Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Análise de influência local no modelo de regressão logística

Édila Cristina de Souza

Dissertação apresentada para obtenção do título de Mestre em Agronomia. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2006

Édila Cristina de Souza
Licenciada em Matemática

Análise de influência local no modelo de regressão logística

Orientador:
Prof. Dr. **EDWIN MOISES MARCOS ORTEGA**

Dissertação apresentada para obtenção do título de Mestre em Agronomia. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2006

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Souza, Édila Cristina de
Análise de influência local no modelo de regressão logística / Édila Cristina de Souza. - - Piracicaba, 2006.
101p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2006.

1. Análise estatística 2. Análise de regressão e de correlação 3. Logística (estatística) 4. Modelagem de dados I. Título

CDD 519.36

Dedicatória

À

DEUS,

que sempre me iluminou e guiou os meus caminhos me dando forças para vencer.

À minha família, em especial aos meus pais,

Benedito Rondon de M. Souza (in memoriam), *pelo exemplo de caráter, dignidade e trabalho e*

Adélia Catarina Souza, *fundamental nesta conquista, pelo apoio, amor e confiança...*

Aos meus irmãos

Ronaldo Benedito de Souza e Nádia Cristina de Souza,

pela motivação, amizade e carinho, especialmente quando eu estava à distância...

Aos meus avós

Romão Baicere (in memoriam) *e*

Rosina Thommen Baicere, *pela ajuda financeira, orações, amor e carinho...*

Minha eterna gratidão...

Agradecimentos

Este período que estive em Piracicaba, muito aprendi. Várias pessoas influenciaram direta ou indiretamente na conclusão deste trabalho. Sou eternamente grata:

- Ao Prof. Dr. **Edwin Moisés Marcos Ortega** pela orientação, pelo crescimento pessoal e profissional na realização desta conquista.
- À Prof. Dr. **Clarice Garcia Borges Demétrio** por aconselhar nos momentos mais adequados.
- Aos Professores **Décio, Maria Cristina, Tadeu, Gabriel, Roseli, Sílvio e Sônia** do curso de Pós-Graduação em Estatística da ESALQ.
- À **Solange** pela disposição e eficiência.
- À **Luciane** e **Expedita** pelo atendimento sempre simpático.
- Ao **Jorge** pelo apoio técnico.
- Aos amigos e colegas do mestrado: **Alexandre, Ana Paula, Angela, Cristiane, Elisabeth, Fernanda, Hélio, Joseane, Juliana, Lúcio, Melissa, Moita, Pâmela e Sandra**.
- Aos amigos e colegas do doutorado: **Afrânio, Ana Maria, Andréia, César, David, Denise, Elizabeth, Genevile, Giovana, Idemauro, João Maurício, Juliana, Luciana, Luciano, Milton e Osmar**.
- Aos amigos do kitinet **Analy, Larissa, Laura, Marcelo e Maurício.**

SUMÁRIO

RESUMO

Análise de influência local no modelo de regressão logística

Uma etapa importante após a formulação e ajuste de um modelo de regressão é a análise de diagnóstico. A regressão logística tem se constituído num dos principais métodos de modelagem estatística de dados; mesmo quando a resposta de interesse não é originalmente do tipo binário, alguns pesquisadores tem dicotomizado a resposta de modo que a probabilidade de sucesso pode ser modelado através da regressão logística. Neste trabalho consideramos um estudo de diagnóstico no modelo da regressão logística, utilizando as medidas proposta por Pregibon (1981) e a técnica de influência local Cook (1986). Investigamos a aplicação da técnica de influência local sob diferentes esquemas de perturbação. Como ilustração, apresentamos a aplicação dos resultados desenvolvidos em dois conjuntos de dados reais.

Palavras-chaves: Regressão logística; análise de diagnóstico; influência local.

ABSTRACT

Analysis of local influence with the logistic regression model

An important stage after the formularization and adjustment of a regression model is the diagnosis analysis. Logistic regression is one of the main methods for modeling data and even when the response of interest is is not originally of the binary type, some researchers have dichotomized the response in a way that the success probability can be modeled through logistic regression. In this work we consider a study of diagnosis methods with logistic regression, using the measures proposed by Pregibon (1981) and the local influence technique of Cook (1986). We investigate the application of the local influence technique of under different types of disturbance. As as illustration, we show the application of the developed results obtained with real data sets.

