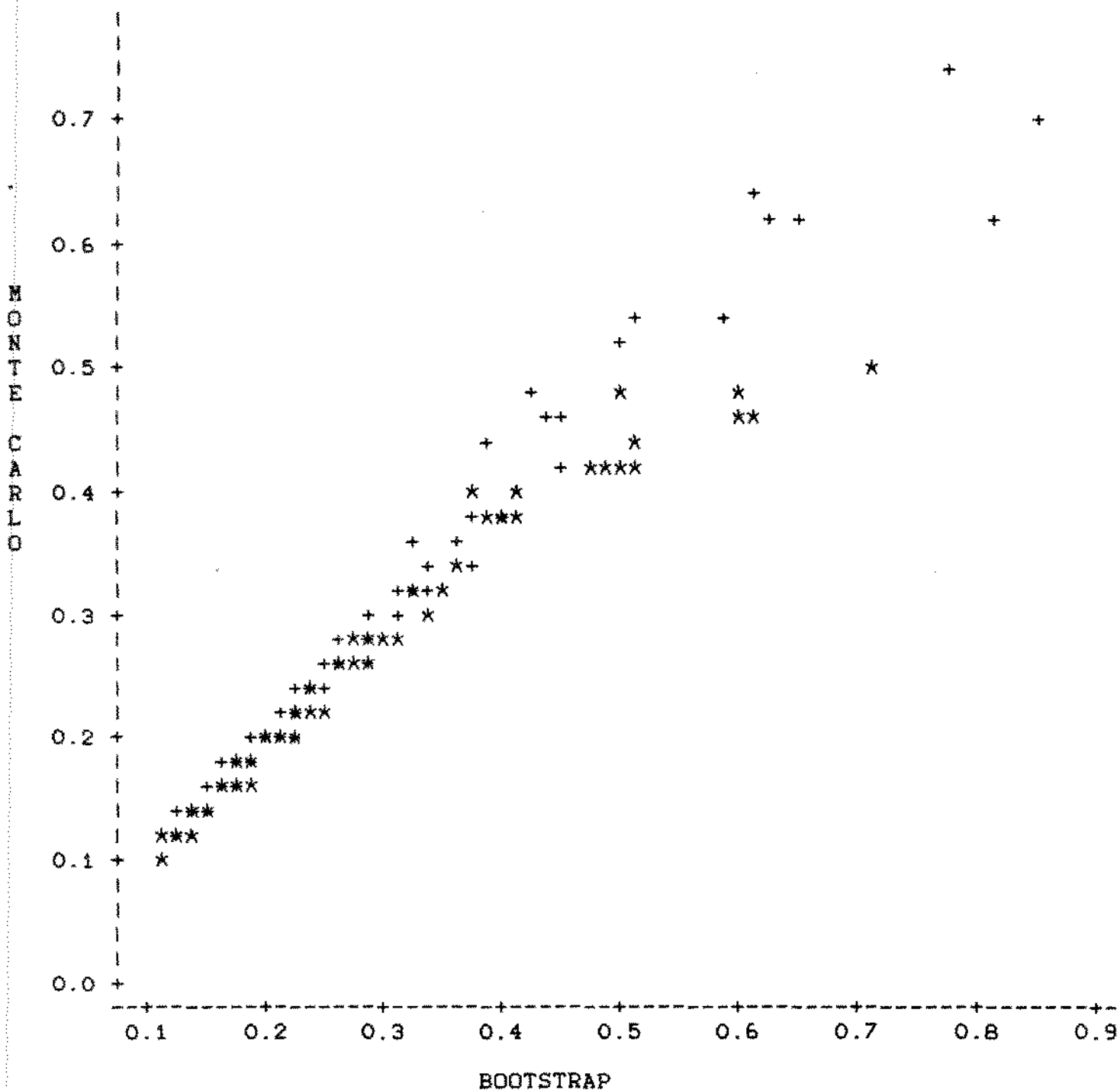


FIGURA 5.9

REAMOSTRAGEM NOS ERROS

MONTE CARLO BETA(0) X BOOTSTRAP BETA(0)

- 1 = tamanho de amostra 10
- 2 = tamanho de amostra 20
- 3 = tamanho de amostra 30
- 5 = tamanho de amostra 50
- 7 = tamanho de amostra 75
- 0 = tamanho de amostra 100

