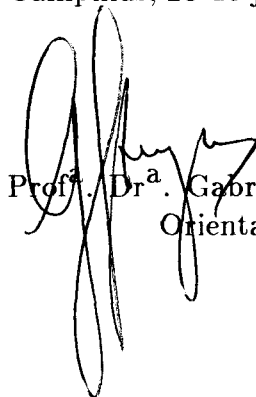


Um Estudo Comparativo de L-Estimadores de Regressão

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida pela Sra. Hildete Prisco Pinheiro e aprovada pela Comissão Julgadora.

Campinas, 28 de julho de 1992.



Prof^a. / Dr^a. Gabriela Stangenhaupt
Orientadora

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

Ao meu marido

Aluísio

Agradecimentos

À Prof^a. Dr^a. Gabriela Stangenhau, pela sua dedicação e trabalho durante a orientação desta dissertação.

Aos Profs. Drs. Reinaldo Charnet e Clóvis Araújo Peres, pelas sugestões dadas para o enriquecimento deste trabalho.

Aos meus amigos e colegas, pelo apoio.

Ao meu tio, Roberto Cláudio Frota Bezerra, pelo incentivo a fazer este mestrado.

À minha mãe, pelo carinho e incentivo.

Ao meu pai, José Tarquínio Prisco, meu exemplo acadêmico.

Sumário

Capítulo 1 - Introdução	1
Capítulo 2 - Regressão Quantílica	6
2.1 - Introdução	6
2.2 - Algumas Propriedades dos Estimadores de Regressão Quantílica	7
2.3 - Teoria Assintótica dos Estimadores de Regressão Quantílica	13
Capítulo 3 - "L-Estimadores de Regressão"	16
3.1 - Introdução	16
3.2 - Alguns L-Estimadores de Regressão	17
3.2.1 - Primeiro Método de Estimação ($A_{11}(\alpha)$)	19
3.2.2 - Segundo Método de Estimação ($A_{21}(\alpha)$)	24
3.2.3 - Terceiro Método de Estimação ($A_{31}(\alpha)$)	25
3.2.4 - Quarto Método de Estimação ($A_{41}(\alpha)$)	26
3.2.5 - Quinto Método de Estimação ($A_{12}(\alpha)$)	28
3.2.6 - Sexto Método de Estimação ($A_{22}(\alpha)$)	28
3.2.7 - Sétimo Método de Estimação ($A_{32}(\alpha)$)	28
3.2.8 - Oitavo Método de Estimação ($A_{42}(\alpha)$)	29
3.2.9 - Nono Método de Estimação ($A_4(\alpha)$)	29
3.2.10 - Trimédia de Tukey (T)	29
3.2.11 - Média de Gastwirth (G)	30
Capítulo 4 - Estudo de Monte Carlo	31
4.1 - Introdução	31
4.2 - Metodologia	31
4.2.1 - Métodos de Estimação	32
4.2.2 - Distribuições	32
4.2.3 - Geração das Distribuições	33
4.2.4 - Número de Repetições de Monte Carlo	35

4.2.5 - Método de Simulação	35
4.3 - Sementes	37
4.4 - Resultados	37
4.5 - Exemplo	40
4.6 - Conclusões	42
Apêndices	43
Apêndice A - Gráficos	44
Apêndice B - Tabelas	57
Apêndice C - Programas	70
Bibliografia	129

Capítulo 1

Introdução

Dentro da modelagem estatística, um dos modelos mais utilizados é o de regressão linear, onde tenta-se ajustar um modelo linear que possa explicar a relação existente entre uma variável dependente e variáveis independentes em estudo.

Entre as técnicas de estimação do modelo de regressão temos o método de Mínimos Quadrados, que é bastante utilizado desde a segunda metade do século XIX, pelo seu fácil cálculo e suas boas propriedades. Estes estimadores têm algumas particularidades interessantes: coincidem com os estimadores de máxima verossimilhança e são eficientes quando a distribuição dos erros é Normal. No entanto, há duas grandes desvantagens nestes estimadores: eles são muito afetados por pequenas contaminações no conjunto de dados, e se a distribuição dos erros tem caudas mais pesadas que a da Normal, não são eficientes. No caso do modelo de posição, o estimador de mínimos quadrados do parâmetro de locação é a média amostral, que é de fácil cálculo, de fácil interpretação, mas é muito sensível a modificações no conjunto de dados. Por exemplo, se tivermos um valor discrepante ("outlier") no conjunto de dados, a estimativa da média será bastante afetada por esse outlier. Assim, diante de situações como esta, é conveniente que busquemos métodos de estimação não muito sensíveis a modificações suaves nas suposições que o sustentam. Tais métodos são chamados *robustos*. Como exemplo de um estimador de locação robusto para o modelo de posição temos a mediana amostral, que na presença de outliers sua estimativa não é muito afetada por esses valores.

A partir da década de 60, alguns estatísticos interessaram-se pelo problema de estimação do parâmetro de locação quando as caudas da distribuição subjacente são mais alongadas que o esperado no caso de normalidade.

Hodges e Lehman (1963) propuseram estimar o parâmetro de locação a partir de testes de postos, que são os R-estimadores. Logo depois, Huber (1964) desenvolveu um estudo dentro da classe dos M-estimadores, os quais são obtidos por máxima verossimilhança. E ainda, quase na mesma época, Chernoff, Gastwirth, Johns e Bickel dedicaram-se ao estudo da teoria assintótica dos L-estimadores, ou seja, combinações lineares de estatísticas de ordem, que apresentam boas propriedades de robustez. Gastwirth (1965) propôs um estimador que consistia na média ponderada dos quantis $\frac{1}{3}$, $\frac{1}{2}$ e $\frac{2}{3}$, com pesos, respectivamente, 0.3, 0.4 e 0.3; Tukey (1977) sugeriu como estimador de locação a média ponderada do primeiro, segundo e terceiro quartis, tendo pesos, respectivamente, 0.25, 0.5 e 0.25. Ainda na década de 70, Andrews e outros fizeram um estudo detalhado sobre alguns estimadores de locação, entre estes: algumas médias aparadas, a média de Gastwirth, a Trimédia de Tukey e alguns M-estimadores. Neste trabalho, conhecido como o estudo de Princeton, verificou-se que a Média de Gastwirth e a Trimédia de Tukey apresentam boas performances, superando as médias aparadas, principalmente diante de distribuições com caudas mais alongadas que a Normal, como a Cauchy.

Mais tarde, Hogg (1975) introduziu a analogia de quantil amostral para regressão, através de um exemplo. Koenker e Bassett (1978) formalizaram o conceito de regressão quantílica, generalizando a definição de quantil amostral para a situação de regressão. Neste artigo, estudaram as propriedades e a teoria assintótica dos estimadores de regressão quantílica, que serão de grande utilidade para o nosso estudo, tendo em vista que muitos dos nossos estimadores são funções dos estimadores de regressões quantílicas.

Na década de 80, Bassett e Koenker continuaram aprofundando seus estudos na teoria de regressão quantílica. Bassett e Koenker (1982) definiram a *função quantílica condicional empírica* ($\hat{Q}(\theta/x)$), que nos diz para que valores de θ teremos iguais ou diferentes retas de regressão quantílica. Neste artigo, Bassett e Koenker comentam a importância da utilização das regressões quantílicas em situações de heteroscedasticidade nos dados.

Várias alternativas ao método de mínimos quadrados foram propostas. Hogg (1979) sugeriu como estimador de um parâmetro de escala a mediana dos desvios absolutos em torno da mediana (MAD). Com a utilização deste estimador de escala e baseado na curva de influência do estimador proposto por Huber em 1964, Hogg propôs um método de estimação alternativo ao de Mínimos Quadrados.

Outras alternativas foram sugeridas por Ruppert e Carroll (1980). Neste trabalho a estimativa dos parâmetros foi feita por *Mínimos Quadrados Aparados*, que é uma extensão do conceito de média aparada para regressão. Eles

consideraram dois métodos de *Mínimos Quadrados Aparados*. O primeiro consiste em aparar as observações de acordo com as regressões quantílicas e posteriormente calcular a estimativa de Mínimos Quadrados com o restante das observações. O segundo método baseia-se em aparar as observações a partir dos resíduos de uma estimativa preliminar e depois calcular a estimativa de Mínimos Quadrados com as observações restantes. Uma alternativa para a estimativa preliminar, proposta neste artigo de Ruppert e Carroll, é a média aritmética das regressões quantílicas de ordem α e $(1 - \alpha)$, $0 < \alpha < 1/2$, quando pretendemos aparar $|2\alpha n|\%$ observações. Um fato interessante é que, utilizando esta alternativa como estimativa preliminar, o comportamento assintótico dos dois métodos de Mínimos Quadrados Aparados são equivalentes. Além destes L-estimadores, alguns M-estimadores também foram estudados, entre eles o estimador de Huber e o de Hampel, cujas funções de influência são, respectivamente, limitada e redescendente. A fim de estudar o comportamento destes estimadores para tamanhos de amostra finitos, Ruppert e Carroll fizeram um estudo de Monte Carlo comparando estes e outros estimadores com o estimador de Mínimos Quadrados através da eficiência relativa de cada um deles. Neste estudo verificaram que os estimadores, cuja estimativa preliminar é Mínimos Quadrados ou L_1 , não são eficientes quando a distribuição dos erros é Normal. No entanto, quando a estimativa preliminar é a média aritmética das regressões quantílicas de ordem α e $(1 - \alpha)$, o estimador é bastante eficiente no modelo Normal. Diante de situações em que a distribuição dos erros tem caudas muito pesadas, os estimadores de *Mínimos Quadrados Aparados* são relativamente eficientes, comparados com Mínimos Quadrados ordinários. Vale ressaltar que o estimador de Huber e o de Hampel têm comportamentos semelhantes, sendo ambos muito eficientes no modelo Normal e com uma performance bastante aceitável diante de distribuições com caudas moderadamente longas.

Mais tarde, Welsh (1987) generalizou a classe dos L-estimadores para o modelo linear e fez um estudo de algumas médias aparadas em modelos de regressão. Ele utiliza estimadores análogos aos estudados por Ruppert e Carroll (1980), sendo que a estimativa final dos parâmetros é baseada na curva de influência da média aparada. Jongh e Wet (1987) comentaram sobre o estimador proposto por Welsh e analisaram um pequeno conjunto de dados, onde alguns estimadores foram comparados, entre estes: os estimadores propostos por Welsh e outro baseado no estimador de Welsh, mas que tem como estimativa preliminar um método L_1 ponderado, que foi definido em Jongh e Wet (1985). Neste estudo, observou-se que o melhor ajuste se deu com a utilização do estimador de Welsh, tendo como estimativa preliminar o método L_1 ponderado. Concluíram que a média aparada de Welsh tem um bom comportamento diante de situações em que os dados provêm de distribuições com caudas alongadas, mas a estimativa preliminar deve ser escolhida com cautela, pois esta tem grande influência no resultado final do estimador de Welsh. Koenker (1987), que também comentou sobre o artigo de Welsh, fez um estudo de Monte Carlo, a fim de comparar alguns

estimadores para o modelo linear. Ele utilizou modelos com 3, 5 e 10 parâmetros, e as observações de \mathbf{Y} e \mathbf{X} foram geradas de acordo com diferentes distribuições. Os métodos de estimação utilizados foram: Mínimos Quadrados Ordinários (MQO); Mínimos Quadrados Aparados (MQA), que foram estudados por Ruppert e Carroll (1980); Regressão Quantílica Aparada (RQA), um método de estimação proposto por Koenker e Portnoy (1987); o estimador de Welsh com estimativa preliminar sendo L_1 ($1SW L_1$); e o estimador de Welsh com estimativa preliminar sendo Mínimos Quadrados ($1SW L_2$). Neste estudo, Koenker pôde concluir que, os outros estimadores são claramente superiores aos MQA, principalmente nos casos em que os dados provêm de distribuições normais. Percebe-se também que, apesar de $1SW L_2$ ter um bom comportamento diante de distribuições com curtoses moderadas, ele não tem uma performance aceitável em situações em que as caudas da distribuição são muito alongadas como a Cauchy. Outro fato a respeito do estimador de MQA, é que sua performance deteriora-se rapidamente a medida em que p (número de parâmetros) se aproxima de n (número de observações), e esta tendência é acentuada quando a distribuição que gerou \mathbf{X} tem um alto grau de curtose. Quanto ao estimador $1SW L_1$, sua performance é extremamente boa, e assim como o estimador RQA, ambos mostram uma tendência muito pequena de piorarem quando p é moderadamente grande.

Na tentativa de buscar alternativas ao método de Mínimos Quadrados, principalmente nos casos em que a distribuição dos erros tem caudas mais alongadas que a Normal, propusemos estudar alguns *L-estimadores de regressão*, isto é, L-estimadores estendidos para o caso de regressão. Alguns estimadores, como a Média de Gastwirth e a Trimédia de Tukey, que já foram estudados no caso do modelo de posição, em nosso estudo, terão seus conceitos estendidos para o modelo de regressão. Entre os L-estimadores de regressão, estudaremos alguns estimadores baseados em regressões quantílicas, que serão combinações lineares dos estimadores de regressão quantílica; os estimadores de mínimos quadrados aparados propostos por Ruppert e Carroll (1980); e outros análogos a estes, sendo que a estimativa final é calculada por L_1 , pois nos casos em que a distribuição dos erros tem caudas muito pesadas, como a Cauchy, mesmo depois de apararmos as observações, ainda podem existir alguns outliers no conjunto de dados, e nestas situações o estimador L_1 seria mais adequado.

Como muitos dos nossos estimadores são funções dos estimadores de regressão quantílica, no capítulo 2 faremos uma revisão do artigo de Koenker e Bassett (1978), onde introduziremos o conceito de regressão quantílica e falaremos sobre as propriedades dos estimadores de regressão quantílica e a teoria assintótica dos mesmos.

No capítulo 3 faremos a descrição de todos os estimadores de interesse para o nosso estudo, onde definiremos cada um deles e trataremos do

comportamento assintótico de alguns destes estimadores.

A fim de observar a performance destes estimadores foi feito um estudo de Monte Carlo, que será tratado no capítulo 4, com a documentação das simulações e apresentação dos resultados e conclusões obtidos neste estudo. Para esse estudo utilizamos o modelo de regressão linear simples. As observações da variável independente foram geradas de acordo com a distribuição Normal Padrão, enquanto que os erros foram gerados de acordo com diferentes distribuições de baixo, médio e alto grau de curtose. A comparação destes estimadores foi feita a partir dos erros quadráticos médios.

Capítulo 2

Regressão Quantílica

2.1 Introdução

O conceito de *Regressão Quantílica*, introduzido por Koenker e Bassett (1978), foi idealizado como uma extensão do conceito de *quantil amostral*, cuja definição coincidirá com a definição de regressão quantílica no caso do modelo de posição.

Um caso particular importante de regressão quantílica é a regressão L_1 , isto é, a minimização dos desvios absolutos. No modelo de posição o estimador de norma L_1 é a mediana, que sabemos ser um estimador melhor do que a média, quando os dados provêm de distribuições com caudas alongadas ou quando temos outliers no conjunto de dados.

A teoria de regressão quantílica será de grande interesse para o nosso estudo, visto que alguns dos estimadores propostos nesta monografia são funções de estimadores de regressão quantílica. Assim, neste capítulo trataremos de algumas propriedades e da teoria assintótica dos estimadores de regressão quantílica. Foge ao objetivo da tese a demonstração de todos os teoremas envolvidos no assunto em questão. Entretanto, para alguns teoremas faremos uso da demonstração, a fim de esclarecer o leitor quanto a possíveis dúvidas ou equívocos da bibliografia.

2.2 Algumas Propriedades dos Estimadores de Regressão Quantílica

Para melhor compreensão do conceito de *Regressão Quantílica*, definiremos primeiramente *quantil amostral*.

Definição 2.1 *Seja y_1, y_2, \dots, y_n uma amostra aleatória de uma variável aleatória Y com função distribuição F .*

O θ -ésimo quantil amostral, $0 < \theta < 1$, pode ser definido como qualquer solução do problema de minimização:

$$\min_{b \in \mathbb{R}} \left[\sum_{i \in \{i: y_i \geq b\}} \theta |y_i - b| + \sum_{i \in \{i: y_i < b\}} (1 - \theta) |y_i - b| \right] \quad (2.2.1)$$

Por exemplo, no caso em que $\theta = 1/2$ o quantil amostral de ordem $1/2$ é a mediana amostral.

A fim de estender a idéia de quantil amostral para regressão, definiremos o modelo de regressão como sendo:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}, \quad (2.2.2)$$

onde $\mathbf{Y}' = (y_1, y_2, \dots, y_n)$, \mathbf{X} é uma matriz $n \times p$ de valores conhecidos, $\beta' = (\beta_1, \dots, \beta_p)$ é um vetor de parâmetros desconhecidos, e $\mathbf{Z}' = (Z_1, \dots, Z_n)$ é um vetor de variáveis aleatórias independentes identicamente distribuídas com função distribuição F .

Assim, temos:

Definição 2.2 *A θ -ésima regressão quantílica, $0 < \theta < 1$, é definida como qualquer solução para o problema de minimização:*

$$\min_{\beta \in \mathbb{R}^p} \left[\sum_{i \in \{i: y_i \geq \mathbf{x}_i \beta\}} \theta |y_i - \mathbf{x}_i \beta| + \sum_{i \in \{i: y_i < \mathbf{x}_i \beta\}} (1 - \theta) |y_i - \mathbf{x}_i \beta| \right], \quad (2.2.3)$$

onde \mathbf{x}_i representa uma linha da matriz \mathbf{X} e β é um vetor coluna de dimensão $p \times 1$.

Observe que no caso do modelo de posição ($p = 1$, $x_i = 1$, para todo i) os problemas de minimização citados acima coincidem.

O parâmetro θ pondera a regressão quantílica, isto é, θ localiza a reta de regressão no conjunto de dados de forma que, no máximo $100\theta\%$ das observações têm resíduos positivos e no máximo $100(1 - \theta)\%$ das observações têm resíduos negativos. Por exemplo, quando $\theta = 1/2$, temos a regressão L_1 , ou seja, a minimização dos desvios absolutos.

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

Seja $\mathcal{T} = \{1, 2, \dots, n\}$ e \mathcal{H} um conjunto de subconjuntos de dimensão p de \mathcal{T} . Os elementos $h \in \mathcal{H}$ e $\bar{h} = \mathcal{T} - h$, ambos servem para particionar \mathbf{Y} e \mathbf{X} . Por exemplo: se temos um vetor $\mathbf{Y}(h)$ de p elementos $\{y_i : i \in h\}$, então $\mathbf{X}(\bar{h})$ será a matriz de dimensão $(n - p) \times p$ com linhas definidas por $\{x_i : i \in \bar{h}\}$. Seja também o conjunto

$$\mathbf{H} = \{h \in \mathcal{H} \mid \text{posto } \mathbf{X}(h) = p\} \quad (2.2.4)$$

e $\mathbf{1}_p$ o vetor composto de p elementos iguais a um.

Teorema 2.1 *Se \mathbf{X} tem posto p então o conjunto $B^*(\theta)$ tem no mínimo um elemento da forma:*

$$\beta^*(\theta) = \mathbf{X}(h)^{-1} \mathbf{Y}(h) \quad (2.2.5)$$

para algum $h \in \mathbf{H}$, onde $\mathbf{X}(h)$ é uma matriz de dimensão $p \times p$ e $B^(\theta)$ é o fecho convexo de todas as soluções possíveis da forma (2.2.5).*

PROVA:

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

$$\min[\theta \mathbf{1}_n' \mathbf{Z}^+ + (1 - \theta) \mathbf{1}_n' \mathbf{Z}^-], \quad (\text{P})$$

sujeito a

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\beta + \mathbf{I}_n \mathbf{Z}^+ - \mathbf{I}_n \mathbf{Z}^- \\ (\beta, \mathbf{Z}^+, \mathbf{Z}^-) &\in \mathbb{R}^p \times \mathbb{R}_+^{2n}, \end{aligned}$$

onde $\mathbf{1}_n$ é um vetor de n uns, \mathbf{I}_n é a matriz identidade de ordem n e \mathbf{Z}^+ , \mathbf{Z}^- são os vetores dos valores absolutos dos erros não negativos e não positivos, respectivamente, sendo ambos completados com zeros e de dimensão $n \times 1$.

O dual de (P) é um problema de variáveis limitadas, e é definido como

$$\max[\mathbf{Y}'\mathbf{d}], \quad (\text{D})$$

sujeito a

$$\mathbf{X}'\mathbf{d} = \mathbf{0}$$

$$\mathbf{d} \in [\theta - 1, \theta]^n,$$

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

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

Este teorema nos mostra que pelo menos uma regressão quantílica passa por pelo menos p observações. Isto significa que se tivermos dois parâmetros no modelo, existirá pelo menos uma regressão quantílica que passa por pelo menos duas observações.

Teorema 2.2 *Se $\beta^*(\theta, \mathbf{Y}, \mathbf{X}) \in B^*(\theta, \mathbf{Y}, \mathbf{X})$ então os elementos abaixo serão soluções dos respectivos problemas transformados:*

- (i) $\beta^*(\theta, \lambda \mathbf{Y}, \mathbf{X}) = \lambda \beta^*(\theta, \mathbf{Y}, \mathbf{X}), \lambda \in [0, \infty)$
- (ii) $\beta^*(1 - \theta, \lambda \mathbf{Y}, \mathbf{X}) = -\lambda \beta^*(\theta, \mathbf{Y}, \mathbf{X}), \lambda \in (-\infty, 0]$
- (iii) $\beta^*(\theta, \mathbf{Y} + \mathbf{X}\gamma, \mathbf{X}) = \beta^*(\theta, \mathbf{Y}, \mathbf{X}) + \gamma, \gamma \in \mathbb{R}^p$
- (iv) $\beta^*(\theta, \mathbf{Y}, \mathbf{X}\mathbf{A}) = \mathbf{A}^{-1}\beta^*(\theta, \mathbf{Y}, \mathbf{X}), \mathbf{A}_{p \times p}$ não singular.

PROVA:

Seja

$$\begin{aligned} \psi(\beta, \theta, \mathbf{Y}, \mathbf{X}) &= \sum_{i \in \{i: y_i \geq \mathbf{x}_i \beta\}} \theta |y_i - \mathbf{x}_i \beta| + \sum_{i \in \{i: y_i < \mathbf{x}_i \beta\}} (1 - \theta) |y_i - \mathbf{x}_i \beta| = \\ &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} |y_i - \mathbf{x}_i \beta|, \end{aligned}$$

onde

$$\text{sign}(u) = \begin{cases} 1 & \text{se } u > 0 \\ 0 & \text{se } u = 0 \\ -1 & \text{se } u < 0 \end{cases}$$

Agora, note que:

- i) $\lambda\psi(\beta; \theta, \mathbf{Y}, \mathbf{X}) = \psi(\lambda\beta; \theta, \lambda\mathbf{Y}, \mathbf{X}), \lambda \in [0, \infty)$
- ii) $-\lambda\psi(\beta; \theta, \mathbf{Y}, \mathbf{X}) = \psi(\lambda\beta; (1 - \theta), \lambda\mathbf{Y}, \mathbf{X}), \lambda \in (-\infty, 0]$
- iii) $\psi(\beta; \theta, \mathbf{Y}, \mathbf{X}) = \psi(\beta + \gamma; \theta, \mathbf{Y} + \mathbf{X}\gamma, \mathbf{X}), \gamma \in \mathbb{R}^p$
- iv) $\psi(\beta; \theta, \mathbf{Y}, \mathbf{X}) = \psi(\mathbf{A}^{-1}\beta; \theta, \mathbf{Y}, \mathbf{XA}), |\mathbf{A}_{p \times p}| \neq 0$

pois,

i) Como $\lambda \in [0, \infty)$,

$$\begin{aligned} \psi(\lambda\beta; \theta, \lambda\mathbf{Y}, \mathbf{X}) &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(\lambda y_i - \mathbf{x}_i \lambda \beta) \right\} \lambda [y_i - \mathbf{x}_i \beta] = \\ &= \lambda \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} [y_i - \mathbf{x}_i \beta] = \lambda \psi(\beta; \theta, \mathbf{Y}, \mathbf{X}); \end{aligned}$$

ii) Como $\lambda \in (-\infty, 0]$,

$$\begin{aligned} \psi(\lambda\beta; (1 - \theta), \lambda\mathbf{Y}, \mathbf{X}) &= \sum_{i=1}^n \left\{ 1 - \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(\lambda y_i - \mathbf{x}_i \lambda \beta) \right\} \lambda [y_i - \mathbf{x}_i \beta] = \\ &= \sum_{i=1}^n \left\{ \frac{1}{2} - \theta - \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} \lambda [y_i - \mathbf{x}_i \beta] = \\ &= -\lambda \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} [y_i - \mathbf{x}_i \beta] = -\lambda \psi(\beta; \theta, \mathbf{Y}, \mathbf{X}); \end{aligned}$$

iii) $\psi(\beta + \gamma; \theta, \mathbf{Y} + \mathbf{X}\gamma, \mathbf{X}) =$

$$\begin{aligned} &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i + \mathbf{x}_i \gamma - \mathbf{x}_i (\beta + \gamma)) \right\} [y_i + \mathbf{x}_i \gamma - \mathbf{x}_i (\beta + \gamma)] = \\ &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} [y_i - \mathbf{x}_i \beta] = \psi(\beta; \theta, \mathbf{Y}, \mathbf{X}), \gamma \in \mathbb{R}^p; \end{aligned}$$

iv) Como $\mathbf{A}_{p \times p}$ é uma matriz não singular,

$$\begin{aligned} \psi(\mathbf{A}^{-1}\beta, \theta, \mathbf{Y}, \mathbf{XA}) &= \\ &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \mathbf{A} \mathbf{A}^{-1} \beta) \right\} [y_i - \mathbf{x}_i \mathbf{A} \mathbf{A}^{-1} \beta] = \\ &= \sum_{i=1}^n \left\{ \theta - \frac{1}{2} + \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta) \right\} [y_i - \mathbf{x}_i \beta] = \psi(\beta; \theta, \mathbf{Y}, \mathbf{X}). \end{aligned}$$

Logo, se $\beta^*(\theta, \mathbf{Y}, \mathbf{X})$ é solução do problema de minimização de $\psi(\beta; \theta, \mathbf{Y}, \mathbf{X})$ segue o teorema. \square

Interpretando os resultados deste teorema diremos que as expressões (i) e (ii) nos mostram que a solução da regressão L_1 é equivariante em escala, (iii) nos mostra que $\beta^*(\theta)$ é equivariante de locação, e (iv) implica que $\beta^*(\theta)$ é equivariante à reparametrização da matriz de planejamento.

Teorema 2.3 (Koenker e Bassett (1978))

Se F é contínua então, com probabilidade um, $\beta^*(\theta) = \mathbf{X}(h)^{-1}\mathbf{Y}(h)$ é a única solução do problema de regressão quantílica se, e somente se,

$$(\theta - 1)\mathbf{1}_p' < \sum_{i \in \bar{h}} \left[\frac{1}{2} - \frac{1}{2} \text{sign}(y_i - \mathbf{x}_i \beta^*(\theta)) - \theta \right] \mathbf{x}_i \mathbf{X}(h)^{-1} < \theta \mathbf{1}_p' \quad (2.2.6)$$

Analisando a expressão (2.2.6) no caso em que $\mathbf{x}_i = 1$ para todo $i = 1, \dots, n$, $\mathbf{H} = \mathcal{T}$ e F é contínua, teremos que $\beta^*(\theta) = \mathbf{Y}(h)$ é o único θ -ésimo quantil amostral se, e somente se,

$$(\theta - 1) < \sum_{i \in \bar{h}} \left[\frac{1}{2} - \frac{1}{2} \text{sign}(y_i - y(h)) - \theta \right] < \theta \quad (2.2.7)$$

Esta expressão implica que o número de y_i 's menores que $y(h)$ estão estritamente entre $n\theta - 1$ e $n\theta$, pois sabemos que o conjunto \bar{h} tem $n - p$ elementos (neste caso particular, $n - 1$, já que $p = 1$).

Sendo a o número de y_i 's menores que $y(h)$ e b o número de y_i 's maiores que $y(h)$, então $a + b = n - 1$. Assim,

Se

$$(\theta - 1) < \sum_{i \in \bar{h}} \left[\frac{1}{2} - \frac{1}{2} \text{sign}(y_i - y(h)) - \theta \right] < \theta,$$

Então,

$$\begin{aligned} (\theta - 1) &< [a(1 - \theta) - b\theta] < \theta \\ (\theta - 1) &< [a(1 - \theta) - (n - 1 - a)\theta] < \theta \\ (\theta - 1) &< [a - n\theta + \theta] < \theta \end{aligned}$$

Logo,

$$n\theta - 1 < a < n\theta.$$

A continuidade de F elimina o problema de empates no modelo de locação, o mesmo ocorrendo para o modelo linear. Nessa condição, a solução única depende da estrutura da matriz \mathbf{X} . No caso do modelo de locação, $n\theta$ terá que ser não inteiro para que a solução seja única. Uma alternativa muito utilizada no modelo de locação, quando a solução não é única, é adotar um critério para escolher um único elemento de um conjunto de soluções existentes.

Teorema 2.4 Seja $P(u^*(\theta))$, $N(u^*(\theta))$, e $Z(u^*(\theta))$ o número de elementos positivos, negativos e iguais a zero no vetor $\mathbf{u}^*(\theta) = \mathbf{Y} - \mathbf{X}\beta^*(\theta)$. Então se \mathbf{X} contém uma coluna de uns,

$$N(u^*) \leq n\theta \leq n - P(u^*) = N(u^*) + Z(u^*) \quad (2.2.8)$$

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

PROVA:

Particionemos a matriz \mathbf{X} como $\mathbf{X} = [\mathbf{1} : \tilde{\mathbf{X}}]$.

