

UM ESTUDO COMPARATIVO
DO DFFITS E DO D DE COOK

FRANCISCO ANTONIO ROJAS ROJAS



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

R638e

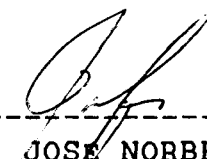
7170/BC

CAMPINAS - SÃO PAULO
BRASIL

U M E S T U D O C O M P A R A T I V O
D O D F F I T S E D O D D E C O O K

Este exemplar corresponde a
redação final da tese defendida
e devidamente corrigida pelo
Senhor FRANCISCO ANTONIO ROJAS
ROJAS e aprovada pela Comissão
Julgadora.

Campinas, 30 de Maio de 1986.



Prof. Dr. JOSÉ NORBERTO W. DACHS
Orientador

Aos meus pais.

As minhas irmãs.

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

Ao Prof. Dr. José Norberto Walter Dachs o apoio que me deu ao longo do mestrado e a segura orientação durante o desenvolvimento deste trabalho.

Aos professores e colegas do IMECC que de forma sincera me ofereceram sua amizade.

Ao Dr. Oscar Landmann, consul honorário da Colômbia em São Paulo, sua amizade e apoio.

A minha futura esposa, Pérsida Rodrigues F., toda a ajuda e a dedicação que me ofereceu na conclusão deste trabalho.

Ao CNPq e a CAPES o apoio financeiro.

Francisco Antonio Rojas Rojas.

S U M A R I O

O problema de diagnóstico em ajustes, lineares ou não, tem obtido grande atenção na literatura nos últimos oito anos, notadamente nos últimos cinco.

Para o caso de ajustes lineares as duas técnicas mais comumente usadas hoje são o D de Cook e o DFFITS.

Neste trabalho são provadas algumas propriedades dessas estatísticas e, principalmente, estudado empiricamente seu desempenho para um conjunto de dados.

Conclue-se que o DFFITS deve ser preferido com relação ao D de Cook por sua capacidade não só de detectar possíveis pontos influentes no espaço dos x's mas também valores aberrantes dos y's.

A B S T R A C T

The problem of diagnostics in linear and non linear fitting has deserved great attention in the technical literature for the past eight years, markedly in the last five ones.

In the linear fitting problem the two most commonly used techniques are Cook's D and DFFITS.

In this work some properties of these statistics are proved but, mainly their performance is studied empirically for a selected data set.

The conclusion is that the DFFITS should be preferred to Cook's D, due to its ability to detect not only potentially influential points in the x's space but also outlying values in the y's.

I N D I C E

	Página
NOTAÇÃO	viii
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xii
LISTA DE QUADROS	xiv
INTRODUÇÃO	1
1- ELEMENTOS BASICOS PARA A ANALISE DE REGRESSÃO LINEAR	
1.1- MODELO LINEAR	4
1.2- AJUSTE PELO MÉTODO DE MÍNIMOS QUADRADOS	5
1.3- INTERPRETAÇÃO GEOMETRICA DO AJUSTE POR MÍNIMOS QUADRADOS	6
1.4- A MATRIZ DE PROJEÇÕES, H	8
1.5- RESTRIÇÕES NOS PARAMETROS E ACRESCIMO NA SOMA DE QUADRADOS DE RESÍDUOS	8
1.6- ESTRUTURA PROBABILÍSTICA DO ERRO	10
1.7- DIAGNÓSTICO	11
2- A DIAGONAL DA MATRIZ DE PROJEÇÕES, H	15
3- O D DE COOK	19
4- O DFFITS	26
5- COMPARAÇÃO DE D DE COOK E DFFITS	
5.1- CONJUNTO DE DADOS DE "STACK-LOSS"	32
5.2- DESEMPENHO DO D DE COOK, DC, E DO DFFITS, DF, PELA VARIAÇÃO DE Y(9) OU DE Y(21)	33
5.3- DESEMPENHO DO D DE COOK E DO DFFITS PELA VARIA- ÇÃO DE X(3,21) OU DE X(3,9)	39
5.4- DESEMPENHO DO DFFITS E DO D DE COOK PELA VARIA-	

AO CONJUNTA DE $Y(i)$ E $X(3,i)$, $i=9$ ou 21	41
6- UMA FORMA ALTERNATIVA DE COMPARAR O DFFITS E O D DE COOK	
6.1- O D RAIZ	60
6.1.A- DESEMPENHO DO D RAIZ, $DR(i)$, COMO FUNÇÃO DE $Y(i)$	61
6.1.B- DESEMPENHO DO D RAIZ COMO FUNÇÃO DE $X(3,i)$	62
6.1.C- DESEMPENHO DO D RAIZ PELA VARIAÇÃO CONJUNTA DE $Y(i)$ E $X(3,i)$	63
6.2- CONCLUSÕES FINAIS	64
6.3- PROGRAMA PARA CÁLCULO DE DIAGNÓSTICOS	74
REFÊNCIAS BIBLIOGRÁFICAS	83

N O T A Ç Ã O

$x(i,j) = x_i^{(j)}$: elemento da i-ésima linha e j-ésima coluna da matriz X do modelo. Valor do j-ésimo preditor na i-ésima observação.

$y(i) = y_i$: i-ésimo elemento do vetor Y. Valor da resposta na i-ésima observação.

$\hat{e}(i) = \hat{e}_i$: i-ésimo elemento do vetor E. Valor do resíduo na i-ésima observação.

$A(i)$: vetor de parâmetros ajustados para o conjunto sem a i-ésima observação.

$h(i,j) = h_{ij}$: elemento da i-ésima linha e j-esima coluna da matriz chapéu.

$h(i,i) = h_{ii} = h_i$: i-ésimo elemento da diagonal da matriz chapéu.

$X_{(i)}$: i-ésima linha da matriz X.

$X^{(j)}$: j-ésima coluna da matriz X.

$D_i = DC(i) = D(i)$: valor do D de Cook para a i-ésima observação.

$DFFITS = DF(i)$: valor do DFFITS para a i-ésima observação.

s^2 : media quadrática de resíduos.

$s^2(i)$: media quadrática de resíduos para o ajuste sem a i-ésima observação.

$\langle \rangle$: diferente de.

$\hat{y}(-i)$: j-ésimo valor ajustado para o conjunto sem a i-ésima observação.

$DR(i) = DR_i$: valor do D Raiz para a i-ésima observação.

L I S T A D E F I G U R A S

Figura nº	Página
5.2.1- D(i) versus Y(21); i=1,2,3,4,9,14,21	43
5.2.2- DFFITS(i) versus Y(21); i=1,2,3,4,9,14,21	43
5.2.3- D(i) versus Y(9); i=1,2,3,4,9,14,21	44
5.2.4- DFFITS(i) versus Y(9); i=1,2,3,4,9,14,21	44
5.3.1- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=35	45
5.3.2- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=35	45
5.3.3- D(i) versus X(3,21) i=1,2,3,4,9,14,21; para Y(21)=30	46
5.3.4- DFFITS(i) versus X(3,21) i=1,2,3,4,9,14,21; para Y(21)=30	46
5.3.5- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=25	47
5.3.6- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=25	47
5.3.7- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=20	48
5.3.8- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=20	48
5.3.9- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=15	49
5.3.10- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=15	49
5.3.11- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=13	50
5.3.12- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=13	50

	x
5.3.13- D(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=5	51
5.3.14- DFFITS(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=5	51
5.3.15- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=30	52
5.3.16- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=30	52
5.3.17- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=25	53
5.3.18- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=25	53
5.3.19- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=20	54
5.3.20- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=20	54
5.3.21- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=17	55
5.3.22- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=17	55
5.3.23- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=15	56
5.3.24- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=15	56
5.3.25- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=13	57
5.3.26- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=13	57
5.3.27- D(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=5	58
5.3.28- DFFITS(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=5	58
5.4.1- DF(21) como função de Y(21) e X(3,21)	59
5.4.2- DF(9) como função de Y(9) e X(3,9)	59
6.1.A.1- DR(i) versus Y(21); i=1,2,3,4,9,14,21	65
6.1.A.2- DR(i) versus Y(9); i=1,2,3,4,9,14,21	65

6.1.B.1-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=35	66
6.1.B.2-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=30	66
6.1.B.3-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=25	67
6.1.B.4-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para (21)=20	67
6.1.B.5-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=17	68
6.1.B.6-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=15	68
6.1.B.7-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=13	69
6.1.B.8-	DR(i) versus X(3,21); i=1,2,3,4,9,14,21; para Y(21)=5	69
6.1.B.9-	DR(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)= 25	70
6.1.B.10-	DR(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=17	70
6.1.B.11-	DR(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)= 15	71
6.1.B.12-	DR(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=13	71
6.1.B.13-	DR(i) versus X(3,9); i=1,2,3,4,9,14,21; para Y(9)=5	72
6.1.C.1-	DR(21) como função de Y(21) e X(3,21)	73
6.1.C.2-	DR(9) como função de Y(9) e X(3,9)	73

LISTA DE TABELAS

Tabela	Página
1.3.1- Tabela de análise de variância	7
5.2.1- Intervalos onde $DF(k)$ como função de $Y(21)$ não detecta como discrepante a k -ésima obser- vação	38
5.2.2- Intervalos onde $DC(k)$ como função de $Y(21)$ naõ detecta como discrepante a k -ésima obser- vação	38
5.2.3- Intervalos onde $DF(k)$ como função de $Y(9)$ não detecta como discrepante a k -ésima obser- vação	38
5.2.4- Intervalos onde $DC(k)$ como função de $Y(9)$ não detecta como discrepante a k -ésima obser- vação	38
5.3.1- Intervalos onde $DF(21)$ como função de $X(3,21)$ não detecta como discrepante a vigésima primei- ra observação	40
5.3.2- Intervalos onde $DC(21)$ como função de $X(3,21)$ não detecta como discrepante a vigésima primei- ra observação	40
5.3.3- Intervalos onde $DF(9)$ como função de $X(3,9)$ não detecta como discrepante a nona observação.	40
5.3.4- Intervalos onde $DC(9)$ como função de $X(3,9)$ não detecta como discrepante a nona observação.	40
5.3.5- Intervalos onde $DF(k)$ como função de $X(3,21)$ não detecta como discrepante a k -ésima obser- vação	42
5.3.6- Intervalos onde $DC(k)$ como função de $X(3,21)$ não detecta como discrepante a k -ésima obser- vação	42
6.1.A.1- Intervalos onde $DR(i)$ como função de $Y(i)$ não detecta como discrepante a k -ésima obser- vação	61

6.1.B.1- Intervalos onde $DR(21)$ como função de $X(3,21)$ não detecta como discrepante a vigésima primei- ra observação	62
6.1.B.2- Intervalos onde $DR(9)$ como função de $X(3,9)$ não detecta como discrepante a nona observação.	62

L I S T A D E Q U A D R O S

Quadro	Página
1.3.1- Representação do j -ésimo vetor coluna da matriz X e do vetor A dos parâmetros	6
1.3.2- Interpretação geométrica do ajuste do vetor Y pelo método de mínimos quadrados	7
1.5.1- Representação de um conjunto de restrições sobre o vetor de parâmetros	9
1.5.2- Representação gráfica do ajuste Y para o modelo com restrições nos parâmetros	9
1.7.1- Dois conjuntos de dados que contém algum tipo de observação discrepante	12
1.7.2- Gráfico de valores de Y contra valores de X para o primeiro conjunto do quadro 1.7.1	13
1.7.3- Gráfico de valores de Y contra valores de X para o segundo conjunto do quadro 1.7.1	13
2.1- Representação da matriz de delineamento para os dois conjuntos do quadro 1.7.1	17
2.2- Cálculo dos elementos da diagonal da matriz chapéu para o primeiro conjunto do quadro 1.7.1	18
3.1- A matriz Z , o vetor dos parâmetros ajustados e o vetor de valores ajustados de Y para o primeiro conjunto do quadro 1.7.1	20
3.2- Valores de $D(i)$ para os pontos dos dois conjuntos do quadro 1.7.1	21
4.1- Valores de $DF(i)$ para os pontos dos dois conjuntos do quadro 1.7.1	29
5.1.1- Conjunto de dados de "Stack-Loss"	32

I N T R O D U Ç Ã O

Ao ajustar um modelo de regressão linear se faz várias suposições tais como: normalidade, variância constante e linearidade. Mas é preciso de alguma forma verificar se as suposições feitas realmente acontecem.

Até uns doze anos atrás, a maneira de conferir as suposições feitas ao ajustar um modelo de regressão linear era através dos gráficos de resíduos contra valores ajustados ou outras quantidades, e os gráficos probabilísticos .

Vinte e cinco anos atrás não se falava em "outliers" (pontos aberrantes) nem "leverage points" (pontos influentes), naturalmente também não havia uma metodologia própria para detectá-los. A recente descoberta, especialmente nos últimos oito anos, de estatísticas chamadas "Diagnósticos" ajudam o analista a decidir se as suposições são corretas ou não.

Hoje dispõe-se de um grande número de diagnósticos, criados pela necessidade de considerar vários aspectos que exigem o uso de técnicas diferentes, por outro lado, diagnósticos que são praticamente idênticos podem ter significados distintos para diferentes analistas.

Ampla literatura dedicada à metodologia de diagnóstico em regressão existe atualmente. Contribuições valiosas são os estudos e comentários de ANSCOMBE and TUKEY (1963), HOCKING (1972), MOSTELLER and TUKEY (1977), COOK (1977 e 1979), HOAGLIN and WELSCH (1978), BESLEY, KUH and WELSCH (1980), DANIEL and WOOD (1980), HUBER (1981), PREGIBON, D. (1981), COOK and WEISBERG (1982), e vários outros.

O propósito principal deste trabalho é mostrar que o diagnóstico DFFITS é eficaz, tanto na detecção de pontos influentes, como na detecção de pontos aberrantes, característica que o D de Cook não possui para a detecção de pontos aberrantes, e que não existe razão válida para preferir o último, diante do primeiro. Apesar do aumento da quantidade de trabalho computacional que exige o cálculo de DFFITS ser mínima, os dois métodos são igualmente usados.

A comprovação do fato foi basicamente prática, pela observa-

ção do comportamento desses dois diagnósticos através dos gráficos gerados pela variação da resposta, ou pela variação de um preditor, ou, pela variação conjunta da resposta e um preditor, na i -ésima observação. Todavia colocou-se algumas hipóteses e estabeleceu-se alguns fatos teóricos, sendo que nem todos são demonstrados aqui.

Considerou-se imprescindível estabelecer uma base teórica para a compreensão do fato de existir diferença entre o desempenho do D de Cook e do DFFITS.

No Capítulo 1 faz-se a descrição de modelo linear e do método por mínimos quadrados usado para seu ajuste, com ou sem restrições nos parâmetros, constroeu-se a estrutura probabilística e descreve-se o método de análise do modelo ajustado.

No Capítulo 2 define-se a diagonal da matriz chapéu, estudam-se as propriedades e ilustra-se o uso dos seus elementos como técnica de diagnóstico da presença de observações influentes num ajuste.

Nos Capítulos 3 e 4, respectivamente, definem-se os diagnósticos D de Cook e DFFITS, e ilustra-se o seu uso.

O estudo dos diagnósticos: Diagonal da matriz chapéu, D de Cook e DFFITS, nessa ordem é proposital. Primeiro, porque permite distinguir claramente a existência de duas classes de pontos discrepantes: aberrantes e influentes. Segundo, porque mostra que a diagonal da matriz chapéu detecta a presença de pontos influentes num ajuste, mas não de pontos especificamente aberrantes. O D de Cook detecta a presença de pontos influentes, mas não destaca a presença de pontos significativamente aberrantes. Entretanto, o DFFITS cumpre a dupla função de detectar ambas as classes de pontos discrepantes.

No Capítulo 5 se faz a análise do desempenho do D de Cook e do DFFITS para os dados de "Stack-Loss" do livro de BROWNLEE, K.A. (1965), pag. 454; primeiro como função de $y(i)$, o i -ésimo valor observado, segundo como função de $x(3,i)$, o valor da terceira variável preditora na i -ésima observação, e depois como função conjunta de $y(i)$ e $x(3,i)$; através dos gráficos de cada uma dessas funções.

No Capítulo 5 constata-se o que se tinha vislumbrado quando se ajustavam os dois conjuntos do quadro 1.7.1: que DFFITS é muito mais eficaz como diagnóstico do que o D de Cook. Mais ainda, em determinados situações o D de Cook deixa de ser confiável, pois para afastamentos grandes de $y(i)$ desde seu "valor exato",

ele não dá o destaque correspondente, e que com alguns valores de $y(i)$ e algum intervalo de valores $x(3,i)$, pontos discrepantes nem são detectados. DFFITS, no entanto, está sempre detectando esses pontos.

No Capítulo 6 é feita a análise do desempenho de um diagnóstico equivalente ao D de Cook, o D Raiz, e é feita a comparação deste com o DFFITS, para estabelecer de maneira equivalente e com maior clareza as diferenças entre o D de Cook e o DFFITS. Neste mesmo Capítulo aparece o programa usado para o cálculo dos valores dos diagnósticos e os detalhes necessários para seu uso.

Com este trabalho também se quer motivar o uso correto de diagnósticos, como instrumento de ajuda para determinar a qualidade do modelo ajustado a um conjunto de dados. Alertando para não se contentar com as informações que alguns deles proporcionam mas, combinar as informações que varios deles fornecem.

C A P I T U L O 1

ELEMENTOS BASICOS PARA A ANALISE DE REGRESSAO LINEAR

Neste capítulo são apresentados conceitos básicos para análise de regressão linear. Brevemente definem-se modelo linear e o método de ajuste de mínimos quadrados, com que vai-se tratar neste trabalho, e descrevem-se as características do modelo a ser ajustado, alguns métodos para determinar sua adequação e outros para determinar as causas que podem estar alterando-o. Há varios textos em que se pode ver o desenvolvimento em detalhe: DACHS (1983), DANIEL and WOOD(1971 e 1980), DRAPER and SMITH(1966 e 1981), MONTGOMERY and PECK(1982), etc.

1.1. MODELO LINEAR.

Um modelo é dito linear quando ele é função linear dos seus parâmetros. Por exemplo:

$$Y = a + bx + e, \quad Y = a + bx^2 + e,$$

são modelos lineares. No entanto,

$$Y = a \cdot \exp(bx) + e, \quad Y = a \cdot \sin(bx) + e,$$

não são modelos lineares.

Em geral, um modelo linear pode ser representado por uma equação da forma:

$$y_i = a_0 + a_1 x_i^{(1)} + \dots + a_p x_i^{(p)} + e_i \quad (1.1.1)$$

y_i são observações, valores das respostas (dados),

$x_i^{(j)}$ são valores dos suportes ou preditores, são valores fixados,

a_j são parâmetros desconhecidos e

e_i são erros (variáveis não observáveis).

Em forma matricial,

$$Y = X A + E, \quad (1.1.2)$$

onde Y é o vetor dos dados, X é a matriz de tamanho $n \times p$ dos $x_i^{(j)}$, que podem ser escritos como $x(i,j)$, e E é o vetor dos resíduos.

1.2. AJUSTE PELO METODO DE MÍNIMOS QUADRADOS.

Dado um conjunto de valores observados:

$$y_1, y_2, \dots, y_n$$

os quais podem ser representados pelo vetor

$$Y' = (y_1, y_2, \dots, y_n),$$

ajustar um modelo a esse conjunto de valores, consiste em obter o modelo parametrizado pelo vetor A que dê a descrição mais próxima dos dados em Y' , de acordo com alguma distância definida. Isto significa encontrar algum vetor A que faça mínima a distância entre os valores observados e os valores ajustados.

A distância a ser usada aqui é a denominada distância euclidiana, definida como a raiz quadrada da soma de quadrados dos desvios ou diferenças entre o valor observado e o valor aproximado.

A obtenção do vetor A que especifica o modelo ajustado pelo método de mínimos quadrados é feita de modo a minimizar a distância euclidiana entre o vetor de observações Y e o vetor ajustado XA .

Equivalente a minimizar a distância euclidiana (norma) é minimizar o seu quadrado. Assim torna-se mais fácil lidar com as expressões algébricas que aparecem.

Interessa minimizar então,

$$\begin{aligned} Q(A) &= (Y - XA)'(Y - XA) \\ &= Y'Y - 2A'X'Y + A'X'XA. \end{aligned} \quad (1.2.1)$$

Derivando $Q(A)$ com relação a cada elemento de A , obtém-se:

$$d/dA (Q(A)) = -2X'Y + 2X'XA, \quad (1.2.2)$$

$$\text{pois } d/dA (2A'X'Y) = -2X'Y \quad \text{e} \quad d/dA (A'X'XA) = 2X'XA.$$

Fazendo $dQ(A)/dA = 0$, tem-se $X'XA = X'Y$, um sistema de equações lineares em A , chamadas comumente de "equações normais", as quais têm a importante propriedade de serem consistentes (terem solução) para todo Y . Com isto, o ajuste para Y pelo método de mínimos quadrados é $\hat{Y} = X\hat{A}$, onde $\hat{A}$ é qualquer solução das equações normais, que para problemas de regressão linear pode ser calculada premultiplicando as equações normais pela inversa de $(X'X)$, porque X é de posto completo. Portanto, $(X'X)^{-1} X'XA = (X'X)^{-1} X'A$ e

$$\hat{A} = (X'X)^{-1} X'Y \quad (1.2.3)$$

1.3. INTERPRETAÇÃO GEOMETRICA DO AJUSTE POR MÍNIMOS QUADRADOS.

Dado que $\hat{Y} = X\hat{A}$, o vetor $\hat{Y}$ pode ser escrito como combinação linear dos vetores coluna da matriz X , da seguinte maneira:

$$\hat{Y} = a_0 X^{(0)} + a_1 X^{(1)} + \dots + a_{p-1} X^{(p-1)} \quad (1.3.1)$$

onde o j -ésimo vetor coluna da matriz X e o vetor de parâmetros são representados como no quadro 1.3.1.

Quadro 1.3.1- Representação do j -ésimo vetor coluna X da matriz X e do vetor A dos parâmetros.

$$X^{(j)} = \begin{bmatrix} x_1^{(j)} \\ x_2^{(j)} \\ \vdots \\ x_n^{(j)} \end{bmatrix} \quad A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix}$$

Os vetores $X^{(j)}$; $j=0,1,\dots,p-1$; geram o espaço $C(X)$, dos vetores de tamanho $n \times 1$, portanto, $\hat{Y}$ é um dos vetores $\tilde{Y}$ de $C(X)$ e o único tal que,

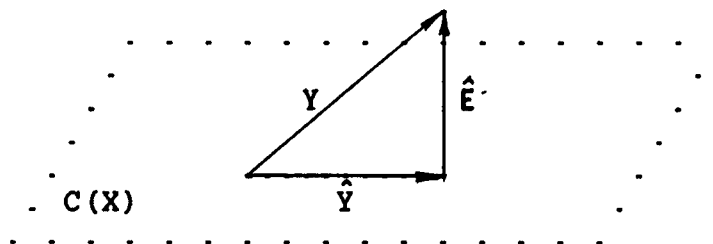
$$\|Y - \hat{Y}\|^2 = \min_{\tilde{Y} \in C(X)} \|Y - \tilde{Y}\|^2. \quad (1.3.2)$$

Ademais os $X^{(j)}$ são linearmente independentes, já que a matriz X é de posto completo e $C(X)$ tem dimensão p .

O vetor $\hat{Y}$ é a projeção ortogonal de Y no espaço $C(X)$, pois $\hat{Y}$ é a melhor aproximação de Y em $C(X)$. Portanto, $\hat{E} = Y - \hat{Y}$ é ortogonal a Y e a todo $\tilde{Y}$ em $C(X)$; pois se $\tilde{Y} \in C(X)$, $\tilde{Y} = XA$ para algum A , e o produto escalar de $\hat{Y}$ e $\hat{E}$ é:

$$\langle \hat{Y}, \hat{E} \rangle = (XA)'(Y - \hat{Y}) = A'X'(Y - XA) = A'(X'Y - X'XA) = 0.$$

Quadro 1.3.2- Interpretação geométrica do ajuste do vetor Y pelo método de mínimos quadrados.



Sabendo que $\hat{Y}$ e $\hat{E}$ são ortogonais, é certo que:

$$\|Y\|^2 = \|\hat{Y}\|^2 + \|\hat{E}\|^2 \quad (1.3.3)$$

ou seja, que a soma de quadrados de valores ajustados mais a soma de quadrados de resíduos é igual à soma de quadrados das respostas. A sua vez, a equação (1.3.3) é a forma mais simples de decomposição da análise de variância. Costuma-se descrever a análise de variância como na tabela 1.3.1.

Tabela 1.3.1- Tabela de análise de variância:

Origem	Graus de liberdade	Soma de quadrados (corrigida)
Modelo	$p-1$	$(\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})$
Resíduos...	$n-p-1$	$(Y - \hat{Y})'(Y - \hat{Y})$
Total.....	$n-1$	$(Y - \bar{Y})'(Y - \bar{Y})$

onde os graus de liberdade são as dimensões dos espaços vetoriais.

1.4. A MATRIZ DE PROJEÇÕES, H.

O vetor $\hat{Y}$ é a projeção ortogonal de Y no espaço $C(X)$ e, pelo método de mínimos quadrados, $\hat{Y} = X\hat{A}$, onde $\hat{A}$ é qualquer solução das equações normais $X'XA = X'Y$, que em problemas de regressão linear é calculada por $\hat{A} = (X'X)^{-1} X'Y$, segundo a equação (1.2.2). Então,

$$\hat{Y} = X(X'X)^{-1} X'Y = HY. \quad (1.4.1)$$

A matriz H, definida por $H = X(X'X)^{-1} X'$, é matriz de projeções ortogonais no espaço $C(X)$, tal que aplicada a um vetor (n x 1) o projeta ortogonalmente em $C(X)$.

Por ser matriz de projeções ortogonais, a matriz H possui as propriedades de simetria:

$$H' = [X(X'X)^{-1} X']' = X(X'X)^{-1} X' = H,$$

e de idempotencia:

$$H = [X(X'X)^{-1} X] [X(X'X)^{-1} X'] = X(X'X)^{-1} X' = H.$$

A matriz H ajusta o vetor Y do modelo $Y = XA + E$ para $\hat{Y}$ e, como $\hat{E} = Y - \hat{Y}$,

$$\hat{E} = Y(I - H). \quad (1.4.2)$$

Isto também mostra que, $\hat{Y}$ e $\hat{E}$ são ortogonais, pois os espaços a que eles pertencem são ortogonais: $\langle H, I-H \rangle = 0$.

1.5. RESTRIÇÕES NOS PARÂMETROS E ACRÉSCIMO NA SOMA DE QUADRADOS DE RESÍDUOS.

O problema de ajuste, aqui, tem sido considerado com completa liberdade para os valores dos parâmetros do modelo. Mas é muito frequente impor restrições a esses valores, por meio de uma hipótese, e determinar se são admitidas ou são contraditas pelos dados.

1.6. ESTRUTURA PROBABILÍSTICA DO ERRO.

Até aqui, o problema de ajuste teve apenas caráter algébrico e geométrico: desenvolveu-se com base na minimização dos erros, sem fazer suposição alguma sobre sua estrutura probabilística.

Julgar se um modelo ajustado é adequado exige determinar se certas quantidades são pequenas (ou grandes), testar hipóteses sobre os parâmetros, calcular intervalos de confiança para os mesmos, etc. Mas, para conseguir isto, é necessário impor uma estrutura probabilística ao vetor de erros, mediante suposições sobre seu comportamento pois, mesmo observações repetidas sob condições supostamente idênticas resultam em valores diferentes, ou seja, dois ou mais pontos com medidas iguais nos correspondentes preditores, produzem resultados distintos.

Daqui por diante ao ajustar $Y = XA + E$ supõe-se que os valores em X são fixos (medidos sem erro, sem serem valores observados de variáveis aleatórias), mas os valores e_i em E são tais que:

1. e_i é uma variável aleatória.
2. $E(e_i) = 0$; $i=1,2,\dots,n$. O valor esperado para cada e_i é zero.
3. $V(e_i) = \sigma^2$ (constante) e $\text{Cov}(e_i, e_j) = 0$ para $i \neq j$ e $1 \leq i, j \leq n$.
Os erros tem variância constante e são não correlacionados.
4. A distribuição de e_i é normal para $i=1,2,\dots,n$.

O modelo $Y = XA + E$ com as suposições 1,2 e 3 é conhecido como o Modelo de Gauss- Markov. Observa-se que 3 e 4 implicam em e 's independentes.

Trabalhando com o modelo de Gauss- Markov tem-se que:

$$E(A) = E[(X'X)^{-1} X'Y] = (X'X)^{-1} X'E(Y) = (X'X)^{-1} X'XA = A, \quad (1.6.1)$$

$$V(A) = V[(X'X)^{-1} X'Y] = [(X'X)^{-1} X'] V(Y) [X'X]^{-1} X'$$

$$= (X'X)^{-1} X' \cdot \sigma^2 I \cdot X(X'X)^{-1} = \sigma^2 (X'X)^{-1}, \quad (1.6.2)$$

$$E(Y) = E(XA + E) = E(XA) + E(E) = XA, \quad (1.6.3)$$

$$V(Y) = V(E) = \sigma^2 I \quad e \quad (1.6.4)$$

$$V(\hat{E}) = V[Y(I-H)] = (I-H)V(Y) = \sigma^2(I-H). \quad (1.6.5)$$

Y tem distribuição normal multivariada com media XA e variância $\sigma^2 I$. E ,

$$s^2 = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{n - p} = \left\{ \sum_{j=1}^n (y_j - \hat{y}_j)^2 \right\} / (n - p) \quad (1.6.6)$$

é o estimador não viciado de mínima variância para σ^2 . Ver demonstração para $p=2$ em CARVALHO E DACHS (1983), pag.22.

1.7. DIAGNÓSTICO.

A verificação de que o uso de estatísticas globais, como R^2 e $\hat{A}$ podem apresentar uma imagem distorcida, falsa, do comportamento de um conjunto de dados, prova a necessidade de recorrer ao uso de outros elementos e/ou técnicas de análise de resíduos para a detecção de problemas de ajuste de um modelo ou de existência de observações discrepantes no conjunto. Uma observação é dita discrepante, quando sua retirada causa alterações importantes no modelo ajustado.

Uma idéia inicial, superficial, da adequação de um ajuste se tem comparando a soma de quadrados de resíduos (SQR) que se espera seja pequena, com a soma de quadrados devida à regressão ou modelo (SQRg); ou alternativamente, comparando a soma de quadrados de regressão com a soma de quadrados total (SQT).

Define-se:

$$R^2 = \frac{SQRg}{SQT} = \frac{\sum_{j=1}^n (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2} \quad (1.7.1)$$

como o Coeficiente de Correlação Múltipla, ao quadrado, entre Y e os X 's. R^2 assume valores entre zero e um.

R^2 é grande (próximo de um) quando o ajuste é adequado, já que a soma de quadrados de resíduos é pequena, e é pequeno (próximo de zero) quando o ajuste é inadequado. Porém, se R^2 é alto convém examinar os resíduos para confirmar que o ajuste é adequado, e se R^2 é pequeno deve-se examinar os resíduos para tentar descobrir porque o ajuste não é adequado e se é possível melhorá-lo.

Em todo problema de ajuste é indispensável a análise de resíduos, por algumas ou varias das técnicas desenvolvidas para este efeito. Entre as mais comuns se têm: gráficos de resíduos contra valores ajustados, ou contra valores dos suportes, ramo e folhas, desenho esquemático e gráfico normal de resíduos. Ampla literatura sobre este assunto se encontra em TUKEY (1977), MOSTELLER and TUKEY (1977), DACHS (1978), e VELLEMAN and HOAGLIN (1981).

As observações discrepantes num conjunto são de duas classes:

1. **Influentes.** Relacionadas com os valores da j -ésima variável preditora na i -ésima observação, $x(i,j)$.

2. **Aberrantes.** Relacionadas com os valores observados $y(i)$.

Para mostrar que as naturezas desses dois tipos de observações são distintos, consideremos dois conjuntos, de cinco pontos cada, descritos no quadro 1.7.1.

Quadro 1.7.1- Dois conjuntos de dados que contém algum tipo de observação discrepante.

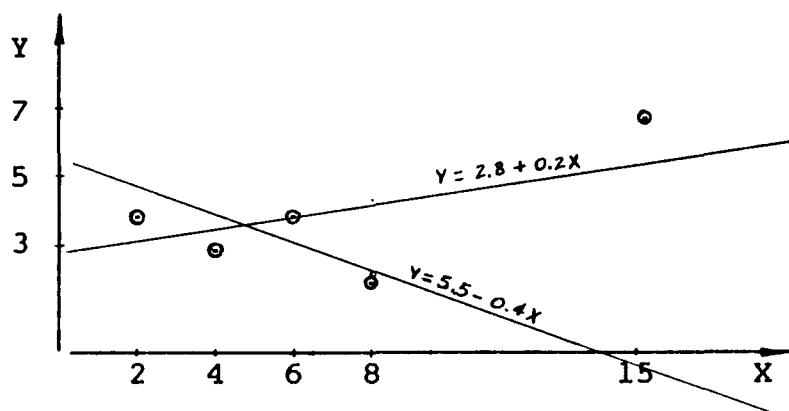
Primeiro conjunto				segundo conjunto			
-----				-----			
i		x		y	i		x y
-----				-----			
1		8		2	1		2 3
2		4		3	2		2 4
3		6		4	3		7 8
4		2		5	4		12 5
5		15		7	5		12 6
-----				-----			

Para ambos os conjuntos:

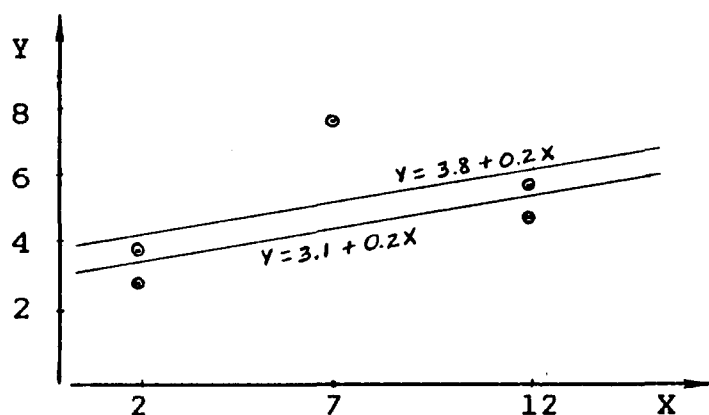
$$(X'X) = \begin{bmatrix} -1 & 0.69 & -0.07 \\ -0.07 & 0.01 \end{bmatrix}$$

Então os modelos ajustados são: $Y = 2.8 + 0.2X$ para o primeiro conjunto e $Y = 3.8 + 0.2X$ para o segundo conjunto, os resíduos para o primeiro conjunto são: -2.4, -0.6, 0, 1.8 e 1.2, para o segundo conjunto são: -1.2, -0.2, 2.8, -1.2 e -0.21. $SQT = 14.8$, $SQRq = 4.0$ e $R^2 = 0.27$, ademais $F(3,2) = 1.11$, que não é significativo nem a 10%, para os dois conjuntos. Esse valor de R^2 de imediato indica que devem ser checados os gráficos de Y contra X ; em outros casos, os de resíduos contra valores ajustados de Y , ou contra valores dos preditores.

Quadro 1.7.2- Gráfico de valores de Y contra valores de X para o primeiro conjunto do quadro 1.7.1.



Quadro 1.7.3- Gráfico de valores de Y contra valores de X para o segundo conjunto do quadro 1.7.1.