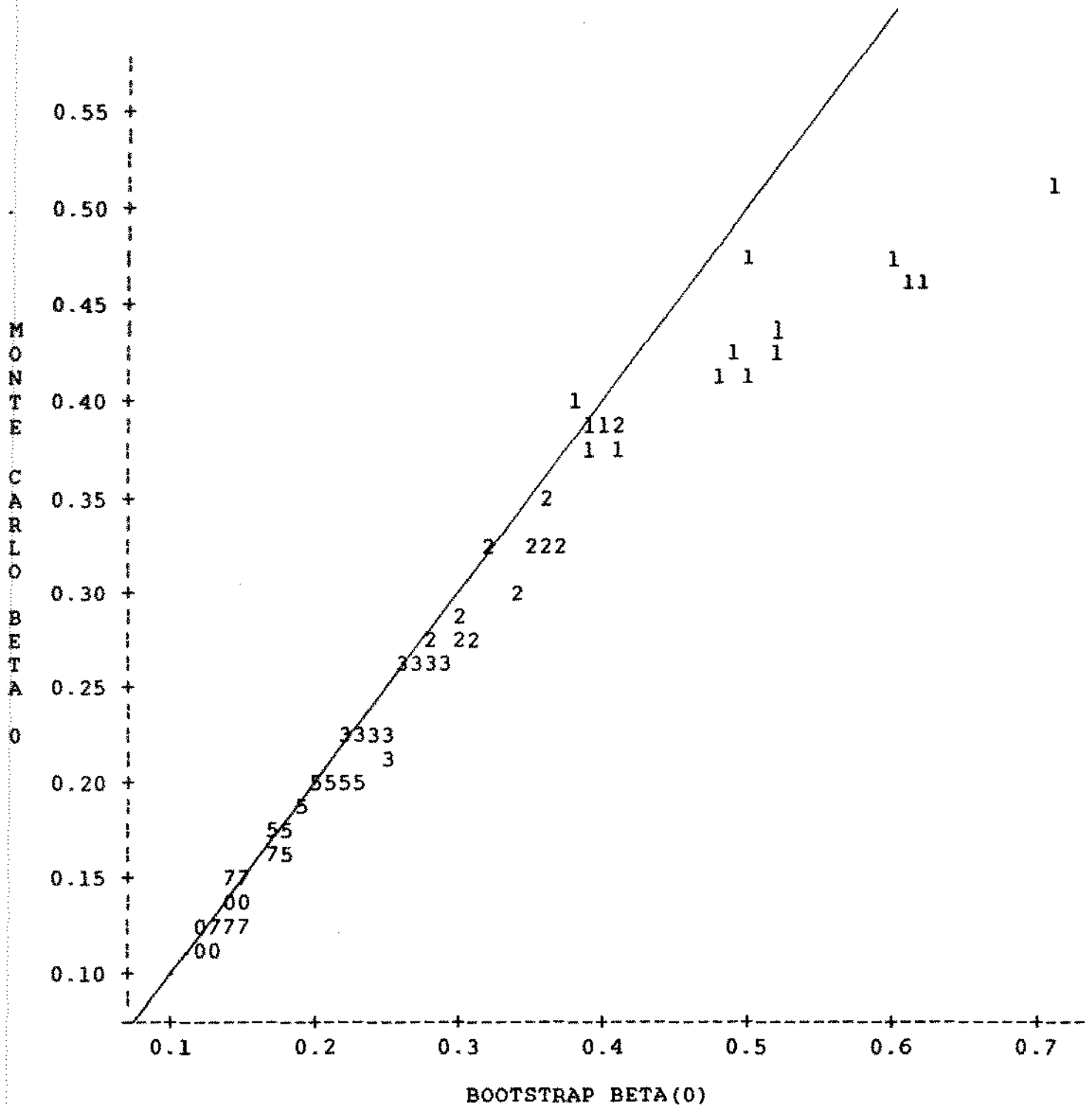
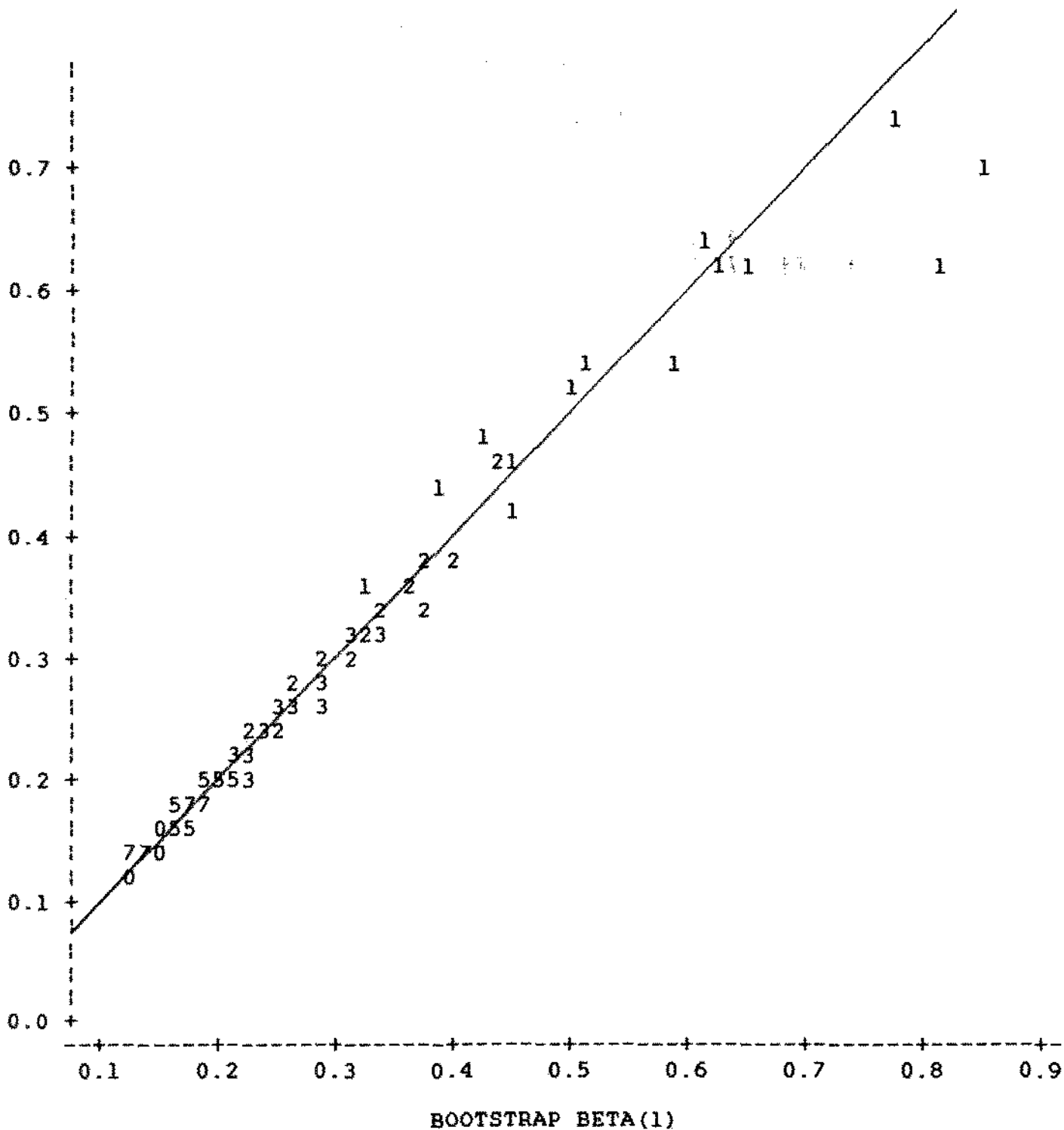