Pelo argumento usado no Teorema 2.3, $\beta^*(\theta) \in B^*(\theta)$ se, e somente se,

$$\psi'(\beta^*(\theta); w) = \sum_{i=1}^n \left[\frac{1}{2} - \frac{1}{2} \text{sign}^*(y_i - \mathbf{x}_i \beta^*(\theta); -\mathbf{x}_i w) - \theta \right] \mathbf{x}_i w \geq 0, \quad (2.2.9)$$

onde $\psi'(\beta^*; w)$ é a derivada de $\psi(\beta^*; \theta, \mathbf{Y}, \mathbf{X})$ na direção w , para todo $w \neq 0$ e

$$\text{sign}^*(u, z) = \begin{cases} \text{sign}(u) & \text{se } u \neq 0 \\ \text{sign}(z) & \text{se } u = 0 \end{cases}$$

Em particular, (2.2.9) vale para $w^+ = (1, 0, \dots, 0)'$ e $w^- = (-1, 0, \dots, 0)'$ em \mathbb{R}^p . Então $\mathbf{x}_i w^+ = 1$ e $\mathbf{x}_i w^- = -1$ para todo \mathbf{x}_i , (2.2.9) implica que

$$\sum_{i=1}^n \pm \left[\frac{1}{2} - \frac{1}{2} \text{sign}^*(y_i - \mathbf{x}_i \beta^*(\theta); \mp 1) - \theta \right] \geq 0. \quad (2.2.10)$$

A desigualdade (2.2.10) é equivalente a:

$$\begin{aligned} i) \quad & -\theta P + (1 - \theta)N + (1 - \theta)Z \geq 0, \\ ii) \quad & -\theta P + (1 - \theta)N - \theta Z \leq 0, \end{aligned}$$

que levam a (2.2.8), pois

$$\begin{aligned} & -\theta P + (1 - \theta)N + (1 - \theta)Z = -\theta(P + N + Z) + N + Z = \\ = & -n\theta + N + Z \geq 0 \Rightarrow n\theta \leq N + Z = n - P \\ \text{e} \\ & -\theta P + (1 - \theta)N - \theta Z = -n\theta + N \leq 0 \Rightarrow n\theta \geq N \end{aligned}$$

Assim,

$$N \leq n\theta \leq n - P = N + Z \quad \square$$

Se $\beta^*(\theta)$ é única, então todas as desigualdades são estritas.

Note que se F é uma função contínua então $Z(u^*) = p$ com probabilidade um, então existirão no máximo $n\theta$ observações com resíduos negativos e no máximo $n(1 - \theta)$ observações com resíduos positivos.

Teorema 2.5 Se $\beta^*(\theta) \in B^*(\theta, \mathbf{Y}, \mathbf{X})$, então

$$\beta^*(\theta) \in B^*(\theta, \mathbf{X}\beta^*(\theta) + \mathbf{D}u^*, \mathbf{X}),$$

onde $u^* = \mathbf{Y} - \mathbf{X}\beta^*(\theta)$ e \mathbf{D} é qualquer matriz diagonal $n \times n$ com elementos não negativos.

PROVA:

Sabemos que $\beta^*(\theta) \in B^*(\theta, \mathbf{Y}, \mathbf{X})$ se, e somente se,

$$\psi'(\beta^*(\theta); w) = \sum_{i=1}^n \left[\frac{1}{2} - \frac{1}{2} \text{sign}^*(y_i - \mathbf{x}_i \beta^*(\theta); -\mathbf{x}_i w) - \theta \right] \mathbf{x}_i w \geq 0$$

para todo $w \neq 0$.

Notemos que:

$$\begin{aligned} & \left[\frac{1}{2} - \frac{1}{2} \text{sign}^*(\mathbf{x}_i \beta^*(\theta) + \mathbf{d}_i(y_i - \mathbf{x}_i \beta^*(\theta)) - \mathbf{x}_i \beta^*(\theta); -\mathbf{x}_i w) - \theta \right] \mathbf{x}_i w = \\ & - \left(\frac{1}{2} - \theta \right) \mathbf{x}_i w - \frac{1}{2} \text{sign}^*(\mathbf{d}_i(y_i - \mathbf{x}_i \beta^*(\theta)); -\mathbf{x}_i w) \mathbf{x}_i w \\ & \geq \left(\frac{1}{2} - \theta \right) \mathbf{x}_i w - \frac{1}{2} \text{sign}^*(y_i - \mathbf{x}_i \beta^*(\theta); -\mathbf{x}_i w) \mathbf{x}_i w, \end{aligned}$$

para $\mathbf{d}_i \geq 0$.

Então:

$$\psi'(\beta^*(\theta); \mathbf{X}\beta^*(\theta) + \mathbf{D}u^*, \mathbf{X}, w) \geq \psi'(\beta^*(\theta); w) \geq 0$$

E daí segue o teorema. \square

Este teorema tem uma interpretação geométrica interessante. Imaginemos várias observações dispersas em \mathbb{R}^p com a θ -ésima regressão quantílica passando através destes pontos. Considerando o efeito do movimento das observações com resíduos positivos ou negativos, de modo que cada observação continue do mesmo lado do hiperplano (isto é, o sinal de cada resíduo não se altera), então a θ -ésima regressão quantílica não se altera.

2.3 Teoria Assintótica dos Estimadores de Regressão Quantílica

Estudaremos nesta seção a distribuição assintótica dos estimadores da θ -ésima regressão quantílica.

Teorema 2.6 (Koenker e Bassett (1978))

Seja $\{\xi_n^*(\theta_1), \dots, \xi_n^*(\theta_m)\}$ com $0 < \theta_1 < \theta_2 < \dots < \theta_m < 1$ uma sequência de quantis amostrais únicos de amostras aleatórias de tamanho n de uma população cuja inversa da função de distribuição é $\xi(\theta) = F^{-1}(\theta)$. Se F é contínua e tem densidade de probabilidade, f , positiva em $\xi(\theta_i)$, $i = 1, \dots, m$, então

$$n^{1/2}[\xi_n^*(\theta_1) - \xi(\theta_1), \dots, \xi_n^*(\theta_m) - \xi(\theta_m)] \xrightarrow{\mathcal{L}} N_m(\mathbf{0}, \Omega)$$

onde $\mathbf{0}$ é um vetor de dimensão $m \times 1$ de elementos iguais a zero, e Ω é a matriz de covariância com elementos da forma:

$$\omega_{ij} = \frac{\theta_i(1 - \theta_j)}{f(\xi(\theta_i))f(\xi(\theta_j))}, \quad 1 \leq i \leq j \leq m \quad (2.3.11)$$

Este teorema proporciona o fundamento para a teoria de grandes amostras dos L -estimadores. A mediana é um caso particular importante, tendo variância assintótica igual a $[2f(\xi(1/2))]^{-2}$. Esse valor será menor do que a variância da média para uma grande classe de distribuições com caudas alongadas.

Teorema 2.7 (Koenker e Bassett (1978))

Seja $\{\beta_n^*(\theta_1), \dots, \beta_n^*(\theta_m)\}$ com $0 < \theta_1 < \dots < \theta_m < 1$ uma sequência de soluções únicas das regressões quantílicas do modelo (2.2.2). Seja $\xi(\theta) = F^{-1}(\theta)$, $\Xi(\theta) = (\xi(\theta), 0, \dots, 0)' \in \mathbb{R}^p$ e $\Xi_n^*(\theta) = \beta_n^*(\theta) - \beta$.

Suponha que:

- (i) F é contínua e tem densidade positiva, f , em $\xi(\theta_i)$, $i = 1, \dots, m$, e
- (ii) $x_{i1} = 1$, $i = 1, 2, \dots, n$ e $\lim_{n \rightarrow \infty} n^{-1} \mathbf{X}' \mathbf{X} = \mathbf{Q}$, uma matriz positiva definida.

Então,

$$n^{1/2}[\Xi_n^*(\theta_1) - \Xi(\theta_1), \dots, \Xi_n^*(\theta_m) - \Xi(\theta_m)] \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Omega \otimes \mathbf{Q}^{-1}),$$

onde $\mathbf{0}$ é um vetor de dimensão $mp \times 1$ de elementos iguais a zero, e Ω é a matriz de covariância correspondente aos m quantis amostrais de amostras aleatórias da distribuição F .

Um caso particular importante é a regressão L_1 , $\beta^*(1/2)$. Sem perda de generalidade, β pode ser alocado tal que $F(0) = 1/2$, então $\xi(1/2) = 0$. Assim, a distribuição assintótica de $n^{1/2}(\beta^*(1/2) - \beta)$ é uma Normal p -variada com média zero e matriz de covariância $[2f(0)]^{-2} \mathbf{Q}^{-1}$.

O paralelismo entre o comportamento assintótico dos quantis amostrais no modelo de locação e regressões quantílicas no modelo linear, sugere a extensão da teoria de grandes amostras dos L-estimadores do modelo de locação para o modelo linear. O próximo teorema vai nos mostrar esta idéia mais claramente.

Teorema 2.8 (*Koenker e Bassett (1978)*)

Seja $\Pi(\theta) = (\pi(\theta_1), \dots, \pi(\theta_m))'$ uma medida de probabilidade discreta, simétrica, com massa concentrada em $\{\theta_i : i = 1, \dots, m \mid 0 < \theta_1 < \dots < \theta_m < 1\}$. Suponha que $F(0) = 1/2$, então $\xi(1/2) = F^{-1}(1/2) = 0$, e as condições (i) e (ii) do Teorema 2.7 são satisfeitas.

Então,

$$\tilde{\beta}_n(\pi) = \sum_{i=1}^m \pi(\theta_i) \beta_n^*(\theta_i)$$

é invariante de posição, escala e reparametrização do modelo, de acordo com o Teorema 2.2, e

$$n^{1/2}(\tilde{\beta}_n(\pi) - \beta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Pi' \Omega \Pi Q^{-1}),$$

onde $\mathbf{0}$ é um vetor de dimensão $p \times 1$ de elementos iguais a zero, e Ω é a matriz de covariância correspondente aos m quantis amostrais.

Este teorema será de grande utilidade para o nosso estudo, pois alguns de nossos estimadores, que serão apresentados no próximo capítulo, são da forma de $\tilde{\beta}_n(\pi)$, sendo apenas seus conceitos estendidos para o caso de regressão linear.

Capítulo 3

”L-estimadores de Regressão”

3.1 Introdução

Entre os métodos robustos de estimação existem vários estimadores, tais como os *L-estimadores*, obtidos através de combinações lineares de estatísticas de ordem, os *R-estimadores*, construídos a partir de testes de postos, e os *M-estimadores*, que são obtidos pelo método de Máxima Verossimilhança.

Nosso estudo baseia-se em alguns L-estimadores, que primeiramente foram definidos no caso do modelo de locação, e posteriormente seus conceitos foram ampliados para regressão. A generalização do conceito de L-estimadores para o modelo de regressão dar-se-á da seguinte forma: os estimadores que envolvem o conceito de Média Aparada ampliado para regressão, denominaremos de ”*regressões aparadas*”, cujo conceito e a maneira de apararmos as observações serão abordados com mais detalhe no decorrer deste capítulo; outros L-estimadores como a Trimédia de Tukey e a Média de Gastwirth, quando estendidos para regressão, serão combinações lineares de estimadores de regressões quantílicas.

Neste capítulo apresentaremos alguns ”L-estimadores de regressão” como alternativa aos estimadores de Mínimos Quadrados, onde definiremos cada um deles e apresentaremos resultados assintóticos sobre alguns destes estimadores em estudo.

3.2 Alguns L-estimadores de Regressão

Consideremos o modelo de regressão

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}, \quad (3.2.1)$$

onde $\mathbf{Y}' = (y_1, \dots, y_n)$, \mathbf{X} é uma matriz $n \times p$ de valores conhecidos, $\beta' = (\beta_1, \dots, \beta_p)$ é um vetor de parâmetros desconhecidos, e $\mathbf{Z}' = (Z_1, \dots, Z_n)$ é um vetor de variáveis aleatórias independentes, identicamente distribuídas com função de distribuição F .

Um caso particular do modelo de regressão é o modelo de posição ($p=1$ e $x_{i1} = 1$ para todo i), isto é,

$$y_i = \gamma + Z_i, \quad (3.2.2)$$

onde $i = 1, \dots, n$ e γ é o parâmetro de locação.

Consideremos agora uma amostra aleatória Y_1, \dots, Y_n de uma distribuição contínua com parâmetro de locação β_0 . Se $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ são as estatísticas de ordem da amostra, então:

Definição 3.1 Definimos como uma *L-estimativa* para γ uma combinação linear dessas estatísticas de ordem, isto é,

$$T_n(Y_1, \dots, Y_n) = \sum_{i=1}^n a_i Y_{(i)}$$

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

Entre os L-estimadores, a média aparada é particularmente atrativa, pela sua facilidade de cálculo e eficiência sob várias circunstâncias. Primeiramente, introduziremos o conceito de média aparada no caso unidimensional, para que posteriormente possamos estendê-lo para regressão.

Definição 3.2 Sejam $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ as estatísticas de ordem de uma amostra aleatória de uma distribuição contínua com parâmetro de locação γ . Definimos como *média α -aparada* ou *média α -truncada*

$$A_n = \frac{1}{[n\alpha_2] - [n\alpha_1]} \sum_{i=[n\alpha_1]+1}^{[n\alpha_2]} Y_{(i)}, \quad (3.2.3)$$

onde $[n\alpha]$ é o maior inteiro tal que $[n\alpha] \leq n\alpha$ e $0 < \alpha_1 < \frac{1}{2} < \alpha_2 < 1$.

A idéia do conceito de L-estimadores para a situação de regressão ("L-estimadores de regressão") é baseada nas definições 3.1 e 3.2.

Entre os L-estimadores de regressão, apresentaremos algumas combinações lineares de regressões quantílicas e alguns estimadores baseados em médias aparadas, que denominaremos de "*regressões aparadas*".

Definição 3.3 *Seja $\{\beta^*(\theta_1), \dots, \beta^*(\theta_m)\}$, com $0 < \theta_1 < \dots < \theta_m < 1$, uma sequência de soluções únicas das regressões quantílicas do modelo (3.2.1). Seja $\Pi(\theta) = (\pi(\theta_1), \dots, \pi(\theta_m))'$ uma distribuição de probabilidade, simétrica, tendo como suporte $\{\theta_i : i = 1, \dots, m\}$.*

Qualquer estimador da forma:

$$\hat{\beta}(\pi) = \sum_{i=1}^m \pi(\theta_i) \beta^*(\theta_i) \quad (3.2.4)$$

é um "L-estimador de regressão".

Estendendo o conceito de média aparada para regressão e considerando que as observações são aparadas simetricamente, podemos definir como *regressão α -aparada*, onde $0 < \alpha < 1$, aquela cuja estimativa final de β é calculada após a remoção de $[n\alpha]$ observações, onde estas são correspondentes aos $[n\alpha]$ erros aleatórios que se encontram nas caudas de sua distribuição subjacente. A remoção dessas observações poderá ser a partir das regressões quantílicas de ordens $\alpha/2$ e $(1 - \alpha/2)$ ou diretamente, retirando-se $[n\alpha/2]$ observações correspondentes aos resíduos de cada extremidade.

Com base no conceito de *regressão aparada*, Ruppert e Carroll (1980) propuseram três estimadores, que citaremos a seguir, como alternativa para o estimador de Mínimos Quadrados (MQ). Os outros métodos de estimação foram baseados nestes e em outros L-estimadores, encontrados na literatura.

A fim de simplificar a notação dos estimadores que descreveremos a seguir, teremos que:

- 1 representa o estimador de mínimos quadrados;
- 2 representa o estimador L_1 ;
- 3 significa que os dados são aparados a partir de regressões quantílicas; e
- 4 representa a média aritmética dos estimadores das regressões quantílicas de ordem $\alpha/2$ e $(1 - \alpha/2)$.

3.2.1 Primeiro Método de Estimação ($A_{11}(\alpha)$)

A primeira alternativa requer uma estimativa preliminar ($\hat{\beta}_0$) calculada pelo método de Mínimos Quadrados para o modelo (3.2.1).

Os resíduos da estimativa preliminar ($\hat{\beta}_0$) são:

$$r_i = y_i - x_i \hat{\beta}_0 = Z_i - x_i(\hat{\beta}_0 - \beta),$$

onde x_i representa uma linha da matriz \mathbf{X} e $\hat{\beta}_0$ é um vetor de estimativas preliminares de β de dimensão $p \times 1$.

Sejam $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(n)}$ os resíduos ordenados. Denotaremos por r_{1n} e r_{2n} os $[n(\frac{\alpha}{2})]$ -ésimo e $[n(1 - \frac{\alpha}{2})]$ -ésimo resíduos ordenados, respectivamente. A estimativa final $A_{11}(\alpha)$ é a estimativa de MQ calculada após a remoção das observações que satisfazem:

$$r_i \leq r_{1n} \text{ ou } r_i \geq r_{2n} \quad (3.2.5)$$

Seja $a_i = 0$ ou 1 de acordo com cada i satisfazendo ou não (3.2.5), e seja \mathbf{A} uma matriz diagonal $n \times n$, com $A_{ii} = a_i$.

A matriz \mathbf{A} indica quais observações são aparadas, dando peso 0 para as observações que serão retiradas e peso 1 para aquelas que não serão. Portanto:

$$A_{11}(\alpha) = (\mathbf{X}' \mathbf{A} \mathbf{X})^{-} \mathbf{X}' \mathbf{A} \mathbf{Y} \quad (3.2.6)$$

onde $(\mathbf{X}' \mathbf{A} \mathbf{X})^{-}$ é uma inversa generalizada para $\mathbf{X}' \mathbf{A} \mathbf{X}$.

Discutiremos agora sobre alguns resultados assintóticos relacionados com $A_{11}(\alpha)$.

Primeiramente esclareceremos algumas notações e suposições.

Seja $\mathbf{e}' = (1, 0, \dots, 0)_{(1 \times p)}$, e seja \mathbf{I}_p uma matriz identidade $p \times p$.

Faremos então as seguintes suposições:

- 1- F tem uma densidade contínua f que é positiva no suporte de F .
- 2- Seja $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ a i -ésima linha de \mathbf{X} , $x_{i1} = 1$ para todo $i = 1, \dots, n$

- e $\sum_{i=1}^n x_{ij} = 0$, para $j = 2, \dots, p$.
- 3- $\lim_{n \rightarrow \infty} \left(\max_{j \leq p \text{ e } i \leq n} (n^{-1/2} |x_{ij}|) \right) = 0$.
- 4- Existe \mathbf{Q} positiva definida tal que $\lim_{n \rightarrow \infty} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q}$.
- 5- $(\hat{\beta}_0 - \beta - \eta \mathbf{e}) = O_p(n^{-1/2})$ para alguma constante η .

Devido à suposição 1, resultados assintóticos não serão afetados pela exigência de desigualdades estritas em (3.2.5).

Mostraremos relações da forma:

$$n^{1/2}(A_{11}(\alpha) - \beta) \approx n^{-1/2} \sum_{i=1}^n G(x_i, Z_i) + n^{1/2}H(\hat{\beta}_0 - \beta), \quad (3.2.7)$$

onde G e H são funções dadas.

Mostraremos que em vários casos especiais (incluindo MQ e L_1) o último termo em (3.2.7) pode ser expandido como:

$$n^{1/2}(A_{11}(\alpha) - \beta) \approx n^{-1/2} \sum_{i=1}^n G(x_i, Z_i) + n^{-1/2} \sum_{i=1}^n H^*(x_i, Z_i), \quad (3.2.8)$$

para alguma função H^* . A obtenção da distribuição limite de $n^{1/2}(A_{11}(\alpha) - \beta)$ a partir de (3.2.8) torna-se um problema mais simples.

Seja

$$\phi(x) = \begin{cases} \xi_1/(1-\alpha) & \text{se } x < \xi_1, \\ x/(1-\alpha) & \text{se } \xi_1 \leq x \leq \xi_2, \\ \xi_2/(1-\alpha) & \text{se } \xi_2 < x \end{cases}, \quad (3.2.9)$$

onde ξ_1 e ξ_2 representam os quantis amostrais de ordens $\alpha/2$ e $(1-\alpha)/2$, respectivamente. Definamos agora,

$$a = \xi_2 f(\xi_2) - \xi_1 f(\xi_1) \text{ e } \mathbf{c}_i = (\mathbf{I} - \mathbf{e}\mathbf{e}')\mathbf{x}_i' = (0, x_{i2}, \dots, x_{ip})'.$$

Então a relação (3.2.7) pode ser escrita da seguinte forma:

Teorema 3.1 (*Ruppert e Carroll (1980)*)

Quando $n \rightarrow \infty$,

$$\begin{aligned} n^{1/2}(A_{11}(\alpha) - \beta) &= (1-\alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i Z_i \mathbb{1}(\xi_1 \leq Z_i \leq \xi_2) + \\ &+ (1-\alpha)^{-1} a n^{1/2} (\mathbf{I} - \mathbf{e}\mathbf{e}') (\hat{\beta}_0 - \beta) + n^{-1/2} \sum_{i=1}^n \mathbf{e} \phi(Z_i) + o_p(1), \end{aligned} \quad (3.2.10)$$

onde

$$\mathbb{I}(\xi_1 \leq Z_i \leq \xi_2) = \begin{cases} 1 & \text{se } \xi_1 \leq Z_i \leq \xi_2, \\ 0 & \text{caso contrário} \end{cases}$$

A premultiplicação de um vetor por $(\mathbf{I} - \mathbf{e}\mathbf{e}')$ troca simplesmente a primeira coordenada da matriz \mathbf{I}_p por zero, os primeiros dois termos do lado direito de (3.2.10) representam as estimativas dos parâmetros de inclinação. Verifiquemos a semelhança (e a diferença) entre o primeiro termo e uma representação do estimador de MQ não aparado, $\hat{\beta}$.

Como

$$(\hat{\beta} - \beta) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z},$$

e

$$(n^{-1}\mathbf{X}'\mathbf{X}) \longrightarrow \mathbf{Q},$$

segue-se que

$$n^{1/2}(\hat{\beta} - \beta) = n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1}\mathbf{x}_i'Z_i + o_p(1). \quad (3.2.11)$$

O segundo termo em (3.2.10) indica a contribuição da estimativa preliminar para a estimativa de MQ aparada; esta condição é somente para as estimativas dos parâmetros de inclinação da reta. Como somente a primeira coordenada de \mathbf{e} é diferente de zero, o terceiro termo do lado direito de (3.2.10) é uma representação da estimativa do intercepto e é idêntica a representação da média aparada no modelo de locação.

Para apresentar a relação (3.2.8) de outra forma, faremos a seguinte suposição adicional:

6- Para alguma função g ,

$$n^{1/2}(\hat{\beta}_0 - \beta) = n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1}\mathbf{x}_i'g(Z_i) + o_p(1). \quad (3.2.12)$$

Como indicado, a suposição 6 é válida com $g(x) = x$ se $\hat{\beta}_0$ é a estimativa de MQ. Pelo Teorema 2.7, a suposição é válida com $g(x) = (f(0))^{-1}(\frac{1}{2} - \mathbb{I}(x < 0))$ se $\hat{\beta}_0$ é a estimativa L_1 .

Como consequência imediata do Teorema 3.1, temos:

Teorema 3.2

$$\begin{aligned}
 n^{1/2}(A_{11}(\alpha) - \beta) &= (1 - \alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i \{Z_i \mathbf{1}(\xi_1 \leq Z_i \leq \xi_2) + ag(Z_i)\} + \\
 &+ n^{-1/2} \sum_{i=1}^n \mathbf{e} \phi(Z_i) + o_p(1)
 \end{aligned} \tag{3.2.13}$$

PROVA:

Substituindo (3.2.12) em (3.2.10) temos:

$$\begin{aligned}
 n^{\frac{1}{2}}(A_{11}(\alpha) - \beta) &= (1 - \alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i Z_i \mathbf{1}(\xi_1 \leq Z_i \leq \xi_2) + \\
 &+ (1 - \alpha)^{-1} a(\mathbf{I} - \mathbf{e} \mathbf{e}') n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{x}_i g(Z_i) + \\
 &+ n^{-1/2} \sum_{i=1}^n \mathbf{e} \phi(Z_i) + o_p(1)
 \end{aligned}$$

Mas sabemos que $\mathbf{c}_i = (\mathbf{I} - \mathbf{e} \mathbf{e}') \mathbf{x}_i$.

Assim,

$$\begin{aligned}
 n^{1/2}(A_{11}(\alpha) - \beta) &= (1 - \alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i Z_i \mathbf{1}(\xi_1 \leq Z_i \leq \xi_2) + \\
 &+ (1 - \alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i ag(Z_i) + n^{-1/2} \sum_{i=1}^n \mathbf{e} \phi(Z_i) + o_p(1) \\
 &= (1 - \alpha)^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{c}_i \{Z_i \mathbf{1}(\xi_1 \leq Z_i \leq \xi_2) + ag(Z_i)\} + \\
 &+ n^{-1/2} \sum_{i=1}^n \mathbf{e} \phi(Z_i) + o_p(1) \quad \square
 \end{aligned}$$

Como um caso especial do Teorema 3.2, temos:

Corolário 3.1 *No caso do modelo de locação ($p = 1$ e $x_{i1} = 1$ para todo i)*

$$n^{1/2}(A_{11}(\alpha) - \beta) = n^{-1/2} \sum_{i=1}^n \phi(Z_i) + o_p(1)$$

Mostraremos agora, que baseado no Teorema 3.2 podemos ter as seguintes conclusões básicas com relação a $A_{11}(\alpha)$:

- O estimador do intercepto é assintoticamente não viciado se F é simétrica;
- Os estimadores dos parâmetros de inclinação são assintoticamente não viciados mesmo quando F é assimétrica;
- A variância assintótica do estimador do intercepto, o qual não depende da escolha de $\hat{\beta}_0$, é a mesma da média aparada no modelo de locação; e
- A matriz de covariância dos estimadores de inclinação dependem de $\hat{\beta}_0$ e em geral são difíceis de estimar.

Seja \mathbf{O} um vetor de zeros de dimensão $(p - 1) \times 1$.

Pela suposição 2, existe uma matriz $\hat{\mathbf{Q}}$ tal que

$$\mathbf{Q} = \begin{bmatrix} 1 & \mathbf{O}' \\ \mathbf{O} & \hat{\mathbf{Q}} \end{bmatrix}$$

e

$$\mathbf{Q}^{-1} = \begin{bmatrix} 1 & \mathbf{O}' \\ \mathbf{O} & \hat{\mathbf{Q}}^{-1} \end{bmatrix}$$

E mais ainda,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{c}_i \mathbf{c}_i' = \begin{bmatrix} 1 & \mathbf{O}' \\ \mathbf{O} & \hat{\mathbf{Q}} \end{bmatrix}$$

e

$$\mathbf{Q} \sum_{i=1}^n \mathbf{c}_i = \begin{bmatrix} \mathbf{O} \\ \mathbf{O} \end{bmatrix} \quad (3.2.14)$$

Se a estimativa de β é $A_{11}(\alpha)$, então o vício assintótico do intercepto é dado por:

$$E\phi(Z_1) = (1 - \alpha)^{-1} \int_{\xi_1}^{\xi_2} x dF(x), \quad (3.2.15)$$

que será zero se F é simétrica em torno de zero.

De (3.2.13) e (3.2.14), segue-se que os estimadores dos parâmetros de inclinação são assintoticamente não viciados, mesmo quando F é assimétrica. A variância assintótica do intercepto normalizado é

$$\begin{aligned} \sigma^2(\alpha, F) = Var(\phi(Z_1)) &= (1 - \alpha)^{-2} \left(\int_{\xi_1}^{\xi_2} (x - \delta(\alpha))^2 dF(x) + \frac{\alpha}{2} k_1^2 + \right. \\ &\quad \left. + (1 - \frac{\alpha}{2}) k_2^2 - ((1 - \frac{\alpha}{2}) k_2 + \frac{\alpha}{2} k_1)^2 \right), \quad (3.2.16) \end{aligned}$$

que é a variância assintótica da média α -aparada normalizada no modelo de locação, onde

$$\begin{aligned}\delta(\alpha) &= (1 - \alpha)^{-1} \int_{\xi_1}^{\xi_2} x dF(x); \\ \text{e} \\ k_j &= (\xi_j - \delta(\alpha)).\end{aligned}$$

O estimador do intercepto é assintoticamente não correlacionado com o restante dos estimadores, e a matriz de covariância dos parâmetros de inclinação é $\hat{Q}^{-1}\sigma^2(\alpha, g, F)$, onde

$$\sigma^2(\alpha, g, F) = (1 - \alpha)^{-2} \text{Var}(Z_1 \mathbb{1}(\xi_1 \leq Z_1 \leq \xi_2) + ag(Z_1)) \quad (3.2.17)$$

Quando $(\hat{\beta}_0 - \beta) = O_p(n^{-1/2})$ podemos ver que a distribuição assintótica do estimador do intercepto não depende de $\hat{\beta}_0$. Por outro lado, podemos perceber de (3.2.13) que os estimadores dos parâmetros de inclinação dependem de $\hat{\beta}_0$, mesmo na situação em que $a = 0$ é excluída pela suposição 1.

Utilizando o Teorema Central do Limite de Lindenberg e o Teorema 3.2, podemos mostrar que, sob as suposições 3 e 4,

$$n^{-1/2}(A_{11}(\alpha) - \beta - \mathbf{e}E\phi(Z_1)) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Sigma),$$

onde

$$\Sigma = \begin{bmatrix} \sigma^2(\alpha, F) & \mathbf{0}' \\ \mathbf{0} & \hat{Q}^{-1}\sigma^2(\alpha, g, F) \end{bmatrix}$$

e $\mathbf{0}$ é uma vetor de zeros de dimensão $(p - 1) \times 1$.

3.2.2 Segundo Método de Estimação ($A_{21}(\alpha)$)

O método para obtenção deste estimador é semelhante ao descrito anteriormente, sendo que a estimativa preliminar ($\hat{\beta}_0$), neste caso, é calculada a partir do estimador da regressão quantílica de ordem 1/2, isto é, a minimização da soma dos desvios absolutos (L_1).

A estimativa final, $A_{21}(\alpha)$, também é obtida por Mínimos Quadrados após a remoção das observações que satisfazem (3.2.5).