Word-keys: Logistic regression; diagnosis analysis; local influence.

LISTA DE FIGURAS

LISTA DE TABELAS

1 INTRODUÇÃO

A análise de regressão é uma técnica estatística que tem como objetivo descrever a relação entre uma variável resposta e um conjunto de variáveis explicativas, através de um modelo que tenha bom ajuste.

Na regressão logística, a variável resposta, pode ser dicotômica ou binária, isto é, aquela que apresenta duas possibilidades de resposta (sucesso ou fracasso), como, por exemplo, o objetivo de um ensaio experimental realizado para testar a sobrevivência ou não de enxertos de um determinado cultivar, ou então, o efeito (sucesso ou fracasso) de um inseticida quando este é aplicado a um determinado número de insetos.

A regressão logística é conhecida desde os anos 50, entretanto, tornou-se mais usual através de Cox (1989) e de Hosmer & Lemeshow (1989). Aspectos teóricos do modelo de regressão logística são amplamente discutidos na literatura, destacando-se Kleinbaum (1994), Agresti (1990), Hosmer & Lemeshow (1989), Cox & Snell (1989), etc.

A modelagem dos dados pode ser feita com base em modelos estatísticos paramétricos supostamente apropriados. A escolha correta de um modelo que se ajuste de forma adequada a um conjunto específico de dados é de grande importância, uma vez que a não tendenciosidade dos resultados da análise depende dessa escolha. Assim, uma etapa importante na análise de um ajuste de regressão logística é o estudo da robustez dos resultados obtidos com relação à presença de pontos extremos. Detectar observações aberrantes e/ou influentes constitui um passo importante na análise do conjunto de dados. Pregibon (1981) aprimorou os métodos de diagnóstico de regressão linear para a regressão logística; desenvolvendo várias medidas para análise de resíduos e diagnóstico, como o resíduo de "*Pearson*" e da "*Deviance*".

Neste trabalho são discutidos alguns procedimentos de diagnóstico aplicados ao modelo de regressão logística; tendo sido utilizadas técnicas que possibilitam medir o quanto pequenas alterações nos dados ou no modelo podem influenciar nos resultados inferências do problema em estudo.

Técnicas simples são bastante utilizadas para tal propósito e se baseiam na retirada individual de casos. Medidas de influência para cada observação da amostra são construídas através da comparação de estimativas calculadas para o conjunto completo de dados e para o conjunto de dados obtido eliminando-se a observação correspondente.

Neste contexto, Cook (1977) sugere uma medida de inflûencia desenvolvida inicial-

mente para modelos de regressão linear com erros normais. Já Cook (1986) apresenta a técnica denominada de influência local, na qual ao invés de retirar uma observação, atribui-se um peso a mesma. Nesta última técnica, são introduzidas, simultaneamente, perturbações em cada um dos casos, sendo que a medida de influência é construída a partir da função do logaritmo da verossimilhança. Diferentes esquemas de perturbação podem ser aplicados, de acordo com o elemento da análise que o pesquisador deseja monitorar. Esta técnica permite detectar observações conjuntamente influentes, o que constitui uma vantagem em relação ao esquema de retirada de casos, visto que, neste último, possíveis observações influentes podem não ser detectadas devido a presença de outras observações.

A presença de observações influentes na amostra pode levar a resultados inferenciais completamente diferentes, sendo importante ao pesquisador conhecer e analisar estes casos para decidir pela retirada, ou não dos mesmos do estudo.

Essa metodologia teve uma grande receptividade entre os pesquisadores de regressão, havendo inúmeras publicações no assunto, como por exemplo, Ortega; Paula e Bolfarine (2003) que aplicam influência local em modelos log-gama generalizados com dados censurados e Hossain (2003) que aplica a metodologia em modelos de regressão logística.

Mediante o exposto, o objetivo do presente trabalho foi pesquisar e analisar as medidas propostas por Pregibon (1981) e a aplicação da metodologia de influência local nos modelos de regressão logística.