FIGURA 5.10

REAMOSTRAGEM NOS ERROS

MONTE CARLO BETA(1) X BOOTSTRAP BETA(1)

- 1 = tamanho de amostra 10
- 2 = tamanho de amostra 20
- 3 = tamanho de amostra 30
- 5 = tamanho de amostra 50
- 7 = tamanho de amostra 75
- 0 = tamanho de amostra 100



pouco abaixo da reta de identidade, indicando desta forma que o bootstrap está super-estimando o desvio padrão, principalmente nas estimativas para β_0 . Como na reamostragem nos pares, procuramos encontrar uma relação funcional entre o resultado do Monte Carlo e do bootstrap. Nas Figuras 5.11 e 5.12 são apresentados os resultados para os tamanhos de amostra com $n \leq 30$. Para estes dados ajustamos algumas possíveis relações entre o Monte Carlo e o bootstrap, as melhores funções obtidas foram um modelo quadrático para β_0 e um modelo linear para β_1 , porém estas correções não melhoraram a qualidade das estimativas, portanto não foram utilizadas.

A partir das estimativas do desvio padrão, construímos intervalos de confiança para os parâmetros usando o método padrão (4.5). As probabilidades de cobertura obtidas são apresentadas nas Tabelas 5.7, 5.8, 5.9, para as distribuições de erros normal, normal padrão e Laplace. Os resultados para a distribuição de erros Cauchy não serão apresentados pois os mesmos não se mostraram corretos, super-estimando totalmente as probabilidades fixadas.