A teoria assintótica descrita para $A_{11}(\alpha)$ também é válida para $A_{21}(\alpha)$.

3.2.3 Terceiro Método de Estimação ($A_{31}(\alpha)$)

Baseados na Teoria de Regressão Quantílica, comentada no capítulo 2, em que Koenker e Bassett (1978) estenderam o conceito de *quantil* para modelos lineares, Ruppert e Carroll (1980) sugeriram aparar as observações a partir de estimadores de regressões quantílicas.

Nosso interesse é aparar $[n\alpha]$ observações. Com este fim, primeiramente calcularemos a $\frac{\alpha}{2}$ -ésima e a $(1 - \frac{\alpha}{2})$ -ésima regressões quantílicas, que serão representadas por $(\beta^*(\frac{\alpha}{2}))$ e $(\beta^*(1 - \frac{\alpha}{2}))$, respectivamente. O próximo passo será remover as observações cujos resíduos, calculados a partir de $\beta^*(\frac{\alpha}{2})$, são negativos ou cujos resíduos, calculados a partir de $\beta^*(1 - \frac{\alpha}{2})$ são positivos. Isto é, diremos que a observação i será removida se satisfaz:

$$y_i - \mathbf{x}_i \beta^*\left(\frac{\alpha}{2}\right) \leq 0 \text{ ou } y_i - \mathbf{x}_i \beta^*\left(1 - \frac{\alpha}{2}\right) \geq 0 \quad (3.2.18)$$

Seja $b_i = 0$ ou 1 de acordo com i satisfazendo ou não (3.2.18), e seja \mathbf{B} uma matriz diagonal $n \times n$ com $B_{ii} = b_i$.

Assim,

$$A_{31}(\alpha) = (\mathbf{X}'\mathbf{B}\mathbf{X})^{-}(\mathbf{X}'\mathbf{B}\mathbf{Y}), \quad (3.2.19)$$

onde $(\mathbf{X}'\mathbf{B}\mathbf{X})^{-}$ é uma inversa generalizada da matriz $(\mathbf{X}'\mathbf{B}\mathbf{X})$.

Falaremos agora sobre alguns resultados assintóticos a respeito de $A_{31}(\alpha)$.

Sabemos que podem existir várias soluções para o problema de minimização (2.2.3); no entanto, os resultados assintóticos que apresentaremos não dependem da regra utilizada para a escolha desta solução. Assim, suponhamos que uma regra específica foi utilizada e denotemos esta solução por $\beta^*(\theta)$.

Como primeiro resultado com respeito a $A_{31}(\alpha)$ temos:

Teorema 3.3 (*Ruppert e Carroll (1980)*)

O estimador $A_{31}(\alpha)$ satisfaz:

$$n^{1/2}(A_{31}(\alpha) - \beta) = \mathbf{Q}^{-1}n^{-1/2}\left\{\sum_{i=1}^n \mathbf{x}_i(\phi(Z_i) - E\phi(Z_i)) + \delta(\alpha)\right\} + o_p(1) \quad (3.2.20)$$

e portanto

$$n^{1/2}(A_{31}(\alpha) - \beta - \delta(\alpha)) \xrightarrow{\mathcal{L}} N_p(\mathbf{O}, \sigma^2(\alpha, F)\mathbf{Q}^{-1}), \quad (3.2.21)$$

onde

$$\begin{aligned} \delta(\alpha) &= (1 - \alpha)^{-1} \int_{\xi_1}^{\xi_2} x dF(x); \\ \sigma^2(\alpha, F) &= (1 - \alpha)^{-2} \left(\int_{\xi_1}^{\xi_2} (x - \delta(\alpha))^2 dF(x) + \frac{\alpha}{2} k_1^2 + \right. \\ &\quad \left. + (1 - \frac{\alpha}{2}) k_2^2 - ((1 - \frac{\alpha}{2}) k_2 + \frac{\alpha}{2} k_1)^2 \right) \\ e \\ k_j &= (\xi_j - \delta(\alpha)). \end{aligned}$$

Notemos que a expressão (3.2.21) satisfaz as hipóteses e teoremas descritos por Koenker e Bassett (1978) com respeito a $A_{31}(\alpha)$. Notemos também que o vício de $A_{31}(\alpha)$ para β envolve somente o intercepto e não os parâmetros de inclinação. $A_{31}(\alpha)$ também é assintoticamente não viciado se F é simétrica.

3.2.4 Quarto Método de Estimação ($A_{41}(\alpha)$)

Vimos que o comportamento de $A_{11}(\alpha)$ e $A_{21}(\alpha)$ dependem de sua estimativa preliminar ($\hat{\beta}_0$), que em geral não tem o mesmo comportamento que a regressão aparada ($A_{31}(\alpha)$). Na tentativa de escolher $\hat{\beta}_0$ de forma a encontrar uma estimativa final para β que tenha a mesma distribuição assintótica de $A_{31}(\alpha)$, Ruppert e Carroll (1980) sugeriram como estimativa preliminar a média aritmética entre a $[\frac{\alpha}{2}]$ -ésima e a $[1 - \frac{\alpha}{2}]$ -ésima regressões quantílicas, isto é,

$$\hat{\beta}_0 = \frac{1}{2}[\beta^*(\frac{\alpha}{2}) + \beta^*(1 - \frac{\alpha}{2})] \quad (3.2.22)$$

A partir de $\hat{\beta}_0$, calculamos os resíduos e removemos as observações correspondentes aos $[n\frac{\alpha}{2}]$ maiores e $[n\frac{\alpha}{2}]$ menores resíduos, isto é, aquelas observações que satisfazem (3.2.5).

Como em $A_{11}(\alpha)$ e em $A_{21}(\alpha)$, definiremos uma matriz \mathbf{A} que funcionará como um operador linear, ou seja, esta matriz indica quais observações são aparadas, dando peso 0 para aquelas que serão retiradas e peso 1 para as que

não serão. Logo, $A_{41}(\alpha)$ será a estimativa de Mínimos Quadrados após a remoção das observações que correspondem aos $[n\frac{\alpha}{2}]$ maiores e menores resíduos, isto é,

$$A_{41}(\alpha) = (\mathbf{X}'\mathbf{A}\mathbf{X})^{-}\mathbf{X}'\mathbf{A}\mathbf{Y}, \quad (3.2.23)$$

onde $(\mathbf{X}'\mathbf{A}\mathbf{X})^{-}$ é uma inversa generalizada para $\mathbf{X}'\mathbf{A}\mathbf{X}$.

Vale ressaltar que, para que $A_{41}(\alpha)$ tenha a mesma distribuição assintótica de $A_{31}(\alpha)$ é necessário que F seja simétrica e que os dados sejam aparados simetricamente.

Portanto, se F é simétrica teremos:

$$\begin{aligned} \xi_1 &= -\xi_2, \\ \text{e} \\ f(\xi_1) &= f(\xi_2). \end{aligned}$$

Pelo Teorema 2.7 podemos ver que $\hat{\beta}_0$ satisfaz a suposição 6 com

$$g(x) = (2f(\xi_1))^{-1}\Psi_{\frac{\alpha}{2}}(x - \xi_1) + (2f(\xi_2))^{-1}\Psi_{(1-\frac{\alpha}{2})}(x - \xi_2), \quad (3.2.24)$$

onde

$$\Psi_{\theta}(u) = \begin{cases} \theta & u \geq 0 \\ (\theta - 1) & \text{caso contrário} \end{cases}$$

Logo,

$$ag(x) = \xi_1 \mathbb{I}(x < \xi_1) + \xi_2 \mathbb{I}(x \geq \xi_2) \quad (3.2.25)$$

De (3.2.13) e (3.2.25) temos:

$$n^{1/2}(A_{41}(\alpha) - \beta) = n^{-1/2} \sum_{i=1}^n \mathbf{Q}^{-1} \mathbf{x}_i \phi(Z_i) + o_p, \quad (3.2.26)$$

e sendo F simétrica $\delta(\alpha) = 0$, então (3.2.20) implicará que

$$n^{1/2}(A_{31}(\alpha) - A_{41}(\alpha)) \xrightarrow{p} 0 \quad (3.2.27)$$

Podemos então dizer que, assintoticamente não há diferença entre $A_{41}(\alpha)$ e $A_{31}(\alpha)$. Mas é bom ressaltar que (3.2.27) não necessariamente ocorre se F é assimétrica.

Agora, de (3.2.20) e (3.2.27) obtemos que:

$$n^{1/2}(A_{41}(\alpha) - \beta) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\alpha, F) \mathbf{Q}^{-1}). \quad (3.2.28)$$

É bom lembrarmos, que a escolha de $\hat{\beta}_0$ deve ser baseada na eficiência do resultado da estimativa final ($A_{11}(\alpha)$, $A_{21}(\alpha)$, ou $A_{41}(\alpha)$) e não na sua semelhança com $A_{31}(\alpha)$.

3.2.5 Quinto Método de Estimação ($A_{12}(\alpha)$)

Ruppert e Carroll (1980) sugerem que a estimativa final de β seja obtida por Mínimos Quadrados. No entanto, sabemos que quando as caudas da distribuição dos erros são muito longas ou muito pesadas, mesmo depois de apararmos algumas observações, há a possibilidade de que ainda existam alguns outliers no conjunto de dados em estudo. Diante desta situação, sugerimos como estimativa final a regressão L_1 .

O processo utilizado para a obtenção de $A_{12}(\alpha)$ é semelhante ao utilizado em $A_{11}(\alpha)$, sendo que, após a remoção das $[n\alpha]$ observações, obteremos $A_{12}(\alpha)$ minimizando a soma dos desvios absolutos das observações restantes.

3.2.6 Sexto Método de Estimação ($A_{22}(\alpha)$)

O procedimento utilizado neste método é análogo a $A_{21}(\alpha)$, sendo que a estimativa final de β é obtida por L_1 .

3.2.7 Sétimo Método de Estimação ($A_{32}(\alpha)$)

Neste caso aparamos as observações a partir de regressões quantílicas, como em $A_{31}(\alpha)$. Após a remoção das observações cujos resíduos calculados a partir da $(\frac{\alpha}{2})$ -ésima regressão quantílica são negativos e cujos resíduos calculados a partir da $(1 - \frac{\alpha}{2})$ -ésima são positivos, minimizaremos a soma dos desvios absolutos das observações restantes para obtermos $A_{32}(\alpha)$.

Os resultados assintóticos sobre $A_{12}(\alpha)$, $A_{22}(\alpha)$ e $A_{23}(\alpha)$ ainda não foram estudados.

3.2.8 Oitavo Método de Estimação ($A_{42}(\alpha)$)

Este estimador é análogo a $A_{41}(\alpha)$. A estimativa preliminar, neste caso, será:

$$\hat{\beta}_0 = \frac{1}{2}[\beta^*(\frac{\alpha}{2}) + \beta^*(1 - \frac{\alpha}{2})],$$

onde $\beta^*(\frac{\alpha}{2})$ e $\beta^*(1 - \frac{\alpha}{2})$ são a $\frac{\alpha}{2}$ -ésima e a $(1 - \frac{\alpha}{2})$ -ésima regressões quantílicas, respectivamente.

3.2.9 Nono Método de Estimação ($A_4(\alpha)$)

Sugerimos como estimativa de β simplesmente a estimativa preliminar da seção 3.2.8, isto é,

$$A_4(\alpha) = \frac{1}{2}[\beta^*(\frac{\alpha}{2}) + \beta^*(1 - \frac{\alpha}{2})] \quad (3.2.29)$$

De acordo com o Teorema 2.8 temos que:

$$n^{1/2}(A_4(\alpha) - \beta) \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \mathbf{\Pi}' \mathbf{\Omega} \mathbf{\Pi} \otimes \mathbf{Q}^{-1}),$$

onde $\mathbf{\Pi} = (1/2, 1/2)'$ e $\mathbf{\Omega}$ é a matriz de covariância correspondente aos dois quantis amostrais ($\theta_1 = \alpha/2, \theta_2 = 1 - \alpha/2$).

3.2.10 Trimédia de Tukey (T)

Este estimador foi proposto por Tukey para o caso do modelo de locação unidimensional, onde é definido como a média ponderada do primeiro, segundo e terceiro quartis com pesos, respectivamente, $\frac{1}{4}$, $\frac{1}{2}$ e $\frac{1}{4}$, isto é,

$$T = 0.25Q(1/4) + 0.5Q(1/2) + 0.25Q(3/4) \quad (3.2.30)$$

Estendendo para regressão, a *Trimédia de Tukey* será uma combinação linear de regressões quantílicas, ou seja,

$$T = 0.25\beta^*(1/4) + 0.5\beta^*(1/2) + 0.25\beta^*(3/4) \quad (3.2.31)$$

Baseado no Teorema 2.8, temos que:

$$n^{1/2}(T - \beta) \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \mathbf{\Pi}' \mathbf{\Omega} \mathbf{\Pi} \otimes \mathbf{Q}^{-1}),$$

onde $\mathbf{\Pi} = (1/4, 1/2, 1/4)'$ e $\mathbf{\Omega}$ é a matriz de covariância correspondente aos 3 quantis amostrais ($\theta_1 = 1/4, \theta_2 = 1/2, \theta_3 = 3/4$).

3.2.11 Média de Gastwirth (G)

A idéia de utilizar este estimador foi proposta por Gastwirth (1965), tentando construir estimadores correspondentes a testes de postos. Ele propôs como estimador do parâmetro de locação no modelo de locação unidimensional, a média ponderada dos quantis $\frac{1}{3}$, $\frac{1}{2}$ e $\frac{2}{3}$ com pesos, respectivamente, 0.3, 0.4 e 0.3.

Este estimador foi obtido a partir de um grupo de testes baseados em postos que aproximam os testes ótimos para a distribuição de Cauchy e Normal. Neste trabalho, verificou-se a eficiência assintótica deste estimador, relativa ao melhor estimador para cada distribuição considerada. Observou-se que a Média de Gastwirth tinha uma eficiência relativa superior a 80% para as distribuições Cauchy, dupla exponencial e logística. Este fato nos dá evidências de que este estimador parece ser bem eficiente para distribuições com caudas mais pesadas que a da Normal.

Poderíamos, então, dizer que este estimador é robusto, pois ele não é muito afetado por outliers.

Como estamos buscando estimadores robustos para o modelo de regressão, propusemos estendê-lo para o caso de regressão. Logo, a *Média de Gastwirth* para o nosso caso será:

$$G = 0.3\beta^*\left(\frac{1}{3}\right) + 0.4\beta^*\left(\frac{1}{2}\right) + 0.3\beta^*\left(\frac{2}{3}\right) \quad (3.2.32)$$

onde $\beta^*(\theta)$ representa o estimador proporcionado pela θ -ésima regressão quantílica.

Assim como em $A_4(\alpha)$ e na Trimédia de Tukey, também podemos encontrar a distribuição assintótica de G, baseado no Teorema 2.8. Portanto,

$$n^{1/2}(G - \beta) \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \mathbf{\Pi}'\mathbf{\Omega}\mathbf{\Pi} \otimes \mathbf{Q}^{-1}),$$

onde $\mathbf{\Pi} = (0.3, 0.4, 0.3)'$ e $\mathbf{\Omega}$ é a matriz de covariância correspondente aos três quantis amostrais ($\theta_1 = 1/3, \theta_2 = 1/2, \theta_3 = 2/3$).

Capítulo 4

Estudo de Monte Carlo

4.1 Introdução

Através de um estudo de Monte Carlo avaliamos a eficiência de alguns estimadores, os quais não dispomos de resultados analíticos para suas variâncias, para pequenos tamanhos de amostra.

O comportamento dos estimadores em estudo foi avaliado de acordo com a distribuição dos erros e os tamanhos de amostra. Foram consideradas várias distribuições para os erros com caudas semelhantes ou mais pesadas que a da Normal. A fim de verificarmos o comportamento dos estimadores para pequenas e médias amostras, vários tamanhos de amostra foram considerados.

4.2 Metodologia

O estudo foi feito para o caso de Regressão simples. O objetivo deste estudo é comparar a eficiência relativa dos estimadores, definidos no capítulo 3, avaliados em termos do seu *Erro Quadrático Médio (EQM)* para as diversas distribuições e cada um dos seguintes tamanhos de amostra: 20, 40 e 100.

As principais características da simulação foram:

4.2.1 Métodos de estimação:

- Regressões aparadas, com $\alpha = 10\%$ e 20% de dados aparados:

$$A_{i,j}(\alpha), \quad i = 1, 2, 3, 4, \quad j = 1, 2 \text{ e } \alpha = 0.10 \text{ e } 0.20,$$

onde i representa o método preliminar para aparar as observações, que pode ser:

1. Mínimos Quadrados;
2. L_1 ;
3. Aparando as observações cujos resíduos, calculados a partir da $[\frac{\alpha}{2}]$ -ésima regressão quantílica, são negativos e cujos resíduos, calculados a partir da $[(1 - \frac{\alpha}{2})]$ -ésima regressão quantílica, são positivos; ou
4. $\hat{\beta}_0 = \frac{1}{2}[\beta^*(\frac{\alpha}{2}) + \beta^*(1 - \frac{\alpha}{2})]$, onde $\hat{\beta}(\alpha)$ representa o estimador proporcionado pela α -ésima regressão quantílica.

j representa a estimativa final de β , que pode ser:

1. Mínimos Quadrados; ou
2. L_1 .

- Combinações lineares de regressões quantílicas:

- $A_4 = \frac{1}{2}[\beta^*(\frac{\alpha}{2}) + \beta^*(1 - \frac{\alpha}{2})]$;
- Trimédia de Tukey
$$T = 0.25\beta^*(\frac{1}{4}) + 0.5\beta^*(\frac{1}{2}) + 0.25\beta^*(\frac{3}{4})$$
- Média de Gastwirth
$$G = 0.3\beta^*(\frac{1}{3}) + 0.4\beta^*(\frac{1}{2}) + 0.3\beta^*(\frac{2}{3})$$

4.2.2 Distribuições:

As distribuições e os respectivos valores de suas curtoses estão dispostos na tabela a seguir:

Tabela 4.1: Distribuições dos erros e os respectivos valores de suas curtoses

Distribuição	Curtose
Normal Padrão	3
t-Student(10)=Tukey ($\lambda_2 = 0.0234$ e $\lambda_3 = 0.0148$)	4
t-Student(6)=Tukey ($\lambda_2 = -0.1377$ e $\lambda_3 = -0.0802$)	6
Exponencial Dupla	6
t-Student(5)=Tukey ($\lambda_2 = -0.2480$ e $\lambda_3 = -0.1358$)	9
$0.15N(0, 25) + 0.85N(0, 1)$	13
$0.10N(0, 25) + 0.90N(0, 1)$	16
$0.15N(0, 100) + 0.85N(0, 1)$	18
$0.10N(0, 100) + 0.90N(0, 1)$	25
Cauchy Padrão	-

Todas as distribuições consideradas são simétricas em torno de zero.

4.2.3 Geração das distribuições:

Todos os programas foram executados num computador IBM/3090 e a linguagem utilizada foi o IML/SAS (*Linguagem matricial do sistema de programas estatísticos SAS*). Os valores para essas distribuições foram gerados de acordo com os seguintes métodos:

UNIFORME:

Para a geração da distribuição Uniforme entre 0 e 1 foi utilizado o método congruencial multiplicativo de módulo primo abordado por Fishman e Moore (1982), que consiste no seguinte:

Seja A um multiplicador e

$$Z_i = AZ_{i-1}(\text{mod}2^{31} - 1),$$

onde $\text{mod}(u)$ é o resto da divisão de Z_{i-1} por u .

Se $\{U_i = Z_i/M; i = 0, 1, \dots\}$ é uma sequência de números aleatórios, diremos que esta sequência será distribuída uniformemente no intervalo $(0, 1)$.

Este método é utilizado pela função CALL RANUNI(*semente, nome da variável*) do SAS, tendo $A = 397204094$ e $M = 2^{31} - 1$.

NORMAL PADRÃO:

Foi utilizado o método polar de Marsaglia, que consiste no seguinte: Dadas duas variáveis aleatórias W_1 e W_2 com distribuição Uniforme entre -1 e 1, se $W_1^2 + W_2^2 \leq 1$, então $\arctan(W_2/W_1)$ e $W_1^2 + W_2^2$ são variáveis aleatórias independentes com distribuição $U(0, 2\pi)$ e $U(0,1)$, respectivamente.

Assim,

$$\begin{aligned} X_1 &= \{-2 \ln(W_1^2 + W_2^2)\}^{1/2} W_1 (W_1^2 + W_2^2)^{-1/2}, \text{ e} \\ X_2 &= \{-2 \ln(W_1^2 + W_2^2)\}^{1/2} W_2 (W_1^2 + W_2^2)^{-1/2}, \end{aligned}$$

são duas variáveis aleatórias normais independentes, com média 0 e variância 1.

Este método é utilizado pela função CALL RANNOR(*semente, nome da variável*) do SAS, onde desta forma é possível controlar a semente.

DISTRIBUIÇÕES DA FAMÍLIA TUKEY:

Se U_1 é uma variável aleatória Uniforme entre 0 e 1, então

$$E = (U_1^{\lambda_3} - (1 - U_1)^{\lambda_3}) / \lambda_2$$

tem distribuição de Tukey com parâmetros λ_2 e λ_3 .

EXPONENCIAL DUPLA:

Se U_1 e U_2 são variáveis aleatórias com distribuição Uniforme entre 0 e 1, e

$$V = -\log(U_1),$$

então

$$E = \text{sign}(U_2 - 1/2)V$$

tem distribuição Exponencial Dupla.

NORMAIS CONTAMINADAS:

Seja σ o desvio padrão de uma variável aleatória com distribuição $N(0, \sigma^2)$ e $\alpha 100$ a porcentagem de contaminação.

Se Z é uma variável aleatória com distribuição Normal Padrão e U_1 é uma variável aleatória Uniforme entre 0 e 1, então

$$E = \begin{cases} Z & \text{se } U_1 > \alpha \\ \sigma Z & \text{caso contrário} \end{cases}$$

tem distribuição Normal Contaminada da forma $\alpha N(0, \sigma^2) + (1 - \alpha)N(0, 1)$

CAUCHY PADRÃO:

Foi utilizado o método de rejeição, que consiste no seguinte:

Se W_1 e W_2 são variáveis aleatórias com distribuição $U(-\frac{1}{2}, \frac{1}{2})$, e se $W_1^2 + W_2^2 \leq \frac{1}{4}$ então W_1/W_2 tem distribuição Cauchy-Padrão.

Este procedimento é utilizado pela função CALL RANCAU(*semente*, *nome da variável*) do SAS, onde desta forma a semente pode ser controlada.

4.2.4 Número de repetições de Monte Carlo:

Primeiramente fizemos 1500, 2000 e 2500 repetições para amostras de tamanho 20 e 40, e 1000, 1500 e 2000 repetições para amostras de tamanho 100. Após este teste, verificamos que 2000 repetições para amostras de tamanho 20 e 40, e 1500 para amostras de tamanho 100 seria suficiente para a estabilização do método de Monte Carlo.

4.2.5 Método de Simulação:

As observações y_1, \dots, y_n , onde

$$y_i = \beta_0 + \beta_1 x_i + Z_i,$$

foram geradas como segue:

Para cada tamanho de amostra considerado no estudo foram fixados os valores x_1, \dots, x_n , gerados de acordo com uma distribuição Normal-Padrão.

Como valores reais dos parâmetros demos $\beta_0 = 1$ e $\beta_1 = 1$, sem perda de generalidade.

Os erros, para cada combinação de tamanho de amostra e distribuição, foram gerados de acordo com 4.2.4. β_0 e β_1 foram estimados conforme os métodos de estimação citados no capítulo 3, para cada combinação considerada. Temos, então, N estimativas para β_0 e β_1 , para cada método em estudo, onde podemos fazer uma comparação entre os estimadores a partir de suas respectivas estimativas de variância, média, desvio-padrão, etc.

A comparação entre os estimadores foi feita a partir dos seus respectivos *Erros Quadráticos Médios* (EQM), considerando o desvio padrão de cada *Erro Quadrático Médio*.

Para obtenção das estimativas de EQM e dos seus desvios padrão, fizemos o seguinte:

Seja $T_n = T_n(x_1, \dots, x_n)$ o estimador de θ para uma amostra de tamanho n e $t_n(j)$ a estimativa de β na replicação j .

Definimos a estimativa do *Erro Quadrático Médio* por:

$$E\hat{Q}M = \frac{1}{N} \sum_{j=1}^N (t_n(j) - \beta)^2,$$

isto é, estamos querendo verificar o quanto, em média, a estimativa t_n está distante do real valor do parâmetro β .

Observemos que em alguns casos, em que a estimativa final de β é L_1 , a solução do problema de minimização dos desvios absolutos não era única. Assim, nessas situações, decidimos não incluir essas estimativas no cálculo de $E\hat{Q}M$. Este problema será mais discutido posteriormente na apresentação dos resultados das simulações.

A estimativa da variância do $E\hat{Q}M$ é dada por:

$$S^2(E\hat{Q}M) = \frac{1}{N} \frac{1}{(N-1)} \sum_{j=1}^N ((t_n(j) - \beta)^2 - E\hat{Q}M)^2$$

4.3 Sementes:

Para a geração das dez distribuições foi necessário a definição de algumas variáveis aleatórias. As variáveis e suas sementes iniciais para o processo da geração dos números pseudo aleatórios estão na seguinte tabela:

Tabela 4.2: Sementes iniciais para a geração dos valores das variáveis aleatórias

VARIÁVEL	SEMENTE
U_1	346789391
U_2	787534995
Z	346789391
X	56879325
Y	346789391

Esclarecemos que U_1 e U_2 tem distribuição Uniforme entre 0 e 1; Z tem distribuição Normal Padrão; Y tem distribuição Cauchy-Padrão; e X tem distribuição Normal Padrão.

Utilizamos para cada tamanho de amostra as mesmas sementes iniciais para a geração das variáveis em estudo. Vale ressaltar que as sementes iniciais para as variáveis aleatórias uniformes foram as mesmas para todas as distribuições. Os valores de X são os mesmos para todas as repetições, pois X representa uma variável independente no nosso estudo de regressão. Os valores das outras variáveis variam a cada simulação, pois o valor da semente inicial para a geração das outras variáveis é acrescentado de uma constante a cada repetição do processo. Para cada distribuição dos erros gerada e cada método de estimação utilizado, obtemos o resultado de uma simulação. Para melhor compreensão do leitor a respeito dos programas verifique o Apêndice C, onde estão dispostos os programas utilizados nesse estudo.

4.4 Resultados:

Foi necessário aumentar a escala dos gráficos referentes a amostras de tamanho 40 e 100 a fim de facilitar a visualização do leitor. Observemos que em todas as tabelas e gráficos, as distribuições dos erros estão em ordem crescente de curtose.

Analisando as tabelas e gráficos, percebemos que para os três tamanhos de amostra considerados, ocorrem fatos semelhantes com os estimadores em estudo.

Quando a porcentagem de dados aparados é de 10% ($\alpha = 0.10$), observamos que:

Diante da distribuição Normal e distribuições com caudas moderadamente mais longas que a Normal, como a $t(10)$, $t(6)$ e $t(5)$, os estimadores de Mínimos Quadrados Aparados (MQA), a Trimédia de Tukey e a Média de Gastwirth apresentam as melhores performances. Os estimadores L_1 Aparados (L_1A), isto é, as *regressões aparadas* cujas estimativas finais são L_1 , não apresentam uma performance muito boa diante de distribuições com caudas moderadamente alongadas, mas o estimador $A_4(\alpha)$ é o que apresenta o pior comportamento nesses casos.

Uma classe de distribuições interessante para o nosso estudo é aquela em que suas caudas apresentam uma medida de pico bem elevada, como é o caso da Exponencial Dupla e das Normais contaminadas em que a porcentagem de contaminação é de 15%. Notemos que, apesar das curtoses da distribuição $t(6)$ e Exponencial Dupla serem iguais, a eficiência de alguns estimadores cai bastante nessas situações. Este fato ocorre porque a Exponencial Dupla tem um grande pico e a curtose não é uma boa medida de pico. A curtose mede principalmente o peso das caudas da distribuição. Os estimadores mais sensíveis a essas mudanças nas medidas de pico são os MQA e $A_4(\alpha)$, que sofrem uma queda nas suas eficiências nessas situações. Percebemos claramente em todos os gráficos e tabelas, uma instabilidade no comportamento desses estimadores, a medida em que alteramos a forma da distribuição. Já os estimadores L_1A não são muito afetados por essas mudanças nas distribuições. Os estimadores L_1A , Trimédia de Tukey e Média de Gastwirth são os mais estáveis diante de mudanças nas formas das distribuições. Mas diante das distribuições Normais contaminadas com apenas 10% de contaminação, a Trimédia de Tukey e a Média de Gastwirth são os mais eficientes.

Diante de distribuições com caudas muito pesadas e variância infinita, como a Cauchy, todos os estimadores deterioram-se, principalmente $A_4(\alpha)$ e os MQA. Por esse motivo não colocamos no gráfico alguns valores de erros quadráticos médios diante de situações em que as caudas da distribuição são muito longas.

Um fato interessante a ser observado é que a Trimédia de Tukey e a Média de Gastwirth têm comportamentos semelhantes frente as diferentes distribuições em estudo. No entanto, quando a distribuição é Cauchy percebemos

que a Média de Gastwirth é sempre mais eficiente que a Trimédia de Tukey. Talvez isso ocorra porque a Média de Gastwirth é baseada em quantis bem mais próximos à mediana, que é um estimador bastante resistente a alterações no conjunto de dados. Assim, diante da distribuição de Cauchy o resultado é um pouco diferente, tendo como estimadores mais eficientes a Média de Gastwirth e $A_{22}(\alpha)$.

Ao aumentarmos a porcentagem de dados aparados para 20% observamos que:

A performance dos estimadores de MQA e $A_4(\alpha)$ melhoram consideravelmente. No entanto, os estimadores L_1A praticamente não se alteram.

