

MÉTODOS DE DIAGNÓSTICO EM REGRESSÃO  
LINEAR E SUA EXTENSÃO PARA O CASO  
DE MAIS DE UMA OBSERVAÇÃO INFLUENTE  
SIMULTANEAMENTE

PAULO HIDEO NAKAMURA



UNICAMP

**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

CAMPINAS - SÃO PAULO  
BRASIL

I/UNICAMP  
N145m

7267

**MÉTODOS DE DIAGNÓSTICO EM REGRESSÃO LINEAR E SUA  
EXTENSÃO PARA O CASO DE MAIS DE UMA OBSERVAÇÃO  
INFLUENTE SIMULTANEAMENTE**

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida pelo Sr. Paulo Hideo Nakamura e aprovada pela Comissão Julgadora.

Campinas, 30 de julho de 1986.

  
Prof. Dr. José Norberto W. Dachs  
Orientador

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

**UNICAMP**  
**BIBLIOTECA CENTRAL**

Aos meus Pais.

## R E S U M O

Em regressão linear, têm sido intensamente discutidas técnicas de diagnóstico que permitam detetar pontos que exercem forte influência no ajuste. Neste trabalho, estão descritas as medidas de diagnóstico comumente utilizados para quantificar a influência de pontos ou subconjunto de pontos.

A identificação de pontos que sejam influentes individualmente é feita, sem grandes dificuldades, analisando as medidas de diagnóstico para cada observação. Um aspecto muito importante do problema é a identificação de mais de uma observação que, conjuntamente, sejam influentes e os problemas em detetar tais subconjuntos são discutidos.

É apresentado um método gráfico proposto por Denby e Mallows para análise de resíduos, baseado em métodos robustos de ajuste. É desenvolvido um método gráfico, semelhante ao de Denby e Mallows, usando e adaptando uma conjetura de Huber, para detecção exploratória de subconjuntos de pontos influentes.

Estes gráficos, denominados gráficos de diagnóstico, foram construídos para vários conjuntos de dados e se mostraram eficazes na identificação de subconjuntos. Após a detecção, pode-se usar o D de Cook múltiplo para uma análise formal.

## A B S T R A C T

In linear regression analysis several diagnostic techniques have been extensively discussed. These techniques are aimed at detecting points that may have a strong influence on the fitted model. In this work are discussed the most commonly used of these techniques with emphasis on the methodology for detection of one or more points that may be simultaneously influential.

The identification of individual points that may be influential is done without major difficulties using diagnostic measures that are evaluated for each observation. A very important aspect of the problem emerges when the attempt is made to identify sets of several points that are simultaneously influential.

A graphical method, proposed by Denby and Mallows for the analysis of residuals, based on robust methods of fitting is presented. With the use and adaptation of a conjecture of Huber a new graphical method, similar to the one of Denby and Mallows, is proposed for the exploratory detection of sets of potentially influential points.

These diagnostic plots were then computed and are present for several sets of data showing a very good potential in the detection of sets of influential points. After detection, Cook's multiple D, or other alternative techniques can be used for more formal analysis.

## ÍNDICE

	pag
1. Introdução.....	01
2. Regressão Linear.....	03
2.1 O ajuste de um modelo linear.....	03
2.2 Estimador de mínimos quadrados.....	04
2.3 Estimador de mínimos quadrados ponderados..	09
2.4 Estimadores Robustos.....	10
3. Detecção de um ponto influente.....	12
3.1 Os Elementos da diagonal da matriz chapéu..	13
3.2 Resíduos studentizados internamente.....	17
3.3 Resíduos studentizados externamente.....	18
3.4 O D de Cook.....	19
3.5 O DFFITS.....	23
4. Influência simultânea de mais de um ponto.....	26
4.1 O DFFIT e D de Cook para o caso de mais de um ponto.....	27
4.2 O problema em encontrar subconjuntos.....	30
4.3 Técnica gráfica de Denby e Mallows.....	31
5. Conclusão.....	81
Apêndice.....	83
Referências.....	97

## C A P Í T U L O    1

### **INTRODUÇÃO**

Num problema de regressão, entre as várias análises para o completo entendimento dos dados e a especificação de um modelo que se ajuste, de forma a traduzir a realidade da melhor maneira possível, está a fase de diagnóstico. É de consenso de todos os pesquisadores, que este estudo é imprescindível e tem sido o assunto de muitos trabalhos publicados na área de regressão.

Dado um problema de ajuste, devemos primeiramente, definir um modelo e escolher um método de estimação dos parâmetros do modelo assumido, baseado em algum critério de distância. Com algumas suposições probabilísticas sobre o modelo, alguns testes de adequação do modelo poderão ser executados, simultaneamente com a análise de resíduos. A análise de resíduos, com as várias técnicas gráficas e testes, permite verificar as suposições assumidas e podem sugerir mudanças no modelo. Além disso, possibilita os primeiros contatos com pontos que se ajustam mal e que possam ter influência desproporcional no modelo. A detecção de problemas com o ajuste começa com a análise de resíduos e portanto, começa o trabalho de diagnose. A identificação de pontos ou conjunto de pontos que têm influência forte no ajuste e a quantificação desta influência será o objetivo seguinte. A preocupação em detectar tais pontos, deve-se ao fato de que a presença destes pode influir decisivamente no ajuste, de maneira que o modelo estimado seja baseado quase que somente nestes poucos pontos em detrimento da maioria dos dados. Isto acarretaria na determinação de um modelo que poderia levar a conclusões totalmente erradas.

A ocorrência de pontos ou conjunto de pontos que influenciam o ajuste, não necessariamente são ruins. Pode ocorrer situações em que são perfeitamente explicáveis no contexto do problema e fornecer conclusões interessantes.

A quantificação da influência de um ponto sobre o ajuste é feita, basicamente, retirando-o e verificando como os parâmetros se alteram com a sua exclusão. Dado que um ponto causa alterações significativas nas estimativas, a exclusão definitiva deste, deve ser feita com muito cuidado. Para tomarmos uma decisão deste tipo, devemos estar convictos de que a observação foi realmente registrada indevidamente.

As técnicas para se detectar influência ponto por ponto, ou seja, a influência de uma observação no ajuste são relativamente simples e não necessitam de um número grande de operações

para serem determinados.

Entretanto, se quisermos estudar a influência de conjunto de pontos, a quantidade de cálculos aumenta muito, e ainda, se o conjunto contiver mais de dois pontos, o problema se torna algebricamente intratável e computacionalmente inviável. Este fato será argumentado no capítulo 3.

O objetivo deste trabalho é estudar este problema e propor um método alternativo para a determinação de conjunto de pontos influentes simultaneamente, baseado em uma técnica proposta por Denby e Mallows num artigo publicado em 1977[13].

Neste trabalho, concentraremos os estudos em modelos de regressão linear de posto completo, ou seja, a matriz de delineamento, que denotaremos  $X$ , tem posto completo e os valores desta matriz serão considerados constantes.

No capítulo que segue, relataremos brevemente o problema de aproximação de modelos lineares, na qual introduziremos técnicas de estimação, definições e notações que serão necessários para os capítulos seguintes.

No capítulo 3, descreveremos as técnicas de diagnóstico de uso mais comum. Os problemas algébricos e computacionais que surgem quando quisermos encontrar grupos de mais de um ponto, que possam, conjuntamente, exercer forte influência no ajuste, será tratado no capítulo 4, assim como a descrição de um método alternativo de detecção de pontos influentes utilizando regressão robusta. Este método poderá, como veremos, trazer grande ajuda na identificação de subconjuntos com mais de um ponto que possa exercer forte influência sobre o ajuste.

Os métodos de diagnóstico descritos aqui, são em geral aplicáveis em situações em que as observações são obtidas sem uma estrutura de delineamento previamente estabelecida. Nos casos em que há um planejamento de maneira que os valores de  $X$  são obtidos conforme o desejo do experimentador, muitas vezes não há necessidade de estudar a influência de pontos.



## C A P Í T U L O    2

### REGRESSÃO LINEAR

Este capítulo tem como objetivo, fornecer subsídios para o acompanhamento deste trabalho. Neste sentido, abordaremos de maneira breve o problema de ajuste de modelos lineares, estimação de parâmetros e formalização de alguns testes.

Um problema de regressão consiste basicamente em, dado um conjunto de observações de um mecanismo ou fenômeno, procurar explicá-lo através de um modelo linear na qual as observações ou resultados experimentais possam ser escritos como função de outras variáveis que são chamadas de regressoras ou suportes, que em nosso caso, serão considerados constantes, ou variáveis experimentais que são controlados pelo pesquisador e que são obtidas com erro negligível.

#### 2.1 O AJUSTE DE UM MODELO LINEAR

Sejam  $y_1, y_2, \dots, y_n$  as observações e  $x_1, x_2, \dots, x_p$  os regressores e suponhamos que  $y_i$  pode ser escrito segundo o seguinte modelo linear

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i \quad (2.1.1)$$

onde  $\beta_1, \dots, \beta_p$  são parâmetros desconhecidos e  $\epsilon$  a flutuação ou erro inerente ao experimento e suponhamos ainda que estes erros são variáveis aleatórias com média zero, variância  $\sigma^2$  e são não correlacionados, ou seja,

$$\begin{aligned} S1) E(\epsilon_i) &= 0, \text{ Var}(\epsilon_i) = \sigma^2 \quad i=1, \dots, n, \\ S2) \text{Cov}(\epsilon_i, \epsilon_j) &= 0, \quad i \neq j \end{aligned} \quad (2.1.2)$$

Este modelo é chamado modelo de Gauss-Markov.

Escrevendo na forma matricial, teremos

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

onde,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} , \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Consideremos a primeira coluna de  $X$  constante igual a 1, ou seja,  $x_{11} = x_{21} = \dots = x_{n1} = 1$  e particularmente neste trabalho, os vetores colunas da matriz  $X$  serão escolhidos de maneira que sejam linearmente independentes. Desta forma,  $X$  terá posto completo igual a  $p$ .

Os parâmetros  $\beta_1, \beta_2, \dots, \beta_p$  devem ser estimados segundo algum critério de distância e na secção seguinte, abordaremos alguns métodos que serão mencionados no decorrer deste texto.

## 2.2 ESTIMADOR DE MÍNIMOS QUADRADOS

Um dos métodos para se obter um estimador de  $\beta$  é o método de mínimos quadrados que consiste em minimizar  $\sum_i \epsilon_i^2$  com respeito a  $\beta$ , isto é, minimizar

$$\epsilon\epsilon = \|y - X\beta\|^2 \quad (2.2.1)$$

que pode ainda ser escrito

$$\begin{aligned} \epsilon\epsilon &= (y - X\beta)'(y - X\beta) \\ &= y'y - 2\beta'X'y + \beta'X'X\beta \end{aligned}$$

e diferenciando com relação a  $\beta$ , teremos

$$-2X'y + 2X'X\beta = 0$$

ou seja,

$$X'X\beta = X'y \quad (2.2.2)$$

Multiplicando ambos os lados por  $(X'X)^{-1}$ , teremos

$$\beta = (X'X)^{-1}X'y \quad (2.2.3)$$

Consequentemente, teremos o vetor  $\hat{y}$  ajustado

$$\hat{y} = X\beta = X(X'X)^{-1}X'y \quad (2.2.4)$$

e chamando  $H = X(X'X)^{-1}X'$  conhecido como matriz chapéu pelo fato de colocar chapéu em  $y$ ,

$$\hat{y} = Hy \quad (2.2.5)$$

e ainda,

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b} = \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y} - \mathbf{H}\mathbf{y} \\
&= (\mathbf{I} - \mathbf{H})\mathbf{y}
\end{aligned} \tag{2.2.6}$$

que são chamados resíduos.

$\mathbf{H}$  é uma matriz de projeções ortogonais e portanto simétrica e idempotente. Por ser idempotente, seu traço é igual a seu posto, ou seja,

$$\sum_{i=1}^n h_{ii} = \text{traço}(\mathbf{H}) = \text{posto}(\mathbf{H}) = p.$$

De (2.2.4), podemos notar que  $\hat{\mathbf{y}}$  é uma combinação linear dos vetores colunas de  $\mathbf{X}$  e estes são supostos linearmente independentes e portanto o espaço gerado pelas colunas de  $\mathbf{X}$ , que chamaremos  $C(\mathbf{X})$  é de dimensão  $p$  e minimizar (2.2.1) consiste em encontrar o vetor  $\hat{\mathbf{y}} \in C(\mathbf{X})$  tal que

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \min_{\tilde{\mathbf{y}} \in C(\mathbf{X})} \|\mathbf{y} - \tilde{\mathbf{y}}\|^2$$

e portanto, o valor mínimo de  $\mathbf{e}'\mathbf{e}$  é dado por

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\
&= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\
&= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'(\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{y}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} \quad \text{por (2.2.2)} \\
&= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}
\end{aligned} \tag{2.2.7}$$

que é chamado de soma de quadrado de resíduos (SQRes).

Os estimadores de mínimos quadrados são não viciados, únicos, de variância mínima e dizemos que são BLUE (Best Linear Unbiased Estimates).

Das suposições em (2.1.2), temos

$$\begin{aligned}
E(\mathbf{b}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}$$

e a matriz de variância e covariância  $\mathbf{V}$  é

$$\begin{aligned}
\mathbf{V}(\mathbf{b}) &= \mathbf{V}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

De (2.2.6) temos

$$\begin{aligned} V(\mathbf{e}) &= V[(\mathbf{I} - \mathbf{H})\mathbf{y}] \\ &= (\mathbf{I} - \mathbf{H})V(\mathbf{y})(\mathbf{I} - \mathbf{H}) \\ &= \sigma^2(\mathbf{I} - \mathbf{H}) \end{aligned}$$

e portanto,

$$\begin{aligned} \text{Var}(e_i) &= \sigma^2(1 - h_{ii}) \\ \text{Cov}(e_i, e_j) &= \sigma^2(-h_{ij}), \quad i \neq j \end{aligned} \quad (2.2.8)$$

De (2.2.5), temos

$$V(\hat{\mathbf{y}}) = V(\mathbf{H}\mathbf{y}) = \mathbf{H}V(\mathbf{y})\mathbf{H}' = \sigma^2\mathbf{H}\mathbf{H}' = \sigma^2\mathbf{H}$$

O estimador de  $\sigma^2$  é dado por

$$s^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{n - p} = \frac{\text{SQRes}}{n - p} \quad (2.2.9)$$

que é o quadrado médio de resíduos e é um estimador não viciado de  $\sigma^2$ .

A soma de quadrados total  $\mathbf{y}'\mathbf{y}$  pode ser decomposta na soma de quadrados de regressão e soma de quadrados de resíduos, ou seja,

$$\text{SQT} = \text{SQR} + \text{SQRes}$$

onde

$$\text{SQT} = \mathbf{y}'\mathbf{y}$$

$$\text{SQR} = \hat{\mathbf{y}}'\hat{\mathbf{y}} = \mathbf{y}'\mathbf{H}\mathbf{H}\mathbf{y} = \mathbf{y}'\mathbf{H}\mathbf{y}$$

$$\begin{aligned} \text{SQRes} &= \mathbf{e}'\mathbf{e} = \mathbf{y}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}. \end{aligned}$$

Então,

$$\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{H}\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \quad (2.2.10)$$

Até agora, nos preocupamos somente em ajustar um modelo linear a um conjunto de dados. Entretanto, se desejamos fazer um estudo mais detalhado com relação aos parâmetros estimados e ao ajuste propriamente, tais como quantificar a magnitude dos parâmetros, construir intervalos de confiança para os parâmetros e testar hipóteses, é necessário estabelecer uma estrutura probabilística no modelo. Para tanto, acrescentaremos uma suposição às estabelecidas em (2.1.2).

$$S4) \quad \epsilon_i \text{'s tem distribuição normal.} \quad (2.2.11)$$

Esta suposição, aliada à S3, implica que que  $\epsilon_i$ 's são independentes e portanto, o vetor aleatório  $\epsilon$  tem distribuição normal n-variada com vetor de médias  $0$  e matriz de variância e covariância  $\sigma^2 I$ .

Com isto, teremos

$$\begin{aligned} \text{i) } y &\sim N(X\beta, \sigma^2 I) \\ \text{ii) } b &\sim N(\beta, \sigma^2 (X'X)^{-1}) \end{aligned} \quad (2.2.12)$$

O Teorema de Fisher-Cochran nos assegura que as formas quadráticas  $y'Hy$  e  $y'(I - H)y$  são independentes e têm distribuição qui-quadrado não central com p e n-p graus de liberdade, respectivamente. A tabela de análise de variância usual fica

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	p	$y'Hy$	$\frac{y'Hy}{p}$
Resíduo	n - p	$y'(I - H)y$	$\frac{y'(I - H)y}{n - p}$
Total	n	$y'y$	

e assim, o quociente do quadrado médio de regressão pelo quadrado médio de resíduos tem distribuição F não central com p e n-p graus de liberdade. O parâmetro de não centralidade será nulo somente quando  $X\beta = 0$  e portanto, um teste de adequação do modelo poderá ser construído testando as hipóteses

$$\begin{aligned} H: X\beta &= 0 \\ K: X\beta &\neq 0. \end{aligned} \quad (2.2.13)$$

Frequentemente, estamos interessados em testar a contribuição de um parâmetro isolado ou de alguns parâmetros no modelo, ou seja, verificar se estes parâmetros são nulos. Isto é equivalente a impor restrições aos parâmetros de que sejam nulos, que significa, em suma, fazer o ajuste retirando as variáveis regressoras correspondentes aos parâmetros e verificar se a retirada destas afetam significativamente o ajuste inicial. Ao retirarmos as variáveis regressoras, a parcela da soma de quadrados de regressão explicada por estas no ajuste inicial, se somará à soma de quadrados de resíduos e assim, uma medida da importância destas variáveis poderá ser dado pelo acréscimo na soma de quadrados de resíduos.

Um conjunto de restrições lineares nos parâmetros pode ser expresso como

$$F\beta = d$$

onde as  $m$  linhas de  $F$  são supostas linearmente independentes. No caso particular em que quisermos testar se  $m$  parâmetros são iguais a zero, para cada linha de  $F$  faremos corresponder a um parâmetro e a linha que cuida do  $j$ -ésimo parâmetro será composta de zeros a menos na  $j$ -ésima coluna onde o valor será 1 e tomaremos o vetor  $d$  nulo.

O problema é encontrar  $b$  tal que

$$\|y - Xb\|^2 = \min \|y - X\beta\|^2$$

$$\text{sujeito a } Fb = d$$

Não exibiremos aqui, o desenvolvimento algébrico para chegarmos à expressão do acréscimo na soma de quadrados devido à restrição  $F\beta = d$ , denotada por  $ASQ(F\beta = d)$ , que poderá ser encontrado com detalhes em [6].

O acréscimo na soma de quadrados devido à restrição  $F\beta = d$  é dado por

$$ASQ(F\beta = d) = (Fb - d)' [F(X'X)^{-1}F']^{-1} (Fb - d) \quad (2.2.14)$$

e se  $d=0$ , pelo teorema de Fisher-Cochran, pode ser mostrado que  $ASQ(F\beta = 0)$  tem distribuição qui-quadrado com número de graus de liberdade igual ao posto de  $F$  e independente da soma de quadrado de resíduos.

Assim, se estamos interessados em testar se um ou alguns parâmetros são iguais a zero, é razoável tomarmos como medida a quantidade (2.2.14). Quanto maior for o acréscimo na soma de quadrados devido à restrição  $F\beta = 0$ , maior será a importância dos parâmetros sujeitos à restrição.

Considerando as hipóteses

$$\begin{aligned} H: F\beta &= 0 \\ K: F\beta &\neq 0, \end{aligned} \quad (2.2.15)$$

é natural que se aceite a hipótese  $H$  se o acréscimo na soma de quadrados devido à restrição for pequeno e a estatística de teste é dada por

$$\frac{ASQ(F\beta = 0)}{m \times s^2} \quad (2.2.16)$$

que tem distribuição  $F$  com  $m$  e  $n-p$  graus de liberdade.



$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad e \quad \delta = \begin{bmatrix} \varepsilon_1/\sqrt{w_1} \\ \vdots \\ \varepsilon_n/\sqrt{w_n} \end{bmatrix}$$

e agora,

$$V(\delta) = \sigma^2 I .$$

e o caso pode ser tratado como de mínimos quadrados ordinários e portanto, o estimador de mínimos quadrados ponderados de  $\beta$ , que denotaremos  $\hat{\beta}$ , será

$$\hat{\beta} = (W'W)^{-1} W'z . \tag{2.3.3}$$

## 2.4 ESTIMADORES ROBUSTOS

Um método de estimação alternativo que é menos sensível a mudança nos dados, ou seja a presença de pontos registrados indevidamente ou erros observacionais, é chamado robusto. Este método faz com que a violação leve de uma suposição, por exemplo a normalidade dos erros, não afete gravemente os testes e intervalos de confiança construídos com base nesta suposição. É portanto usado, quando a distribuição do erro é simétrica mas tem caudas mais longas que a da distribuição normal, ou seja, quando os desvios observados são maiores do que os esperados se os erros fossem normalmente distribuídos. O procedimento robusto faz com que as observações que deram origem a desvios maiores do que esperávamos tenham menor importância ou menor influência nas estimativas dos parâmetros.

Uma classe importante de estimadores robustos pertence à classe de estimadores M que são do tipo máxima verossimilhança.

No caso específico de regressão, estes estimadores são obtidos minimizando uma função de resíduos diferente da soma de quadrados. Sejam  $b^*$  o estimador de  $\beta$  e  $x_i$  a i-ésima linha da matriz  $X$ . Então, para alguma função  $\rho$ ,  $b^*$  é aquele que minimiza

$$\sum_{i=1}^n \rho(y_i - x_i \beta) \tag{2.4.1}$$

Se tomarmos a função  $\rho(x)=|x|^k$ , onde  $k$  é constante positiva e tomarmos  $k=1$ , teremos o estimador  $L_1$  de mínimos desvios absolutos e se  $k=2$ , teremos o estimador de mínimos quadrados.

Huber[17] considerou a classe de estimadores  $b(c)$  de  $\beta$  obtido minimizando

$$\sum_{i=1}^n \rho(y_i - x_i \beta)$$

onde



$$\rho(t) = \begin{cases} t^2/2 & |t| \leq c \\ c*|t| - c^2/2 & |t| > c \end{cases} \quad (2.4.2)$$

e  $c$  é uma constante positiva fixa. Equivalentemente, resolvendo as  $p$  equações

$$\sum_{i=1}^n x_{im} \psi(y_i - \sum_{j=1}^p x_{ij} b_j(c)) = 0, \quad m=1, \dots, p$$

onde

$$\psi(t) = \begin{cases} t & |t| \leq c \\ c*sinal(t) & |t| > c \end{cases} \quad (2.4.3)$$

Na prática,  $c$  deve ser estimado e Huber recomenda a escolha de  $c$  da forma  $c=v*s$  onde  $v$  é uma constante pré-selecionada e  $s$  é um estimador da escala.

Esta escolha de  $c$  faz com que a casos com resíduos grandes sejam atribuídos pesos de forma a reduzir a influência destes nas estimativas.

## CAPÍTULO 3

### DETECCÃO DE UM PONTO INFLUENTE

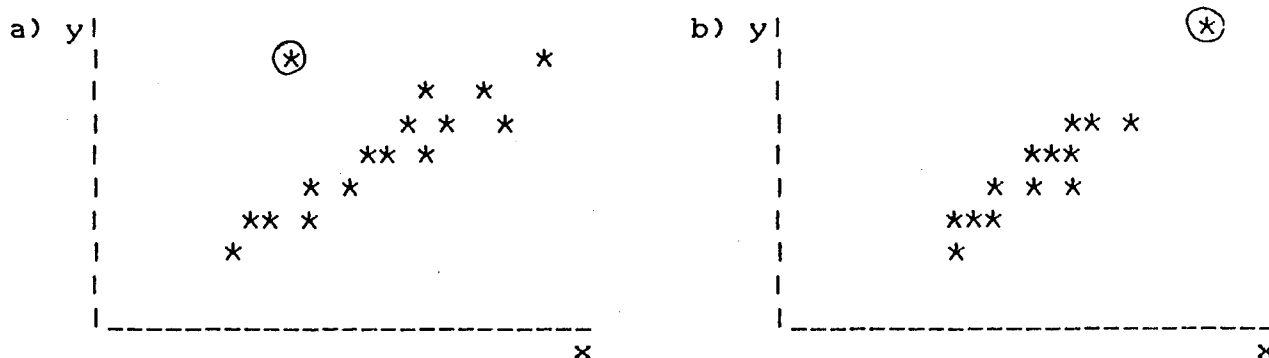
Neste capítulo, descreveremos os métodos comumente utilizados para a detecção de pontos influentes. As técnicas exibidas aqui, estarão restritas a modelos lineares, embora tenham se estendidos a outros modelos.

Relataremos, nesta secção, as técnicas usuais de determinação de pontos influentes individualmente, ou seja, de pontos exercendo forte influência no ajuste, um a um.

Num problema de regressão linear simples, a detecção destes pontos pode até mesmo ser feita através de gráficos. Simplesmente através de um gráfico  $x$  versus  $y$  e a reta ajustada, podemos ter uma noção de quais pontos influenciam o modelo.

Entretanto, em regressão linear múltipla esta prática torna-se inviável pois pode haver pontos, que plotados com cada um dos componentes (variáveis regressoras), não aparentem qualquer influência e, no entanto, considerando o conjunto de componentes, exerçam forte influência no modelo. Faz-se necessária, portanto, técnicas que permitam detectar tais pontos num espaço multidimensional.

Para motivar a introdução destas técnicas, consideremos um problema de regressão linear simples, cuja dispersão dos dados se apresente de uma das seguintes maneiras:



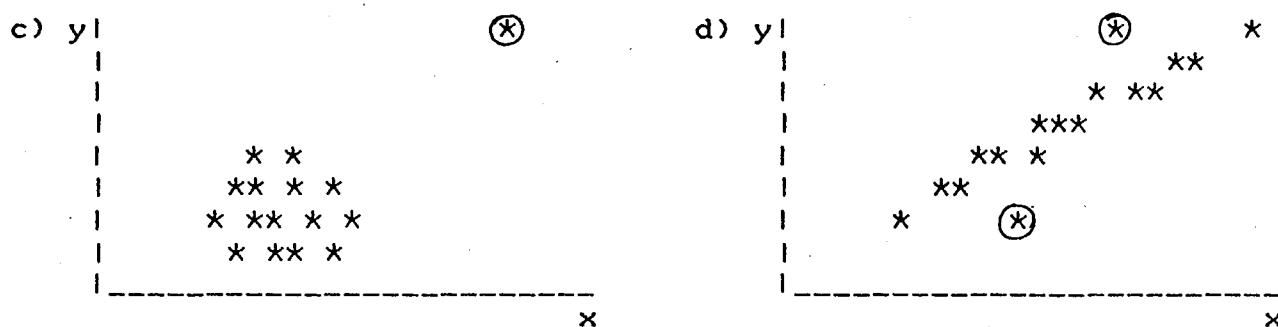


Figura 3.1

Na figura a), o ponto assinalado se distancia da núvem de pontos. Entretanto, a sua posição, próxima à média da variável explicativa, fará com que a sua retirada não cause acentuada mudança na estimativa do coeficiente angular. O intercepto será afetado pois o ajuste estaria sendo puxado para cima por este ponto. Na figura b), o ponto anômalo não afetará as estimativas dos parâmetros pois se posiciona de tal maneira que as informações nele contidas com relação à inclinação e intercepto, são as mesmas do restante dos dados. A figura c) mostra o caso em que um ponto exerce forte influência no modelo. Note que se se excluir este ponto, os parâmetros serão totalmente diferentes daqueles encontrados com a sua presença. Na figura d), os pontos assinalados parecem exercer, conjuntamente, alguma influência no ajuste. Este caso será estudado no capítulo 4, onde trataremos de influência múltipla.

### 3.1 OS ELEMENTOS DA DIAGONAL DA MATRIZ CHAPÉU

Os elementos da diagonal da matriz chapéu nos fornecem uma primeira indicação da existência de pontos anômalos e veremos adiante, que são elementos de grande importância na determinação de outras medidas de diagnóstico.

Hoaglin e Welsch[16], propuseram como uma medida de influência sobre um ajuste, os elementos da diagonal da matriz chapéu. Basicamente, é um método para distinguir pontos anômalos entre as variáveis explicativas, uma vez que a matriz chapéu, que denotaremos  $H$ , não depende de  $y$ .

Cada valor ajustado  $y$  é dado como uma combinação linear dos valores observados pois,

$$\hat{y} = Xb = X(X'X)^{-1}X'y = Hy \quad (3.1.1)$$

e portanto, a matriz  $H$  contém informações que podem revelar pontos anômalos(multivariados) entre as variáveis explicativas.

O elemento  $h_{ij}$  de  $H$  tem uma interpretação direta como o grau de influência que  $y_j$  exerce sobre  $\hat{y}_i$ . Desta forma, o estudo de  $H$  pode revelar pontos sensíveis. Pontos para os quais, o valor

de  $y$  tem grande impacto no ajuste. Os elementos da diagonal de  $H$  podem também ser relacionados à distância entre  $x'_i$  e  $\bar{x}'$  onde  $x'_i$  é o vetor linha da variável explicativa da  $i$ -ésima observação e  $\bar{x}'$  o vetor linha de médias das variáveis explicativas.

A influência do valor observado  $y_i$  no ajuste está mais diretamente refletido no correspondente valor ajustado  $\hat{y}_i$ , e esta informação está precisamente contida em  $h_{ii}$ , que simplificando denotaremos  $h_i$ .

$H$  é simétrica e idempotente e numa matriz de projeção, a soma dos elementos de qualquer linha da matriz é 1, ou seja,

$$\sum_{j=1}^n h_{ij} = 1, \text{ para todo } i \quad (3.1.2)$$

Portanto,

$$h_i = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 \quad (3.1.3)$$

e assim, fica claro que  $0 \leq h_i \leq 1$ .

Seja  $M$  a matriz chapéu obtida centrando as variáveis explicativas. Então,

$$\hat{y} - \bar{y} = Hy - \bar{y} = My$$

e os elementos da matriz  $M$  são

$$h_{ij} = h_{ij} - 1/n$$

que implica

$$1/n \leq h_i \leq 1. \quad (3.1.4)$$

Sabemos que os autovalores de uma matriz de projeção são zero ou 1 e o número de autovalores iguais a 1 é o posto da matriz (Rao[20]). Neste caso,

$$\text{posto}(H) = \text{posto}(X) = p.$$

E assim,

$$\text{traço}(H) = p, \text{ isto é, } \sum_{i=1}^n h_i = p.$$

Portanto, o tamanho médio do elemento da diagonal é  $p/n$ .

Hoaglin e Welsh[16], baseado em experiências, sugerem que um procedimento razoável é estabelecer como o ponto de corte, a quantidade  $2p/n$ , ou seja, se  $h_i > 2p/n$ , a  $i$ -ésima observação é considerada um ponto anômalo.

Belsley, Kuh e Welsch[3] sustentaram este mesmo procedimento argumentando da seguinte maneira: se as variáveis explicativas são distribuídas independentemente segundo a distribuição

normal multivariada, é possível calcular a distribuição exata de certas funções de  $h_i$ 's. Sob esta suposição, mostra que

$$\frac{(n-p)(h_i - 1/n)}{(p-1)(1 - h_i)}$$

tem distribuição F com  $p-1$  e  $n-p$  graus de liberdade.

Para  $p > 10$  e  $n-p > 50$ , o valor de F para 95% é menor que 2 e assim,  $2p/n$  é proposto como um bom ponto de corte grosseiro.

Desta forma, como um indicativo inicial, diremos, segundo critério acima, que a  $i$ -ésima observação tem a sua posição distinta do conjunto de pontos quando  $h_i$  exceder  $2p/n$ .

Huber[17] diz ainda que  $1/h_i$  é o número equivalente de observações na determinação de  $\hat{y}_i$ . De fato, pois de (3.1.2) temos que a soma dos elementos da  $i$ -ésima linha de  $H$  é 1 e assim, se  $h_i = 0.5$ , por exemplo, os outros elementos da linha deverão somar 0.5 e portanto  $1/h_i = 2$  é o número equivalente de observações. Isto é,  $y_i$  recebe uma ponderação 0.5 na determinação de  $\hat{y}_i$  e o restante dos  $y_j, j \neq i$ , fatores de ponderação que deverão somar 0.5.

Huber propõe duas regras básicas para indicação da presença de pontos anômalos:

- a) Estudar com muito cuidado, pontos para os quais o número equivalente de observações é menor que 2, ou seja,  $h_i > 0.5$ .
- b) Estudar com cuidado, pontos para os quais o número equivalente de observações é menor que 5, ou seja,  $h_i > 0.2$ .

Ainda em [16], Hoaglin e Welsch propuseram estudar conjuntamente os  $h_i$ 's e resíduos, embora não tenham chegado a uma relação com as duas quantidades simultaneamente.

Para motivar a necessidade de se fazer tal estudo, tomemos um exemplo simples elaborado por Carvalho e Dachs[ 6].