FIGURA 5.11

REAMOSTRAGEM NOS ERROS

MONTE CARLO BETA(0) X BOOTSTRAP BETA(0)

TAMANHO DE AMOSTRA MENOR QUE 30

1 = tamanho de amostra 10 2 = tamanho de amostra 20
3 = tamanho de amostra 30

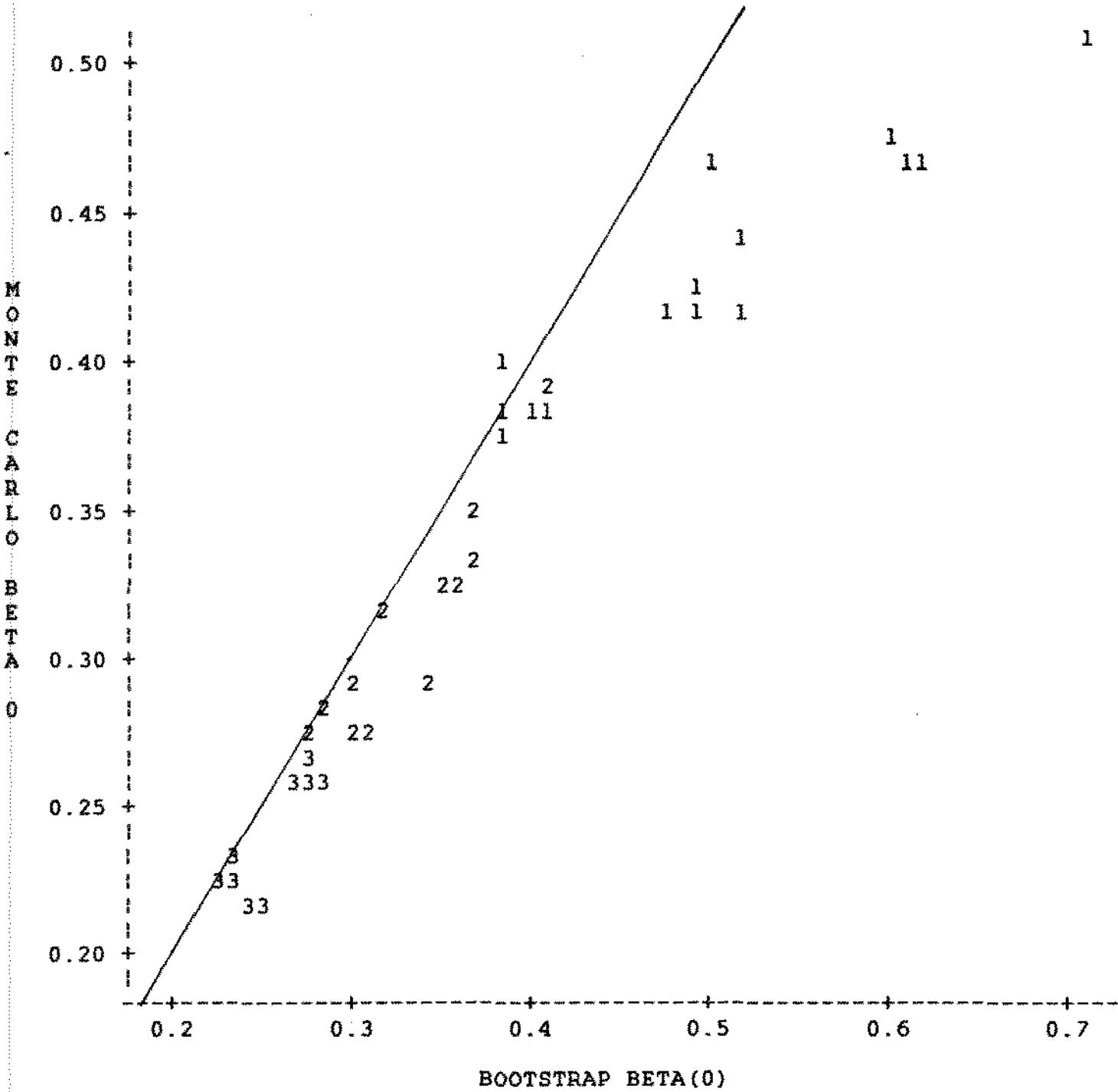


FIGURA 5.12

REAMOSTRAGEM NOS ERROS

MONTE CARLO BETA(1) X BOOTSTRAP BETA(1)