Este trabalho está organizado da seguinte forma: no capítulo II são apresentados alguns conceitos relacionados aos modelos de regressão logística simples e múltipla, assim como a parte inferencial. Também discutindo-se as medidas de diagnóstico propostas por Pregibon (1981). Ainda neste capítulo, a metodologia de influência local é descrita e aplicada no modelo de regressão logística considerando os diferentes esquemas de pertubação. Considerando dois conjuntos de dados reais, os resultados desta teoria são aplicadas no capítulo III. As considerações finais do trabalho são apresentadas no capítulo IV como uma discussão dos resultados obtidos e proposta de possíveis pesquisas futuras. Os resultados das análises, os dados e os programas correspondentes encontram-se no anexo.

2 DESENVOLVIMENTO

2.1 Regressão Logística

2.1.1 Fatos históricos

Um breve resumo histórico do modelo de regressão logística foi desenvolvido por Jan Salomon Cramer em 2002 no seu livro *"Logit Models from Economics and Other Fields"*. A Regressão Logística foi descoberta no século XIX para descrever o crescimento das populações e as reações químicas no curso de autocatálise.

O modelo logístico definido na época era razoável para se estudar o crescimento de países jovens, como os Estados Unidos. Assim sendo, já em 1789, Thomas Robert Malthus (1766-1834) defendia a hipótese de que a população aumentava em uma progressão geométrica. Enquanto Alphonse Quetelet (1795-1874), astrônomo belga, preocupava-se com a extrapolação do crescimento exponencial que iria conduzir a valores impossíveis; experimentando assim, vários ajustes da equação, que seriam estudados por seu aluno, Pierre-François Verhulst (1804-1849).

Verhulst publicou três artigos entre 1838 e 1847. O primeiro, uma breve nota na revista "*Correspondance Mathématique et Physique*", editado por Quetelet em 1838, contém a essência do argumento em quatro páginas. Neste artigo, Verhulst não mostra como ajustar a curva de crescimento, porém, demonstra que a mesma está em concordância com o curso atual da população da França, Bélgica, Essex e Russia para o período de 1833.

O segundo artigo, publicado na revista "*Proceedings*" da Belgian Royal Academy em 1845, define a função de uma forma mais complexa e com todas as suas propriedades. Verhulst nomeia a função por "logística", devido ao diagrama da curva ser parecido com a "*courbe logarithmique*", atualmente conhecida como exponencial.

A função logística foi analisada novamente em 1920 por Raymond Pearl (1879-1940) e Lowell J. Reed (1886-1966) no estudo do crescimento da população dos Estados Unidos. Ambos desconheciam o trabalho de Verhulst e conseguiram chegar à curva logística.

Pearl era biólogo, tendo adquirido conhecimento estatístico no período de 1905 à 1906 em Londres, com Karl Pearson. Tornou-se um prodigioso investigador e escreveu sobre grande variedade de fenômenos como longevidade, fertilidade, contracepção e os efeitos do consumo do álcool e do tabaco na saúde. Reed era matemático, tinha interesse pela biostatística; era excelente professor e administrador. Ambos trabalhavam na *Johns Hopkins University*.

O termo *logistic* não era usado até redescobrirem o trabalho de Verhulst, citado por Pearl e Reed em trabalhos publicados em 1922 e 1923.

A idéia básica do desenvolvimento logístico é simples e efetiva, usada nos dias atuais, para modelar o crescimento populacional e na introdução de novos produtos e tecnologias no mercado, como por exemplo telefones celulares por um processo de autocatálise (reações em cadeia), assim como muitos outros produtos e técnicas usadas na indústria.

A invenção do modelo *probit* é atribuído a Gaddum (1933) e Bliss (1934). Mas a origem do método, em particular, a transformação da distribuição normal foi traçado pelo estudante alemão Fechner (1801-1887).

Tabela 1 - Número de artigos em jornais estatísticos contendo a palavra *probit* ou *logit* (CRAMER, 2002)

Período	probit	logit
1935-39	6	-
1940-44	3	1
1945-49	22	6
1950-54	50	15
1955-59	53	23
1960-64	41	27
1965-69	43	41
1970-74	48	61
1975-79	45	72
1980-84	93	147
1985-89	98	215
1990-94	127	311
Total	629	919

A tabela 1 ilustra o desenvolvimento geral de publicações no **JSTOR** eletrônico que contém os doze principais jornais estatísticos no idioma inglês. A tabela apresenta o número de artigos que possuem a palavra *probit* ou *logit*. Percebe-se que a partir de 1970 há um crescimento do uso do termo *logit* em artigos.