**EXEMPLO\_3.1** Dois conjuntos de dados com mesmo ajuste por mínimos quadrados ao usar  $y = \beta_1 + \beta_2 x + \epsilon$

conjunto de dados A					conjunto de dados B				
x	y	$\hat{y}$	e	h	x	y	$\hat{y}$	e	h
1	4	2.2	1.8	0.45	1	1	2.2	-1.2	0.45
3	2	2.6	-0.6	0.29	1	2	2.2	-0.2	0.45
5	3	3.0	0.0	0.21	6	6	3.2	2.8	0.20
7	1	3.4	-2.4	0.21	11	3	4.2	-1.2	0.45
14	6	4.8	1.2	0.84	11	4	4.2	-0.2	0.45

Estes dois conjuntos produzem um mesmo ajuste

$$\hat{y} = 2.0 + 0.2x$$

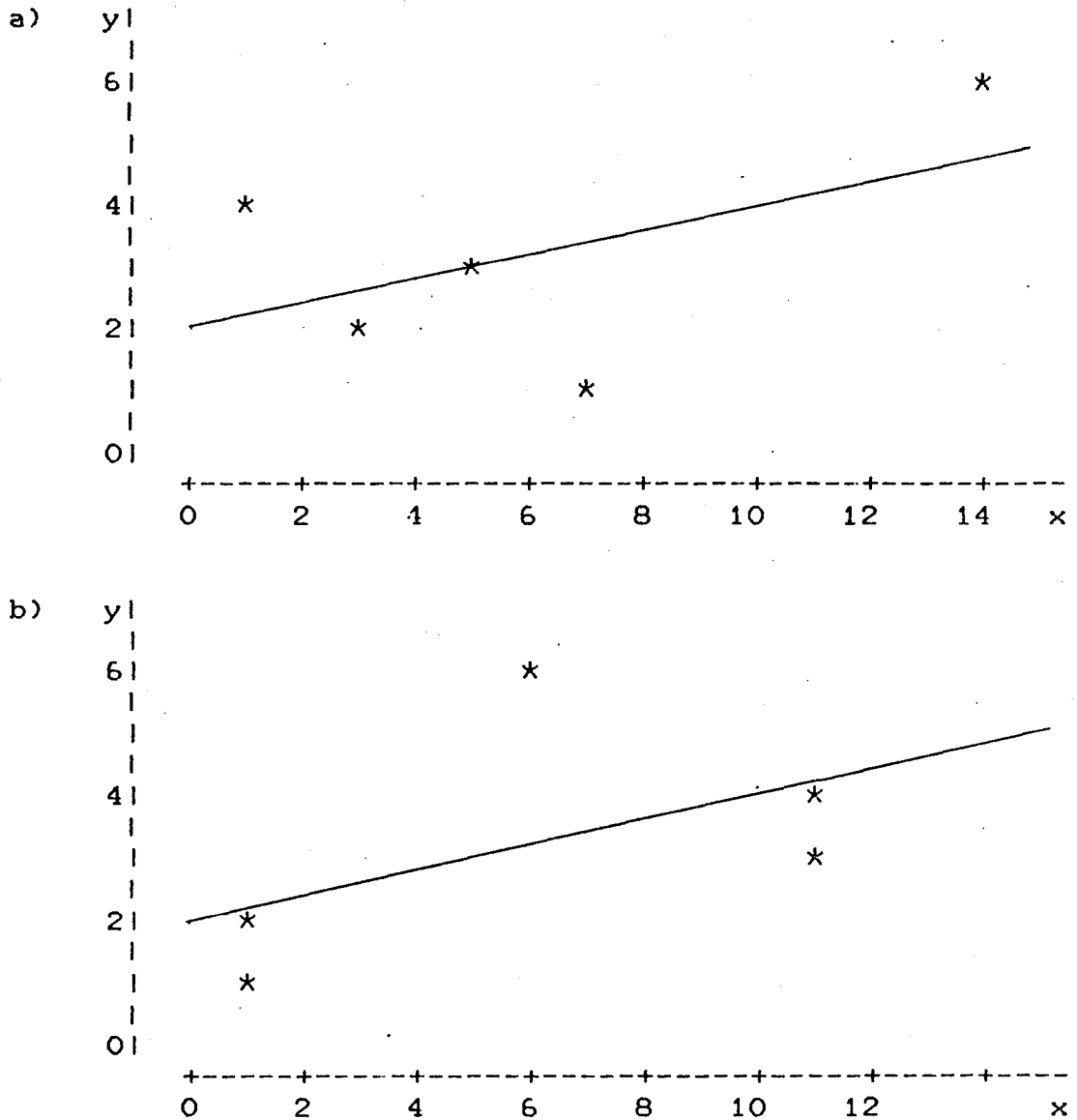


Figura 3.2

Notemos que os gráficos de  $y$  versus  $x$  destes dois conjuntos de dados, figura 3.2 a) e b), guardam uma semelhança com os da figura 3.1 a) e c). Na figura 3.2 a) correspondente ao conjunto de dados A, através de um simples exame visual, podemos ver que a retirada de um único ponto pode alterar de maneira drástica o ajuste. Este ponto corresponde ao par (14,6), que examinando apenas os resíduos, não se detectaria sua influência decisiva no ajuste. Neste problema,  $2p/n=2 \times 2/5=0.8$  e na coluna de  $h'$ s,  $h=0.84$  que dá devida importância ao ponto (14,6). Na figura 3.2 b), correspondente ao conjunto de dados B, o maior resíduo corresponde ao ponto (6,6) do qual poderíamos suspeitar que a sua

retirada causaria uma melhora no ajuste. Entretanto, podemos notar pela coluna dos  $h$ 's que neste caso, o ponto (6,6), segundo este critério não mereceria atenção. Isto deve-se ao fato de que a alteração provocada pelo ponto (6,6) não é devido à posição de  $x$  mas sim do valor observado de  $y$ .

Fica claro, dessa forma, que se faz necessário um estudo conjunto de resíduos e da diagonal da matriz  $H$ .

Passaremos, então, a estudar os resíduos e suas variâncias.

### 3.2 RESÍDUOS STUDENTIZADOS INTERNAMENTE

Lembrando que,

$$\hat{y} = Hy \quad (3.1.5)$$

a variância de  $\hat{y}$  será

$$\text{Var}(\hat{y}) = H \times \text{Var}(y) \times H'$$

e sob a suposição de que  $\text{Var}(y_i) = \sigma^2$  e  $\text{Cov}(y_i, y_j) = 0$ ,  $\text{Var}(y) = \sigma^2 \times I$  e como  $H$  é simétrica e idempotente,

$$\text{Var}(\hat{y}) = \sigma^2 \times H. \quad (3.1.6)$$

Como

$$e = y - \hat{y} = (I - H)y,$$

então,

$$\text{Var}(e) = \sigma^2 \times (I - H). \quad (3.1.7)$$

Temos então que a variância de  $e_i$  é

$$\text{Var}(e_i) = \sigma^2 \times (1 - h_i). \quad (3.1.8)$$

É importante notarmos que  $e_i$ 's são correlacionados e com variâncias diferentes. Como  $s^2$  é um estimador não viciado de  $\sigma^2$ , um estimador não viciado de  $\text{Var}(e_i)$  é

$$\text{var}(e_i) = s^2 \times (1 - h_i) \quad (3.1.9)$$

Pela suposição de que  $E(e) = 0$ , definimos  $t_i$  como

$$t_i = \frac{e_i}{s \times \sqrt{1 - h_i}} \quad (3.1.10)$$

que tem a forma de uma variável aleatória t-de-Student e os  $t_i$ 's são chamados de resíduos studentizados internamente.

Os resíduos studentizados calculados para os dados do exemplo 3.1 são:

para o conjunto de dados A. 1.28, -0.38, 0.00, -1.48 e 1.58  
 para o conjunto de dados B. -0.85, -0.14, 1.65, -0.85 e -0.14

O resíduo correspondente ao último ponto no caso A ganha uma importância maior, pois de terceiro, passou a primeiro em valor absoluto. No conjunto de dados B, permanece a terceira observação com resíduo grande, comparada aos outros.

### 3.3 RESÍDUOS STUDENTIZADOS EXTERNAMENTE

A falta de ajuste num ponto  $i$  pode ser modelado construindo uma nova matriz, digamos  $Z$ , adicionando à matriz  $X$  uma coluna  $p+1$  de zeros a menos na  $i$ -ésima linha onde o valor é 1. O modelo então será  $y = X\beta$  a menos do  $i$ -ésimo ponto onde  $y_i = x_i'\beta + \alpha$ , onde  $x_i'$  é a  $i$ -ésima linha de  $X$ , e  $\alpha$  é o novo parâmetro que cuida somente do ponto  $i$ .

Testar o ajuste no ponto  $i$  é equivalente a testar, portanto, se  $\alpha = 0$ .

Neste novo modelo então,

$$Z'Z = \begin{bmatrix} X'X & | & x_i' \\ \hline x_i' & | & 1 \end{bmatrix} \quad (3.1.11)$$

e a inversa de  $Z'Z$  é

$$(Z'Z)^{-1} = \begin{bmatrix} (X'X - x_i x_i')^{-1} & | & -(X'X)^{-1} x_i / (1-h_i) \\ \hline -x_i' (X'X)^{-1} / (1-h_i) & | & 1/(1-h_i) \end{bmatrix} \quad (3.1.12)$$

Assim, podemos obter o estimador de  $\alpha$ , multiplicando a última linha de  $(Z'Z)^{-1}$  por  $Z'y$ . O estimador será

$$a = \frac{-x_i' (X'X)^{-1} X'y}{1-h_i} + \frac{y_i}{1-h_i} = \frac{-x_i' b + y_i}{1-h_i} = \frac{y_i - \hat{y}_i}{1-h_i} = \frac{e_i}{1-h_i}$$

e portanto o acréscimo na soma de quadrados devido à restrição  $\alpha = 0$  é

$$\begin{aligned} ASQ(\alpha=0) &= a^2 (1/(1-h_i))^{-1} \\ &= (y_i - \hat{y}_i)^2 (1/(1-h_i)) \\ &= e_i^2 / (1-h_i) \end{aligned} \quad (3.1.13)$$



A estatística para testar  $\alpha = 0$  é

$$\tau_i^2 = \frac{ASQ(\alpha = 0)/1}{s^2(1)} = \frac{(y_i - \hat{y}_i)^2}{s^2(1)(1-h_i)} = \frac{e_i^2}{s^2(1)(1-h_i)} \quad (3.1.14)$$

onde  $\tau_i^2$  tem distribuição F com 1 e n-p-1 graus de liberdade e  $s^2(1)$  é a média quadrática de resíduos no modelo

$$y = Z \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \delta ,$$

ou seja, o estimador da variância no modelo aumentado. Como a soma de quadrados com a restrição é igual à soma de quadrados sem a restrição mais o acréscimo na soma de quadrados devido à restrição, a obtenção de  $s^2(1)$  torna-se muito fácil fazendo,

$$(n-p) * s^2 = (n-p-1) s^2(1) + e^2 / (1-h_i) ,$$

ou seja,

$$s^2(1) = \frac{n-p}{n-p-1} * s^2 - \frac{e_i^2}{(n-p-1)(1-h_i)} \quad (3.1.15)$$

De (3.1.14) temos que  $\tau_i = \sqrt{\tau_i^2}$  tem distribuição t-de-Student com n-p-1 graus de liberdade, que chamaremos de resíduos studentizados externamente.

A vantagem de  $\tau_i$ 's sobre  $t_i$ 's é de que temos uma distribuição de referência e ao testar falta de ajuste em um ponto, estamos quantificando o grau de influência no ajuste exercido pelo i-ésimo ponto.

### 3.4 O D DE COOK

A medida proposta por Cook [8], é uma função do resíduo e sua variância, do elemento da diagonal da matriz H e da alteração provocada nos parâmetros estimados quando se retira uma observação.

A influência de um ponto no ajuste pode ser verificada através da alteração que a retirada deste causa nos parâmetros estimados. Esta alteração causada pela retirada da i-ésima observação será quantificada por

$$DFBETA_i = b - b(i) ,$$

onde  $b$  é a estimativa dos parâmetros obtida considerando-se todos os casos e  $b(i)$  a estimativa obtida com a exclusão da i-ésima observação.

Denotaremos, também, por  $y(i)$  e  $X(i)$  o vetor de  $y$ 's ajustados e a matriz de delineamento  $X$  sem a  $i$ -ésima observação, respectivamente.

Para medirmos a influência de um ponto no ajuste, seria necessário retirar uma observação de cada vez, digamos  $i$ , e calcular  $b(i)$ . Para tanto, precisaríamos inverter a matriz  $X'(i)X(i)$  para cada ponto  $i$ . Inverter uma matriz de ordem  $p$  tantas vezes quanto forem o número de observações, torna-se caro e frequentemente inviável. É possível, entretanto, algébricamente chegarmos a uma fórmula que evite este trabalho todo. O teorema que segue é devido a Gauss(1821), havendo também uma versão dada por Sherman, Morrison e Woodbury[21] e usado por Bringham[ 4] para várias finalidades em regressão.

**TEOREMA 3.1** Seja  $A$  uma matriz  $p \times p$ , não singular, simétrica e suponha que  $u$  e  $v$  são duas matrizes  $q \times p$  de posto  $q$ . Então, a inversa existe e é dada por

$$(A + u'v)^{-1} = A^{-1} - A^{-1}u'(I + vA^{-1}u')^{-1}vA^{-1}$$

Em nosso caso, consideremos  $A = X'X$ ,  $u = -x_i$  e  $v = x_i'$ . Então,

$$(X'(i)X(i))^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_ix_i'(X'X)^{-1}}{1 - h_i} \quad (3.1.16)$$

Com isso, basta termos  $(X'X)^{-1}$  para se obter  $(X'(i)X(i))^{-1}$ .

Os parâmetros estimados depois de retirado o  $i$ -ésimo ponto será

$$b(i) = (X'(i)X(i))^{-1}(X'y - x_i'y_i) \quad (3.1.17)$$

Substituindo (3.1.16) em (3.1.17), teremos

$$\begin{aligned} b(i) &= [(X'X)^{-1} + \frac{(X'X)^{-1}x_ix_i'(X'X)^{-1}}{1 - h_i}] * (X'y - x_i'y_i) = \\ &= [(X'X)^{-1}X'y + \frac{(X'X)^{-1}x_ix_i'(X'X)^{-1}X'y}{1 - h_i}] - (X'X)^{-1}x_i'y_i - \\ &\quad \frac{(X'X)^{-1}x_ix_i'(X'X)^{-1}x_i'y_i}{1 - h_i} \end{aligned}$$

Sabemos que  $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  e  $\hat{y}_i = \mathbf{x}_i' \mathbf{b} = \mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . Assim,

$$\begin{aligned} \mathbf{b}(1) &= \mathbf{b} + \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i \hat{y}_i}{1 - h_i} - [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i y_i + \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i h_i y_i}{1 - h_i}] \\ &= \mathbf{b} + \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i \hat{y}_i - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_i} = \mathbf{b} + \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i (\hat{y}_i - y_i)}{1 - h_i} \end{aligned}$$

Desta forma, a expressão de DFBETA fica

$$\text{DFBETA}_i = \mathbf{b} - \mathbf{b}(1) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i \frac{e_i}{1 - h_i} \quad (3.1.18)$$

Cook propôs então, a medida denotada por  $D_i$ , como a medida de influência do ponto  $i$  sobre o ajuste.

$$D_i = \frac{(\mathbf{b} - \mathbf{b}(1))' \mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}(1))}{p s^2} \quad (3.1.19)$$

Substituindo (3.1.18) em (3.1.19), temos

$$\begin{aligned} D_i &= \frac{\mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1} e_i \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i e_i}{p s^2 * (1 - h_i)^2} = \frac{\mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i e_i^2}{p s^2 * (1 - h_i)^2} = \\ &= \frac{h_i e_i^2}{p s^2 * (1 - h_i)^2} = \left[ \frac{e_i}{s * (1 - h_i)^{1/2}} \right]^2 \frac{h_i}{p * (1 - h_i)} \end{aligned}$$

Notemos aqui que a expressão elevada ao quadrado é exatamente  $t_i$  (3.1.10), o  $i$ -ésimo resíduo studentizado internamente, e assim, a expressão final de  $D_i$  fica

$$D_i = \frac{t_i^2 h_i}{p * (1 - h_i)} \quad (3.1.20)$$

ou ainda, de (3.1.6) e (3.1.8), a razão  $h_i / (1 - h_i)$  é a razão de  $\text{Var}(\hat{y}_i)$  por  $\text{Var}(e_i)$  e portanto,

$$D_i = \frac{t_i \text{Var}(\hat{y}_i)}{p \text{Var}(e_i)} \quad (3.1.21)$$

De (3.1.20), podemos notar que  $D_i$  pode ser grande se a  $i$ -ésima observação é muito influente e portanto a razão  $h_i/(1-h_i)$  será grande. Ou ainda, se este ponto produz um resíduo grande de maneira que  $t_i$  seja grande.

Sob a suposição de normalidade, a expressão (3.1.19) é semelhante à que define uma região de confiança para  $\beta$  dada por

$$\frac{(\mathbf{b} - \beta)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta)}{s^2 \times p} \leq F(p, n-p, 1-\alpha)$$

que define o elipsóide de confiança de  $(1-\alpha)100\%$  para um vetor  $\beta$ .

Desta forma, é razoável tomarmos como um valor de comparação dos  $D_i$ 's, os valores da distribuição  $F$ , ou seja, tomar cuidado com os pontos cuja remoção afaste  $\mathbf{b}$  para próximo da fronteira da região de confiança com  $(1-\alpha) \times 100\%$ . Escolhe-se, por exemplo, uma região de confiança com  $(1-\alpha) \times 100\% = 10\%$  e usando a relação

$$F(n, m, 1-\alpha) = 1/F(m, n, \alpha) ,$$

encontramos o valor de comparação dos  $D_i$ 's.

Para os dados do exemplo 3.1, os  $D_i$ 's encontrados são

Conjunto de dados A

$$\begin{aligned} D_1 &= 0.67 \\ D_2 &= 0.01 \\ D_3 &= 0.00 \\ D_4 &= 0.30 \\ D_5 &= 6.56 \end{aligned}$$

Conjunto de dados B

$$\begin{aligned} D_1 &= 0.30 \\ D_2 &= 0.12 \\ D_3 &= 0.34 \\ D_4 &= 0.30 \\ D_5 &= 0.12 \end{aligned}$$

Como o valor de  $F(3, 2, 0.9) = 9.1618$ , o valor de comparação é  $1/9.1618 = 0.109$ . Embora esta medida chame a atenção para todos os pontos do conjunto de dados B, o  $D_3$  correspondente ao ponto (6,6) não mereceu destaque, quando comparado aos outros  $D_i$ 's. Isto deve-se ao fato de que temos no denominador da expressão de  $D$  o fator  $s^2$ . Como neste caso,  $s^2$  é grande, exatamente pela presença do ponto (6,6), faz com que a quantidade  $D_3$  seja não muito diferente dos outros.

Uma outra medida, muito parecida com  $D$  de Cook foi proposta por Belsley, Kuh e Welsch [3], onde este fato é levado em consideração.

### 3.5 O DFFITS

Belsley et al[ 3] propuseram a medida de influência do  $i$ -ésimo ponto no ajuste, denotada por  $DFFITS_i$ , dada por

$$DFFITS_i = \frac{e_i}{s(1)(1-h_i)^{1/2}} * \left[ \frac{h_i}{1-h_i} \right]^{1/2} \quad (3.1.22)$$

Agora, no denominador temos  $s(1)$  no lugar de  $s$ .  $s(1)$  é obtida facilmente usando (3.1.15) e pode ser bem menor que  $s$  se o ponto em questão se situa de maneira distinta do conjunto de observações. Isto faz com que esta medida atribua a importância que um ponto tem no ajuste, de maneira mais apropriada.

Usando (3.1.14), temos

$$DFFITS_i = \tau_i * \left[ \frac{h_i}{1-h_i} \right]^{1/2} \quad (3.1.23)$$

Vimos que  $\tau_i \sim t$ -de Student com  $n-p-1$  graus de liberdade. Se o delineamento fosse completamente balanceado, teríamos  $h_i$ 's iguais a  $p/n$ . O ponto de corte de 5% da distribuição  $t$ -de Student com 20 graus de liberdade é aproximadamente igual a 2. Assim, um valor de comparação para os DFFITS pode ser dado por

$$2 * \left[ \frac{p}{n} / \left( 1 - \frac{p}{n} \right) \right]^{1/2} = 2 * \left[ \frac{p}{n-p} \right]^{1/2}$$

ou seja, devemos estudar com cuidado os pontos para os quais

$$|DFFITS_i| > 2 * (p/(n-p))^{1/2} \quad (3.1.24)$$

Para os dados do exemplo 3.1, os DFFITS encontrados são

Conjunto de dados A

$$\begin{aligned} DFFITS_1 &= 1.40 \\ DFFITS_2 &= -0.20 \\ DFFITS_3 &= 0.00 \\ DFFITS_4 &= -1.05 \\ DFFITS_5 &= 7.25 \end{aligned}$$

Conjunto de dados B

$$\begin{aligned} DFFITS_1 &= -0.72 \\ DFFITS_2 &= -0.11 \\ DFFITS_3 &= 2.21 \\ DFFITS_4 &= -0.72 \\ DFFITS_5 &= -0.11 \end{aligned}$$

As comparações neste exemplo ficam prejudicadas pois cinco observações é muito pouco. Mesmo assim, podemos chamar atenção para os pontos (14,6) correspondente ao 5º ponto do conjunto A e (6,6) correspondente ao 3º ponto do conjunto B.

EXEMPLO 3.2 Os dados são devidos a Brownlee[ 5] e consiste na oxidação de amônia para obtenção de ácido nítrico. As observações foram tomadas em 21 dias sucessivos e registrou-se as seguintes informações:  $x_2$  corrente de ar,  $x_3$  a temperatura da água de resfriamento,  $x_4$  a concentração de ácido nítrico no líquido de absorção. A resposta  $y$  é 10 vezes a porcentagem de amônia não absorvida.

Quadro 3.1 Dados de amônia não absorvida

Obs.	$x_2$	$x_3$	$x_4$	$y$
1	80	27	89	42
2	80	27	88	37
3	75	25	90	37
4	62	24	87	28
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

Este mesmo conjunto de dados foi estudado pelo próprio Brownlee, por Draper e Smith[14], Daniel e Wood[12], Denby e Mal-lows[13] e tomados como exemplo em muitas publicações.

O modelo assumido para este exemplo é dado por

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

e os parâmetros de mínimos quadrados estimados são

$$\begin{aligned} b_1 &= -39.9197 \\ b_2 &= 0.7156 \\ b_3 &= 1.2953 \\ b_4 &= -0.1521 \end{aligned}$$

Neste exemplo e na maioria dos que sequem, não estaremos preocupados em encontrar um melhor modelo que represente os dados e também não concentraremos esforços em testes de hipóteses sobre

os parâmetros ou de regressão. A intensão dos exemplos será explorar de maneira mais abrangente possível as medidas de diagnóstico, assumindo um determinado modelo. Assim, temos no quadro 3.2 os valores de  $\hat{y}$  estimados, os resíduos, os elementos da diagonal da matriz de projeções, resíduos studentizados externamente e as medidas de influência D de Cook e DFFITS.

No quadro 3.2, na coluna dos  $h_j$ 's, adotando o critério sugerido por Huber, descrito na seção 3.1.1, os elementos assinalados com '\*' são maiores que 0.2, devendo portanto serem estudados com cuidado os pontos 1,2,7,8,12,14,17 e 21. Na coluna dos resíduos studentizados ( $\tau$ ), os pontos assinalados com '\*' são significativos a 5% e '\*\*' a 1% e portanto, os pontos 4 e 21 são, segundo este critério, pontos aberrantes. Na coluna dos D, para os pontos assinalados com '\*' a estimativa  $b(\cdot)$ , ou seja, a estimativa de  $\beta$  sem o respectivo ponto, desloca-se para um contorno de 10% em relação a  $b$  e os assinalados com '\*\*' para um contorno de no mínimo 25%. Na coluna DFFITS, o valor de comparação é encontrado conforme (3.1.24) e o ponto assinalado é dado como influente. Tanto a medida D quanto o DFFITS, apontaram como influente, somente o ponto 21. Todos os valores de comparação estão dispostos logo abaixo do quadro. #

Quadro 3.2 Medidas de influência - dados de amonia não absorvidos.

caso	$\hat{y}$	e	h	$\tau$	D	DFFITS
1	38.7654	3.2346	0.3016*	1.2095	0.1537	0.7947
2	38.9175	-1.9175	0.3178*	-0.7051	0.0597	-0.4813
3	32.4445	4.5555	0.1746	1.6179	0.1264	0.7442
4	22.3022	5.6978	0.1285	2.0518*	0.1305	0.7879
5	19.7117	-1.7117	0.0522	-0.5305	0.0040	-0.1245
6	21.0069	-3.0069	0.0775	-0.9632	0.0196	-0.2792
7	21.3895	-2.3895	0.2192*	-0.8259	0.0488	-0.4377
8	21.3895	-1.3895	0.2192*	-0.4737	0.0165	-0.2510
9	18.1444	-3.1444	0.1402	-1.0486	0.0446	-0.4234
10	12.7328	1.2672	0.2000	0.4262	0.0119	0.2131
11	11.3637	2.6363	0.1550	0.8783	0.0359	0.3762
12	10.2205	2.7795	0.2172*	0.9667	0.0651	0.5092
13	12.4286	-1.4286	0.1575	-0.4687	0.0108	-0.2027
14	12.0505	-0.0505	0.2058*	-0.0170	0.0000	-0.0086
15	5.6386	2.3614	0.1905	0.8006	0.0385	0.3883
16	6.0949	0.9051	0.1311	0.2912	0.0034	0.1131
17	9.5200	-1.5200	0.4121*	-0.5996	0.0655	-0.5020
18	8.4551	-0.4551	0.1606	-0.1487	0.0011	-0.0650
19	9.5983	-0.5983	0.1745	-0.1972	0.0022	-0.0907
20	13.5879	1.4121	0.0802	0.4431	0.0045	0.1308
21	22.2377	-7.2377	0.2845*	-3.3305**	0.6920**	-2.1003*

$$t(16,0.05) = 1.746$$

$$t(16,0.01) = 2.583$$

$$F(4,17,0.10) = 0.2604$$

$$F(4,17,0.25) = 0.8469$$

$$|DFFITS| > 0.97$$

## C A P Í T U L O    4

### INFLUÊNCIA SIMULTÂNEA DE MAIS DE UM PONTO

Vimos, no capítulo anterior, formas de detectar influência de um único ponto no ajuste e meios para quantificar esta influência.

Entretanto, para um estudo mais detalhado, é necessário considerarmos a possibilidade de que um conjunto de pontos possa exercer forte influência no ajuste. Este conjunto de pontos, se existir, pode não ser detectado com as técnicas mencionadas até aqui, pois a influência de um ponto isolado deste conjunto pode ser mascarado por outros deste mesmo conjunto.

Para ilustrarmos este fato, consideremos novamente, um problema de regressão linear simples, cuja dispersão dos dados é dado como na figura 4.1.

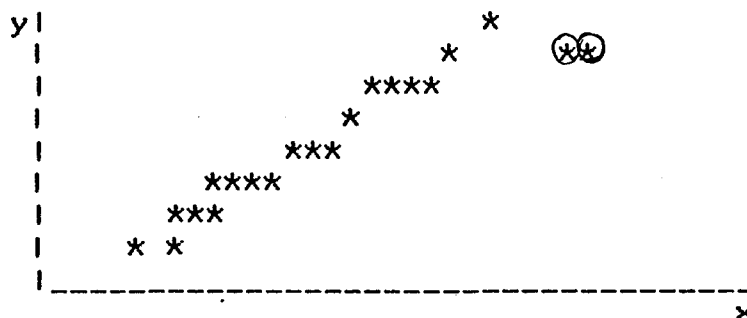


Figura 4.1

Na análise ponto por ponto, provavelmente a importância dos pontos assinalados não será ressaltada devidamente, pois a proximidade de ambos, faz com que ao retirarmos um deles, os parâmetros estimados não sofram mudanças consideráveis devido a presença do outro. Entretanto, parece-nos que o ajuste será afetado de maneira significativa se considerarmos estes dois pontos simultaneamente.

De maneira semelhante, podem ocorrer situações em que um número maior de pontos exerça, conjuntamente, forte influência no ajuste.



Discutiremos a seguir, como quantificar esta influência e os problemas que enfrentaremos para identificar estes conjuntos.

#### 4.1 O DFFIT E D DE COOK PARA O CASO DE MAIS DE UM PONTO

Belsley et al propõe como uma medida natural, uma simples generalização do caso de um ponto, dado por

$$\frac{|b_j - b_j(M)|}{\text{escala}} \quad (4.1.1)$$

para  $j=1, \dots, p$ , onde  $M$  é o conjunto de índices de  $m$  pontos,  $m=2,3,4, \dots$  indicando as linhas a serem retiradas e a escala indica alguma medida apropriada do erro padrão.

Sugerem, ainda, como medidas sumárias, as normas quadráticas tais como

$$[b - b(M)]' [b - b(M)] \quad (4.1.2)$$

ou

$$[b - b(M)]' X' X [b - b(M)] \quad (4.1.3)$$

A mudança no ajuste que ocorre depois de retirados os  $m$  pontos é quantificada modificando (4.1.3) para

$$\text{MDFFIT} = [b - b(M)]' X' (M) X (M) [b - b(M)] \quad (4.1.4)$$

e Bringham[4] mostra que esta expressão pode ser escrita como

$$\text{MDFFIT} = e_M' X_M [X' (M) X (M)]^{-1} X_M' e_M \quad (4.1.5)$$

onde  $e$  é o vetor coluna de resíduos por mínimos quadrados e  $M$  usado como índice, denota uma matriz ou vetor com linhas cujos índices estão contidos em  $M$ .

Esta medida não é exatamente a generalização do caso de um ponto. Também não temos uma distribuição de referência para efeito de comparações, ou ainda, para encontrar pontos de corte. Os autores usam esta medida de forma exploratória e o fazem encontrando um índice em relação ao maior MDFFIT para um conjunto de tamanho  $m$ . Explicando melhor, dado que o tamanho do conjunto é  $m$ , encontra-se o maior MDFFIT com  $m$  pontos e divide-se todos os outros por este. As comparações são portanto feitas em relação ao subconjunto que forneceu maior MDFFIT.

Outras técnicas são mencionadas por Belsley et al, tais como diferenciação, aproximação geométrica e a estatística de Andrews e Pregibon[1].

Cook e Weisberg[11] propõe também medidas que são generalizações do caso de um ponto.

Seja  $M$  o conjunto de  $m$  índices. O modelo para falta de

ajuste em  $m$  pontos pode ser escrito

$$y = X\beta + D\alpha + \epsilon$$

onde  $D$  é uma matriz  $n \times m$  com a  $k$ -ésima coluna  $d_{ik}$ ,  $i$  pertencente a  $M$ , ou seja, uma coluna de zeros, a menos da linha  $i$ , onde o valor é 1 e  $\alpha$  é um vetor de parâmetros desconhecidos.

Sob a suposição de normalidade, a estatística de teste para pontos aberrantes  $\alpha = 0$  é obtido da mesma forma que 3.1.14.

$$\tau_M^2 = \frac{e'(I - H_M)^{-1} e_M (n - p - m)}{((n - p) * s^2 - e_M'(I - H_M)^{-1} e_M)_m} \quad (4.1.6)$$

que, sob a hipótese nula, tem distribuição  $F$  com  $m$  e  $n-p-m$  graus de liberdade. Nesta expressão,  $I$  é a matriz identidade de dimensão  $m$ .

A medida de distância para detecção de conjuntos influentes, denotada por  $D_M$ , é dada por

$$D_M = \frac{(b(H) - b)' X' X (b(H) - b)}{p * s^2} \quad (4.1.7)$$

uma generalização do caso de um ponto. A expressão (4.1.7) pode ser escrita em termos de resíduos e elementos da matriz  $H$  como no caso unidimensional. Para tanto, primeiro expressaremos  $b(H)$  como uma função de  $b$ .

$$\begin{aligned} b(H) &= (X'(H)X(H))^{-1} X(H)y(H) \\ &= ((X'X) - X'_M X_M)^{-1} (X'y - X'_M y_M) \end{aligned} \quad (4.1.8)$$

A inversa em (4.1.8) é calculada usando o teorema 3.1 e assim,

$$\begin{aligned} b(H) &= [(X'X)^{-1} + (X'X)^{-1} X'_M (I - H_M)^{-1} X_M (X'X)^{-1}] [X'y - X'_M y_M] \\ &= b - (X'X)^{-1} X'_M [-(I - H_M)^{-1} X_M b + (I + (I - H_M)^{-1} H_M) y_M] \end{aligned}$$

e usando novamente o teorema 3.1 fazendo  $u = H$ ,  $v = -I$  e  $A = I$ , teremos

$$(I - H_M)^{-1} = I + (I - H_M)^{-1} H_M$$

e assim,

$$b(H) = b - (X'X)^{-1} X'_M (I - H_M)^{-1} e_M$$

Substituindo em (4.1.7),

$$D_M = \frac{e_M'(I - H_M)^{-1} H_M (I - H_M)^{-1} e_M}{p * s^2} \quad (4.1.9)$$

e esta quantidade pode ser interpretada de maneira análoga ao caso de um ponto. O problema é calcularmos esta quantidade para todos os conjuntos com  $m$  pontos, para  $m$  fixado. A quantidade de cálculos seria muito grande a ponto de se tornar impraticável.

Para contornar este problema, Cook e Weisberg sugerem aplicar algumas desigualdades, encontrando um limite superior para  $D_M$  e calculando o valor exato somente quando este limite for muito grande.

Consideremos a decomposição espectral  $H_M = \Gamma \Lambda \Gamma'$ , onde  $\Gamma$  é uma matriz ortogonal de autovetores e  $\Lambda$  é uma matriz diagonal de autovalores  $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq 1$ .