Com a retirada da quinta observação do primeiro conjunto, o novo ajuste fica $Y = 5.5 - 0.4X$, os resíduos são: -0.3, -0.9, 0.9, 0.3 e $R^2 = 0.64$. A retirada da terceira observação do segundo conjunto muda o ajuste para $Y = 3.1 + 0.2X$, os novos resíduos são: -0.5, 0.5, -0.5, 0.5 e $R^2 = 0.80$.

Nota-se que com a simples análise de resíduos não se detectam as alterações importantes que causa a retirada do quinto ponto no primeiro conjunto. Entretanto no segundo conjunto se podia desconfiar de que a retirada do terceiro ponto alterasse o modelo ajustado por ele possuir o maior resíduo.

A retirada de qualquer outro ponto em cada conjunto produz mudanças muito pequenas, o que também se pode verificar mediante o cálculo da norma quadrática do vetor da diferença entre $\hat{A}$ antes e $\hat{A}(\cdot)$ depois da retirada da observação. Note-se que com a retirada do ponto $x = 4$ e $y = 3$, no primeiro conjunto, e $x = 2$ e $y = 4$, no segundo, os respectivos ajustes mudam para: $Y = 3.2 + 0.14X$ e $Y = 4.0 + 0.17X$.

No primeiro conjunto a retirada do ponto:

$x = 15$ e $y = 7$ tem um efeito de $\|\hat{A} - \hat{A}(5)\|_2^2 = 7.65$, e a de $x = 4$ e $y = 3$ tem um efeito de $\|\hat{A} - \hat{A}(2)\|_2^2 = 0.16$.

No segundo conjunto a retirada do ponto:

$x = 7$ e $y = 8$ tem um efeito de $\|\hat{A} - \hat{A}(3)\|_2^2 = 0.49$, e a de $x = 2$ e $y = 4$ tem um efeito de $\|\hat{A} - \hat{A}(2)\|_2^2 = 0.04$.

Os gráficos de Y contra X mostram que o ponto $x = 15$ e $y = 7$, que causa a maior alteração no ajuste para o primeiro conjunto pela sua retirada, é muito influente e que, no entanto, o ponto $x = 7$ e $y = 8$, que causa a maior alteração no ajuste para o segundo conjunto pela sua retirada, é aberrante.

C A P I T U L O 2

A DIAGONAL DA MATRIZ DE PROJEÇÕES, H .

Pontos influentes num ajuste podem ser detectados de forma simples e segura, considerando apenas os valores dos suportes ou preditores, através dos elementos da diagonal da matriz de projeções, mais comumente conhecida com o nome de "matriz chapéu", porque $\hat{Y} = HY$, pois como se sabe, a aparição desses pontos se deve a mau comportamento de valores dos preditores.

Em ajustes com uma variável preditora, duas contando a constante, gráficos de Y contra X revelam qualquer ponto influente, como no caso do gráfico para o primeiro conjunto do quadro 1.7.1. Verifica-se que esses pontos têm associado um valor $h(i,i) = h_i$, o i-ésimo elemento da diagonal da matriz chapéu, grande. Mas quando o número de variáveis preditoras p é maior que dois, incluída a constante, tais gráficos podem não mostrar pontos que são influentes no espaço de dimensão p e não aparecem como discrepantes no espaço de duas dimensões. Nesses casos, a diagonal da matriz chapéu representa um valioso elemento de diagnóstico. $h(i,i)$ também será denotado h_{ii} ou $h(i)$, exceto nos casos onde possa haver confusão.

É importante destacar aqui algumas propriedades dos elementos da diagonal da matriz de projeções para compreender sua utilidade como elemento de diagnóstico. Estabeleceu-se atrás que a matriz, H, de projeções é idempotente, de posto p e simétrica. Logo, se cumpre:

(1) A soma dos elementos da diagonal da matriz chapéu é igual a p, o posto da matriz X:

$$\sum_{i=1}^n h(i) = p. \quad (2.1)$$

(2) A soma dos elementos da i-ésima linha da matriz chapéu é igual a um:

$$\sum_{j=1}^n h(i,j) = 1. \quad (2.2)$$

(3) O i -ésimo elemento da diagonal da matriz chapéu é igual à soma dos quadrados dos elementos da i -ésima linha da mesma matriz:

$$h(i) = \sum_{j=1}^n h^2(i,j). \quad (2.3)$$

Ademais, como $\hat{Y} = X\hat{A} = HY$, onde $\hat{A} = (X'X)^{-1} X'Y$,

$$\hat{y} = h_{i1} y_1 + h_{i2} y_2 + \dots + h_{ii} y_i + \dots + h_{in} y_n. \quad (2.4)$$

Portanto $h(i,j)$, ou h_{ij} , representa o peso que y_j , denotado também por $y(j)$, tem na determinação de $y(i)$. Em particular $h(i)$ é o peso que $y(i)$ tem no seu próprio ajuste.

A projeção sobre o subespaço $C(X)$ pode ser expressada como a soma de duas ou mais projeções, segundo COOK and WEISBERG (1980). Então,

$$H = \mathbb{1} \mathbb{1}'/n + W(W'W)^{-1} W' \quad \text{e} \quad h_{ii} = 1/n + X_{(i)} (W'W)^{-1} X_{(i)}',$$

onde W é a matriz de tamanho $n \times p$ de variáveis preditoras centradas, $X_{(i)}$ é a i -ésima linha de W e $\mathbb{1}$ é o vetor de n uns correspondente à constante do modelo.

Tem-se que $h(i) \geq 1/n$. Para o modelo que não contém a constante o limite inferior de $h(i)$ é zero. O limite superior para $h(i)$ é calculado levando em conta que se a i -ésima e a j -ésima linhas de X são iguais então,

$$h_{ii} = h_{ij} \quad \text{e} \quad h_{ii} = \sum_{j=1}^n h_{ij}^2 \geq c h_{ii}^2, \text{ logo } h_{ii} \leq 1/c,$$

onde c é o número de vezes que a i -ésima linha de X aparece repetida. h_{ii} atinge seu valor máximo igual a um quando a i -ésima linha não aparece repetida, e assim $h(i,j) = 0$ para $i < j$.

Se $h(i) = 0$, $\hat{y}(i) = 0$ pela propriedade (3), $\hat{y}(i)$ não é afetado por $y(i)$ ou por qualquer outro $y(j)$. Um ponto com $x = 0$, quando o modelo é uma reta que passa pela origem, é um exemplo. Se $h(i) = 1$, $\hat{y}(i) = y(i)$: o modelo ajusta exatamente o i -ésimo ponto; aqui, o modelo dedica uma variável preditora para o i -ésimo ponto em particular, como quando se adiciona uma variável à

matriz dos preditores para testar a falta de ajuste num ponto.

O conhecimento desses limites ajuda a interpretar $h(i)$, mas não dá uma caracterização da sua grandeza. HUBER (1981), tendo em conta a propriedade (2), define a quantidade $1/h(i)$ como o "número equivalente de observações" e com base nela, propõe dois critérios para a identificação de pontos influentes:

(a) prestar atenção aos pontos com número equivalente de observações menor ou igual que cinco, ou seja com $h(i) > 1/5$.

(b) prestar muita atenção aos pontos com $h(i) > 1/2$.

HOAGLIN and WELSCH (1978) levando em conta a propriedade (1) e observando que o valor médio dos elementos da diagonal da matriz chapéu é p/n , sugerem caracterizar como pontos influentes aqueles cujo $h(i)$ seja maior que $2p/n$.

Usando os dados do quadro 1.7.1 exemplifica-se a notação usual, no quadro 2.1.

Quadro 2.1- Representação da matriz X de delineamento para o primeiro e para o segundo conjunto do quadro 1.7.1.

Para o primeiro conjunto:

Para o segundo conjunto:

$$X = \begin{bmatrix} 1 & 8 \\ 1 & 4 \\ 1 & 6 \\ 1 & 2 \\ 1 & 15 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 7 \\ 1 & 12 \\ 1 & 12 \end{bmatrix}$$

O i-ésimo vetor linha da matriz X será denotado $X_{(i)}$.

$$X_{(5)} = (1 \ 15) \quad \text{e} \quad X_{(5)} = (1 \ 12) \quad (2.5)$$

são, respectivamente, o quinto vetor linha da matriz X do primeiro e do segundo conjunto.

O i-ésimo elemento da diagonal da matriz chapéu é calculado pela expressão:

$$h(i) = h_i = X_{(i)} (X'X)^{-1} X_{(i)}' \quad (2.6)$$

Por exemplo, para o primeiro conjunto o cálculo do quinto elemento da diagonal da matriz chapéu aparece no quadro 2.2.

Quadro 2.2- Cálculo do quinto elemento da diagonal da matriz chapéu para o primeiro conjunto do quadro 1.7.1.

$$h_5 = [1 \quad 15] \begin{bmatrix} - & - & - & - \\ | & 0.69 & -0.07 & | & | & 1 & | \\ | & -0.07 & 0.01 & | & | & 15 & | \\ - & - & - & - & - & - & - \end{bmatrix} = 0.84$$

Os outros elementos $h(i)$ para esse conjunto são:

$$h_1 = h_3 = 0.21, \quad h_2 = 0.29 \quad \text{e} \quad h_4 = 0.45.$$

E para o segundo conjunto são:

$$h_1 = h_2 = h_4 = h_5 = 0.45 \quad \text{e} \quad h_3 = 0.20.$$

Cabe advertir contra o uso indiscriminado dos critérios enunciados para a caracterização de pontos influentes com base nos valores dos elementos da matriz chapéu. Apenas se garante sua eficácia trabalhando com conjuntos com mais de dez casos. Já, em situações como as dos conjuntos de cinco pontos do quadro 1.7.1 esses critérios são inválidos, porque de entrada se estaria violando o segundo deles, pois $h(i) > 1/5$, para todo i , no primeiro conjunto

Fica verificado que, através dos valores dos elementos da diagonal da matriz chapéu são claramente detectáveis pontos influentes, mas não o são os pontos cuja natureza é aberrante. No primeiro dos conjuntos de cinco pontos, o ponto $x = 15$ e $y = 7$, influente, que causa a grande alteração no ajuste, pode ser detectado pelo valor $h(5) = 0.84$ (grande), enquanto que o ponto $x = 7$ e $y = 8$, que causa a maior alteração no ajuste para o segundo conjunto, não é destacado, pois $h(3) = 0.20$.

Medidas que detectem essas duas situações simultaneamente se fazem necessárias para dar maior segurança no trabalho de diagnóstico. Nos dois capítulos seguintes tratar-se-á de duas medidas propostas como solução ao mencionado inconveniente, são elas: o D de Cook e o DFFITS.

C A P I T U L O 3

O D DE COOK

Para determinar o grau de importância, ou peso, que um ponto tem no ajuste de um modelo para um conjunto de dados, é razoável medir a alteração causada no ajuste pela sua retirada do conjunto de dados.

Uma medida dessa alteração é fornecida pela norma quadrática da diferença entre os vetores ajustados, $\hat{A}$, antes e, $\hat{A}(i)$, depois da retirada do i -ésimo ponto; mas, esta medida tem o inconveniente de depender das unidades em que são expressadas as variáveis preditoras. Evita-se essa inconveniência usando a norma quadrática ponderada:

$$[\hat{A} - \hat{A}(i)]'X'X[\hat{A} - \hat{A}(i)], \quad (3.1)$$

de acordo com as unidades usadas para medir as variáveis preditoras.

COOK (1977 e 1979) propõe como medida dessa alteração:

$$DC(i) = D_i^2 = [\hat{A} - \hat{A}(i)]'X'X[\hat{A} - \hat{A}(i)] / p s^2, \quad (3.2)$$

que tem a forma de uma norma quadrática ponderada e padronizada, a qual é chamada de D de Cook.

O D de Cook expressado como em (3.2) tampouco descreve de maneira clara o que está medindo. Uma expressão bem mais simples se consegue calculando a diferença $\hat{A} - \hat{A}(i)$; mas, para isto, é preciso calcular antes $\hat{A}(i)$.

$$\hat{A}(i) = [X'X - X'_{(i)}X_{(i)}]^{-1} [X'Y - X'_{(i)}y_{(i)}] \quad (3.3)$$

exige calcular a inversa de uma matriz de ordem p para a retirada de cada observação, fazendo inaplicável o método, pelo grande au-

mento de trabalho e tempo computacional.

Esse obstáculo pode ser afastado usando a matriz do modelo que descreve a falta de ajuste no i -ésimo ponto, ou seja, quando $\hat{y}_i = x_{(i)} \hat{A} + e$, e $\hat{y}_j = x_{(j)} \hat{A}$ para $j \neq i$. Essa matriz, Z , tem $p + 1$ colunas, das quais as p primeiras formam a matriz X e a coluna $p+1$ é toda de zeros, menos na i -ésima posição onde há um valor igual a um, e satisfaz:

$\hat{y}_j = x_{(j)} \hat{A}$ se $j \neq i$, e, $\hat{y}_i = y_i$ se $j = i$. Aqui, $Y = ZA + E$ e então, $\hat{A} = (Z'Z)^{-1}Z'Y$.

Quadro 3.1- A matriz Z , o vetor dos parâmetros ajustados para o modelo aumentado e o vetor de valores ajustados de Y para o primeiro conjunto do quadro 1.7.1.

$$Z = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 4 & 0 \\ 1 & 6 & 0 \\ 1 & 8 & 0 \\ 1 & 15 & 1 \end{bmatrix} \quad \hat{A} = \begin{bmatrix} 5.5 \\ -0.4 \\ 7.5 \end{bmatrix} \quad \hat{Y} = \begin{bmatrix} 2.3 \\ 3.9 \\ 3.1 \\ 4.7 \\ 7.0 \end{bmatrix}$$

Constata-se que $\hat{y}_j = x_{(j)} \hat{A}$ para $j \neq 5$ e $\hat{y}_5 = 7.0 = y_5$.

Usando a matriz Z encontra-se:

$$\hat{A} - \hat{A}(i) = (X'X)^{-1} X' e_{(i)} \quad (3.4)$$

e substituindo na equação (3.1),

$$DC(i) = \frac{1}{p} \left(\frac{h_{ii}}{1-h_{ii}} \right) \left[\frac{\hat{e}_i}{s(1-h_{ii})} \right]^2 \quad (3.5)$$

Duas situações podem fazer com que o D de Cook, denotado também por $D(i)$ e $DC(i)$, assumam valores grandes: que o i -ésimo ponto seja muito influente e então, $h_{ii} / 1-h_{ii}$ terá um valor grande, ou que o i -ésimo ponto seja significativamente aberrante e então, $\hat{e}_i / s(1-h_{ii}) = t(i)$, o chamado i -ésimo resíduo estudatizado internamente, terá valor grande.

Falar de valores grandes de $DC(i)$ deve ser feito com referência a algum valor. É preciso portanto, encontrar esse valor de comparação: limite a partir do qual se possa dizer se um ponto é, ou não, discrepante. $DC(i)$ na expressão (3.1) tem a mesma forma da expressão que fornece o elipsoide de confiança para os p parâ-

metros do modelo pois, nesse caso, $F = I$ e então,

$$\frac{[F\hat{A}-\hat{A}(i)]'[F(X'X)^{-1}F'] [F\hat{A}-\hat{A}(i)]/p}{Y'(I-H)Y/(n-p)} = \frac{[\hat{A}-\hat{A}(i)]'(X'X)^{-1}[\hat{A}-\hat{A}(i)]/p}{Y'(I-H)Y/(n-p)}$$

$$< F(p, n-p, 1-\alpha) \quad (3.6)$$

onde $Y'(I-H)Y/(n-p) = s^2$ e $F(p, n-p, 1-\alpha)$ é o $(1-\alpha)$ -ésimo quantil da distribuição F com p e $n-p$ graus de liberdade.

O mesmo COOK (1977 e 1979) propõe $F(p, n-p, 10\%)$ como valor de comparação para DC, assim os valores de $DC(i)$ são transformados para uma escala mais familiar que facilita a análise posto que, $F(p, n-p, 10\%) = 1 / F(n-p, p, 90\%)$. Com isso, um valor de $DC(i)$ maior que $F(p, n-p, 10\%)$ corresponde a um ponto discrepante.

Quadro 3.2- Valores de $DC(i)$ para os correspondentes pontos dos conjuntos do quadro 1.7.1.

No primeiro conjunto	No segundo conjunto
$DC(1) = 0.37$	$DC(1) = 0.31$
$DC(2) = 0.04$	$DC(2) = 0.01$
$DC(3) = 0.00$	$DC(3) = 0.48$
$DC(4) = 0.70$	$DC(4) = 0.31$
$DC(5) = 1.11$	$DC(5) = 0.01$

Com esses valores de $DC(i)$, apenas se destaca o quinto ponto no primeiro conjunto, como era de se esperar, $DC(5) = 1.11$ é um valor grande, enquanto que, no segundo conjunto todos os $DC(i)$ são pequenos, mesmo o $DC(3) = 0.48$ é pequeno; porque $\hat{\epsilon}(3)$ sendo grande contribui a aumentar s^2 e então, a importância desse valor aberrante fica encoberta.

Neste exemplo, do mesmo modo que se trata com os valores da diagonal da matriz chapéu, não é correto usar o valor de comparação para identificar pontos discrepantes, por ser demasiado pequeno o número de pontos nos conjuntos, tanto que, no primeiro conjunto três dos cinco valores de $DC(i)$ são maiores que o valor de comparação tomado por regra e, no segundo conjunto, todos os $DC(i)$ o são.

APENDICE DO CAPÍTULO 3

Demonstrar-se-á aqui, usando a matriz Z que descreve a falta de ajuste no i -ésimo ponto, que a diferença entre o vetor de parâmetros $\hat{A}$ antes, e o vetor de parâmetros $\hat{A}(i)$ depois da retirada do i -ésimo ponto do conjunto de dados é

$$\hat{A} - \hat{A}(i) = [(X'X)^{-1} X' \hat{e}_{(i)}] / (1-h_i) .$$

Demonstração:

$$Z = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ X \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} X'X & X' \\ \hline X & 1 \end{bmatrix}_{(i)}$$

Tomando $A_{11} = X'X$

$A_{12} = X'$

$A_{21} = X$

$A_{22} = 1$

e fazendo inversão por blocos se obtém

$$(Z'Z)^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{bmatrix} \quad \text{onde}$$

$$B_{11}^{-1} = A_{11}^{-1} + (A_{11} A_{12})^{-1} \text{ @ } (A_{21} A_{22})^{-1}$$

$$B_{12}^{-1} = -(A_{11} A_{12})^{-1} \text{ @ } A_{22}^{-1}$$

$$B_{21} = - \begin{matrix} -1 & -1 \\ @ & (A \ A) \\ 21 & 21 \ 11 \end{matrix}$$

$$B_{22} = \begin{matrix} -1 \\ @ \\ 22 \end{matrix}$$

$$e \quad @ = A_{22} - A_{21} (A_{11} \ A_{12})^{-1} \quad \text{Então, } @ = (1 - h_i),$$

$$B_{11} = \{ (X'X)^{-1} + [(X'X)^{-1} X' X (X'X)^{-1}] \} / (1 - h_i)$$

$$B_{11} = X'X - X' X_{(i)} = (X' X_{(i)})^{-1} \quad (\text{RAO, 1965, pág.29})$$

$$B_{12} = - (X'X)^{-1} X'_{(i)} / (1 - h_i)$$

$$B_{21} = - X_{(i)} (X'X)^{-1} / (1 - h_i) \quad e \quad B_{22} = 1 / (1 - h_i)$$

$$(Z'Z)^{-1} = \begin{array}{c|c} \begin{array}{c} - \\ | \\ [X'X - X' X_{(i)}]^{-1} \\ | \end{array} & \begin{array}{c} - \\ | \\ -(X'X)^{-1} X'_{(i)} / (1 - h_i) \\ | \end{array} \\ \hline \begin{array}{c} | \\ - X_{(i)} (X'X)^{-1} / (1 - h_i) \\ | \end{array} & \begin{array}{c} | \\ 1 / (1 - h_i) \\ | \end{array} \\ \hline \end{array}$$

$$\text{Então, } I = (Z'Z)^{-1} (Z'Z),$$

$$I = \begin{array}{c|c|c|c} \begin{array}{c} - \\ | \\ [X'X - X' X_{(i)}]^{-1} \\ | \end{array} & \begin{array}{c} - \\ | \\ -(X'X)^{-1} X'_{(i)} / (1 - h_i) \\ | \end{array} & \begin{array}{c} - \\ | \\ X'X \\ | \end{array} & \begin{array}{c} - \\ | \\ X'_{(i)} \\ | \end{array} \\ \hline \begin{array}{c} | \\ - X_{(i)} (X'X)^{-1} / (1 - h_i) \\ | \end{array} & \begin{array}{c} | \\ 1 / (1 - h_i) \\ | \end{array} & \begin{array}{c} | \\ X_{(i)} \\ | \end{array} & \begin{array}{c} | \\ 1 \\ | \end{array} \\ \hline \end{array}$$

logo, o elemento que está na primeira "linha" e primeira "coluna" desta matriz identidade é a matriz identidade de tamanho $p \times p$:

$$I = [X'X - X'_{(i)} X_{(i)}]^{-1} (X'X)^{-1} - (X'X)^{-1} X'_{(i)} X_{(i)} / (1-h_i)$$

que multiplicada à direita por $(X'X)^{-1}$ fica

$$(X'X)^{-1} = [X'X - X'_{(i)} X_{(i)}]^{-1} - (X'X)^{-1} X'_{(i)} X_{(i)} (X'X)^{-1} / (1-h_i).$$

Usando esta última expressão, e sabendo que

$$\hat{A}(i) = [X'X - X'_{(i)} X_{(i)}]^{-1} [X'Y - X'_{(i)} y_i],$$

$$\hat{A}(i) = [(X'X)^{-1} + (X'X)^{-1} X'_{(i)} X_{(i)} (X'X)^{-1} / (1-h_i)] [X'Y - X'_{(i)} y_i]$$

$$= (X'X)^{-1} X'Y + (X'X)^{-1} X'_{(i)} X_{(i)} (X'X)^{-1} X'Y / (1-h_i) -$$

$$(X'X)^{-1} X'_{(i)} y_i - (X'X)^{-1} X'_{(i)} X_{(i)} (X'X)^{-1} X'_{(i)} y_i / (1-h_i)$$

Ademais,

$$\hat{A} = (X'X)^{-1} X'Y, \quad h_i = X_{(i)}' (X'X)^{-1} X_{(i)} \quad \text{e} \quad y_i = X_{(i)}' (X'X)^{-1} X'Y$$

logo,

$$\hat{A}(i) = \hat{A} + [(X'X)^{-1} X'_{(i)} y_i - (X'X)^{-1} X'_{(i)} y_i (1-h_i)] / (1-h_i)$$

$$- [(X'X)^{-1} X'_{(i)} h_i y_i] / (1-h_i).$$

$$\text{Ent\~ao, } \hat{A} - \hat{A}(i) = (X'X)^{-1} X'_{(i)} (y_i - \hat{y}_i) / (1-h_i)$$

$$= (X'X)^{-1} X'_{(i)} e_i / (1-h_i) .$$

C A P I T U L O 4

O DFFITS

BESLEY, KUH e WELSCH (1980) propõem medir o peso que uma observação num conjunto tem, na determinação do correspondente modelo ajustado, usando uma medida que resume as alterações nos coeficientes e ajuda a compreender os efeitos da retirada da observação.

Com base na alteração causada sobre o valor ajustado:

$$\begin{aligned}
 \text{DFFIT}_i &= \hat{y}_i - \hat{y}_i(-i) = \sum_{(i)} X_i [A - A(i)] \\
 &= \sum_{(i)} X_i [(X'X)^{-1} X_i'] \hat{e}_i / (1-h_i) \\
 \text{DFFIT}_i &= \frac{h_i \hat{e}_i}{1-h_i}, \quad (4.1)
 \end{aligned}$$

onde $\hat{y}_i(-i)$ é o valor ajustado para o conjunto sem a i -ésima observação, B.K.W. definem a medida com o nome de DFFITS, como:

$$\text{DFFITS}_i = \frac{1/2 \frac{h_i \hat{e}_i}{(1-h_i) s(i)}}{1}, \quad (4.2)$$

onde $\sigma\sqrt{h_i}$, o desvio padrão de $\hat{y}_i$ é estimado por $s(i)\sqrt{h_i}$, e

$$s^2(i) = \left\{ \sum_{k \neq i} [y_k - X_k A(i)]^2 \right\} / (n-p-1) \quad (4.3)$$

é a média quadrática de resíduos para o ajuste sem o i -ésimo ponto. Assim como o D de Cook, o DFFITS conta com a vantagem de não depender das unidades em que são medidas as variáveis preditoras.

Já, a alteração experimentada pelo j -ésimo valor ajustado quando a i -ésima observação é retirada, com $j > i$, é expressada por:

$$\begin{aligned} \text{DFFITS}_j(-i) &= \hat{y}_j - \hat{y}_j(-i) \\ &= \frac{X_{(j)} [\hat{A} - \hat{A}(i)]}{s(i) h_j^{1/2}} \\ \text{DFFITS}_j(-i) &= \frac{h_{ij} e_i}{(1-h_i) h_j^{1/2} s(i)} \end{aligned} \quad (4.4)$$

onde $\hat{y}_j(-i)$ é o j -ésimo valor ajustado para o conjunto sem o i -ésimo ponto.

Essa alteração é sempre menor que a alteração provocada sobre o valor ajustado correspondente à observação retirada. Formalmente,

$$\begin{aligned} \sup_{\lambda} \frac{|\lambda' [\hat{A} - \hat{A}(i)]|}{s(i) [\lambda' (X'X)^{-1} \lambda]^{1/2}} &= \frac{\{[\hat{A} - \hat{A}(i)]' X' X [\hat{A} - \hat{A}(i)]\}^{1/2}}{s(i)} \\ &= \frac{h_i^{1/2} |e_i|}{(1-h_i) s(i)} = |\text{DFFITS}_i| \end{aligned}$$

Isto é, $|\text{DFFITS}_j(i)| < |\text{DFFITS}_i|$ para todo $j > i$.

Logo, só é necessário prestar atenção às alterações causadas pela retirada da i -ésima observação, nos restantes valores ajustados, quando o DFFITS é grande.

Assim como definido pela expressão (4.1), o DFFITS apresenta a dificuldade de ter que se calcular de cada vez a correspondente média quadrática de resíduos para a retirada de cada ponto. Porém, de forma versátil, pode-se calcular esta média a partir do ajuste para o conjunto original de dados (conjunto completo), pela fórmula que se obtém usando o modelo aumentado.

No modelo aumentado $\hat{Y} = X \hat{A}(-i)$, exceto para a i -ésima observação onde $\hat{y}_i = X \hat{A}(-i) + \hat{\theta}$. Ou $Y = Z(A|\theta) + D$, e então: $\hat{y}_i = y_i$ e $\hat{\theta} = y_i - \hat{y}_i$, sendo $\hat{y}_i$ o i -ésimo valor ajustado pelo modelo $Y = XA + E$.

Pelo modelo aumentado, o valor ajustado de y_i , $\hat{y}_i$, é o próprio y_i e

$$\hat{\theta} = -X_{(i)}' (X'X)^{-1} X'Y / (1-h_i) + y_i / (1-h_i)$$

$$\hat{\theta} = (-X_{(i)}' \hat{A} + y_i) / (1-h_i) = (y_i - \hat{y}_i) / (1-h_i)$$

$$\hat{\theta} = e_i / (1-h_i)$$

Logo, o acréscimo na soma de quadrados pela restrição $\theta = 0$ é:

$$ASQ(\theta = 0) = \left[\frac{e_i}{(1-h_i)} \right] \left[\frac{1}{(1-h_i)} \right]^{-1} \left[\frac{e_i}{(1-h_i)} \right]$$

$$ASQ(\theta = 0) = \frac{e_i}{(1-h_i)},$$

e levando-se em conta que a soma de quadrados de resíduos com a restrição é igual à soma de quadrados de resíduos sem a restrição mais o acréscimo na soma de quadrados de resíduos devida à restrição,

$$(n-p)s^2 = (n-p-1)s^2(i) + \left[\frac{e_i}{(1-h_i)} \right], \text{ e então,}$$

$$s^2(i) = \frac{n-p}{n-p-1} s^2 - \frac{\hat{e}_i}{(n-p-1)(1-h_i)} \quad (4.4)$$

Quadro 4.1- Valores de DFFITS, denotado também como DF, para os dois conjuntos do quadro 1.7.1 são:

No primeiro conjunto	No segundo conjunto
DF(1) = -1.05	DF(1) = -0.72
DF(2) = -0.20	DF(2) = -0.11
DF(3) = 0.00	DF(3) = 2.21
DF(4) = 1.40	DF(4) = -0.72
DF(5) = 7.25	DF(5) = -0.11

No primeiro conjunto, a quinta observação é destacada pelo seu grande valor de DFFITS, 7.25, e no segundo conjunto a terceira observação é destacada, também, pelo seu valor de DFFITS, 2.21.

Nota-se que diferente do D de Cook, o DFFITS tem capacidade de detectar esses pontos aberrantes, que por terem resíduo grande, aumentam muito a soma de quadrados dos mesmos e diminuem bastante a magnitude do D de Cook. Por exemplo, o que ocorre com a terceira observação do segundo conjunto do quadro 1.7.1, que faz aumentar a soma de quadrados de resíduos, de 1.00 (ela excluída no ajuste) para 10.80 (ela incluída no ajuste), impedindo que seja detectada como discrepante através do correspondente valor do D de Cook.

Agora, é necessário definir um critério para caracterizar pontos discrepantes num conjunto de dados com o uso dos valores de DFFITS. Como

$$DFFITS_i = \frac{\hat{e}_i^{1/2}}{(1-h_i)^{1/2}} * \frac{\hat{e}_i^{1/2}}{(1-h_i)^{1/2} s(i)} \quad (4.5)$$

onde $[\hat{e}_i / s(i) (1-h_i)^{1/2}] = t'_i$

é o chamado i -ésimo resíduo studentizado externamente, o qual tem distribuição t -student com $(n-p-1)$ graus de liberdade, posto que é tem distribuição normal com média zero e variância $(1-h(i))$, e como num delineamento perfeitamente balanceado os $h(i)$ valem p/n ,

$$DFFITS_i = 2 \left[\frac{p/n}{1 - p/n} \right]^{1/2} = 2 \left[\frac{p}{n-p} \right]^{1/2} \quad (4.6)$$

onde 2 corresponde a aproximadamente 5% de significância para 20 graus de liberdade ou mais. Ou seja,

$$|DFFITS_i| > 2 \left[\frac{p}{n-p} \right]^{1/2}$$

indica que a i -ésima observação é discrepante e, então, se deve prestar muita atenção a ela porque pode estar pesando bastante na determinação do ajuste.

C A P I T U L O 5

COMPARAÇÃO DE D DE COOK E DFFITS

Neste capítulo analisa-se comparativamente o desempenho dos diagnósticos D de Cook e DFFITS e apresentam-se os resultados observados. A comparação do desempenho desses diagnósticos é feita principalmente sobre gráficos, os quais permitem formular algumas hipóteses a respeito, nem sempre acompanhadas de suas provas. Algumas provas são apresentadas aqui.

Determinando sobre os gráficos a região onde o diagnóstico detecta, ou a região onde não detecta, como discrepante a correspondente observação e comparando-as, encontra-se que elas são geralmente diferentes e, especificamente, que as regiões onde o D de Cook não detecta são maiores, ou pelo menos iguais porém, não menores, que as regiões onde o DFFITS não detecta. Com base nisso, conclue-se que DFFITS parece ser mais eficaz e confiável como diagnóstico, que o D de Cook.

Na primeira parte do capítulo descreve-se o conjunto original de vinte e uma observações e três preditores ou variáveis explicativas, utilizado como base para o estudo. Na segunda parte faz-se a análise dos valores dos diagnósticos e dos seus gráficos bidimensionais, obtidos pela variação dos valores da resposta Y na nona e na vigésima primeira observações, individualmente. Na terceira parte observa-se o desempenho dos diagnósticos através dos gráficos bidimensionais, obtidos pela variação do valor do preditor X(3) na nona e na vigésima primeira observação, individualmente. E, na quarta parte descreve-se o desempenho dos diagnósticos segundo os gráficos tridimensionais obtidos pela variação do valor da variável resposta e da variável explicativa, X(3), na nona e na vigésima primeira observação, individualmente.

A construção dos gráficos foi feita localizando os valores das variáveis e dos respectivos valores dos diagnósticos no espaço bidimensional ou tridimensional, segundo o caso, e traçando as curvas com a ajuda de elementos próprios de desenho técnico com o fim de conseguir uma melhor visualização. No caso das curvas do D de Cook como função de $x(3,i)$, algumas delas aparecem interrompidas, por não dispor de uma quantidade maior de valores calculados para dar-lhes uma aproximação completa porém, elas dão a ideia necessária.

Os valores dos diagnósticos foram calculados usando um programa em MBASIC num microcomputador I-7000. O programa foi elaborado especificamente para o propósito deste trabalho e é apresentado em detalhe no capítulo 6.

5.1. CONJUNTO DE DADOS DE "STACK-LOSS".

Para o desenvolvimento da parte computacional, de cálculo de valores dos diagnósticos necessários para a construção dos gráficos que serão analisados adiante, foi usado o conjunto de dados de "Stack-Loss" do livro de K.A.Brownlee's. "Statistical Theory and Methodology in science and Engineering". Willey, New York, 2a.edição, 1965, pag.454. Estes dados foram obtidos de 21 dias de operação de uma planta de conversão de amônia (NH_3) para ácido nítrico (HNO_3). As variáveis são:

X1: Corrente de ar.

X2: Temperatura da água de resfriamento na entrada.

X3: Concentração de ácido nítrico.

Y : Porcentagem de amônia não absorvida, vezes 10.

Os dados são apresentados no quadro 5.1.1.

Quadro 5.1.1- Conjunto de dados de "Stack-Loss".

i	X(1,i)	X(2,i)	X(3,i)	Y(i)
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

5.2. DESEMPENHO DO D DE COOK, DC, E DO DFFITS, DF, PELA VARIAÇÃO DE Y(9) OU DE Y(21).

Observando os gráficos de DF(21) e de DF(9), obtidos pela variação da resposta na vigésima primeira, ou na nona observação (y(21) ou y(9)), figuras 5.2.2 e 5.2.4, respectivamente, com valores fixos nos preditores encontra-se, que eles são retas com pendente positiva, maior para DFFITS(21) do que para DFFITS(9).

O fato anterior verifica-se de modo geral, isto é, o gráfico de DF(i) como função de y(i) é uma reta. Formalmente:

Proposição 5.2.1. DFFITS(i) é função linear de y(i).

Demonstração.

$$\begin{aligned}
 \text{DFFITS}(i) &= \frac{h_i^{1/2}}{1-h_i} * \frac{\hat{e}_i}{s(i)} \\
 &= \frac{h_i^{1/2}}{[1-h_i] s(i)} * [y_i - \hat{y}_i] \\
 &= \frac{h_i^{1/2}}{[1-h_i] s(i)} * [y_i - \sum_{k=1}^n h_{ik} y_k] \\
 &= \frac{h_i^{1/2}}{[1-h_i] s(i)} * [(1-h_i) y_i - \sum_{k \neq i}^n h_{ik} y_k] \\
 &= \frac{h_i^{1/2}}{s(i)} y_i - \frac{h_i^{1/2}}{[1-h_i] s(i)} \sum_{k \neq i}^n h_{ik} y_k.
 \end{aligned}$$

A inclinação da reta DF(i) é $h_i^{1/2} / s(i)$, crescente com relação a h(i).