As análises que relacionam respostas discretas binárias a várias covariáveis ficaram conhecidas como regressão logística, tendo ampla aceitação devido a utilização de computadores e dos pacotes desenvolvidos para a estimação da máxima verossimilhança para os modelos *logit* e *probit*. A **BMDP** (*Biomedical Data Processing*), em 1977, foi a primeira a oferecer esta facilidade, se tornando uma característica padrão para a maioria dos pacotes estatísticos.

2.1.2 Modelo de Regressão Logística Simples

Os métodos de regressão têm como objetivo descrever as relações entre a variável resposta (Y) e a variável explicativa (X). Na Regressão Logística, a variável resposta (Y) é dicotômica, isto é, atribui-se o valor 1 para o acontecimento de interesse ("*sucesso*") e o valor 0 para o acontecimento complemetar ("$fracasso$"), com probabilidades $\pi_i = P(Y = 1|X = x_i)$ e $1 - \pi_i = P(Y = 0|X = x_i)$, respectivamente. Para descrever a média condicional de Y dado X com a distribuição logística, é utilizada a notação π_i (HOSMER; LEMESHOW, 1989).

Considera-se uma série de eventos binários, em que $(Y_1, Y_2, \ldots, Y_n)$ são variáveis aleatórias independentes com distribuição Bernoulli, com probabilidade de sucesso (π_i), isto é, $Y_i \sim Ber(\pi_i)$ e denota-se $\boldsymbol{x}_i^T = (1, x_i)$ a i-ésima linha da matriz ($\boldsymbol{X}$) em que $i = 1, 2, \ldots, n$.

A probabilidade de sucesso do modelo logístico simples é definida como:

$$\pi_i = \pi(\boldsymbol{x}_i) = P(Y = 1|X = \boldsymbol{x}_i) \quad = \quad \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \tag{1}$$

e a probabilidade de fracasso:

$$1 - \pi_i = 1 - \pi(\boldsymbol{x}_i) = P(Y = 0|X = \boldsymbol{x}_i) \quad = \quad \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}, \tag{2}$$

em que $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ é o vetor de parâmetros desconhecidos.

Uma diferença importante entre o modelo de regressão logística e o modelo de regressão linear pode ser notada, quando diz respeito à natureza da relação entre a variável resposta e as variáveis independentes. Em qualquer problema de regressão, a quantidade a ser modelada é o valor médio da variável resposta dado os valores das variáveis independentes. Esta quantidade é chamada de média condicional, denotada por $E(Y|X = \boldsymbol{x}_i)$, em que Y é a variável resposta e $\boldsymbol{x}_i$, os valores das variáveis independentes. Na regressão linear tem-se $-\infty < E(Y|X = \boldsymbol{x}_i) < +\infty$ e na regressão logística, devido à natureza da variável resposta, $0 \leq E(Y|X = \boldsymbol{x}_i) \leq 1$.

Na regressão linear, $E(Y|X = \boldsymbol{x}_i) = \beta_0 + \beta_1 x_i$ e na regressão logística usando a

definição de variáveis aleatórias discretas, tem-se:

$$E(Y|X = \boldsymbol{x}_i) = 1P(Y_i = 1|X = \boldsymbol{x}_i) + 0P(Y_i = 0|X = \boldsymbol{x}_i) = \pi_i.$$

Outra diferença importante entre um modelo de regressão linear e o modelo de regressão logístico refere-se à distribuição condicional da variável resposta. No modelo de regressão linear assume-se que uma observação da variável resposta pode ser expressa por $Y_i = E(Y|X = x_i) + \varepsilon_i$, em que ε_i é chamado de erro, com distribuição Normal, média zero e variância constante. Isto não ocorre, quando a resposta é dicotômica. O valor da variável resposta dado $\boldsymbol{x}_i$, é expresso por $Y_i = \pi_i + \varepsilon_i$, como a quantidade ε_i, que pode assumir somente um de dois possíveis valores, isto é, $\varepsilon_i = 1 - \pi_i$ para $Y_i = 1$ ou $\varepsilon_i = -\pi_i$ para $Y_i = 0$, segue que ε_i tem distribuição com média zero e variância dada por $\pi_i(1 - \pi_i)$ (HOSMER; LEMESHOW, 1989).