Visto que  $\lambda_m / (1 - \lambda_m)^2 \geq \lambda_l / (1 - \lambda_l)^2$ ,  $l=1, 2, \dots, m$ , o primeiro limite superior para  $D_M$  seria

$$D_M \leq \frac{1}{pks^2} * \frac{\lambda_m}{(1 - \lambda_m)^2} * \sum (\gamma_i' e_M)^2 = \frac{1}{pks^2} * \frac{\lambda_m}{(1 - \lambda_m)^2} * e_M' \Gamma \Gamma' e_M$$

e como  $\Gamma \Gamma' = 1$ ,

$$D_M \leq \frac{1}{pks^2} * \frac{\lambda_m}{(1 - \lambda_m)^2} * \sum_{i \in M} e_i^2 \quad (4.1.10)$$

Para (4.1.10) ser útil, devemos substituir  $\lambda_m$  por uma aproximação de tal forma que possa ser calculado sem obter  $H_M$ . Supondo que  $\text{tr}(H_M) < 1$ , a aproximação mais simples é  $\lambda_m < \text{tr}(H_M)$ . Assim,

$$D_M \leq \frac{1}{pks^2} * \frac{\text{tr}(H_M)}{(1 - \text{tr}(H_M))^2} * \sum_{i \in M} e_i^2$$

ou equivalentemente,

$$D_M \leq \frac{1}{pks^2} * \frac{\sum_{i \in M} h_i}{(1 - \sum_{i \in M} h_i)^2} * \sum_{i \in M} e_i^2 \quad (4.1.11)$$

Para os subconjuntos com  $\text{tr}(H_M) \geq 1$ , é necessário calcularmos  $H_M$ .

Para  $m$  fixado, seja  $T = \max(\sum_{i \in M} h_i)$  e  $R^2 = \max(\sum_{i \in M} e_i^2)$ , onde  $M$  varia sobre todos os conjuntos de tamanho  $m$ . Dois limites superiores para (4.1.11) são então

$$D_M \leq \frac{\text{tr}(H_M)}{(1 - \text{tr}(H_M))^2} * \frac{R^2}{p \times s^2} \quad (4.1.12)$$

e se  $T < 1$ ,

$$D_M \leq \frac{1}{p \times s^2} * \frac{T}{(1 - T)^2} \sum_{i \in M} e_i^2 \quad (4.1.13)$$

e estas duas últimas combinadas,

$$D_M \leq \frac{T}{(1 - T)^2} * \frac{R^2}{p \times s^2} \quad (4.1.14)$$

Claramente, (4.1.11)  $\leq$  (4.1.12)  $\leq$  (4.1.14) e (4.1.11)  $\leq$  (4.1.13)  $\leq$  (4.1.14).

O procedimento proposto por Cook e Weisberg para se encontrar todos os subconjuntos relevantes com  $m$  fixado, pode ser baseado nestas aproximações e se resume no seguinte: as desigualdades podem ser aplicadas aos subconjuntos com  $\text{tr}(H_M) < 1$  na ordem (4.1.14) então (4.1.13) ou (4.1.12) e finalmente (4.1.11). O cálculo exato de  $D$  se faz necessário se (4.1.11) for grande.

## 4.2 O PROBLEMA EM ENCONTRAR SUB-CONJUNTOS

Até agora, todas as medidas propostas supunham um conjunto de pontos com um número fixado de elementos  $m$  ou um conjunto específico com  $m$  pontos.

Se determinarmos um conjunto específico com  $m$  pontos, o cálculo das medidas MDFFIT e  $D_M$  serão reduzidos. Entretanto, se decidirmos que o tamanho do conjunto será  $m$ , teremos combinação de  $n$  pontos tomados  $m$  a  $m$  conjuntos possíveis e uma investigação completa como foi realizada no caso de um ponto, se torna impraticável.

Mas, para uma análise completa, é necessário considerarmos subconjuntos de vários tamanhos, ou seja, para  $m=1,2,\dots$ , o que torna ainda mais proibitiva a determinação das medidas. Portanto, a identificação de conjuntos de pontos que eventualmente possa ter influência no ajuste, torna-se muito mais complicada.

Se considerarmos todos os conjuntos possíveis de serem formados com  $n$  observações, excluindo os conjuntos com apenas uma observação, o conjunto vazio e o conjunto com todas as observações, teremos

$$\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n-1} = 2^n - 2 - n$$

conjuntos possíveis.

é claro, entretanto, que estamos procurando conjuntos de pontos que tenham comportamento distinto em relação à maioria dos dados e, desta forma, não faz sentido estudar conjuntos com um número grande de pontos. Por exemplo, se um conjunto com 80% dos pontos tem comportamento distinto dos demais, é claro que o conjunto complementar é que tem comportamento atípico.

Desta forma, poderíamos sugerir estudar conjuntos com até 10% ou 20% dos dados. Ainda assim, mesmo em conjunto de dados pequeno, o número de subconjuntos é grande e este número cresce violentamente com o número de observações aumentando. Por exemplo, suponhamos que temos 20 observações e queremos estudar conjuntos com até 10% dos dados. Então devemos tomar todos os con-

juntos com 2 pontos e teremos  $\binom{20}{2} = 190$  conjuntos e se decidirmos tomar conjuntos com até 20% dos dados, teremos  $\binom{20}{2} + \binom{20}{3} + \binom{20}{4} = 6175$  :

Fica evidente, portanto, a impraticabilidade de se obter medidas para todos os conjuntos possíveis.

Faz-se necessário, portanto, procurar formas para nos fornecer informações que permitam sugerir conjuntos de pontos que possam exercer influência.

Em regressão linear simples, apenas um gráfico de  $x$  versus  $y$  é suficiente para que identifiquemos tais conjuntos, como podemos ver pela figura 4.1. Em regressão linear múltipla, poderíamos sugerir o gráfico de regressão parcial. Entretanto, estes gráficos ainda podem mascarar pontos e se tivermos muitas variáveis regressoras, o estudo conjunto de todos os gráficos de regressão parcial torna-se um pouco confuso.

Belsley et al [3] sugerem escolher o subconjunto  $M$  de pontos potencialmente influentes relaxando os pontos de corte de  $h_i$ ,  $\tau_i$ ,  $D_i$ , e  $DFFITs_i$ , e pelo exame dos gráficos de regressão parcial.

Cook e Weisberg [11] propõe colocar os pontos em ordem decrescente com relação aos  $e_i$  e  $h_i$  para se encontrar subconjuntos com  $m$  fixado. Desta forma, subconjuntos que são, provavelmente, mais influentes serão considerados primeiro. Calcula-se então os limites superiores para  $D_M$  conforme (4.1.11) a (4.1.14) e quando estes limites forem suficientemente pequenos, os subconjuntos restantes podem ser desconsiderados. Ainda assim, como ilustram os próprios autores em [11], a quantidade de desigualdades calculadas mesmo num conjunto pequeno de observações (20), é grande.

### 4.3 TÉCNICA GRÁFICA DE DENBY E MALLOW'S

Abordaremos nesta secção, as técnicas gráficas propostas por Denby e Mallows[13], para a detecção de pontos influentes em regressão robusta, e em particular, como estas técnicas podem auxiliar na determinação de subconjuntos que possam exercer influência no ajuste.

A classe de estimadores considerada pelos autores é a denominada m-estimadores, particularmente a proposta por Huber[17], descrita na secção 2.4 do capítulo 2. O trabalho não tem por objetivo encontrar melhores estimadores de  $\beta$  e  $\sigma^2$ , e sim, exibir como os m-estimadores  $b(c)$  e os resíduos dependem do parâmetro escala  $c$  em (2.4.2).

Para cada valor não negativo de  $c$ , podemos escrever a expressão (2.4.3) na forma

$$\sum_{i=1}^n x_{im} * e_i(c) * w(c, e_i) \quad m=1, 2, \dots, p \quad (4.3.1)$$

onde o resíduo  $e_i(c)$  é

$$e_i(c) = y_i - x_i b(c) \quad (4.3.2)$$

e

$$w(c, t) = \psi(t)/t = \begin{cases} 1 & |t| < c \\ c/|t| & |t| > c \end{cases} \quad (4.3.3)$$

Note que, se  $c$  for muito grande, as equações (2.4.3) ou (4.3.1) são as equações de mínimos quadrados usuais.

Sejam  $b$  os estimadores de mínimos quadrados de  $\beta$ , e  $e$  os resíduos correspondentes

$$c_1 = \max_i |e_i|$$

Assim, para  $c > c_1$ , a solução de mínimos quadrados será mantida.

Para a escolha de algum conjunto  $C$  de valores  $c < c_1$ , digamos  $C = (c_1, c_2, \dots, c_g)$ , podemos calcular os estimadores  $b(c_i)$  e respectivos resíduos, usando o método de mínimos quadrados ponderados repetidamente, ajustando as ponderações conforme (4.3.3).

Dado o conjunto  $C$  e calculadas as estimativas para cada elemento de  $C$ , a idéia é verificar como os parâmetros e os resíduos se comportam com o parâmetro escala  $c$  decrescendo.

Os elementos de  $C$  são calculados sequencialmente, a partir da obtenção do último  $c$ , tomando como primeiro ponto o maior resíduo de mínimos quadrados em valor absoluto, ou seja,  $c_1$ .

Suponhamos que o índice do maior resíduo de mínimos quadrados em valor absoluto seja  $u_1$  e  $s_{u_1}$  o seu sinal, de maneira que

$$e_{u_1} = s_{u_1} c_1 .$$

Para algum valor de  $c$ , menor que  $c_1$ , as equações de estimação em (2.4.3) são

$$\sum_{i \neq u_1} x_{im} (y_i - x_i' b(c)) + c s_{u_1} x_{u_1 m} \quad m=1, \dots, p \quad (4.3.4)$$

e para esta variação dos valores de  $c$ , teremos

$$b(c) = (X'(u_1)X(u_1))^{-1} (X'(u_1)Y(u_1) + c s_{u_1} x_{u_1}) \quad (4.3.5)$$

e portanto,  $b(c)$  é linear em  $c$  e, assim, cada um dos resíduos

$$e_j(c) = y_j - x_j' b(c) \quad (4.3.6)$$

também o são.

Com isso, é possível encontrarmos o valor de  $c$  para o qual o segundo maior resíduo se torna igual a  $c$ , resolvendo as  $2 \times n - 2$  equações

$$\begin{aligned} e_j(c) &= c & i \neq u_1 \\ e_j(c) &= -c & i \neq u_1 \end{aligned} \quad (4.3.7)$$

O valor de  $c$  é o maior destas raízes que não seja maior que  $c_1$ . Isto define o valor  $c_2$  e o índice correspondente  $u_2$ . Com isso, temos o começo de uma indução e uma sequência monotônica  $c_1, c_2, \dots$  de valores de  $c$ , determinada de maneira que para cada um destes valores, um dos resíduos é igual a  $c$  e entre estes valores, cada um dos coeficientes estimados e cada um dos resíduos, é uma função linear de  $c$ .

De posse dos valores de  $c$  e respectivos parâmetros estimados e resíduos, recomenda-se os gráficos de:

- i) elementos de  $b(c_g)$  versus  $c_g, g=1, 2, \dots, G$
- ii) os resíduos  $(e_j(c_g))$  versus  $c_g, g=1, 2, \dots, G$

Juntando os sucessivos pontos de i) e ii), obtemos os gráficos de  $b(c)$  e  $(e_j(c_g))$  versus  $c$  para  $c_g < c < c_1$ .

Não há dificuldade em apresentarmos todos os resíduos em um mesmo gráfico, mas para apresentarmos os gráficos em i) num único gráfico, será necessário reduzi-los a uma mesma escala para que possamos compará-los diretamente. Denby e Mallows, escolheram padronizar com o intervalo interquartis. Desta forma, nos gráficos em i), teremos dispostos

$$\beta_j^*(c) = \beta_j(c) (x_j' - x_j) , \quad j=2, \dots, p,$$

onde  $x_j^*$  e  $x_j^*$  são os quartis superior e inferior de  $x_{1j}, \dots, x_{nj}$ . Para padronizarmos a constante  $\beta_1$ , fazemos

$$\beta_1^*(c) = \beta_1(c) - \beta_1(c_1),$$

pois o nosso interesse maior é a variação deste em função de  $c$ .

**EXEMPLO 4.1** Consideremos os dados do exemplo 3.2. Denby e Mallows utilizaram este mesmo exemplo e para efeito de ilustração, reproduziremos aqui os resultados encontrados pelos mesmos. O modelo proposto é

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Na figura 4.2, temos o gráfico de resíduos versus o parâmetro escala  $c$ , onde os 21 resíduos são identificados pelas letras A a U da seguinte maneira: 'A' corresponde ao resíduo da observação 1, 'B' ao resíduo da observação 2 e assim por diante.

Observe que o estado de cada resíduo para cada valor de  $h$  pode ser identificado através das linhas através da origem a +45 e a -45 graus. Estas linhas interceptam os traços de resíduos num ponto, onde o resíduo é igual ao parâmetro  $c$ , conforme as equações em (4.3.7). Para cada valor de  $c$ , os resíduos que estão fora das linhas de 45 graus foram atenuados, e aos resíduos entre as linhas de 45 graus foram atribuídos pesos integrais.

Da figura 4.2, podemos concluir que os pontos A, C, D e U tem resíduos de mínimos quadrados grandes e aumentam à medida que  $c$  diminui e o restante permanece amontoado e se concentram mais ainda com  $c$  diminuindo. Os quatro pontos A, C, D e U, correspondem aos pontos 1, 3, 4 e 21, citados por Daniel e Wood [12] como possíveis pontos aberrantes.

A figura 4.3 mostra o gráfico de  $b(c)$  versus o parâmetro escala  $c$ , onde a letra 'A' corresponde ao parâmetro  $\beta_1$ , 'B' ao parâmetro  $\beta_2$ , 'C' ao parâmetro  $\beta_3$  e 'D' ao parâmetro  $\beta_4$ . Nesta figura, podemos notar que os parâmetros  $\beta_2$  e  $\beta_4$  não sofrem mudanças consideráveis com  $c$  decrescendo enquanto que os parâmetros  $\beta_1$  e  $\beta_3$  são instáveis, particularmente  $\beta_3$  que mostra uma trajetória descendente bastante acentuada com  $c$  decrescendo, mantendo esta tendência mesmo para valores pequenos de  $c$ . Este distúrbio pode ser devido à escolha da função de Huber que não elimina os pontos que resultam em resíduos grandes, atribuindo a estes apenas pesos menores, de maneira que os pontos considerados aberrantes ainda continuem a exercer influência sobre os parâmetros, mesmo depois de serem atenuados. Observamos ainda que o parâmetro  $\beta_4$  é pequeno para todo  $c$ . #



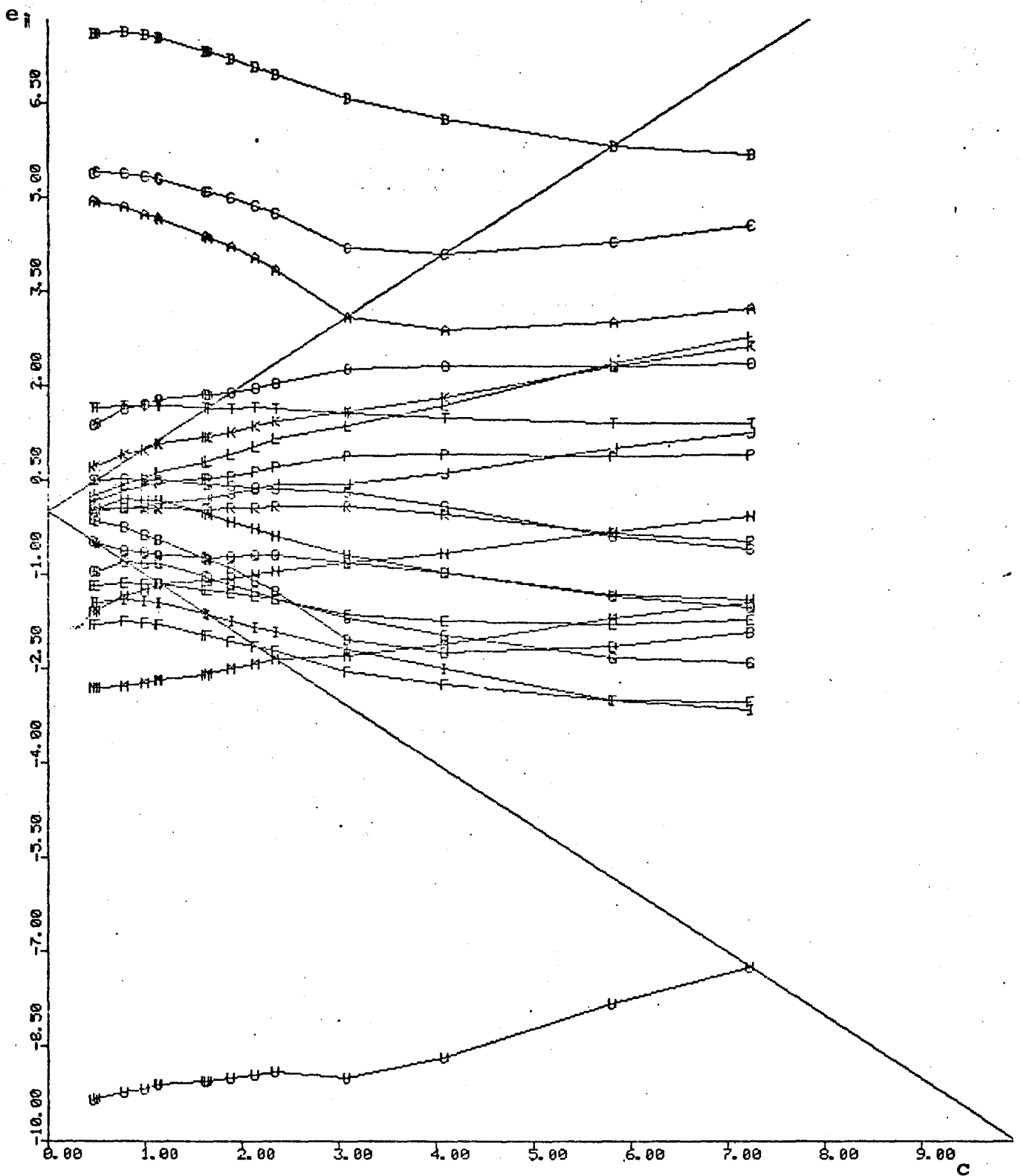


Figura 4.2 Gráfico de resíduos x c - Dados de amônia não absorvida

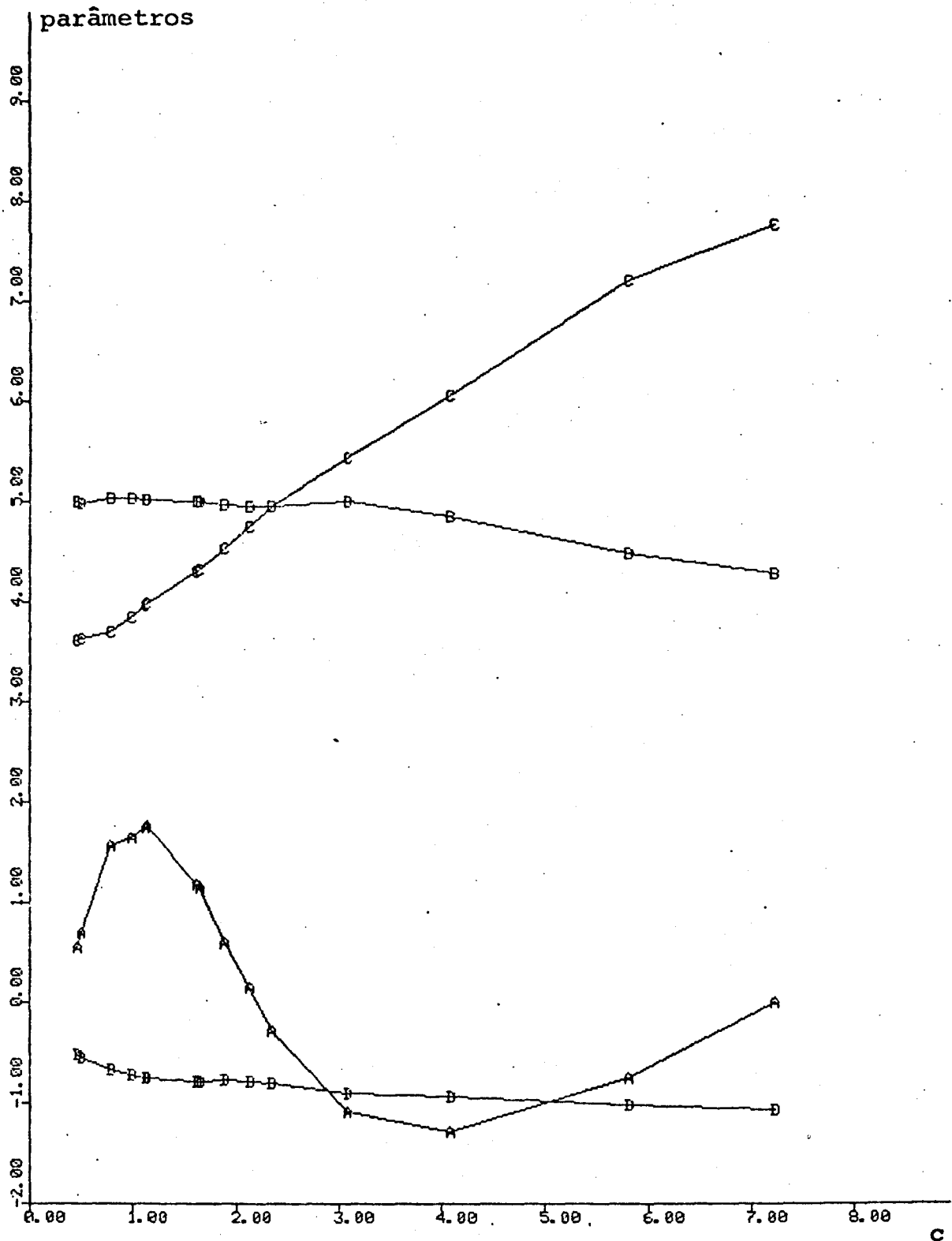


Figura 4.3 Gráfico de parâmetros x c - Dados de amonia não absorvida

Estes gráficos, como foram propostos, são eficientes para a determinação de pontos aberrantes e conforme os autores, ainda possuem um aspecto interessante que podem auxiliar na determinação de subconjuntos de pontos que exercem influência no ajuste. Este fato pode ser notado no exemplo anterior, observando os gráficos de resíduos versus o parâmetro  $c$  e o gráfico de parâmetros versus o parâmetro  $c$ , conjuntamente. Observemos primeiramente, na figura 4.2, a linha correspondente ao resíduo do ponto 1, denotado pela letra 'A', na abscissa 3.09, correspondente a  $c(4)$ . Esta linha sofre uma inflexão forte quando o ponto 1 é atenuado e a mesma tendência é observada nas linhas B,C e talvez ainda a linha D, correspondentes aos resíduos das observações 2,3 e 4 respectivamente. Nesta mesma abscissa,  $c(4)=3.09$ , na figura 4.3, a linha 'A' correspondente ao parâmetro  $\beta_1$ , sofre também uma mudança de tendência acentuada e a linha 'C' correspondente ao parâmetro  $\beta_3$ , mantém a sua trajetória decrescente. Podemos concluir, através destas observações, que os pontos 1,2,3 e possivelmente 4 exercem forte influência no parâmetro  $\beta_1$ , e talvez também no parâmetro  $\beta_3$ , quando o parâmetro de atenuação  $c$  é 3.09. Esta influência, presumivelmente, deve se estender ao caso de mínimos quadrados e neste sentido, exploraremos os pontos sugeridos por estes gráficos, encontrando as medidas de influência múltipla,  $D_M$  para o caso de mínimos quadrados.

Quadro 4.1 Medidas de influência múltipla - dados de amonia não absorvida.

Subconjunto	D
(1,2,3,4)	7.9789
(1,2,3)	3.2034
(1,2,4)	0.6464
(1,3,4)	2.1203
(2,3,4)	0.4188
(1,2)	0.0790
(1,3)	1.1140
(1,4)	0.5232
(2,3)	0.0612
(2,4)	0.1028
(3,4)	0.4183

No quadro acima, temos os valores do D de Cook calculado para o conjunto dos quatro pontos e de todos os subconjuntos menores envolvendo estes mesmos quatro pontos. O valor de comparação é 0.2604, o mesmo encontrado para o caso de um ponto, no exemplo 3.2. Vemos que o subconjunto formado pelas observações 1,2,3 e 4 exerce forte influência no ajuste e dos subconjuntos menores, todos aqueles formados com 3 destes pontos são influentes, notadamente o subconjunto (1,2,3). Sabemos que os dados foram tomados em 21 dias consecutivos e é importante notar que as observações pertencentes ao subconjunto mais influente é justamente dos primeiros quatro dias. Este fato pode ser de grande interesse e seria recomendável um estudo conjunto com o pequisa-

dor responsável pelo experimento.

Apesar de, neste exemplo, termos conseguido isolar um subconjunto com quatro pontos exercendo forte influência no ajuste e ilustrar como estes gráficos podem sugerir tais subconjuntos, há situações em que certamente deixarão de ser eficientes. Consideremos uma situação em que um ponto ou um subconjunto de pontos seja muito influente e nestes casos, em geral, o(s) resíduo(s) correspondente(s) é(são) pequeno(s). Como estamos trabalhando com o estimador de Huber onde os resíduos determinam a constante a partir da qual serão atribuídos pesos menores e os pontos que serão atenuados primeiro são aqueles que tem resíduos maiores, é possível e bastante provável que os pontos que são influentes e têm resíduos pequenos não sejam atenuados no intervalo considerado pelo método proposto.

Ilustraremos esta situação através do exemplo seguinte, que deixará evidente a necessidade de certos ajustes para o aproveitamento destas técnicas gráficas como uma alternativa para a detecção de subconjunto de pontos influentes.

**EXEMPLO 4.2** Consideremos um conjunto de dados hipotéticos. Estes números foram gerados da seguinte maneira: os oito primeiros pontos foram encontrados, por um processo aleatório, sobre a circunferência de raio 0.1, com centro no ponto (0.5,0.5) e aos pontos 9 e 10, atribuímos os valores (0.0,0.0) e (1.0,1.0) respectivamente. Claramente, isto fará com que os pontos 9 e 10 sejam de grande importância no ajuste, que podemos confirmar com um simples exame da figura 4.4.

**Quadro 4.2** Dados gerados sobre a circunferência de raio 0.1 e centro (0.5,0.5).

caso	x	y
1	0.5935	0.4645
2	0.4494	0.5863
3	0.5893	0.4550
4	0.5279	0.5960
5	0.5095	0.5995
6	0.5847	0.5531
7	0.4749	0.5968
8	0.5017	0.6000
9	0.0000	0.0000
10	1.0000	1.0000

Os parâmetros estimados por mínimos quadrados são:

$$b_1 = 0.0609$$

$$b_2 = 0.9256$$

e o coeficiente de correlação múltipla  $R^2 = 0.8421$ .



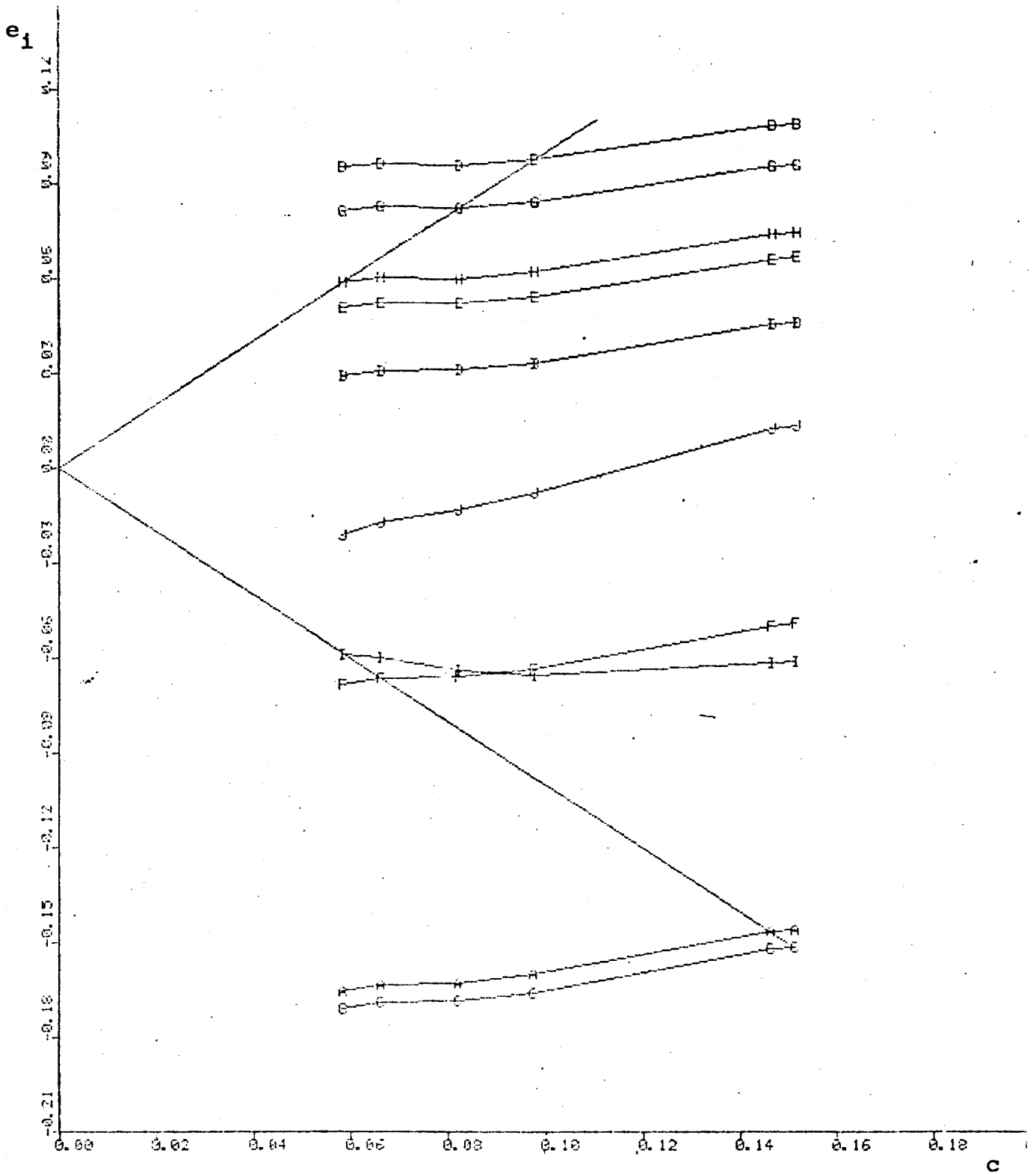


Figura 4.5 Gráfico de resíduos x c - exemplo 4.2

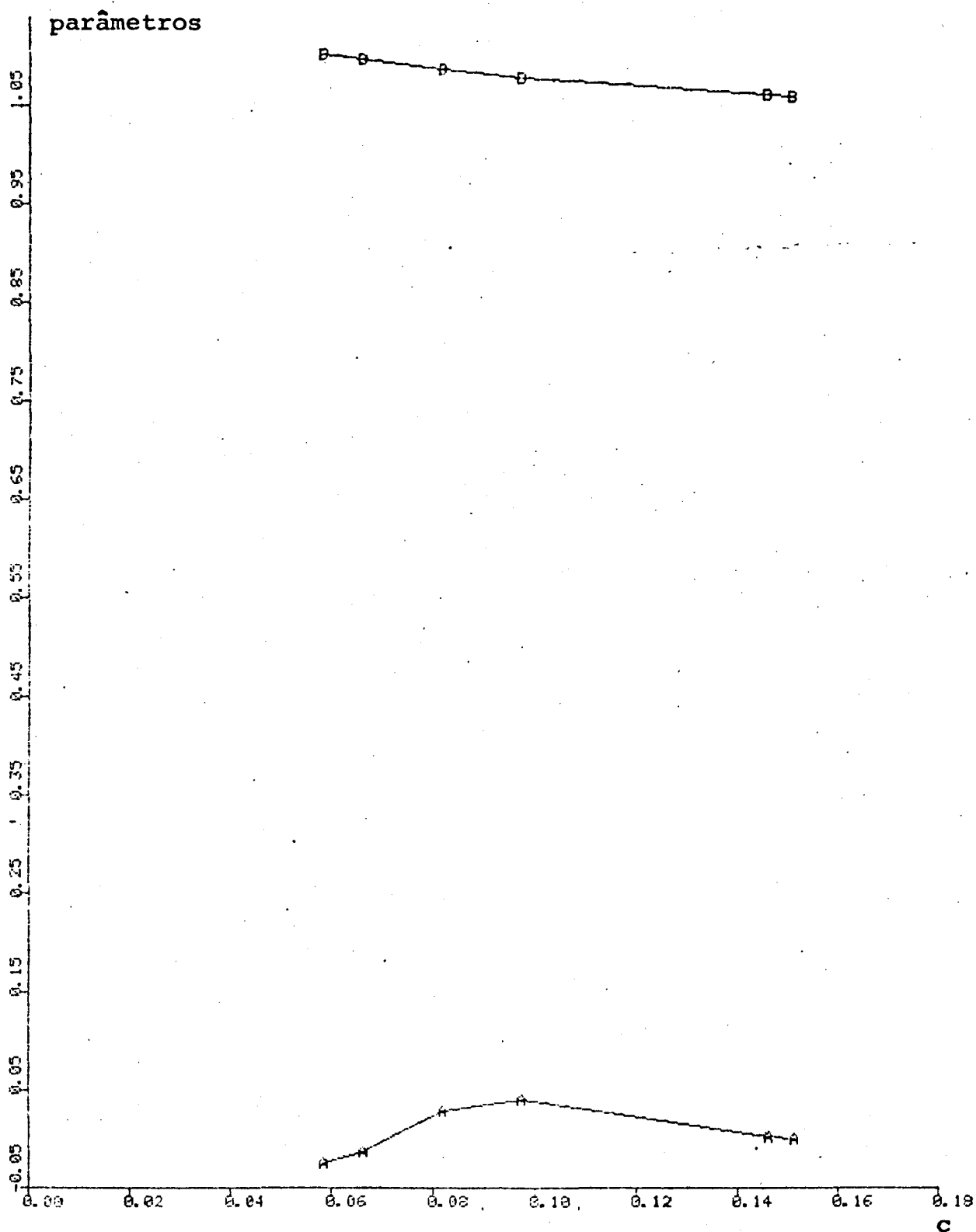


Figura 4.6 Gráfico de parâmetros x c - exemplo 4.2

Ainda na figura 4.4, podemos notar que a retirada do ponto 9 não deve causar mudança considerável no ajuste, o mesmo acontecendo com o ponto 10. Isto pode ser constatado pelo quadro 4.3, onde temos as medidas de influência de cada ponto. Entretanto, parece evidente que se retirarmos os pontos 9 e 10 simultaneamente, o ajuste sofrerá uma mudança significativa.