Diante da distribuição Normal e daquelas com caudas moderadamente mais longas que a Normal podemos perceber a formação de três grupos de estimadores. No primeiro grupo estão os mais eficientes, que são os MQA, a Trimédia de Tukey e a Média de Gastwirth. No segundo, temos $A_{12}(\alpha)$, $A_{22}(\alpha)$ e $A_{32}(\alpha)$. Em um terceiro grupo está $A_{42}(\alpha)$ com a pior performance. O estimador $A_4(\alpha)$, apesar de ter melhorado sua performance com $\alpha = 0.20$, ainda tem um comportamento muito instável. Ele pertence ao primeiro grupo quando a distribuição é Normal, mas quando alongamos as caudas da distribuição, sua eficiência cai muito, levando-o para o segundo grupo.

Nas situações em que as distribuições têm na sua forma grandes picos, todos os estimadores pioram o seu comportamento, principalmente $A_4(\alpha)$ e os MQA. Os melhores estimadores, nessas situações, são a Média de Gastwirth e a Trimédia de Tukey, cujos comportamentos são bem estáveis frente a mudanças nas formas das distribuições. Os estimadores L_1A não são muito afetados frente a essas mudanças, chegando em alguns casos ($n = 40$ e $n = 100$) a terem as segundas melhores eficiências, perdendo apenas para a Trimédia de Tukey e Média de Gastwirth.

Quando a distribuição dos erros tem caudas muito pesadas, os melhores estimadores ainda são a Trimédia de Tukey e a Média de Gastwirth. Em alguns casos, como as Normais contaminadas com 10% de contaminação, os MQA e os L_1A têm performances razoáveis, mas não tão boas quanto os dois primeiros. O estimador $A_4(\alpha)$ continua sendo o pior estimador diante dessas distribuições.

No caso da distribuição de Cauchy os estimadores mais eficientes são novamente a Média de Gastwirth e $A_{22}(\alpha)$. Os outros L_1A e a Trimédia de Tukey também têm uma boa performance, mas ainda um pouco inferiores aos outros dois.

Um fato interessante que merece destaque, é que apesar dos es-

timadores, cujas estimativas finais são L_1 , apresentarem muitas vezes uma boa performance, eles têm uma grande desvantagem. Em muitos casos, a solução do problema de minimização dos desvios absolutos para a estimativa final não é única. Diante de situações como essa, teríamos que adotar um critério para a escolha da solução. Este problema ocorre pelo fato de que ao retirarmos algumas observações do conjunto de dados, alteramos a estrutura da matriz \mathbf{X} , e como já vimos no Teorema 2.3, a solução única depende da estrutura dessa matriz.

Principalmente quando $\alpha = 10\%$, A_4 não tem uma boa performance, e esta vai piorando quando as caudas vão ficando mais longas. Quando $\alpha = 20\%$, sua performance melhora bastante, mas ainda não supera o comportamento de muitos outros estimadores como a Trimédia de Tukey e a Média de Gastwirth. Este fato ocorre para todos os tamanhos de amostra.

Outro fato que devemos ressaltar é que há evidências de que todos os estimadores aumentam sua eficiência e precisão a medida em que aumentamos o tamanho da amostra.

De um modo geral, a performance do estimador do intercepto e do parâmetro de inclinação são semelhantes diante dos diferentes tamanhos de amostra, diferentes distribuições dos erros e das diferentes porcentagens de dados aparados.

4.5 Exemplo

Utilizaremos um conjunto de dados, onde tenta-se explicar as médias bissemanais de salinidade de um rio no período de tempo i , pela salinidade medida há duas semanas atrás, por um dos períodos bissemanais, representado por i , e pelo volume de água do rio despejada no tempo i .

Os dados estão dispostos na tabela 4.3:

As estimativas dos parâmetros foram calculadas pelos métodos: $A_{22}(0.20)$, G e T. A escolha do método A_{22} se deu pelo fato de que, no nosso estudo de Monte Carlo o comportamento desse estimador foi estável e muito bom diante de distribuições com caudas muito pesadas como a Cauchy. A Trimédia de Tukey e a Média de Gastwirth, de acordo com o nosso estudo foram os estimadores que apresentaram as melhores performances diante das diversas distribuições consideradas.

Tabela 4.3: Conjunto de Dados sobre Salinidade

OBS	Y	X_1	X_2	X_3	ANO
1	7.6	8.2	4	23.005	72
2	7.7	7.6	5	23.873	
3	4.3	4.6	0	26.417	73
4	5.9	4.3	1	24.868	
5	5.0	5.9	2	29.895	
6	6.5	5.0	3	24.200	
7	8.3	6.5	4	23.215	
8	8.2	8.3	5	21.862	
9	13.2	10.1	0	22.274	74
10	12.6	13.2	1	23.830	
11	10.4	12.6	2	25.144	
12	10.8	10.4	3	22.430	
13	13.1	10.8	4	21.785	
14	12.3	13.1	5	22.380	
15	10.4	13.3	0	23.927	75
16	10.5	10.4	1	33.443	
17	7.7	10.5	2	24.859	
18	9.5	7.7	3	22.686	
19	12.0	10.0	0	21.789	76
20	12.6	12.0	1	22.041	
21	13.6	12.1	4	21.033	
22	14.1	13.6	5	21.005	
23	13.5	15.0	0	25.865	77
24	11.5	13.5	1	26.290	
25	12.0	11.5	2	22.932	
26	13.0	12.0	3	21.313	
27	14.1	13.0	4	20.769	
28	15.1	14.1	5	21.393	

NOTA: Y: médias bissemanais de salinidade medidas no tempo i ; X_1 : salinidade medida há duas semanas atrás; X_2 : um dos seis períodos bissemanais de março a maio; e X_3 : volume de água despejada no tempo i .

Tabela 4.4: Estimativas dos Parâmetros e a Mediana dos Desvios Absolutos (MAD) para os diferentes métodos utilizados

Estimadores	Intercepto	X_1	X_2	X_3	MAD
$A_{22}(0.20)$	18.4196	0.7042	-0.1839	-0.6178	0.48
T	15.9064	0.7361	-0.1148	-0.5322	0.46
G	14.2696	0.7301	-0.1138	-0.4596	0.49

Através da Mediana dos Desvios Absolutos (MAD) para os resíduos correspondentes a cada um dos três métodos utilizados, percebemos que há evidências de que o melhor estimador foi a Trimédia de Tukey. Mas, os outros dois estimadores também foram bons.

Este conjunto de dados foi analisado por Ruppert e Carroll (1980). Foram utilizados os métodos de mínimos quadrados aparados; o estimador L_1 ; o M-estimador de Huber; e o M-estimador de Andrews. O estimador que obteve a menor MAD (0.47) foi o M-estimador de Andrews, o segundo menor valor (0.50) de MAD foi do estimador L_1 . Quanto aos demais estimadores, os valores de MAD foram superiores a 0.56.

Diante desse exemplo, vimos que esses estimadores têm um bom desempenho quando comparados com outros estimadores robustos encontrados na literatura.

4.6 Conclusões

De acordo com esse estudo, os estimadores mais eficientes diante das diversas situações (diferentes tamanhos de amostra e diferentes distribuições dos erros) são a Trimédia de Tukey e a Média de Gastwirth, cujos comportamentos são bem estáveis para diferentes tipos de caudas da distribuição dos erros. Resaltemos também, que diante de situações em que as caudas da distribuição são muito alongadas, como a Cauchy, estes estimadores apresentam uma performance muito boa, comparada com a dos demais estimadores em estudo. Outra vantagem desses estimadores é que eles não exigem o cálculo de uma estimativa preliminar e a ordenação dos resíduos. Além disso, outra vantagem é que não precisamos desperdiçar nenhuma informação, visto que essas estimativas utilizam todos as observações do nosso conjunto de dados.

APÊNDICES

Apêndice A

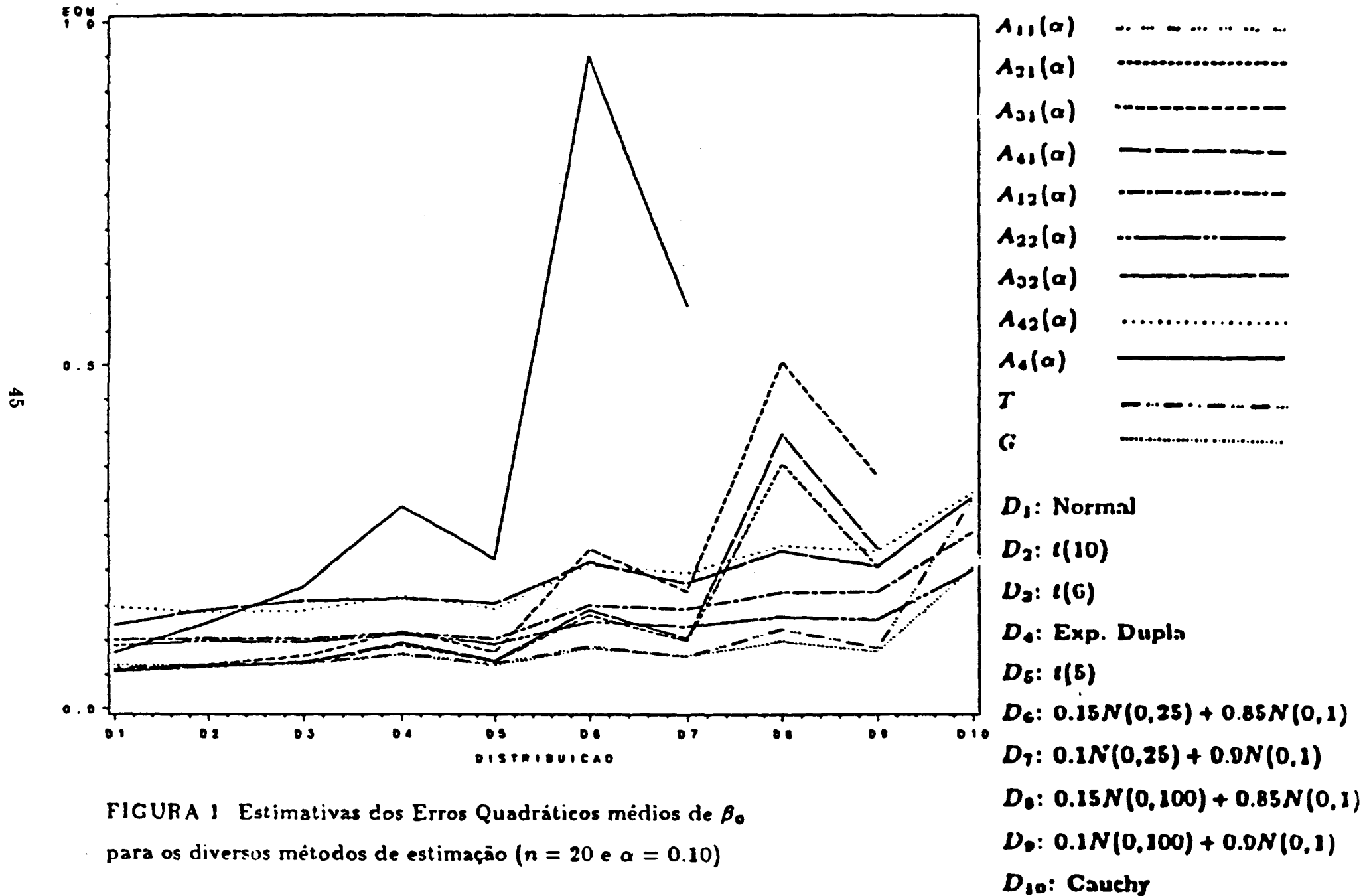
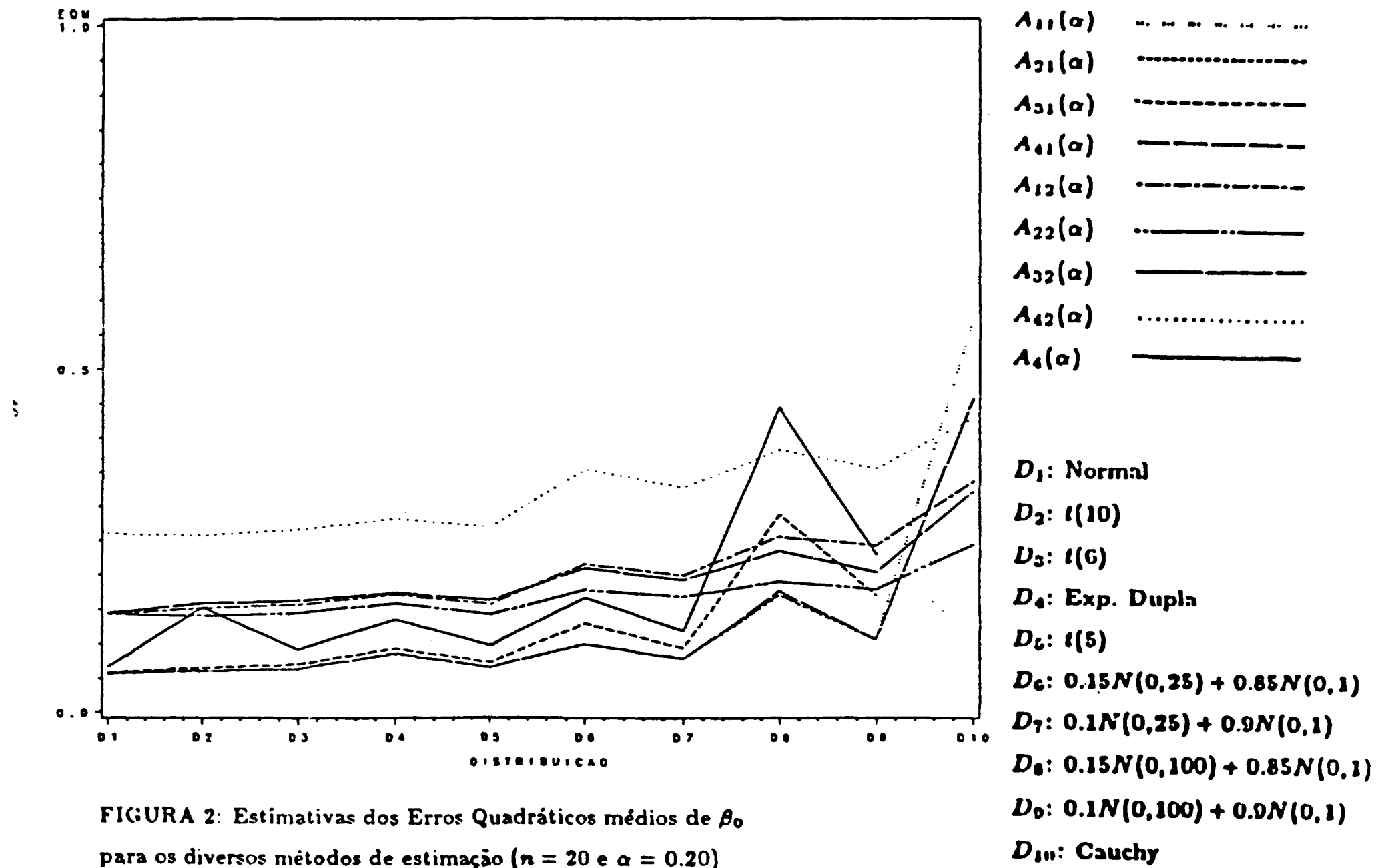


FIGURA 1 Estimativas dos Erros Quadráticos médios de β_0
para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.10$)



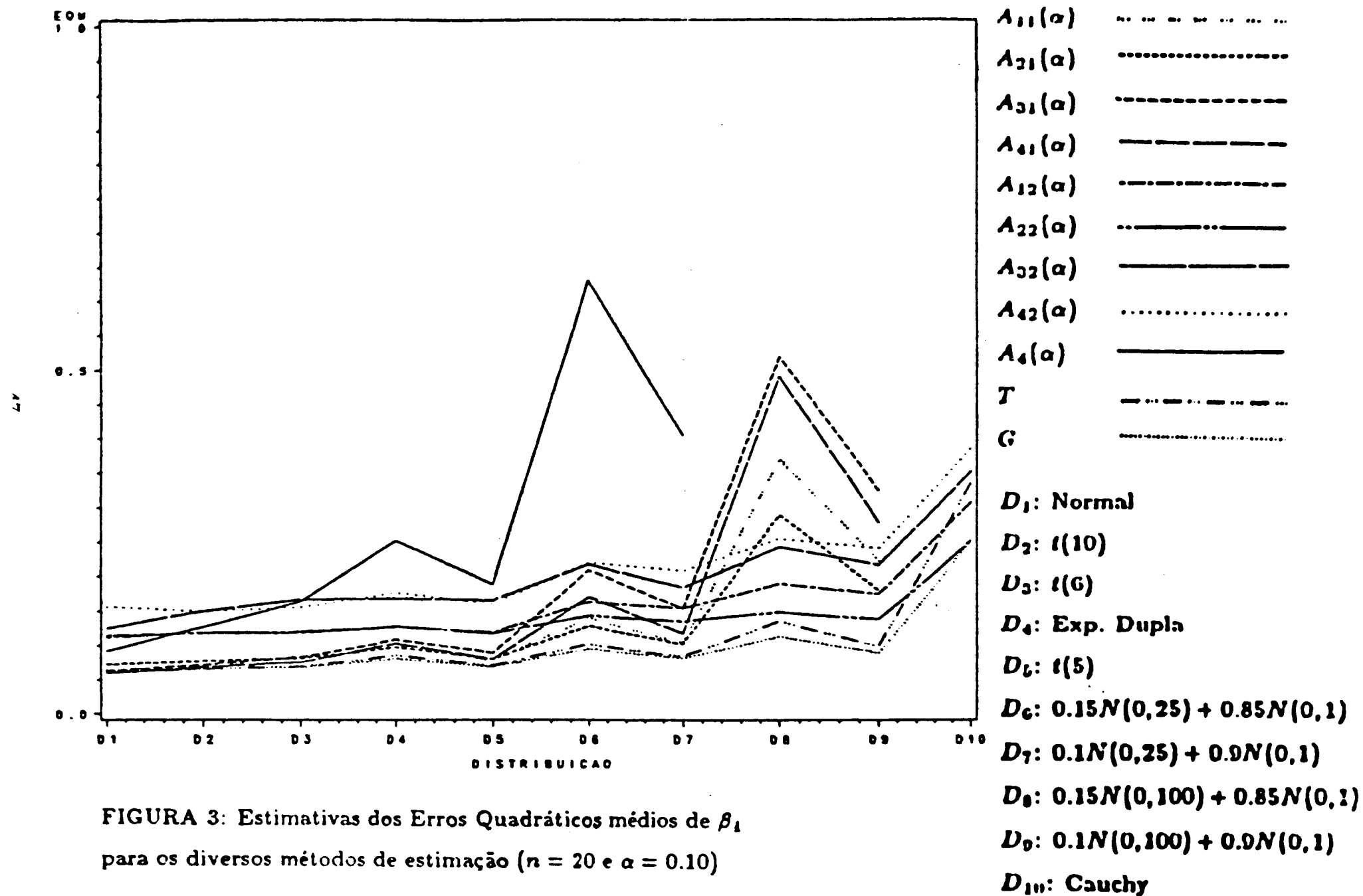
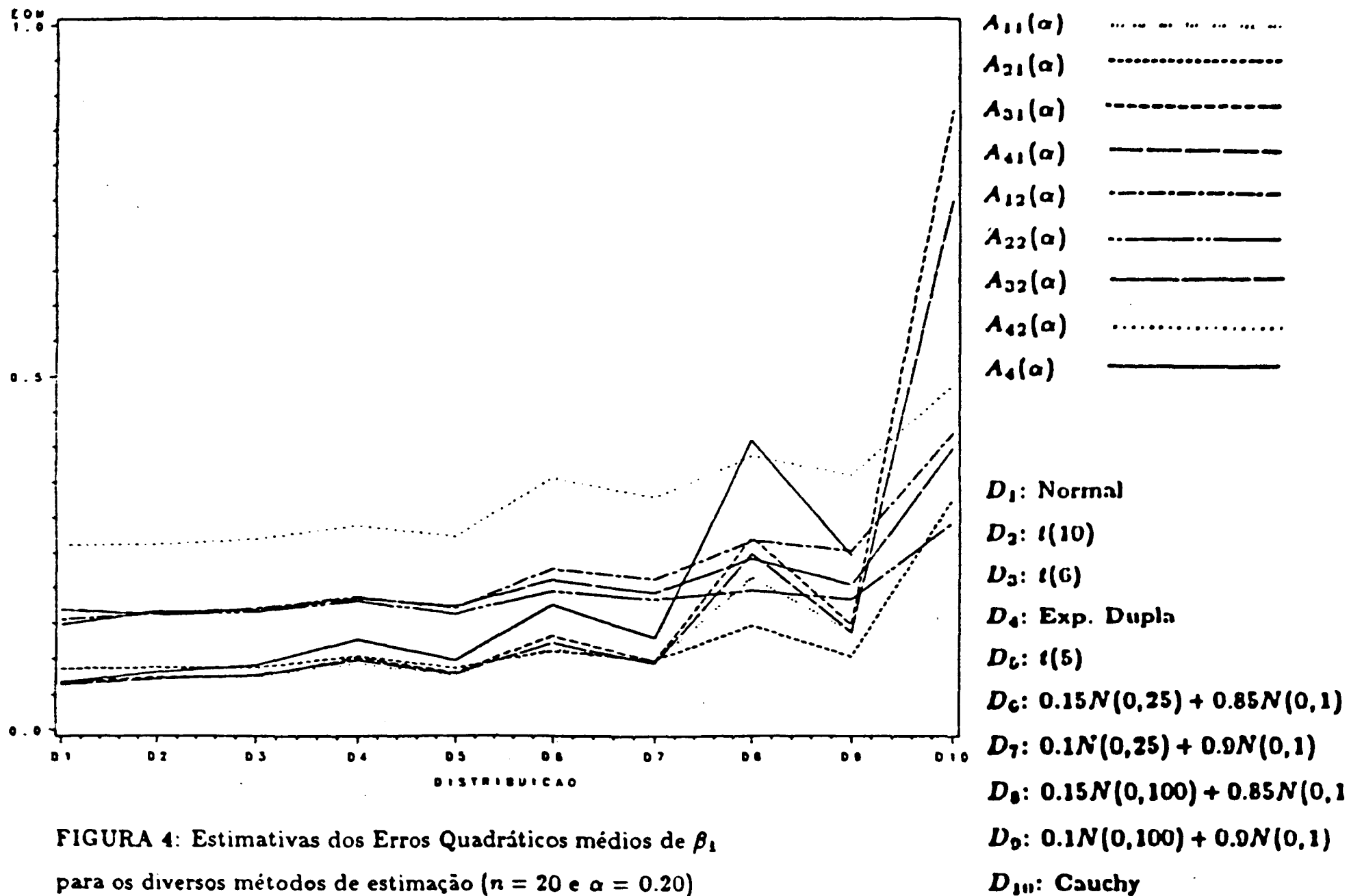
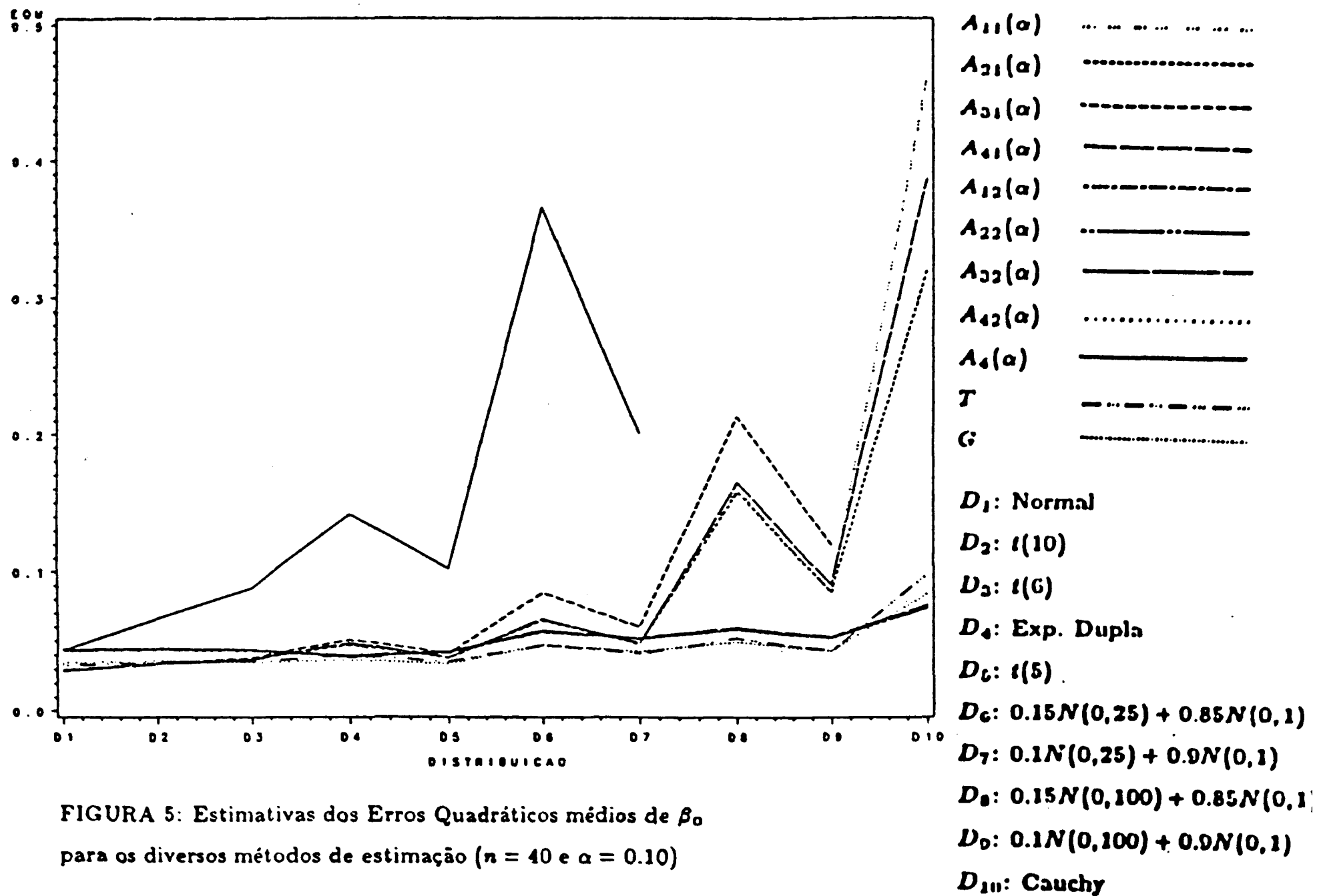


FIGURA 3: Estimativas dos Erros Quadráticos médios de β_1
para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.10$)