Quanto ao desempenho dos DF(k); para $k = 1, 2, 3, 4, 9, 14, 21$; pela variação de y(i), $i=21$ ou 9, $k < i$, seus gráficos nas figuras 5.2.2 e 5.2.4, mostram que eles assumem comportamento quase constante para valores afastados do y(i) onde DF(i) = 0, ao qual se chamará de "valor exato" de y(i). Isto levou a supor e verifi-

car o seguinte fato:

Proposição 5.2.1. DF(k) como função de Y(i), $k \neq i$, é assintoticamente constante.

Demonstração. Provar-se-á, equivalentemente, que inclinação da curva DF(k) como função de y(i) é nula quando y(i) tende para mais (+) ou menos (-) infinito.

$$\lim_{y_i \rightarrow \pm\infty} \text{DFFITS}(k) =$$

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{h_k^{1/2}}{(1-h_k) s(k)} * (y_k - \sum_{j=1}^n h_{kj} y_j) \right] =$$

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{(n-p-1) h_k^{1/2}}{(1-h_k) [(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + \sum_{j \neq i,k} (y_j - \hat{y}_j)^2]^{1/2}} * [y_k - \sum_{j=1}^n h_{kj} y_j] \right] =$$

$$K_1 * \lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{y_k - \sum_{j=1}^n h_{kj} y_j}{[(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + K_2]^{1/2}} \right], \text{ onde}$$

$$K_1 = \frac{(n-p-1) h_k^{1/2}}{(1-h_k)} \quad \text{e} \quad K_2 = \sum_{j \neq i,k} (y_j - \hat{y}_j)^2, \text{ ficando então}$$

$$K_1 * \lim_{y_i \rightarrow \pm\infty} \left[\frac{-[(y_i - \sum_{j=1}^n h_{ij} y_j) + K_2] h_{ki}}{[(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + K_2]^2} - \frac{(y_k - \sum_{j=1}^n h_{kj} y_j) [2 (y_i - \sum_{j=1}^n h_{ij} y_j)] (1-h_i)}{[(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + K_2]^2} \right]$$

Agora, dividindo numerador e denominador da última expressão por y_i^4 , maior potência de y_i no seu denominador, obtem-se:

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \text{ DFFITS}(k) = 0$$

Entende-se com isso que, grandes afastamentos de $y(i)$ desde seu valor exato fixam o comportamento do k -ésimo ponto, quando $k > i$.

Os gráficos DC(21) versus Y(21) e de DC(9) versus Y(9) nas figuras 5.2.1. e 5.2.3. têm aproximadamente forma de parábola pois, quando $y(21)$ ou $y(9)$ se afastam de seu respectivo valor exato, a curva vai declinando e vira reta nos extremos. Aliás, o gráfico de DC(i) versus Y(i) tem esse comportamento.

Proposição 5.2.3. DC(i) como função de Y(i) é assintoticamente limitado.

Demonstração. provar-se-á que a inclinação da curva DC(i) como função de $y(i)$ é nula quando $y(i)$ tende para mais (+) ou menos (-) infinito.

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \text{ DC}(i) =$$

$$\frac{h_i}{p(1-h_i)^2} \left\{ \lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{e_i^2}{s^2} \right] \right\} =$$

$$K_1 * \lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left\{ \frac{[y_i - X(i)(X'X)^{-1}X'Y]^2}{[y_i - X(i)(X'X)^{-1}X'Y]^2 + \sum_{k > i}^n [y_k - X(k)(X'X)^{-1}X'Y]^2} \right\}$$

$$\text{onde } K_1 = \frac{h_i}{p(1-h_i)^2}, \text{ ou}$$

$$K_1 * \lim_{y_i \rightarrow \pm\infty} \frac{(A^2 + K_2) \frac{d}{dy_i} [A^2] - A^2 \left[\frac{d}{dy_i} (A^2 + K_2) \right]}{[A^2 + K_2]}, \text{ onde}$$

$$A^2 = [y_i - X(i)(X'X)^{-1}X'Y]^2 \text{ e } K_2 = \sum_{k > i}^n [y_k - X(k)(X'X)^{-1}X'Y]^2$$

ficando então,

$$K_1 * \lim_{y_i \rightarrow \pm\infty} \frac{K_2}{[A^2 + K_2]} \left[\frac{d}{d y_i} A^2 \right] =$$

$$2 K_1 K_2 (1-h_i) * \lim_{y_i \rightarrow \pm\infty} \frac{2A}{[A^2 + K_2]^2} =$$

$$K_3 * \lim_{y_i \rightarrow \pm\infty} \frac{Y_i(1-h_i) - \sum_{k \neq i}^n h_{ik} y_k}{[(1-h_i)y_i - \sum_{k \neq i}^n h_{ik} y_k] + K_2}, \quad K_3 = 2 K_1 K_2 (1-h_i),$$

e se dividir numerador e denominador desta última expressão por y_i^4 , a maior potência de y_i no denominador, e logo calcular o limite tem-se :

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{d y_i} DC(i) = 0.$$

Os gráficos de $DC(k)$ versus $Y(i)$; $k=1,2,3,4,9,14,21$; $i=9$ ou 21 ; $k \neq i$; apresentam uma curvatura nas proximidades do valor exato de $y(i)$, maior em alguns, menor em outros. Mas quando $y(i)$ se afasta do seu valor exato, $DC(k)$ também vai tornando-se constante. Segundo isto, grandes afastamentos de $y(i)$ desde seu valor exato estabilizam o comportamento do k -ésimo ponto.

Proposição 5.2.4. $DC(k)$ como função de $y(i)$ assume comportamento assintoticamente constante.

Demonstração. Demonstra-se-á também aqui, que a inclinação da curva de $DC(k)$ como função de $y(i)$ é zero, no limite.

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{d y_i} DC(i) =$$

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{h_k}{p(1-h_k)} \right] * \left[\frac{e_k^2}{S^2} \right] =$$

$$K_1 \lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} \left[\frac{(y_k - \sum_{j=1}^n h_{kj} y_j)^2}{(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + \sum_{j \neq i}^n (y_j - \hat{y}_j)^2} \right],$$

$$\text{onde } K_1 = \frac{(n-p)h_k}{p(1-h_k)}, \text{ ficando então}$$

$$K_1 \lim_{y_i \rightarrow \pm\infty} \left[\frac{-2 \left[y_k - \sum_{j=1}^n h_{kj} y_j \right] \left[(y_i - \sum_{j=1}^n h_{ij} y_j)^2 + K_2 \right] h_{ki}}{\left[K_2 + (y_i - \sum_{j=1}^n h_{ij} y_j)^2 \right]^2} - \right. \\ \left. \frac{(y_k - \sum_{j=1}^n h_{kj} y_j)^2 \left[2 (y_i - \sum_{j=1}^n h_{ij} y_j) \right] (1-h_i)}{\left[K_2 + (y_i - \sum_{j=1}^n h_{ij} y_j)^2 \right]^2} \right],$$

$$\text{onde } K_2 = \sum_j^n (y_j - \hat{y}_j)^2.$$

Ao dividir numerador e denominador da última expressão por y_i^4 a maior potência de y_i no seu denominador, para então tomar o limite e obter:

$$\lim_{y_i \rightarrow \pm\infty} \frac{d}{dy_i} DC(k) = 0.$$

O fato do D de Cook ser limitado, como foi provado na proposição 5.2.3, mostra que ele não é um diagnóstico confiável pois, afastamentos grandes de $y(i)$ desde seu valor exato não são destacados por ele. Entretanto, DFFITS destaca suficientemente esses afastamentos, por ser função linear de $y(i)$.

Por outro lado, determinando as regiões onde $DF(i)$ e $DC(i)$, como funções de $y(k)$, detectam a i -ésima observação como discrepante, vê-se que elas são diferentes. Verifiquemos:

DFFITS(k) como função de $y(21)$; $5 < y(21) < 35$; não detecta como discrepante a k-ésima observação :

Para	Desde	Até
k = 21	20.4	29.8
k = 2	-	24
k = 4	-	18

Tabela 5.2.1

DC(k) como função de $y(21)$; $5 < y(21) < 35$; não detecta como discrepante a k-ésima observação :

Para	Desde	Até
k = 21	20	30.4
k = 2	-	25.6
k = 4	-	24
	e 29	-

Tabela 5.2.2

A primeira, terceira, nona e décima quarta observações não chegam a ser discrepantes pela variação do valor da vigésima primeira resposta.

DFFITS(k) como função de $y(9)$; $2 < y(9) < 30$; não detecta como discrepante a k-ésima observação :

Para	Desde	Até
k = 9	10	27
k = 21	-	-

Tabela 5.2.3

DC(k) como função de $y(9)$; $2 < y(9) < 30$; não detecta como discrepante a k-ésima observação :

Para	Desde	Até
k = 9	8	29.4
k = 21	-	-

Tabela 5.2.4

A primeira, segunda, terceira, quarta e décima quarta observações não chegam a ser discrepantes pela variação do valor da nona resposta.

Nota. O simbolo - indica algum valor que está fora do intervalo de variação de $y(21)$ ou de $y(9)$.

Isto mostra que não só para valores afastados de $y(i)$, mas também, para valores não muito distantes do seu valor exato, DFFITS é mais eficaz na detecção de observações discrepantes do que o D de Cook.

5.3. DESEMPENHO DO D DE COOK E DO DFFITS PELA VARIAÇÃO DE $x(3,21)$ OU DE $x(3,9)$.

Quando se faz variar o valor do terceiro preditor na vigésima primeira observação, ou na nona, os gráficos de DF, figuras 5.3.2 à 5.3.28 com último índice par, são aproximadamente retas com inclinação positiva nos extremos e apresentam uma deformação ao redor do valor original do terceiro preditor.

Essa deformação é assimétrica à direita e tem vértice para cima quando $y(i)$ é menor que seu valor exato; é assimétrica à esquerda e tem vértice para baixo quando $y(i)$ é maior que seu valor exato. Ademais, ela cresce com o aumento da distância entre $y(i)$ e seu valor exato, e só se anula quando $y(i)$ coincide com ele.

Até agora não se encontrou uma razão explícita para a ocorrência da deformação no gráfico de DF pela variação de $x(3,i)$ nas proximidades do seu valor original mas, parece razoável que aconteça já que DF não é função linear de $x(i,j)$. Isso por sua vez faz difícil explicar porque o gráfico vira uma reta quando $y(i)$ coincide com seu valor exato. A análise da complicada expressão escrita por extenso nada esclareceu sobre a forma como $x(3,i)$ determina $DF(i)$.

Também para $1 < x(3,i) < 180$, os gráficos de DC, figuras 5.3.1 à 5.3.27 com último índice ímpar, resultam ser aproximadamente parábolas que têm uma deformação com vértice para abaixo, perto do $x(3,i)$ original.

O gráfico de DC tem o eixo deslocado para a direita do $x(3,i)$ original e apresenta a deformação no ramo esquerdo, quando $y(i)$ é menor que seu valor exato, e tem o eixo deslocado para a esquerda do $x(3,i)$ original e a deformação no ramo direito, quando $y(i)$ é maior que seu valor exato. Quando $y(i)$ coincide com seu valor exato, a curva de DC é uma legítima parábola com eixo perto do $x(3,i)$ original, ou talvez sobre o mesmo.

Essa deformação em DC cresce com $y(i)$ deslocando-se à direita, ou à esquerda, do seu valor exato e se anula quando $y(i)$ coincide com esse valor exato.

Não se encontrou uma justificativa para o aparecimento da deformação no gráfico de DC, tampouco neste caso, a análise da fórmula de DC escrita por extenso fornece justificativa alguma.

Pela simples observação dos gráficos de DC e DF como funções

de $x(3,i)$ não se consegue encontrar diferença importante entre seus desempenhos mas, determinando os intervalos onde eles detectam, ou onde não detectam, a correspondente observação como discrepante, encontram-se que eles são diferentes:

DFITS(21) como função de $x(3,21)$ não detecta como discrepante a vigésima primeira observação para:

y(21)	Desde	Até
15	162	-
20	106	162
25	63	120
30	21	90
35	-	23

Tabela 5.3.1

DC(21) como função de $x(3,21)$ não detecta como discrepante a vigésima primeira observação para:

y(21)	Desde	Até
15	161	-
20	89	162
25	62	120
30	21	93
35	-	25

Tabela 5.3.2

DFITS(9) como função de $x(3,9)$ não detecta como discrepante a nona observação para:

y(9)	Desde	Até
13	78	151
15	72	137
17	66	124
25	18	92
30	-	30

Tabela 5.3.3

DC(9) como função de $x(3,9)$ não detecta como discrepante a nona observação para:

y(9)	Desde	Até
13	77	151
15	72	138
17	66	124
25	18	94
30	-	33
	e 84	85

Tabela 5.3.4

Essas diferenças geralmente se apresentam nos extremos desses intervalos e não são muito grandes mas, há casos em que o são, por exemplo, quando $y(21) = 20$ o intervalo onde DC não detecta a vigésima primeira observação é bem maior que o intervalo onde DF não a detecta. E, quando $y(9) = 30$, DC não consegue detectar ademais os pontos isolados com valores de $x(3,9)$ entre 84 e 85.

A forma como a variação de $x(3,21)$, ou de $x(3,9)$, afeta os $DF(k)$ e os $DC(k)$ se pode inferir dos gráficos correspondentes, figuras 5.3.1. à 5.3.28. Nos extremos do intervalo $1 < x(3,i) < 180$ esses gráficos tendem para retas constantes, o que indica que afastamentos grandes de $x(3,i)$ desde seu valor original não os alteram mas, com $x(3,i)$ próximos do seu original os $DF(k)$ e os $DC(k)$ sofrem também uma deformação. Em particular, têm uma deformação maior: $DF(2)$, $DF(4)$, $DC(2)$ e $DC(4)$, quando $x(3,21)$ varia, e $DF(1)$, $DF(2)$, $DC(1)$ e $DC(2)$, quando $x(3,9)$ varia.

Em geral, as deformações nos gráficos de $DF(k)$ e de $DC(k)$, como funções de $X(3,i)$, aumentam ou diminuem de acordo com o maior ou menor afastamento de $y(i)$ desde seu valor exato.

Tratando com os $DF(k)$ e os $DC(k)$, pela variação de $x(3,i)$, $i > k$, observa-se uma diferença notável no seu comportamento, quando se comparam as regiões onde os mesmos não detectam a k -ésima observação como discrepante, por exemplo:

$DF(k)$ como função de $x(3,21)$ não detecta como discrepante a k -ésima observação, $1 < x(3,21) < 180$.

$DC(k)$ em função de $x(3,21)$ não detecta como discrepante a k -ésima observação com $1 < x(3,21) < 180$

Para	Desde	Até
$k = 2$	-	110
$k = 4$	45	106

Tabela 5.3.5

Para	Desde	Até
$k = 2$	- e 175	115 -
$k = 4$	-	133

Tabela 5.3.6

Nota. O símbolo - indica um número que está fora do intervalo $1 < x(3,21) < 180$.

Isso mostra que na situação de apenas $x(3,9)$ ou $x(3,21)$ variando, $DF(k)$ pode detectar pontos discrepantes que o D de Cook não consegue detectar. Também nesta situação $DF(k)$ é mais eficaz na detecção de observações discrepantes num ajuste.

5.4. DESEMPENHO DO $DF(k)$ E DO D DE COOK PELA VARIAÇÃO CONJUNTA DE $y(i)$ E $x(3,i)$, $i = 21$ OU 9 .

A variação conjunta de $y(i)$ e $x(3,i)$ em $DF(i)$, $i = 21$ ou $i = 9$ gera uma superfície quase plana, com uma deformação ao redor do valor original de $x(3,i)$ e que tem vértice sobre um valor $x(3,i)$

próximo dele, talvez sobre o próprio. A superfície é antisimétrica com relação ao plano $\hat{y}(i) = y(i)$, onde a i -ésima observação é ajustada exatamente pelo modelo, como se ve nas figuras 5.4.1 e 5.4.2.

A deformação vai crescendo para valores de $y(i)$ que se afastam do seu valor exato à direita, ou à esquerda, e anula-se quando $y(i)$ assume o próprio. Percorrendo $y(i)$, desde valores menores até valores maiores que seu valor exato, a deformação muda de assimétrica à direita para assimétrica à esquerda e o vértice que aponta para cima passa à apontar para abaixo. Com isto entende-se que existe uma região de valores $x(3,i)$ para cada valor de $y(i)$, que faz com que o peso da observação correspondente tenha menor variabilidade na determinação do modelo ajustado. Esta região neste caso, está localizada em volta do valor original de $x(3,i)$.

Para $i=9$ e $i=21$, a superfície gerada pela variação conjunta de $y(i)$ e $x(3,i)$ em $DF(i)$ apenas tem uma diferença visível: que a deformação no caso de $i=9$ cresce mais rápido do que com $i=21$. De resto, elas são semelhantes.

Devido à dificuldade na construção do gráfico da superfície de $DC(i)$ gerada pela variação conjunta de $y(i)$ e $x(3,i)$ optou-se por não apresentá-la aqui. Porém, os perfís de $DC(i)$ como função de $x(3,i)$ para valores fixados de $y(i)$ dão uma idéia suficientemente clara da forma da superfície do $DC(i)$ quando, ambos, $y(i)$ e $x(3,i)$ variam.

O comportamento de $DC(9)$ como função de $y(9)$ e $x(3,9)$ e de $DC(21)$ como função de $y(21)$ e $x(3,21)$, que correspondem a duas observações de naturezas diferentes, opostas, nos dados originais (a nona é bem comportada, a vigésima primeira é discrepante), indica que há uma região de valores $x(3,i)$ próximos do seu valor original, tal que o peso do ponto correspondente no conjunto de dados tem menor variação e mais, com valores grandes de $y(i)$ o peso tende a estabilizar-se. Como pode observar-se nas figuras 5.3.1 à 5.3.27 de último índice impar.

Nessas condições, $DC(i)$ como função conjunta de $y(i)$ e $x(3,i)$ é ineficaz na detecção de grandes afastamentos de $y(i)$ desde seu valor original e também de afastamentos conjuntos médios ou grandes de $y(i)$ e $x(3,i)$ desde seus valores exatos.

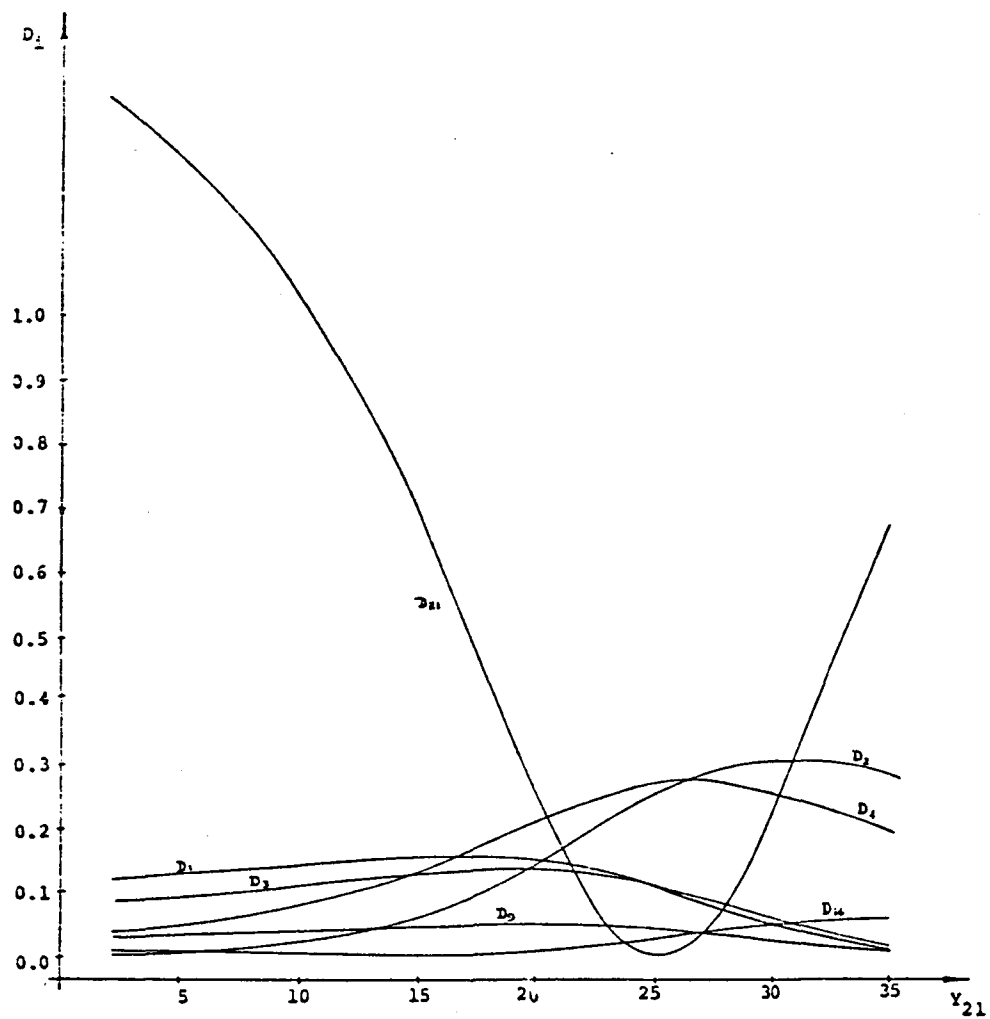


Figura 5.2.1 - D_i versus Y_{21} ; $i=1,2,3,4,9,14,21$.

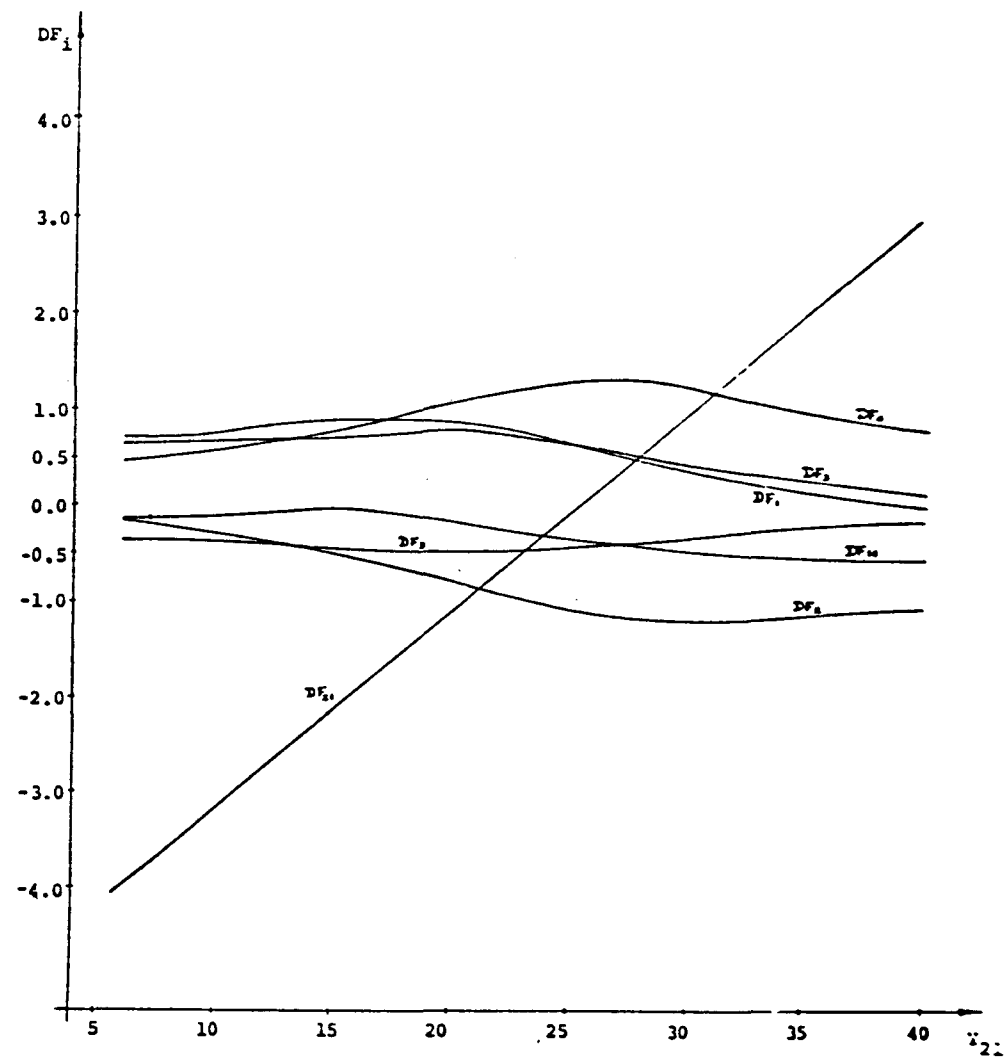


Figura 5.2.2 - $DFFITS_i$ versus Y_{21} ; $i=1,2,3,4,9,14,21$.

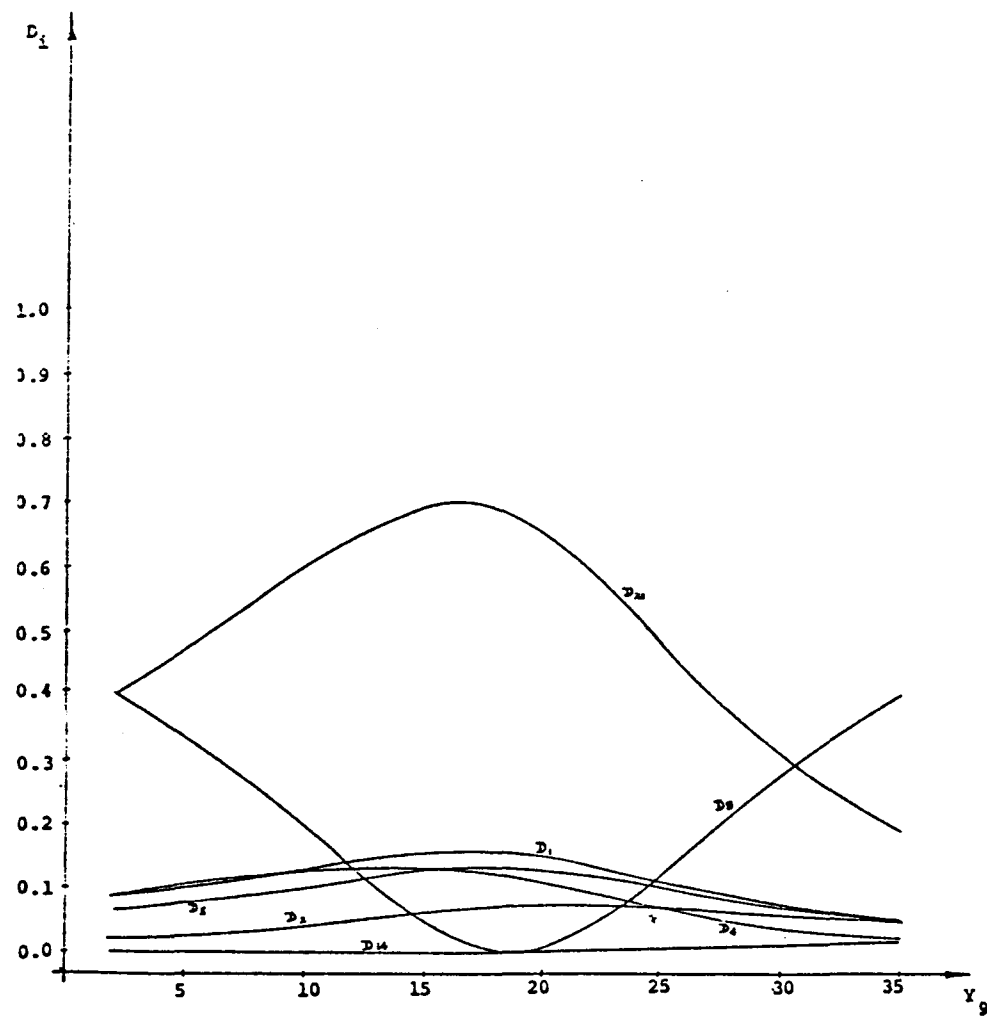


Figura 5.2.3 - D_i versus Y_9 , $i=1,2,3,4,9,14,21$.

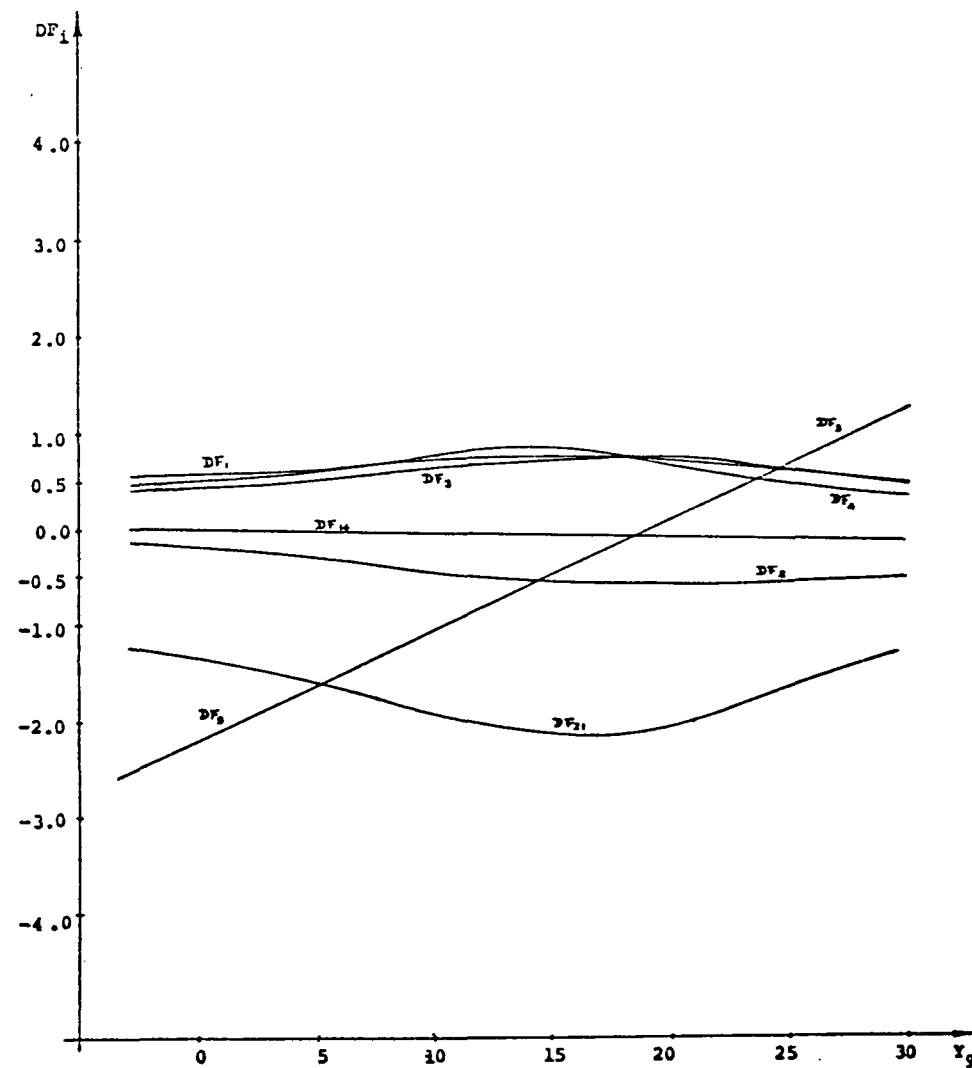


Figura 5.2.4 - $DFFITS_i$ versus Y_9 , $i=1,2,3,4,9,14,21$.

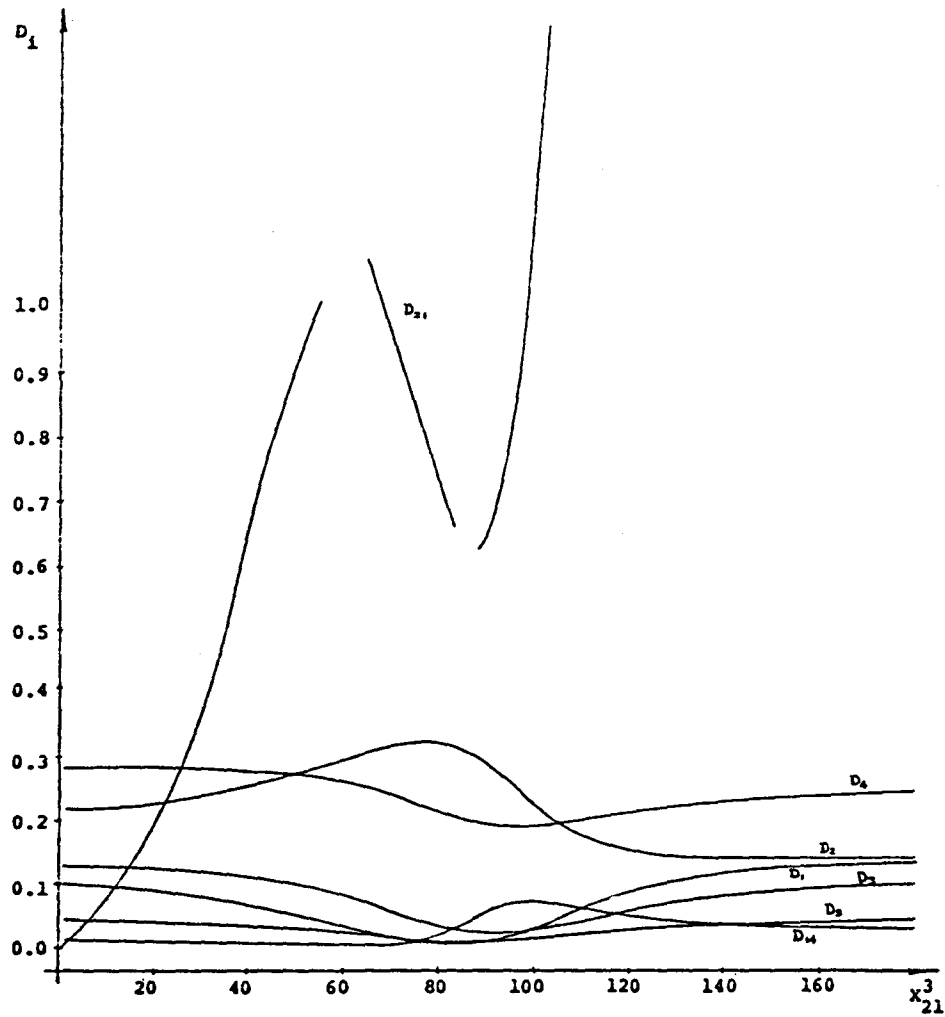


Figura 5.3.1 - D_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=35$.