No quadro 4.3, temos as medidas de influência para estes dados e vemos que individualmente, segundo D de Cook, os pontos 1,3 e 9 são influentes e segundo DFFITS, o ponto 9. Os valores de comparação estão dados abaixo do quadro.

Nas figuras 4.5 e 4.6, temos os gráficos de diagnóstico e como era de se esperar, os pontos 9 e 10 por terem resíduos pequenos, não chegaram a serem atenuados fazendo com que não consigam ser detectados por esta técnica, da maneira que foi sugerida. #

Huber[17] menciona este mesmo problema e para que pontos influentes tenham menor peso na determinação do modelo, sugere o uso do resíduo dividido pela raiz quadrada de  $1-h$  onde  $h$  é o elemento da diagonal da matriz chapéu correspondente. Desta forma, a função em (2.4.2), fica

$$\sum_{i=1}^n \rho(y_i - x_i \beta)$$

onde

$$\rho(t) = \begin{cases} t^2/2 & |t| < c \times (1-h_i)^{1/2} \\ c \times |t| - c^2/2 & |t| > c \times (1-h_i)^{1/2} \end{cases} \quad (4.3.8)$$

e as equações em (4.3.7),

$$\begin{aligned} e_i(c) &= c \times (1 - h_i)^{1/2} & i \neq u_j \\ e_i(c) &= -c \times (1 - h_i)^{1/2} & i \neq u_j \end{aligned} \quad (4.3.9)$$

Com isso, a posição de cada ponto no espaço gerado pelas colunas da matriz de  $X$  é considerada, fazendo com que a pontos mais distante do centóide dos dados sejam atribuídas ponderações maiores. Quanto mais distante do centóide dos dados estiver um ponto, mais próximo de um será o elemento correspondente da diagonal da matriz chapéu e portanto  $1-h$  será pequeno e dividindo o resíduo deste ponto pela raiz quadrada desta quantidade, faz com que este ponto adquira importância maior.

**EXEMPLO 4.2** Continuação. Voltando ao exemplo anterior e utilizando esta sugestão, construímos os gráficos de diagnóstico, figuras 4.7 e 4.8, notamos que ainda não é possível identificar os pontos 9 e 10 como conjuntamente influentes. Embora estes pontos tenham os elementos da diagonal da matriz chapéu grandes, não são suficientemente grandes para que façam os resíduos correspondentes crescerem após o ajuste proposto em (4.3.9), a ponto de serem



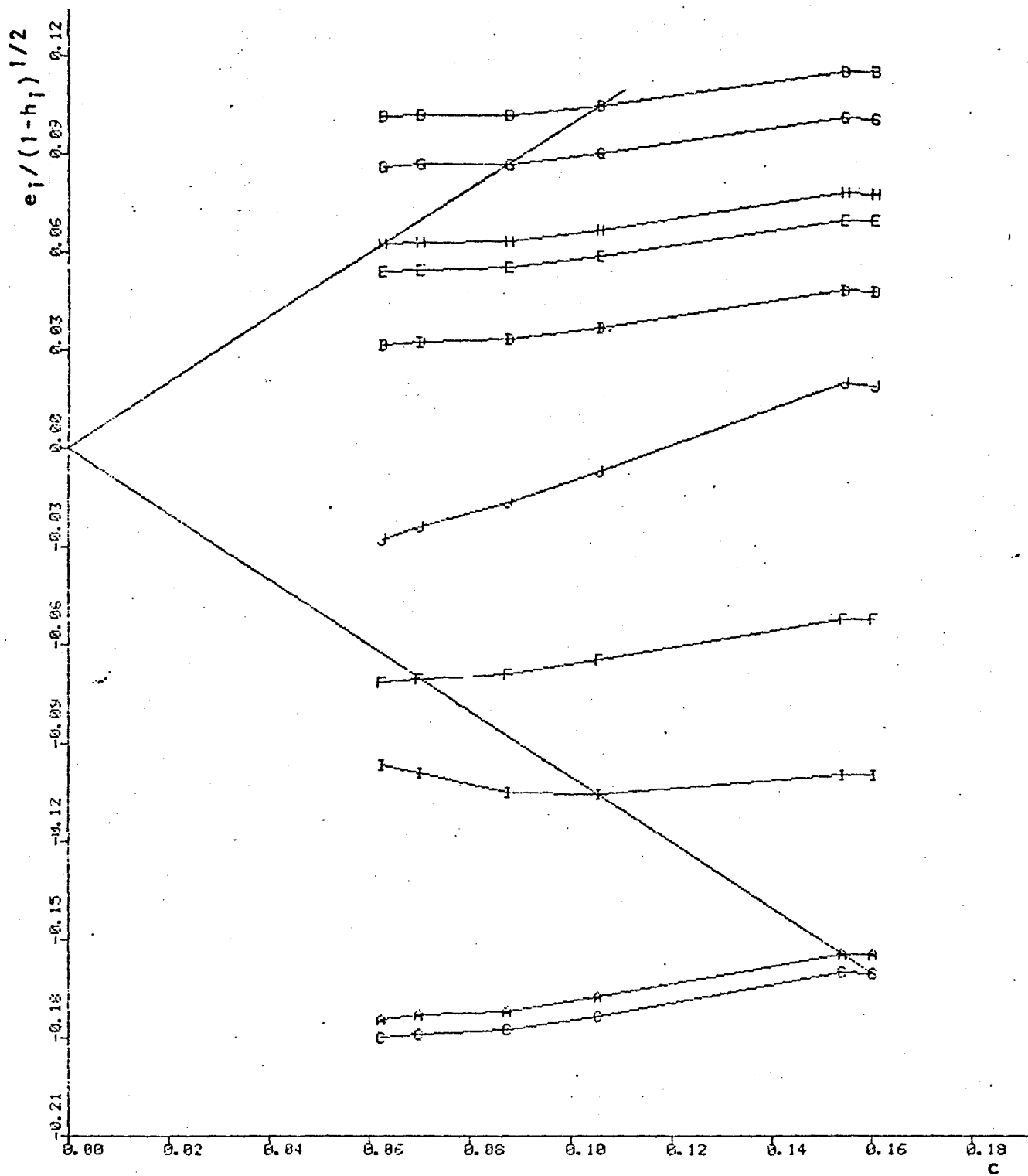


Figura 4.7 Gráfico de resíduos x c - exemplo 4.2(cont.)

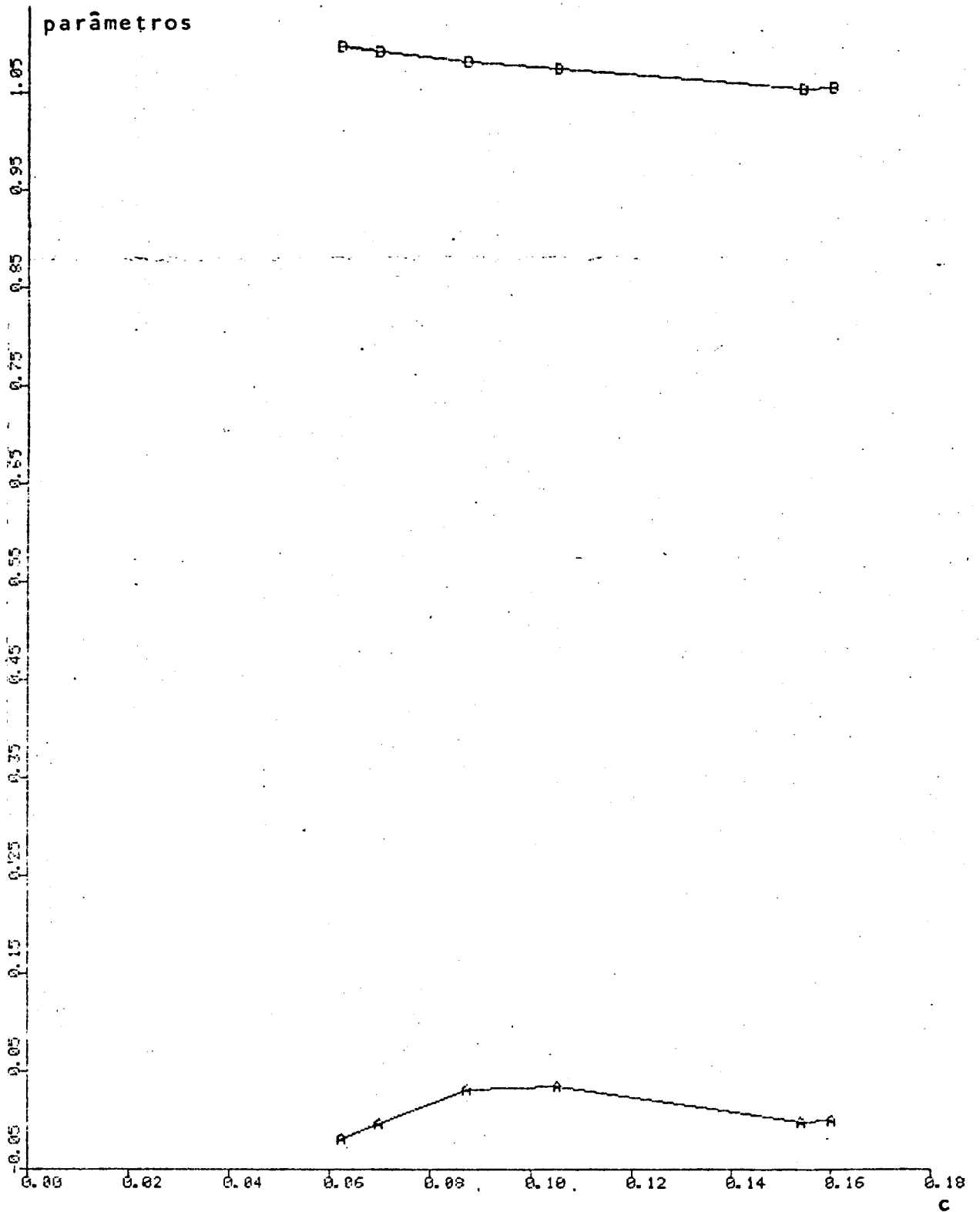


Figura 4.8 Gráfico de parâmetros x c - exemplo 4.2 (cont.)

atenuados. Neste caso em particular, podemos ver na figura 4.7 o ponto 9, rotulado pela letra 'I' cruzar a linha de referência de -45 graus na abscissa  $c=0.11$  aproximadamente. Entretanto, este ponto não foi atenuado nesta abscissa pois o ponto 2, rotulado 'B', apresentou resíduo praticamente igual em valor absoluto ao do ponto 9, sendo o resíduo do primeiro um pouco maior. Assim, ao atenuar o ponto 2, o resíduo do ponto 9 na próxima etapa passou a ser maior em valor absoluto que o valor de  $c$  da etapa anterior e como o valor de  $c$  consecutivo deve ser menor ao anterior, a linha correspondente ao ponto 9 cruza a linha de -45 graus sem que este ponto tenha sofrido corte exatamente sobre ele.

A sequência de entrada dos pontos no conjunto  $C$  é: 3,1,2,7,6 e B, o que corresponde no gráfico de resíduos às letras C,A,B,G,F e H. #

Para que o método gráfico em estudo seja eficiente, seria necessário que pontos com comportamento particular fossem atenuados no início do processo de indução de forma que possamos ver a variação que sofre seu resíduo e a perturbação que este pode eventualmente causar em outros. Para tanto, tentaremos dar maior importância aos pontos cujo  $h$  seja grande. Foram realizados estudos utilizando o resíduo dividido por  $1-h$  correspondente elevado à potência 1 e 2. Ou seja, generalizando a expressão em (4.3.8) e (4.3.9),

$$\sum_{i=1}^n \rho(y_i - x_i' \beta)$$

onde

$$\rho(t) = \begin{cases} t^2/2 & |t| < c \times (1-h_i)^k \\ c \times |t| - c^2/2 & |t| > c \times (1-h_i)^k \end{cases} \quad (4.3.10)$$

e as equações em (4.3.7),

$$\begin{aligned} e_i(c) &= c \times (1 - h_i)^k & i \neq u_1 \\ e_i(c) &= -c \times (1 - h_i)^k & i \neq u_1 \end{aligned} \quad (4.3.11)$$

e tomando  $k=1$  e  $k=2$ .

**EXEMPLO 4.2** Continuação. Nas figuras 4.9 e 4.10, temos os gráficos de diagnóstico resíduos versus  $h$  e parâmetros versus  $h$ , respectivamente, com a potência de  $1-h$ ,  $k=1$ , ou seja, os resíduos aqui são divididos por 1 menos o elemento da diagonal da matriz chapéu correspondente. Aqui, a sequência de entrada é: 3,1,9,7,10 e 2, que corresponde às letras C,A,I,G,J e B na figura 4.11. Devido a uma variação grande de escala, a figura 4.9 fica difícil de ser interpretada, mas nos fornece a noção de variabilidade que a atenuação de um ponto está causando. A figura 4.11 é a mesma da 4.9, com escala maior e nesta podemos ver que a atenuação do ponto 9, rotulado 'I' é que causa a perturbação em todos os pontos, notadamente no ponto 10, rotulado 'J'. Nesta mesma abscissa,

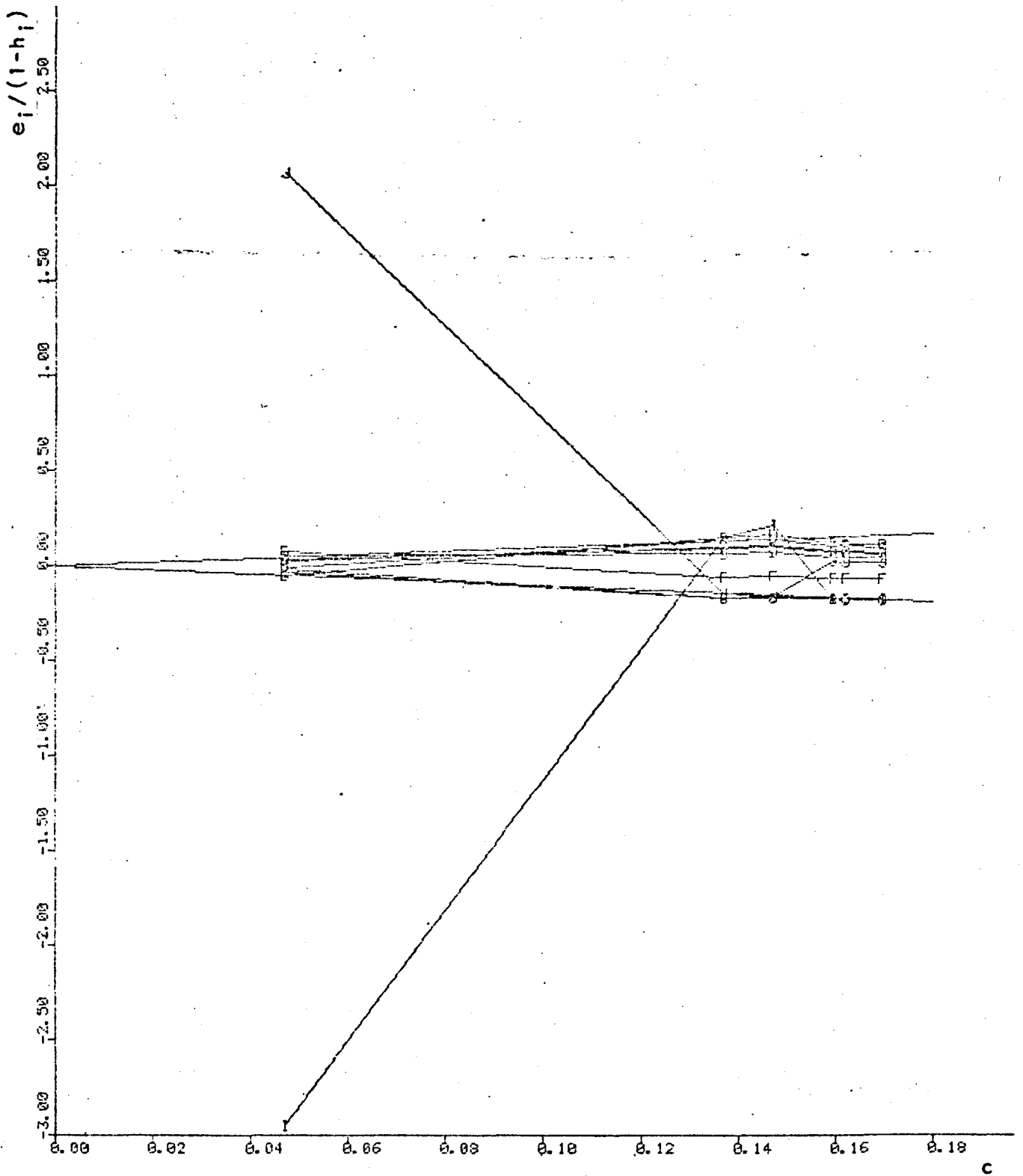


Figura 4.9 Gráfico de resíduos x c - exemplo 4.2(cont.)

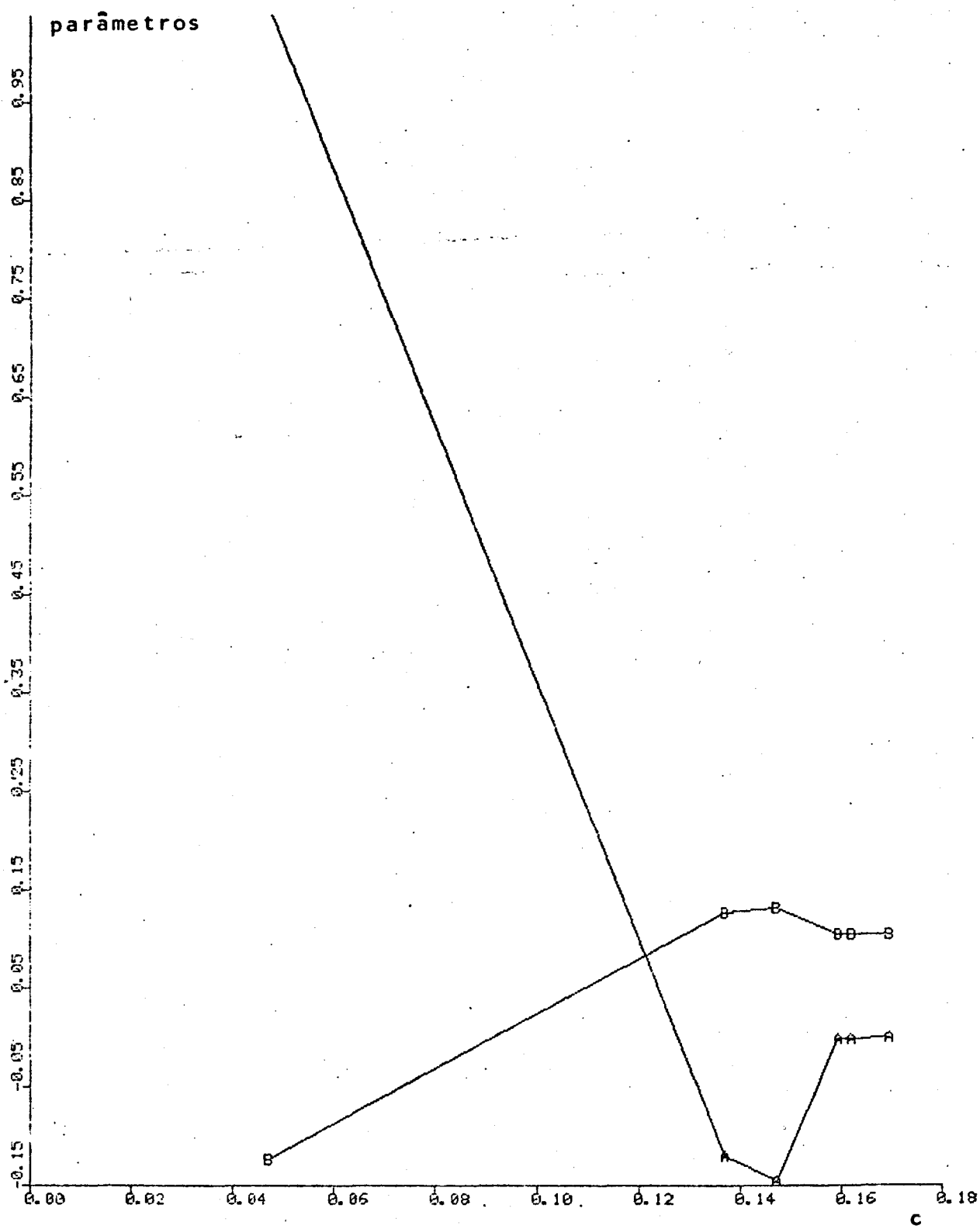


Figura 4.10 Gráfico de parâmetros x c - exemplo 4.2(cont)

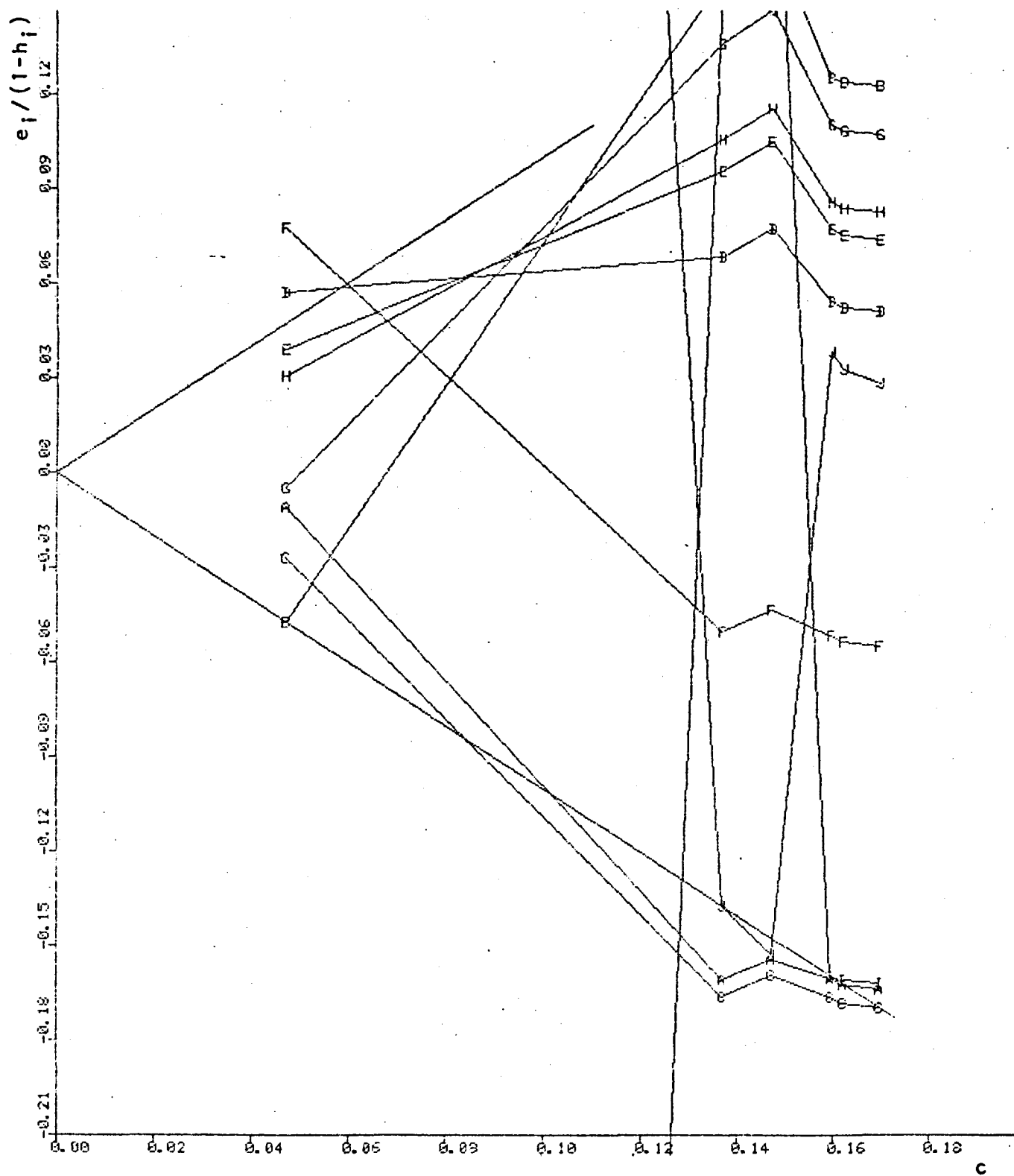


Figura 4.11 Gráfico de resíduos x c - exemplo 4.2(cont.)

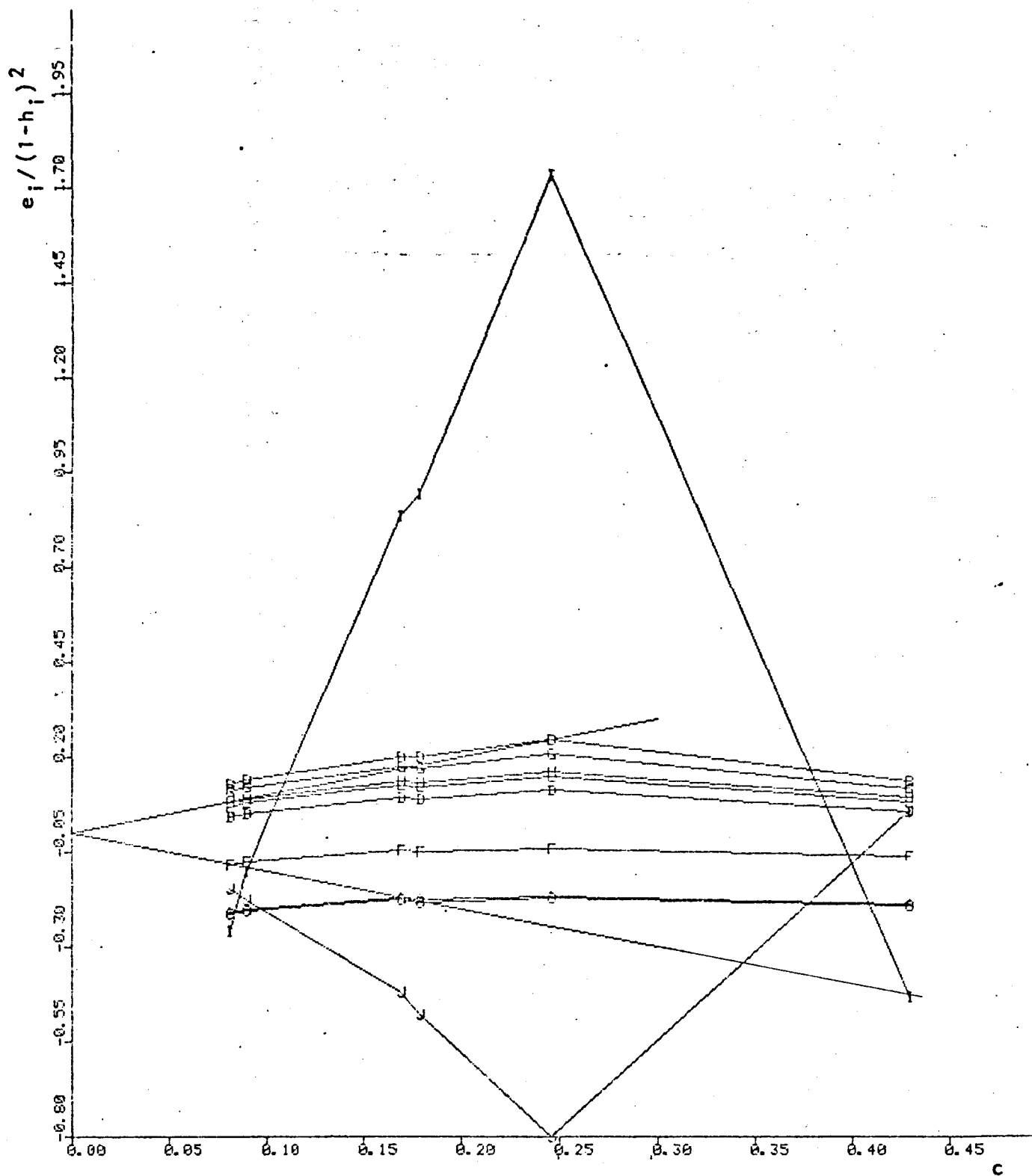


Figura 4.12 Gráfico de resíduos x c - Exemplo 4.2(cont.)

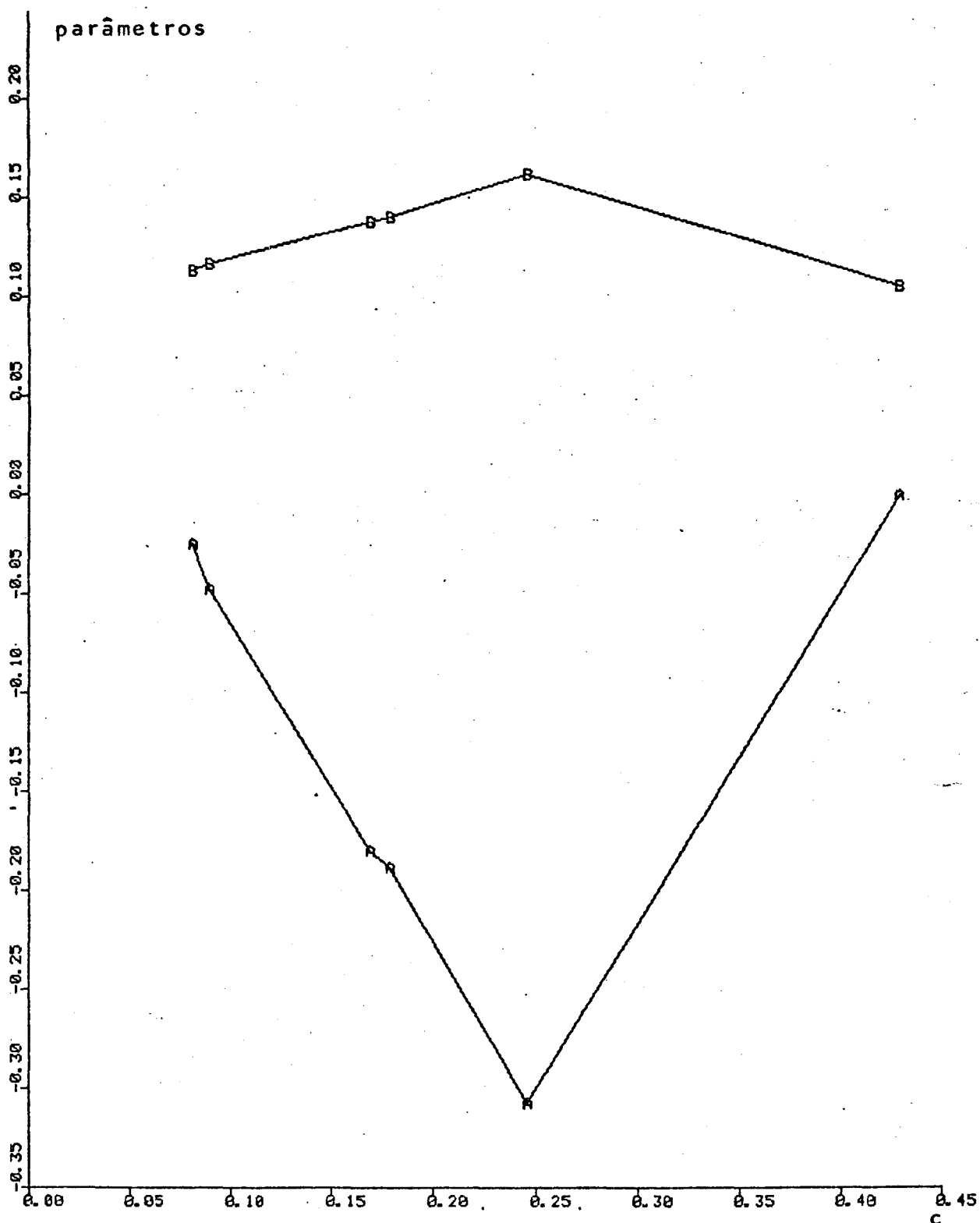


Figura 4.13 Gráfico de parâmetros x c - exemplo 4.2(cont.)



aproximadamente 0.16 na figura 4.10, a linha 'A' correspondente ao parâmetro  $\beta_1$ , sofre forte desvio para valor negativo e a linha 'B' correspondente ao parâmetro  $\beta_2$ , uma inflexão moderada para um valor positivo maior. Ainda, na sequência, o ponto 10, rotulado 'J' é atenuado, causando novamente grandes perturbações, notadamente no ponto 9 e também nos parâmetros, conforme figura 4.10. Podemos concluir, portanto, que os pontos 9 e 10 são conjuntamente influentes, como era de se esperar.

Nas figuras 4.12 e 4.13, temos os gráficos de diagnóstico com  $k=2$ , ou seja, aqui os resíduos são divididos por  $1-h$  ao quadrado e estas mostram de maneira mais clara ainda a influência dos pontos 9 e 10. A sequência de entrada dos pontos é 9,2,3,1,8 e 6 que corresponde às letras I,B,C,A,H e F. Como a importância do ponto 9 cresce bastante com a ponderação  $k=2$ , este é o primeiro a ser atenuado e com isso, faz a presença do ponto 10 ser de grande influência, causando o desvio acentuado em seu resíduo, conforme podemos ver pela figura 4.12. Na figura 4.13 temos os parâmetros sofrendo também mudanças consideráveis no momento em que o ponto 9 é atenuado. \*

Realizamos um estudo análogo aumentando o raio da circunferência sobre o qual os 8 pontos foram escolhidos para 0.2. Pretendemos com isso, fazer com que a influência dos pontos 9 e 10, também tomados como (0.0,0.0) e (1.0,1.0), sejam menores pois estes estarão mais próximos do centro dos dados. Assim, é de se esperar que os gráficos de diagnóstico para este caso consigam detetar os pontos 9 e 10 como conjuntamente influentes somente quando a potência de  $1-h$  for maior.