TAMANHO DE AMOSTRA MENOR QUE 30

1 = tamanho de amostra 10 2 = tamanho de amostra 20
3 = tamanho de amostra 30

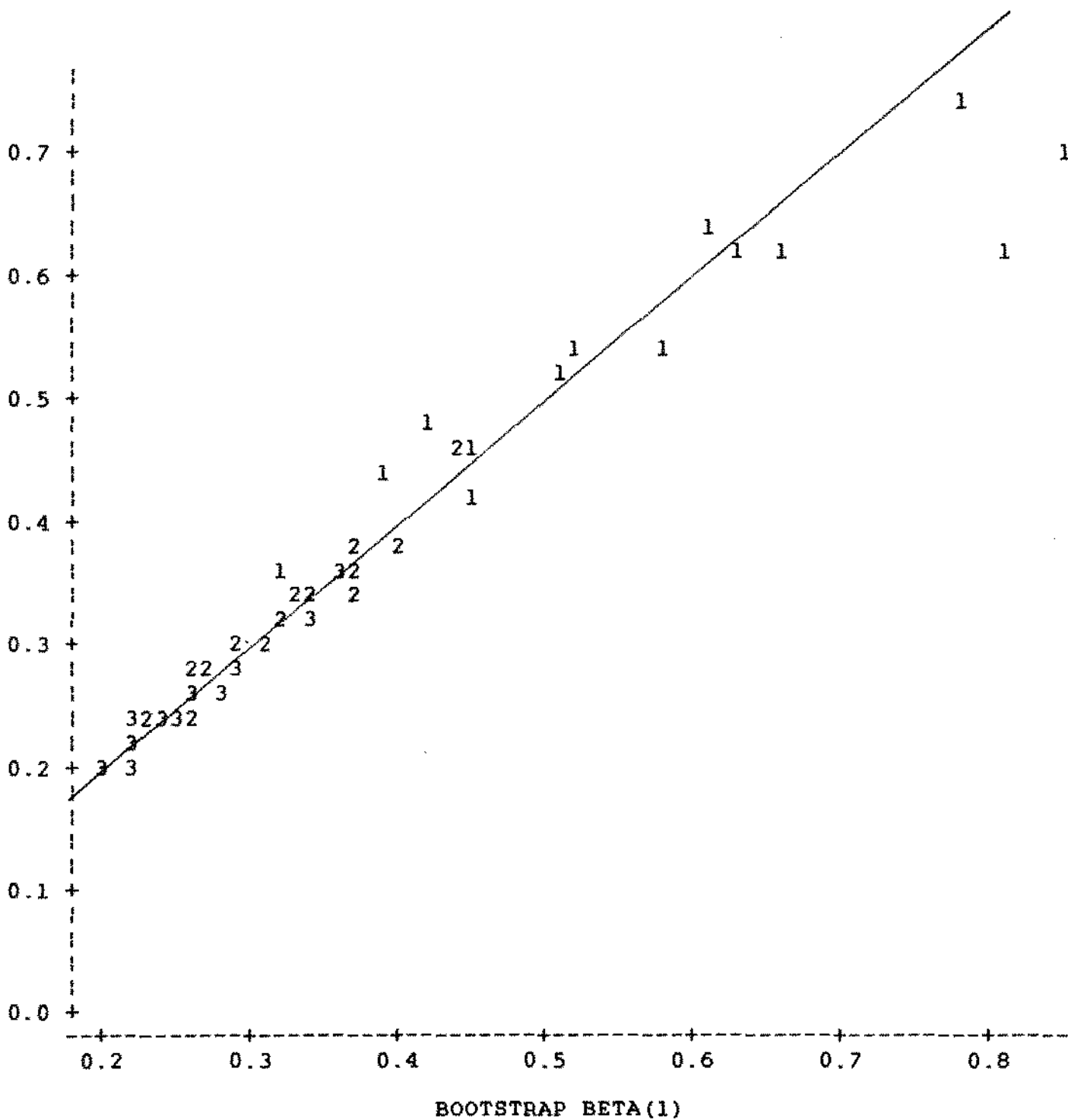


TABELA 5.7 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS ERROS. DISTRIBUIÇÃO DE ERROS NORMAL PADRÃO.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.896	.832	.920	.884	.964	.956
20	.932	.874	.966	.930	.984	.984
30	.908	.868	.946	.922	.984	.972
50	.910	.880	.950	.926	.992	.980
75	.918	.886	.958	.928	.996	.980
100	.916	.898	.956	.95	.990	.986

TABELA 5.8 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS ERROS. DISTRIBUIÇÃO DE ERROS NORMAL CONTAMINADA.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.912	.900	.946	.942	.988	.978
20	.914	.880	.946	.946	.988	.988
30	.904	.882	.944	.944	.984	.980
50	.920	.882	.962	.934	.996	.980
75	.930	.876	.964	.932	1	.986
100	.930	.916	.958	.934	.99	.980

TABELA 5.9 - PROBABILIDADES DE COBERTURA. REAMOSTRAGEM NOS ERROS. DISTRIBUIÇÃO DE ERROS LAPLACE.