Na Tabela 2, verifica-se a diferença entre o modelo de regressão linear simples e o modelo de regressão logística simples.

Tabela 2 - Comparação o Modelo de Regressão Linear Simples e o Modelo de Regressão Logística Simples (FARHAT, 2003)

Regressão Linear Simples	**Regressão Logística Simples**
$E(Y\|X = \boldsymbol{x}_i) = \beta_0 + \beta_1 x_i$	$E(Y\|X = \boldsymbol{x}_i) = \pi_i$
$-\infty < E(Y\|X = \boldsymbol{x}_i) < +\infty$	$0 \leq E(Y\|X = \boldsymbol{x}_i) \leq 1$
$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$	$Y_i = \pi_i + \varepsilon_i$
$\varepsilon_i \sim N(0, \sigma^2)$	$\varepsilon_i = \begin{cases} 1 - \pi_i \ \text{ com } \ P(Y_i = 1\|X = \boldsymbol{x}_i) \\ -\pi_i \ \text{ com } \ P(Y_i = 0\|X = \boldsymbol{x}_i) \end{cases}$
	$E(\varepsilon_i) = 0$ e $Var(\varepsilon_i) = \pi_i(1 - \pi_i)$
$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$	$Y_i \sim Ber(\pi_i)$

A transformação de π_i, é interpretada como o logaritmo da razão das chances entre π_i e $1 - \pi_i$. Esta transformação é bastante empregada em estudos toxicológicos, epidemiológicos e de outras áreas, sendo definida como:

$$g(\boldsymbol{x}_i) \ = \ \ln\left[\frac{\pi_i}{1 - \pi_i}\right] = \beta_0 + \beta_1 x_i. \tag{3}$$

2.1.2.1 Estimação dos Parâmetros

Supondo que (x_i, y_i) seja uma amostra independente com n pares de observações, y_i representa o valor da variável resposta dicotômica e x_i é o valor da variável independente da i-ésima observação em que $i = 1, 2, ...n$. Para o ajuste do modelo de regressão logística simples, segundo a equação (1), é necessário estimar os parâmetros desconhecidos: β_0 e β_1. O método mais utilizado para estimar esses parâmetros considerando uma regressão linear é o de mínimos quadrados. Neste método, a escolha de β_0 e β_1 é dada pelos valores que minimizam a soma de quadrados dos desvios para os valores observados (y_i) em relação ao valor predito $(\hat{y}_i)$ baseado no modelo. No entanto, quando o método de mínimos quadrados é aplicado para um modelo com variável dicotômica, os estimadores não seguem as mesmas pressuposições do modelo de regressão linear.

O método de máxima verossimilhança é utilizado para estimar os parâmetros. A função de distribuição da probabilidade de Y_i para o modelo de regressão logística simples com $Y_i \sim Ber(\pi_i)$ é dada por:

$$f(y_i, \pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}.$$

Como as observações são independentes, a função de distribuição de probabilidade conjunta de $y_1, y_2, \ldots, y_n$ será:

$$\prod_{i=1}^{n} f(y_i, \pi_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i},\ y_i \in [0, 1].$$

Então, a função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i},\ \boldsymbol{\beta} \in \mathbb{R}^{(2)}. \tag{4}$$

O princípio da máxima verossimilhança é estimar o valor de $\boldsymbol{\beta}$ que maximiza $L(\boldsymbol{\beta})$. Aplicando logaritmo, a expressão é definida como:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \ln[L(\boldsymbol{\beta})] = \ln\left[\prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i}\right] \\ &= \sum_{i=1}^{n} [y_i \ln(\pi_i) + (1 - y_i)\ln(1 - \pi_i)] \\ &= \sum_{i=1}^{n} [y_i \ln(\pi_i) + \ln(1 - \pi_i) - y_i \ln(1 - \pi_i)] \\ &= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi_i}{1 - \pi_i}\right) + \ln(1 - \pi_i)\right]. \end{aligned} \tag{5}$$

Substituindo pelas equações (2) e (3), temos:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\left[y_i(\beta_0+\beta_1 x_i)+\ln\left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)\right] \\ &= \sum_{i=1}^{n}\left[y_i(\beta_0+\beta_1 x_i)-\ln(1+\exp(\beta_0+\beta_1 x_i))\right]. \end{aligned} \tag{6}$$