**EXEMPLO 4.3** 8 pontos escolhidos ao acaso sobre a circunferência de raio 0.2, centrado em (0.5,0.5) e os pontos 9 e 10 como sendo (0.0,0.0) e (1.0,1.0) respectivamente.

**Quadro 4.4** Dados gerados sobre a circunferência de raio 0.2 e centro (0.5,0.5).

caso	x	y
1	0.4293	0.3129
2	0.6461	0.6366
3	0.5360	0.6967
4	0.4081	0.6777
5	0.6032	0.3287
6	0.3991	0.3273
7	0.4114	0.3207
8	0.6472	0.6354
9	0.0000	0.0000
10	1.0000	1.0000

Os parâmetros estimados por mínimos quadrados são:

$$\begin{aligned} b_1 &= 0.0079 \\ b_2 &= 0.9560 \end{aligned}$$

e o coeficiente de correlação múltipla  $R^2 = 0.7302$ .

No quadro 4.5, temos as medidas de influência individualmente e podemos notar que isoladamente os pontos 9 e 10 não são influentes e ainda, que os resíduos destes pontos são bastante pequenos. Os elementos da diagonal da matriz chapéu destes dois pontos são grandes, entretanto menores que ao do exemplo anterior, como era de se esperar. Os valores de comparação são os mesmos do exemplo anterior, dados logo abaixo da tabela no quadro 4.3.

Quadro 4.5 Medidas de influência - Dados do exemplo 4.3.

caso	$\hat{y}$	e	h	$\tau$	D	DFITS
1	0.4183	-0.1054	0.1106	-0.6871	0.0314	-0.2423
2	0.6256	0.0110	0.1325	0.0704	0.0004	0.0275
3	0.5203	0.1764	0.1013	1.2185	0.0789	0.4092
4	0.3981	0.2796	0.1171	2.3824*	0.2374*	0.8674
5	0.5846	-0.2559	0.1155	-2.0459*	0.1954*	-0.7392
6	0.3895	-0.0622	0.1203	-0.3987	0.0121	-0.1474
7	0.4012	-0.0805	0.1159	-0.5191	0.0194	-0.1880
8	0.6266	0.0088	0.1331	0.0560	0.0003	0.0219
9	0.0079	-0.0079	0.5406**	-0.0695	0.0032	-0.0754
10	0.9639	0.0361	0.5132**	0.3098	0.0570	0.3180

Na figura 4.14, temos a dispersão dos pontos do quadro 4.4 e a reta ajustada por mínimos quadrados e novamente é fácil ver que os pontos 9 e 10 determinam fortemente a inclinação da reta.

As figuras 4.15 a 4.22 são os gráficos de diagnóstico com a potência de  $1-h$ ,  $k=0$  e sucessivamente, 0.5, 1 e 2. Os pontos 9 e 10 são detetados como conjuntamente influentes somente quando  $k=2$ , figuras 4.21 e 4.22. Na figura 4.21, o ponto 9, rotulado 'I' é atenuado na abscissa 0.125 causando forte desvio em si próprio e no ponto 10, rotulado 'J'. Na figura 4.22, nesta mesma abscissa, os parâmetros sofrem desvios consideráveis e isto nos leva a concluir que os pontos 9 e 10 são conjuntamente influentes.

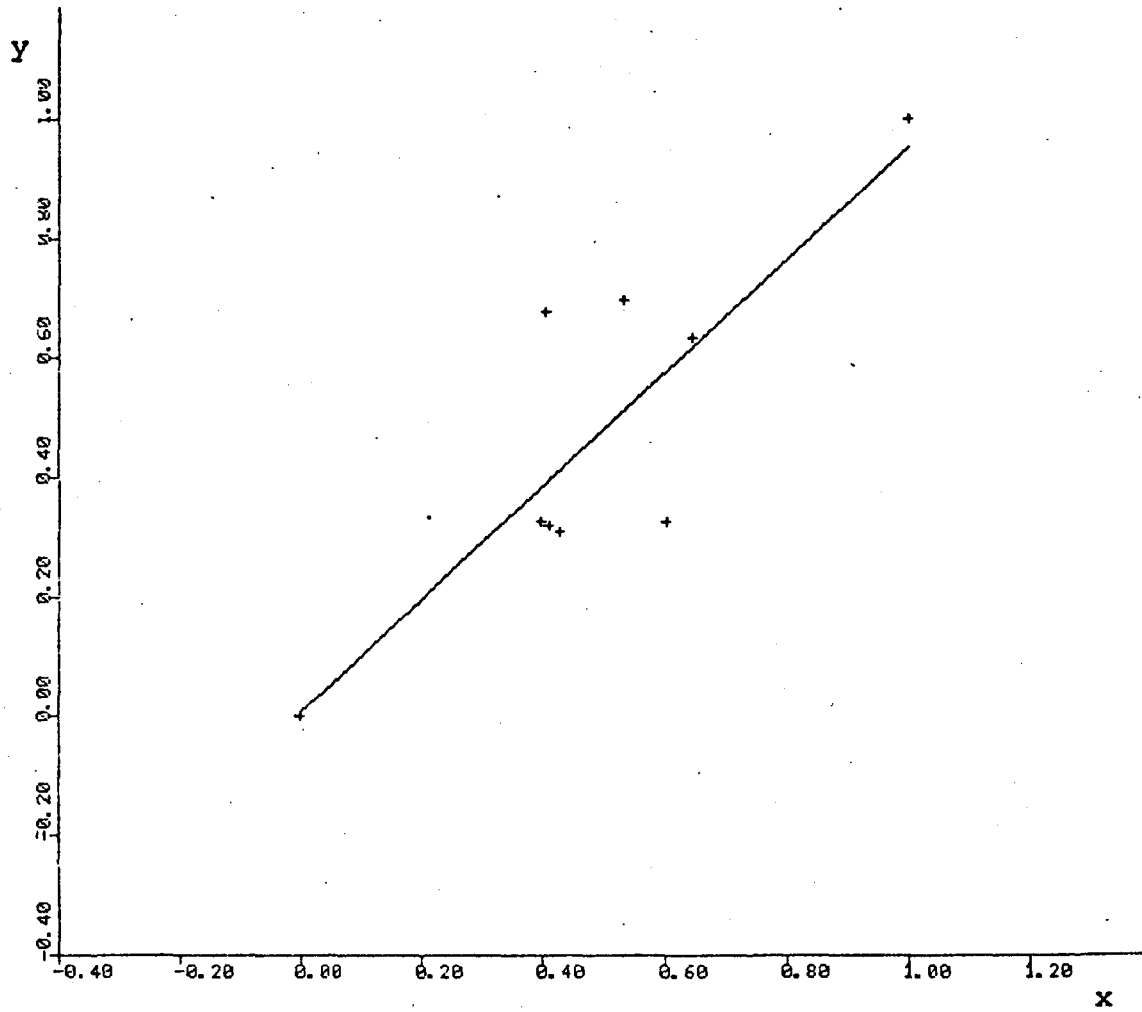


Figura 4.14 Gráfico x versus y - Exemplo 4.3

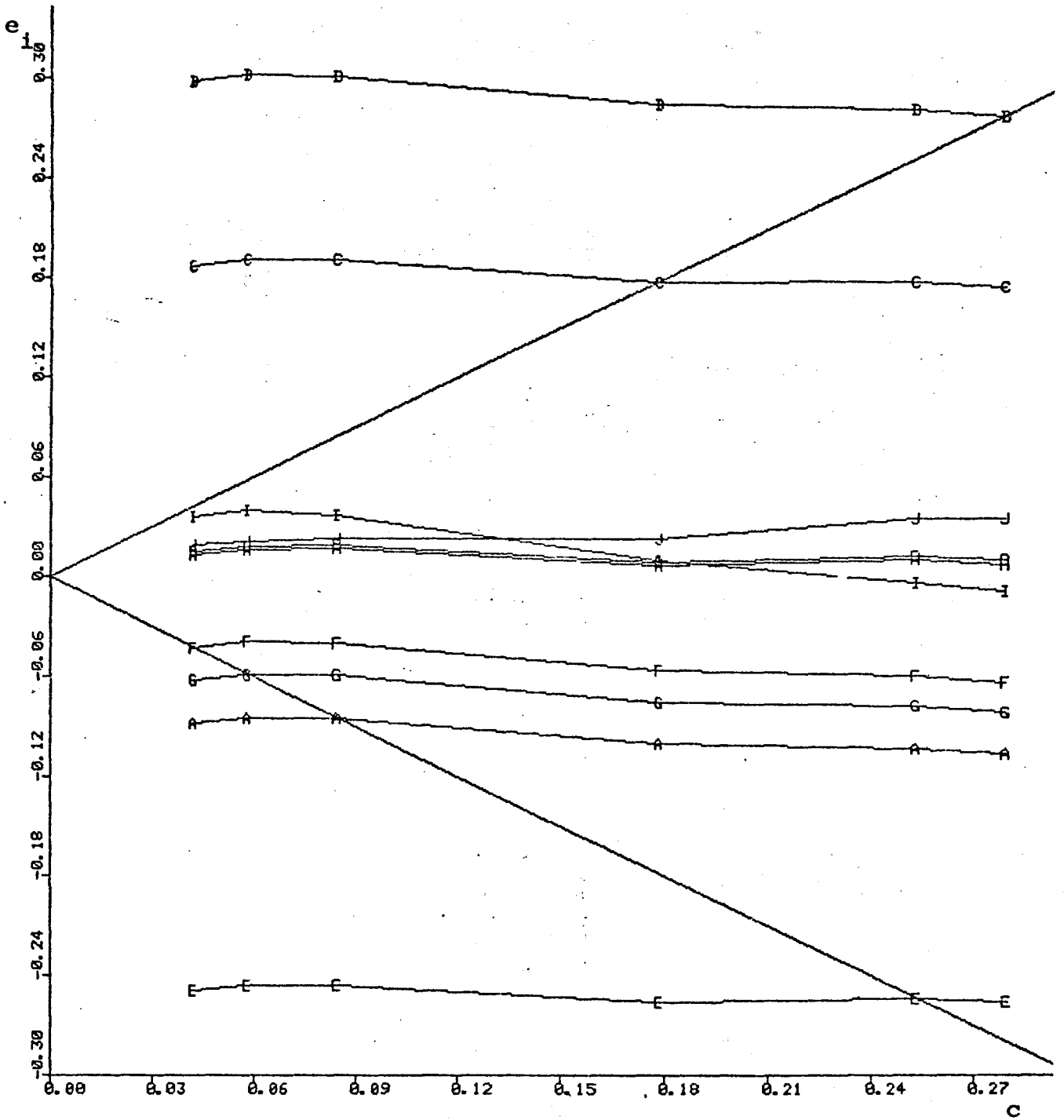


Figura 4.15 Gráfico de resíduos x c - exemplo 4.3

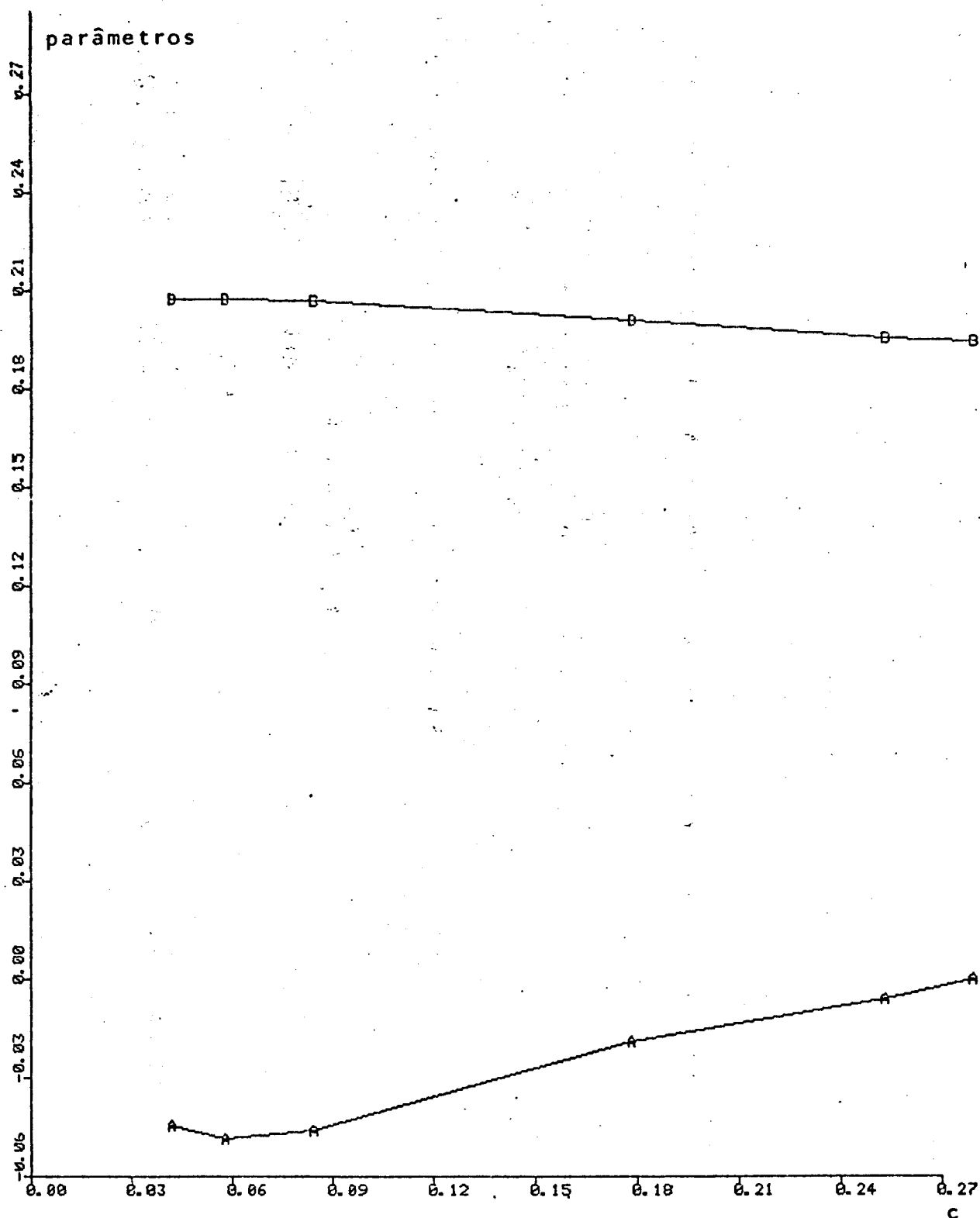


Figura 4.16 Gráfico de parâmetros x c - exemplo 4.3

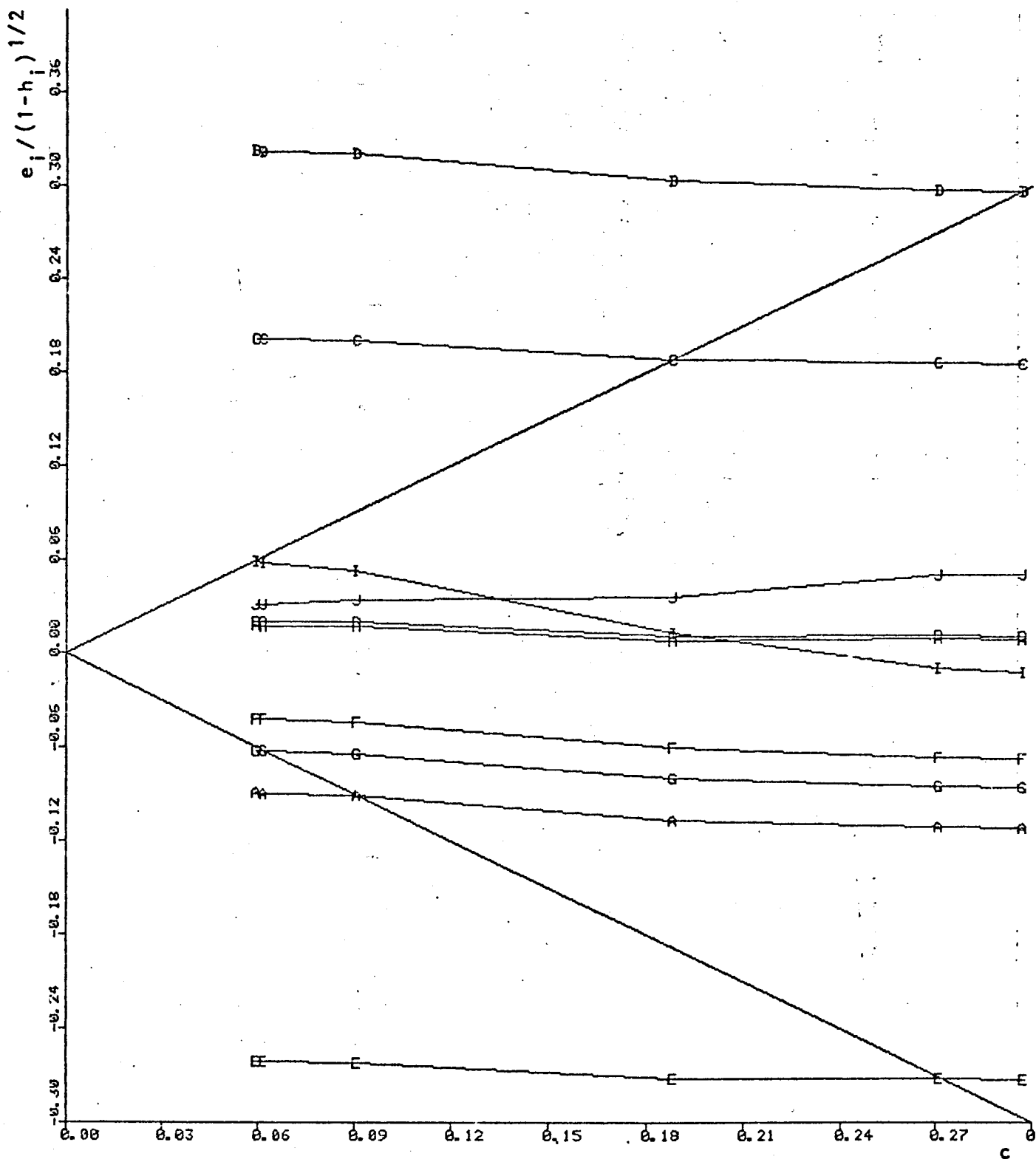


Figura 4.17 Gráfico de resíduos x c - exemplo 4.3

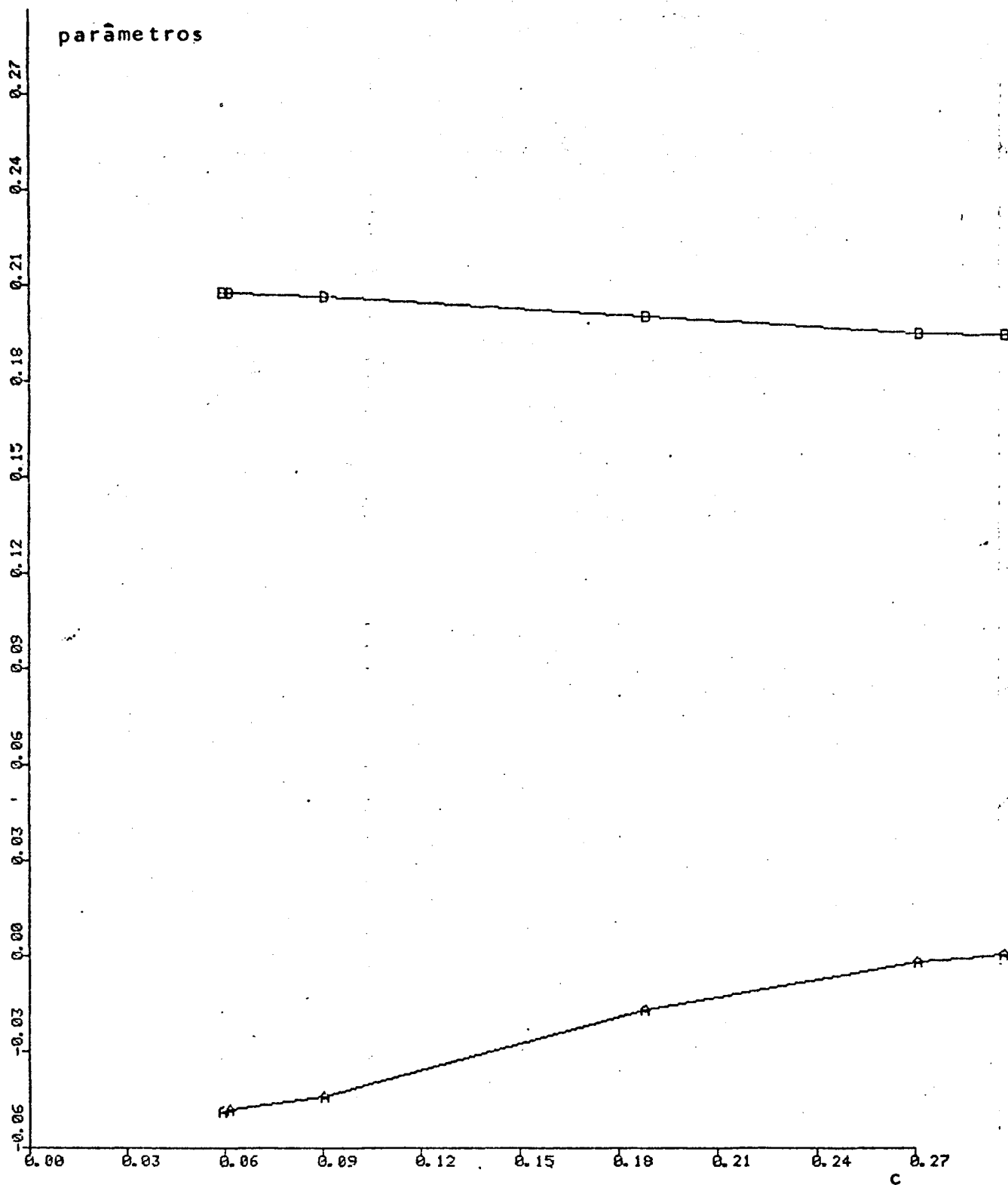


Figura 4.18 Gráfico de parâmetros x c - exemplo 4.3

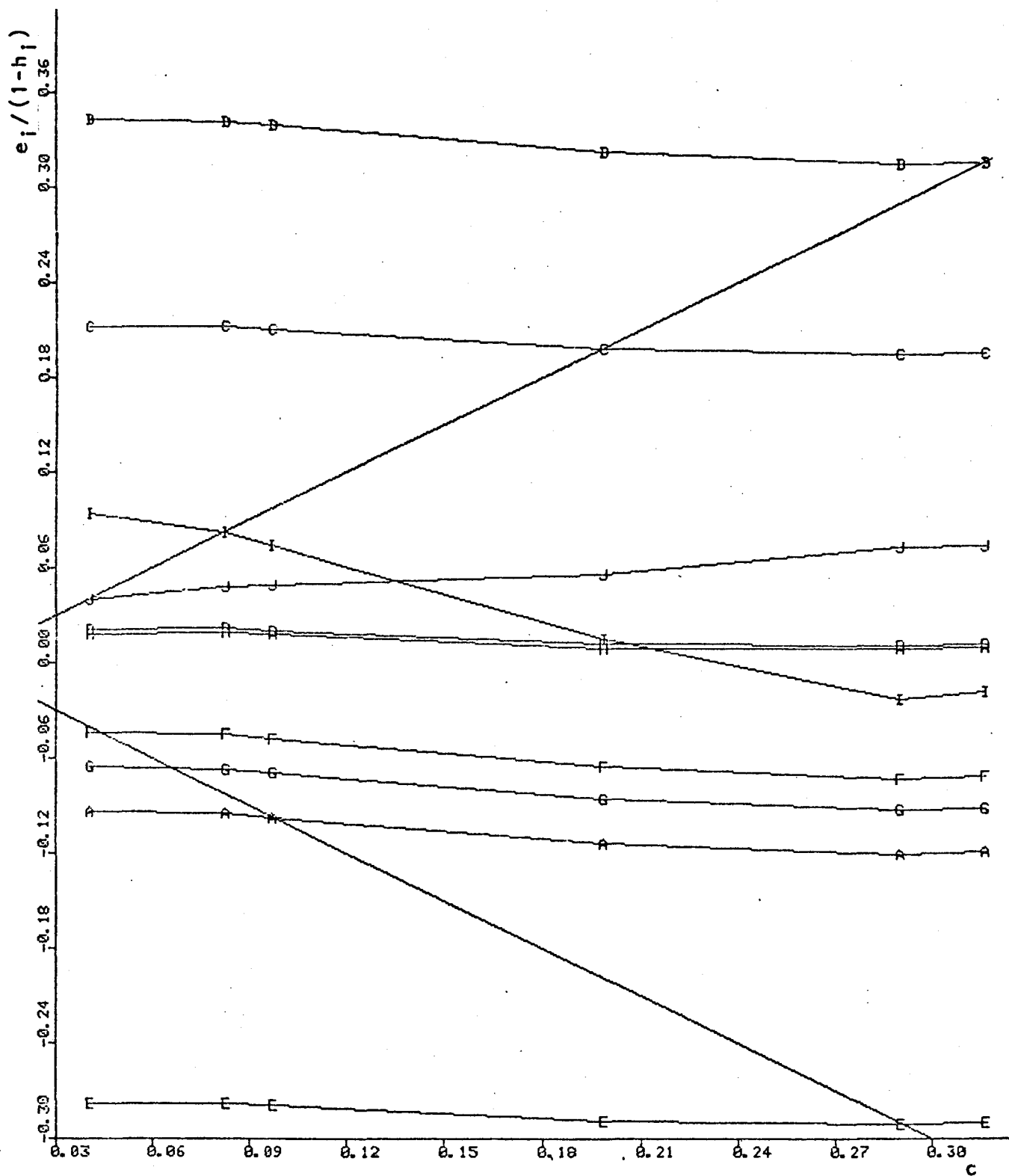


Figura 4.19 Gráfico de resíduos x c - exemplo 4.3



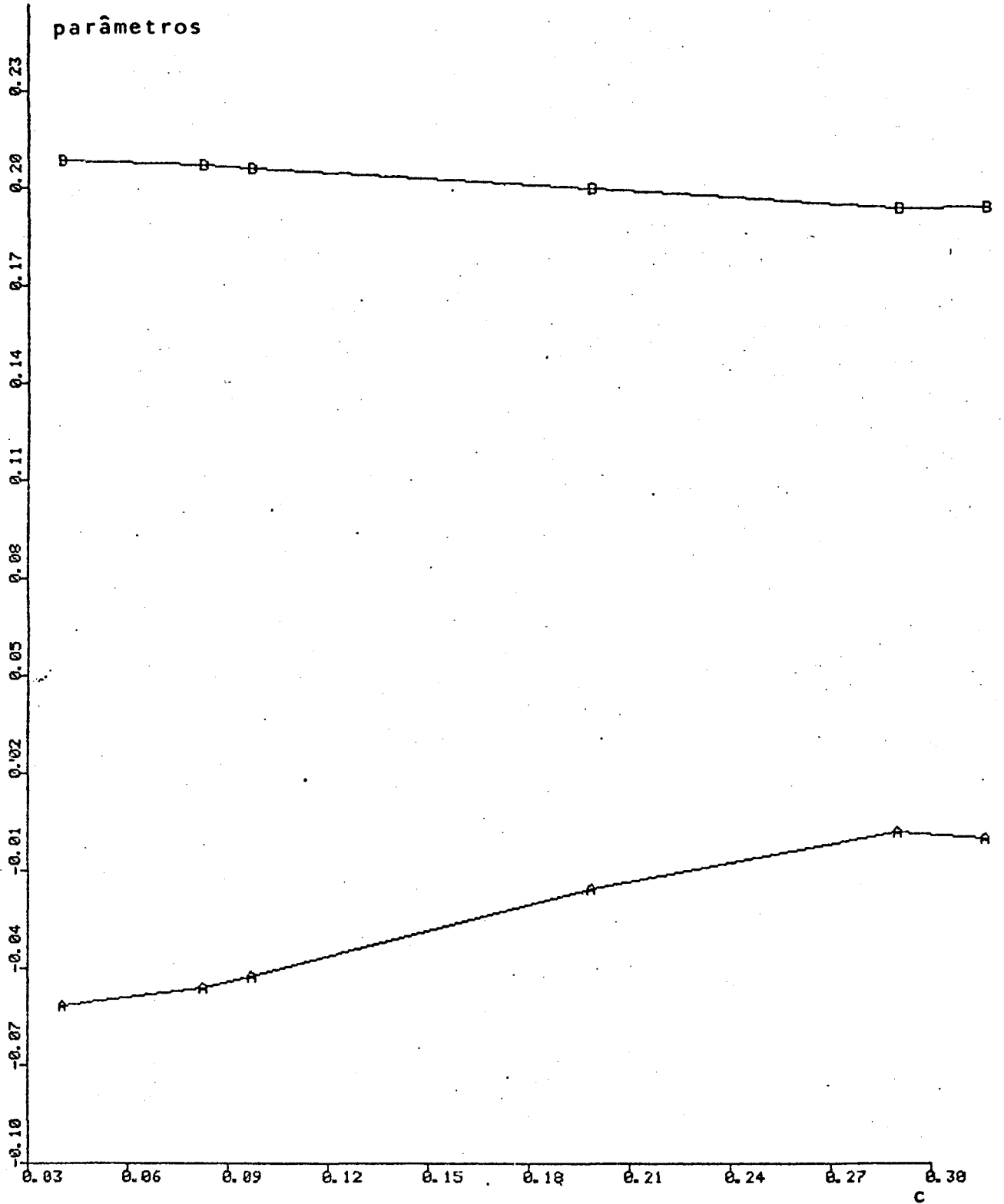


Figura 4.20 Gráfico de parâmetros x c - exemplo 4.3

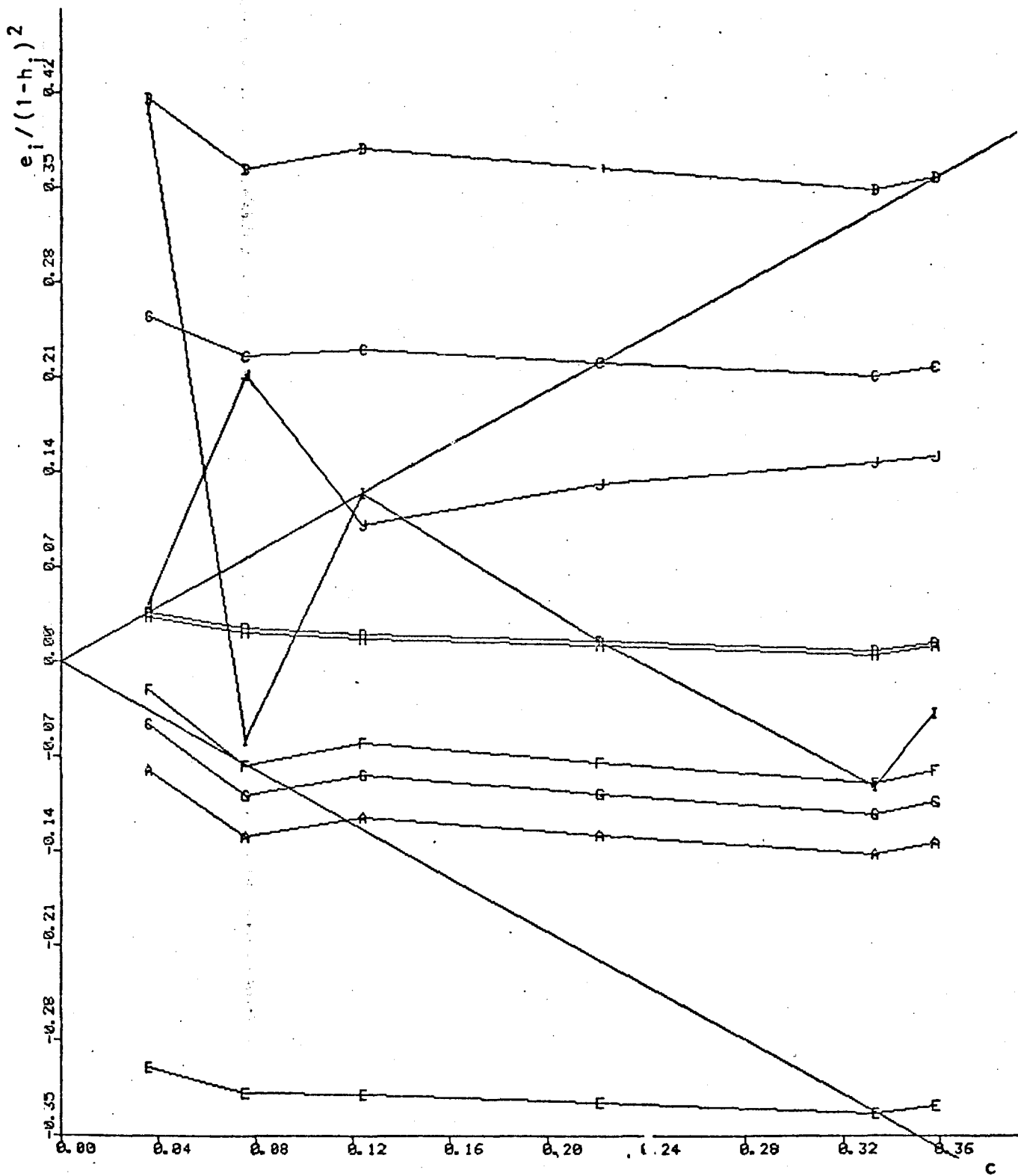


Figura 4.21 Gráfico de resíduos x c - exemplo 4.3

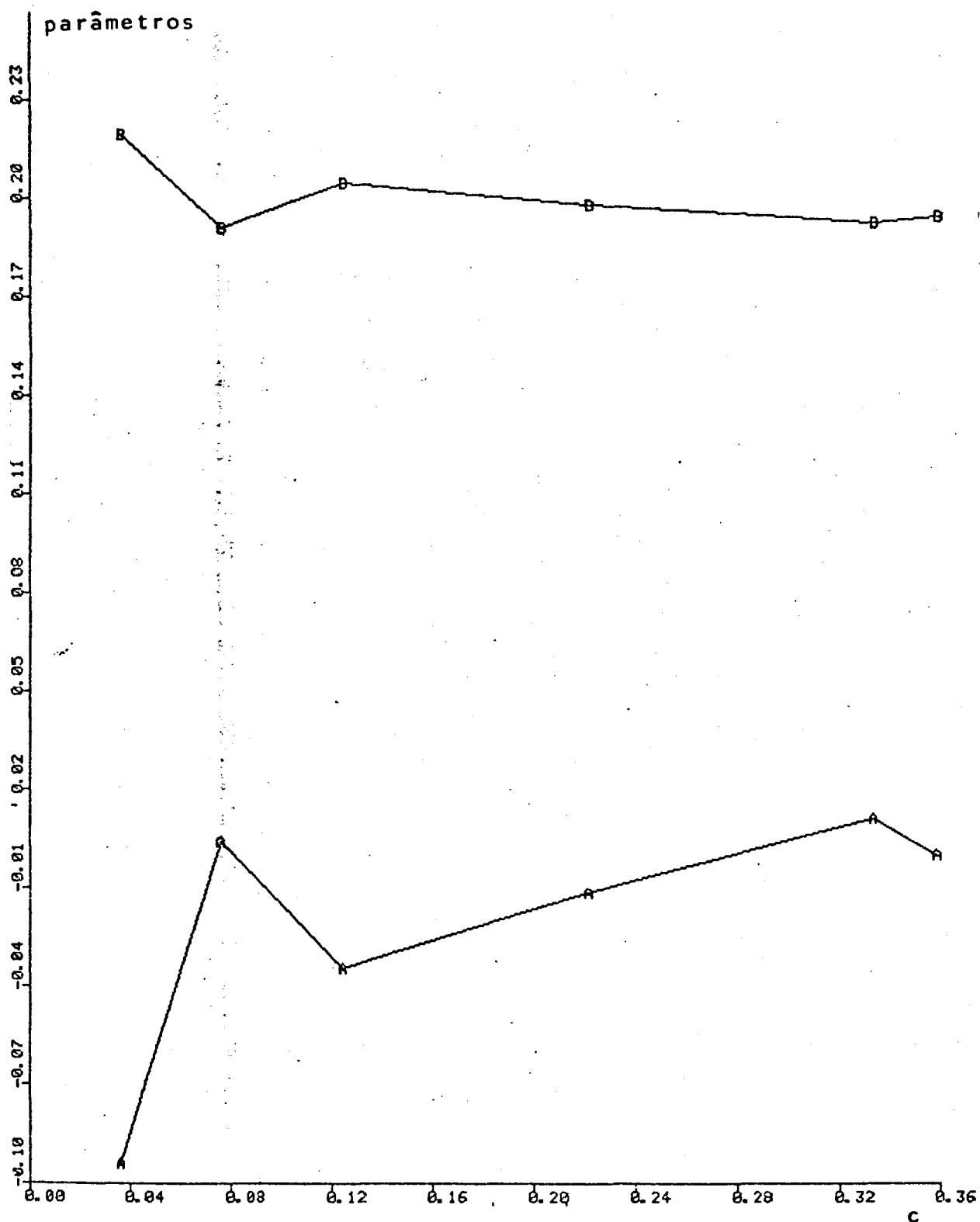


Figura 4.22 Gráfico de parâmetros x c - exemplo 4.3

No exemplo seguinte, estudaremos uma outra situação e veremos novamente que, os gráficos com os ajustes devidos, são capazes de nos fornecer informações valiosas.