n	PROBABILIDADES FIXADAS					
	90%		95%		99%	
	β_0	β_1	β_0	β_1	β_0	β_1
10	.92	.878	.954	.922	.982	.970
20	.938	.904	.978	.948	.992	.984
30	.928	.884	.966	.934	.994	.990
50	.914	.912	.952	.952	.986	.992
75	.920	.922	.96	.954	.996	.992
100	.938	.916	.974	.952	.990	.980

A partir destes valores verificamos que o método pa
drão utilizando a estimativa do desvio padrão obtida pelo boots
trap com reamostragem nos resíduos apresenta probabilidades de
cobertura com comportamento semelhante ao da reamostragem nos
pares. Os erros com relação ao valor fixado, com algumas exce
ções, estão ao redor de 2%. Porém, no caso de probabilidade fi
xada de 90% ou para tamanhos de amostras menores do que 30 a va
riabilidade em torno dos valores fixados é bem maior, indicando
uma maior instabilidade do método padrão para estes casos.

C A P Í T U L O 6

CONCLUSÕES

Pelos resultados apresentados no capítulo anterior fica evidente a viabilidade do uso do método bootstrap como procedimento para fins práticos na obtenção do desvio padrão das estimativas dos parâmetros da regressão L_1 e o seu uso na construção de intervalos de confiança, no caso de pequenas amostras.

A maior dificuldade na aplicação do método, ocorre quando os erros apresentam distribuição com variância infinita. No bootstrap com reamostragem nos erros, se a distribuição apresenta esta característica o método bootstrap não se mostra viável em qualquer uma das situações estudadas. Quando a reamostragem é feita nos pares, é esta característica que apresenta maior influência na obtenção da correção para os valores originais, e, mesmo com esta correção, os intervalos de confiança obtidos através do método padrão não são satisfatórios quando o tamanho de amostra é menor do que 30.

Finalmente, resta-nos a escolha do tipo de reamostragem que devemos utilizar na aplicação do método bootstrap. Pelos resultados apresentados, os dois métodos propostos tem comportamento semelhante, sendo que cada um deles apresenta problemas, levando à algumas restrições no seu uso.

No caso de reamostragem nos pares, necessitamos de

uma correção nos valores obtidos e, conseqüentemente estudos para casos com mais de uma variável independente precisam ser feitos para verificar o comportamento destas correções. A reamostragem nos erros mostra-se superior à medida que não necessita de correção dos valores obtidos, porém, torna-se inviável de ser utilizada quando a distribuição dos erros tem variância infinita.

Baseados nestes aspectos, podemos dividir a resposta da questão acima em duas partes:

- i) Se o tamanho de amostra for maior ou igual a 30 e se for possível mostrar que a distribuição dos erros não apresenta variância infinita, a reamostragem nos erros é recomendada.
- ii) Quando o tamanho de amostra for menor do que 30 ou a distribuição dos erros apresentar variância infinita, a reamostragem nos pares deve ser usada.

A P Ê N D I C E I

RECURSOS COMPUTACIONAIS

O método de regressão L_1 está disponível no procedimento PROC LAV do SAS (Statistical Analysis System), porém a aplicação do método bootstrap através do mesmo é inviável pela impossibilidade de geração de um arquivo com as estimativas dos parâmetros obtidas a cada execução do algoritmo.

Está em fase de desenvolvimento uma macro para obtenção da regressão L_1 utilizando o SOC (Software Científico) desenvolvido pela NTIA/EMBRAPA que é compatível com diferentes ambientes de hardware (superminis, supermicros e micros de 16 bits). Para construção desta macro estão sendo utilizados os recursos do módulo CM (Cálculo Matricial) do SOC.

Um programa em linguagem Fortran 77 que calcula a regressão L_1 e executa o método bootstrap com reamostragem nos pares ou nos erros obtendo no final um intervalo de confiança para os parâmetros da regressão L_1 pode ser obtido com o autor deste trabalho no seguinte endereço:

. Departamento de Estatística
UFSCar - Universidade Federal de São Carlos
Via Washington Luiz, Km 235
Caixa Postal 676
13.560 - São Carlos, S.P.