Para encontrar o valor de $\boldsymbol{\beta}$ que maximiza $l(\boldsymbol{\beta})$, deriva-se $l(\boldsymbol{\beta})$ em a relação cada parâmetro (β_0, β_1), obtendo-se duas equações:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^{n}\left[y_i - \frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\exp(\beta_0+\beta_1 x_i)\right] \\ \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1} &= \sum_{i=1}^{n}\left[y_i x_i - \frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\exp(\beta_0+\beta_1 x_i)x_i\right], \end{aligned}$$

que igualadas a zero, geram o seguinte sistema de equações:

$$\sum_{i=1}^{n}(y_i-\pi_i) = 0 \tag{7}$$

$$\sum_{i=1}^{n} x_i(y_i-\pi_i) = 0 \tag{8}$$

em que $i = 1, \ldots, n$ e $\pi_i = \dfrac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}$.

Como as equações (7) e (8) não são lineares em β_0 *e* β_1, são necessários métodos iterativos para resolução, estes disponíveis em vários programas computacionais, a serem discutidos na parte de regresão logística múltipla.

2.1.2.2 Interpretação dos Coeficientes

Para iniciar a discussão dos coeficientes do modelo logístico, inicialmente será demonstrada a situação na qual a variável independente também é dicotômica. Neste caso, a variável x será codificada como 0 ou 1. Em relação ao modelo, existem dois valores para π_i que equivalem a dois valores para $(1-\pi_i)$.

A chance da resposta quando $x = 1$ é definida como $\pi(1)/[1-\pi(1)]$. Da mesma forma, a chance da resposta quando $x = 0$ é definida como $\pi(0)/[1-\pi(0)]$. O logaritmo da razão é dado por:

$$g(1) = \ln \pi(1)/[1-\pi(1)] \quad \text{e} \quad g(0) = \ln \pi(0)/[1-\pi(0)].$$

Tabela 3 - Valores do Modelo de Regressão Logística quando a variável independente é dicotômica

Variável resposta Y	**Variável Independente X**	
	$x = 1$	$x = 0$
$y = 1$	$\pi(1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
$y = 0$	$1 - \pi(1) = \dfrac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \dfrac{1}{1 + \exp(\beta_0)}$
Total	1.0	1.0

A razão das chances (*"Odds ratio"*), denotada por Ψ, é definida por:

$$\Psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]}. \tag{9}$$

O logaritmo da razão das chances (*"log-odds"*) é:

$$\ln(\Psi) = \ln\left[\frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]}\right] = g(1) - g(0). \tag{10}$$

Usando a expressão para o modelo de regressão logística como mostrado na tabela 2, a razão de chances é definida por:

$$\Psi = \frac{\left(\dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}\right) \Big/ \left(\dfrac{1}{1 + \exp(\beta_0 + \beta_1)}\right)}{\left[\dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)}\right] \Big/ \left[\dfrac{1}{1 + \exp(\beta_0)}\right]} = \frac{\exp(\beta_0 + \beta 1)}{\exp(\beta_0)} = \exp(\beta_1), \tag{11}$$

e o logaritmo da razão de chances é dado por:

$$\ln(\Psi) = \ln[\exp(\beta_1)] = \beta_1.$$

A razão de chances é uma medida de associação muito utilizada em muitas áreas. Por exemplo, se Y representa a presença ou ausência de câncer no pulmão e se X representa se a pessoa é ou não fumante, um valor $\hat{\Psi} = 2$ pode ser interpretada como a chance de uma pessoa que fuma adquirir câncer no pulmão é duas vezes maior que uma pessoa que não fuma.

A razão de chances é um parâmetro de grande interesse no modelo de regressão logística devido sua fácil interpretação. A distribuição assimétrica de $\hat{\Psi}$ é devida ao fato de seus limites tenderem a zero (PAULA, 2004). As inferências são freqüentemente baseadas na distribuição do $\ln(\hat{\Psi}) = \hat{\beta}_1$, o qual tende a seguir uma distribuição normal, mesmo para pequenas amostras.

Assim sendo, a razão de chances é definida como a chance de ocorrência de um evento entre indivíduos que têm um fator de risco, comparados com indivíduos não expostos, sujeitos ao evento.