EXEMPLO 4.4 Tomemos novamente 8 pontos escolhidos ao acaso sobre a circunferência de raio 0.1, agora centrado no ponto (0.1,0.1) e o ponto 9 e 10 como sendo (0.5,0.5) e (1.0,1.0), respectivamente. Os pontos encontrados estão listados no quadro 4.6 e na figura 4.14 temos a dispersão dos mesmos num gráfico de x versus y.

Quadro 4.6 Dados gerados sobre a circunferência de raio 0.1 e centro (0.1,0.1).

caso	x	y
1	0.1858	0.1514
2	0.1742	0.0330
3	0.0713	0.0042
4	0.0101	0.0562
5	0.0142	0.0487
6	0.0393	0.1795
7	0.0085	0.1403
8	0.1126	0.0008
9	0.5000	0.5000
10	1.0000	1.0000

Os parâmetros estimados por mínimos quadrados são:

$$b_1 = 0.0094$$

$$b_2 = 0.9545$$

e o coeficiente de correlação múltipla  $R^2 = 0.9134$ .

Na figura 4.23, temos a dispersão dos dados e a reta ajustada por mínimos quadrados e por este gráfico podemos ver que o ponto 10 parece exercer forte influência no ajuste e que os pontos 9 e 10 formariam, provavelmente, um subconjunto muito influente.

No quadro 4.7, temos as medidas de influência onde são destacados o ponto 10 segundo os  $h$ 's, os pontos 2,6,7 e 10 segundo  $D$  e o ponto 10 segundo  $DFIT$ S.

Como nos casos anteriores, os gráficos de diagnóstico com potência 0 e 0.5 em  $(1-h)$  não conseguem detetar os pontos 9 e 10 como conjuntamente influentes. Ao usarmos a potência 1, é possível identificar o conjunto formado por estes dois pontos como um provável subconjunto influente através das figuras 4.24 e 4.25. Deixaremos de apresentar os gráficos referentes às outras ponderações por considerá-las redundantes.

Quadro 4.7 Medidas de Influência - Dados do exemplo 4.4

caso	$\hat{y}$	e	h	$\tau$	D	DFFITs
1	0.1868	-0.0354	0.1008	-0.3599	0.0081	-0.1205
2	0.1757	-0.1427	0.1016	-1.7149	0.1338*	-0.5766
3	0.0775	-0.0733	0.1222	-0.7795	0.0445	-0.2908
4	0.0191	0.0371	0.1458	0.3879	0.0144	0.1603
5	0.0230	0.0257	0.1439	0.2669	0.0068	0.1094
6	0.0470	0.1325	0.1335	1.5870	0.1630*	0.6228
7	0.0176	0.1227	0.1465	1.4475	0.1582*	0.5997
8	0.1169	-0.1161	0.1111	-1.3144	0.0989	-0.4646
9	0.4867	0.0133	0.1938	0.1420	0.0028	0.0696
10	0.9639	0.0361	0.8009**	0.8082	1.3734**	1.6211*

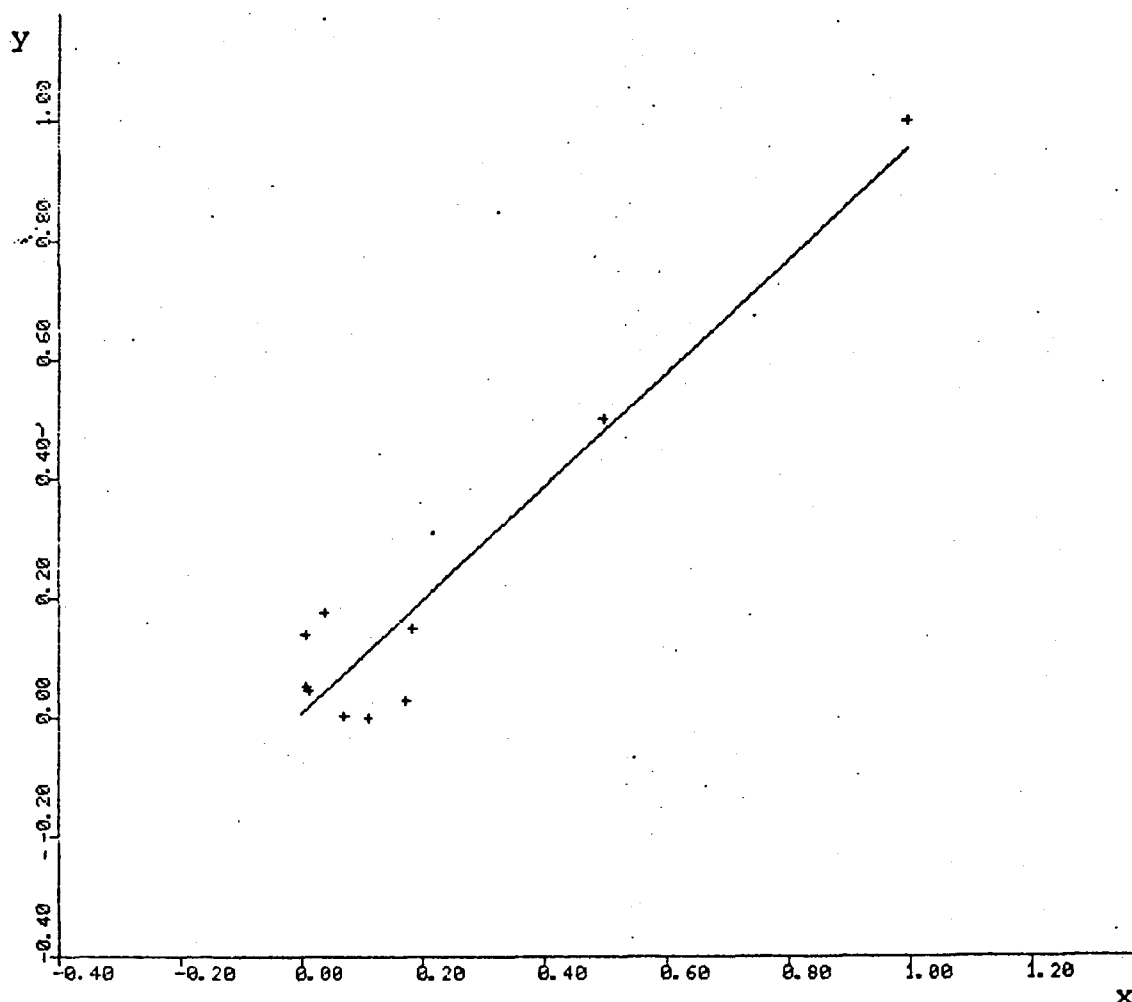


Figura 4.23 Gráfico x versus y - exemplo 4.4

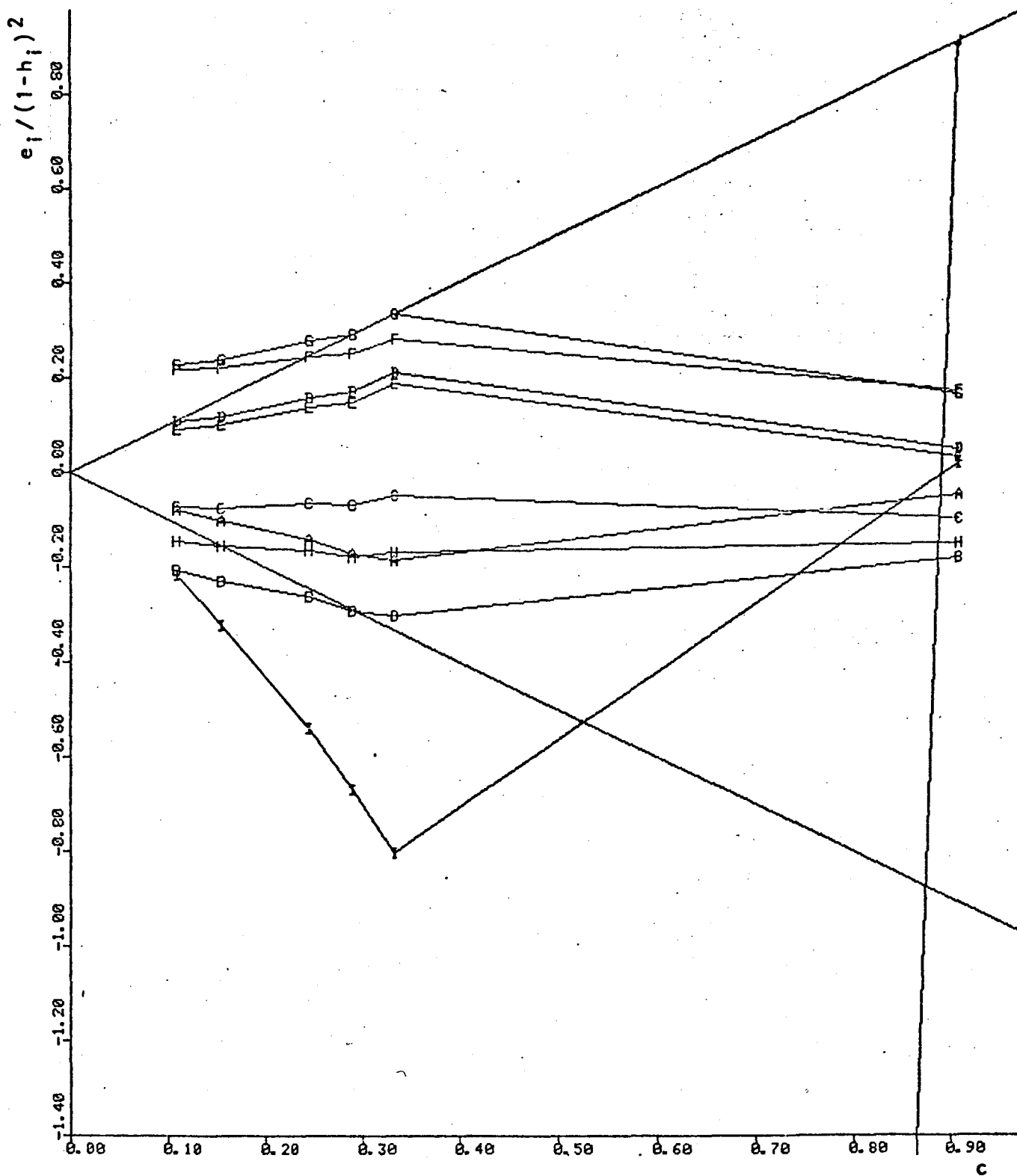


Figura 4.24 Gráfico de resíduos x c - exemplo 4.4

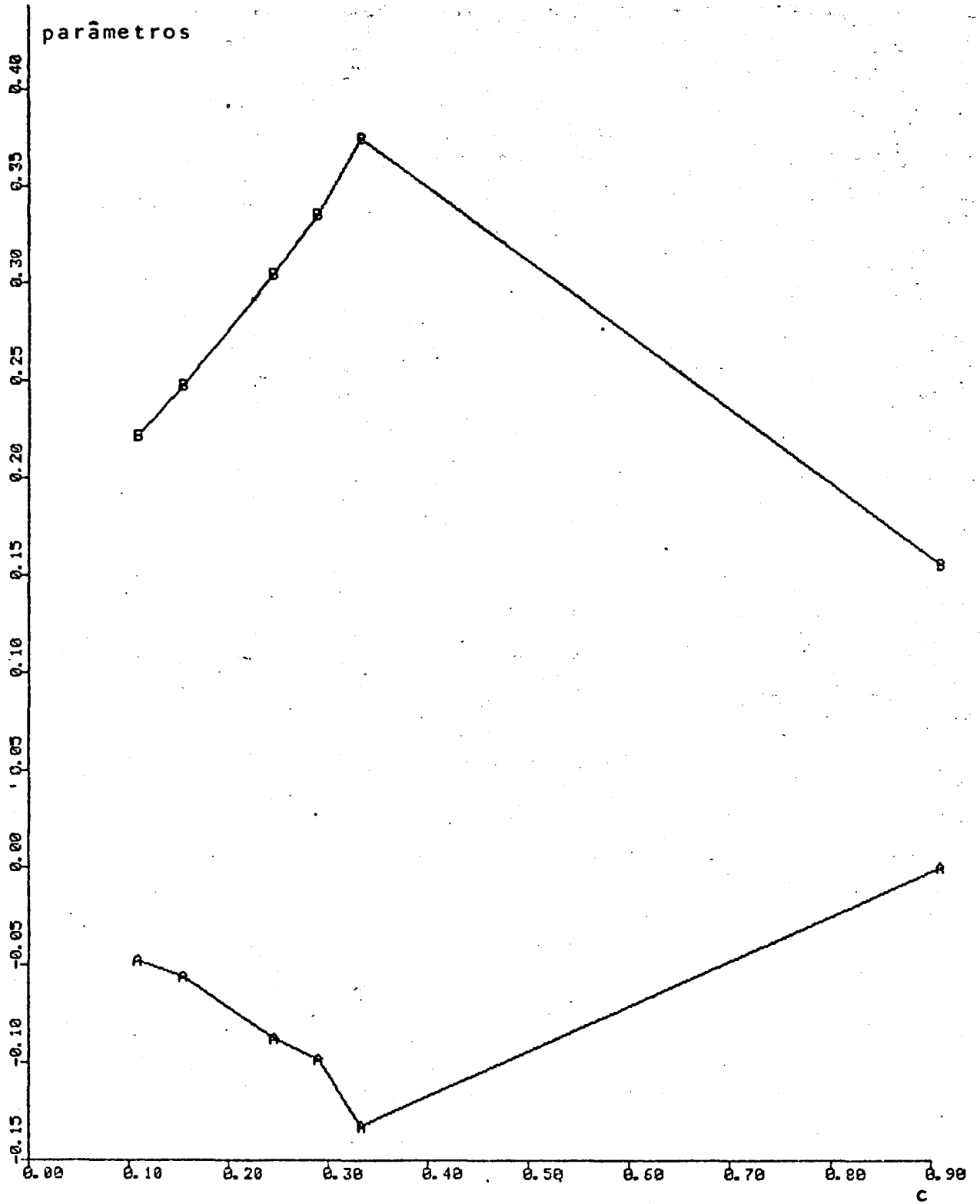


Figura 4.25 Gráfico de parâmetros x c - exemplo 4.4

Repetimos novamente a mesma situação, aumentando o raio da circunferência para 0.2.

EXEMPLO 4.5 8 pontos escolhidos ao acaso sobre a circunferência de raio 0.2 e centro (0.2,0.2). Os pontos 9 e 10 são, respectivamente, (0.5,0.5) e (1.0,1.0).

Quadro 4.8 Dados gerados sobre a circunferência de raio 0.2 e centro (0.2,0.2)

caso	x	y
1	0.1252	0.3855
2	0.3666	0.0894
3	0.2100	0.0003
4	0.0182	0.1166
5	0.3496	0.3327
6	0.1239	0.0151
7	0.0021	0.1713
8	0.2072	0.3999
9	0.5000	0.5000
10	1.0000	1.0000

Os parâmetros estimados por mínimos quadrados são:

$$b_1 = 0.0560$$

$$b_2 = 0.8444$$

e o coeficiente de correlação múltipla  $R^2 = 0.6854$ .

Na figura 4.26, temos a dispersão destes dados e a reta ajustada por mínimos quadrados. Note que o ponto 9 não está tão distante do restante dos pontos e que o ponto 10 determina fortemente a inclinação da reta. No quadro 4.9, temos ressaltado este fato pela coluna dos h's. Neste mesmo quadro, a medida D destaca os pontos 1,2,3 e 10 como influentes e o DFFITS, o ponto 10.



Quadro 4.9 Medidas de influência - Dados do exemplo 4.5.

caso	$\hat{y}$	$e$	$h$	$\tau$	D	DFITS
1	0.1617	0.2238	0.1348	1.4302	0.1410*	0.5646
2	0.3655	-0.2761	0.1074	-1.8721	0.1607*	-0.6496
3	0.2333	-0.2330	0.1082	-1.4777	0.1154*	-0.5148
4	0.0713	0.0453	0.1946	0.2651	0.0096	0.1303
5	0.3512	-0.0185	0.1045	-0.1021	0.0007	-0.0349
6	0.1606	-0.1455	0.1354	-0.8603	0.0599	-0.3404
7	0.0577	0.1136	0.2061	0.6886	0.0659	0.3509
8	0.2309	0.1690	0.1088	1.0006	0.0611	0.3497
9	0.4782	0.0218	0.1562	0.1244	0.0016	0.0535
10	0.9004	0.0996	0.7438**	1.1172	1.7573**	1.9036*

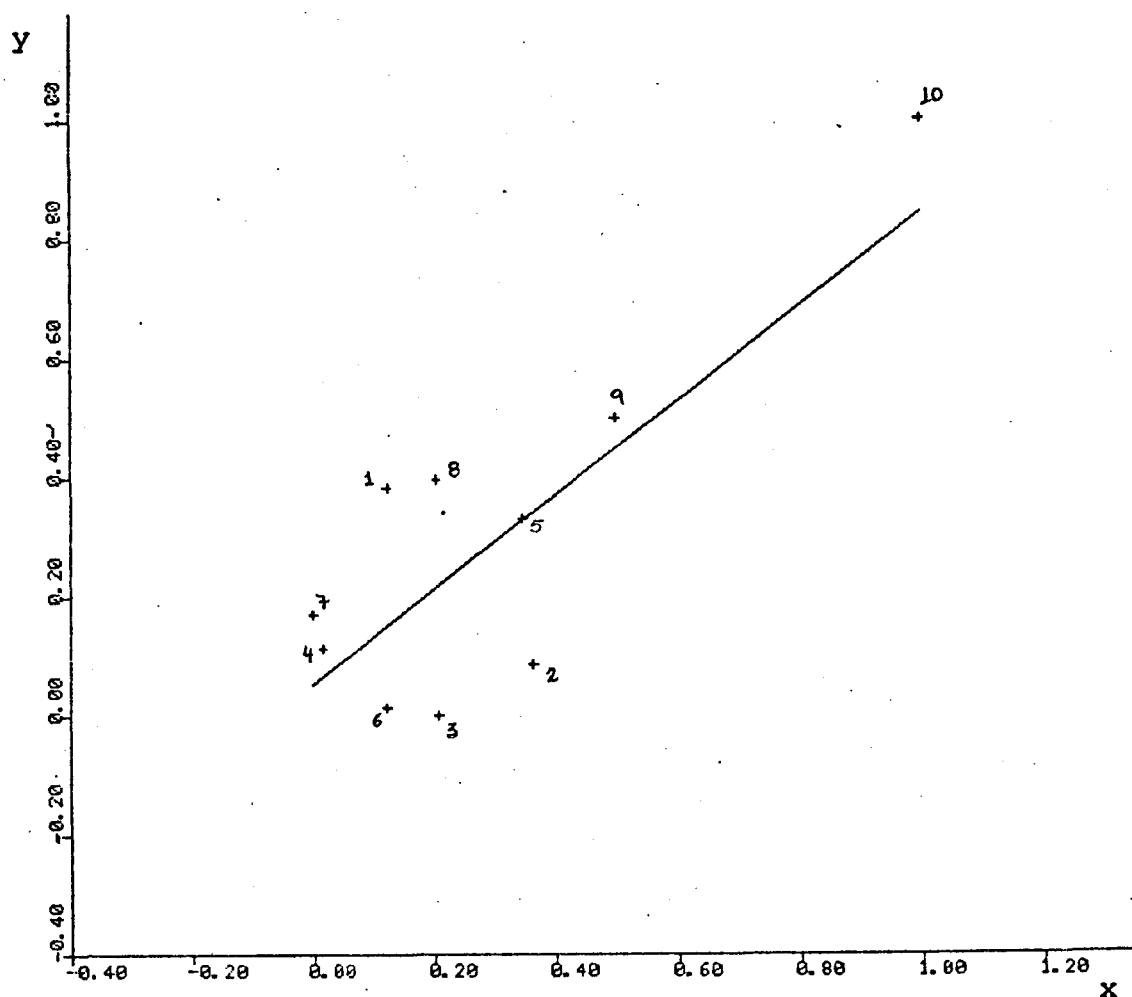


Figura 4.26 Gráfico x versus y - exemplo 4.5

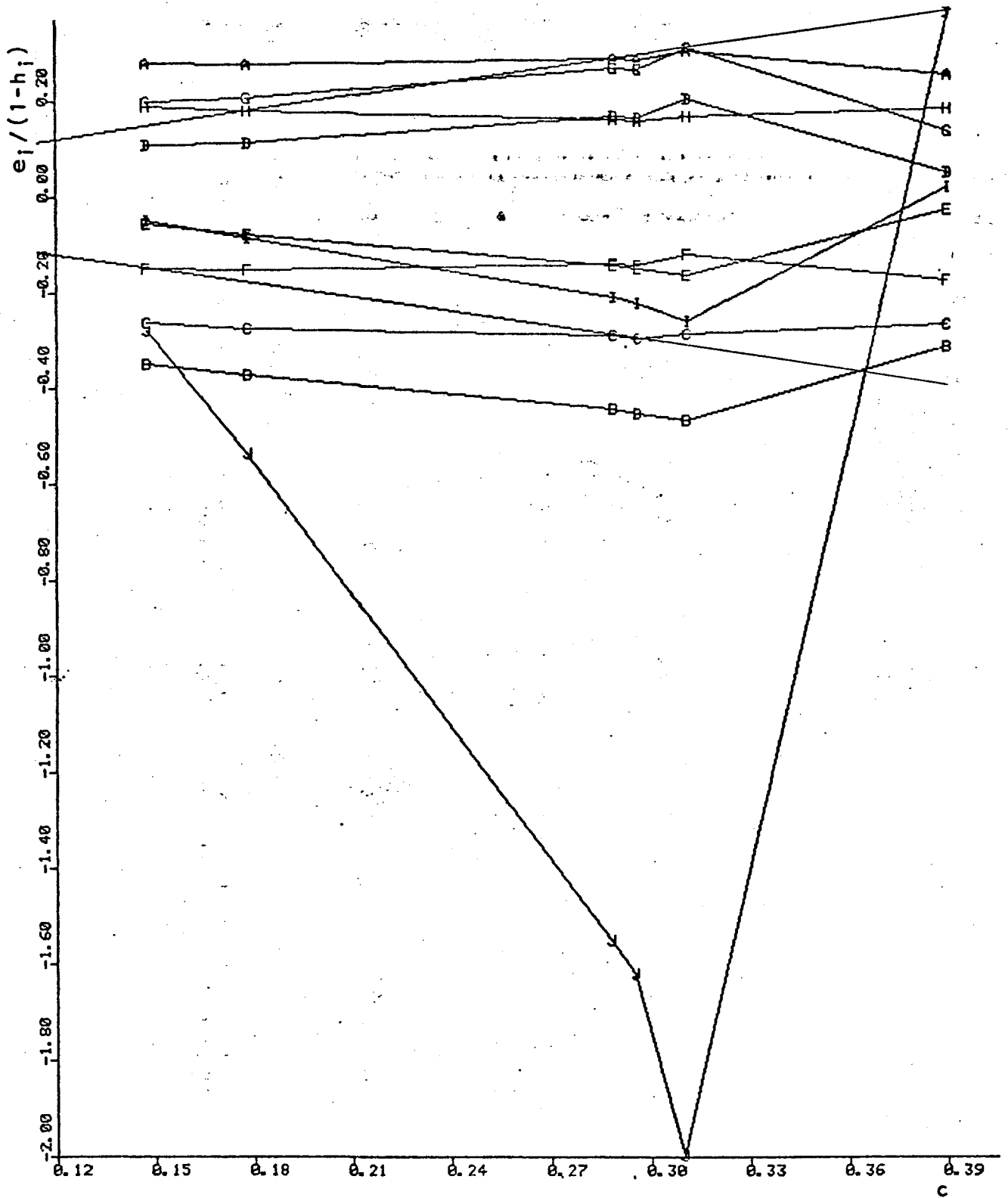


Figura 4.27 Gráfico de resíduos x c - exemplo 4.5

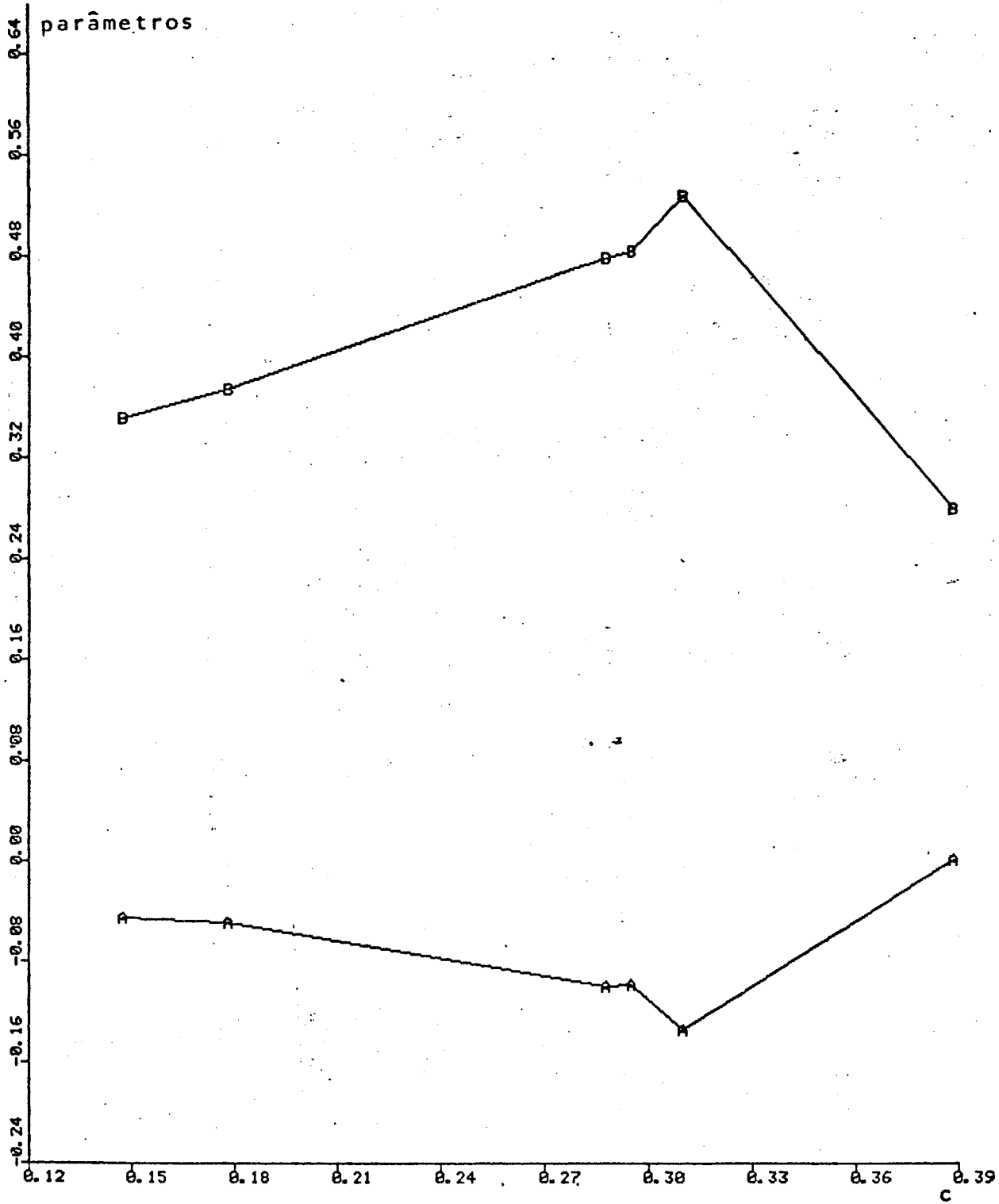


Figura 4.28 Gráfico de parâmetros x c - exemplo 4.5

Nas figuras 4.27 e 4.28, temos os gráficos de diagnóstico com a potência de  $(1-h)$  igual a 1. A figura 4.27 mostra um aspecto interessante. Neste caso, devido a particularidade destes dados, é possível distinguirmos dois subconjuntos que possivelmente são influentes. Note que o ponto 10, rotulado 'J', ao ser atenuado, faz com que o ponto 9 rotulado 'I' e numa escala menor, o ponto 5 rotulado 'E' tenham a mesma tendência, indicando que estes três pontos podem formar um subconjunto. Da mesma maneira, temos os pontos 4,6 e 7, rotulados 'D','F' e 'G' com um ligeiro acréscimo em seu resíduo, indicando também que podem formar outro conjunto. Estes seis pontos poderiam formar um único conjunto se tivéssemos um número maior de observações. Como em nosso caso temos apenas 10 pontos, não teria sentido considerar um subconjunto com a maioria dos pontos. #

Para confirmarmos os subconjuntos encontrados nos exemplos 4.2 a 4.5 como influentes, aconselhamos calcular a medida D para influência múltipla. Abaixo, exibimos as medidas calculadas para cada um dos subconjuntos detetados pelos métodos gráficos como influentes.

Quadro 4.10 Medidas de influência múltipla - Exemplos 4.2 a 4.5.