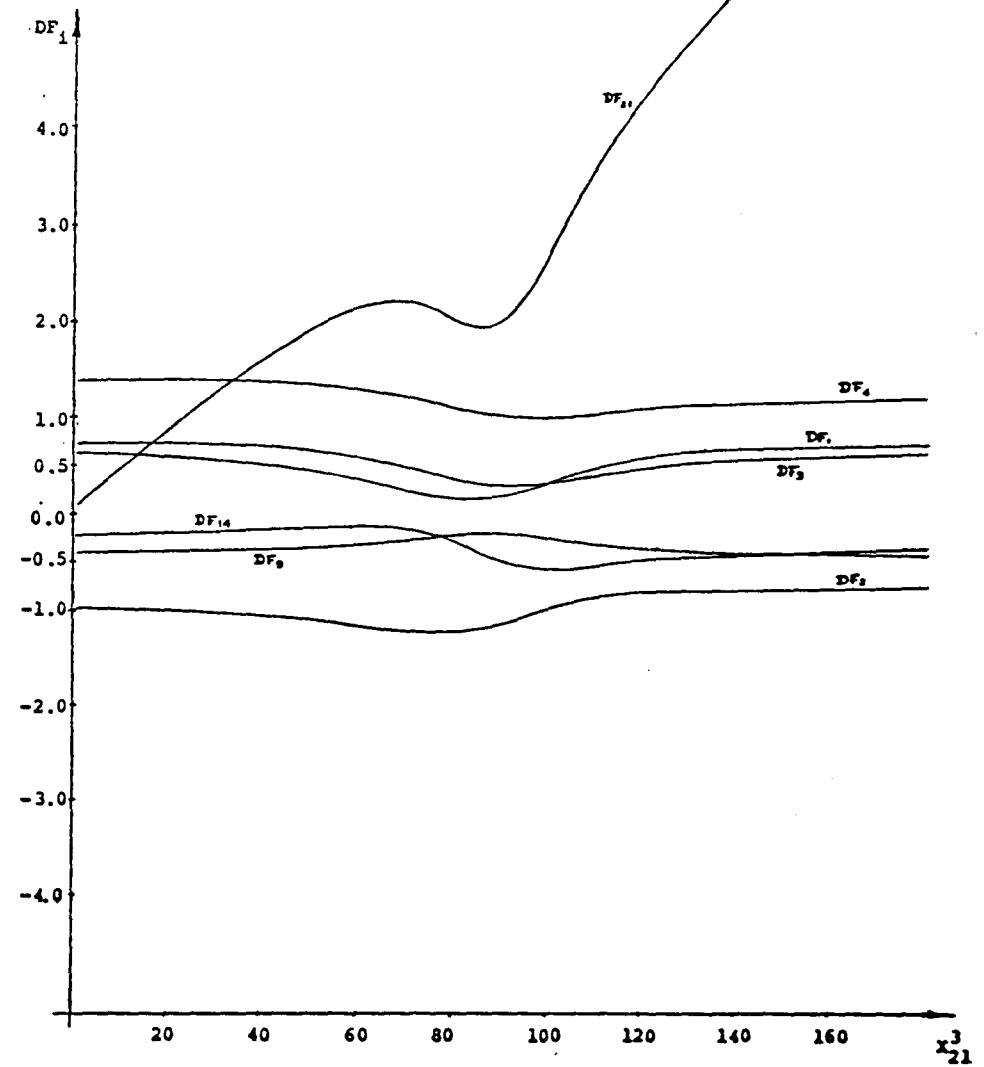


Figura 5.3.2 - $DFFITS_i$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=35$.

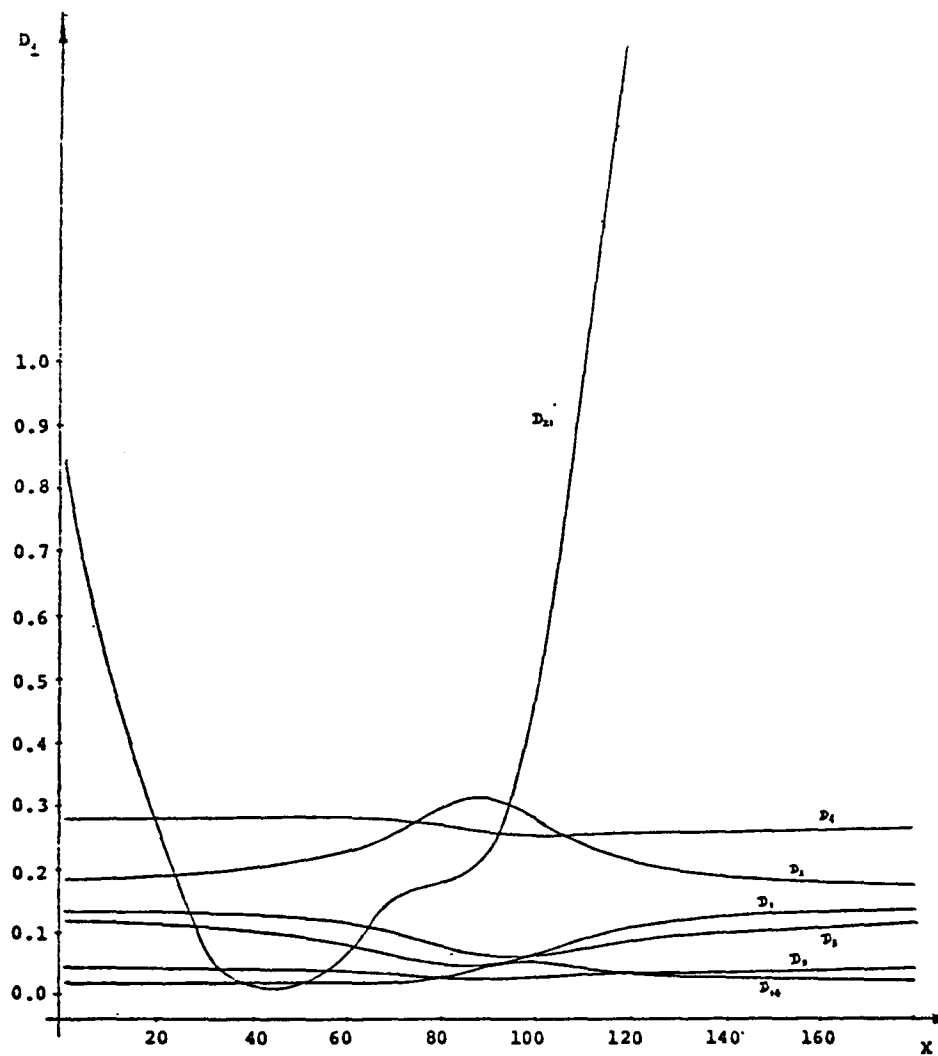


Figura 5.3.3 - D_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=30$.

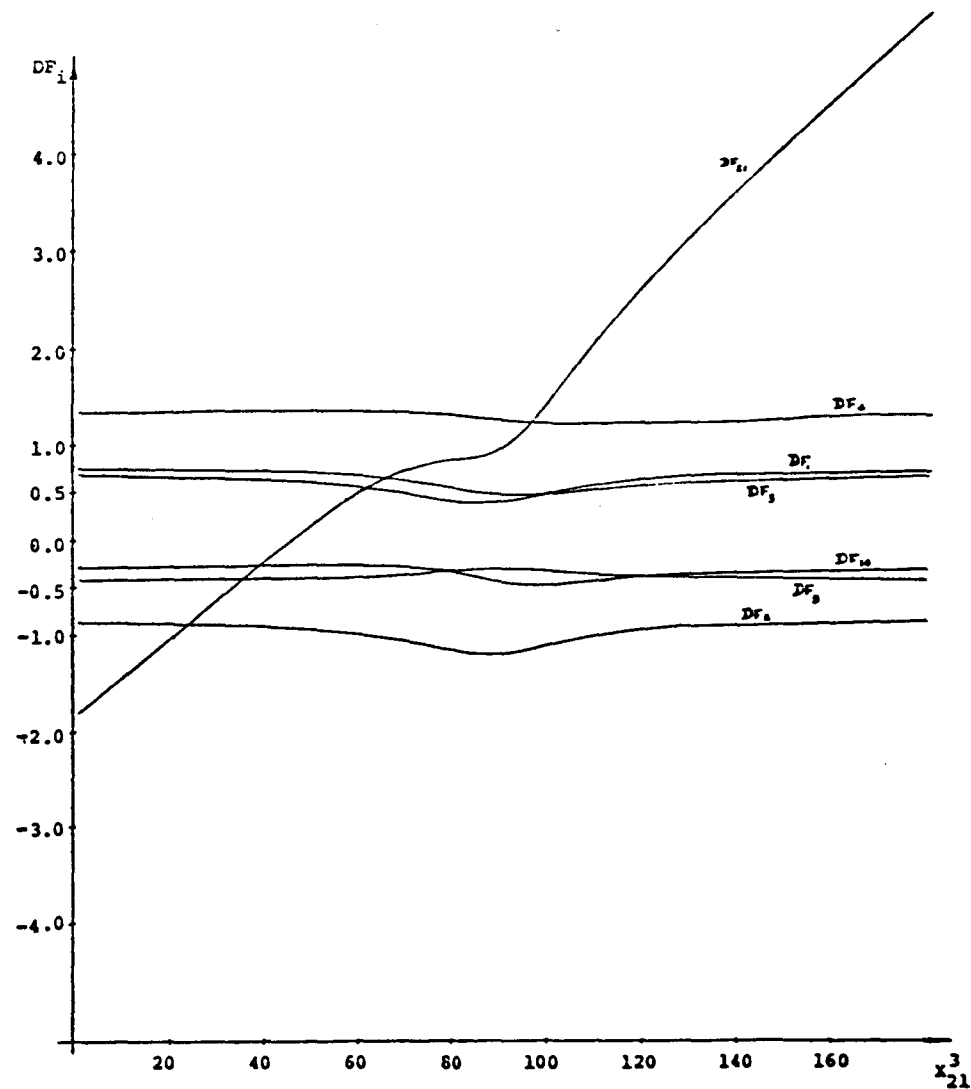


Figura 5.3.4 - $DFFITS_i$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=30$.

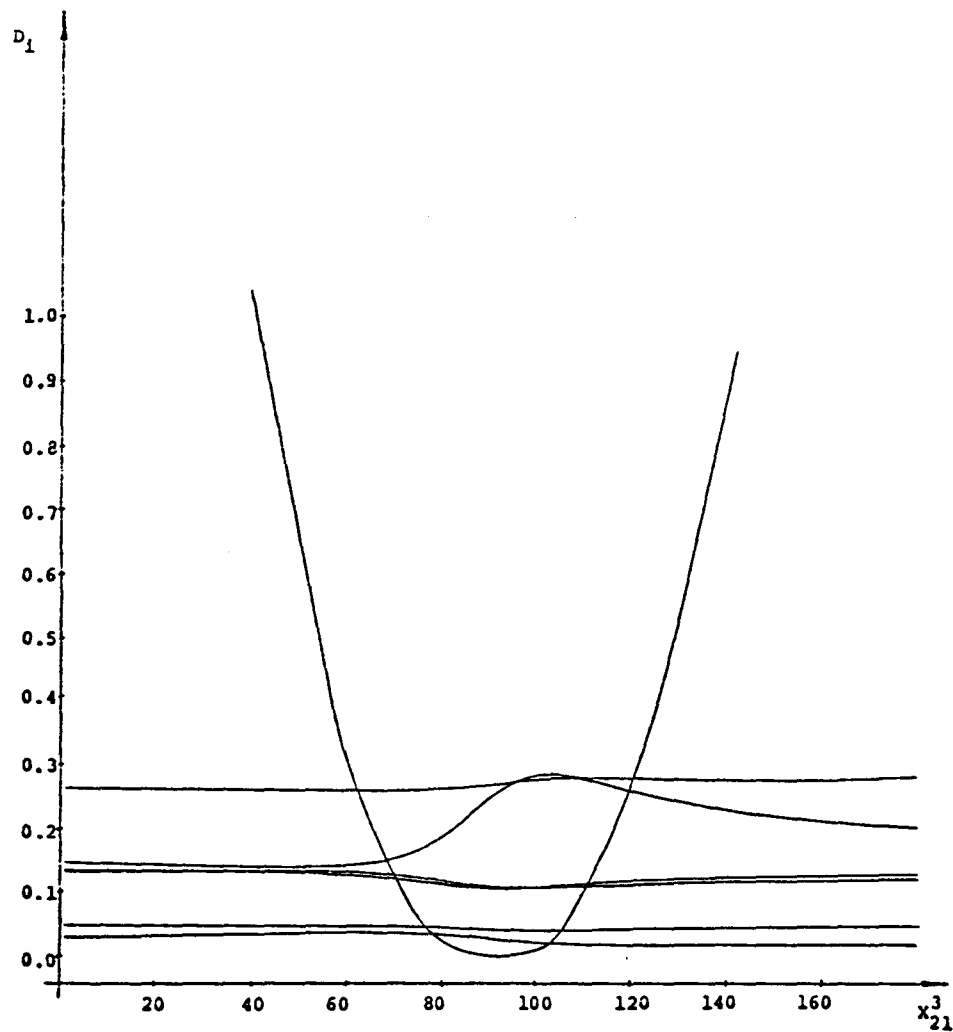


Figura 5.3.5 - D_1 versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=25$.

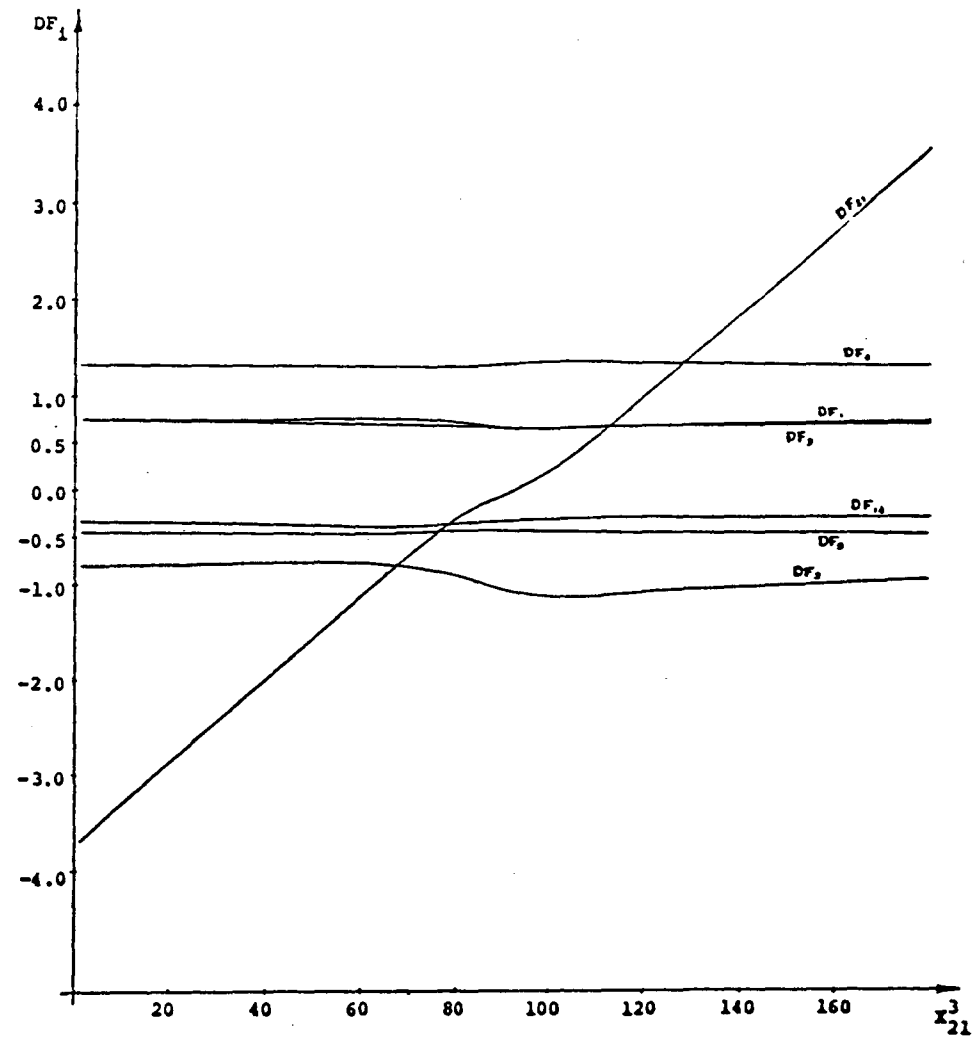


Figura 5.3.6 - $DFFITS_1$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=25$.

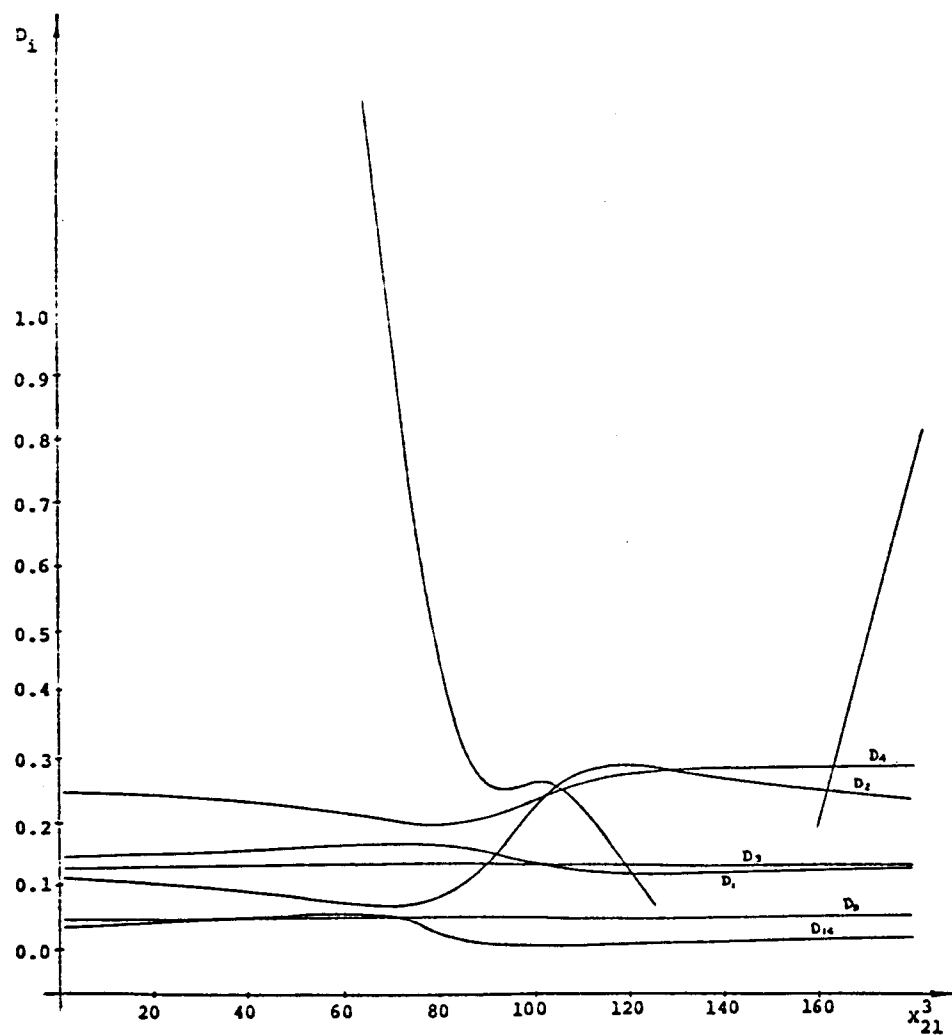


Figura 5.3.7 - D_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=20$.

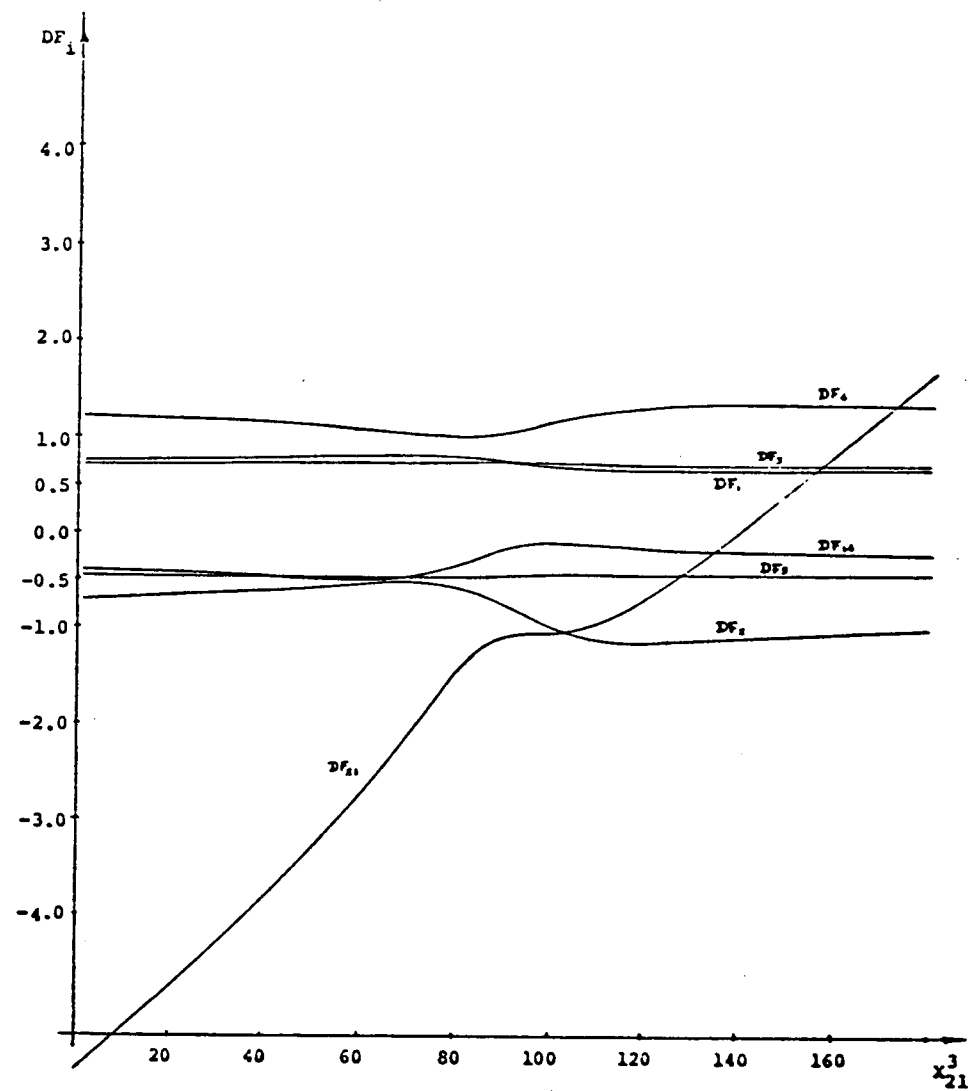


Figura 5.3.8 - $DFFITS_i$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=20$.

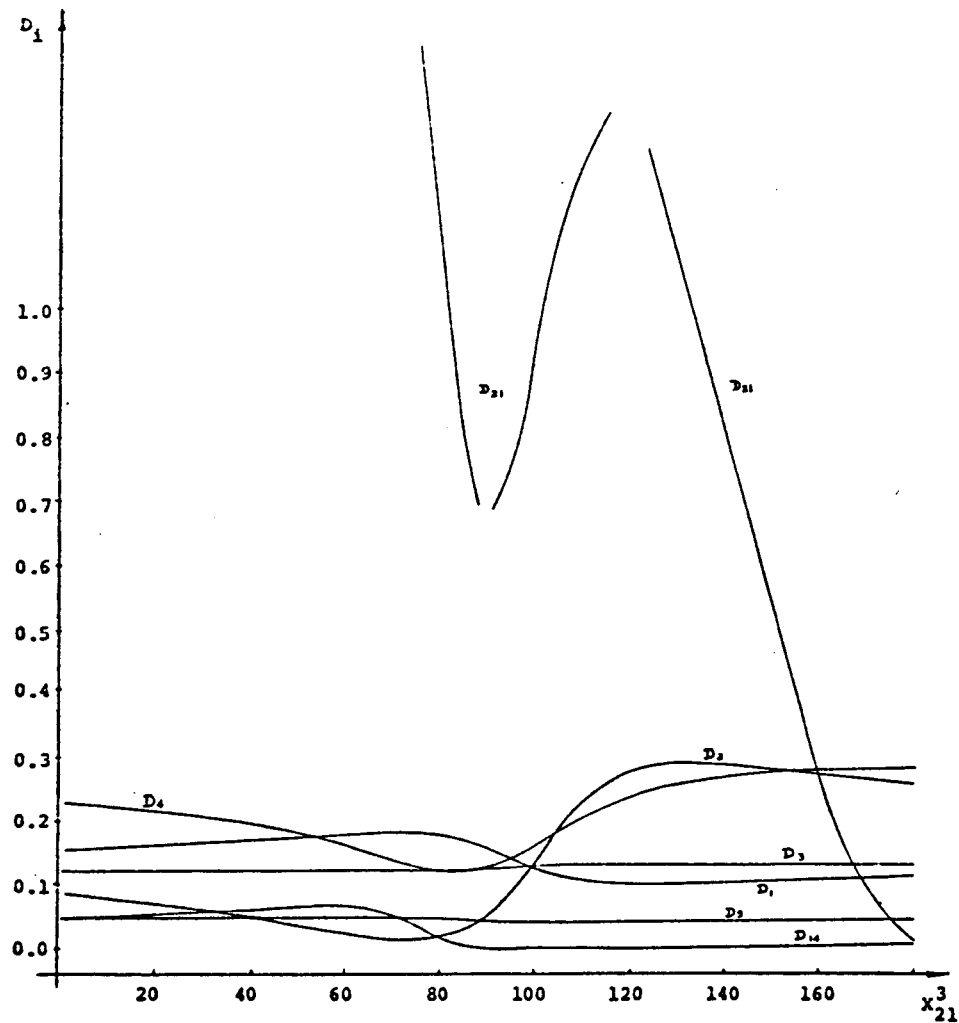


Figura 5.3.9 - D_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=15$.

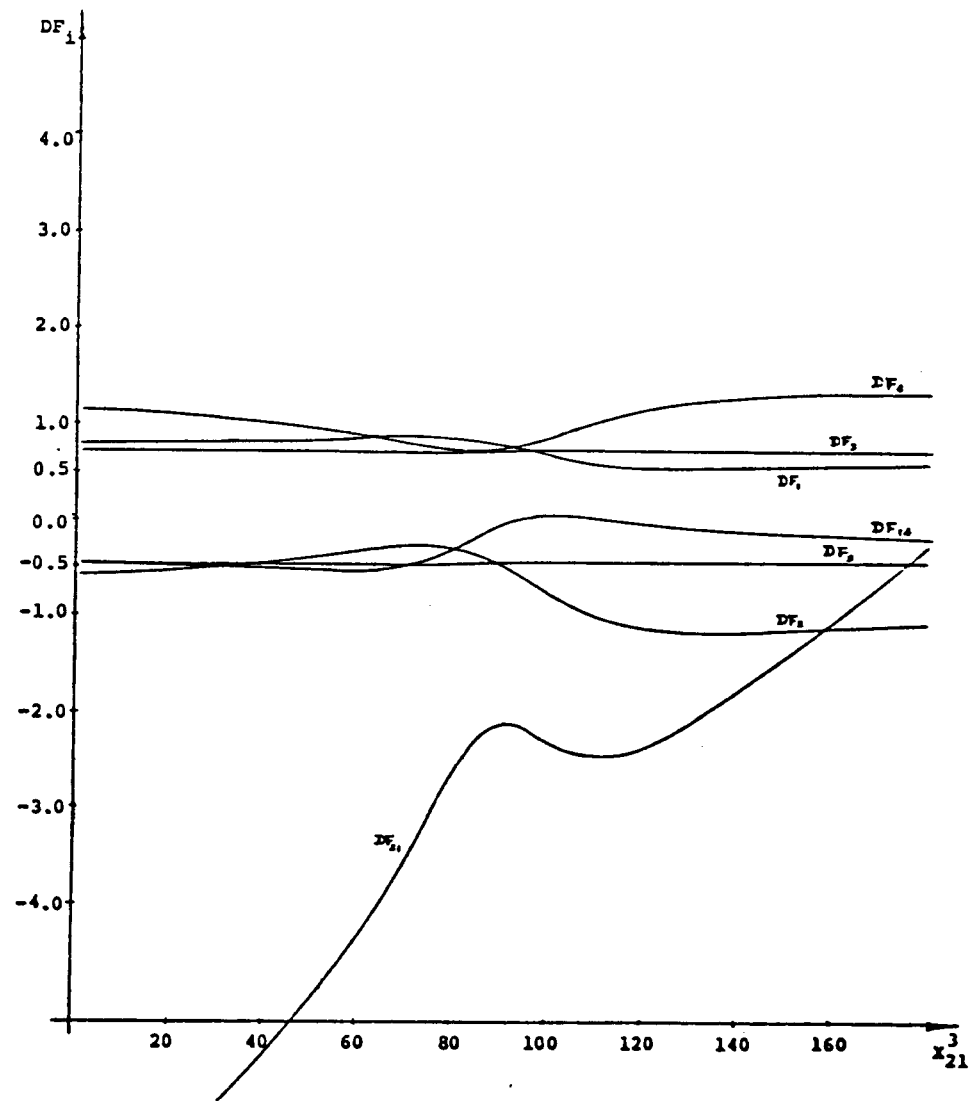


Figura 5.3.10- $DFFITS_i$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=15$.

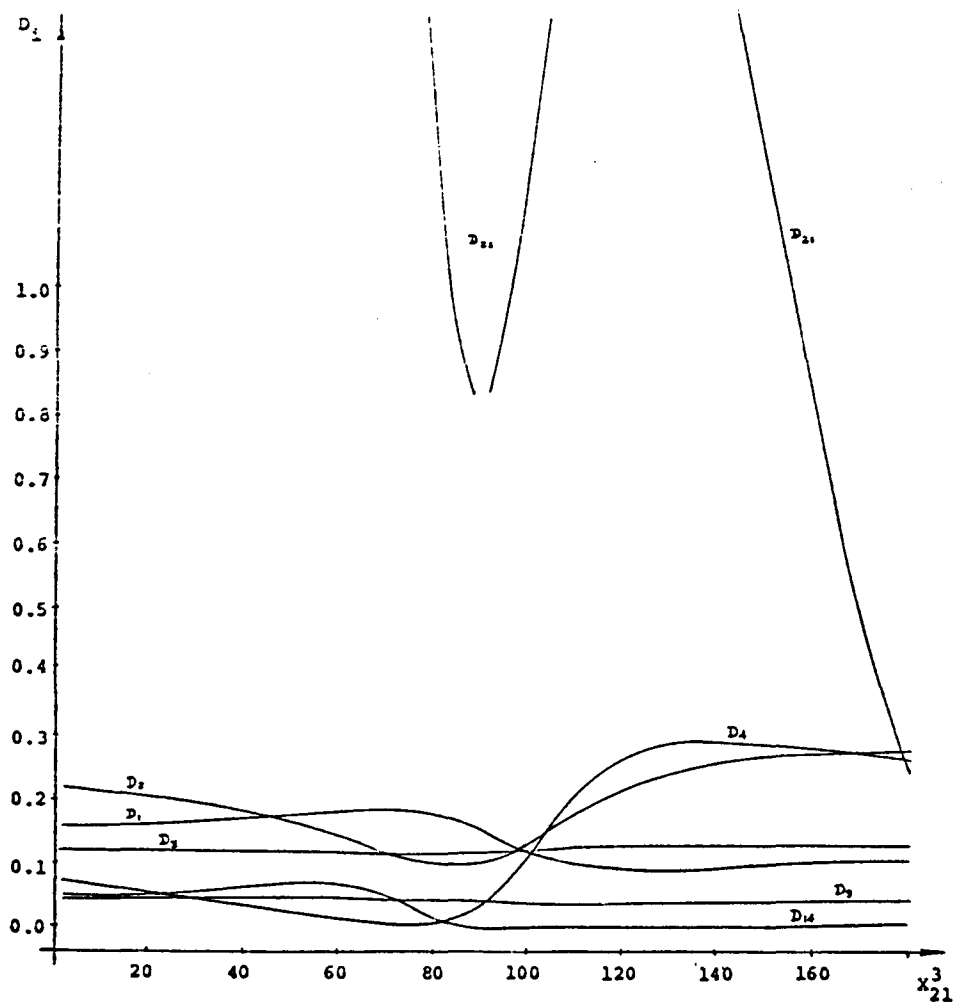


Figura 5.3.11- D_i versus X_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=13$.

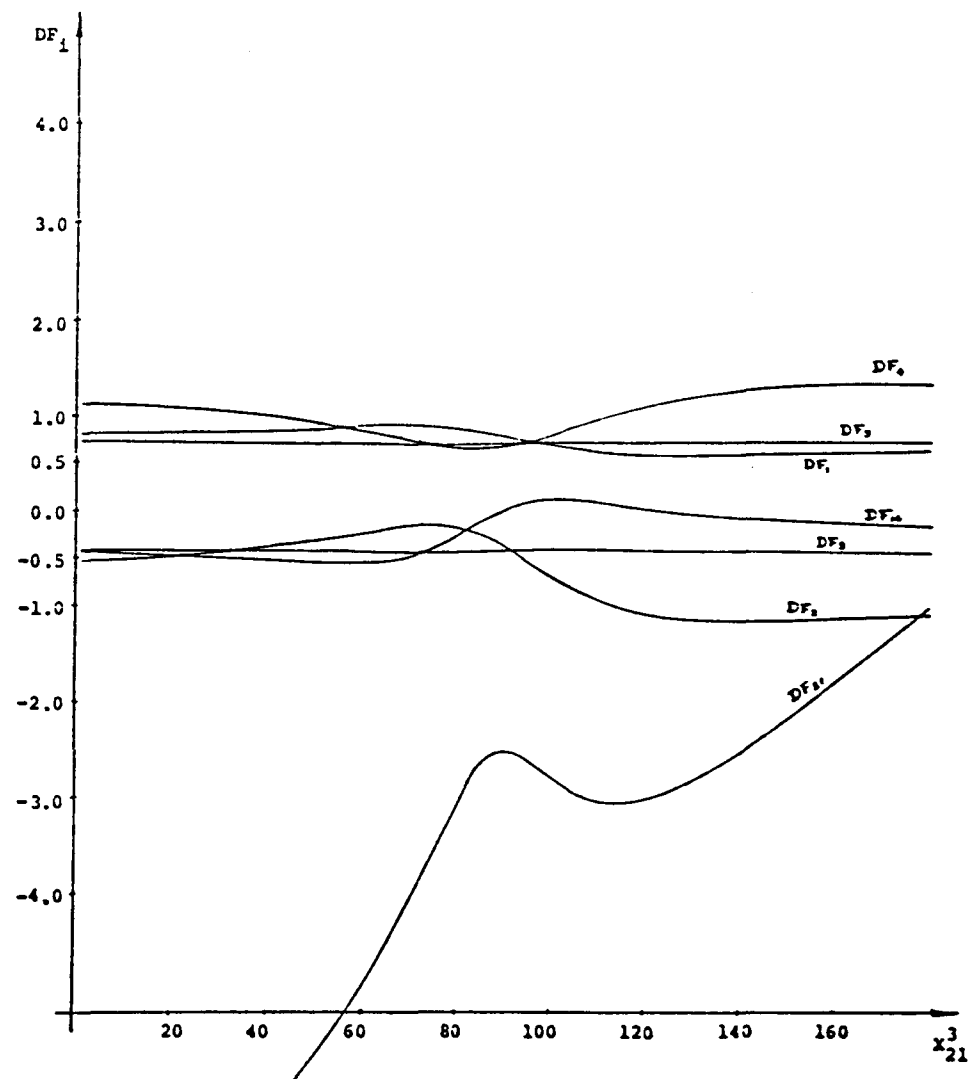


Figura 5.3.12 - $DFFITS_i$ versus X_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=13$.