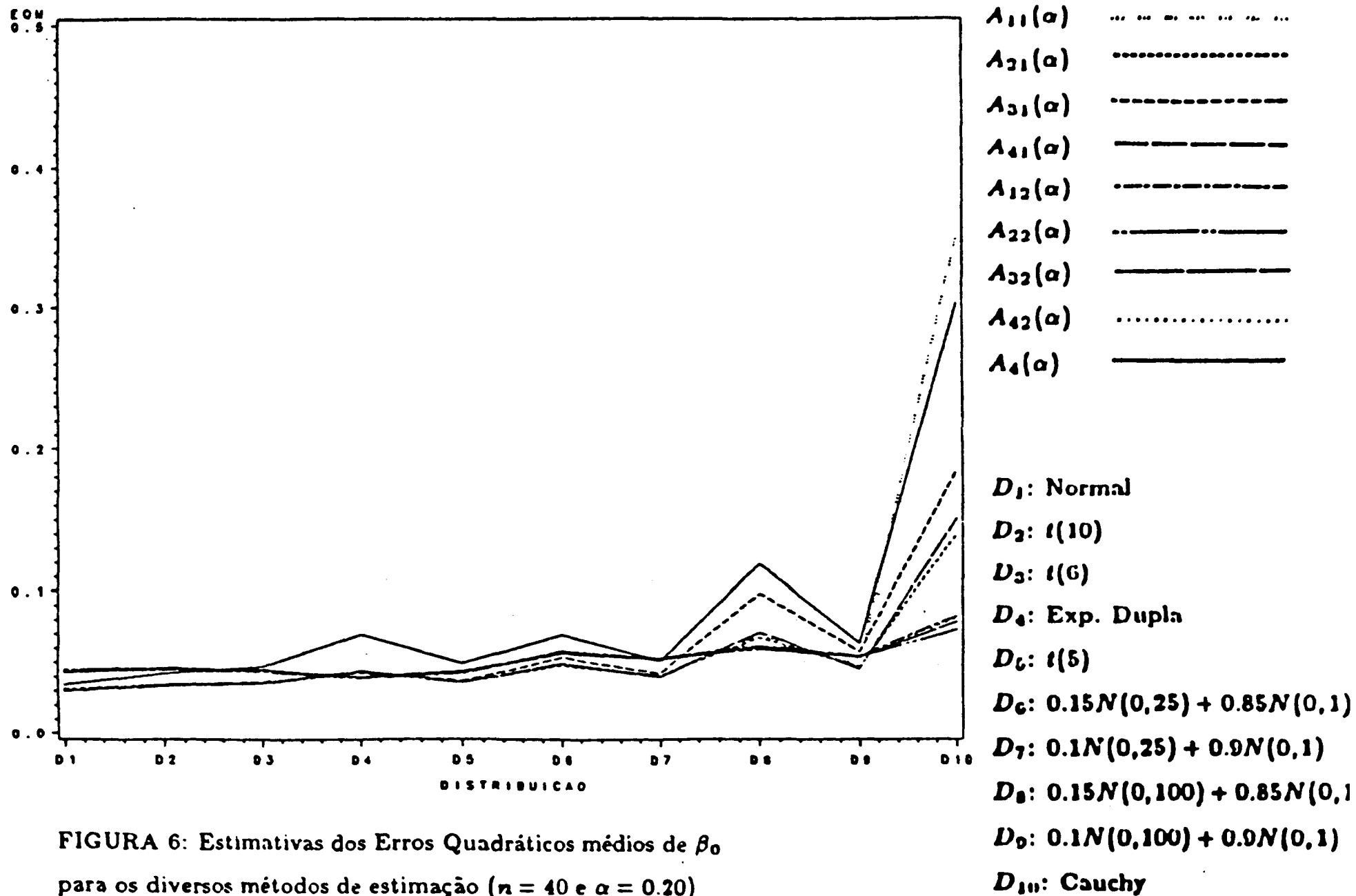
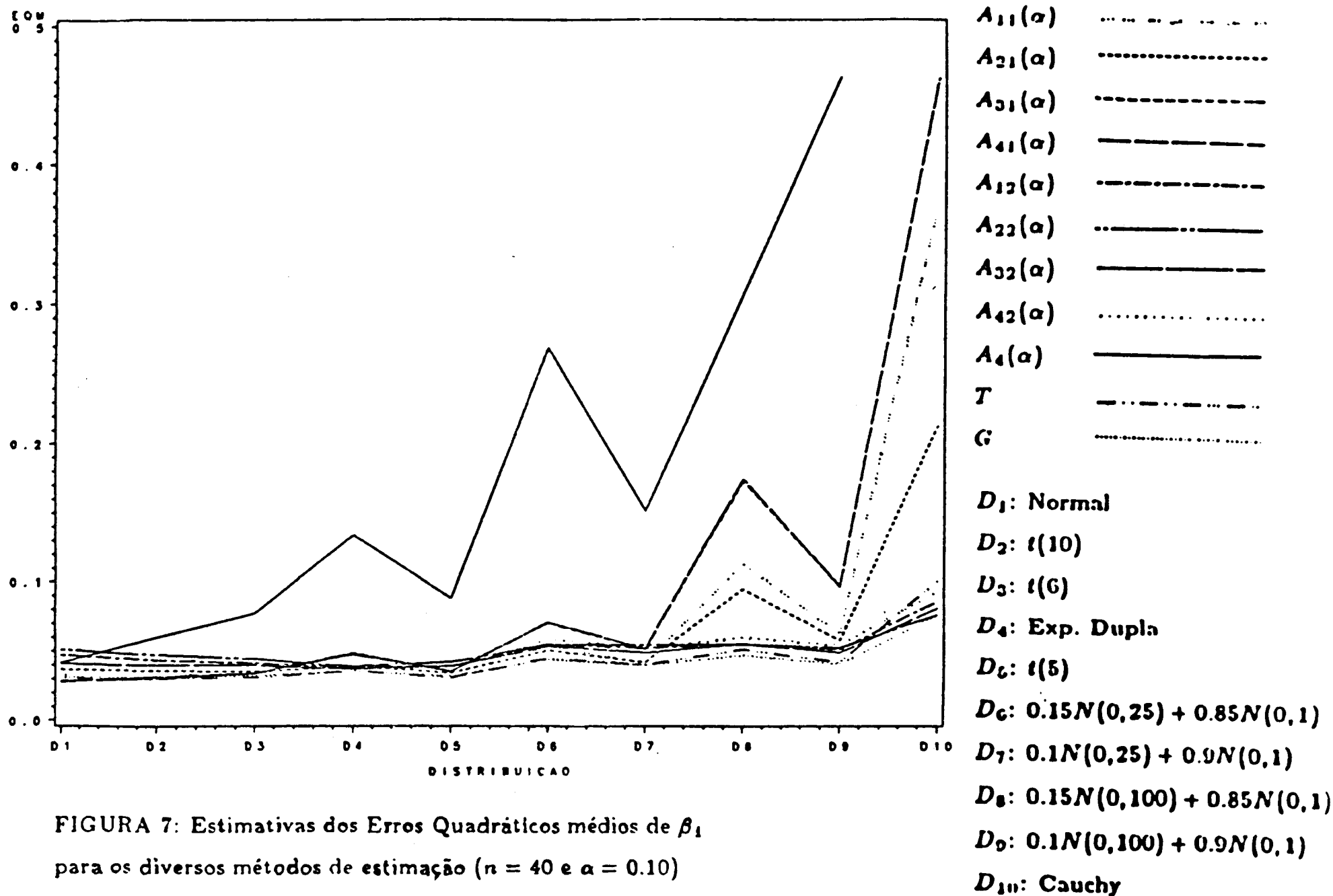
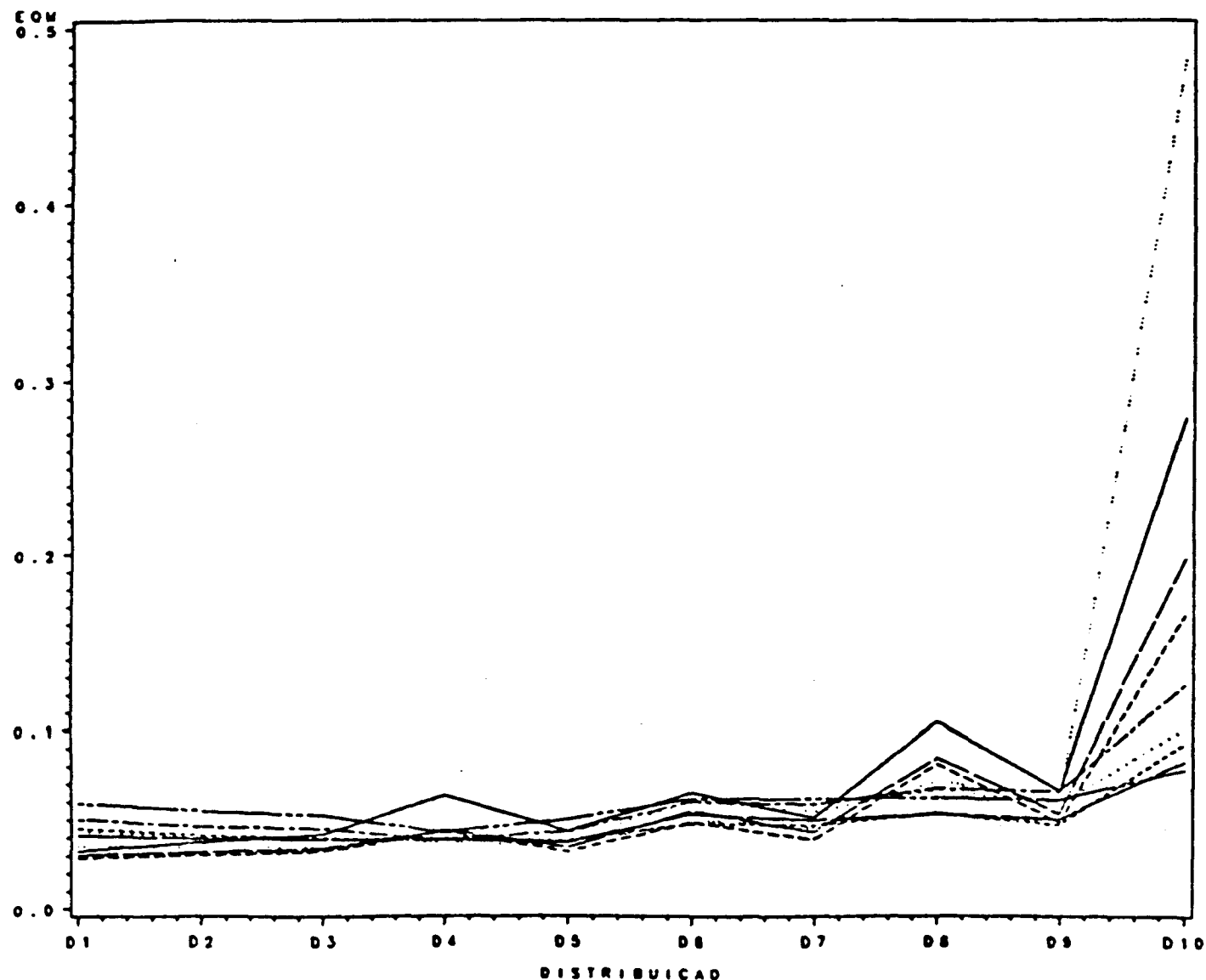


FIGURA 6: Estimativas dos Erros Quadráticos médios de β_0
para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.20$)





$A_{11}(\alpha)$
 $A_{21}(\alpha)$
 $A_{31}(\alpha)$
 $A_{41}(\alpha)$
 $A_{12}(\alpha)$
 $A_{22}(\alpha)$
 $A_{32}(\alpha)$
 $A_{42}(\alpha)$
 $A_4(\alpha)$

D_1 : Normal

D_2 : $t(10)$

D_3 : $t(6)$

D_4 : Exp. Dupla

D_5 : $t(5)$

D_6 : $0.15N(0,25) + 0.85N(0,1)$

D_7 : $0.1N(0,25) + 0.9N(0,1)$

D_8 : $0.15N(0,100) + 0.85N(0,1)$

D_9 : $0.1N(0,100) + 0.9N(0,1)$

D_{10} : Cauchy

FIGURA 8: Estimativas dos Erros Quadráticos médios de β_1
para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.20$)

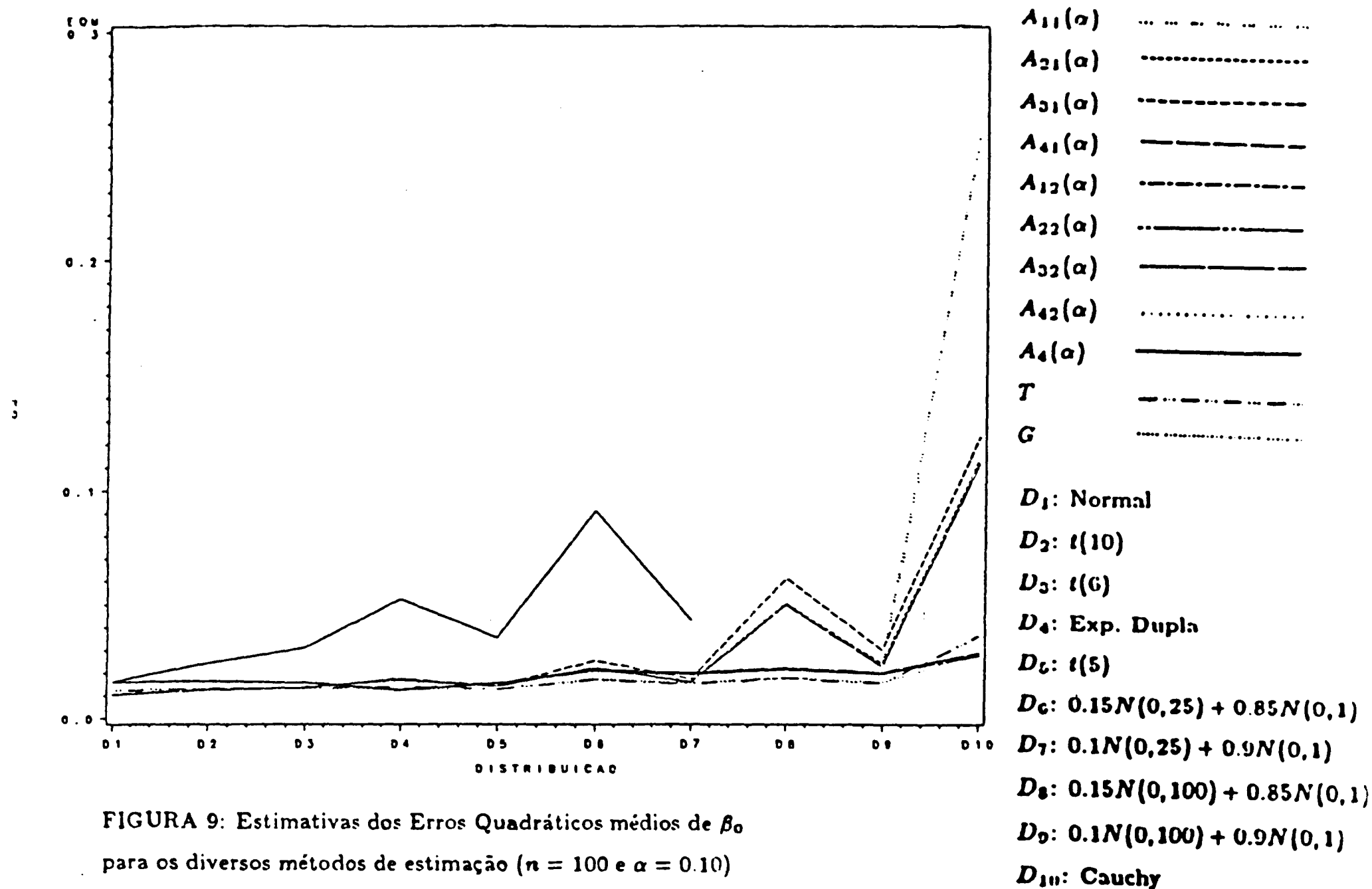


FIGURA 9: Estimativas dos Erros Quadráticos médios de β_0
para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.10$)

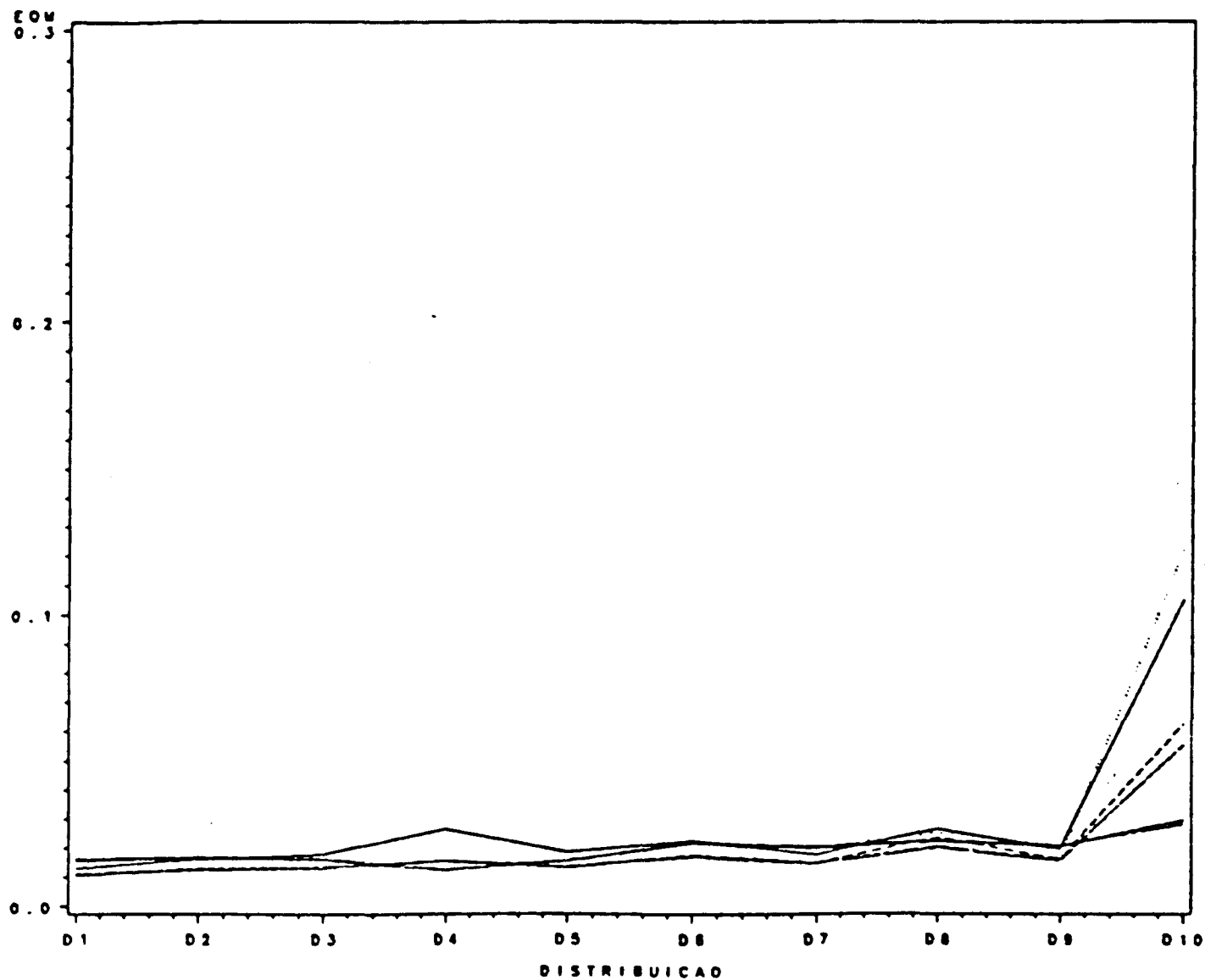


FIGURA 10: Estimativas dos Erros Quadráticos médios de β_0
para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.20$)

$A_{11}(\alpha)$ -
 $A_{21}(\alpha)$
 $A_{31}(\alpha)$ - - - - -
 $A_{41}(\alpha)$ - - - - -
 $A_{12}(\alpha)$ -
 $A_{22}(\alpha)$ -
 $A_{32}(\alpha)$ - - - - -
 $A_{42}(\alpha)$
 $A_4(\alpha)$ - - - - -

D_1 : Normal
 D_2 : $t(10)$
 D_3 : $t(6)$
 D_4 : Exp. Dupla
 D_5 : $t(5)$
 D_6 : $0.15N(0,25) + 0.85N(0,1)$
 D_7 : $0.1N(0,25) + 0.9N(0,1)$
 D_8 : $0.15N(0,100) + 0.85N(0,1)$
 D_9 : $0.1N(0,100) + 0.9N(0,1)$
 D_{10} : Cauchy

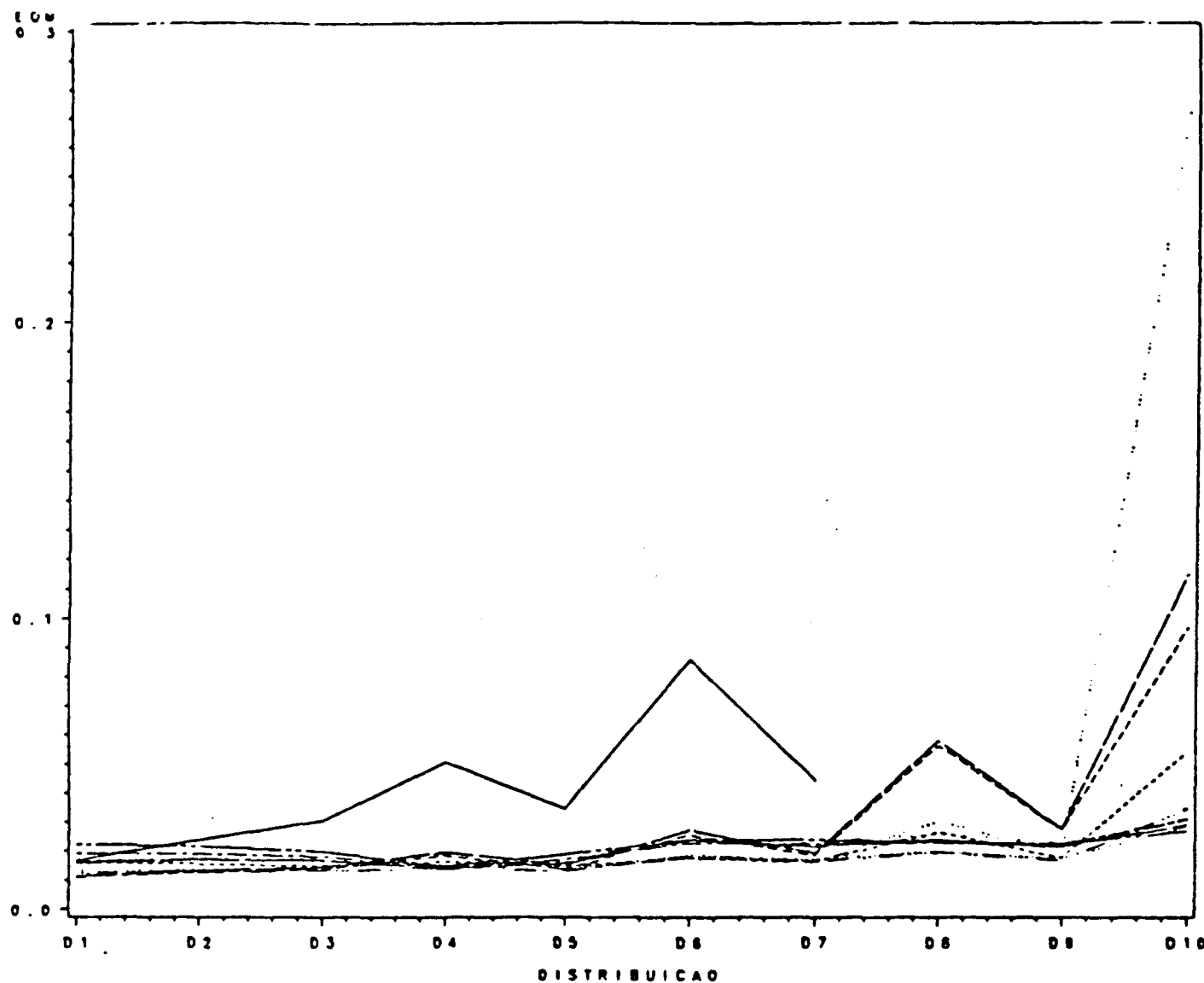


FIGURA 11: Estimativas dos Erros Quadráticos médios de β_1
para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.10$)

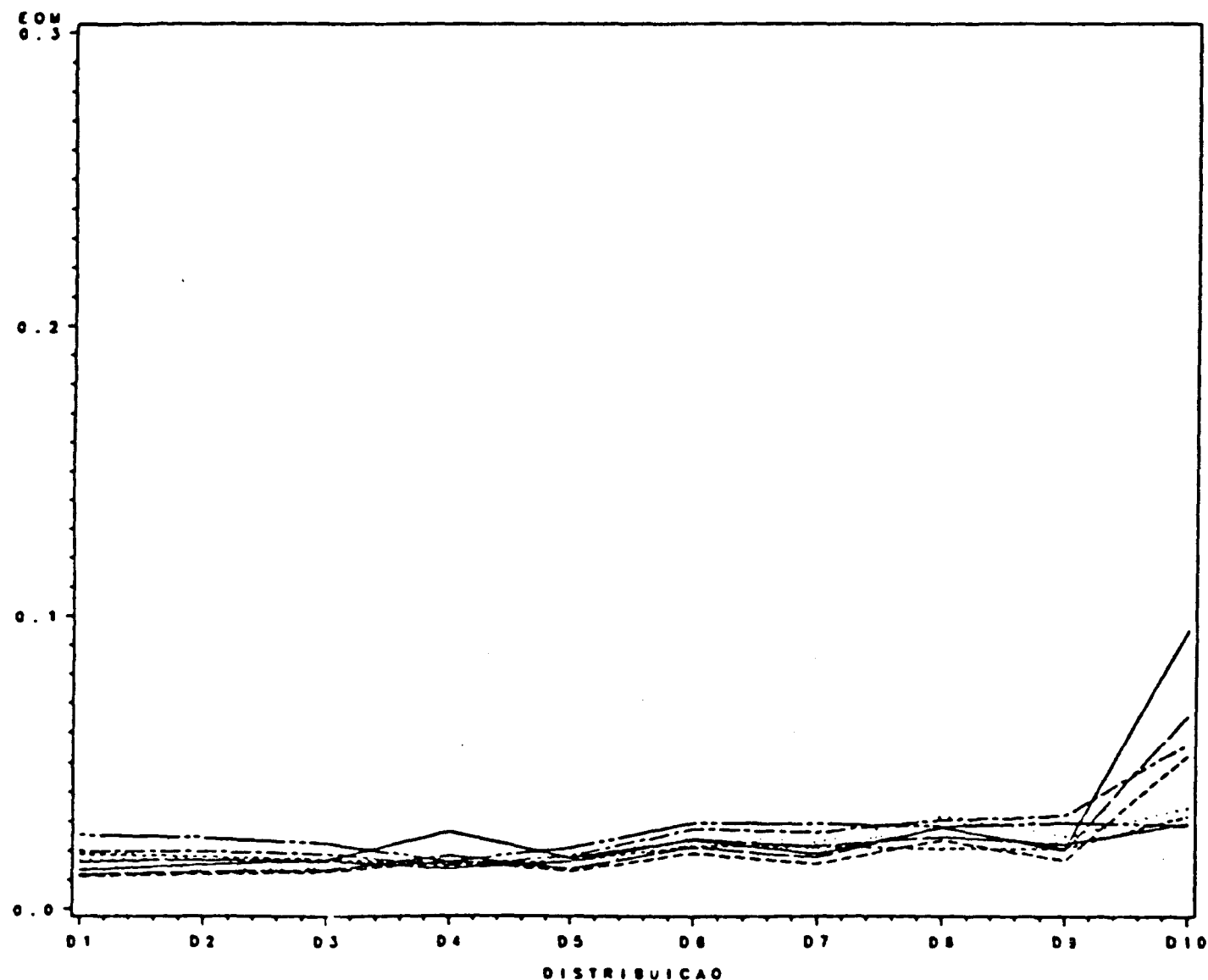


FIGURA 12: Estimativas dos Erros Quadráticos médios de β_1
para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.20$)

$A_{11}(\alpha)$
 $A_{21}(\alpha)$
 $A_{31}(\alpha)$
 $A_{41}(\alpha)$
 $A_{12}(\alpha)$
 $A_{22}(\alpha)$
 $A_{32}(\alpha)$
 $A_{42}(\alpha)$
 $A_4(\alpha)$

D_1 : Normal
 D_2 : $t(10)$
 D_3 : $t(6)$
 D_4 : Exp. Dupla
 D_5 : $t(5)$
 D_6 : $0.15N(0,25) + 0.85N(0,1)$
 D_7 : $0.1N(0,25) + 0.9N(0,1)$
 D_8 : $0.15N(0,100) + 0.85N(0,1)$
 D_9 : $0.1N(0,100) + 0.9N(0,1)$
 D_{10} : Cauchy

Apêndice B

Tabela 1: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0547 (0.0016)	0.0550 (0.0016)	0.0580 (0.0018)	0.0556 (0.0017)	0.1013 (0.0041)	0.0932 (0.0036)	0.1234 (0.0051)	0.1491 (0.0062)	0.0830 (0.0027)	0.0612 (0.0018)	0.0646 (0.0019)
D_2	0.0607 (0.0019)	0.0610 (0.0019)	0.0652 (0.0021)	0.0623 (0.0019)	0.1027 (0.0041)	0.0992 (0.0039)	0.1447 (0.0059)	0.1412 (0.0059)	0.1257 (0.0045)	0.0648 (0.0020)	0.0644 (0.0020)
D_3	0.0671 (0.0021)	0.0673 (0.0021)	0.0774 (0.0026)	0.0686 (0.0022)	0.1026 (0.0042)	0.0975 (0.0039)	0.1578 (0.0065)	0.1439 (0.0060)	0.1773 (0.0072)	0.0669 (0.0021)	0.0651 (0.0021)
D_4	0.0925 (0.0032)	0.0924 (0.0032)	0.1123 (0.0038)	0.0958 (0.0033)	0.1110 (0.0049)	0.1082 (0.0048)	0.1605 (0.0068)	0.1641 (0.0070)	0.2907 (0.0108)	0.0799 (0.0028)	0.0785 (0.0030)
D_5	0.0697 (0.0022)	0.0699 (0.0022)	0.0837 (0.0031)	0.0713 (0.0023)	0.1024 (0.0043)	0.0954 (0.0039)	0.1523 (0.0064)	0.1465 (0.0062)	0.2151 (0.0097)	0.0666 (0.0021)	0.0638 (0.0021)
D_6	0.1378 (0.0050)	0.1383 (0.0050)	0.2310 (0.0082)	0.1457 (0.0052)	0.1515 (0.0057)	0.1279 (0.0048)	0.2083 (0.0075)	0.2071 (0.0076)	0.9514 (0.0444)	0.0927 (0.0031)	0.0894 (0.0030)
D_7	0.0999 (0.0036)	0.1005 (0.1005)	0.1718 (0.0063)	0.1040 (0.0037)	0.1471 (0.0058)	0.1221 (0.0047)	0.1839 (0.0068)	0.1977 (0.0075)	0.5887 (0.0345)	0.0781 (0.0026)	0.0787 (0.0025)
D_8	0.3547 (0.0157)	0.3555 (0.0156)	0.5053 (0.0208)	0.3988 (0.0177)	0.1706 (0.0063)	0.1355 (0.0049)	0.2294 (0.0082)	0.2368 (0.0083)	4.6526 (0.2120)	0.1172 (0.0052)	0.0990 (0.0038)
D_9	0.2065 (0.0103)	0.2067 (0.0102)	0.3387 (0.0154)	0.2324 (0.0120)	0.1702 (0.0065)	0.1300 (0.0050)	0.2067 (0.0075)	0.2296 (0.0083)	2.7423 (0.1659)	0.0883 (0.0038)	0.0831 (0.0030)
D_{10}	1.8133 (0.3123)	1.7607 (0.3086)	50.5961 (25.3406)	2.0003 (0.3024)	0.2546 (0.0086)	0.2008 (0.0071)	0.3071 (0.0102)	0.3138 (0.0100)	‡ (‡)	0.3041 (0.0197)	0.2047 (0.0080)