O risco relativo (RR) é utilizado em estudos prospectivos, fornecendo o risco de desenvolvimento de uma determinada condição (frequentemente uma doença) para um grupo quando comparado a outro grupo. O risco relativo é a relação entre $\pi(1)$ e $\pi(0)$:

$$RR = \frac{\dfrac{\pi(1)}{\pi(1)[1-\pi(1)]}}{\dfrac{\pi(0)}{\pi(0)[1-\pi(0)]}} = \frac{\pi(1)}{\pi(0)}.$$

O intervalo de confiança, com nível de confiança $100(1-\alpha)\%$ para a razão de chances é obtido inicialmente calculando o intervalo para β_1 e aplicando exponencial, tem-se:

$$\exp[\hat{\beta}_1 \pm z_{1-\alpha/2} SE(\hat{\beta}_1)],$$

em que $SE(\hat{\beta}_1)$ é o erro padrão de $\hat{\beta}_1$.

2.1.3 Modelo de Regressão Logística Múltipla

Hosmer e Lemeshow (1989) generalizam o modelo de regressão logística para o caso de mais de uma variável independente.

Seja um conjunto com p variáveis independentes, denotadas por $\boldsymbol{x}_i^T = (x_{i0}, x_{i1}, x_{i2}, \ldots, x_{ip})$, o vetor da i-ésima linha da matriz $(\boldsymbol{X})$ das variáveis explicativas, em que cada elemento da matriz corresponde ao ij-ésimo componente (x_{ij}), em que $i = 1, 2, \ldots, n$ e $j = 0, 1, \ldots, p$, com $x_{i0} = 1$. Denota-se por $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, o vetor de parâmetros desconhecidos e β_j é o j-ésimo parâmetro associado a variável explicativa x_j. No modelo de regressão logística múltipla a probabilidade de sucesso é dada por:

$$\begin{aligned} \pi_i = \pi(\boldsymbol{x}_i) = P(Y_i = 1 | \boldsymbol{X} = \boldsymbol{x}_i) &= \frac{\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})} \\ &= \frac{\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \end{aligned} \tag{12}$$

e a probabilidade de fracasso por:

$$1-\pi_i = 1-\pi(\boldsymbol{x}_i) = P(Y_i=0|\boldsymbol{X}=\boldsymbol{x}_i) = \frac{1}{1+\exp\left(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}\right)} = \frac{1}{1+\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)}$$

No modelo de regressão múltipla assume-se que Y_i tem uma distribuição de Bernoulli com parâmetro de sucesso π_i.

O *"logit"* para o modelo de regressão múltipla é dado pela equação:

$$g(\boldsymbol{x}_i) = \ln\left[\frac{\pi_i}{1-\pi_i}\right] = \boldsymbol{x}_i^T\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}$$

Assim, o logaritmo da função de verossimilhança pode ser escrito como:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[y_i\boldsymbol{x}_i^T\boldsymbol{\beta} - \ln(1+\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\})\right]. \tag{13}$$

2.1.3.1 Estimação dos parâmetros

Para poder estimar os parâmetros foi utilizado o método de máxima verossimilhança, similar ao caso da regressão logística simples.

Para encontrar o valor de $\boldsymbol{\beta}$ que maximiza $l(\boldsymbol{\beta})$, foi utilizado o processo iterativo de Newton-Raphson, e para isso fez-se necessário derivar $l(\boldsymbol{\beta})$ em relação a cada parâmetro,

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\beta_j} = \sum_{i=1}^{n}\left[y_i x_{ij} - \frac{\exp\left(x_i^T\boldsymbol{\beta}\right)}{1+\exp\left(x_i^T\boldsymbol{\beta}\right)}x_{ij}\right] = \sum_{i=1}^{n}\left[y_i-\pi_i\right]x_{ij}$$

dessa maneira, o vetor escore $\boldsymbol{U}(\boldsymbol{\beta})$ pode ser escrito como

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{\pi} = \boldsymbol{X}^T(\boldsymbol{y}-\boldsymbol{\pi}).$$

A matriz de informação de Fischer é dada por:

$$\boldsymbol{I}(\boldsymbol{\beta}) = E\left[-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right] = \boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X},$$

sendo $\boldsymbol{Q} = diag[\boldsymbol{\pi}_i(1-\boldsymbol{\pi}_i)]$ e $\boldsymbol{X}$ a matriz de dados, e sua inversa $[\boldsymbol{I}(\boldsymbol{\beta})]^{-1}$, a matriz de variâncias e covariância das estimativas de máxima verossimilhança dos parâmetros (SILVA, 1992).