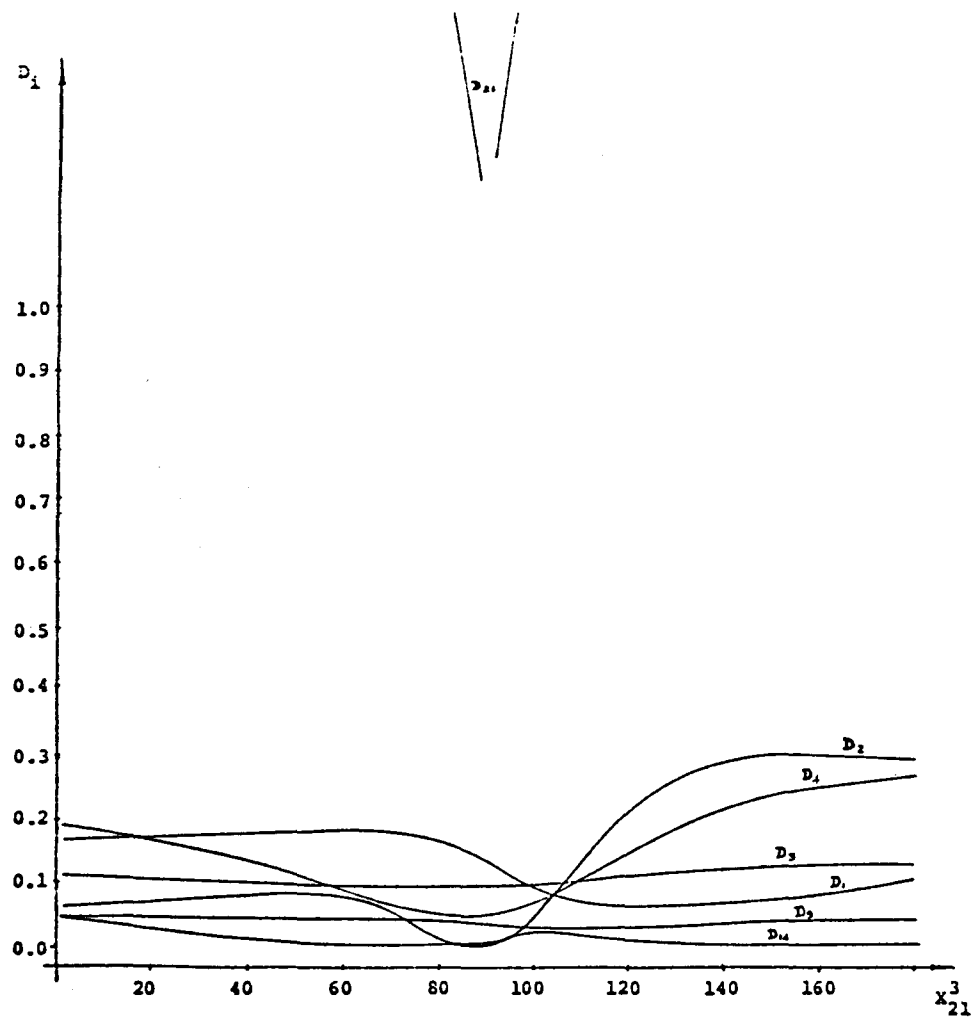


Figura 5.3.13 - D_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=5$.

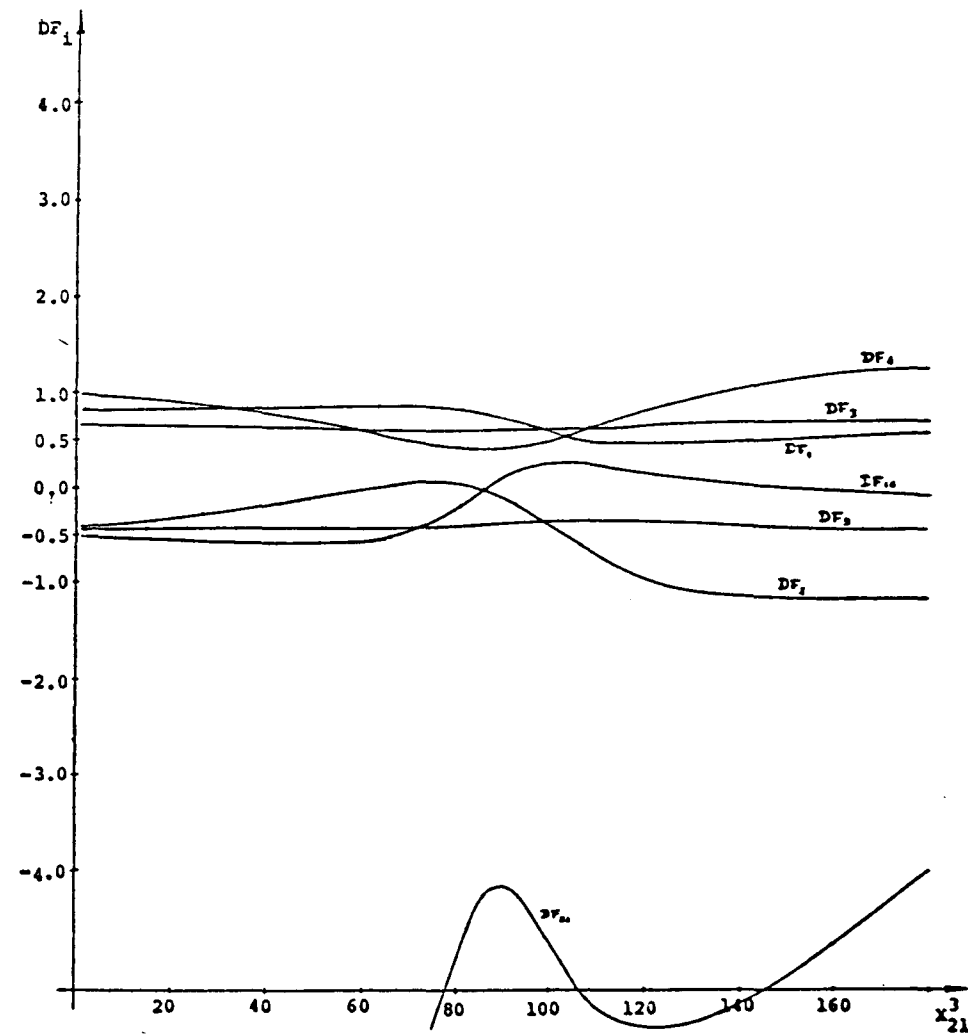


Figura 5.3.14 - $DFFITS_i$ versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=5$.

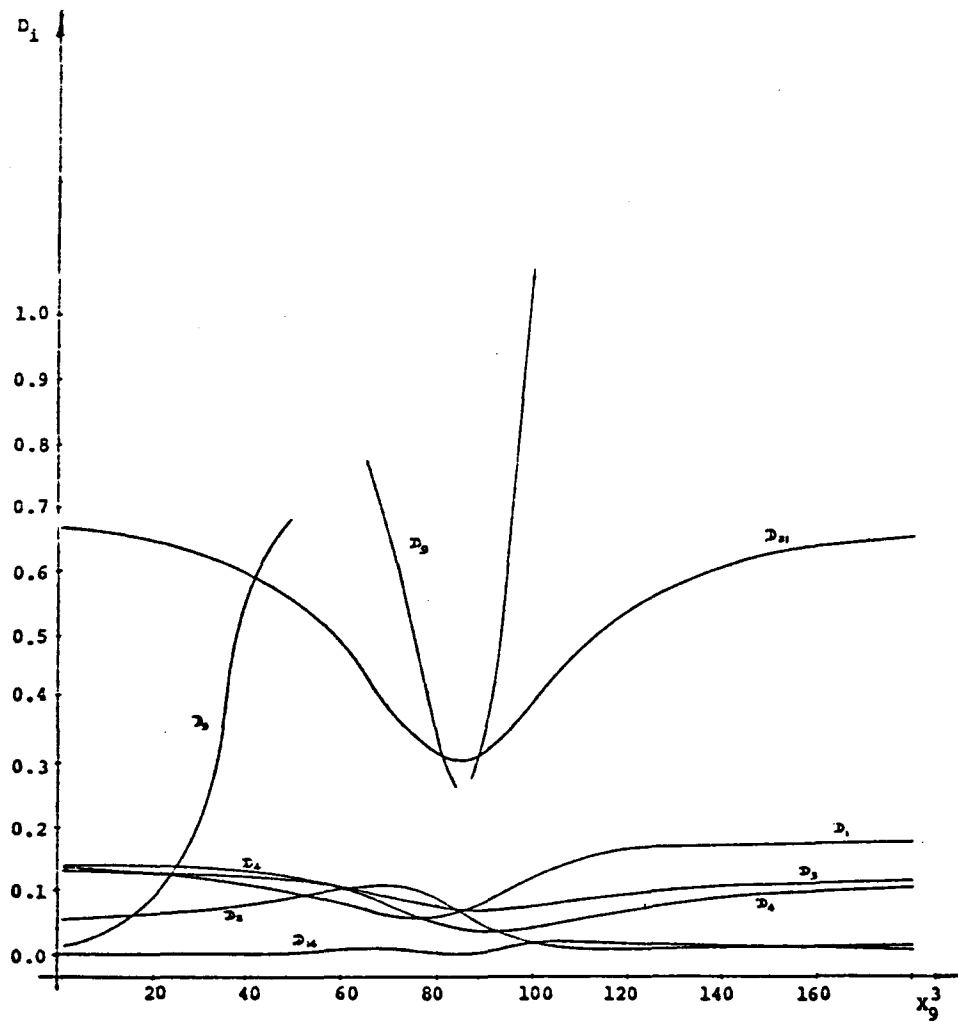


Figura 5.3.15 - D_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=30$.

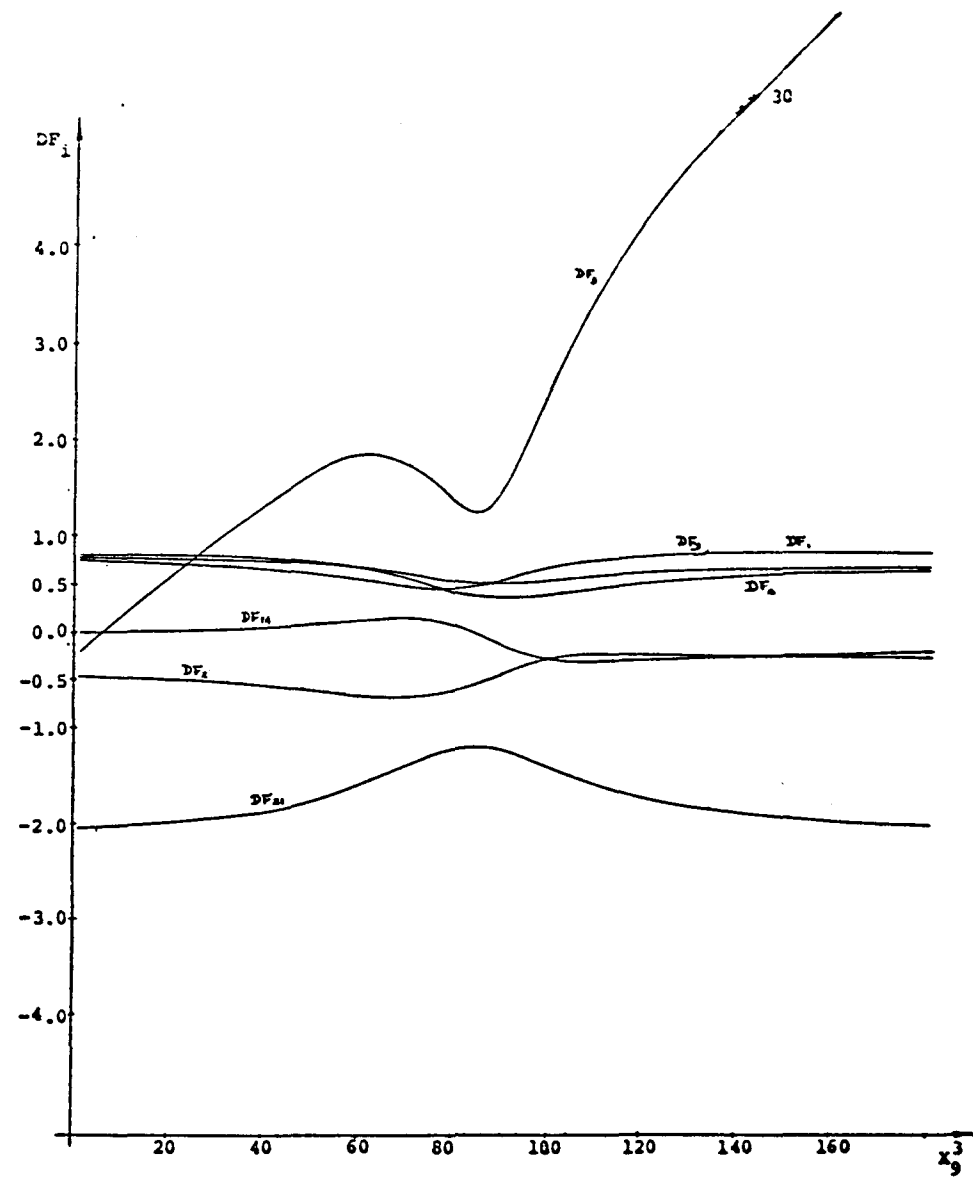


Figura 5.3.16 - $DFFITS_i$ versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=30$.

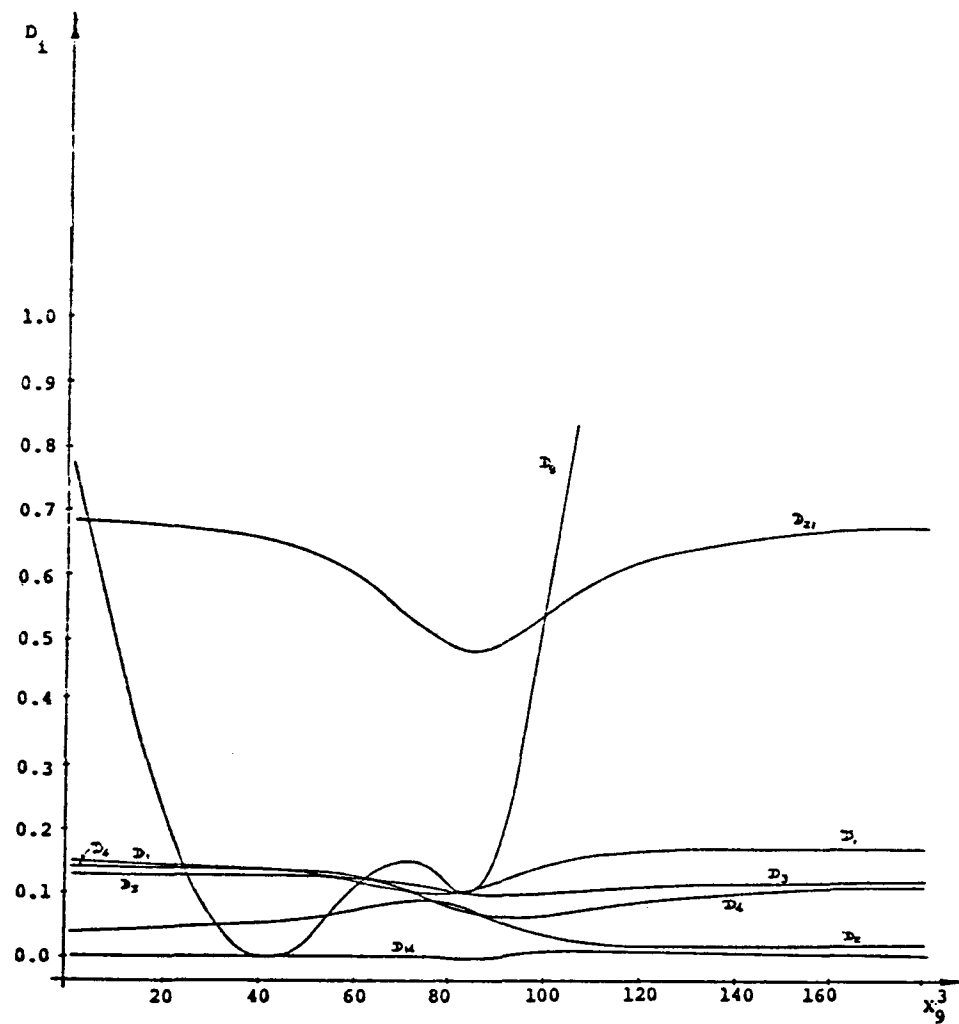


Figura 5.3.17 - D_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=25$.

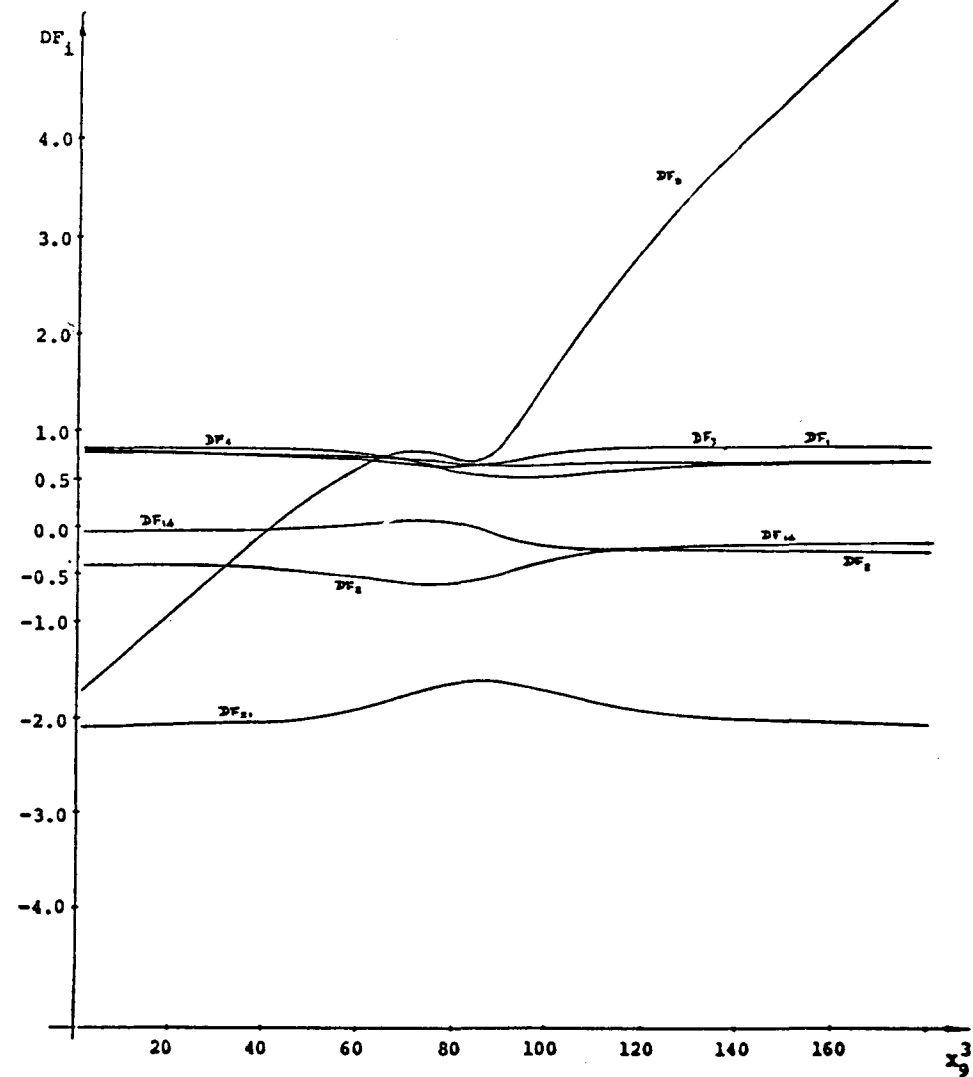


Figura 5.3.18 - $DFFITS_i$ versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=25$.

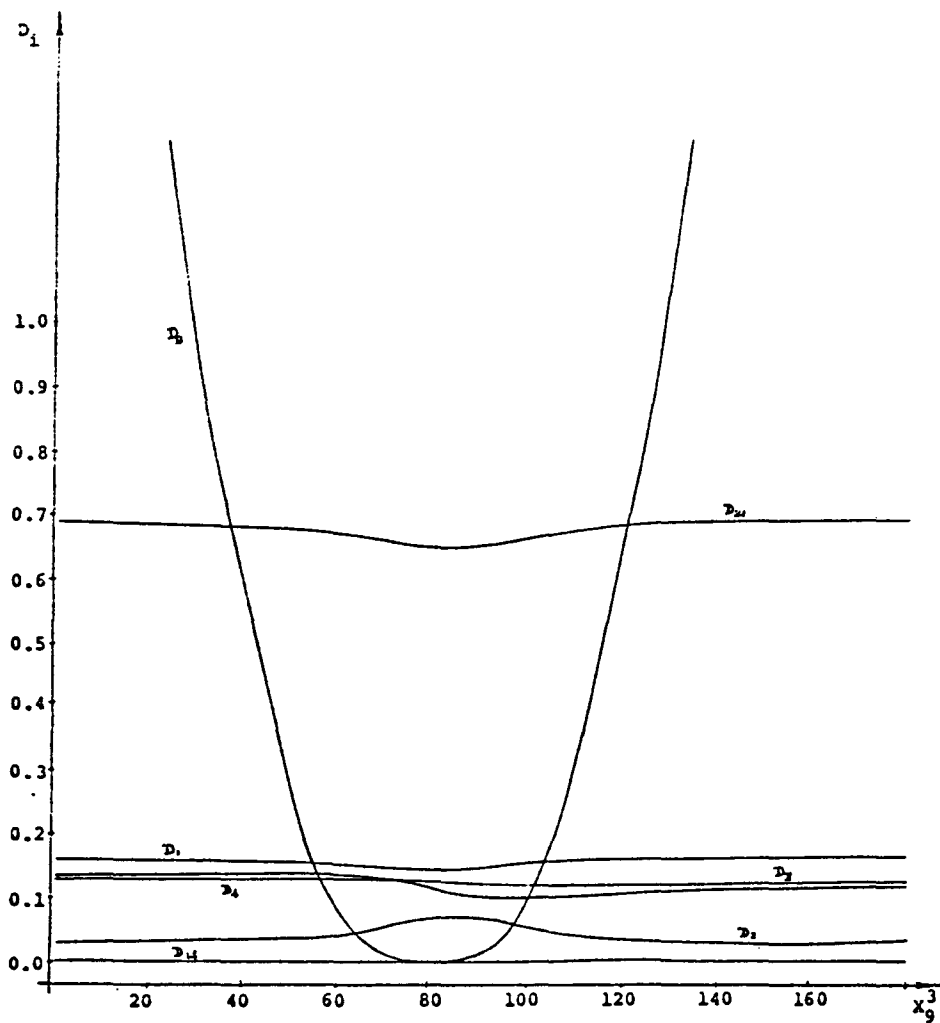


Figura 5.3.19 - D_i versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=20$.

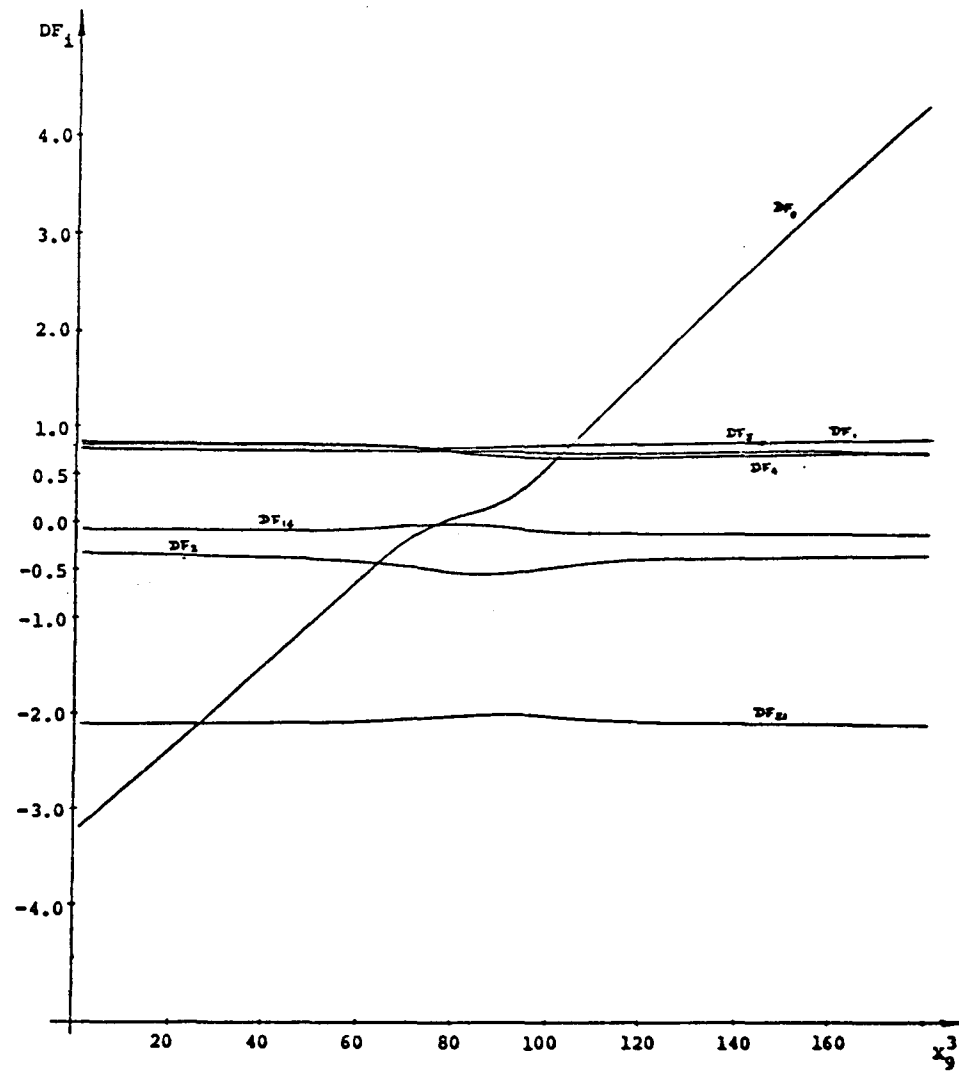


Figura 5.3.20 - $DFFITS_i$ versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=20$.

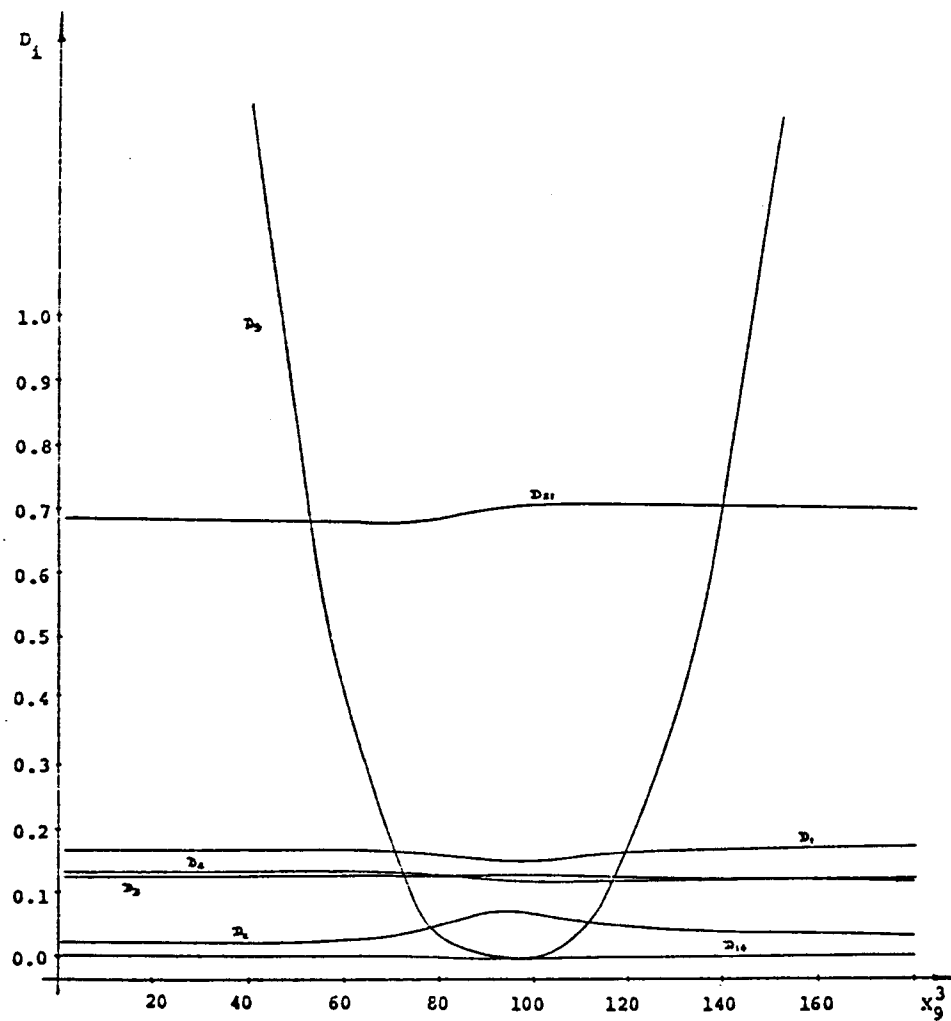


Figura 5.3.21 - D_i versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=17$.

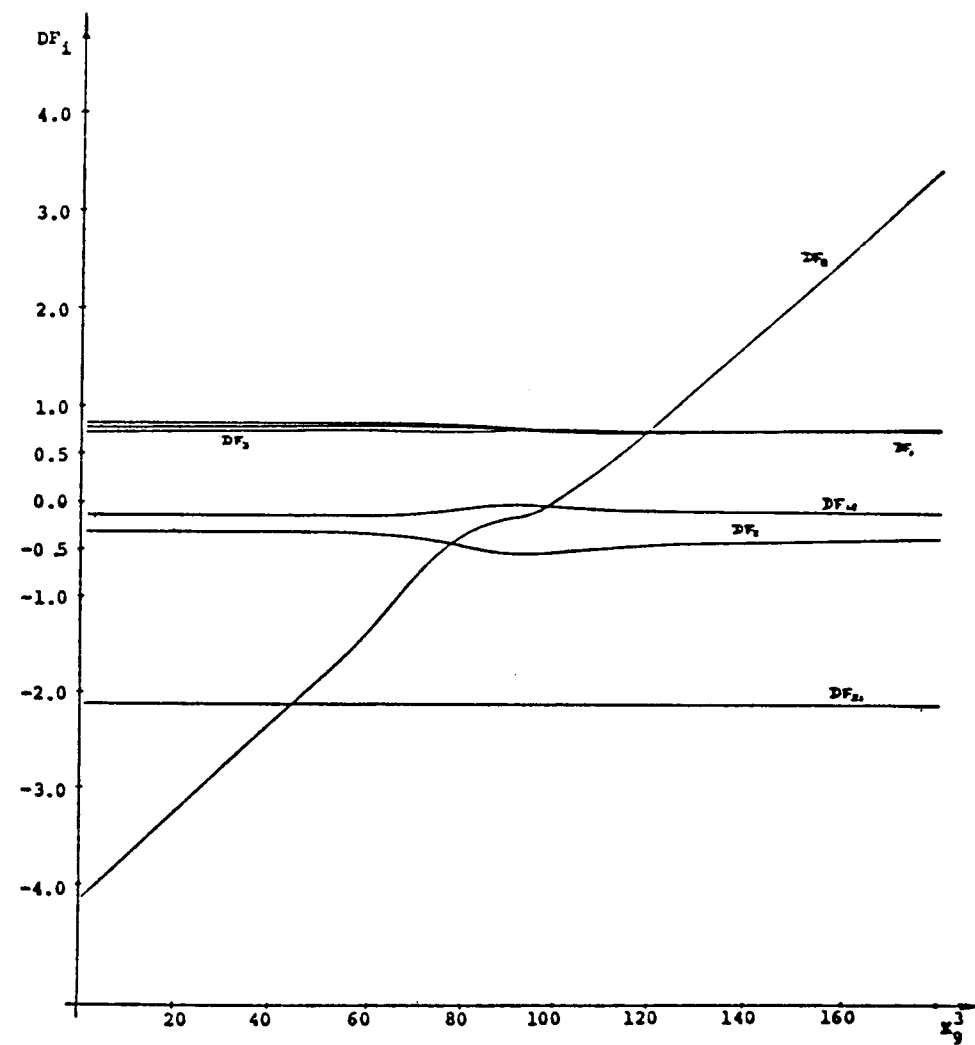


Figura 5.3.22 - $DFFITS_i$ versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=17$.

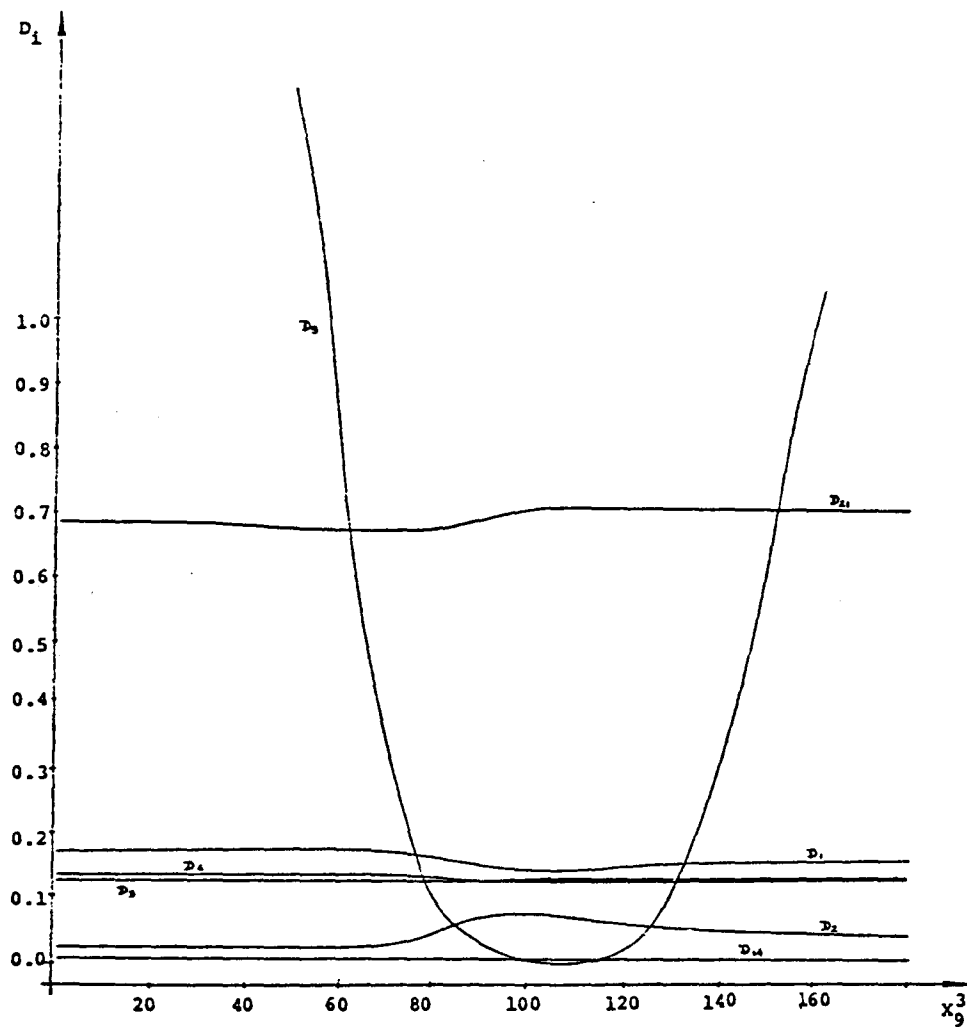


Figura 5.3.23 - D_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=15$.