‡: valor acima de 100

(): desvio padrão do EQM de β_0

Tabela 2: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0648 (0.0021)	0.0726 (0.0023)	0.0626 (0.0021)	0.0602 (0.0019)	0.1145 (0.0044)	0.1122 (0.0041)	0.1251 (0.0051)	0.1567 (0.0064)	0.0922 (0.0030)	0.0604 (0.0019)	0.0609 (0.0019)
D_2	0.0709 (0.0024)	0.0774 (0.0026)	0.0722 (0.0024)	0.0685 (0.0022)	0.1177 (0.0047)	0.1192 (0.0046)	0.1499 (0.0061)	0.1502 (0.0062)	0.1274 (0.0043)	0.0659 (0.0022)	0.0657 (0.0022)
D_3	0.0760 (0.0027)	0.0815 (0.0029)	0.0832 (0.0030)	0.0761 (0.0026)	0.1187 (0.0049)	0.1196 (0.0047)	0.1678 (0.0067)	0.1557 (0.0064)	0.1641 (0.0058)	0.0697 (0.0024)	0.0685 (0.0023)
D_4	0.0993 (0.0038)	0.0975 (0.0039)	0.1087 (0.0039)	0.1033 (0.0034)	0.1278 (0.0055)	0.1278 (0.0054)	0.1684 (0.0068)	0.1771 (0.0071)	0.2514 (0.0086)	0.0856 (0.0031)	0.0808 (0.0029)
D_5	0.0775 (0.0028)	0.0801 (0.0029)	0.0891 (0.0034)	0.0796 (0.0028)	0.1186 (0.0050)	0.1167 (0.0047)	0.1655 (0.0068)	0.1596 (0.0066)	0.1884 (0.0070)	0.0704 (0.0025)	0.0686 (0.0024)
D_6	0.1409 (0.0066)	0.1292 (0.0056)	0.2104 (0.0092)	0.1718 (0.0077)	0.1647 (0.0067)	0.1445 (0.0057)	0.2182 (0.0081)	0.2216 (0.0081)	0.6327 (0.0272)	0.1023 (0.0040)	0.0957 (0.0037)
D_7	0.1046 (0.0045)	0.1027 (0.0041)	0.1552 (0.0071)	0.1183 (0.0053)	0.1556 (0.0062)	0.1363 (0.0053)	0.1864 (0.0071)	0.2095 (0.0078)	0.4070 (0.0207)	0.0844 (0.0031)	0.0817 (0.0029)
D_8	0.3713 (0.0235)	0.2907 (0.0175)	0.5198 (0.0326)	0.4913 (0.0281)	0.1916 (0.0101)	0.1504 (0.0063)	0.2436 (0.0109)	0.2559 (0.0106)	2.4732 (0.1156)	0.1370 (0.0073)	0.1144 (0.0060)
D_9	0.2237 (0.0155)	0.1813 (0.0110)	0.3256 (0.0228)	0.2791 (0.0180)	0.1760 (0.0069)	0.1388 (0.0052)	0.2179 (0.0082)	0.2418 (0.0088)	1.4620 (0.0859)	0.0988 (0.0047)	0.0894 (0.0037)
D_{10}	1.7631 (0.2930)	1.2673 (0.2574)	11.3135 (4.0333)	2.2996 (0.3327)	0.3076 (0.0132)	0.2514 (0.0107)	0.3530 (0.0146)	0.3867 (0.0165)	† (†)	0.3372 (0.0247)	0.2523 (0.0141)

†: valor acima de 100

():desvio padrão do EQM de β_1

Tabela 3: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0568 (0.0017)	0.0571 (0.0017)	0.0590 (0.0018)	0.0574 (0.0017)	0.1429 (0.0060)	0.1439 (0.0060)	0.1454 (0.0060)	0.2595 (0.0095)	0.0671 (0.0020)	0.0612 (0.0018)	0.0646 (0.0019)
D_2	0.0605 (0.0019)	0.0604 (0.0019)	0.0653 (0.0020)	0.0607 (0.0019)	0.1520 (0.0063)	0.1410 (0.0058)	0.1591 (0.0065)	0.2578 (0.0094)	0.1533 (0.0024)	0.0648 (0.0020)	0.0644 (0.0020)
D_3	0.0642 (0.0020)	0.0643 (0.0020)	0.0707 (0.0022)	0.0642 (0.0020)	0.1575 (0.0066)	0.1450 (0.0060)	0.1628 (0.0066)	0.2652 (0.0095)	0.0916 (0.0028)	0.0669 (0.0021)	0.0651 (0.0021)
D_4	0.0848 (0.0030)	0.0846 (0.0029)	0.0928 (0.0034)	0.0857 (0.0029)	0.1706 (0.0072)	0.1581 (0.0068)	0.1729 (0.0073)	0.2803 (0.0101)	0.1355 (0.0046)	0.0799 (0.0028)	0.0785 (0.0030)
D_5	0.0650 (0.0021)	0.0652 (0.0021)	0.0725 (0.0024)	0.0652 (0.0021)	0.1573 (0.0066)	0.1419 (0.0060)	0.1630 (0.0067)	0.2684 (0.0097)	0.0969 (0.0031)	0.0666 (0.0021)	0.0638 (0.0021)
D_6	0.1002 (0.0035)	0.0997 (0.0035)	0.1301 (0.0051)	0.1000 (0.0034)	0.2154 (0.0078)	0.1781 (0.0066)	0.2095 (0.0077)	0.3543 (0.0111)	0.1678 (0.0075)	0.0927 (0.0031)	0.0894 (0.0030)
D_7	0.0794 (0.0027)	0.0792 (0.0027)	0.0943 (0.0037)	0.0791 (0.0027)	0.1989 (0.0076)	0.1687 (0.0065)	0.1922 (0.0072)	0.3273 (0.0108)	0.1187 (0.0049)	0.0781 (0.0026)	0.0787 (0.0025)
D_8	0.1711 (0.0084)	0.1701 (0.0084)	0.2871 (0.0147)	0.1759 (0.0086)	0.2550 (0.0088)	0.1893 (0.0069)	0.2335 (0.0081)	0.3823 (0.0115)	0.4436 (0.0317)	0.1172 (0.0052)	0.0990 (0.0038)
D_9	0.1061 (0.0053)	0.1057 (0.0053)	0.1704 (0.0100)	0.1058 (0.0054)	0.2422 (0.0087)	0.1787 (0.0068)	0.2035 (0.0075)	0.3548 (0.0113)	0.2286 (0.0174)	0.0883 (0.0038)	0.0831 (0.0030)
D_{10}	0.5712 (0.0468)	0.4536 (0.0251)	1.0197 (0.1428)	0.4565 (0.0336)	0.3362 (0.0109)	0.2436 (0.0086)	0.3214 (0.0103)	0.4278 (0.0124)	2.7218 (1.2039)	0.3041 (0.0197)	0.2047 (0.0080)

():desvio padrão do EQM de β_0

Tabela 4: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 20$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0683 (0.0021)	0.0861 (0.0028)	0.0655 (0.0021)	0.0634 (0.0020)	0.1553 (0.0062)	0.1686 (0.0064)	0.1481 (0.0061)	0.2606 (0.0095)	0.0664 (0.0022)	0.0604 (0.0019)	0.0609 (0.0019)
D_2	0.0712 (0.0024)	0.0876 (0.0030)	0.0739 (0.0025)	0.0710 (0.0024)	0.1632 (0.0064)	0.1617 (0.0062)	0.1659 (0.0067)	0.2620 (0.0094)	0.0805 (0.0026)	0.0659 (0.0022)	0.0657 (0.0022)
D_3	0.0751 (0.0026)	0.0877 (0.0030)	0.0762 (0.0022)	0.0773 (0.0027)	0.1705 (0.0068)	0.1653 (0.0064)	0.1682 (0.0067)	0.2700 (0.0096)	0.0914 (0.0031)	0.0697 (0.0024)	0.0685 (0.0023)
D_4	0.0932 (0.0034)	0.1037 (0.0042)	0.1014 (0.0040)	0.0978 (0.0036)	0.1871 (0.0074)	0.1801 (0.0072)	0.1843 (0.0075)	0.2884 (0.0101)	0.1270 (0.0042)	0.0856 (0.0031)	0.0808 (0.0029)
D_5	0.0752 (0.0027)	0.0865 (0.0031)	0.0800 (0.0029)	0.0781 (0.0027)	0.1702 (0.0069)	0.1628 (0.0064)	0.1732 (0.0070)	0.2730 (0.0098)	0.0966 (0.0033)	0.0704 (0.0025)	0.0686 (0.0024)
D_6	0.1139 (0.0051)	0.1107 (0.0046)	0.1318 (0.0059)	0.1220 (0.0052)	0.2265 (0.0083)	0.1947 (0.0075)	0.2105 (0.0077)	0.3555 (0.0112)	0.1754 (0.0076)	0.1023 (0.0040)	0.0957 (0.0037)
D_7	0.0951 (0.0038)	0.0976 (0.0036)	0.0963 (0.0042)	0.0930 (0.0038)	0.2119 (0.0079)	0.1831 (0.0070)	0.1927 (0.0073)	0.3283 (0.0110)	0.1295 (0.0055)	0.0844 (0.0031)	0.0817 (0.0029)
D_8	0.2145 (0.0167)	0.1467 (0.0100)	0.2719 (0.0180)	0.2484 (0.0164)	0.2677 (0.0094)	0.1963 (0.0074)	0.2414 (0.0094)	0.3869 (0.0135)	0.4092 (0.0252)	0.1370 (0.0073)	0.1144 (0.0060)
D_9	0.1367 (0.0103)	0.1023 (0.0046)	0.1475 (0.0103)	0.1366 (0.0103)	0.2530 (0.0090)	0.1831 (0.0069)	0.2037 (0.0076)	0.3602 (0.0131)	0.2472 (0.0182)	0.0988 (0.0047)	0.0894 (0.0037)
D_{10}	0.7528 (0.0778)	0.3267 (0.0179)	0.8772 (0.1240)	0.7481 (0.1107)	0.4200 (0.0219)	0.2934 (0.0168)	0.3974 (0.0251)	0.4878 (0.0230)	2.6092 (1.1888)	0.3372 (0.0247)	0.2523 (0.0141)

(): desvio padrão do EQM de β_1

Tabela 5: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0291 (0.0009)	0.0296 (0.0009)	0.0292 (0.0009)	0.0287 (0.0009)	0.0435 (0.0013)	0.0440 (0.0013)	0.0434 (0.0013)	0.0422 (0.0013)	0.0436 (0.0013)	0.0328 (0.0010)	0.0346 (0.0011)
D_2	0.0340 (0.0011)	0.0341 (0.0011)	0.0343 (0.0011)	0.0341 (0.0011)	0.0447 (0.0015)	0.0451 (0.0015)	0.0443 (0.0014)	0.0438 (0.0014)	0.0668 (0.0023)	0.0351 (0.0011)	0.0363 (0.0012)
D_3	0.0370 (0.0012)	0.0369 (0.0012)	0.0382 (0.0013)	0.0371 (0.0012)	0.0438 (0.0015)	0.0440 (0.0015)	0.0439 (0.0015)	0.0429 (0.0014)	0.0887 (0.0033)	0.0358 (0.0012)	0.0348 (0.0011)
D_4	0.0472 (0.0015)	0.0473 (0.0015)	0.0511 (0.0017)	0.0486 (0.0016)	0.0386 (0.0014)	0.0390 (0.0014)	0.0399 (0.0015)	0.0388 (0.0014)	0.1419 (0.0048)	0.0398 (0.0014)	0.0367 (0.0012)
D_5	0.0381 (0.0013)	0.0380 (0.0013)	0.0413 (0.0014)	0.0384 (0.0013)	0.0422 (0.0014)	0.0423 (0.0014)	0.0427 (0.0014)	0.0414 (0.0014)	0.1025 (0.0039)	0.0354 (0.0011)	0.0339 (0.0011)
D_6	0.0660 (0.0024)	0.0653 (0.0023)	0.0856 (0.0032)	0.0663 (0.0024)	0.0570 (0.0018)	0.0568 (0.0018)	0.0582 (0.0018)	0.0565 (0.0018)	0.3660 (0.0167)	0.0473 (0.0015)	0.0474 (0.0015)
D_7	0.0480 (0.0017)	0.0480 (0.0016)	0.0601 (0.0023)	0.0484 (0.0017)	0.0517 (0.0017)	0.0518 (0.0016)	0.0522 (0.0016)	0.0518 (0.0016)	0.2013 (0.0119)	0.0412 (0.0013)	0.0426 (0.0014)
D_8	0.1601 (0.0069)	0.1583 (0.0067)	0.2134 (0.0089)	0.1652 (0.0072)	0.0586 (0.0019)	0.0584 (0.0019)	0.0601 (0.0019)	0.0587 (0.0019)	1.9371 (0.0905)	0.0523 (0.0018)	0.0497 (0.0016)
D_9	0.0860 (0.0040)	0.0860 (0.0039)	0.1208 (0.0055)	0.0916 (0.0045)	0.0531 (0.0017)	0.0530 (0.0017)	0.0536 (0.0017)	0.0530 (0.0017)	0.8351 (0.0609)	0.0435 (0.0014)	0.0439 (0.0014)
D_{10}	0.4599 (0.0827)	0.3206 (0.0150)	0.5414 (0.0445)	0.3865 (0.0247)	0.0754 (0.0027)	0.0740 (0.0027)	0.0768 (0.0030)	0.0756 (0.0030)	4.3997 (0.4109)	0.0989 (0.0036)	0.0847 (0.0031)

():desvio padrão do EQM de β_0

Tabela 6: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0328 (0.0010)	0.0367 (0.0012)	0.0282 (0.0009)	0.0276 (0.0009)	0.0471 (0.0015)	0.0512 (0.0016)	0.0409 (0.0013)	0.0412 (0.0013)	0.0417 (0.0013)	0.0302 (0.0010)	0.0316 (0.0010)
D_2	0.0332 (0.0011)	0.0360 (0.0012)	0.0302 (0.0010)	0.0313 (0.0010)	0.0436 (0.0014)	0.0474 (0.0016)	0.0397 (0.0013)	0.0399 (0.0013)	0.0599 (0.0019)	0.0300 (0.0010)	0.0313 (0.0010)
D_3	0.0337 (0.0011)	0.0352 (0.0012)	0.0339 (0.0011)	0.0341 (0.0011)	0.0415 (0.0014)	0.0443 (0.0015)	0.0398 (0.0013)	0.0395 (0.0013)	0.0770 (0.0025)	0.0308 (0.0010)	0.0323 (0.0010)
D_4	0.0415 (0.0014)	0.0384 (0.0013)	0.0483 (0.0017)	0.0477 (0.0017)	0.0381 (0.0014)	0.0390 (0.0015)	0.0367 (0.0014)	0.0378 (0.0014)	0.1335 (0.0046)	0.0355 (0.0013)	0.0353 (0.0013)
D_5	0.0335 (0.0011)	0.0341 (0.0011)	0.0349 (0.0011)	0.0351 (0.0011)	0.0388 (0.0013)	0.0419 (0.0014)	0.0387 (0.0013)	0.0378 (0.0012)	0.0876 (0.0029)	0.0305 (0.0010)	0.0317 (0.0010)
D_6	0.0534 (0.0021)	0.0506 (0.0019)	0.0705 (0.0028)	0.0708 (0.0031)	0.0548 (0.0018)	0.0548 (0.0017)	0.0533 (0.0017)	0.0583 (0.0019)	0.2689 (0.0110)	0.0449 (0.0015)	0.0439 (0.0014)
D_7	0.0411 (0.0014)	0.0415 (0.0014)	0.0508 (0.0020)	0.0511 (0.0019)	0.0516 (0.0016)	0.0538 (0.0017)	0.0485 (0.0015)	0.0530 (0.0016)	0.1503 (0.0066)	0.0397 (0.0013)	0.0391 (0.0012)
D_8	0.1123 (0.0058)	0.0944 (0.0041)	0.1720 (0.0090)	0.1738 (0.0096)	0.0550 (0.0018)	0.0546 (0.0017)	0.0551 (0.0018)	0.0596 (0.0019)	1.0446 (0.0464)	0.0510 (0.0019)	0.0469 (0.0015)
D_9	0.0628 (0.0033)	0.0578 (0.0029)	0.0964 (0.0056)	0.0969 (0.0053)	0.0512 (0.0016)	0.0526 (0.0016)	0.0489 (0.0017)	0.0546 (0.0017)	0.4641 (0.0263)	0.0420 (0.0014)	0.0406 (0.0013)
D_{10}	0.3737 (0.0288)	0.2122 (0.0120)	0.5116 (0.0735)	0.4622 (0.0315)	0.0856 (0.3224)	0.0756 (0.0035)	0.0806 (0.0027)	0.0924 (0.0030)	3.2413 (0.0040)	0.1002 (0.0038)	0.0755 (0.0028)

(): desvio padrão do EQM de β_1

Tabela 7: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0303 (0.0009)	0.0312 (0.0010)	0.0300 (0.0009)	0.0299 (0.0009)	0.0438 (0.0013)	0.0449 (0.0014)	0.0428 (0.0013)	0.0429 (0.0013)	0.0343 (0.0011)	0.0328 (0.0010)	0.0346 (0.0011)
D_2	0.0335 (0.0011)	0.0341 (0.0011)	0.0341 (0.0011)	0.0335 (0.0011)	0.0446 (0.0015)	0.0458 (0.0015)	0.0454 (0.0015)	0.0443 (0.0014)	0.0420 (0.0014)	0.0351 (0.0011)	0.0363 (0.0012)
D_3	0.0353 (0.0012)	0.0357 (0.0012)	0.0361 (0.0012)	0.0355 (0.0012)	0.0436 (0.0014)	0.0449 (0.0015)	0.0447 (0.0015)	0.0434 (0.0014)	0.0469 (0.0017)	0.0358 (0.0012)	0.0348 (0.0011)
D_4	0.0425 (0.0014)	0.0427 (0.0014)	0.0436 (0.0015)	0.0438 (0.0015)	0.0388 (0.0014)	0.0394 (0.0015)	0.0396 (0.0015)	0.0391 (0.0015)	0.0695 (0.0023)	0.0398 (0.0014)	0.0367 (0.0012)
D_5	0.0357 (0.0012)	0.0357 (0.0012)	0.0367 (0.0012)	0.0358 (0.0012)	0.0420 (0.0014)	0.0430 (0.0014)	0.0433 (0.0014)	0.0418 (0.0014)	0.0490 (0.0018)	0.0354 (0.0011)	0.0339 (0.0011)
D_6	0.0490 (0.0016)	0.0482 (0.0015)	0.0533 (0.0020)	0.0492 (0.0017)	0.0582 (0.0019)	0.0573 (0.0018)	0.0565 (0.0018)	0.0581 (0.0020)	0.0695 (0.0033)	0.0473 (0.0015)	0.0474 (0.0015)
D_7	0.0408 (0.0013)	0.0406 (0.0013)	0.0421 (0.0014)	0.0397 (0.0013)	0.0529 (0.0017)	0.0522 (0.0017)	0.0521 (0.0016)	0.0520 (0.0017)	0.0511 (0.0017)	0.0412 (0.0013)	0.0426 (0.0014)
D_8	0.0686 (0.0029)	0.0669 (0.0027)	0.0984 (0.0052)	0.0707 (0.0032)	0.0613 (0.0021)	0.0590 (0.0019)	0.0605 (0.0019)	0.0633 (0.0023)	0.1188 (0.0134)	0.0523 (0.0018)	0.0497 (0.0016)
D_9	0.0476 (0.0017)	0.0465 (0.0017)	0.0571 (0.0026)	0.0451 (0.0017)	0.0542 (0.0018)	0.0534 (0.0017)	0.0533 (0.0017)	0.0537 (0.0018)	0.0630 (0.0032)	0.0435 (0.0014)	0.0439 (0.0014)
D_{10}	0.3496 (0.0998)	0.1399 (0.0052)	0.1846 (0.0072)	0.1511 (0.0060)	0.0831 (0.0033)	0.0730 (0.0027)	0.0787 (0.0030)	0.0816 (0.0034)	0.3027 (0.0123)	0.0989 (0.0036)	0.0847 (0.0031)

(): desvio padrão do EQM de β_0

Tabela 8: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 40$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0348 (0.0011)	0.0449 (0.0014)	0.0283 (0.0009)	0.0300 (0.0009)	0.0502 (0.0016)	0.0590 (0.0019)	0.0412 (0.0013)	0.0430 (0.0014)	0.0323 (0.0010)	0.0302 (0.0010)	0.0316 (0.0010)
D_2	0.0346 (0.0011)	0.0416 (0.0014)	0.0304 (0.0010)	0.0321 (0.0010)	0.0464 (0.0015)	0.0553 (0.0018)	0.0396 (0.0013)	0.0402 (0.0013)	0.0375 (0.0012)	0.0300 (0.0010)	0.0313 (0.0010)
D_3	0.0351 (0.0012)	0.0396 (0.0013)	0.0323 (0.0011)	0.0339 (0.0011)	0.0451 (0.0015)	0.0525 (0.0018)	0.0391 (0.0013)	0.0398 (0.0013)	0.0418 (0.0014)	0.0308 (0.0010)	0.0323 (0.0010)
D_4	0.0370 (0.0013)	0.0388 (0.0014)	0.0430 (0.0015)	0.0439 (0.0015)	0.0386 (0.0015)	0.0428 (0.0017)	0.0382 (0.0015)	0.0383 (0.0015)	0.0634 (0.0021)	0.0355 (0.0013)	0.0353 (0.0013)
D_5	0.0346 (0.0012)	0.0380 (0.0013)	0.0321 (0.0011)	0.0345 (0.0011)	0.0432 (0.0014)	0.0499 (0.0017)	0.0374 (0.0013)	0.0388 (0.0013)	0.0435 (0.0014)	0.0305 (0.0010)	0.0317 (0.0010)
D_6	0.0504 (0.0017)	0.0482 (0.0015)	0.0488 (0.0019)	0.0553 (0.0022)	0.0605 (0.0019)	0.0623 (0.0019)	0.0533 (0.0017)	0.0606 (0.0020)	0.0657 (0.0028)	0.0449 (0.0015)	0.0439 (0.0014)
D_7	0.0454 (0.0015)	0.0466 (0.0015)	0.0387 (0.0013)	0.0431 (0.0015)	0.0583 (0.0018)	0.0614 (0.0019)	0.0494 (0.0016)	0.0544 (0.0018)	0.0512 (0.0019)	0.0397 (0.0013)	0.0391 (0.0012)
D_8	0.0719 (0.0031)	0.0535 (0.0020)	0.0810 (0.0047)	0.0845 (0.0060)	0.0675 (0.0023)	0.0621 (0.0020)	0.0530 (0.0016)	0.0669 (0.0023)	0.1049 (0.0080)	0.0510 (0.0019)	0.0469 (0.0015)
D_9	0.0596 (0.0024)	0.0469 (0.0016)	0.0484 (0.0024)	0.0530 (0.0026)	0.0657 (0.0023)	0.0612 (0.0020)	0.0500 (0.0016)	0.0585 (0.0020)	0.0660 (0.0043)	0.0420 (0.0014)	0.0406 (0.0013)
D_{10}	0.4837 (0.0587)	0.0926 (0.0034)	0.1658 (0.0072)	0.1975 (0.0101)	0.1254 (0.0060)	0.0780 (0.0029)	0.0828 (0.0033)	0.1013 (0.0044)	0.2790 (0.0110)	0.1002 (0.0038)	0.0755 (0.0028)

() : desvio padrão do EQM de β_1

Tabela 9: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0104 (0.0004)	0.0105 (0.0004)	0.0106 (0.0004)	0.0103 (0.0004)	0.0161 (0.0006)	0.0161 (0.0006)	0.0160 (0.0006)	0.0157 (0.0006)	0.0160 (0.0006)	0.0121 (0.0004)	0.0128 (0.0005)
D_2	0.0128 (0.0005)	0.0129 (0.0005)	0.0127 (0.0005)	0.0127 (0.0005)	0.0166 (0.0006)	0.0167 (0.0006)	0.0166 (0.0006)	0.0166 (0.0006)	0.0245 (0.0009)	0.0133 (0.0005)	0.0135 (0.0005)
D_3	0.0139 (0.0005)	0.0139 (0.0005)	0.0139 (0.0005)	0.0138 (0.0005)	0.0162 (0.0006)	0.0161 (0.0006)	0.0162 (0.0006)	0.0162 (0.0006)	0.0315 (0.0012)	0.0135 (0.0005)	0.0134 (0.0005)
D_4	0.0171 (0.0006)	0.0171 (0.0006)	0.0177 (0.0006)	0.0173 (0.0006)	0.0124 (0.0005)	0.0125 (0.0005)	0.0127 (0.0005)	0.0124 (0.0005)	0.0527 (0.0019)	0.0141 (0.0005)	0.0128 (0.0004)
D_5	0.0142 (0.0005)	0.0143 (0.0005)	0.0144 (0.0005)	0.0142 (0.0005)	0.0155 (0.0006)	0.0155 (0.0006)	0.0157 (0.0006)	0.0155 (0.0005)	0.0358 (0.0014)	0.0132 (0.0005)	0.0130 (0.0005)
D_6	0.0225 (0.0009)	0.0225 (0.0009)	0.0258 (0.0010)	0.0225 (0.0009)	0.0212 (0.0008)	0.0213 (0.0008)	0.0219 (0.0008)	0.0213 (0.0008)	0.0915 (0.0051)	0.0172 (0.0007)	0.0175 (0.0007)
D_7	0.0164 (0.0006)	0.0164 (0.0006)	0.0177 (0.0007)	0.0160 (0.0006)	0.0197 (0.0007)	0.0198 (0.0007)	0.0202 (0.0007)	0.0195 (0.0007)	0.0434 (0.0024)	0.0153 (0.0006)	0.0158 (0.0006)
D_8	0.0503 (0.0022)	0.0506 (0.0022)	0.0619 (0.0026)	0.0503 (0.0023)	0.0218 (0.0008)	0.0218 (0.0008)	0.0225 (0.0008)	0.0218 (0.0008)	0.4414 (0.0293)	0.0181 (0.0007)	0.0183 (0.0007)
D_9	0.0243 (0.0011)	0.0245 (0.0011)	0.0306 (0.0014)	0.0234 (0.0011)	0.0200 (0.0007)	0.0199 (0.0007)	0.0205 (0.0008)	0.0200 (0.0007)	0.1034 (0.0098)	0.0157 (0.0006)	0.0162 (0.0006)
D_{10}	0.2591 (0.1280)	0.1135 (0.0052)	0.1235 (0.0052)	0.1117 (0.0050)	0.0280 (0.0010)	0.0278 (0.0010)	0.0292 (0.0011)	0.0280 (0.0010)	0.8888 (0.0552)	0.0366 (0.0014)	0.0289 (0.0010)

():desvio padrão do EQM de β_0

Tabela 10: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.10$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0136 (0.0005)	0.0158 (0.0006)	0.0110 (0.0004)	0.0112 (0.0004)	0.0192 (0.0007)	0.0223 (0.0008)	0.0161 (0.0006)	0.0164 (0.0006)	0.0167 (0.0006)	0.0122 (0.0004)	0.0127 (0.0004)
D_2	0.0139 (0.0005)	0.0157 (0.0006)	0.0127 (0.0005)	0.0134 (0.0005)	0.0193 (0.0007)	0.0219 (0.0008)	0.0172 (0.0006)	0.0177 (0.0006)	0.0239 (0.0009)	0.0132 (0.0005)	0.0134 (0.0005)
D_3	0.0141 (0.0005)	0.0151 (0.0005)	0.0138 (0.0005)	0.0146 (0.0005)	0.0183 (0.0007)	0.0201 (0.0007)	0.0169 (0.0006)	0.0173 (0.0006)	0.0305 (0.0011)	0.0133 (0.0005)	0.0134 (0.0005)
D_4	0.0171 (0.0006)	0.0164 (0.0006)	0.0186 (0.0007)	0.0197 (0.0007)	0.0144 (0.0006)	0.0149 (0.0006)	0.0139 (0.0006)	0.0142 (0.0006)	0.0505 (0.0018)	0.0147 (0.0005)	0.0136 (0.0005)
D_5	0.0140 (0.0005)	0.0148 (0.0005)	0.0138 (0.0005)	0.0151 (0.0006)	0.0172 (0.0006)	0.0189 (0.0007)	0.0158 (0.0006)	0.0167 (0.0006)	0.0344 (0.0013)	0.0131 (0.0005)	0.0130 (0.0005)
D_6	0.0179 (0.0007)	0.0178 (0.0007)	0.0257 (0.0010)	0.0272 (0.0011)	0.0229 (0.0009)	0.0241 (0.0009)	0.0239 (0.0009)	0.0244 (0.0009)	0.0861 (0.0039)	0.0183 (0.0007)	0.0189 (0.0007)
D_7	0.0161 (0.0006)	0.0166 (0.0006)	0.0185 (0.0007)	0.0193 (0.0007)	0.0224 (0.0008)	0.0239 (0.0009)	0.0217 (0.0008)	0.0228 (0.0009)	0.0443 (0.0020)	0.0164 (0.0006)	0.0172 (0.0007)
D_8	0.0302 (0.0014)	0.0265 (0.0012)	0.0563 (0.0026)	0.0580 (0.0028)	0.0230 (0.0009)	0.0231 (0.0009)	0.0239 (0.0009)	0.0258 (0.0010)	0.3015 (0.0166)	0.0196 (0.0007)	0.0198 (0.0007)
D_9	0.0196 (0.0008)	0.0179 (0.0007)	0.0275 (0.0013)	0.0277 (0.0012)	0.0222 (0.0008)	0.0227 (0.0008)	0.0217 (0.0008)	0.0231 (0.0009)	0.0887 (0.0061)	0.0170 (0.0006)	0.0177 (0.0007)
D_{10}	0.2755 (0.1331)	0.0542 (0.0023)	0.0970 (0.0044)	0.1151 (0.0052)	0.0311 (0.0013)	0.0269 (0.0010)	0.0292 (0.0011)	0.0310 (0.0012)	0.7040 (0.0299)	0.0347 (0.0014)	0.0285 (0.0011)

():desvio padrão do EQM de β_1

Tabela 11: Estimativas dos Erros Quadráticos Médios de β_0 para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0108 (0.0004)	0.0110 (0.0004)	0.0109 (0.0004)	0.0108 (0.0004)	0.0159 (0.0006)	0.0160 (0.0006)	0.0158 (0.0006)	0.0159 (0.0006)	0.0130 (0.0005)	0.0121 (0.0004)	0.0128 (0.0005)
D_2	0.0128 (0.0005)	0.0130 (0.0005)	0.0127 (0.0005)	0.0126 (0.0005)	0.0169 (0.0006)	0.0168 (0.0006)	0.0167 (0.0006)	0.0167 (0.0006)	0.0162 (0.0006)	0.0133 (0.0005)	0.0135 (0.0005)
D_3	0.0134 (0.0005)	0.0136 (0.0005)	0.0134 (0.0005)	0.0132 (0.0005)	0.0164 (0.0006)	0.0162 (0.0006)	0.0162 (0.0006)	0.0163 (0.0006)	0.0179 (0.0007)	0.0135 (0.0005)	0.0134 (0.0005)
D_4	0.0156 (0.0005)	0.0156 (0.0005)	0.0156 (0.0005)	0.0156 (0.0005)	0.0125 (0.0005)	0.0126 (0.0005)	0.0124 (0.0005)	0.0124 (0.0005)	0.0264 (0.0010)	0.0141 (0.0005)	0.0128 (0.0004)
D_5	0.0134 (0.0005)	0.0136 (0.0005)	0.0134 (0.0005)	0.0133 (0.0005)	0.0157 (0.0006)	0.0155 (0.0006)	0.0156 (0.0006)	0.0156 (0.0006)	0.0185 (0.0007)	0.0132 (0.0005)	0.0130 (0.0005)
D_6	0.0173 (0.0007)	0.0174 (0.0007)	0.0177 (0.0007)	0.0171 (0.0007)	0.0218 (0.0008)	0.0216 (0.0008)	0.0218 (0.0008)	0.0223 (0.0010)	0.0224 (0.0009)	0.0172 (0.0007)	0.0175 (0.0007)
D_7	0.0145 (0.0006)	0.0147 (0.0006)	0.0148 (0.0006)	0.0145 (0.0006)	0.0198 (0.0007)	0.0199 (0.0007)	0.0202 (0.0008)	0.0198 (0.0007)	0.0177 (0.0007)	0.0153 (0.0006)	0.0158 (0.0006)
D_8	0.0207 (0.0009)	0.0205 (0.0009)	0.0235 (0.0011)	0.0199 (0.0008)	0.0225 (0.0008)	0.0222 (0.0008)	0.0223 (0.0008)	0.0251 (0.0017)	0.0264 (0.0011)	0.0181 (0.0007)	0.0183 (0.0007)
D_9	0.0159 (0.0007)	0.0156 (0.0006)	0.0162 (0.0007)	0.0154 (0.0006)	0.0205 (0.0008)	0.0208 (0.0010)	0.0203 (0.0008)	0.0206 (0.0010)	0.0195 (0.0008)	0.0157 (0.0006)	0.0162 (0.0006)
D_{10}	0.1226 (0.0254)	0.0554 (0.0022)	0.0625 (0.0025)	0.0554 (0.0022)	0.0296 (0.0011)	0.0280 (0.0010)	0.0291 (0.0011)	0.0296 (0.0015)	0.1051 (0.0043)	0.0366 (0.0014)	0.0289 (0.0010)

():desvio padrão do EQM de β_0

Tabela 12: Estimativas dos Erros Quadráticos Médios de β_1 para os diversos métodos de estimação ($n = 100$ e $\alpha = 0.20$).