Exemplo	Subconjunto	D
4.2	(9,10)	81.8891
4.3	(9,10)	0.7637
4.4	(9,10)	62.2867
4.5	(9,10)	7.2482
	(5,9,10)	14.1134
	(4,6,7)	0.0284

Como podemos ver, todos os subconjuntos encontrados pelo método gráfico podem ser considerados altamente influentes, a menos do subconjunto (4,6,7) do exemplo 4.5 o qual tomamos apenas para averiguação. Os valores de comparação são os mesmos utilizados para a medida D do caso unidimensional e que consta no quadro 4.3. Para termos uma melhor idéia de quanto um dado subconjunto é influente, podemos interpretar os números do quadro acima da seguinte maneira: calculando a probabilidade de que uma variável com distribuição F com 2 e 8 graus de liberdade seja maior que a quantidade D encontrada e esta probabilidade nos fornecerá a idéia de quanto a estimativa  $b(.)$  se afasta de  $b$ , ou seja, para um contorno de quanto em relação a  $b$ . No exemplo 4.2, a retirada dos pontos 9 e 10 desloca  $b(9,10)$  para um contorno de praticamente 100% em relação a  $b$ . No exemplo 4.3, para um contorno de 50% e no exemplo 4.4, para um contorno de 100%. No exemplo 4.5, o subconjunto (9,10) para um contorno de 98% e o subconjunto (5,9,10) para 99,8%.

Tomaremos mais um exemplo para ilustrar a eficiência destes gráficos.

**EXEMPLO 4.6** Os dados do quadro abaixo são referentes a um estudo realizado na Universidade de Califórnia em Los Angeles sobre cianose. A variável X é a idade, em meses, em que a criança pronunciou a primeira palavra e Y é a pontuação de adaptação Gesell para cada uma das 21 crianças. Os dados foram obtidos por Mickey, Dunn e Clark[18] e são também estudados por Cook e Weisberg. Esta é a razão pela qual tomaremos como exemplo pois teremos assim, como checar nossos métodos.

Quadro 4.11 Dados de pontuação de adaptação.

caso	x	y	caso	x	y
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	94	20	11	86
			21	10	100

Na figura 4.29 temos a dispersão destes pontos e o ajuste linear simples por mínimos quadrados traçado. Podemos evidenciar o ponto 19 como possível ponto aberrante e os pontos 18 e 2 parecem exercer forte influência no ajuste.

Os parâmetros estimados por mínimos quadrados são:

$$b_1 = 109.8738$$

$$b_2 = -1.1270$$

No quadro 4.12, temos as medidas de diagnóstico onde na coluna h, a observação 18 se destaca bastante. Os resíduos studentizados ressaltam o ponto 19 e conforme o teste respectivo, é considerado ponto aberrante. As medidas de influência D e DFFITS destacam os pontos 18 e 19 como influentes, individualmente, com maior evidência para o primeiro.

Nas figuras 4.30 e 4.31 temos gráficos de diagnóstico como foram propostos por Denby e Mallows. No primeiro, resíduos versus o parâmetro escala c, podemos ressaltar somente o ponto 19, rotulado 'S', como ponto aberrante e todos os resíduos permanecem quase constantes com c decrescendo. A ordem de entrada dos pontos no conjunto C é 19,3,14,11,20,5,17,4,2,10,15,7,12,9,8,1 e 16. No segundo gráfico, parâmetros versus c, a linha 'A' correspondente ao parâmetro b1 sofre ligeiros desvios que não estão associados à atenuação de nenhum ponto, conforme o gráfico de resíduos e o parâmetro b2, representado pela linha 'B' permanece constante com c decrescendo.

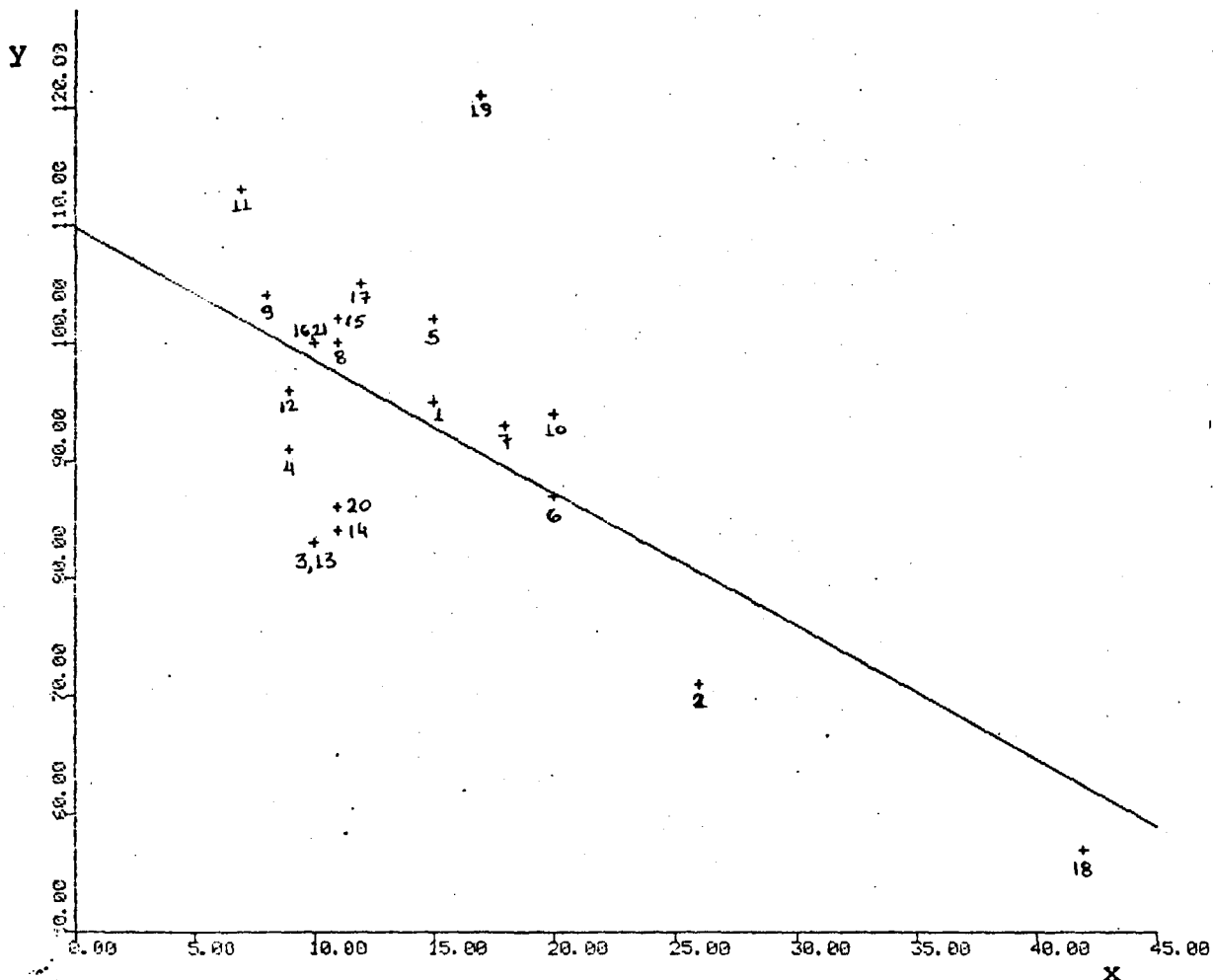


Figura 4.29 Gráfico x versus y - Dados de pontuação de adaptação

Quadro 4.12 Medidas de Influência - Dados do exemplo 4.6

caso	$\hat{y}$	e	h	$\tau$	D	DFFITs
1	92.9690	2.0310	0.0479	0.1840	0.0009	0.0413
2	80.5721	-9.5721	0.1545	-0.9416	0.0815	-0.4025
3	98.6040	-15.6040	0.0628	-1.5108	0.0717	-0.3911
4	99.7309	-8.7309	0.0705	-0.8143	0.0256	-0.2243
5	92.9690	9.0310	0.0479	0.8329	0.0177	0.1869
6	87.3341	-0.3341	0.0726	-0.0306	0.0000	-0.0086
7	89.5880	3.4120	0.0580	0.3112	0.0031	0.0772
8	97.4770	2.5230	0.0567	0.2297	0.0017	0.0563
9	100.8579	3.1421	0.0799	0.2899	0.0038	0.0854
10	87.3341	6.6659	0.0726	0.6177	0.0154	0.1728
11	101.9849	11.0151	0.0908	1.0508	0.0548	0.3320
12	99.7309	-3.7309	0.0705	-0.3428	0.0047	-0.0944
13	98.6040	-15.6040	0.0628	-1.5108	0.0717	-0.3911
14	97.4770	-13.4770	0.0567	-1.2798	0.0476	-0.3137
15	97.4770	4.5230	0.0567	0.4132	0.0054	0.1013
16	98.6040	1.3960	0.0628	0.1274	0.0006	0.0330
17	96.3500	8.6500	0.0521	0.7983	0.0179	0.1872
18	62.5403	-5.5403	0.6516**	-0.8451	0.6781**	-1.1558*
19	90.7150	30.2850	0.0531	3.6070**	0.2233*	0.8537*
20	97.4770	-11.4770	0.0567	-1.0765	0.0345	-0.2638
21	98.6040	1.3960	0.0628	0.1274	0.0006	0.0330

$t(18, 0.05) = 1.734$   
 $t(18, 0.01) = 2.552$

$F(2, 19, 0.10) = 0.1059$   
 $F(2, 19, 0.25) = 0.2915$

$|DFFITs| > 0.6489$

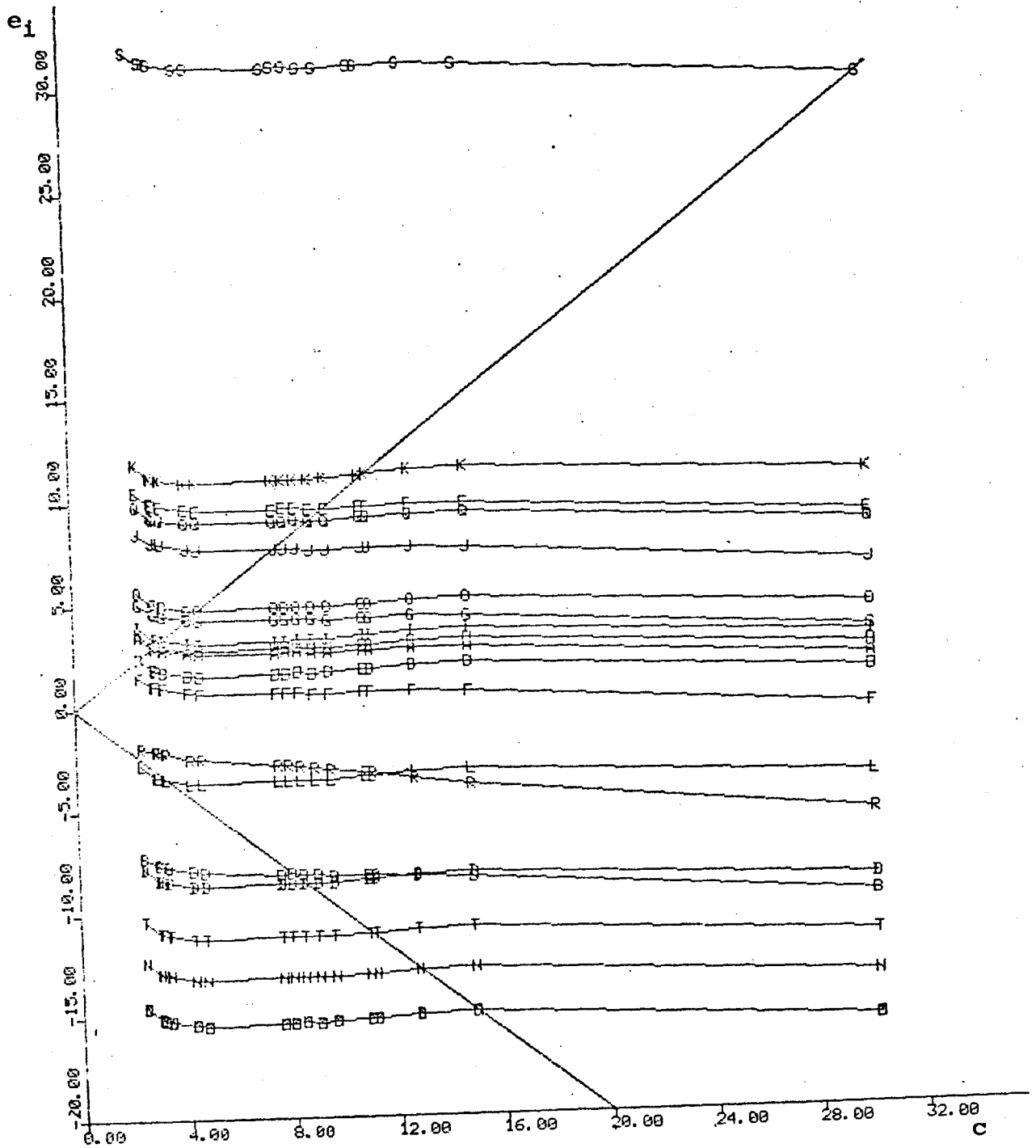


Figura 4.30 Gráfico de resíduos x  $\xi$  - Dados de pontuação de adaptação

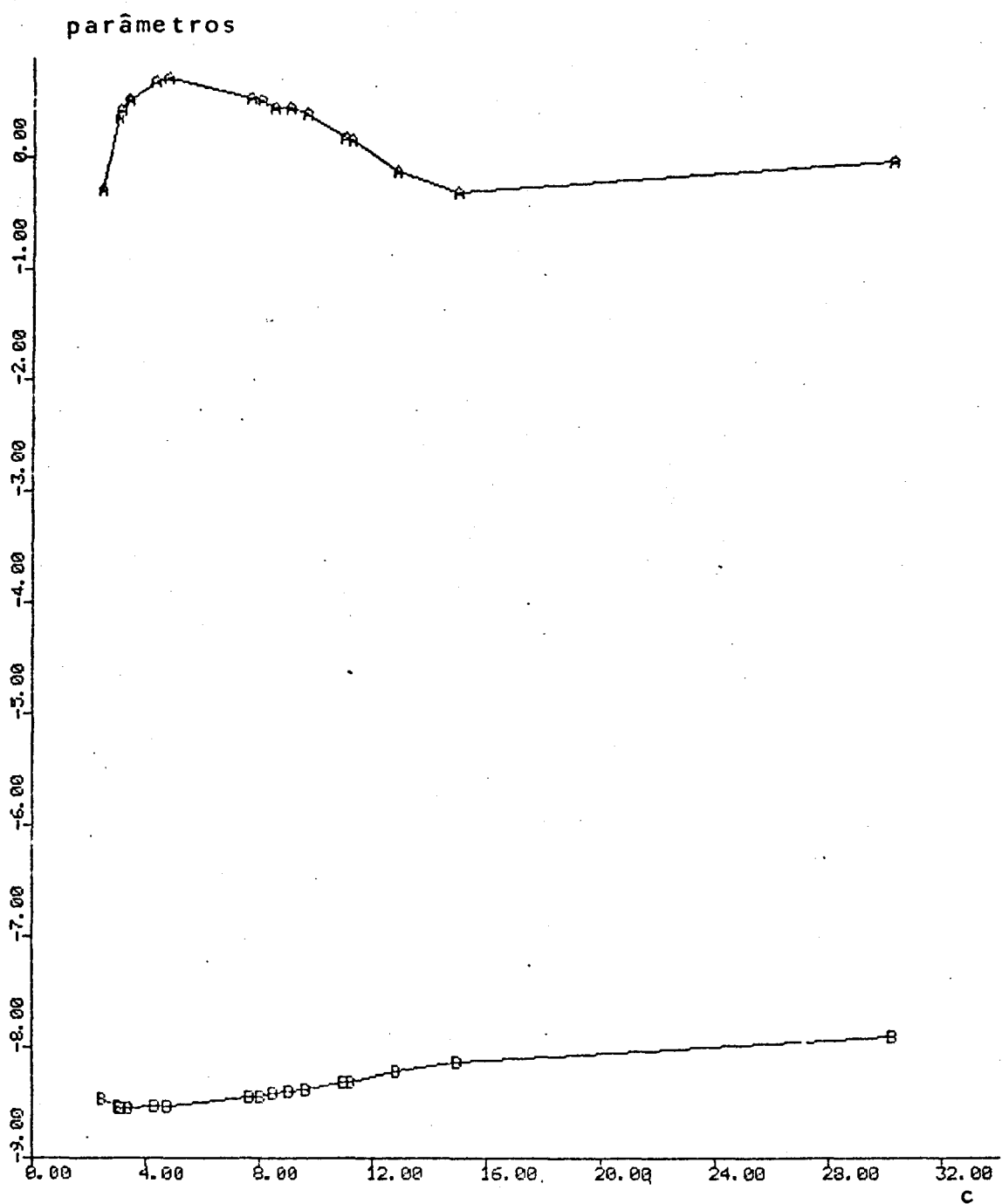


Figura 4.31 Gráfico de parâmetros x c - Dados de pontuação de adaptação



Para detetarmos subconjunto de pontos influente, foram construídos gráficos de diagnóstico com  $(1-h)$  elevado às potências 0.5, 1 e 2. Apresentaremos aqui, somente os gráficos com a potência 2 as quais permitiram visualizar e identificar de maneira mais clara os possíveis subconjuntos influentes.

Nas figuras 4.32 e 4.33, temos os gráficos de resíduos versus  $c$  e parâmetros versus  $c$  respectivamente. Vemos na figura 4.32 que o primeiro ponto atenuado foi o 18, representado pela letra 'R', destacado como influente individualmente no quadro 4.12. Este ponto, ao receber ponderação menor, tem seu resíduo bastante aumentado e a mesma tendência é notada nos pontos 2, 6, 7 e 10, destacadamente o ponto 2. No gráfico são representados respectivamente pelas linhas 'B', 'F', 'G' e 'J'. Na figura 4.33, vemos que os parâmetros sofrem grande variação no momento em que o ponto 18 é atenuado.

Com isso, podemos sugerir o subconjunto formado pelos pontos 2, 6, 7, 10 e 18 como possivelmente influente. No quadro que segue, calculamos a medida de influência múltipla  $D$  para o subconjunto e também para todos os menores constituídos por elementos deste subconjunto.

Quadro 4.13 Medidas de influência múltipla - Dados de pontuação de adaptação.

subconjunto	D	subconjunto	D
(2, 6, 7, 10, 18)	30.7035	(6, 7, 18)	0.8591
(2, 6, 7, 10)	0.0209	(6, 10, 18)	0.5963
(2, 6, 7, 18)	15.3387	(7, 10, 18)	0.3231
(2, 6, 10, 18)	17.1768	(2, 6)	0.1000
(2, 7, 10, 18)	9.4900	(2, 7)	0.0599
(6, 7, 10, 18)	0.4818	(2, 10)	0.0300
(2, 6, 7)	0.0742	(2, 18)	6.3688
(2, 6, 10)	0.0375	(6, 7)	0.0029
(2, 6, 18)	11.3879	(6, 10)	0.0164
(2, 7, 10)	0.0173	(6, 18)	0.9304
(2, 7, 18)	7.4568	(7, 10)	0.0369
(2, 10, 18)	7.4049	(7, 18)	0.6063
(6, 7, 10)	0.0409	(10, 18)	0.4137

Os valores de comparação são as mesmas dado no quadro 4.12. Confirmamos, assim, que o subconjunto formado pelos pontos 2, 6, 7, 10 e 18 é muito influente. Dos subconjuntos com 4 pontos, o formado pelos pontos (2, 6, 10, 18) é o mais influente. Dos formados com 3, o subconjunto (2, 6, 18) e dos formados com 2, o par 2 e 18 são mais influentes. Note que todos estes subconjuntos incluem os pontos 2 e 18. Entretanto, começando com o par (2, 18), a inclusão do ponto 6 causa um aumento considerável no valor de  $D$ . Da mesma forma, ao incluir o ponto 10 e finalmente o ponto 7.

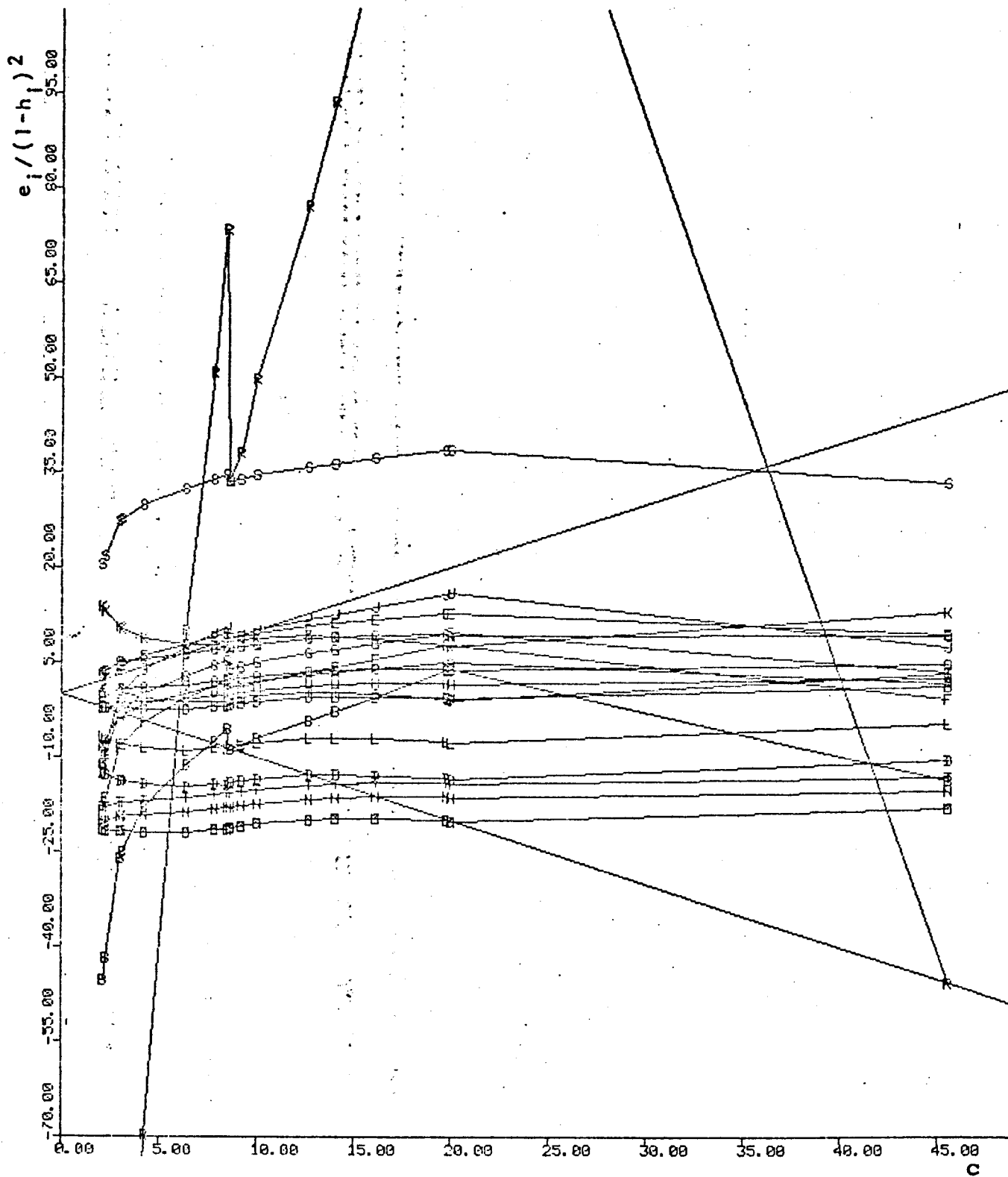


Figura 4.32 Gráfico de resíduos x c - Dados de pontuação de adaptação

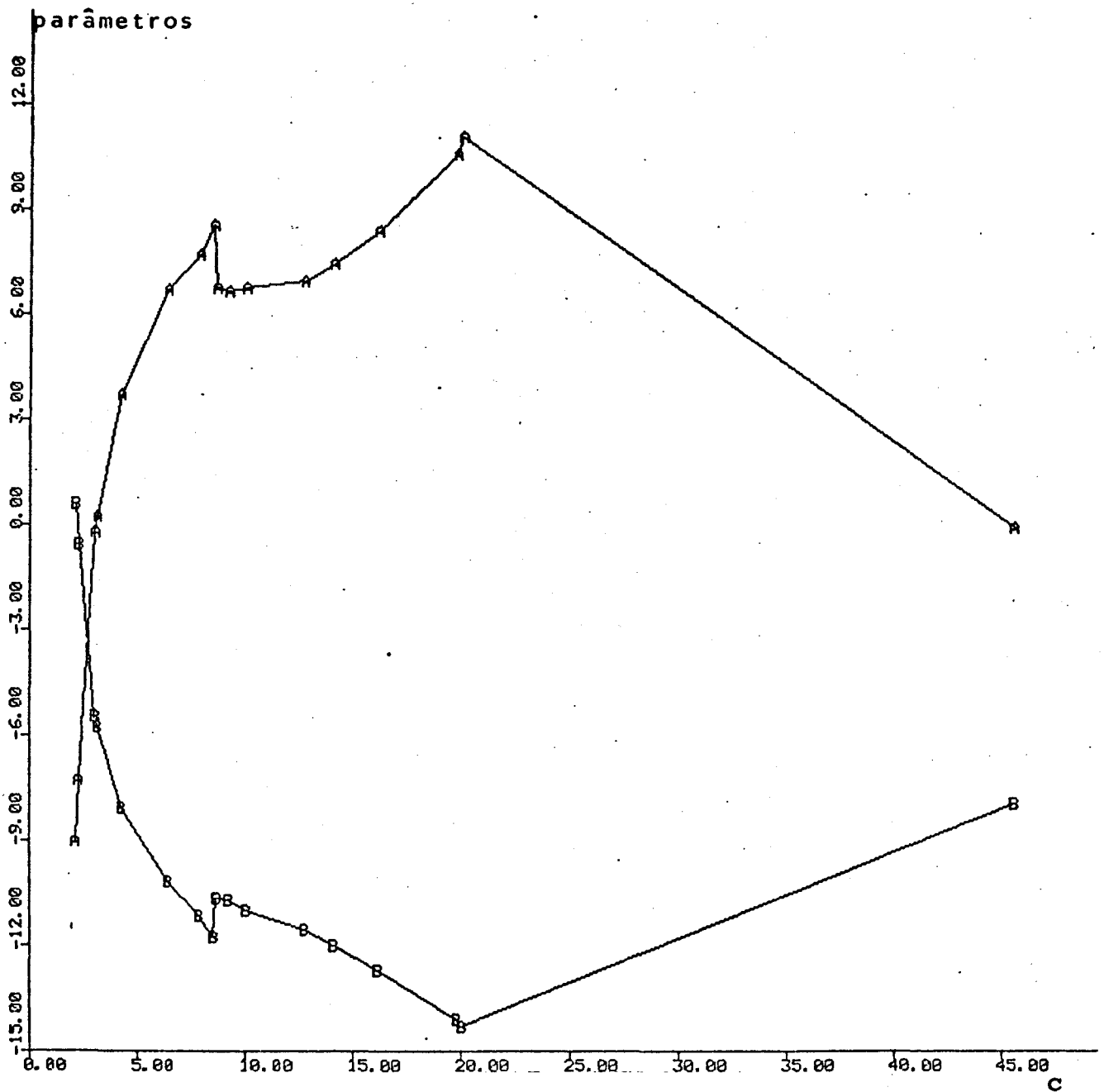


Figura 4.33 Gráfico de parâmetros x c - Dados de pontuação de adaptação

Cook e Weisberg[11], estudando estes mesmos dados, chegaram a conclusão de que o par (2,18) é altamente influente. Não mencionam, entretanto, qualquer subconjunto com mais de 2 pontos.

#  
Voltando ao exemplo 4.1, dados de amônia não absorvida, construímos os gráficos de diagnóstico com a ponderação  $1-h$  às potências 0.5, 1 e 2. Nas figuras 4.34 e 4.35, temos os gráficos de diagnóstico resíduos/(1-h) versus c e parâmetros versus c, respectivamente. Podemos ver de maneira mais clara na figura 4.34, que os resíduos das primeiras quatro observações possuem comportamento semelhante com c diminuindo e principalmente sofrem grandes desvios quando o ponto 2 (B) é atenuado. Nesta figura, notamos um fato interessante. As linhas E,F,G,H e I, correspondentes aos pontos 5,6,7,8 e 9 também se comportam de maneira semelhante aos pontos 1,2,3 e 4, embora os desvios sofridos por estes sejam em menor escala. Mais ainda, as linhas J,K,L,M e N, correspondentes aos pontos 10,11,12,13 e 14, comportam-se de maneira muito semelhante depois que o ponto 4 é atenuado. Em seguida, os resíduos dos pontos 15 e 16, representados pelas letras O e P, desvia-se para valores menores no momento em que 2 é atenuado. Todos consecutivos. Isto deve-se, talvez, ao fato de as observações terem sido tomados em dias consecutivos.

Calculamos mais alguma medidas de influência múltipla envolvendo estes novos pontos e constatamos que os subconjuntos formados pelos pontos 10,11,12,13,14 e 5,6,7,8,9 são influentes. Ainda, se unirmos os subconjuntos (1,2,3,4) com (5,6,7,8,9) temos um aumento considerável no valor da estatística D e a remoção destes pontos moverá a estimativa para um contorno de praticamente 100% em relação a b. Os pontos 15 e 16 não constituem um subconjunto influente. Entretanto, o subconjunto formado por estes dois pontos e 1,2,3,4 também causa um crescimento considerável em D.

subconjunto	D
(1,2,3,4,5,6,7,8,9)	27.1469
(5,6,7,8,9)	2.5621
(10,11,12,13,14)	0.9870
(15,16)	0.0914
(1,2,3,4,15,16)	11.5548

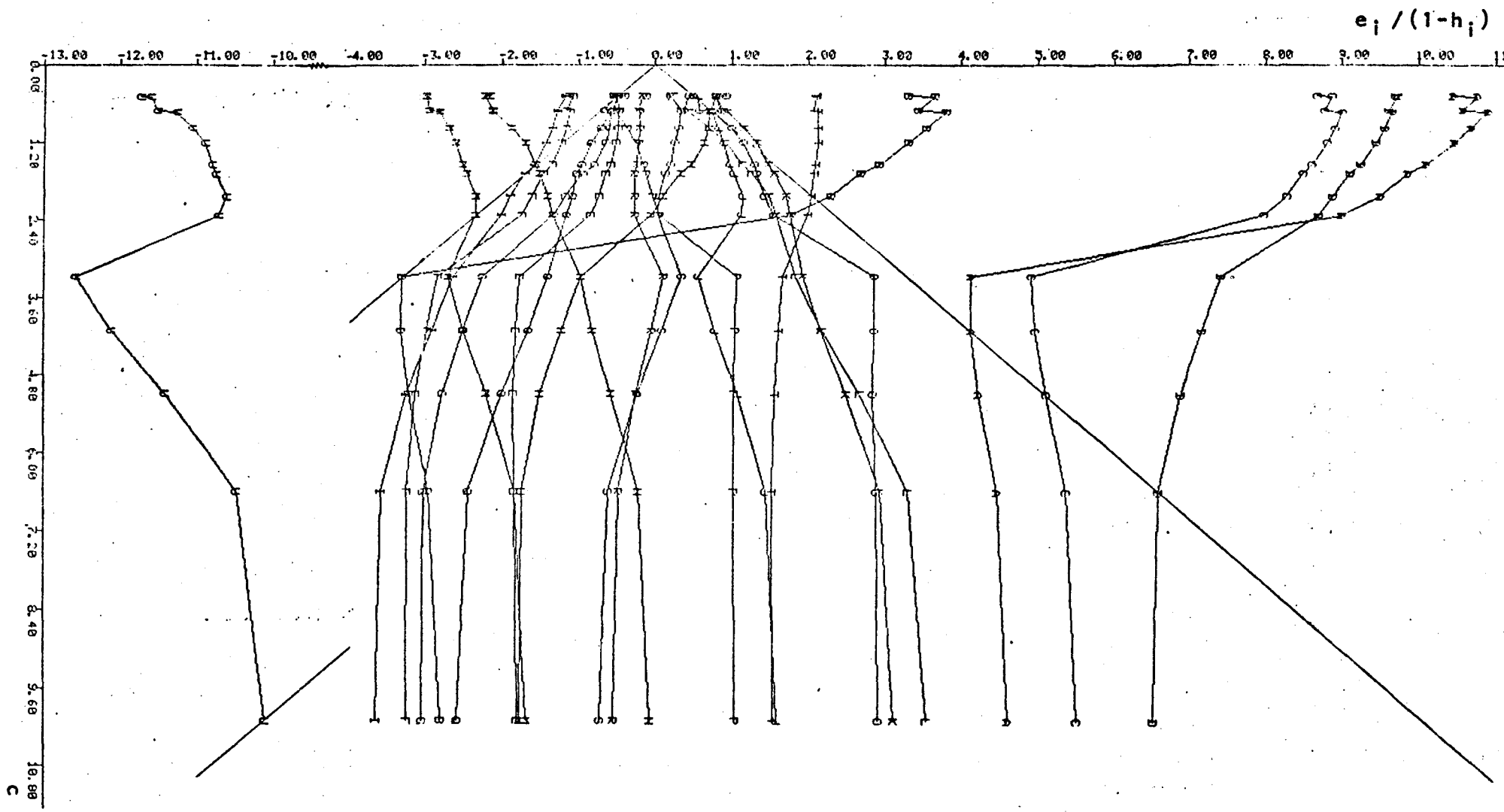


Figura 4.34 Gráfico de resíduos x c - Dados de amonia não absorvida

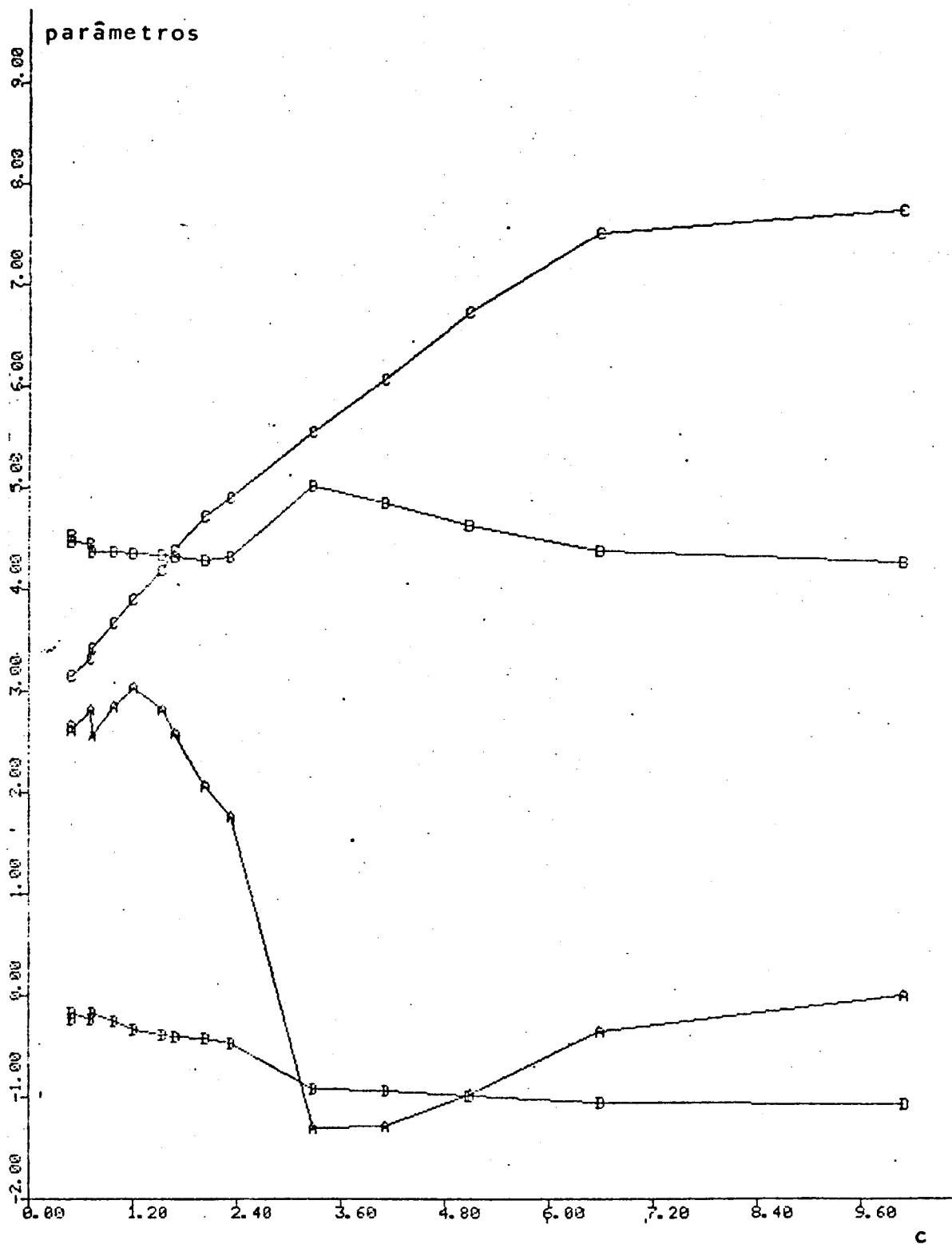


Figura 4.35 Gráfico de parâmetros x c - Dados de amonia não absorvida

## C A P Í T U L O 5

### CONCLUSÃO

Como foi discutido em 4.2, a detecção de subconjunto de pontos influente esbarra em limitações sérias em sua identificação dado o número muito grande de cálculos que deveriam ser feitos. Preocupamo-nos então em encontrar um método alternativo para a determinação de tais subconjuntos. Um método que reduzisse significativamente a quantidade de cálculo. Identificado o subconjunto, o cálculo da medida de influência múltipla para este subconjunto específico é feito sem maiores problemas.