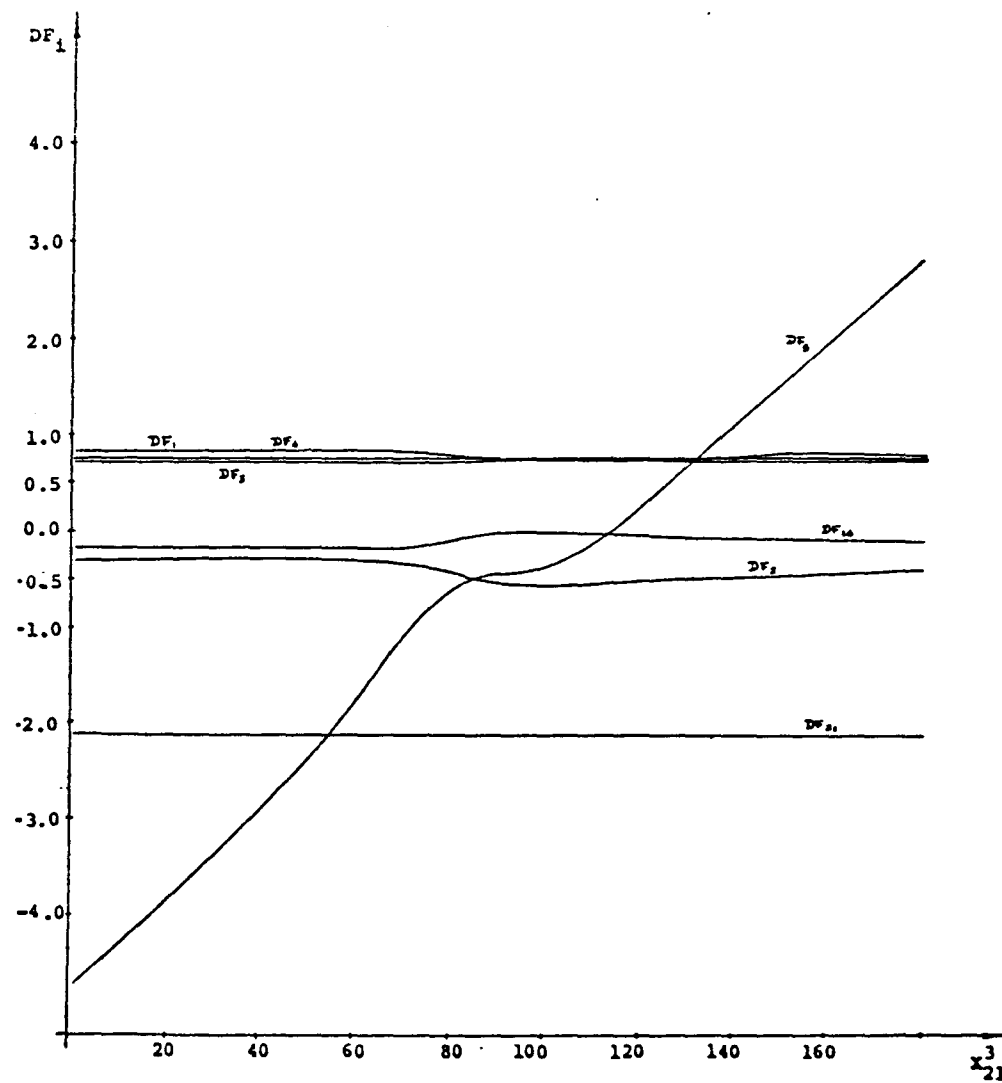


Figura 5.3.24 - $DFFITS_i$ versus X_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=15$.

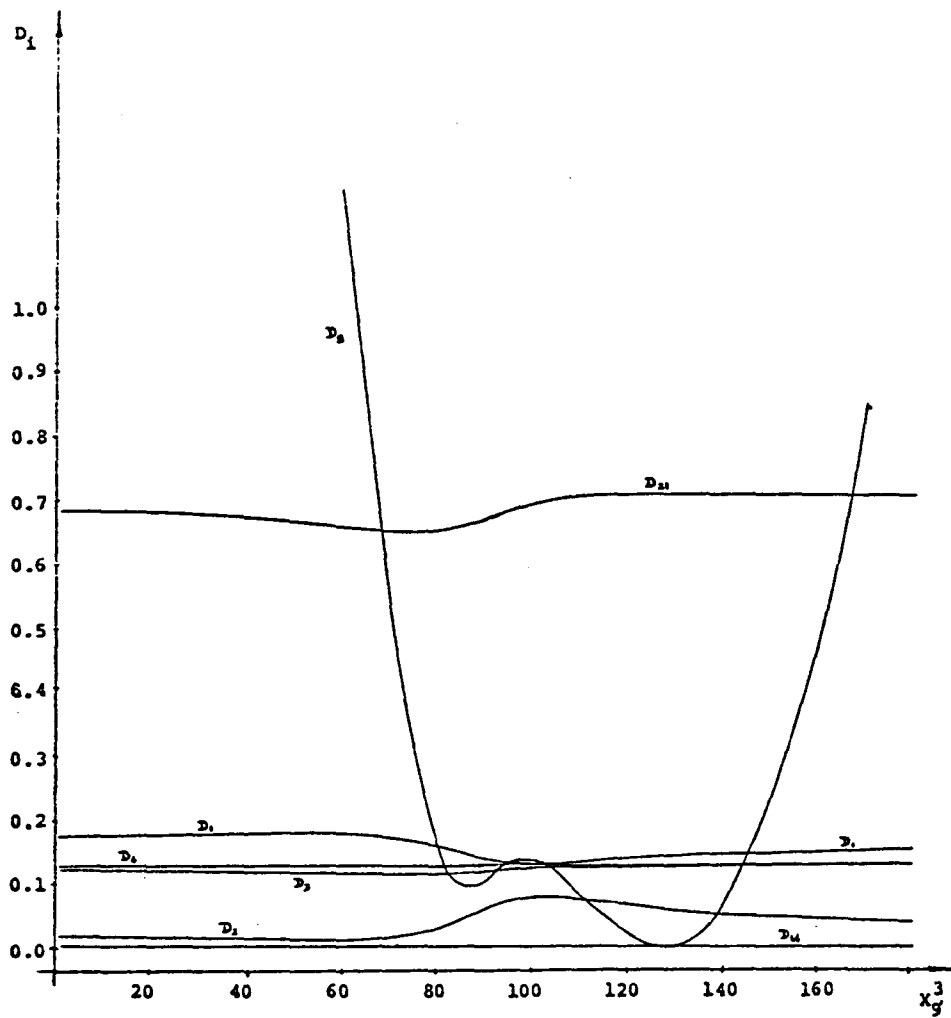


Figura 5.3.25 - D_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=13$.

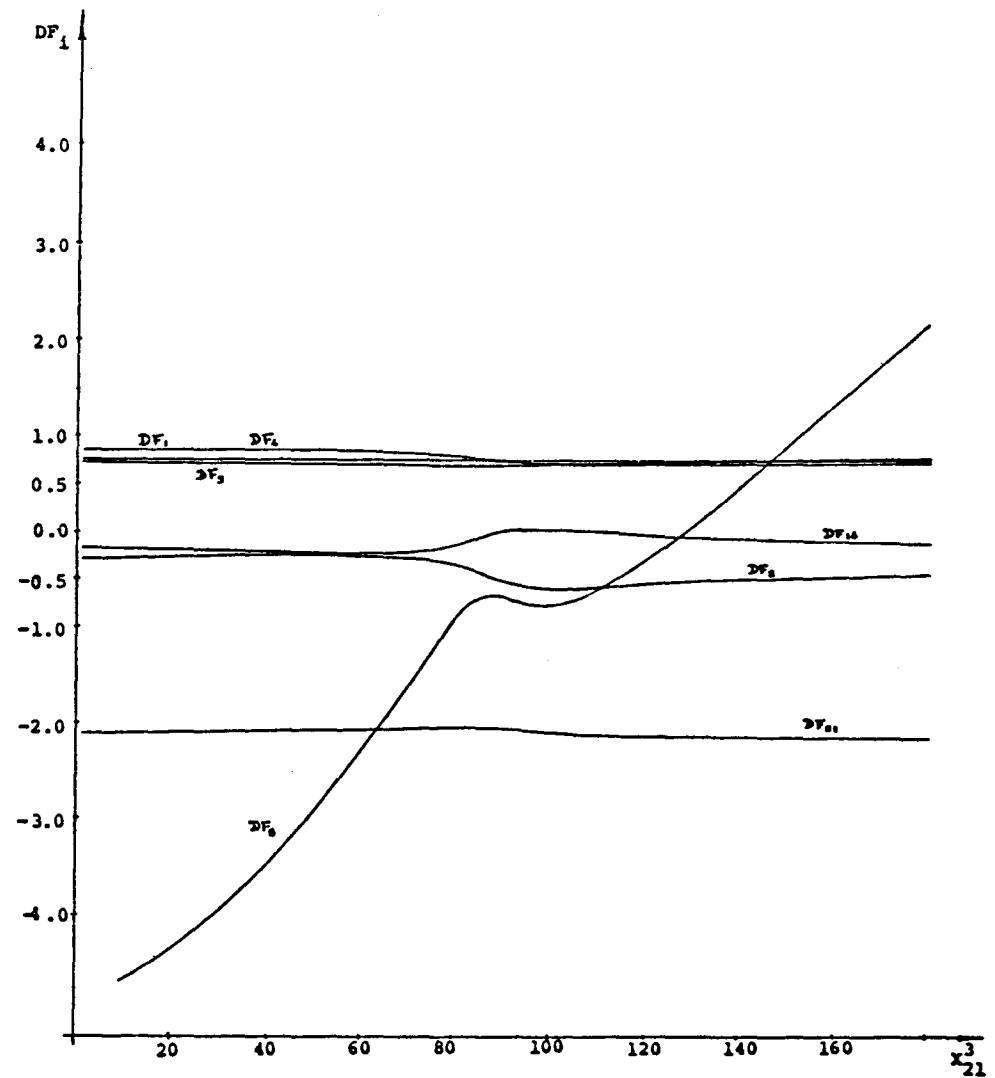


Figura 5.3.26 - $DFFITS_i$ versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=13$.

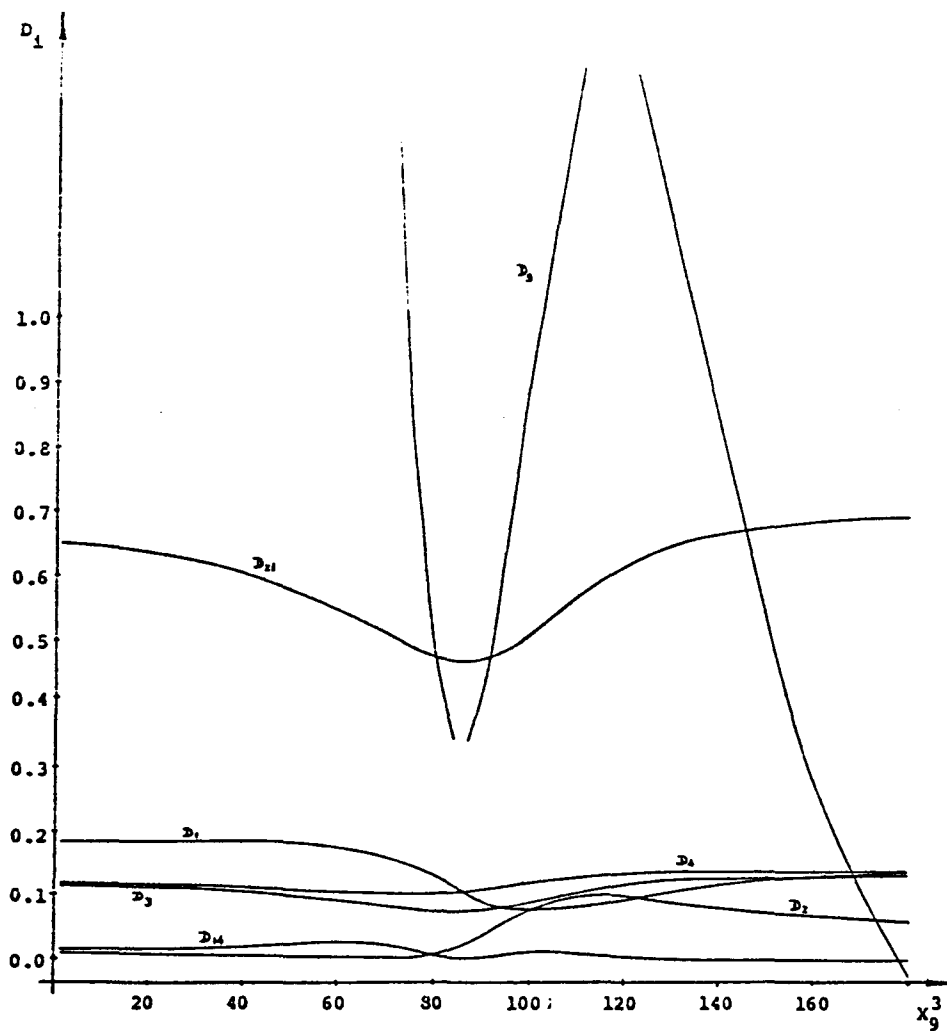


Figura 5.3.27 - D_i versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=5$.

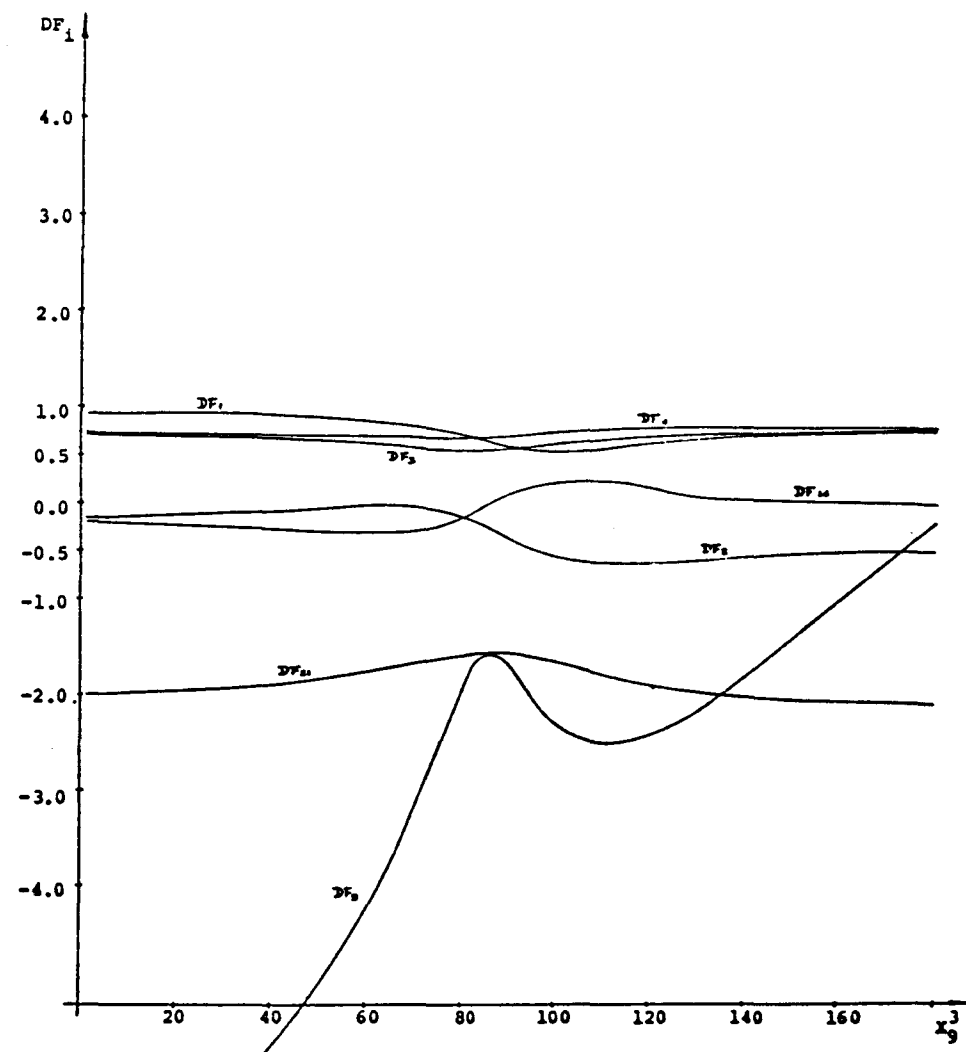


Figura 5.3.28 - $DFFITS_i$ versus x_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=5$.

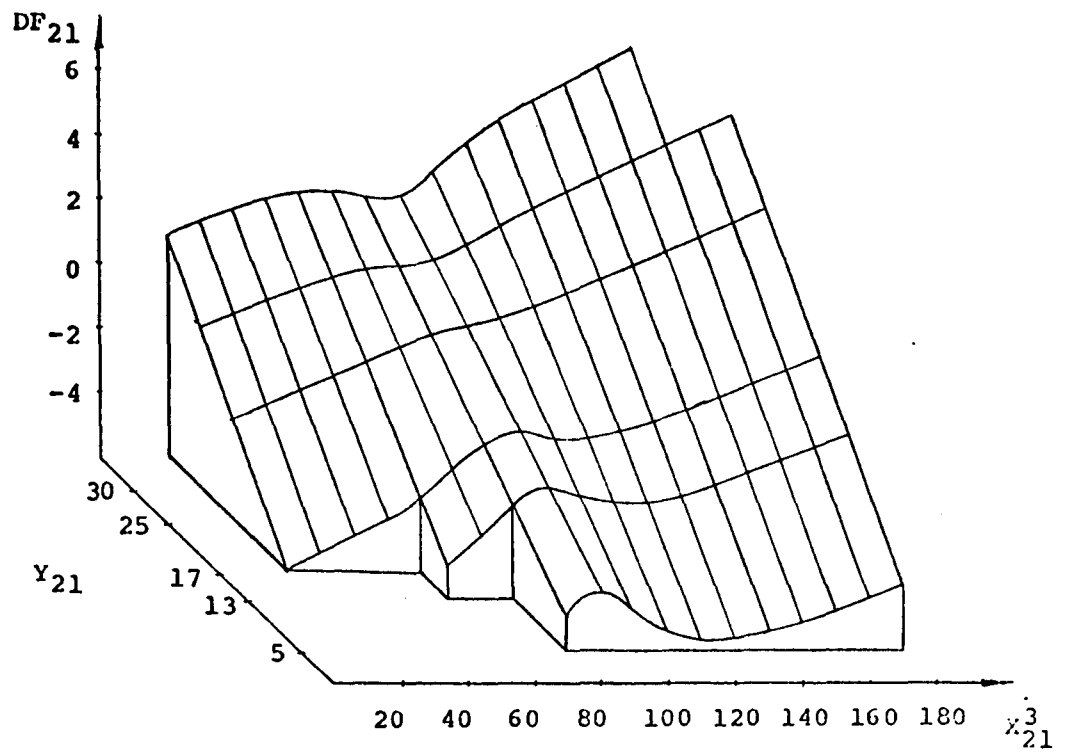


Figura 5.4.1 - DF_{21} como função de Y_{21} e x_{21}^3

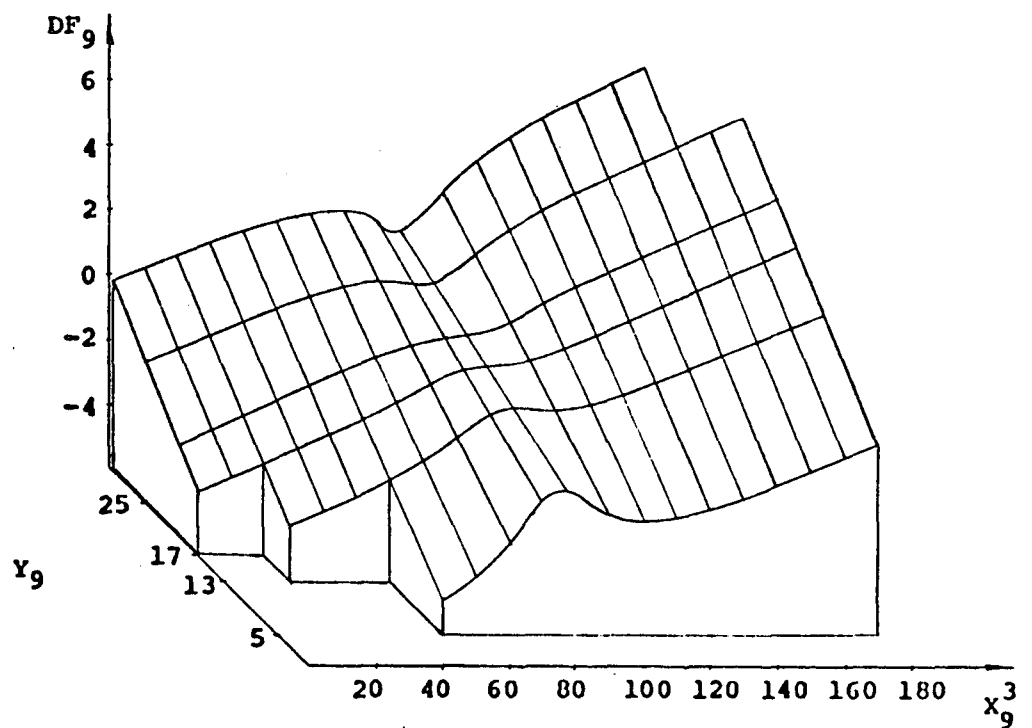


Figura 5.4.2 - DF_9 como função de Y_9 e x_9^3

C A P Í T U L O 6

UMA FORMA ALTERNATIVA DE COMPARAR O DFFITS E O D DE COOK

Na primeira parte deste capítulo trata-se do desempenho do D Raiz, diagnóstico equivalente ao D de Cook, pela forma como é calculado, que facilita a comparação do DFFITS e do D de Cook. Na segunda parte do capítulo apresentam-se as conclusões finais dessa comparação e na terceira o programa aqui usado para o cálculo dos valores dos diagnósticos.

6.1. O D RAIZ.

Usando o valor de comparação para o D de Cook proposto em WEISBERG (1980, pag.108) e analisando o DFFITS sob a forma da equação do D de Cook, com o fim de encontrar um novo valor de comparação para o DFFITS, VELLEMAN and WELSCH (1981) observam que, a raiz quadrada do D de Cook com o sinal do residuo tem aproximadamente o mesmo comportamento do DFFITS.

Tomando como seu valor de comparação, a raiz quadrada do valor de comparação para o D de Cook escolhido por COOK (1977 e 1979), estabeleceu-se aqui que, o D raiz definido pela expressão:

$$DR(i) = \text{sinal}(\hat{\epsilon}(i)) * \{DC(i)\}^{1/2}$$

$$DR(i) = \frac{1}{ps} * \frac{[h(i)]^{1/2}}{[1 - h(i)]} * \hat{\epsilon}(i) \quad (6.1.1)$$

tem desempenho parecido com o desempenho do DFFITS com valor de comparação definido pela expressão (4.7).

O critério de comparação aqui adotado diz que:

$$|DR(i)| > \{ F(p, n-p, 10\%) \}^{1/2}, \quad (6.1.2)$$

indica que o i -ésimo ponto no conjunto de dados é discrepante, e devemos dar muita atenção a ele.

6.1.A. DESEMPENHO DO D RAIZ, $DR(i)$, COMO FUNÇÃO DE $y(i)$.

O gráfico de $DR(i)$ como função de $y(i)$, figuras 6.1.A.1 e 6.1.A.2, com valores de $y(i)$ próximos de seu valor exato é quase uma reta de inclinação positiva, vai declinando com $y(i)$ afastando-se daquele e tende para uma reta constante nos extremos do intervalo $1 < y(i) < 40$. Isso porque, com $y(i)$ próximo do seu valor exato, s^2 (a média quadrática de resíduos para o conjunto completo) aproxima-se de $s^2(i)$, a média para o conjunto sem o i -ésimo ponto, e então, $DR(i)$ é aproximadamente linear mas, com $y(i)$ distante daquele, S cresce rapidamente e $DR(i)$ vai tornando-se constante.

$DR(i)$ como função de $y(i)$ é assintoticamente constante. De fato, isto acontece pois, DC como função de $y(i)$ é assintoticamente constante, como provado na proposição 5.1.3. Este comportamento é observado nos gráficos 6.1.A.1 e 6.1.A.2.

O gráfico de $DR(i)$ versus $y(i)$ mostra que, para afastamentos não muito grandes de $y(i)$ desde seu valor exato, seu desempenho é similar ao de $DF(i)$. Grandes afastamentos de $y(i)$ desde seu valor exato não são detectados por $DR(i)$, tampouco por $DC(i)$, mas sim por $DF(i)$. Faz-se a verificação disto determinando os intervalos onde $DR(i)$ não detecta, ou onde $DR(i)$ detecta, a i -ésima observação como discrepante, e comparando-os com os correspondentes intervalos para $DC(i)$ e os intervalos para $DF(i)$; ver tabelas 5.2.1 e 5.2.2.

$DR(i)$ como função de $y(i)$
 não detecta como discrepante a i -ésima observação

Para	Desde	Até
$i = 9$	8.0	29.5
$i = 21$	20.0	30.5

Tabela 6.1.A.1

DR(k) como função de Y(i); $k = 1, 2, 3, 4, 9, 14, 21$; $k > i$; como pode ser visto nos seus gráficos, figuras 6.1.A.1 e 6.1.A.2, vai tornando-se constante com valores afastados de y(i), e assume esse comportamento definitivamente com valores extremos de y(i). Com valores de y(i) próximos de seu valor exato, DR(i) sofre uma deformação semelhante à experimentada por DF(k) e maior pela variação de y(21) do que pela variação de y(9). Portanto, grandes afastamentos de y(i) desde seu valor exato estabilizam o resto das observações do conjunto.

6.1.B. DESEMPENHO DO D RAIZ COMO FUNÇÃO DE X(3,i).

DR(i) como função de X(3,i) tem comportamento semelhante com o de DF(i). Ver gráficos nas figuras 6.1.B.1 à 6.1.B.6. Seus gráficos apenas se diferenciam em que, a deformação em DR(i) é menor do que em DF(i) e em que, fora do intervalo onde ela ocorre, a inclinação da curva é menor para DR do que para DF. De resto, tudo ocorre como em DF(i): a deformação aparece em DR(i) ao redor do valor original de x(3,i), seu vértice que aponta para cima passa a apontar para baixo, quando y(i) varia desde valores menores até valores maiores que o seu valor exato. E nos extremos do intervalo $1 < x(3,i) < 180$, igual a DF(i), DR(i) aproxima-se a uma reta de inclinação positiva, que parece ser a forma definitiva que toma com valores extremos de x(3,i).

Por todo o anterior, DR(i) e DF(i), como funções de X(3,i), parecem ser equivalentes.

DR(9) como função de x(3,9)
não detecta como discrepante
a nona observação, com
 $1 < x(3,9) < 180$:

Y(9)	Desde	Até
5	160	-
13	77	151
17	66	124
25	18	93
30	-	31
	e 83	86

Tabela 6.1.B.1

DF(9) como função de x(3,9)
não detecta como discrepante
a nona observação, com
 $1 < x(3,9) < 180$:

Y(9)	Desde	Até
5	160	-
13	77	151
17	66	124
25	18	93
30	-	31
	e 83	86

Tabela 6.1.B.2

O comportamento de $DR(k)$ como função de $x(3,i)$; $k = 1,2,3,4, 9,14,21$; $i = 9$ e 21 ; $k <> i$; pelo que em seus gráficos nas figuras 6.1.B.1 à 6.1.B.6 pode-se observar, é quase idêntico com o de $DF(k)$ correspondente. O gráfico de $DR(k)$, como o de $DF(k)$, é quase uma reta constante, tem uma pequena deformação nas proximidades do valor original de $x(3,i)$, que aumenta ou diminui segundo $y(i)$ esteja longe ou perto do seu valor exato. $DR(k)$ como função de $x(3,i)$ parece ser assintoticamente constante.

6.1.C. DESEMPENHO DO D RAIZ PELA VARIAÇÃO CONJUNTA DE $y(i)$ E $x(3,i)$, $i = 21$ OU 9 .

Os gráficos de $DR(21)$ e de $DR(9)$ gerados pela variação conjunta de $y(21)$ e $x(3,21)$ e de $y(9)$ e $x(3,9)$, respectivamente, figuras 6.1.C.1 e 6.1.C.2, são superfícies quase planas, semelhantes às de $DF(21)$ e $DF(9)$ mas, com ondulações menores que as destes.

As ondulações nas superfícies de $DR(i)$ como função de $y(i)$ e $x(3,i)$ apresentam-se quase como nas superfícies correspondentes de DF . Percorrendo $y(i)$, desde valores menores até valores maiores que seu valor exato, elas mudam de assimétricas à direita para assimétricas à esquerda, seu vértice muda o sentido de cima para abaixo, e quando $y(i)$ coincide com seu valor exato elas se anulam. Mas, quando $y(i)$ aproxima-se dos extremos do intervalo $1 < y(i) < 40$, vão tornando-se constantes.

6.2. CONCLUSÕES FINAIS.

Em resumo, DR (ou DC) e o DF são praticamente equivalentes na detecção de possíveis problemas causados por afastamentos grandes em um dos valores de uma das variáveis preditoras, $x(i, j)$. Já o DF apresenta a grande vantagem de que, além disso, detecta situações potencialmente sérias advindas de grandes afastamentos num valor da resposta $y(i)$ de forma muito mais eficaz do que DR (ou DC).

Com os resultados apresentados, principalmente a comparação das proposições 5.2.1 e 5.2.3 e das proposições 5.2.2 e 5.2.4 e o exame das figuras, principalmente 5.4.1, 5.4.2, 6.1.C.1 e 6.1.C.2 pode-se concluir que:

- 1- Não há indicação alguma de haver qualquer razão para se usar simultaneamente DF e DC ou DF e DR como elementos de diagnóstico.
- 2- Deve-se preferir DF por sua capacidade de detectar problemas potenciais não só pelo afastamento nos x 's mas também em valores dos y 's.

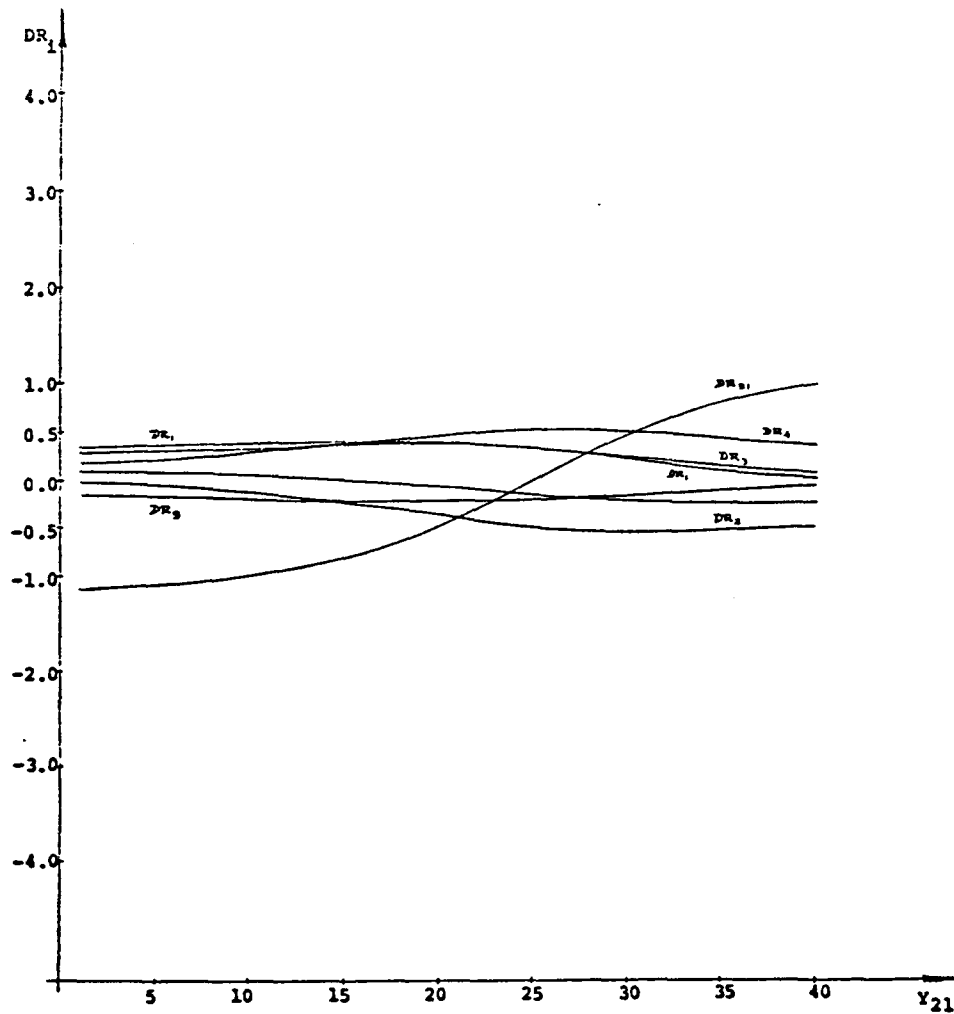


Figura 6.1.A.1 - DR_i versus Y_{21} ; $i=1,2,3,4,9,14,21$.

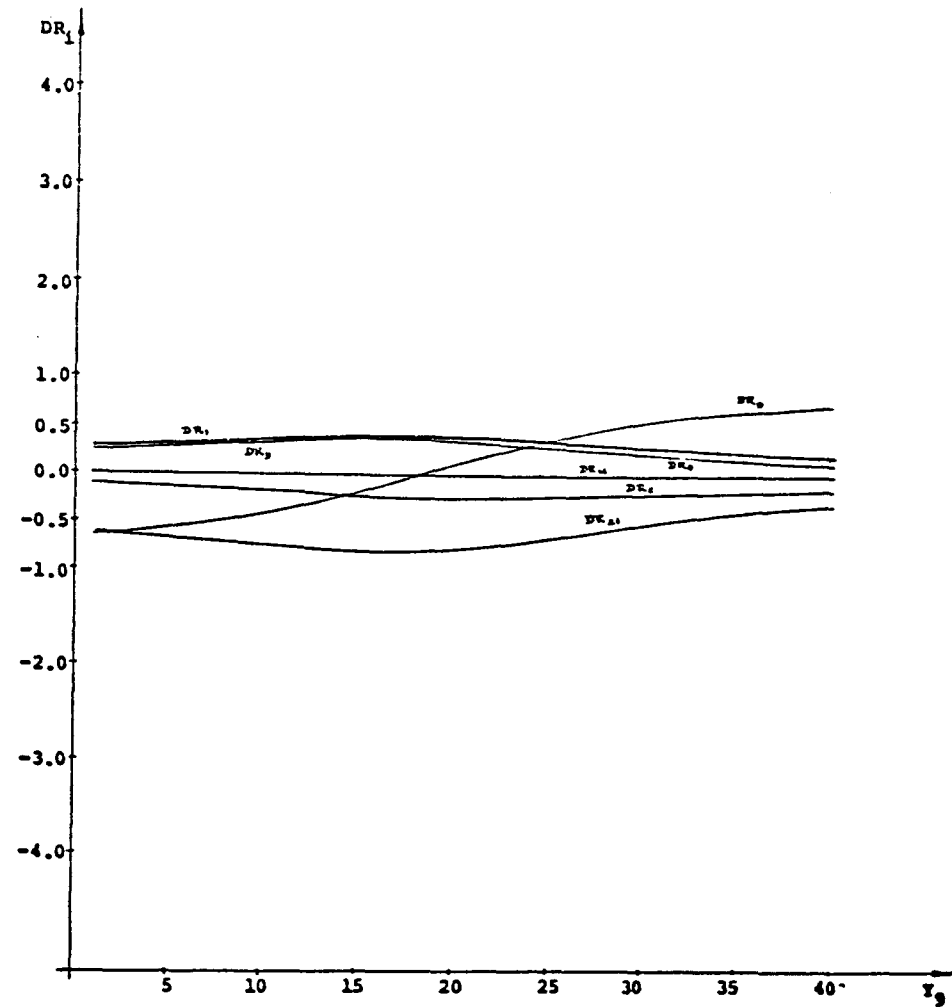


Figura 6.1.A.2 - DR_i versus Y_9 ; $i=1,2,3,4,9,14,21$.