Dist.	$A_{11}(\alpha)$	$A_{21}(\alpha)$	$A_{31}(\alpha)$	$A_{41}(\alpha)$	$A_{12}(\alpha)$	$A_{22}(\alpha)$	$A_{32}(\alpha)$	$A_{42}(\alpha)$	$A_4(\alpha)$	T	G
D_1	0.0137 (0.0005)	0.0188 (0.0007)	0.0113 (0.0004)	0.0120 (0.0004)	0.0199 (0.0007)	0.0254 (0.0009)	0.0162 (0.0006)	0.0171 (0.0006)	0.0133 (0.0005)	0.0122 (0.0004)	0.0127 (0.0004)
D_2	0.0141 (0.0005)	0.0179 (0.0007)	0.0119 (0.0004)	0.0127 (0.0005)	0.0198 (0.0007)	0.0244 (0.0009)	0.0168 (0.0006)	0.0171 (0.0006)	0.0151 (0.0006)	0.0132 (0.0005)	0.0134 (0.0005)
D_3	0.0142 (0.0005)	0.0171 (0.0006)	0.0127 (0.0005)	0.0134 (0.0005)	0.0189 (0.0007)	0.0225 (0.0008)	0.0165 (0.0006)	0.0167 (0.0006)	0.0167 (0.0006)	0.0133 (0.0005)	0.0134 (0.0005)
D_4	0.0155 (0.0006)	0.0157 (0.0006)	0.0162 (0.0006)	0.0181 (0.0006)	0.0148 (0.0006)	0.0162 (0.0006)	0.0137 (0.0005)	0.0145 (0.0006)	0.0263 (0.0010)	0.0147 (0.0005)	0.0136 (0.0005)
D_5	0.0140 (0.0005)	0.0161 (0.0006)	0.0128 (0.0005)	0.0135 (0.0005)	0.0180 (0.0007)	0.0207 (0.0008)	0.0160 (0.0006)	0.0161 (0.0006)	0.0173 (0.0007)	0.0131 (0.0005)	0.0130 (0.0005)
D_6	0.0221 (0.0008)	0.0217 (0.0008)	0.0191 (0.0007)	0.0212 (0.0008)	0.0274 (0.0010)	0.0294 (0.0011)	0.0240 (0.0009)	0.0258 (0.0011)	0.0236 (0.0009)	0.0183 (0.0007)	0.0189 (0.0007)
D_7	0.0205 (0.0008)	0.0215 (0.0008)	0.0155 (0.0006)	0.0175 (0.0006)	0.0262 (0.0010)	0.0291 (0.0011)	0.0214 (0.0008)	0.0227 (0.0009)	0.0189 (0.0007)	0.0164 (0.0006)	0.0172 (0.0007)
D_8	0.0291 (0.0010)	0.0205 (0.0007)	0.0234 (0.0011)	0.0279 (0.0011)	0.0302 (0.0011)	0.0280 (0.0010)	0.0245 (0.0009)	0.0318 (0.0018)	0.0275 (0.0010)	0.0196 (0.0007)	0.0198 (0.0007)
D_9	0.0293 (0.0011)	0.0207 (0.0008)	0.0167 (0.0006)	0.0199 (0.0008)	0.0318 (0.0012)	0.0294 (0.0012)	0.0223 (0.0008)	0.0248 (0.0011)	0.0208 (0.0008)	0.0170 (0.0006)	0.0177 (0.0007)
D_{10}	0.3013 (0.0400)	0.0316 (0.0013)	0.0522 (0.0022)	0.0657 (0.0030)	0.0563 (0.0024)	0.0286 (0.0011)	0.0290 (0.0011)	0.0344 (0.0017)	0.0947 (0.0040)	0.0347 (0.0014)	0.0285 (0.0011)

(): desvio padrão do EQM de β_1

Apêndice C

PROGRAMAS

```
/* Este programa utiliza como estimativa      */
/* inicial para beta0 e beta1 o metodo de      */
/* minimos quadrados e apos a remocao das      */
/* observacoes correspondentes aos maiores     */
/* e menores residuos, calcula-se a estimativa */
/* final de beta0 e beta1 tambem por minimos  */
/* quadrados.                                  */
```

```
OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE Beta0 E Beta1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=100;
*---NUMERO DE SIMULACOES = NS-----;
NS=1500;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*--PARA A EXPONENCIAL DUPLA-----;
  SEED3=SEED3+4;
  DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO;
  *   CALL RANNOR(SEED1,E);
  *-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
```

```

*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO DE TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO DE TUKEY=T(10)-----;
*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO DE TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;

```

```

*-----;
*9)-DISTRIBUICAO DE TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA---;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W = -LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO DE TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*   CALL RANCAU(SEED1,E);
*-----;
*   CALL RANNOR(SEED2,XX);
*   Y=B0 + B1*XX + E;
*   OUTPUT;KEEP J I E XX Y;
*   END;
END;
*--TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;

```

```

PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=100;
*-----NUMERO DE SIMULACOES-----;
NS=1500;
*-----PORCENTAGEM DE DADOS APARADOS-----;
  ALFA=0.10;
  PC=ALFA*100;
PRINT 'METODO A11';
PRINT 'ESTIMATIVA INICIAL : MQ';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : MQ';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*--DEFINIR QUAL A DISTRIBUICAO DOS ERROS--;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
  PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: B0=1 B1=1';
USE NEWARQ1;
READ ALL INTO MA;
*PRINT MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
*PRINT X Y;
*-----CALCULO DE BETAO POR MINIMOS QUADRADOS-----;
  BETAO=INV(X'*X)*X'*Y;
*-----CALCULO DOS RESIDUOS-----;

```

```

RESID = Y - X*BETA0;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
X1=X[,2];
RANKRES = RANK(RESID);
NEWRES = RESID[LOC(RANKRES),];RESORD = NEWRES;
NEWY = Y[LOC(RANKRES),];YORD = NEWY;
NEWX = X1[LOC(RANKRES),];XORD = NEWX;
RANKRES = RANKRES[LOC(RANKRES),];
RESORD[RANKRES,] = NEWRES;
YORD[RANKRES,] = NEWY;
XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
*-----CALCULO DA MATRIZ 'A'(OPERADOR LINEAR)-----;
N=NROW(RESORD);
D=SHAPE(1,N,1);
DO K=1 TO N;
    IF K<INT(N*(ALFA/2))+0.01 ] K>N-INT(N*(ALFA/2))
        THEN D[K,1]=0;
END;
*PRINT D;
A=DIAG(D);
*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
XNOVA = REPEAT(1,N,1)||XORD;
BETAF = INV(XNOVA'*A*XNOVA)*XNOVA'*A*YORD;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
YPREDITO = X*BETAF;
RESIDUO = Y - YPREDITO;
*PRINT X Y YPREDITO RESIDUO;
IF J=1 THEN BETAG=BETAF';
ELSE BETAG=BETAG//BETAF';
END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS ESTIMATIVAS';
*PRINT 'INICIAIS SAO CALCULADAS POR MINIMOS QUADRADOS';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;

```



```

QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B1A10012;
SET ARQBETA;
RUN;

```

```

/* Este programa utiliza como estimativa */
/* inicial para beta0 e beta1 L1 */
/* e apos a remocao das observacoes */
/* correspondentes aos maiores e menores */
/* residuos calcula-se a estimativa final */
/* de beta0 e beta1 por minimos quadrados */

```

```

OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=100;
*---NUMERO DE SIMULACOES = NS-----;
NS=1500;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*---PARA A EXPONENCIAL DUPLA-----;
SEED3=SEED3+4;
DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
* CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;

```

```

*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO TUKEY=T(10)-----;
*L2=0.0234;

```

```

*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W = -LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*   CALL RANCAU(SEED1,E);
*-----;
*   CALL RANNOR(SEED2,XX);
*   Y=B0 + B1*XX + E;
*   OUTPUT;KEEP J I E XX Y;
*   END;
*   END;
*---TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;

```

```

OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=100;
*-----NUMERO DE SIMULACOES-----;
NS=1500;
*--PORCENTAGEM DE DADOS APARADOS-----;
ALFA = 0.10;
PC=ALFA*100;
PRINT 'METODO A21';
PRINT 'ESTIMATIVA INICIAL : L1';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : MQ';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DEFINIR QUAL A DISTRIBUICAO DOS ERROS--;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: B0=1 B1=1';
USE NEWARQ1;

```

```

READ ALL INTO MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIABEL DEPENDENTE
* Y:          VARIABEL DEPENDENTE
* XNAME:      NOME DAS VARIABEIS INDEPENDENTES
* X:          VARIABEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*               $Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E$ 
*              ONDE  $E = (XB \leq Y)$ .
*              ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.0E10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(0,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (          Y'          || REPEAT(0,1,M) || { -1 0 }          ) //
    ( REPEAT(0,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
    (          COEF          || I(M)          || REPEAT(0,M,2) ) ;
BASIS = N+M+2 - (0:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R.,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || {0} ;
U = REPEAT(Q,1,N) || REPEAT(0,1,M) || { . 0 } ;

```

```

CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*----- REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
      'VARIABEL DEPENDENTE: ' YNAME ,
      'REGRESSAO QUANTILICA: ' Q ,
      'NUMERO DE OBSERVACOES: ' N ,
      'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
      VARIABLE B;
*PRINT X Y PREDICT ERROR;
FINISH;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.','Y',{'INTERCEPT' 'VAR.INDEP.'},X,B1,PRED,RESID,.5);
BETA0=B1;
*-----;
*-----CALCULO DOS RESIDUOS-----;
RESID = Y - X*BETA0;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
X1=X[,2];
RANKRES = RANK(RESID);
NEWRES = RESID[LOC(RANKRES),];RESORD = NEWRES;
NEWY = Y[LOC(RANKRES),];YORD = NEWY;
NEWX = X1[LOC(RANKRES),];XORD = NEWX;
RANKRES = RANKRES[LOC(RANKRES),];
RESORD[RANKRES,] = NEWRES;
YORD[RANKRES,] = NEWY;
XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
*-----CALCULO DA MATRIZ 'A'(OPERADOR LINEAR)-----;
N=NROW(RESORD);
D=SHAPE(1,N,1);
DO K=1 TO N;
  IF K<INT(N*(ALFA/2))+0.01 ] K>N-INT(N*(ALFA/2))
  THEN D[K,1]=0;
END;
*PRINT D;

```

```

A=DIAG(D);
*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
XNOVA = REPEAT(1,N,1)||XORD;
BETAF = INV(XNOVA'*A*XNOVA)*XNOVA'*A*YORD;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
YPREDITO = X*BETAF;
RESIDUO = Y - YPREDITO;
*PRINT X Y YPREDITO RESIDUO;
IF J=1 THEN BETAG=BETAF';
ELSE BETAG=BETAG//BETAF';
END;
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B2A10012;
SET ARQBETA;
RUN;

```

```

/* Este programa apara as observacoes a */
/* partir das regressoes quantilicas de */
/* ordem alpha/2 e (1-alpha/2), depois */
/* calcula-se a estimativa final de beta0 e */
/* beta1 por minimos quadrados. */

```

```

OPTIONS LS=75 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;

```

```

*-----TAMANHO DA AMOSTRA = N-----;
N=100;
*----NUMERO DE SIMULACOES = NS-----;
NS=1500;
*-----GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
    SEED2=56879325;
    SEED1=SEED1+2;
*----PARA A EXPONENCIAL DUPLA-----;
    SEED3=SEED3+4;
    DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
        CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
* DP=10;
* ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*DP=10;
*ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*DP=5;
*ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;

```



```

*5)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*DP=5;
*ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO DE TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO DE TUKEY=T(10)-----;
*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO DE TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO DE TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W=-LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO DE TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;

```

```

*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO-----;
*   CALL RANCAU(SEED1,E);
*-----;
      CALL RANNOR(SEED2,XX);
      Y=BO + B1*XX + E;
      OUTPUT;
      KEEP J I E XX Y;
      END;
END;
*PROC PRINT;
*VAR J I E XX Y;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
PROC DELETE DATA=ARQ1;
*---ENTRADA NO SAS/IML---;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=100;
*-----NUMERO DE SIMULACOES-----;
NS=1500;
*--PORCENTAGEM DE DADOS APARADOS-----;
ALFA=0.10;
PC=ALFA*100;
PRINT 'METODO A31';

```

```

PRINT 'DADOS APARADOS A PARTIR DAS REGRESSOES QUANTILICAS';
PRINT 'ESTIMATIVAS FINAIS CALCULADAS POR : MQ';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DISTRIBUICAO DOS ERROS-----;
  PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
*PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY-PADRAO';
PRINT 'PARAMETROS INICIAIS BO=1 B1=1';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
USE NEWARQ1;
READ ALL INTO MA;
*PRINT MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
*PRINT X Y;
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA----
* YNAME:      NOME DA VARIABEL DEPENDENTE
* Y:          VARIABEL DEPENDENTE
* XNAME:      NOME DAS VARIABEIS INDEPENDENTES
* X:          VARIABEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*               $q * (y - Xb) * e + (1-q) * (y - Xb) * \hat{e}$ 
*              ONDE  $E = (XB \leq Y)$ . ONDE  $Q = .5$  QUE E EQUIVALENTE A
*              MINIMIZAR A SOMA DOS DESVIOS ABSOLUTOS.
*              ESTA SUBROTINA SEQUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.

```

```

-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
*PRINT M N;
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(O,M+2,1);
*PRINT R;
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
*PRINT L;
*PRINT U;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = ( Y' || REPEAT(O,1,M) || { -1 0 } ) //
( REPEAT(O,1,N) || REPEAT(-1,1,M) || { 0 -1 } ) //
( COEF || I(M) || REPEAT(O,M,2) );
*PRINT A;
BASIS = N+M+2 - (O:N+M+1);
*PRINT BASIS;
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,.,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || {O} ;
U = REPEAT(Q,1,N) || REPEAT(O,1,M) || { . 0 } ;
*PRINT L;
*PRINT U;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*-----REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
'VARIABEL DEPENDENTE: ', YNAME ,
'REGRESSAO QUANTILICA: ' Q ,
'NUMERO DE OBSERVACOES: ' N ,
'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
VARIABLE B;
*PRINT X Y PREDICT ERROR;

```

```

FINISH;
Q1=ALFA/2;
Q2=1-Q1;
*-----PORCENTAGEM DE DADOS APARADOS-----;
ALFA=0.10;
Q1=ALFA/2;
Q2=1-ALFA/2;
*----EXECUTAR O PROGRAMA DE REGRESSAO QUANTILICA---;
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B1,PRED1,RES1,Q1);
*----EXECUTAR O PROGRAMA DE REGRESSAO QUANTILICA---;
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B2,PRED2,RES2,Q2);
*-----CALCULO DA MATRIZ 'A'(OPERADOR LINEAR)-----;
D=SHAPE(1,N,1);
DO K=1 TO N;
  IF RES1[K,1]<=0 ] RES2[K,1]>=0
  THEN D[K,1]=0;
END;
*D1=D';
*PRINT D1;
A=DIAG(D);
*--CALCULO DA SEGUNDA ESTIMATIVA DO VETOR DE PARAMETROS--;
BETAF=INV(X'*A*X)*X'*A*Y;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
*YPREDITO=X*BETAF;
*RESIDUO=Y-YPREDITO;
*PRINT X Y YPREDITO RESIDUO;
  IF J=1 THEN BETAG=BETAF';
  ELSE BETAG=BETAG//BETAF';
END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS';
*PRINT ' OBSERVACOES SAO APARADAS DE ACORDO COM';
*PRINT 'AS REGRESSOES QUANTILICAS .05 E .95';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*-----ENTRADA NO SAS-----;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;

```

```

PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B3A1001;
SET ARQBETA;
RUN;

```

```

/* este programa utiliza como estimativa */
/* inicial para o vetor de parametros BETA */
/* a media aritmetica dos estimadores das */
/* regressoes quantilicas de ordem alpha/2 */
/* e (1-alpha/2). Apos a remocao das */
/* observacoes correspondentes aos maiores */
/* e menores residuos, calcula-se a */
/* estimatica final de BETA por minimos */
/* quadrados. */

```

```

OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=100;
*---NUMERO DE SIMULACOES = NS-----;
NS=1500;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*---PARA A EXPONENCIAL DUPLA-----;
SEED3=SEED3+4;
DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
* CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;

```

```

*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO TUKEY=T(10)-----;
*L2=0.0234;

```

```

*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W = -LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*   CALL RANCAU(SEED1,E);
*-----;
*   CALL RANNOR(SEED2,XX);
*   Y=B0 + B1*XX + E;
*   OUTPUT;KEEP J I E XX Y;
*   END;
*   END;
*---TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;

```



```

OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=100;
*-----NUMERO DE SIMULACOES-----;
NS=1500;
*--PORCENTAGEM DE DADOS APARADOS-----;
ALFA = 0.10;
PC=ALFA*100;
PRINT 'METODO A41';
PRINT 'ESTIMATIVA INICIAL : BO= (B(ALFA/2)-B(1-ALFA/2))/2';
PRINT 'DADOS APARADOS A PARTIR DOS RESIDUOS DE BO';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : MQ';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DEFINIR QUAL A DISTRIBUICAO DOS ERROS--;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: BO=1 B1=1';

```

```

USE NEWARQ1;
READ ALL INTO MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIABEL DEPENDENTE
* Y:          VARIABEL DEPENDENTE
* XNAME:      NOME DAS VARIABEIS INDEPENDENTES
* X:          VARIABEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*               $Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E$ 
*              ONDE  $E = (XB \leq Y)$ .
*              ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(0,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (          Y'          || REPEAT(0,1,M) || { -1 0 }          ) //
    ( REPEAT(0,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
    (          COEF          ||          I(M)          || REPEAT(0,M,2)          ) ;
BASIS = N+M+2 - (0:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || {0} ;

```

```

U = REPEAT(Q,1,N) || REPEAT(O,1,M) || { . O } ;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*----- REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT , 'REGRESSAO QUANTILICA ESTIMADA' ,
        'VARIABEL DEPENDENTE: ' YNAME ,
        'REGRESSAO QUANTILICA: ' Q ,
        'NUMERO DE OBSERVACOES: ' N ,
        'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
        VARIABLE B;
*PRINT X Y PREDICT ERROR;
FINISH;
Q1=ALFA/2;
Q2=1-Q1;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B1,PRED,RESID,Q1);
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B2,PRED,RESID,Q2);
BETAO=(B1 + B2)/2;
*PRINT BETAO;
*-----;
*-----CALCULO DOS RESIDUOS-----;
RESID = Y - X*BETAO;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
X1=X[,2];
RANKRES = RANK(RESID);
NEWRES = RESID[LOC(RANKRES),]; RESORD = NEWRES;
NEWY = Y[LOC(RANKRES),]; YORD = NEWY;
NEWX = X1[LOC(RANKRES),]; XORD = NEWX;
RANKRES = RANKRES[LOC(RANKRES),];
RESORD[RANKRES,] = NEWRES;
YORD[RANKRES,] = NEWY;
XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
*-----CALCULO DA MATRIZ 'A'(OPERADOR LINEAR)-----;
N=NROW(RESORD);
D=SHAPE(1,N,1);

```

```

DO K=1 TO N;
  IF K<INT(N*(ALFA/2))+0.01 ] K>N-INT(N*(ALFA/2))
    THEN D[K,1]=0;
  END;
*PRINT D;
A=DIAG(D);
*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
XNOVA = REPEAT(1,N,1)||XORD;
BETAF = INV(XNOVA'*A*XNOVA)*XNOVA'*A*YORD;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
YPREDITO = X*BETAF;
RESIDUO = Y - YPREDITO;
*PRINT X Y YPREDITO RESIDUO;
IF J=1 THEN BETAG=BETAF';
ELSE BETAG=BETAG//BETAF';
END;
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B4A10012;
SET ARQBETA;
RUN;

```

```

/* Este programa calcula a estimativa      */
/* inicial do vetor de parametros BETA      */
/* por minimos quadrados e apos a remocao  */
/* das observacoes correspondentes aos      */
/* maiores e menores residuos, calcula-se */

```

/* a estimativa final de BETA por L1. */

```
OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=20;
*---NUMERO DE SIMULACOES = NS-----;
NS=2000;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*--PARA A EXPONENCIAL DUPLA-----;
  SEED3=SEED3+4;
  DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO;
*  CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*  DP=10;
*  ALFA1=0.15;
*  CALL RANUNI(SEED1,U);
*  CALL RANNOR(SEED1,Z);
*  IF U>ALFA1 THEN E=Z;
*  ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*  DP=10;
*  ALFA1=0.10;
*  CALL RANUNI(SEED1,U);
*  CALL RANNOR(SEED1,Z);
*  IF U>ALFA1 THEN E=Z;
*  ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*  DP=5;
```

```

*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO DE TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO DE TUKEY=T(10)-----;
*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO DE TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO DE TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA---;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W = -LOG(U1);

```

```

*      IF U2>=0.5 THEN E=W;
*      ELSE E=-W;
*-----;
*11)-DISTRIBUICAO DE TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*      CALL RANUNI(SEED1,U);
*      E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*      CALL RANCAU(SEED1,E);
*-----;
*      CALL RANNOR(SEED2,XX);
*      Y=B0 + B1*XX + E;
*      OUTPUT;KEEP J I E XX Y;
*      END;
END;
*--TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=20;
*-----NUMERO DE SIMULACOES-----;

```

```

NS=2000;
*-----PORCENTAGEM DE DADOS APARADOS-----;
  ALFA=0.10;
  PC=ALFA*100;
PRINT 'METODO A12';
PRINT 'ESTIMATIVA INICIAL : MQ';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : L1';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*--DEFINIR QUAL A DISTRIBUICAO DOS ERROS---;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: B0=1 B1=1';
USE NEWARQ1;
READ ALL INTO MA;
*PRINT MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
*PRINT X Y;
*-----CALCULO DE BETAO POR MINIMOS QUADRADOS-----;
  BETAO=INV(X'*X)*X'*Y;
*-----CALCULO DOS RESIDUOS-----;
  RESID = Y - X*BETAO;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
  X1=X[,2];
  RANKRES = RANK(RESID);
  NEWRES = RESID[LOC(RANKRES),];RESORD = NEWRES;
  NEWY = Y[LOC(RANKRES),];YORD = NEWY;
  NEWX = X1[LOC(RANKRES),];XORD = NEWX;
  RANKRES = RANKRES[LOC(RANKRES),];
  RESORD[RANKRES,] = NEWRES;
  YORD[RANKRES,] = NEWY;

```



```

XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
LI = INT(N*ALFA/2) + 1;
LS = N - INT(N*ALFA/2);
XX = XORD[LI:LS,1];
YY = YORD[LI:LS,1];
*PRINT XX;
*PRINT YY;
N1 = NROW(XX);
*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
XNOVA = REPEAT(1,N1,1)||XX;
*PRINT XNOVA;
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIABEL DEPENDENTE
* Y:          VARIABEL DEPENDENTE
* XNAME:      NOME DAS VARIABEIS INDEPENDENTES
* X:          VARIABEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*               $Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E$ 
*              ONDE  $E = (XB \leq Y)$ .
*              ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(0,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/

```

```

A = (          Y'          || REPEAT(0,1,M)  || { -1 0 }          ) //
      ( REPEAT(0,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
      (          COEF          ||          I(M)          || REPEAT(0,M,2)          ) ;
BASIS = N+M+2 - (0:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M)  || -BOUND || {0} ;
U = REPEAT(Q,1,N)    || REPEAT(0,1,M)  || { . 0 } ;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*-----REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
      'VARIABEL DEPENDENTE: ' YNAME ,
      'REGRESSAO QUANTILICA: ' Q ,
      'NUMERO DE OBSERVACOES: ' N ,
      'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
      VARIABLE B;
*PRINT X Y PREDICT ERROR;
FINISH;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.',YY,{'INTERCEPT' 'VAR.INDEP.'},XNOVA,B2,PRED,RESID,.5);
*-----;
BETAF = B2;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
YPREDITO = XNOVA*BETAF;
RESIDUO = YY - YPREDITO;
*PRINT XNOVA YY YPREDITO RESIDUO;
IF J=1 THEN BETAG=BETAF';
ELSE BETAG=BETAG//BETAF';
END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS ESTIMATIVAS';
*PRINT 'INICIAIS SAO CALCULADAS POR MINIMOS QUADRADOS';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];

```

```

APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B6A2012;
SET ARQBETA;
RUN;

```

```

/* Este programa calcula as estimativas de beta0 e beta1 */
/* aparando as observacoes a partir dos residuos de uma */
/* estimativa preliminar: L1, tendo como estimativa      */
/* final tambem L1.                                       */

```

```

OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=20;
*---NUMERO DE SIMULACOES = NS-----;
NS=2000;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*--PARA A EXPONENCIAL DUPLA-----;
  SEED3=SEED3+4;
  DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO;
*  CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;

```

```

*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO DE TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO DE TUKEY=T(10)-----;
*L2=0.0234;

```

```

*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO DE TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO DE TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA---;
*       CALL RANUNI(SEED1,U1);
*       CALL RANUNI(SEED3,U2);
*       W = -LOG(U1);
*       IF U2>=0.5 THEN E=W;
*       ELSE E=-W;
*-----;
*11)-DISTRIBUICAO DE TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*   CALL RANCAU(SEED1,E);
*-----;
*       CALL RANNOR(SEED2,XX);
*       Y=B0 + B1*XX + E;
*       OUTPUT;KEEP J I E XX Y;
*       END;
END;
*--TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;

```

```

OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=20;
*-----NUMERO DE SIMULACOES-----;
NS=2000;
*-----PORCENTAGEM DE DADOS APARADOS-----;
  ALFA=0.10;
  PC=ALFA*100;
PRINT 'METODO 7';
PRINT 'ESTIMATIVA INICIAL : L1';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : L1';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*--DEFINIR QUAL A DISTRIBUICAO DOS ERROS---;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
  PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
*PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: B0=1 B1=1';
USE NEWARQ1;

```

```

READ ALL INTO MA;
*PRINT MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
*PRINT X Y;
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIAVEL DEPENDENTE
* Y:          VARIAVEL DEPENDENTE
* XNAME:      NOME DAS VARIAVEIS INDEPENDENTES
* X:          VARIAVEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*               $Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E$ 
*              ONDE  $E = (XB \leq Y)$ .
*              ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(0,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (      Y'      || REPEAT(0,1,M) || { -1 0 }      ) //
    ( REPEAT(0,1,N) || REPEAT(-1,1,M) || { 0 -1 }      ) //
    (      COEF      ||      I(M)      || REPEAT(0,M,2)  ) ;
BASIS = N+M+2 - (0:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R.,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/

```

```

L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || {O} ;
U = REPEAT(Q,1,N) || REPEAT(O,1,M) || { . O } ;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*----- REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<O , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
      'VARIABEL DEPENDENTE: ' YNAME ,
      'REGRESSAO QUANTILICA: ' Q ,
      'NUMERO DE OBSERVACOES: ' N ,
      'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
      VARIABLE B;
*PRINT X Y PREDICT ERROR;
FINISH;
*-----CALCULO DE BETAO POR L1-----;
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,BO,PRED,RESID,.5);
BETAO=BO;
IF J=1 THEN BETAI=BETAO';
ELSE BETAI=BETAI//BETAO';
*-----CALCULO DOS RESIDUOS-----;
RESID = Y - X*BETAO;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
X1=X[,2];
RANKRES = RANK(RESID);
NEWRES = RESID[LOC(RANKRES),];RESORD = NEWRES;
NEWY = Y[LOC(RANKRES),];YORD = NEWY;
NEWX = X1[LOC(RANKRES),];XORD = NEWX;
RANKRES = RANKRES[LOC(RANKRES),];
RESORD[RANKRES,] = NEWRES;
YORD[RANKRES,] = NEWY;
XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
LI = INT(N*ALFA/2) + 1;
LS = N - INT(N*ALFA/2);
XX = XORD[LI:LS,1];
YY = YORD[LI:LS,1];
*PRINT XX;