As técnicas gráficas propostas neste trabalho, embora sejam de característica exploratória, mostraram-se bastante eficazes na identificação de subconjuntos de pontos que exercem influência sobre o ajuste.

Com relação à medida de influência múltipla, optamos por trabalhar com a medida proposta por Cook a qual permite fazer comparações através de uma distribuição de referência, no caso a distribuição F.

Um outro aspecto muito importante que o estudo destes gráficos nos permitiu visualizar é com relação à ponderação nos resíduos proposta por Huber [17]. Recordando, Huber propôs tomar o ponto de corte da função em (2.4.2) como  $c \times (1-h_i)^{1/2}$  como forma de atribuir pesos menores a pontos com influência grande no ajuste. Utilizando esta sugestão, construímos os nossos gráficos de diagnóstico e notamos que esta ponderação não é suficiente para que pontos influentes sejam atenuados. Pudemos notar que um ponto conseguiria um destaque maior, somente quando o elemento da diagonal da matriz de projeções correspondente for muito próximo de um. Propusemos, então, generalizar a potência de  $(1-h)$  e tomamos para os nossos propósitos, a potência 1 e 2. Com isso, os pontos influentes passaram a ser melhor evidenciados e conseguimos fazer com que estes fossem considerados no início do processo de indução para a determinação do conjunto C.

Evidentemente, ao aumentarmos a potência de  $1-h$ , podemos estar dando importância maior a pontos que não sejam de fato importantes. Entretanto, em nosso caso, isto pode ser colocado em cheque calculando as medidas de influência múltipla para os subconjuntos encontrados.

Em todos os exemplos deste trabalho, foram construídos os gráficos de diagnóstico com as potências  $k=0, 0.5, 1$  e  $2$  em

1-h. Em todos, os gráficos que permitiram visualizar melhor e conseqüentemente identificar possíveis subconjuntos influentes, foram aqueles com  $k=1$  ou  $2$ , dependendo do caso.

Em suma, dado um problema de regressão linear e encontrado o melhor ajuste, recomendamos em seguida encontrar as medidas de influência unidimensionais, construir os gráficos de diagnóstico com  $k=0, 1$  e  $2$  e calcular as medidas de influência múltipla para cada possível subconjunto que os gráficos indicarem.

Uma desvantagem deste método é que quando o número de observações for grande, o gráfico de resíduos versus o parâmetro escala  $c$  pode se tornar difícil de ser interpretado. Nestes casos, sugerimos construir dois gráficos, ou quantos forem necessários, com os dados particionados.

Os programas computacionais elaborados para este trabalho, encontram-se no apêndice e são em número de três. O primeiro, chamado 'regressão', encontra o ajuste por mínimos quadrados e fornece os arquivos de entrada para os outros programas. O segundo, chamado 'gráficos', fornece os dados para a construção dos gráficos de diagnóstico e o terceiro, calcula a medida de influência múltipla.

Estes programas, embora envolvam quantidade considerável de cálculos e de memória, não são demorados. Num problema com 24 observações e 10 parâmetros, o programa 'gráficos' que é o mais lento, precisou de aproximadamente 2 minutos.

Todos os programas foram escritos em pascal e o equipamento utilizado foi o Itautec I-7000. Por terem sido escritos e executados em microcomputador, estes programas sofrem uma restrição séria quanto ao número de observações e de parâmetros em virtude da capacidade de memória do equipamento.

Os gráficos foram construídos utilizando a rotina VPLOTT e a impressora gráfica do equipamento VAX/Sistema VMS versão 4.2.



## A P E N D I C E

### PROGRAMAS

#### 1. Regressão.

Este programa foi elaborado exclusivamente para efetuar o ajuste por mínimos quadrados, fornecer as medidas de diagnóstico unidimensionais e gerar arquivos que servirão de entrada para os outros dois programas.

```

program regressao;
type mat=array[1..32,1..32] of real;
   vetor=array[1..32] of real;
   vetori=array[1..11] of real;
var
   x,h           :mat;
   y,e           :vetor;
   dados,saida,saida1,saida2 :text;
   arqsai,arquivo,arqdat,arqi:string[14];
   nlin,ncol,i,j,k,l,chave   :integer;
   colin         :array[1..11] of boolean;
   beep,lf,cr,chv   :char;
   ivar          :vetori;
   sqres,s,si,dmin  :real;
procedure multmv(c:mat;d:vetor;nlc,ncc:integer;chv:char);
var
   ik,ij        :integer;
begin
   for ik:=1 to nlc do
      begin
         e[ik]:=0;
         ij:=1;
         repeat
            e[ik]:=e[ik]+c[ik,ij]*d[ij];
            ij:=ij+1;
         until ij>ncc;
         if ((chv='s') or (chv='S')) then
            writeln(saida,e[ik]:15:4);
      end;
   end;
procedure multiplica(r,t:mat;nlr,nlt,nct,f1,f2:integer;chv:char);
var
   ik,il,ij     :integer;
begin
   for ik:=1 to nlr do
      begin
         for il:=1 to nct do

```

```

begin
  h[ik,il]:=0;
  ij:=1;
  repeat
    h[ik,il]:=h[ik,il]+r[ik,ij]*t[ij,il];
    ij:=ij+1;
  until ij>nlt;
  if ((chv='s') or (chv='S')) then
    write(saida,h[ik,il]:f1:f2);
  end;
  if ((chv='s') or (chv='S')) then
    writeln(saida);
  end;
end;
end;
procedure transpoe(g:mat);
var
  ij,ik      : integer;
begin
  for ij:=1 to nlin do
    for ik:=1 to ncol do
      x[ik,ij]:=g[ij,ik];
    end;
  end;
{$i sweep2.pas}
begin
  clrscr;
  beep:=chr(7);l:=chr(10);cr:=chr(13);
  writeln(beep,lf,lf,' ':10,'DIGITE O NOME DO SEU ARQUIVO DE
    DADOS E TECLE ENTER',lf);
  readln(arquivo);
  writeln(beep,lf,lf,' ':10,'DIGITE O NOME DO SEU ARQUIVO DE
    SAIDA E TECLE ENTER',lf);
  readln(arqsai);
  writeln(beep,lf,lf,cr,' ':10,'QUAL O NOME DO ARQUIVO DE DADOS
    PARA OS GRAFICOS?',lf,cr);
  readln(arq1);
  assign(dados,arquivo);assign(saida,arqsai);assign(saida2,arq1);
  reset(dados);rewrite(saida);rewrite(saida2);
  writeln(beep,lf,lf,' ':15,'A REGRESSAO E COM CONSTANTE?
    (S/N)',lf);
  readln(chv);
  writeln(saida2,chv);
  readln(dados,nlin,ncol);
  if ((chv='s') or (chv='S')) then
    begin
      chave:=2;
      ncol:=ncol+1;
      for i:=1 to nlin do
        begin
          x[i,1]:=1;h[1,i]:=1;
        end;
      end;
    end
  else
    chave:=1;
  for j:=1 to ncol do
    ivar[j]:=1;
  end;
end;

```

```

dmin:=1e-06;
for i:=1 to nlin do
  begin
    for j:=chave to ncol do
      begin
        read(dados,x[i,j]);
        h[j,i]:=x[i,j];
      end;
    read(dados,y[i]);
    readln(dados);
  end;
write(saida,1f,1f,1f,' ':25,'MATRIZ XTX',1f,1f,cr);
multiplica(h,x,ncol,nlin,ncol,15,4,'s');
for i:=1 to ncol do
  sweep(ncol,i,h);
write(saida,1f,1f,1f,' ':15,'MATRIZ INVERSA DE XTX ',1f,1f,cr);
for i:=1 to ncol do
  begin
    for j:=i to ncol do
      begin
        if j>i then h[j,i]:=h[i,j];
        write(saida,h[i,j]:15:6);
        writeln(saida2,h[i,j]);
      end;
    writeln(saida);
  end;
transpoe(x);
multiplica(h,x,ncol,ncol,nlin,8,4,'n');
write(saida,1f,1f,1f,' ':15,'OS PARAMETROS ESTIMADOS
SAO',1f,1f,cr);
multmv(h,y,ncol,nlin,'s');
for i:=1 to ncol do
  writeln(saida2,e[i]);
transpoe(x);
multiplica(x,h,nlin,ncol,nlin,8,4,'n');
multmv(h,y,nlin,nlin,'n');
sqres:=0;
writeln(beep,1f,1f,' ':15,'QUAL O NOME DO ARQUIVO DE
SAIDA?(MATRIZ CHAPEU)',1f,cr);
readln(arqdat);assign(saida1,arqdat);rewrite(saida1);
writeln(saida1,nlin:5,ncol:5);
for i:=1 to nlin do
  begin
    for k:=1 to nlin do
      writeln(saida1,h[i,k]:8:6);
      x[i,1]:=e[i];
      x[i,2]:=y[i]-e[i];
      x[i,3]:=h[i,i];
      writeln(saida2,x[i,2],x[i,3]);
      sqres:=sqres+sqr(x[i,2]);
    end;
s:=sqres/(nlin-ncol);
write(saida,1f,1f,1f,' ':10,'yaj', ' ':14,'e', ' ':14,'h', ' '
:13,'t', ' ':13);
write(saida,'tal', ' ':13,'D', ' ':12,'DFFITS',1f,1f,cr);

```

```

for i:=1 to nlin do
  begin
    writeln(saida1,x[i,2]);
    x[i,4]:=x[i,2]/sqrt(s*(1-h[i,i]));
    si:=(nlin-ncol)*s/(nlin-ncol-1)-sqr(x[i,2])/
      ((nlin-ncol-1)*(1-h[i,i]));
    x[i,5]:=x[i,2]/sqrt(si*(1-h[i,i]));
    x[i,6]:=sqr(x[i,4])*h[i,i]/(ncol*(1-h[i,i]));
    x[i,7]:=x[i,5]*(sqrt(h[i,i]/(1-h[i,i])));
    for j:=1 to 7 do
      write(saida,x[i,j]:15:4);
    writeln(saida);
  end;
writeln(saida1,s);
close(saida2);
close(saida1);
close(saida);
end.

```

## 2. Graficos.

Fornece os dados para a construção dos gráficos de diagnóstico. Este programa foi elaborado segundo o algoritmo proposto por Denby e Mallows em [13].

```

program denby_mallows;
type
  mat      =array[1..10,1..10] of real;
  mati     =array[1..31,1..10] of real;
  vet      =array[1..11] of real;
  vetor    =array[1..31] of real;
  nome     =string[14];

var
  i,j,g,u,nl,nc,modo :integer;
  indice,chave,sinal :integer;
  h,h1             :mat;
  b,b1,q,xi,bh,bh1 :vet;
  r,y,diag,r1     :vetor;
  arquivo,arqsai,arqi :nome;
  grafpar,grafres  :nome;
  dados,saida,data :text;
  grf,grfi         :text;
  beep,lf,cr,chv,cont :char;
  mm,max,interc    :real;
  quart           :array[1..2,1..11] of real;
  ind             :array[1..31] of integer;
  x              :mati;

($i multmv.pas)
procedure prepara;
begin
  clrscr;beep:=chr(7);lf:=chr(10);cr:=chr(13);
  writeln(beep,lf,lf,' ':10,'ENTRE COM O NOME DO ARQUIVO DE DADOS
    E TECLE ENTER',lf,cr);
  readln(arquivo);

```

```

writeln(beep,1f,1f,' ':10,'ENTRE COM O NOME DO ARQUIVO DE DADOS
      (MATRIZ X E VETOR Y)',1f,cr);
readln(arq1);
assign(dados,arquivo);assign(data,arq1);reset(data);
reset(dados);
end;
procedure prepara1;
begin
  writeln(beep,1f,1f,' ':10,'QUAL A POTENCIA DE (1-h)?
      (0 se 0,5 se 0.5,1 se 1 e 2 se 2)',1f,cr);
  readln(modo);
  writeln(beep,1f,1f,' ':10,'ENTRE COM O NOME DO ARQUIVO DE SAIDA
      E TECLE ENTER',1f,cr);
  readln(arqsai);
  assign(saida,arqsai);rewrite(saida);
  for i:=1 to nl do
    begin
      case modo of
        0:r[i]:=r1[i];
        5:r[i]:=r1[i]/sqrt(1-diag[i]);
        1:r[i]:=r1[i]/(1-diag[i]);
        2:r[i]:=r1[i]/sqr(1-diag[i]);
      end;
      ind[i]:=0;
    end;
  for i:=1 to nc do
    begin
      for j:=1 to nc do
        h[i,j]:=h1[i,j];
        b[i]:=b1[i];
        bh1[i]:=0;
        bh[i]:=0;
        q[i]:=0;
      end;
      interc:=b[i];
      for i:=2 to nc do
        bh1[i]:=b[i]*(quart[2,i]-quart[1,i]);
      g:=1;
      writeln(beep,1f,1f,' ':10,'ENTRE COM O NOME DO ARQUIVO PARA
          PLOTTER(RESIDUOS)',1f,cr);
      readln(grafres);
      writeln(beep,1f,1f,' ':10,'ENTRE COM O NOME DO ARQUIVO PARA
          PLOTTER(PARAMETROS)',1f,cr);
      readln(grafpar);
      assign(grf,grafres);rewrite(grf);assign(grf1,grafpar);
      rewrite(grf1);
    end;
  procedure leitura;
  begin
    readln(dados,chv);
    readln(data,nl,nc);
    if chv in ['s','S'] then
      begin
        chave:=2;
        nc:=nc+1;
      end;
  end;
end;

```

```

        for i:=1 to n1 do
            x[i,1]:=1;
        end
    else
        chave:=1;
    for i:=1 to nc do
        for j:=1 to nc do
            readln(dados,h[i,j]);
        for i:=1 to nc do
            readln(dados,b[i]);
        for i:=1 to n1 do
            begin
                readln(dados,r[i],diag[i]);
                for j:=chave to nc do
                    read(data,x[i,j]);
                read(data,y[i]);
                readln(data);
            end;
        end;
    procedure quartis(ab:mat1);
    var
        w      :real;
        troca,v :integer;
    begin
        for i:=2 to nc do
            begin
                ab[22,i]:=1e+06;
                repeat
                    troca:=0;
                    for j:=1 to n1 do
                        begin
                            if ab[j+1,i]<ab[j,i]
                                then
                                    begin
                                        w:=ab[j,i];
                                        ab[j,i]:=ab[j+1,i];
                                        ab[j+1,i]:=w;
                                        troca:=1;
                                    end;
                                end;
                            until troca=0;
                        end;
                    u:=trunc(n1/4);
                    if n1/4-u>0
                        then
                            for i:=2 to nc do
                                begin
                                    quart[1,i]:=ab[u+1,i];
                                    quart[2,i]:=ab[n1-u,i];
                                end
                            else
                                for i:=2 to nc do
                                    begin
                                        quart[1,i]:=(ab[u,i]+ab[u+1,i])/2;
                                        quart[2,i]:=(ab[n1-u-1,i]+ab[n1-u,i])/2;

```

```

end;
end;
procedure maxabs(z:vetor);
begin
  max:=z[1]; indice:=1;
  for i:=2 to n1 do
    if abs(z[i])>abs(max)
      then
        begin
          max:=z[i];
          indice:=i;
          sinal:=trunc(max/abs(z[i]));
        end;
  for i:=1 to nc do
    xi[i]:=x[indice,i];
  ind[i]:=indice;
end;
procedure maximo(z:vetor);
begin
  max:=z[1]; indice:=1;
  for i:=2 to n1 do
    if z[i]>max
      then
        begin
          max:=z[i];
          indice:=i;
        end;
end;
procedure itera(a:mat;aux:vet);
var
  d,f,f1 :real;
begin
  d:=0; f:=0; f1:=0;
  for i:=1 to nc do
    begin
      f:=f+aux[i]*b[i];
      f1:=f1+aux[i]*q[i];
    end;
  multmv(nc,nc,a,aux);
  for i:=1 to nc do
    d:=d+aux[i]*xi[i];
  for i:=1 to nc do
    begin
      for j:=1 to nc do
        begin
          h[i,j]:=aux[i]*aux[j];
          h[i,j]:=h[i,j]/(1-d);
          h[i,j]:=a[i,j]+h[i,j];
        end;
      b[i]:=b[i]-(y[indice]-f)*aux[i]/(1-d);
      q[i]:=q[i]+(f1+sinal)*aux[i]/(1-d);
    end;
end;
end;
procedure calcula;
var

```

```

f, f1, max1, max2, w           :real;
aga1, aga2                     :vetor;
indice1                         :integer;
begin
  for u:=1 to n1 do
    begin
      f:=0; f1:=0;
      for i:=1 to nc do
        begin
          f:=f+x[u, i]*b[i];
          f1:=f1+x[u, i]*q[i];
        end;
      case modo of
        0:w:=1+f1;
        5:w:=sqrt(1-diag[u])+f1;
        1:w:=(1-diag[u])+f1;
        2:w:=sqr(1-diag[u])+f1;
      end;
      if w=0 then aga1[u]:=0 else aga1[u]:=(y[u]-f)/w;
      case modo of
        0:w:=-1+f1;
        5:w:=-sqrt(1-diag[u])+f1;
        1:w:=-1-diag[u]+f1;
        2:w:=-sqr(1-diag[u])+f1;
      end;
      if w=0 then aga2[u]:=0 else aga2[u]:=(y[u]-f)/w;
    end;
  for i:=1 to g-1 do
    begin
      j:=ind[i];
      aga1[j]:=0;
      aga2[j]:=0;
    end;
  mm:=abs(max);
  repeat
    maximo(aga1);
    if max>=mm-1e-06 then aga1[indice]:=0;
  until max<mm-1e-06;
  max1:=max;
  indice1:=indice;
  repeat
    maximo(aga2);
    if max>=mm-1e-06 then aga2[indice]:=0;
  until max<mm-1e-06;
  max2:=max;
  if max1>max2
    then
      begin
        max:=max1;
        indice:=indice1;
      end
    else
      max:=max2;
  ind[g]:=indice;
  for i:=1 to nc do

```



```

begin
  xi[i]:=x[indice,i];
  bh[i]:=b[i]+abs(max)*q[i];
  bh1[i]:=bh[i]*(quart[2,i]-quart[1,i]);
end;
bh1[1]:=bh[1]-interc;
for i:=1 to n1 do
begin
  f:=0;
  for j:=1 to nc do
    f:=f+x[i,j]*bh[j];
  case modo of
    0:r[i]:=(y[i]-f);
    5:r[i]:=(y[i]-f)/sqrt(1-diag[i]);
    1:r[i]:=(y[i]-f)/(1-diag[i]);
    2:r[i]:=(y[i]-f)/sqr(1-diag[i]);
  end;
end;
sinal:=trunc(r[indice]/abs(r[indice]));
end;
procedure escreve;
begin
  for i:=1 to 74 do
    write(saida,#45);
  writeln(saida,1f,1f,' ':35,'g = ',g);
  write(saida,1f,1f,' ':6,'PARAMETROS',' ':6,'PARAMETROS(h)');
  writeln(saida,' ':5,'PARAMETROS(h) PADR.','
    ':5,'RESIDUOS',1f,1f,cr);
  for i:=1 to nc do
    writeln(saida,' ':2,'b',i,' ',b[i]:10:4,' ':6,bh[i]:10:4,
      ' ':8,bh1[i]:10:4,' ':10,i:2,r[i]:10:4);
  for i:=nc+1 to n1 do
    writeln(saida,' ':60,i:2,r[i]:10:4);
  write(saida,1f,1f);
  writeln(saida,' ':10,'h(',g,') = ',abs(max):8:4,' ':5,'linha',
    indice:4);
  write(saida,1f,1f,1f,cr);
  write(grf,abs(max):8:4);
  write(grf1,abs(max):8:4);
  for i:=1 to n1 do
    write(grf,r[i]:8:4);
  writeln(grf);
  for i:=1 to nc do
    write(grf1,bh1[i]:10:4);
  writeln(grf1);
end;
begin(programa principal)
  prepara;
  leitura;
  quartis(x);
  cont:='s';
  while cont in ['s','S'] do
    begin
      prepara1;
      maxabs(r);
    end;
end;

```

```

        escreve;
        itera(h,xi);
        g:=2;
        repeat
            calcula;
            escreve;
            itera(h,xi);
            g:=g+1;
        until g>nl-nc-2;
        close(saida);
        close(grf);
        close(grf1);
        writeln(beep,lf,lf,' ':10,'DESEJA CONTINUAR?
                (S/N)',lf,cr);
        readln(cont);
    end;
end.

```

### 3. Influência múltipla.

Calcula a medida de influência múltipla sugerida por Cook e Weisberg em [11].

```

program minfl;
type mat=array[1..10,1..10] of real;
     vetor=array[1..10] of real;
var
    i,j,k,l,m,ncol : integer;
    ei,ivar         : vetor;
    indice          : array[1..10] of integer;
    arqsai,arquivo  : string[14];
    beep,lf,cr,cont : char;
    hi,ident,ab     : mat;
    dados,saida     : text;
    colin           : array[1..10] of boolean;
    d,s,dmin        : real;
    e               : array[1..32] of real;
    h               : array[1..32,1..32] of real;
procedure prepara;
begin
    clrscr;beep:=chr(7);lf:=chr(10);cr:=chr(13);
    writeln(beep,lf,' ':15,'QUAL O NOME DO ARQUIVO DE
              DADOS?(MATRIZ CHAPEU)',lf);
    readln(arquivo);
    assign(dados,arquivo);reset(dados);
    writeln(beep,lf,lf,' ':15,'QUAL O NOME DO ARQUIVO DE
              SAIDA?',lf);
    readln(arqsai);
    assign(saida,arqsai);rewrite(saida);
end;
procedure prepara1;
begin
    writeln(beep,lf,' ':15,'QUAL O TAMANHO DO CONJUNTO?',lf);
    readln(m);

```

```

writeln(beep,lf,lf,' ':15,'ENTRE COM O NUMERO DA LINHA E
TECLE ENTER',lf);
for i:=1 to m do
  readln(indice[i]);
end;
procedure leitura;(le a matriz chapeu)
var
  n,kk          : integer;
begin
  readln(dados,n,ncol);
  for i:=1 to n do
    for j:=1 to n do
      readln(dados,h[i,j]);
    for i:=1 to n do
      readln(dados,e[i]);
  readln(dados,s);
  for i:=1 to 10 do
    for j:=1 to 10 do
      ident[i,j]:=0;
  dmin:=1e-06;
end;
procedure menor;(constroi o menor de dimensao m)
var
  kk          : integer;
begin
  for j:=1 to m do
    begin
      k:=indice[j];
      ei[j]:=e[k];
      ivar[j]:=1;
      for l:=1 to m do
        begin
          kk:=indice[l];
          hi[j,l]:=h[k,kk];
          if j=1
            then
              ident[j,l]:=1
            else
              ident[j,l]:=0;
          ident[j,l]:=ident[j,l]-hi[j,l];
        end;
      end;
  end;
end;
($i sweep2.pas)
($i multmm.pas)
($i multmvi.pas)
procedure calcula;
begin
  for k:=1 to m do
    sweep(m,k,ident);
  for i:=1 to m do
    begin
      for j:=1 to m do
        if j>i then ident[j,i]:=ident[i,j];
      end;
    end;
end;

```

```

multiplica(m,m,m,ident,hi);
multiplica(m,m,m,hi,ident);
multmv(ident,ei,m,m);
d:=0;
for i:=1 to m do
  d:=d+ei[i]*ivar[i];
d:=d/(s*ncol);
write(saida,' ':10,'INFLUENCIA MULTIPLA DOS PONTOS');
for i:=1 to m do
  write(saida,indice[i]:4);
writeln(saida);writeln(saida);
writeln(saida,' ':20,d:10:4,lf,lf,lf,cr);
end;
begin(programa principal)
  prepara;
  leitura;
  cont:='s';
  while ((cont='s') or (cont='S')) do
    begin
      prepara1;
      menor;
      calcula;
      writeln(' ':10,'QUER CALCULAR PARA MAIS CONJUNTO?
              (S/N)');
      readln(cot);
    end;
  close(saida);
end.

```

#### 4. Procedures inclusos nos programas anteriores.

##### 4.1 Sweep. Inverte uma matriz simétrica.

```

PROCEDURE sweep (numesc,kk:INTEGER;var matriz:MAT) ;
VAR
  ii,jj      : INTEGER ;
  de,be,ce   : REAL ;

BEGIN
  de := matriz[kk,kk] ;
  IF ((de<dmin) AND (ivar[kk]=1)) THEN colin[kk] := TRUE ;
  IF NOT colin[kk] THEN BEGIN
    FOR ii:=1 TO numesc+1 DO BEGIN
      IF ii<>kk THEN BEGIN
        IF ii>kk THEN be :=
          ivar[ii]*ivar[kk]*matriz[kk,ii]/de ;
        IF ii<kk THEN be := matriz[ii,kk]/de ;
        FOR jj:=ii TO numesc+1 DO BEGIN
          IF jj<>kk THEN BEGIN
            IF jj<kk THEN ce :=
              ivar[jj]*ivar[kk]*matriz[jj,kk] ;
            IF jj>kk THEN ce := matriz[kk,jj] ;
            matriz[ii,jj] := matriz[ii,jj] - be*ce ;
          END ;
        END ;
      END ;
    END ;
  END ;

```

```

        END ;
    END ;
END ;
IF kk<>1 THEN FOR ii:=1 TO kk-1 DO
    matriz[ii,kk] := -matriz[ii,kk]/de ;
FOR ii:=kk+1 TO numesc+1 DO
    matriz[kk,ii] := matriz[kk,ii]/de ;
matriz[kk,kk] := 1/de ;
ivar[kk] := - ivar[kk] ;
END ;
END ;

```

#### 4.2 Multiplica. Multiplica duas matrizes.

```

procedure multiplica(nla,nlb,ncb:integer;a:mat;var b:mat);
var
    ik,il,ij    :integer;
    aux         :mat;
begin
    for ik:=1 to nla do
        for il:=1 to ncb do
            begin
                aux[ik,il]:=0;
                ij:=1;
                repeat
                    aux[ik,il]:=aux[ik,il]+a[ik,ij]*b[ij,il];
                    ij:=ij+1;
                until ij>nlb;
            end;
        for ik:=1 to nla do
            for il:=1 to ncb do
                b[ik,il]:=aux[ik,il];
            end;
        end;
end;

```

#### 4.3 Multmv. Multiplica matriz por vetor.

```

procedure multmv(nlh,nch:integer;hh:mat;var yy:vet);
var
    ik,ij       :integer;
    iv          :array[1..31] of real;
begin
    for ik:=1 to nlh do
        begin
            iv[ik]:=0;
            ij:=1;
            repeat
                iv[ik]:=iv[ik]+hh[ik,ij]*yy[ij];
                ij:=ij+1;
            until ij>nch;
        end;
        for ik:=1 to nlh do
            yy[ik]:=iv[ik];
        end;
end;

```

#### 4.4 Multmv. Multiplica matriz por vetor(usado em influência múltipla).

```
procedure multmv(h:mat;y:vetor;n1h,nch:integer);
var
    ik,ij      :integer;
begin
    for ik:=1 to n1h do
        begin
            ivar[ik]:=0;
            ij:=1;
            repeat
                ivar[ik]:=ivar[ik]+h[ik,ij]*y[ij];
                ij:=ij+1;
            until ij>nch;
        end;
    end;
end;
```

REFERÊNCIAS

- [ 1 ] ANDREWS D.F. e PREGIBON D.(1978), "Finding the outliers that matter", Journal of the Royal Statistical Society, série B, 40, pp. 85.
- [ 2 ] ATKINSON A.C.(1981), "Robustness, transformations and two graphical displays for outlying and influential observations in regression", Biometrika, 68, pp. 13.
- [ 3 ] BELSLEY D.A., KUH E. e WELSCH R.E.(1980), "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity", New York, Wiley.
- [ 4 ] BRINGHAN C.(1977), "Some Identities Useful in the Analysis of Residuals from Linear Regression", Technical Report 300, School of Statistics, University of Minnesota.
- [ 5 ] BROWNLEE K.A.(1965), "Statistical Theory and Methodology", 2a. ed., New York, Wiley.
- [ 6 ] CARVALHO J.F. e DACHS J.N.W.(19 ), "Regressão", São Carlos, Gráfica UFScar.
- [ 7 ] CHAMBERS R.L. e HEATHCOTE C.R.(1981), "On the estimation of Slope and the Identification of Outliers in Linear Regression", Biometrika, 68, 1, pp. 21.
- [ 8 ] COOK R.D.(1977), "Detection of Influential Observation in Linear Regression", Technometrics, 19, nº 1, pp. 15.
- [ 9 ] COOK R.D.(1979), "Influential Observation in Linear Regression", Journal of American Statistical Association, 74, pp. 169.
- [10] COOK R.D. e WEISBERG S.(1980), "Characterization of an Empirical Influence Function for Detecting Influential Cases in Regression", Technometrics, 22, nº 4, pp. 495.
- [11] COOK R.D. e WEISBERG S.(1982), "Residuals and Influence in Regression", New York, Wiley.
- [12] DANIEL C. e WOOD F.(1980), "Fitting Equations to Data", 2a. ed., New York, Wiley.
- [13] DENBY L. e MALLOWS C.L.(1977), "Two Diagnostic Displays for Robust Regression Analysis", Technometrics, 19, nº 1, pp. 01.
- [14] DRAPER N.R. e SMITH H.(1966), "Applied Regression Analysis", New York, Wiley.

- [15] DRAPER N.R. e SMITH H.(1966), "Applied Regression Analysis", 2a. ed., New York, Wiley.
- [16] HOAGLIN D.C. e WELSH R.E.(1978), "The Hat Matrix in Regression and ANOVA", The American Statistician, 32, n° 1, pp. 17.
- [17] HUBER P.(1981), "Robust Statistics", New York, Wiley.
- [18] MICKEY M.R., DUNN O.J. e CLARK V.(1967), "Note on the Use of Stepwise Regression in Detecting Outliers", Computers and Biomedical Research, 1, pp. 105.
- [19] MONTGOMERY D.C. e PECK E.A.(1982), "Introduction to Linear Regression Analysis", New York, Wiley.
- [20] RAD C.R.(1973), "Linear Statistical Inference and its Applications", 2a. ed., New York, Wiley.
- [21] SHERMAN J. e MORRISON W.J.(1949), "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix"(abstract), Annals of Mathematical Statistics, 20, pp.621.
- [22] WEISBERG S.(1980), "Applied Linear Regression", New York, Wiley.