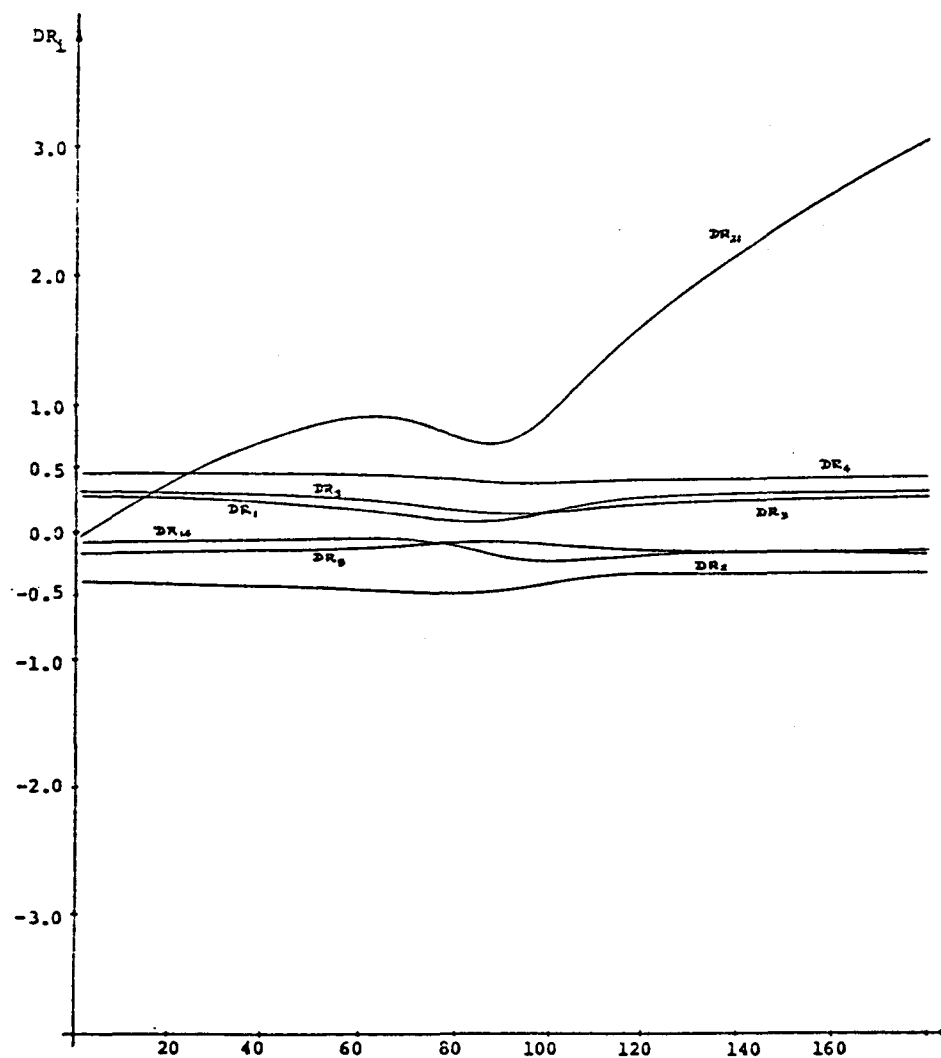


Figura 6.1.B.1 - DR_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=35$.

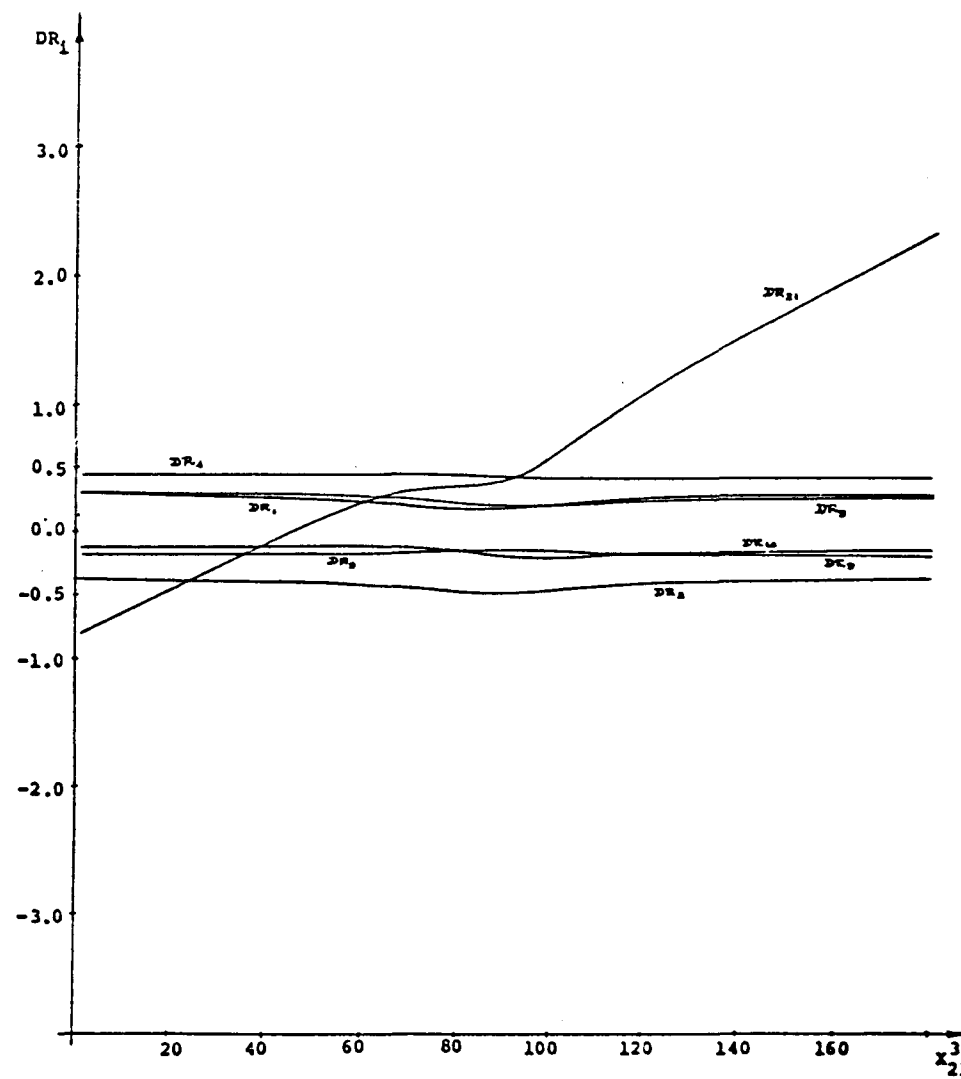


Figura 6.1.B.2 - DR_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=30$.

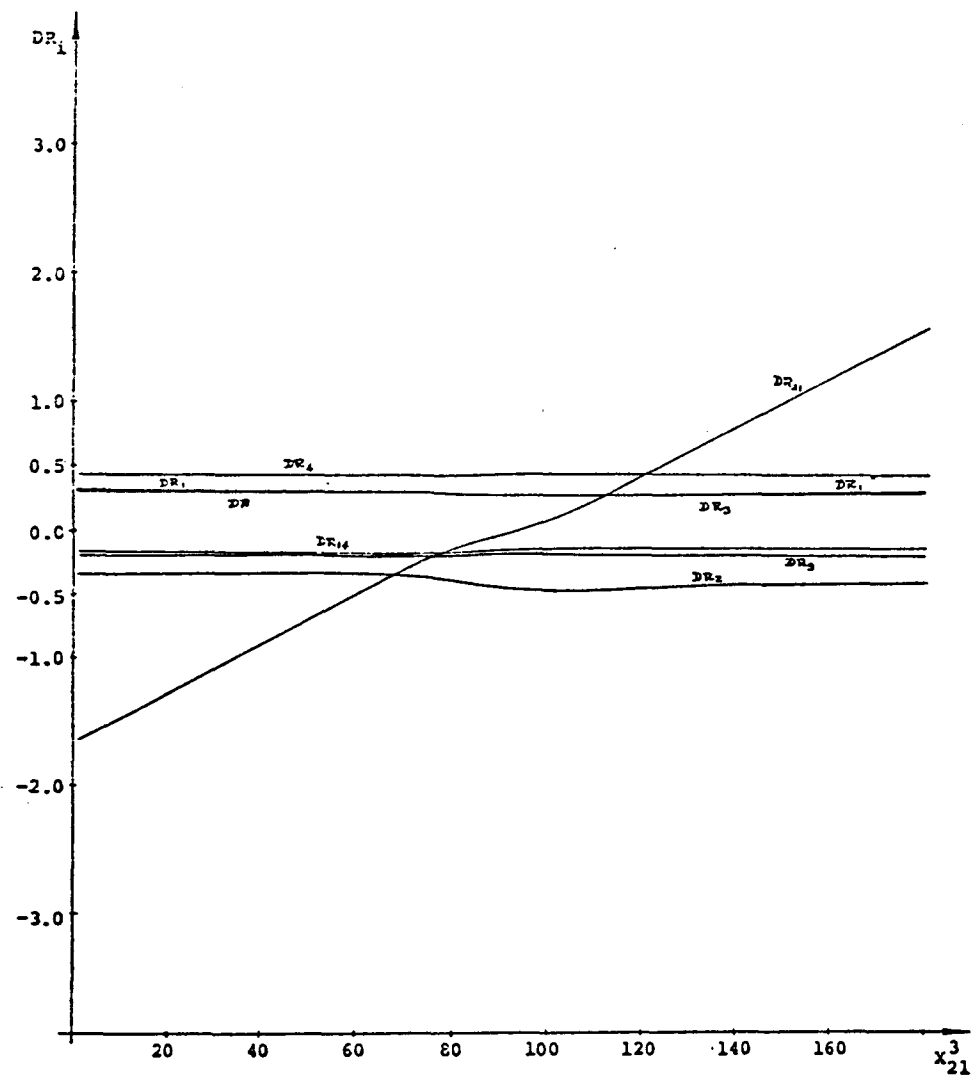


Figura 6.1.B.3 - DR_i versus X_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=25$.

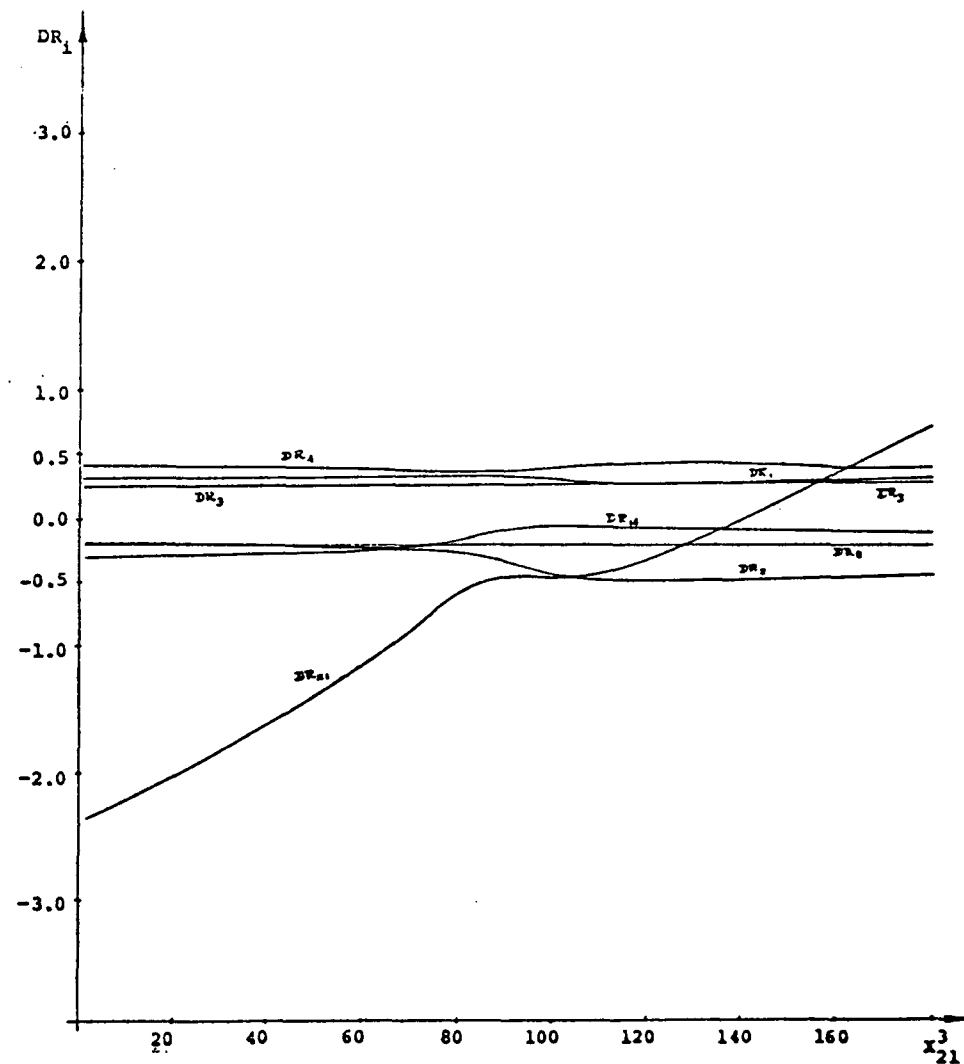


Figura 6.1.B.4 - DR_i versus X_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=20$.

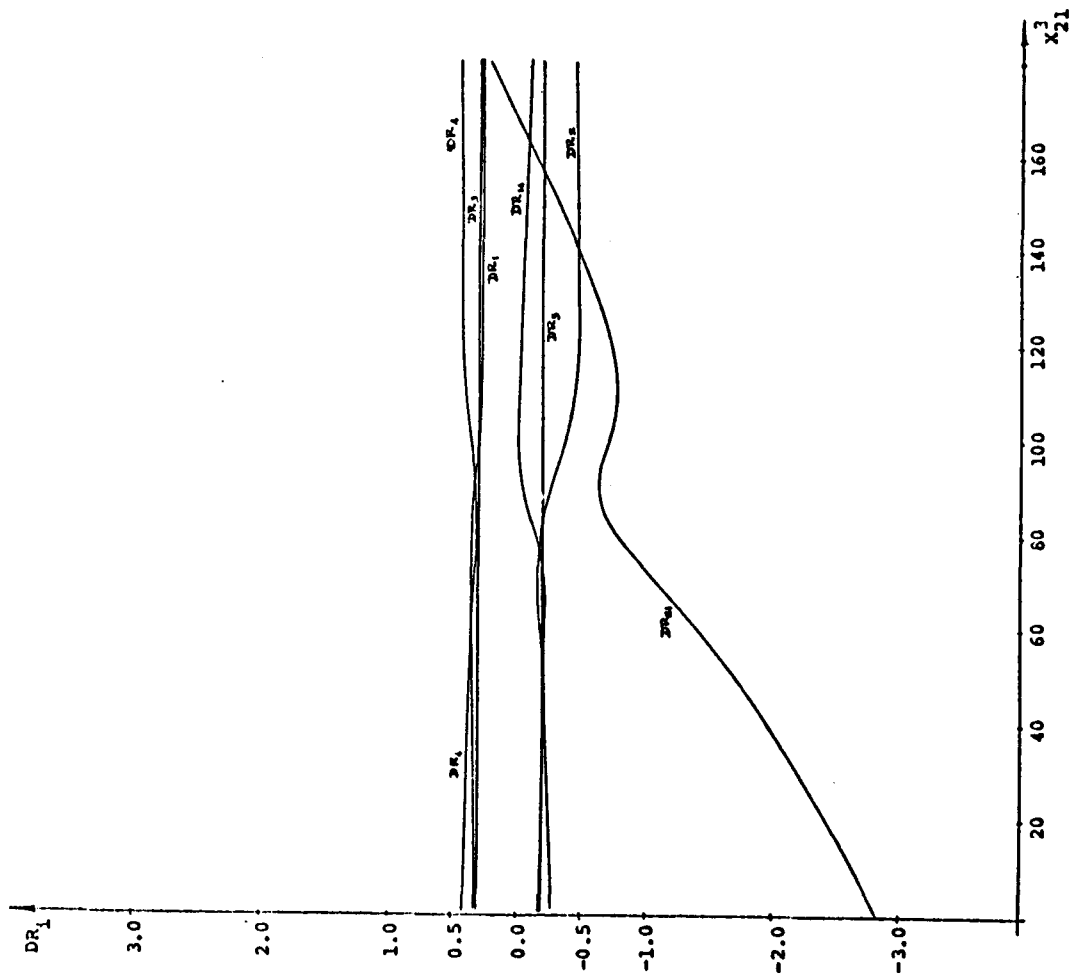


Figura 6.1.B.5 - DR_1 versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=17$.

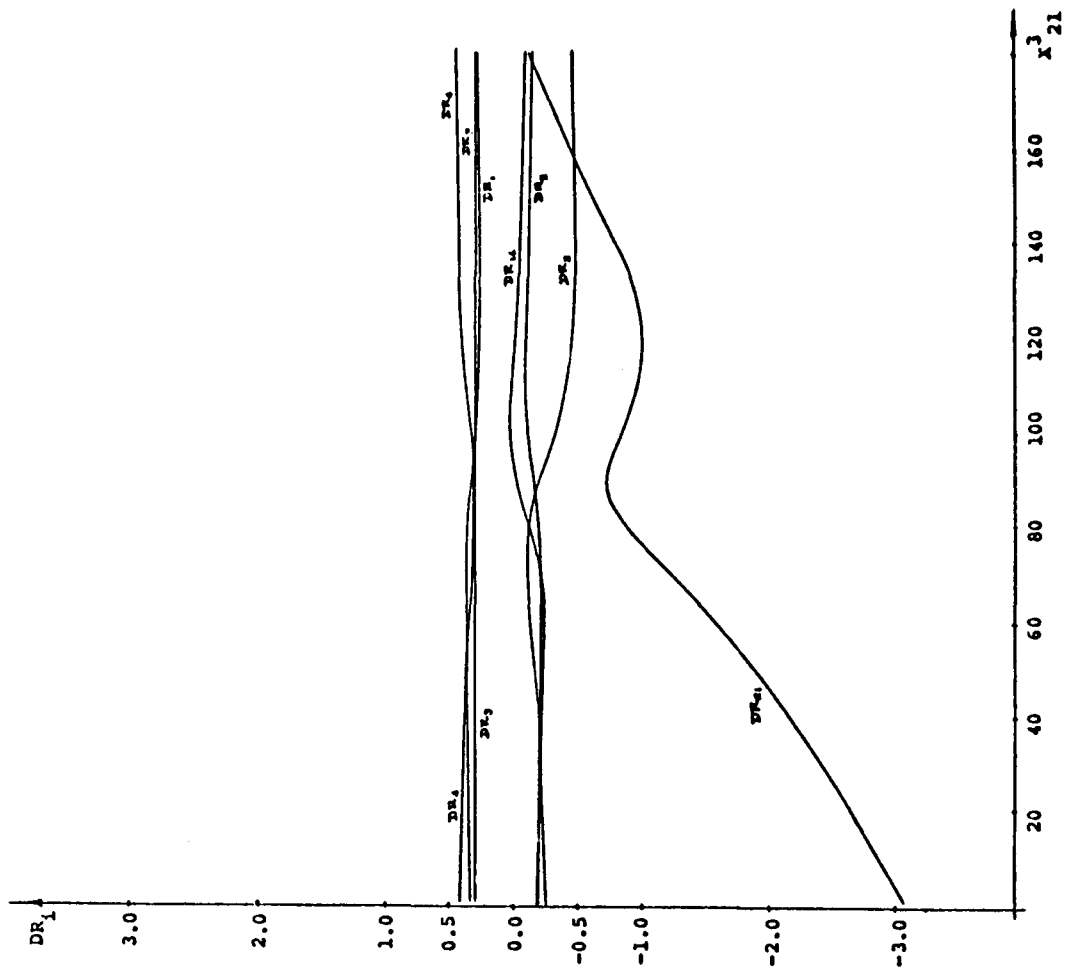


Figura 6.1.B.6 - DR_1 versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=15$.

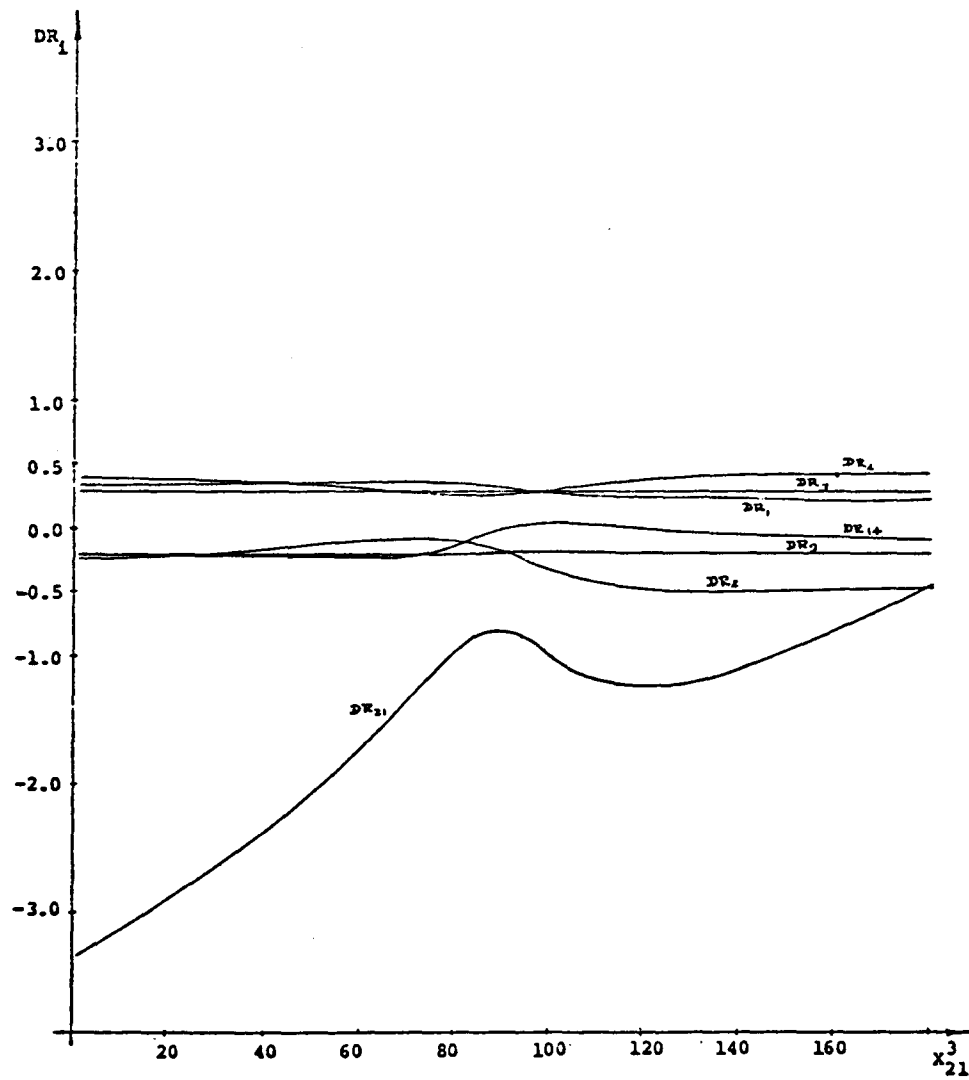


Figura 6.1.B.7 - DR_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=13$.

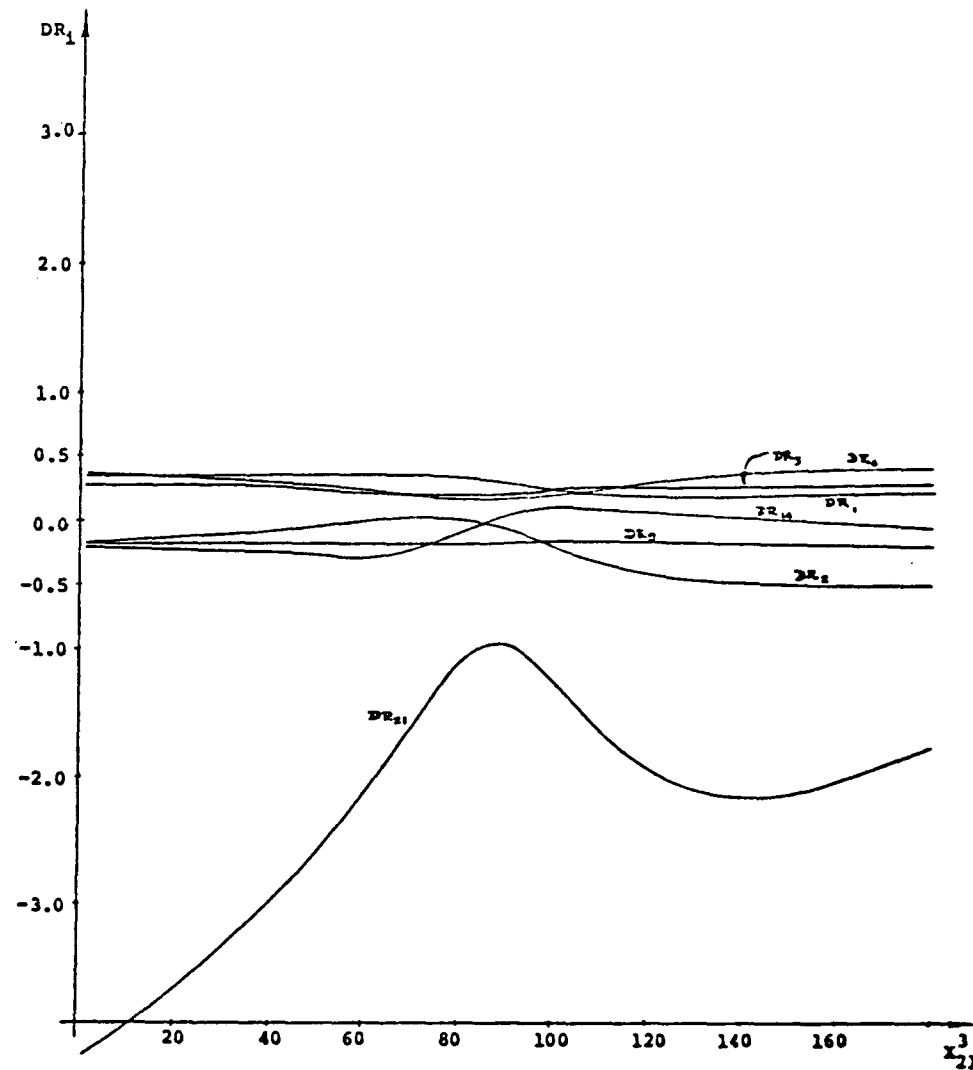


Figura 6.1.B.8 - DR_i versus x_{21}^3 ; $i=1,2,3,4,9,14,21$; para $Y_{21}=5$.

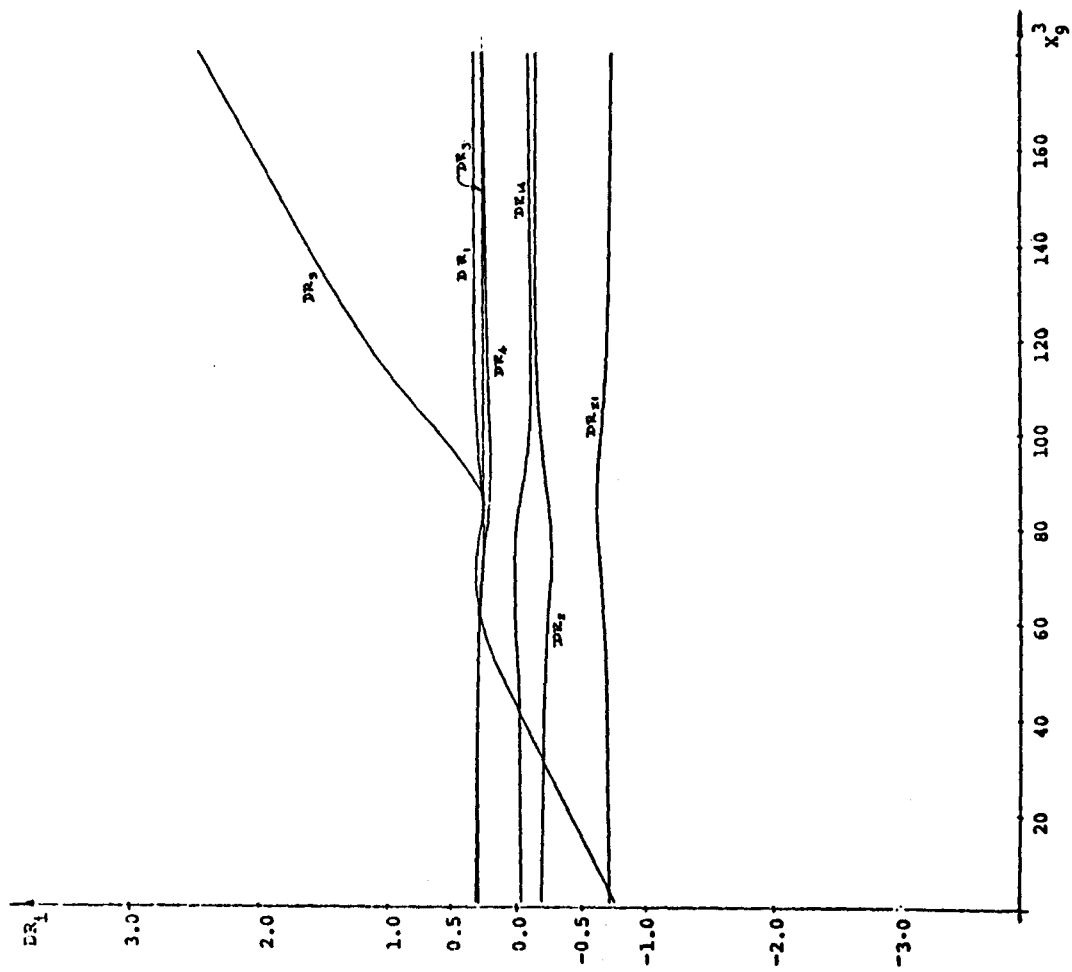


Figura 6.1.B.9 - DR_1 versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=25$.

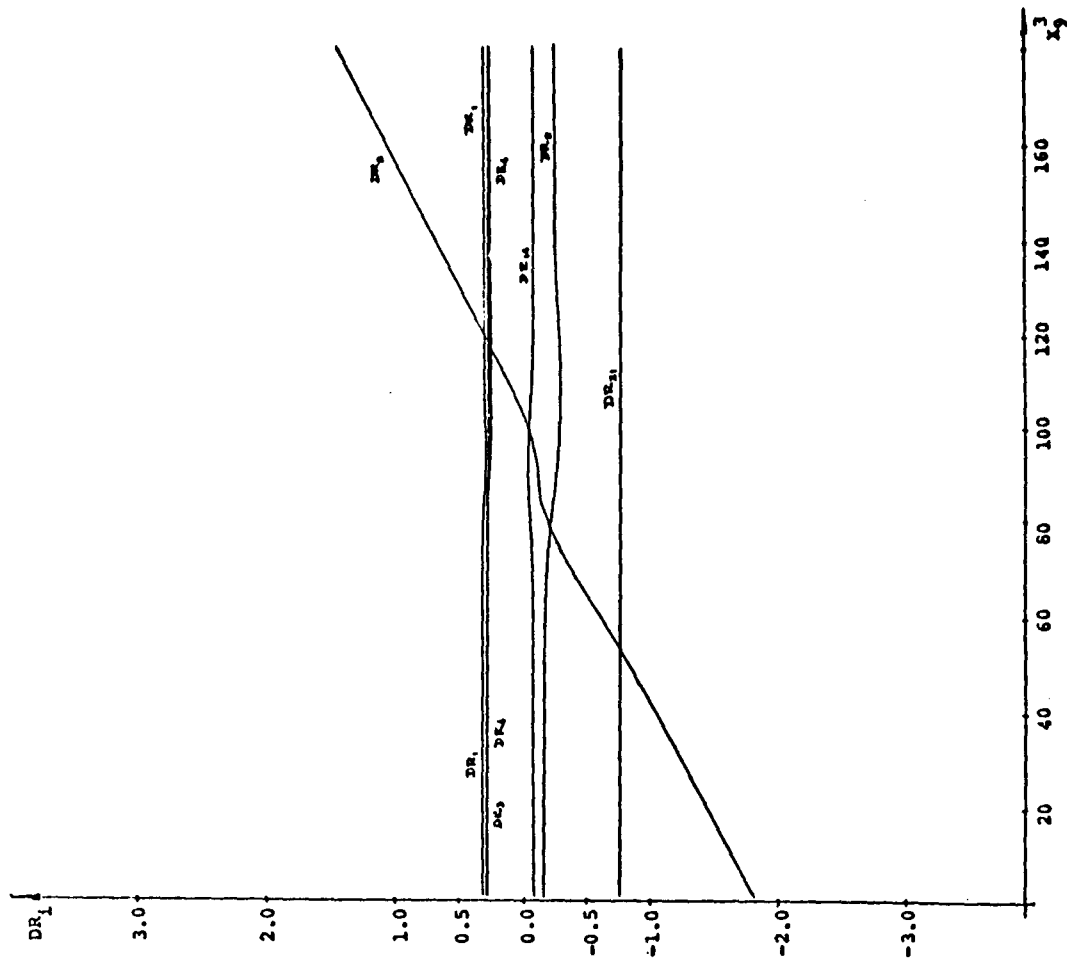


Figura 6.1.B.10 - DR_1 versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=17$.