```



```

*PRINT YY;
N1 = NROW(XX);
XNOVA = REPEAT(1,N1,1)||XX;
*PRINT XNOVA;

*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.',YY,{'INTERCEPT' 'VAR.INDEP.'},XNOVA,B2,PRED,RESID,.5);
*-----;

BETAF = B2;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
YPREDITO = XNOVA*BETAF;
RESIDUO = YY - YPREDITO;
*PRINT XNOVA YY YPREDITO RESIDUO;
IF J=1 THEN BETAG=BETAF';
ELSE BETAG=BETAG//BETAF';
END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS ESTIMATIVAS';
*PRINT 'INICIAIS SAO CALCULADAS POR L1';
*PRINT 'VALORES DAS ESTIMATIVAS INICIAIS';
*PRINT BETAI;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
PROC PRINT DATA=ARQBETA;
VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B7A2010;
SET ARQBETA;
RUN;

```

```

/* Este programa apara as observacoes a      */
/* partir das regressoes quantilicas de      */
/* ordem alpha/2 e (1-alpha/2), depois a      */
/* estimativa final do vetor de parametros */
/* BETA e calculada por L1.                  */

OPTIONS LS=75 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*-----TAMANHO DA AMOSTRA = N-----;
N=100;
*----NUMERO DE SIMULACOES = NS-----;
NS=1500;
*-----GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
    SEED2=56879325;
    SEED1=SEED1+2;
*----PARA A EXPONENCIAL DUPLA-----;
    SEED3=SEED3+4;
    DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
        CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
        DP=10;
        ALFA1=0.15;

        CALL RANUNI(SEED1,U);
        CALL RANNOR(SEED1,Z);
        IF U>ALFA1 THEN E=Z;
        ELSE E=Z*DP;

*-----;

*3)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
        DP=10;
        ALFA1=0.10;

```

```

*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;

*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*DP=5;
*ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA--;
*---DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*DP=5;
*ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO DE TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO DE TUKEY=T(10)-----;
*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO DE TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO DE TUKEY=T(6)-----;

```

```

*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W=-LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO DE TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3-(1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO-----;
*   CALL RANCAU(SEED1,E);
*-----;
*   CALL RANNOR(SEED2,XX);
*   Y=BO + B1*XX + E;
*   OUTPUT;
*   KEEP J I E XX Y;
*   END;
END;
*PROC PRINT;
*VAR J I E XX Y;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;

```

```

PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
PROC DELETE DATA=ARQ1;
*---ENTRADA NO SAS/IML---;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=100;
*-----NUMERO DE SIMULACOES-----;
NS=1500;
*--PORCENTAGEM DE DADOS APARADOS-----;
ALFA=0.10;
PC=ALFA*100;
PRINT 'METODO A32';
PRINT 'DADOS APARADOS A PARTIR DAS REGRESSOES QUANTILICAS';
PRINT 'PORCENTAGEM DE DADOS APARADOS' PC;
PRINT 'ESTIMATIVAS FINAIS CALCULADAS POR : L1';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DISTRIBUICAO DOS ERROS-----;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY-PADRAO';
PRINT 'PARAMETROS INICIAIS BO=1 B1=1';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
USE NEWARQ1;
READ ALL INTO MA;
*PRINT MA;
DO J=1 TO NS;
Y=MA[(J-1)*N+1:J*N,2];
X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
*PRINT X Y;
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIÁVEL DEPENDENTE
* Y:          VARIÁVEL DEPENDENTE

```

```

* XNAME:    NOME DAS VARIÁVEIS INDEPENDENTES
* X:        VARIÁVEIS INDEPENDENTES
* B:        ESTIMATIVAS
* PREDICT:  VALORES PREDITOS
* ERROR:    DIFERENÇA ENTRE Y E VALOR PREDITO
* Q:        QUANTIL
* NOTAS:    ESTA SUBRTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*            $q * (y - Xb) * e + (1-q) * (y - Xb) * ^e$ 
*           ONDE  $E = (XB \leq Y)$ . ONDE  $Q = .5$  QUE É EQUIVALENTE A
*           MINIMIZAR A SOMA DOS DESVIOS ABSOLUTOS.
*           ESTA SUBROTINA SEQUE AS DIREÇÕES DADAS EM:
*           KOENKER, R. AND G. BASSETT (1978). REGRESSION
*           QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*           BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*           QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*           JASA. VOL. 77. NO. 378. 407-415.
-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
*PRINT M N;
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(0,M+2,1);
*PRINT R;
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
*PRINT L;
*PRINT U;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (          Y'          || REPEAT(0,1,M) || { -1 0 }          ) //
    ( REPEAT(0,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
    (          COEF          ||          I(M)          || REPEAT(0,M,2) ) ;
*PRINT A;
BASIS = N+M+2 - (0:N+M+1);
*PRINT BASIS;
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(0,1,M) || -BOUND || {0} ;
U = REPEAT(Q,1,N) || REPEAT(0,1,M) || { . 0 } ;

```

```

*PRINT L;
*PRINT U;
  CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*----- REPORT THE SOLUTION-----*/
  VARIABLE = XNAME'; B=D[3:M+2];
  PREDICT = X*B;
  ERROR = Y - PREDICT;
  WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,, 'REGRESSAO QUANTILICA ESTIMADA' ,
  'VARIABEL DEPENDENTE: ', YNAME ,
  'REGRESSAO QUANTILICA: ' Q ,
  'NUMERO DE OBSERVACOES: ' N ,
  'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
  VARIABLE B;
*PRINT X Y PREDICT ERROR;
  FINISH;
Q1=ALFA/2;
Q2=1-Q1;
*-----PORCENTAGEM DE DADOS APARADOS-----;
ALFA=0.10;
Q1=ALFA/2;
Q2=1-ALFA/2;
*-----EXECUTAR O PROGRAMA DE REGRESSAO QUANTILICA---;
RUN RQ('VAR.DEP.',Y,{'INTERCEPT' 'VAR.INDEP.'},X,B1,PRED1,RES1,Q1);
*-----EXECUTAR O PROGRAMA DE REGRESSAO QUANTILICA---;
RUN RQ('VAR.DEP.',Y,{'INTERCEPT' 'VAR.INDEP.'},X,B2,PRED2,RES2,Q2);
*-----CALCULO DA MATRIZ 'A'(OPERADOR LINEAR)-----;
D=SHAPE(1,N,1);
DO K=1 TO N;
  IF RES1[K,1]<=0 ] RES2[K,1]>=0
    THEN D[K,1]=0;
END;
*D1=D';
*PRINT D1;
*PRINT D X Y;
NZERO = 0;
DO H=1 TO N;
  IF D[H,1]=0 THEN NZERO = NZERO + 1;
END;
*PRINT NZERO;
D2=D[LOC(D),];
*PRINT D2;

```

```

X1 = X[,2];
RANKD = RANK(D);
*PRINT RANKD;
NEWD = D[LOC(RANKD),];DORD=NEWD;
NEWY = Y[LOC(RANKD),];YORD=NEWY;
NEWX = X1[LOC(RANKD),];XORD=NEWX;
RANKD = RANKD[LOC(RANKD),];
DORD[RANKD,] = NEWD;
YORD[RANKD,] = NEWY;
XORD[RANKD,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT DORD XORD YORD;
N2 = NROW(D2);
*PRINT N2;
*PRINT D2;
YNOVO = YORD[NZERO + 1:N,1];
XNOVA = REPEAT(1,N2,1)||XORD[NZERO + 1:N,1];
*PRINT D2 XNOVA YNOVO;
XN=XNOVA;
YN=YNOVO;
*--CALCULO DA SEGUNDA ESTIMATIVA DO VETOR DE PARAMETROS--;
*----EXECUTAR O PROGRAMA DE REGRESSAO QUANTILICA----;
RUN RQ('VAR.DEP.',YN,{'INTERCEPT' 'VAR.INDEP.'},XN,BF,PRED,RES,.5);
*-----;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
BETAF = BF;
*-----CALCULO DOS VALORES PREDITOS-----;
*YPREDITO=XN*BETAF;
*RESIDUO=YN-YPREDITO;
*PRINT XN YN YPREDITO RESIDUO;
  IF J=1 THEN BETAG=BETAF';
    ELSE BETAG=BETAG//BETAF';
  END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS';
*PRINT ' OBSERVACOES SAO APARADAS DE ACORDO COM';
*PRINT 'AS REGRESSOES QUANTILICAS .05 E .95';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;

```



```

*-----ENTRADA NO SAS-----;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B8A10012;
SET ARQBETA;
RUN;

```

```

/* Este Programa utiliza como estimativa */
/* inicial para o vetor de parametros BETA */
/* a media aritmetica dos estimadores das */
/* regressoes quantilicas de ordem alpha/2 */
/* e (1-alpha/2). Apos a remocao das */
/* observacoes correspondentes aos maiores */
/* e menores residuos, a estimativa final */
/* de BETA e calculada por L1. */

```

```

OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=20;
*---NUMERO DE SIMULACOES = NS-----;
NS=2000;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*---PARA A EXPONENCIAL DUPLA-----;
SEED3=SEED3+4;
DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
* CALL RANNOR(SEED1,E);
*-----;

```

```

*2)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=10;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO TUKEY=T(10)-----;

```

```

*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*   CALL RANUNI(SEED1,U1);
*   CALL RANUNI(SEED3,U2);
*   W = -LOG(U1);
*   IF U2>=0.5 THEN E=W;
*   ELSE E=-W;
*-----;
*11)-DISTRIBUICAO TUKEY=T(5)-----;
*   L2=-0.2480;
*   L3=-0.1358;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*   CALL RANCAU(SEED1,E);
*-----;
*   CALL RANNOR(SEED2,XX);
*   Y=B0 + B1*XX + E;
*   OUTPUT;KEEP J I E XX Y;
*   END;
END;
*---TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;

```

```

PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;
*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=20;
*-----NUMERO DE SIMULACOES-----;
NS=2000;
*--PORCENTAGEM DE DADOS APARADOS-----;
ALFA = 0.10;
PC=ALFA*100;
PRINT 'METODO A42';
PRINT 'ESTIMATIVA INICIAL : BO= (B(ALFA)-B(1-ALFA))/2';
PRINT 'PORCENTAGEM DE DADOS APARADOS:' PC;
PRINT 'ESTIMATIVA FINAL : L1';
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DEFINIR QUAL A DISTRIBUICAO DOS ERROS--;
*PRINT 'DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25) + 0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25) + 0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
*PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: BO=1 B1=1';

```

```

USE NEWARQ1;
READ ALL INTO MA;
DO J=1 TO NS;
  Y=MA[(J-1)*N+1:J*N,2];
  X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME:      NOME DA VARIABEL DEPENDENTE
* Y:          VARIABEL DEPENDENTE
* XNAME:      NOME DAS VARIABEIS INDEPENDENTES
* X:          VARIABEIS INDEPENDENTES
* B:          ESTIMATIVAS
* PREDICT:    VALORES PREDITOS
* ERROR:      DIFERENCA ENTRE Y E VALOR PREDITO
* Q:          QUANTIL
* NOTAS:      ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM
*              
$$Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E$$

*              ONDE  $E = (XB \leq Y)$ .
*              ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*              KOENKER, R. AND G. BASSETT (1978). REGRESSION
*              QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*              BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*              QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*              JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(O,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (          Y'          || REPEAT(O,1,M) || { -1 0 }          ) //
    ( REPEAT(O,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
    (          COEF          ||          I(M)          || REPEAT(O,M,2)          ) ;
BASIS = N+M+2 - (O:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || {0} ;

```

```

U = REPEAT(Q,1,N) || REPEAT(O,1,M) || { . O } ;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*----- REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<0 , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
      'VARIABEL DEPENDENTE: ' YNAME ,
      'REGRESSAO QUANTILICA: ' Q ,
      'NUMERO DE OBSERVACOES: ' N ,
      'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
      VARIABLE B;
*PRINT X Y PREDICT ERROR;
FINISH;
Q1=ALFA/2;
Q2=1-Q1;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.','Y',{'INTERCEPT' 'VAR.INDEP.'},X,B1,PRED,RESID,Q1);
RUN RQ('VAR.DEP.','Y',{'INTERCEPT' 'VAR.INDEP.'},X,B2,PRED,RESID,Q2);
BETAO=(B1 + B2)/2;
*PRINT BETAO;
*-----;
*-----CALCULO DOS RESIDUOS-----;
RESID = Y - X*BETAO;
*-----;
*-----ORDENACAO DOS RESIDUOS-----;
X1=X[,2];
RANKRES = RANK(RESID);
NEWRES = RESID[LOC(RANKRES),];RESORD = NEWRES;
NEWY = Y[LOC(RANKRES),];YORD = NEWY;
NEWX = X1[LOC(RANKRES),];XORD = NEWX;
RANKRES = RANKRES[LOC(RANKRES),];
RESORD[RANKRES,] = NEWRES;
YORD[RANKRES,] = NEWY;
XORD[RANKRES,] = NEWX;
*PRINT 'VETORES ORDENADOS';
*PRINT RESORD XORD YORD;
*-----COMO APARAR AS OBSERVACOES-----;
LI = INT(N*ALFA/2) + 1;
LS = N - INT(N*ALFA/2);
XX = XORD[LI:LS,1];

```

```

YY = YORD[LI:LS,1];
*PRINT XX;
*PRINT YY;
N1 = NROW(XX);
  XNOVA = REPEAT(1,N1,1)||XX;
*PRINT XNOVA;
*-----CALCULO DA SEGUNDA ESTIMATIVA DE BETA-----;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.',YY,{'INTERCEPT' 'VAR.INDEP.'},XNOVA,B2,PRED,RESID,.5);
*-----;
  BETAF = B2;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
  YPREDITO = XNOVA*BETAF;
  RESIDUO = YY - YPREDITO;
*PRINT XNOVA YY YPREDITO RESIDUO;
  IF J=1 THEN BETAG=BETAF';
  ELSE BETAG=BETAG//BETAF';
END;
*PRINT 'VALORES DAS ESTIMATIVAS FINAIS QUANDO AS ESTIMATIVAS';
*PRINT 'INICIAIS SAO CALCULADAS POR MINIMOS QUADRADOS';
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B9A2011;
SET ARQBETA;
RUN;

```

/* Este programa calcula a estimativa do */

```

/* vetor de parametros BETA como sendo */
/* a media de GASTWIRTH, isto e, a media */
/* ponderada dos estimadores das regressoes */
/* quantilicas de ordem 1/3, 1/2 e 2/3 com */
/* pesos respectivamente iguais a 0.3, 0.4 */
/* e 0.3. */

OPTIONS LS=72 PS=60;
DATA ARQ1;
*-----VALORES DE BO E B1-----;
BO=1;B1=1;
*---TAMANHO DA AMOSTRA = N-----;
N=20;
*---NUMERO DE SIMULACOES = NS-----;
NS=2000;
*---GERACAO DOS ERROS-----;
SEED1=346789391;
SEED3=787534995;
DO J=1 TO NS;
SEED2=56879325;
SEED1=SEED1+2;
*---PARA A EXPONENCIAL DUPLA-----;
    SEED3=SEED3+4;
    DO I=1 TO N;
*1)-DISTRIBUICAO NORMAL PADRAO-----;
*    CALL RANNOR(SEED1,E);
*-----;
*2)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*    DP=10;
*    ALFA1=0.15;
*    CALL RANUNI(SEED1,U);
*    CALL RANNOR(SEED1,Z);
*    IF U>ALFA1 THEN E=Z;
*    ELSE E=Z*DP;
*-----;
*3)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
    DP=10;
    ALFA1=0.10;
    CALL RANUNI(SEED1,U);
    CALL RANNOR(SEED1,Z);
    IF U>ALFA1 THEN E=Z;

```



```

ELSE E=Z*DP;
*-----;
*4)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.15;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*5)-DISTRIBUICAO NORMAL CONTAMINADA---;
*--DADOS SOBRE A CONTAMINACAO DAS OBSERVACOES---;
*   DP=5;
*   ALFA1=0.10;
*   CALL RANUNI(SEED1,U);
*   CALL RANNOR(SEED1,Z);
*   IF U>ALFA1 THEN E=Z;
*   ELSE E=Z*DP;
*-----;
*6)-DISTRIBUICAO TUKEY=T(16)-----;
*L2=0.0956;
*L3=0.0625;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*7)-DISTRIBUICAO TUKEY=T(10)-----;
*L2=0.0234;
*L3=0.0148;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*8)-DISTRIBUICAO TUKEY=T(7)-----;
*L2=-0.0735;
*L3=-0.0443;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;
*-----;
*9)-DISTRIBUICAO TUKEY=T(6)-----;
*L2=-0.1377;
*L3=-0.0802;
*   CALL RANUNI(SEED1,U);
*   E=(U**L3 - (1-U)**L3)/L2;

```

```

*-----;
*10)-DISTRIBUICAO EXPONENCIAL DUPLA----;
*  CALL RANUNI(SEED1,U1);
*  CALL RANUNI(SEED3,U2);
*  W = -LOG(U1);
*  IF U2>=0.5 THEN E=W;
*  ELSE E=-W;
*-----;
*11)-DISTRIBUICAO TUKEY=T(5)-----;
*L2=-0.2480;
*L3=-0.1358;
*  CALL RANUNI(SEED1,U);
*  E=(U**L3 - (1-U)**L3)/L2;
*-----;
*12)-DISTRIBUICAO CAUCHY-PADRAO----;
*  CALL RANCAU(SEED1,E);
*-----;
      CALL RANNOR(SEED2,XX);
      Y=B0 + B1*XX + E;
      OUTPUT;KEEP J I E XX Y;
      END;
END;
*---TESTE DE CONTROLE DA SIMULACAO---;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ2 MEAN=MEDIA MEDIAN=MEDIANA;
VAR E;
BY J;
PROC UNIVARIATE NOPRINT;
OUTPUT OUT=ARQ3 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIA;
PROC PRINT DATA=ARQ3;
PROC DELETE DATA=ARQ3;
PROC UNIVARIATE NOPRINT DATA=ARQ2;
OUTPUT OUT=ARQ4 MEAN=MEDIAG STD=DESVIO MEDIAN=MEDG;
VAR MEDIANA;
PROC PRINT DATA=ARQ4;
PROC DELETE DATA=ARQ4;
PROC DELETE DATA=ARQ2;
DATA NEWARQ1;
SET ARQ1;
DROP J I E;
*PROC PRINT DATA=NEWARQ1;
PROC DELETE DATA=ARQ1;

```

```

*-----ENTRADA NO SAS/IML-----;
PROC IML;
*-----TAMANHO DE CADA AMOSTRA-----;
N=20;
*-----NUMERO DE SIMULACOES-----;
NS=2000;
*-----VALORES DOS QUANTIS-----;
Q1=1/3;
Q2=0.50;
Q3=2/3;
PRINT 'MEDIA DE GASTWIRTH: 0.3*B(Q1)+0.4*B(Q2)+0.3B(Q3)';
PRINT 'Q1=' Q1;
PRINT 'Q2=' Q2;
PRINT 'Q3=' Q3;
PRINT 'NUMERO DE SIMULACOES' NS;
PRINT 'AMOSTRA DE TAMANHO' N;
*----DEFINIR QUAL A DISTRIBUICAO DOS ERROS--;
*PRINT ' DISTRIBUICAO DOS ERROS: NORMAL PADRAO';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,100)+0.85*N(0,1)';
PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,100)+0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.15*N(0,25)+0.85*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: 0.10*N(0,25)+0.90*N(0,1)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(10)';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(6)';
*PRINT 'DISTRIBUICAO DOS ERROS: EXPONENCIAL DUPLA';
*PRINT 'DISTRIBUICAO DOS ERROS: TUKEY = T(5)';
*PRINT 'DISTRIBUICAO DOS ERROS: CAUCHY PADRAO';
PRINT 'PARAMETROS INICIAIS: B0=1 B1=1';
USE NEWARQ1;
READ ALL INTO MA;
DO J=1 TO NS;
Y=MA[(J-1)*N+1:J*N,2];
X=REPEAT(1,N,1)||MA[(J-1)*N+1:J*N,1];
/*-----ROTINA PARA O PROBLEMA DE REGRESSAO QUANTILICA-----
* YNAME: NOME DA VARIABEL DEPENDENTE
* Y: VARIABEL DEPENDENTE
* XNAME: NOME DAS VARIABEIS INDEPENDENTES
* X: VARIABEIS INDEPENDENTES
* B: ESTIMATIVAS
* PREDICT: VALORES PREDITOS
* ERROR: DIFERENCA ENTRE Y E VALOR PEDITO
* Q: QUANTIL
* NOTAS: ESTA SUBROTINA ENCONTRA AS ESTIMATIVAS B QUE MINIMIZAM

```

```

*          Q * (Y - XB) * E + (1-Q) * (Y - XB) * ^E
*      ONDE E = ( XB <= Y ).
*      ESTA SUBROTINA SEGUE AS DIRECOES DADAS EM:
*      KOENKER, R. AND G. BASSETT (1978). REGRESSION
*      QUANTILES. ECONOMETRICA. VOL. 46. NO. 1. 33-50.
*      BASSETT, G. AND R. KOENKER (1982). AN EMPIRICAL
*      QUANTILE FUNCTION FOR LINEAR MODELS WITH IID ERRORS.
*      JASA. VOL. 77. NO. 378. 407-415.
*-----*/
START RQ( YNAME, Y, XNAME, X, B, PREDICT, ERROR, Q);
BOUND=1.OE10;
COEF = X';
M = NROW(COEF);
N = NCOL(COEF);
/*-----BUILD RHS AND BOUNDS-----*/
R = REPEAT(O,M+2,1);
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || -BOUND ;
U = REPEAT(Q,1,N) || REPEAT(.,1,M) || { . . } ;
/*-----BUILD COEFFICIENT MATRIX AND BASIS-----*/
A = (          Y'          || REPEAT(O,1,M) || { -1 0 }          ) //
    ( REPEAT(O,1,N) || REPEAT(-1,1,M) || { 0 -1 }          ) //
    (          COEF          ||          I(M)          || REPEAT(O,M,2) ) ;
BASIS = N+M+2 - (O:N+M+1);
/*-----FIND A FEASIBLE SOLUTION-----*/
CALL LP(RC,P,D,A,R,,U,L,BASIS);
*PRINT RC;
/*-----FIND THE OPTIMAL SOLUTION-----*/
L = REPEAT(Q-1,1,N) || REPEAT(O,1,M) || -BOUND || {O} ;
U = REPEAT(Q,1,N) || REPEAT(O,1,M) || { . 0 } ;
CALL LP(RC,P,D,A,R,N+M+1,U,L,BASIS);
*PRINT RC;
/*-----REPORT THE SOLUTION-----*/
VARIABLE = XNAME'; B=D[3:M+2];
PREDICT = X*B;
ERROR = Y - PREDICT;
WSUM = SUM ( CHOOSE(ERROR<O , (Q-1)*ERROR , Q*ERROR) );
*PRINT ,,'REGRESSAO QUANTILICA ESTIMADA' ,
    'VARIABEL DEPENDENTE: ' YNAME ,
    'REGRESSAO QUANTILICA: ' Q ,
    'NUMERO DE OBSERVACOES: ' N ,
    'SOMA DOS ERROS ABSOLUTOS PONDERADOS: ' WSUM ,
    VARIABLE B;
*PRINT X Y PREDICT ERROR;

```

```

FINISH;
*-----EXECUCAO DO PROGRAMA DE REGRESSAO QUANTILICA-----;
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B1,PRED,RESID,Q1);
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B2,PRED,RESID,Q2);
RUN RQ('VAR.DEP.',Y,{ 'INTERCEPT' 'VAR.INDEP.' },X,B3,PRED,RESID,Q3);
*-----VALORES DOS PESOS-----;
P1=0.3;
P2=0.4;
P3=0.3;
*-----ESTIMATIVA FINAL DE BETA-----;
BETAF=P1*B1 + P2*B2 + P3*B3;
*PRINT BETAF;
*-----;
*PRINT 'VALOR DA ESTIMATIVA FINAL DE BETA PARA A SIMULACAO' J;
*PRINT BETAF;
*-----CALCULO DOS VALORES PREDITOS-----;
  YPREDITO = X*BETAF;
  RESIDUO = Y - YPREDITO;
*PRINT X Y YPREDITO RESIDUO;
  IF J=1 THEN BETAG=BETAF';
  ELSE BETAG=BETAG//BETAF';
END;
*PRINT BETAG;
VARNOMES = {BO B1};
CREATE ARQBETA FROM BETAG[COLNAME=VARNOMES];
APPEND FROM BETAG;
CLOSE ARQBETA;
QUIT;
*PROC PRINT DATA=ARQBETA;
*VAR BO B1;
PROC UNIVARIATE PLOT NORMAL DATA=ARQBETA;
VAR BO B1;
LIBNAME INSAS 'K';
DATA INSAS.B11A203;
SET ARQBETA;
RUN;

```

BIBLIOGRAFIA

- ANDREWS, D.F. et. al. (1972) *Robust Estimates of Location*. Princeton University Press.
- BASSETT, G.J. and KOENKER, R. (1982) An Empirical Quantile Function for Linear Models with i.i.d. errors. *Journal of American Statistical Association*. 77: 407-415.
- BEZERRA, M.I.S. (1980) *Métodos para análise de regressão L_1* (Dissertação de tese de Mestrado). Campinas, Universidade Estadual de Campinas.
- BISHOP, Y. M. M.; FIENBERG, S.E. and HOLLAND, P.W. (1977) *Discrete Multivariate Analysis*. Cambridge, The MIT press.
- BUSTOS, O. e JAMES, K. L. (1980) *Procedimentos Robustos: 4 SINAPE*. Rio de Janeiro, S.C.P.
- COX, D.R. and HINKLEY, D.V. (1974) *Theoretical Statistics*. London, Chapman and Hall.
- DAGPUNAR, J. (1988) *Principles of Random Variate Generation*. Claredon Press.
- DE JONGH, P.J. and DE WET, T. (1985) Trimmed Mean and Bounded Influence Estimates for the Parameters of the AR(1) Process. *Communications in Statistics - Theory Methods*. 14: 1361-1375.
- DRUMOND, M.F.B. (1989) *Métodos Robustos em Análise de Regressão* (Dissertação de tese de Mestrado). São Paulo, Universidade de São Paulo.
- FISHMAN, G.S. and MOORE, L.R. (1982) A Statistical Evaluation of Multiplicative Congruential Generators with Modulus($2^{31} - 1$). *Journal of American Statistical Association*. 77: 129-136.
- GARWIN, W.W. (1960) *Introduction to Linear Programming*. New York, McGraw-Hill Book Company.
- GASTWIRTH, J.L. (1966) On Robust Procedures. *Journal of American Statistical Association*. 929-948.
- HAMPEL, F.R. et. el. (1985) *Robust Statistics: The Approach Based on In-*

fluence Functions. New York, John Wiley & Sons.

HODGES, J.L. and LEHMANN, E.L. (1963) Estimates of Location Based on Rank Tests. *Annals of Mathematical Statistics* 34: 598-611.

HOGG, R.V. (1975) Estimates of Percentile Regression Lines Using Salary Data. *Journal of American Statistical Association*. 70: 56-59.

HOGG, R.V. (1979) Statistical Robustness: One View of Its Use in Applications Today. *The American Statistician*. 33: 108-115.

HUBER, P.J. (1964) Robust Estimation of Location Parameter. *Annals of Mathematical Statistics* 35: 73-101.

HÜSLER, J. (1988) On the asymptotic Behaviour of Selector Statistics. *Communications in Statistics - Theory Methods*. 17: 3569-3590.

JOHNSON, N. and KOTZ, S. (1970) *Continuous Univariate Distribution vol.1*. Houghton Mifflin Company.

KOENKER, R. and BASSETT, G.J. (1978) Regression Quantiles. *Econometrica*. 46: 33-50.

KOENKER, R. and PORTNOY, S. (1987) L-Estimation for Linear Models. *Journal of American Statistical Association*. 82: 851-857.

LIRA FILHO, S. (1991) *Um estudo comparativo de estimadores adaptáveis do parâmetro de locação* (Dissertação de tese de Mestrado). Campinas, Universidade Estadual de Campinas.

MOSTELLER, F. and TUKEY, J.W. (1977) *Data Analysis and Regression: a second course in Statistics*. Addison - Wesley Publishing Company.

PORTNOY, S. and KOENKER, R. (1989) Adaptive L-Estimation for Linear Model. *Annals of Statistics*. 17: 362-381.

RANBERG, J.S. and SCHMEISER, B.W. (1972) An Approximate Method for Generating Symmetric Random Variables. *Communications of ACM*. 15: 987-990.

RIPLEY, B.D. (1987) *Stochastic Simulation*. New York, John Wiley & Sons.

ROTHENBERG, T.J., FISHER, F.M. and TILANUS, C.B. (1964) A Note on

Estimation from a Cauchy Sample. *Journal of American Statistical Association*. 460-463.

RUPPERT, D. and CARROLL, R.J. (1980) Trimmed Least Squares Estimation in the Linear Model *Journal of American Statistical Association*. 75: 828-838.

SAS Institute Inc. (1985) *SAS/IML User's Guide*, Version 5 Edition. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1988) *SAS/GRAPH User's Guide*, Release 6.03 Edition. Cary, NC: SAS Institute Inc.

TUKEY, J.W. (1977) *Exploratory Data Analysis*. Addison - Wesley Publishing Company.

WELSH, A.H. (1987) The Trimmed Mean in the Linear Model. *Annals of Statistics*. 15: 20-45.

WELSH, A.H. (1987) On Step L-Estimates for the Linear Model. *Annals of Statistics*. 15: 626-641.