BIBLIOGRAFIA

- [1] ABDELMALEK, N.N. (1975). An Efficient Method for the Discrete Linear L_1 Approximation Problem. *Math.Comp.*, 29: 844-850.
- [2] ARMSTRONG, R.D. and KUNG, H.T. (1978). Algorithm AS132: Least Absolute Value Estimates for a Simple Linear Regression Problem. *Applied Statistics* 27: 363-366.
- [3] ARMSTRONG, R.D., FROME, E.L. and KUNG, D.S. (1979). A Revised Simplex Algorithm for the Absolute Deviation Curve-Fitting Problem. *Comun. Statist.-Simul. Comp.*, B8: 175-190.
- [4] ASHAR, V.G. and WALLACE, T.D. (1963). A Sampling Study of Minimum Absolute Deviations Estimators. *Operations Research*, 11: 747-758.
- [5] BARRODALE, I. and YOUNG, A. (1966). Algorithm for Best L_1 and L_∞ Linear Approximations on a Discrete Set. *Numer. Math.*, 8: 295-306.
- [6] BARRODALE, I. and ROBERTS, F.D.K. (1973). An Improved Algorithm for Discrete L_1 Linear Approximation. *SIAM J. Numer. Anal.*, 10: 839-848.
- [7] BASSET, G. and KOENKER, R. (1978). Asymptotic Theory of Least Absolute Error Regressions. *J. Amer. Ass.*, 73: 618-622.

- [8] BLATTBERG, R. and SARGENT, T. (1971). Regression with Non Gaussian Stable Disturbance: Some Sampling Results. *Econometrica*. 39: 501-510.
- [9] BLOOMFIELD, P. and STEIGER, W. (1980). Least Absolute Deviations Curve-Fitting. *SIAM J. Scient. and Statist. Comp.* 1: 290-301.
- [10] BOSCOVICH, R.J. (1757). De Litteraria Expeditione per Pontifician Ditionem et Synopsis Amplioris Operis, ac Habentur Plura Ejus ex Exemplaria Etian Sensorum Impressa. *Bononiensi Scientiarum et Artum Instituto Atque Academia Commentarii*. 4: 353-396.
- [11] COX, D.R. and HINKLEY, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.
- [12] CHARNES, A., COOPER, W.W. and FERGUSON, R. (1955). Optimal Estimation of Executive Compensation by Linear Programming. *Manag. Science*. 1: 138-151.
- [13] CRAMER, H. (1946). *Mathematical Methods of Statistics*, Princeton Statistics Press, Princeton.
- [14] DIELMAN, T. and PFAFFEMBERGER, R. (1982). An Examination of Inference Procedures for Least Absolute Value Regression, *Working Paper #4, Texas Christian University, Fort Worth*.
- [15] DIELMAN, T. and PFAFFEMBERGER, R. (1980). A Simulation Study of Inference Procedures for Least Absolute Value Regression. *Working Paper #7, Texas Christian University, Fort Worth*.

- [16] DIELMAN, T. and PFAFFEMBERGER, R. (1982). LAV (Least Absolute Value) Estimation in Linear Regression: A Review. *TIMS Studies in Manag. Science.* 19: 31-52.
- [17] EDGEWORTH, F.Y. (1987). On Observations Relating to Several Quantities. *Hermathena.* 6: 279-285.
- [18] EFROM, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics.* 7: 1-26.
- [19] EFROM, B. (1981). Nonparametric Standard Errors and Confidence Intervals. *The Canadian Journal of Statistics.* 9: 139-172.
- [20] EFROM, B. (1982). The Jackknife, the Bootstrap, and other Resampling Plaus. *SIAM CBMS-NSF, Monograph* 38.
- [21] EFROM, B. and TIBSHIRANI, R. (1985). The Bootstrap Method for Assessing Statistical Accuracy. *Technical Report #101, Stanford University, Stanford.*
- [22] EISENHART, C. (1961). *Boscovitch and the Combination of Observations, Roger J. Boscovitch, L.L. Whyte.* Fordham University Press, New York.
- [23] FOURIER, J.B.J. (1824). Solution d'une Question Particuliere au Calcul des Inegalites, second extrait. *Histoire de l'Academie des Sciences pour 1824:* 47-55.
- [24] GENTLE, J.E. (1977). Least Absolute Value Estimation: An Introduction. *Comun. Statist.-Simul. Comp.* B6: 313-328.
- [25] GENTLE, J.E., NARULA, S.C. and SPOSITO, V.A. (1987). A

gorithms for Unconstrained L_1 Linear Regression. *Statistical Data Analysis Based on the L_1 Norm and Related Methods*, ed. Y. Dodge, Elsevier Science Publishers: 83-94.