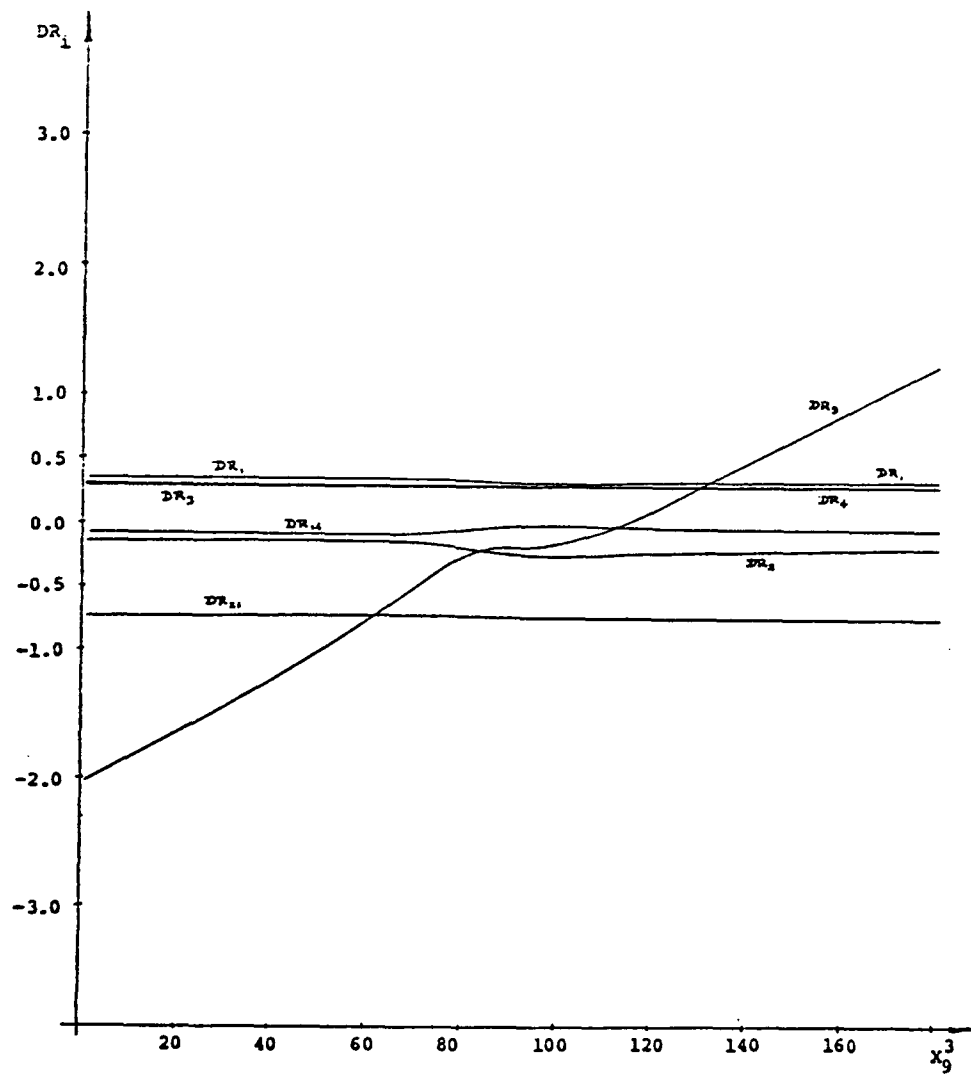


Figura 6.1.B.11 - DR_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=15$.

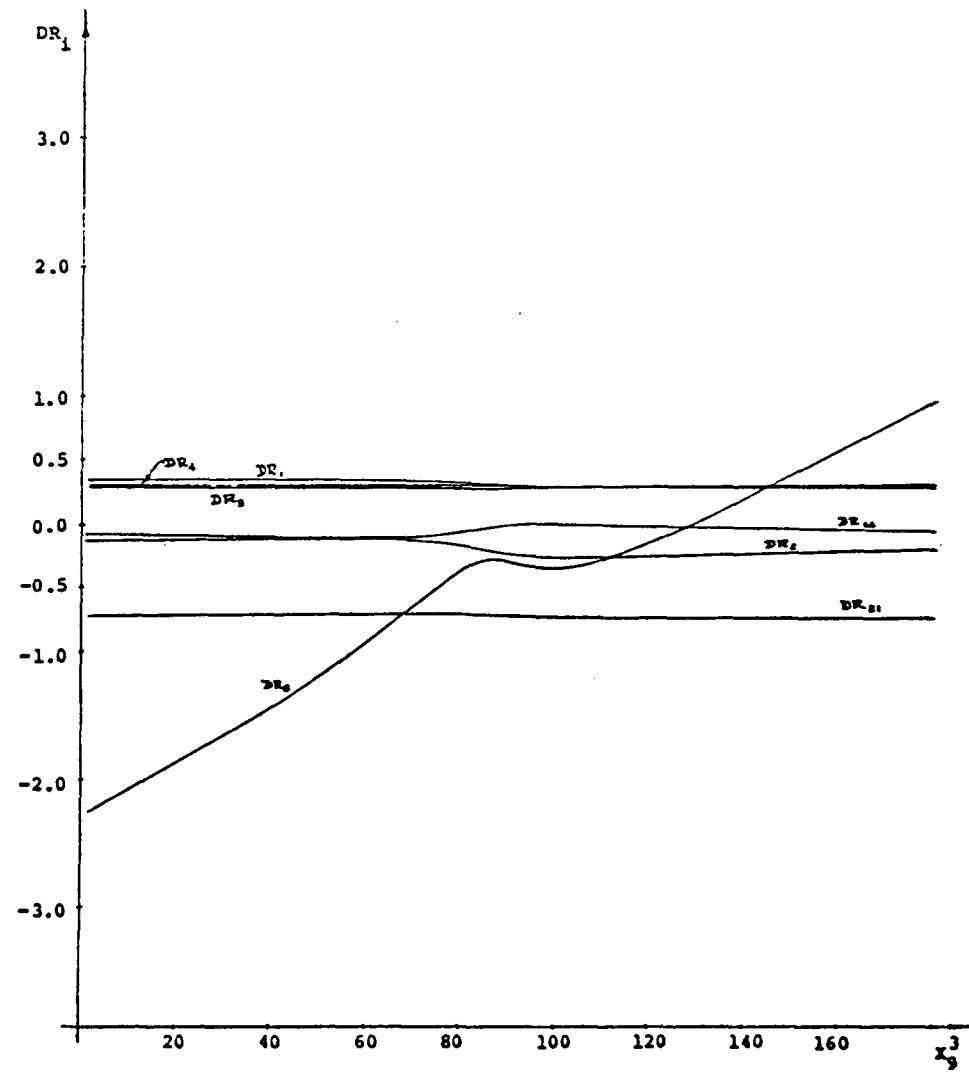


Figura 6.1.B.12 - DR_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=13$.

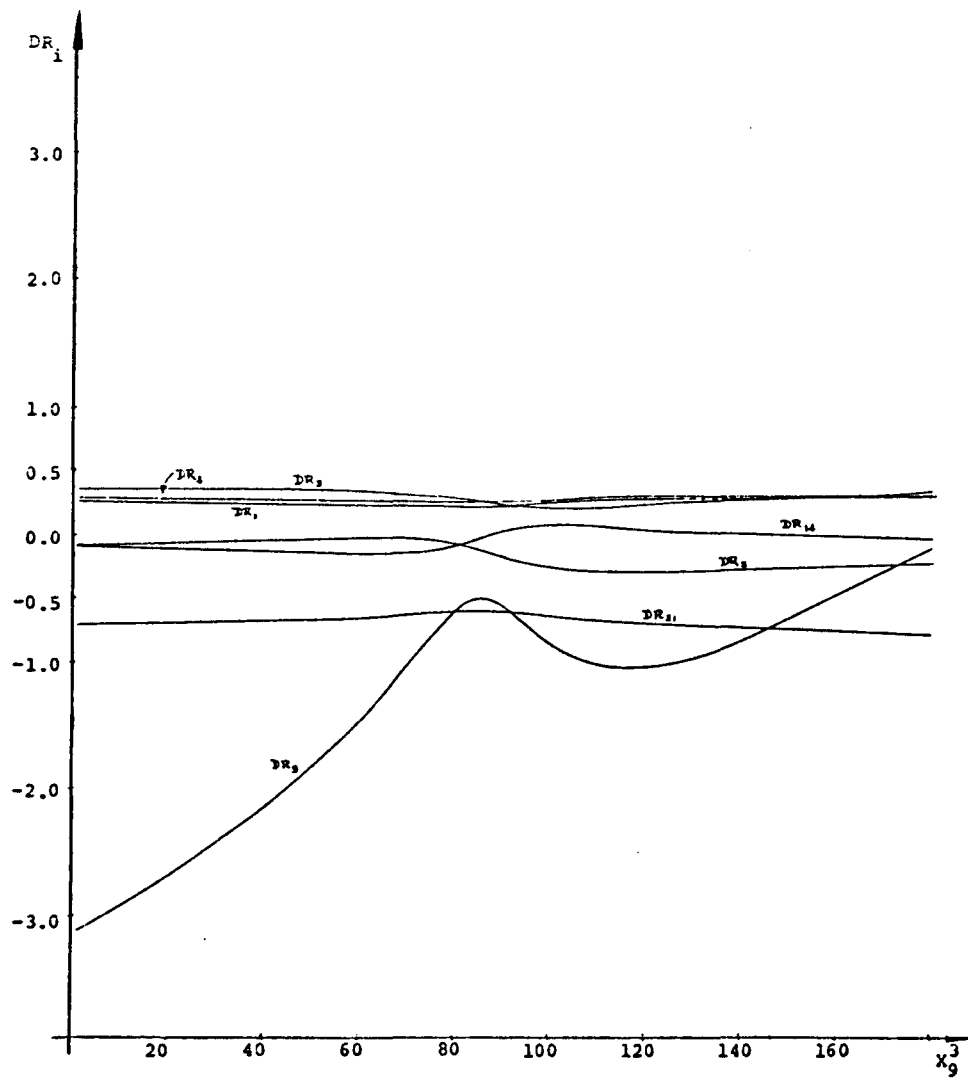


Figura 6.1.B.13 - DR_i versus X_9^3 ; $i=1,2,3,4,9,14,21$; para $Y_9=5$.

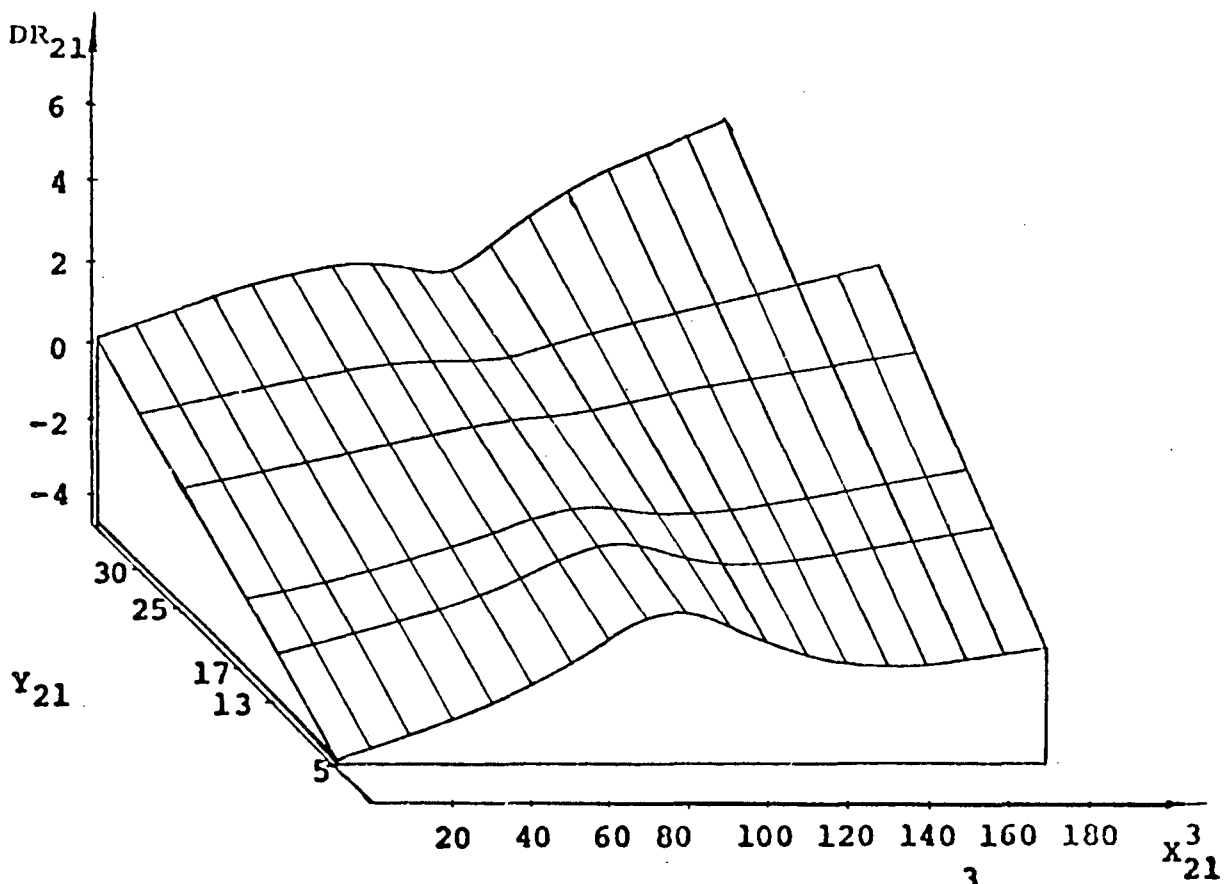


Figura 6.1.C.1 - DR_{21} como função de Y_{21} e X_{21}^3

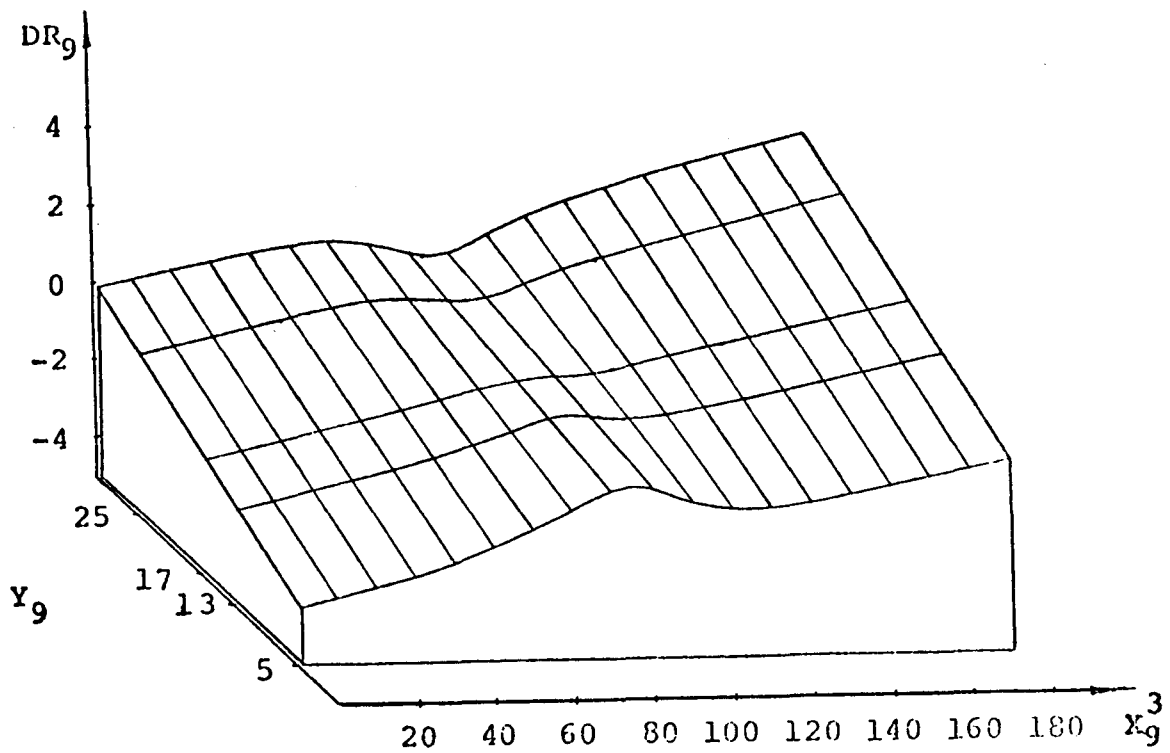


Figura 6.1.C.2 - DR_9 como função de Y_9 e X_9^3

6.2. PROGRAMA PARA CÁLCULO DE DIAGNÓSTICOS.

Descreve-se e apresenta-se aqui, o programa usado para o cálculo do modelo ajustado e dos diagnósticos: diagonal da matriz chapéu, D de Cook, D raiz e DFFITS correspondentes aos diferentes conjuntos de dados considerados no desenvolvimento deste trabalho. O programa foi construído usando o Microsoft Basic; versão BASIC-80, Rev. 5.21.

A utilização deste programa é fácil, têm requerimentos mínimos. A obtenção do modelo ajustado e dos valores dos diagnósticos através dele é rápida, e com boa precisão (os resultados finais têm quatro cifras significativas).

A execução do programa exige apenas dar entrada dos números de linhas menos um (IL) e de colunas menos um (IC), da matriz de delineamento, dos valores da variável resposta $Y(I)$, dos valores das variáveis preditoras $X(I,J)$, começando pela primeira e em diante, e ajustar as dimensões aos vetores e matrizes calculados ao longo do programa.

Os vetores e matrizes de maior interesse, aparecem declarados no programa como segue:

XX é a matriz produto da matriz X e sua transposta e depois a matriz inversa de XX.

Z é o vetor de dos parâmetros do modelo ajustado.

H é o vetor dos elementos da diagonal da matriz chapéu.

YAJ é o vetor ajustado de respostas.

RE é o vetor de resíduos.

DR é o vetor de valores do D raiz.

DC é o vetor de valores do D de Cook.

SE(2) é o vetor de valores da média quadrática de resíduos para o conjunto sem a i-ésima observação.

DF é o vetor de valores do DFFITS.

Talvez seja de interesse observar que a matriz inversa do produto da matriz X com sua transposta é calculada usando o operador "Sweep", (GOODNIGHT, 1979), com o fim diminuir a quantidade de trabalho computacional e melhorar precisão dos cálculos.

Tendo dado entrada dos dados e comandado a execução do programa, pode-se verificar através da tela, ou da impressora, que os dados introduzidos estejam corretos, para então prosseguir com a execução do programa. A impressão dos resultados é feita em sequência horizontal, com quatro resultados por linha, sempre precedidos do número de ordem, menos um, que a observação correspondente tem no conjunto.

O programa listado a seguir é, pela sua vez, um exemplo da sua utilização. Ele contém como dados as dimensões, menos um, da matriz de delineamento: 20 e 3, o conjunto de observações descrito na tabela 5.1.1 e as dimensões ajustadas às matrizes e vetores a serem calculados. Aparecem, depois os resultados obtidos para esse conjunto.

```

1000 REM *** PROGRAMA PARA O CALCULO DO MODELO DE REGRESSAO AJUSTADO E DOS DIA
GNOSTICOS H(I), D DE COOK, D RAIZ E DFFITS ***
1010 REM *** Entrar o numero de linhas menos hum e de colunas menos hum da matr
iz de delineamento ***
1020 DATA 20,3
1030 READ IL,IC
1040 REM *** Entrar os valore da variavel resposta ***
1050 DATA 42,37,37,28,18,18,19,20,15,14,14,13,11,12,8,7,8,8,9,15,15
1060 REM *** Entrar os valores do primeiro preditor ***
1070 DATA 80,80,75,62,62,62,62,62,58,58,58,58,58,50,50,50,50,50,56,70
1080 REM *** Entrar os valores do segundo preditor ***
1090 DATA 27,27,25,24,22,23,24,24,23,18,18,17,18,19,18,19,20,20,20
1100 REM *** Entrar os valores do terceiro preditor ***
1110 DATA 89,88,90,87,87,87,93,93,87,80,89,88,82,93,89,86,72,79,80,82,91
1120 REM *** Entrar as dimensoes do vetores e matrizes a serem calculados ***
1130 DIM X(50,20),Y(50),XT(20,50),XX(20,20),W(20,50),H(50),VT(20)
1140 DIM YAJ(50),RE(50),DC(50),DR(50),SE2(50),DFT(50)
1150 FOR I=0 TO IL
1160   READ Y(I)
1170 NEXT I
1180 FOR I=0 TO IL
1190   X(I,0)=1
1200   XT(0,I)=1
1210 NEXT I
1220 REM *** Calculo da matriz transposta ***
1230 FOR J=1 TO IC
1240   FOR I=0 TO IL
1250     READ X(I,J)
1260     XT(J,I)=X(I,J)
1270   NEXT I
1280 NEXT J
1290 REM *** Apresentacao dos dados ***
1300 FOR I=0 TO IL
1310   FOR J=1 TO IC
1320     PRINT USING "##### ";X(I,J);
1330   NEXT J
1340 PRINT USING "\ \##### "; "y= ";Y(I);
1350 NEXT I
1360 CARCS = INPUT$(1)
1370 REM *** Apresentacao da matriz transposta de X ***
1380 FOR I=0 TO IC
1390   FOR J=0 TO IL
1400     PRINT XT(I,J)
1410   NEXT J
1420 NEXT I
1430 FOR I=0 TO IC
1440   FOR K=0 TO IC
1450     FOR J=0 TO IL
1460       XX(I,K)=XX(I,K)+XT(I,J)*X(J,K)
1470     NEXT J

```

```

1480     PRINT XX(I,K)
1490     NEXT K
1500     NEXT I
1510     REM *** Calculo da matriz inversa de X'X ***
1520     FOR K=0 TO IC
1530         GOSUB 1610
1540     NEXT K
1550     FOR I=0 TO IC
1560         FOR J=0 TO IC
1570             PRINT XX(I,J)
1580         NEXT J
1590     NEXT I
1600     GOTO 1760
1610     D=XX(K,K)
1620     IF D<1E-10 THEN GOTO 3300
1630     FOR J=0 TO IC+1
1640         XX(K,J)=XX(K,J)/D
1650     NEXT J
1660     FOR I =0 TO IC+1
1670         IF I=K THEN 1740
1680         B=XX(I,K)
1690         FOR J=0 TO IC+1
1700             XX(I,J)=XX(I,J)-B*XX(K,J)
1710         NEXT J
1720         XX(I,K)=-B/D
1730         XX(K,K)=1/D
1740     NEXT I
1750     RETURN
1760     TOT=0
1770     FOR I=0 TO IC
1780         FOR K=0 TO IL
1790             FOR J=0 TO IC
1800                 TOT=TOT+XX(I,J)*XT(J,K)
1810             NEXT J
1820             W(I,K)=TOT
1830             PRINT W(I,K)
1840             TOT=0
1850         NEXT K
1860     NEXT I
1870     REM *** Calculo dos parametros ajustados ***
1880     TOT1=0
1890     FOR K=0 TO IC
1900         FOR I=0 TO IL
1910             TOT1=TOT1+W(K,I)*Y(I)
1920         NEXT I
1930         Z(K)=TOT1
1940         PRINT Z(K)
1950         TOT1=0
1960     NEXT K

```

```

1970 LPRINT "      VALORES DOS PARAMETROS AJUSTAD
O S"
1980 LPRINT
1990 REM *** Apresentacao dos parametros ajustados ***
2000 FOR I=0 TO IC
2010 LPRINT USING "  ## #####.#### " ; I ; Z(I) ;
2020 NEXT I
2030 LPRINT "      DIAGONAL DA MATRIZ CHAPEU"
2040 LPRINT
2050 LPRINT "      I      H(I)      I      H(I)      I      H(I)      I
H(I)"
2060 LPRINT
2070 TOT2=0
2080 FOR I=0 TO IL
2090   FOR K=0 TO IC
2100     TOT2=TOT2+X(I,K)*W(K,I)
2110   NEXT K
2120 H(I)=TOT2
2130 IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 2160
2140 LPRINT USING "  ## #####.#### " ; I-3 ; H(I-3) ; I-2 ; H(I-2) ; I-1 ; H(I-1) ; I ; H(I)
2150 PRINT
2160 TOT2=0
2170 NEXT I
2180 INTI = FIX((IL+1)/4)
2190 II=IL-4*INTI+1
2200   IF II=0 GOTO 2240
2210   FOR I=1 TO II
2220     LPRINT USING "  ## #####.#### " ; IL-II+I ; H(IL-II+I) ;
2230   NEXT I
2240   PRINT
2250 LPRINT
2260 LPRINT "      VALORES AJUSTADOS DE Y"
2270 LPRINT
2280 LPRINT "      I      YAJ(I)      I      YAJ(I)      I      YAJ(I)      I
YAJ(I)"
2290 LPRINT
2300 TOT3=0
2310 FOR I=0 TO IL
2320   FOR J=0 TO IC
2330     TOT3=TOT3+X(I,J)*Z(J)
2340   NEXT J
2350 YAJ(I)=TOT3
2360 IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 2390
2370 LPRINT USING "  ## #####.#### " ; I-3 ; YAJ(I-3) ; I-2 ; YAJ(I-2) ; I-1 ; YAJ(I-1)
; I ; YAJ(I)
2380 PRINT
2390 TOT3=0
2400 NEXT I
2410   IF II=0 GOTO 2450
2420   FOR I=0 TO II
2430     LPRINT USING "  ## #####.#### " ; IL-II+I ; YAJ(IL-II+I)
2440   NEXT I
2450 LPRINT

```

```

2460 LPRINT "                                R E S I D U O S      RE(I)"
2470 LPRINT
2480 LPRINT "          I      RE(I)          I      RE(I)          I      RE(I)          I
      RE(I)"
2490 LPRINT
2500 FOR I=0 TO IL
2510   RE(I)=Y(I)-YAJ(I)
2520   IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 2550
2530   LPRINT USING "      ## ####.####      ";I-3;RE(I-3);I-2;RE(I-2);I-1;RE(I-1)
;I;RE(I)
2540   PRINT
2550 NEXT I
2560   IF II=0 GOTO 2610
2570   FOR I=0 TO II
2580     LPRINT USING "      ## ####.####      ";IL-II+I;RE(IL-II+I)
2590     PRINT
2600   NEXT I
2610 SOMAT=0
2620 FOR I=0 TO IL
2630   SOMAT=SOMAT+(RE(I)^2)
2640 NEXT I
2650 S2=SOMAT/(IL-IC)
2660 PRINT S2
2670 PRINT
2680 LPRINT "                                V A L O R E S   D O   D   D E   C O O K   "
2690 LPRINT
2700 LPRINT "          I      DC(I)          I      DC(I)          I      DC(I)          I      DC(I)
      "
2710 LPRINT
2720 FOR I=0 TO IL
2730   DC(I)=((RE(I)^2)*H(I))/(S2*((1-H(I))^2)*IC)
2740   IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 2770
2750   LPRINT USING "      ## ##.##^####      ";I-3;DC(I-3);I-2;DC(I-2);I-1;DC(I-1);I
;DC(I)
2760   PRINT
2770 NEXT I
2780   IF II=0 GOTO 2830
2790   FOR I=0 TO II
2800     LPRINT USING "      ## ##.##^####      ";IL-II+I;DC(IL-II+I)
2810     PRINT
2820   NEXT I
2830 LPRINT
2840 LPRINT "                                V A L O R E S   D O   D   R A I Z   "
2850 LPRINT
2860 LPRINT "          I      DR(I)          I      DR(I)          I      DR(I)          I      DR(I)
      "
2870 LPRINT
2880 FOR I=0 TO IL
2890   DR(I)=(SGN(RE(I)))*(DC(I)^.5)
2900   IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 2930
2910   LPRINT USING "      ## ##.##^####      ";I-3;DR(I-3);I-2;DR(I-2);I-1;DR(I-1);I
;DR(I)

```

```

2920     PRINT
2930     NEXT I
2940     IF II=0 THEN GOTO 2980
2950     FOR I=0 TO II
2960         LPRINT USING "  ## ##.##^~~~ " ; IL-II+I; DR(IL-II+I)
2970     NEXT I
2980     LPRINT
2990     LPRINT "  V A L O R E S   D A   M E D I A   Q U A D R A T I C A   E X T E
R N A   S E ( I ) "
3000     LPRINT
3010     LPRINT "      I      SE2(I)      I      SE2(I)      I      SE2(I)      I
SE2(I)"
3020     LPRINT
3030     FOR I=0 TO IL
3040         SE2(I)=(((IL-IC)*S2)/(IL-IC-1))- (RE(I)^2)/(((IL-IC-1)*(1-H(I)))
3050         IF I+1-4*FIX((I+1)/4)<>0 GOTO 3080
3060         LPRINT USING "      ## ##.##.#### " ; I-3; SE2(I-3); I-2; SE2(I-2); I-1; SE2(I-
1); I; SE2(I)
3070     PRINT
3080     NEXT I
3090     IF II=0 GOTO 3140
3100     FOR I=0 TO II
3110         LPRINT USING "      ## ##.##.#### " ; IL-II+I; SE2(IL-II+I);
3120     LPRINT
3130     NEXT I
3140     LPRINT "      V A L O R E S   D E   D F F I T S "
3150     LPRINT
3160     LPRINT "      I      DF(I)      I      DF(I)      I      DF(I)      I
DF(I)"
3170     LPRINT
3180     FOR I=0 TO IL
3190         DFT(I)=(RE(I)*(H(I)^.5))/((SE2(I)^.5)*(1-H(I)))
3200         IF I+1-4*FIX((I+1)/4)<>0 THEN GOTO 3230
3210         LPRINT USING "      ## ##.##.#### " ; I-3; DFT(I-3); I-2; DFT(I-2); I-1; DFT(I
-1); I; DFT(I)
3220     PRINT
3230     NEXT I
3240     IF II=0 GOTO 3300
3250     FOR I=0 TO II
3260         LPRINT USING "      ## ##.##.#### " ; IL-II+I; DFT(IL-II+I);
3270     PRINT
3280     NEXT I
3290     LPRINT
3300     LPRINT 'D='; D
3310     STOP

```

VALORES DOS PARAMETROS AJUSTADOS

0 -39.9200 1 0.7156 2 1.2953 3 -0.1521
 DIAGONAL DA MATRIZ CHAPEU

I	H(I)	I	H(I)	I	H(I)	I	H(I)
0	0.3016	1	0.3178	2	0.1746	3	0.1285
4	0.0522	5	0.0775	6	0.2192	7	0.2192
8	0.1402	9	0.2000	10	0.1550	11	0.2172
12	0.1575	13	0.2058	14	0.1905	15	0.1311
16	0.4121	17	0.1606	18	0.1745	19	0.0802
20	0.2845						

VALORES AJUSTADOS DE Y

I	YAJ(I)	I	YAJ(I)	I	YAJ(I)	I	YAJ(I)
0	38.7649	1	38.9171	2	32.4441	3	22.3018
4	19.7113	5	21.0066	6	21.3891	7	21.3891
8	18.1440	9	12.7324	10	11.3633	11	10.2202
12	12.4282	13	12.0501	14	5.6382	15	6.0946
16	9.5196	17	8.4547	18	9.5979	19	13.5875
19	13.5875						
20	22.2373						

RESIDUOS RE(I)

I	RE(I)	I	RE(I)	I	RE(I)	I	RE(I)
0	3.2351	1	-1.9171	2	4.5560	3	5.6982
4	-1.7113	5	-3.0066	6	-2.3891	7	-1.3891
8	-3.1440	9	1.2676	10	2.6367	11	2.7798
12	-1.4282	13	-0.0501	14	2.3618	15	0.9054
16	-1.5196	17	-0.4547	18	-0.5979	19	1.4125
19	1.4125						
20	-7.2373						

VALORES DO D DE COOK

I	DC(I)	I	DC(I)	I	DC(I)	I	DC(I)
0	2.05E-01	1	7.95E-02	2	1.69E-01	3	1.74E-01
4	5.39E-03	5	2.61E-02	6	6.50E-02	7	2.20E-02
8	5.94E-02	9	1.59E-02	10	4.78E-02	11	8.68E-02
12	1.43E-02	13	2.60E-05	14	5.14E-02	15	4.51E-03
16	8.73E-02	17	1.49E-03	18	2.90E-03	19	5.99E-03
19	5.99E-03						
20	9.23E-01						

VALORES DO DRAIZ

I	DR(I)	I	DR(I)	I	DR(I)	I	DR(I)
0	4.53E-01	1	-2.82E-01	2	4.11E-01	3	4.17E-01
4	-7.34E-02	5	-1.61E-01	6	-2.55E-01	7	-1.48E-01
8	-2.44E-01	9	1.26E-01	10	2.19E-01	11	2.95E-01
12	-1.20E-01	13	-5.10E-03	14	2.27E-01	15	6.72E-02
16	-2.95E-01	17	-3.86E-02	18	-5.39E-02	19	7.74E-02
19	7.74E-02						
20	-9.60E-01						

VALORES DA MEDIA QUADRATICA EXTERNA SE(I)

I	SE2(I)	I	SE2(I)	I	SE2(I)	I	SE2(I)
0	10.2404	1	10.8402	2	9.6051	3	8.8483
4	10.9838	5	10.5645	6	10.7200	7	11.0224
8	10.4584	9	11.0513	10	10.6627	11	10.5599
12	11.0256	13	11.1767	14	10.7462	15	11.1179
16	10.9314	17	11.1615	18	11.1498	19	11.0413
19	11.0413						
20	6.6013						

VALORES DE DFFITS

I	DF(I)	I	DF(I)	I	DF(I)	I	DF(I)
0	0.7948	1	-0.4812	2	0.7442	3	0.7879
4	-0.1245	5	-0.2791	6	-0.4376	7	-0.2509
8	-0.4233	9	0.2132	10	0.3763	11	0.5092
12	-0.2026	13	-0.0086	14	0.3884	15	0.1131
16	-0.5019	17	-0.0650	18	-0.0906	19	0.1309
19	0.1309	20	-2.1001				

B I B L I O G R A F I A

- ANDREWS, D.F. and PREGIBON, D. (1978), "Finding the Outliers that Matter". J. Royal Stt. Soc., s.B, Vol. 40, pag. 85-93.
- BECKMAN, R.J. and TRUSSEL, H.J. (1974), "The Distribution of the Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression". J. Am. Stt. Assc., Vol. 69, pag.199-201.
- BESLEY, D.A., KUH, E. and WELSCH, R. (1980), "Regression Diagnostics: Identifying Influential Data and Sources of Colinearity". John Wiley, New York.
- CARVALHO, J. e DACHS, J.N. (1984), "Diagnóstico em Regressão". Sexto SINAPE, Rio de Janeiro.
- COOK, R.D. (1977), "Detection of Influential Observations in Linear Regression". Technometrics, Vol. 19, pag. 15-18.
- COOK, R.D. (1979), "Influential Observations in Linear Regression". J. AM. Stt. Assc., Vol. 74, pag. 169-174.
- COOK, R.D. and WEISBERG, S. (1980), "Characterizations of Empirical Influence Function for Detecting Influential Cases". Technometrics, Vol. 22, pag. 495-502.
- DACHS, J. (1978), "Análise de Dados e Regressão". Dpto. de Estatística IMECC-UNICAMP.
- DANIEL, C. and WOOD, F.S. (1981), "Fitting Equations to Data". John Wiley, New York, 2a. edição.
- DRAPER, N.R. and JOHN, J.A. (1981), "Influential Observations and Outliers in Regression". Technometrics, Vol. 23, pag. 21-26.
- DRAPER, N.R. and SMITH, H. (1981), "Applied Regression Analysis". John Wiley, New York, 2a. edição.

- GOODNIGHT, J.A. (1979), "A Tutorial on the SWEEP Operator". The American Statistician , Vol. 33, pag. 144-158.
- HOAGLIN, D.C. and WELSCH, R.E. (1978), "The Hat Matrix in Regression and ANOVA". The American Statistician , Vol.32, pag.17-22.
- HOCKING, R.D. (1983), "Developments in Linear Regression Methodology 1959-1982". Technometrics, Vol. 25, pag.219-247.
- MONTGOMERY, D.C. and PECK, E.A. (1982), "Introduction to Linear Regression Analysis". John Wiley, New York.
- MOSTELLER, F. and TUKEY, J.W. (1977), "Data Analysis and Regression". Reading. Mass.: Addison-Wesley.
- PREGIBON, D. (1981), "Logistic Regression Diagnostics". Annals of Statistics, Vol 9, pag. 705-724.
- TUKEY, J.W. (1977), "Exploratory Data Analysis". Reading Mass.: Addison-Wesley.
- VELLEMAN ,P.F. and HOAGLIN, D.C. (1981), "Applications Basics, and Computing of Exploratory Data Analysis". Mass.: Duxbury Press, Boston.
- VELLEMAN, P.F. and WELSCH, R.E. (1981), "Efficient Computing of the Regression Diagnostics". The American Statistician, Vol. 35, pag. 234-242.
- WEISBERG, S. (1980), "Applied Linear Regression". John Wiley, New York.