- [26] HADLEY, G. (1962). *Linear Programming*. Addison Wesley, Reading, M.A.
- [27] HOFFMAN, K.L. and SHIER, D.R. (1980). A Test Problem Generator for Discrete Linear L_1 Approximation Problems. *ACM Trans. Math. Software*. 6: 587-593.
- [28] HOGG, R.V. (1979). Statistical Robustness: One View of It's use in Applications Today. *The American Statistician*. 33: 108-115.
- [29] JACKSON, D. (1921). Note on the median of a Set of Numbers. *Bull. Amer. Math. Soc.* 27: 160-164.
- [30] JOSVANGER, L.A. and SPOSITO, V.A. (1983). L_1 -norm Estimates for the Simple Regression Problem. *Comun. Statist. Simul. Comp.* B12: 215-221.
- [31] KARST, O.J. (1958). Linear Curve-Fitting Using Least Deviations. *J. Amer. Statist. Ass.* 53: 118-132.
- [32] KENNEDY, W.J., GENTLE, J.E. and SPOSITO, V.A. (1977). A Computer Oriented Method for Generating Tests Problems for L_1 Regression. *Comun. Statist.-Simul. Comp.* B6: 21-27.
- [33] KOENKER, R. and BASSETT, G. (1982). Tests of Linear Hypotheses and L_1 Estimation. *Econometrica*. 50: 1577-1583.

- [34] LEGENDRE, A.M. (1805). *Nouvelles Methodes pour la Determination des Orbites des Cometes*, Paris: *Courcier* (esp. Appendice sur la Methode des Moindres Quarres): 72-80.
- [35] NARULA, S.C. and WELLINGTON, J.F. (1979). Selection of Variables in Linear Regression Using the Sum of Weighted Absolute Errors Criterion. *Technometrics*. 21: 299-306.
- [36] NARULA, S.C. and WELLINGTON, J.F. (1985). Interior Analysis for the Minimum Sum of Absolute Errors Regression. *Technometrics*. 27: 181-188.
- [37] PFAFFEMBERGER, R.C. and DINKEL, J.J. (1978). Absolute Deviations Curve Fitting: An Alternative to Least Squares. *Contributions to Survey Sampling and Applied Statistics*. Ed. H.A. David, Academic Press: 279-294.
- [38] RICE, J.R. and WHITE, J.S. (1964). Norms for Smoothing and Estimation. *SIAM Review*. 6: 243-256.
- [39] ROSEMBERG, B. and CARLSON, D. (1970). The Sampling Distribution of Least Absolute Residuals Regression Estimates. *Working Paper IP-164, Center of Research in Manag. Science*, University of California, Berkley.
- [40] ROSEMBERG, B. and CARLSON, D. (1977). A Simple Approximation of the Sampling Distribution of Least Absolute Regression Estimates. *Comun. Statist.-Simul. Comp.* B6: 421-437.
- [41] SCHLOSSMACHER, E.J. (1973). An Iterative Technique for Absolute Deviations Curve-Fitting. *J. Amer. Statist. Ass.* 68: 857-859.

- [42] SIELKEN, R.L. and HARTLEY, H.O. (1973). Two Linear Algorithms for Unbiased Estimation of Linear Models. *J. Amer. Statist. Ass.* 68: 639-641.
- [43] SINGLETON, J. (1940). A Method of Minimizing the Sum of Absolute Values of Deviations. *Annals of Math. Statist.* 11: 301-310.
- [44] SPOSITO, V.A. (1975). *Linear and Nonlinear Programming*, The Iowa State University Press, Ames, Iowa.
- [45] SPOSITO, V.A. (1976). Remark AS R14: A Remark on Algorithm AS74: L_1 -norm Fit of a Straight Line. *Applied Statistics.* 25: 96-97.
- [46] SPOSITO, V.A. (1987). Some Properties of L_p -Estimators, in *Robust Procedures*, Marcel Dekker, in print.
- [47] SPOSITO, V.A., SMITH, W. and MCCORMICK, G. (1978). *Minimizing the Sum of Absolute Deviations*. Göttingen: Vandenhoeck und Ruprecht.
- [48] TALARICO, R.P.M. (1984). *Regressão Linear Múltipla Através da Minimização da Soma dos Desvios Absolutos*, Tese de Mestrado, ICM-USP, São Carlos.
- [49] TAYLOR, L.D. (1974). Estimating by Minimizing the Sum of Absolute Errors. *Frontiers in Econometrics*, Academic Press, New York: 169-190.
- [50] WAGNER, H.M. (1959). Linear Programming Techniques for Regression Analysis. *J. Amer. Statist. Ass.* 54: 206-212.

- [51] WESOŁOWSKY, G.O. (1981). A New Descendent Algorithm for the Least Absolute Value Regression Problem. *Comun. Statist.-Simul. Comp.* B10: 641-648.
- [52] WILSON, H.G. (1978). Least Squares versus Minimum Absolute Deviations Estimation in Linear Models. *Decision Science.* 9: 322-335.