A solução para as equações de verossimilhança é obtida usando o método iterativo de Newton Raphson. O conjunto de equações iterativas é dado por:

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + [\boldsymbol{I}(\boldsymbol{\beta}^{(t)})]^{-1}\boldsymbol{U}(\boldsymbol{\beta}^{(t)}); t = 0, 1, 2, \ldots \\ &= \boldsymbol{\beta}^{(t)} + [\boldsymbol{X}^T\boldsymbol{Q}^{(t)}\boldsymbol{X}]^{-1}\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{\pi}^{(t)}). \end{aligned} \tag{14}$$

sendo que $\boldsymbol{\beta}^{(t)}$ e $\boldsymbol{\beta}^{(t+1)}$ são vetores de parâmetros estimados nos passos t e $t+1$, respectivamente.

O chute inicial é dado com todos os coeficientes igualados a zero. Esses valores iniciais são substituídos no lado direito da equação (14), que dará o resultado para a primeira iteração, $\beta^{(1)}$. Os valores então são novamente substituídos no lado direito, $\boldsymbol{U}(\boldsymbol{\beta})$ e $\boldsymbol{I}(\boldsymbol{\beta})$ são recalculados, encontrando $\beta^{(2)}$. Esse processo é repetido, até que a máxima mudança em cada parâmetro estimado do próximo passo seja menor que um critério. Se o valor absoluto do corrente parâmetro estimado $\beta^{(t)}$ é menor ou igual a 0,01, o critério para convergência é: $\left|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right| < 0,0001$. Se o parâmetro estimado for maior que 0,01, assume-se o seguinte critério: $\left|\frac{\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}}{\boldsymbol{\beta}^{(t)}}\right| < 0,0001$, conforme Allison (1999).

2.1.3.2 Inferência

A etapa de inferência tem como objetivo principal verificar a adequação do modelo e realizar um estudo detalhado das discrepâncias. Estas podem levar a eleger outro modelo ou a aceitação da existência de possíveis pontos influentes.

Nesta etapa, deve-se verificar a precisão dos parâmetros estimados, construir intervalos de confiança, testar hipóteses e por último realizar análise de diagnóstico e de resíduos.

Geralmente não é possível encontrar distribuições exatas para os estimadores, assim sendo, trabalha-se com resultados assintóticos considerando-se que o modelo escolhido irá satisfazer as condições de regularidade.

Cox e Hinkley (1986) demostram que, em problemas regulares, a função *Escore* $\boldsymbol{U}(\boldsymbol{\beta}) = \dfrac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ tem valor esperado igual a zero e a estrutura de covariância é igual a matriz de informação de Fischer $\boldsymbol{I}(\boldsymbol{\beta}) = E\left[-\dfrac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right] = \boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}$.

Assim, a distribuição assintótica dos $\boldsymbol{\beta}$ é dada por:

$$\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \boldsymbol{I}(\boldsymbol{\beta})^{-1}).$$

Os métodos de inferência são baseados na teoria de máxima verossimilhança. Con-

forme esta teoria, existem três estatísticas para testar hipóteses relativas aos $\boldsymbol{\beta}$´s, que são deduzidas de distribuições assintóticas de funções adequadas dos $\boldsymbol{\beta}$´s (DEMÉTRIO, 2002).

Supondo-se interesse em testar as hipóteses:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$$
$$H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

As três estatísticas são:

i) A estatística da razão da verossimilhança que é dada pela diferença de:

$$\Lambda = -2 \ln \left[\frac{L(\boldsymbol{\beta}_0)}{L(\widehat{\boldsymbol{\beta}})} \right] = 2[l(\widehat{\boldsymbol{\beta}}) - l(\boldsymbol{\beta}_0)]$$

em que $\widehat{\boldsymbol{\beta}}$ é o estimador da máxima verossimilhança sob todo espaço paramétrico.

ii) A estatística Wald que é dada por:

$$W = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \boldsymbol{I}(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

em que $\boldsymbol{I}(\widehat{\boldsymbol{\beta}})$ é a matriz de informação de Fischer avaliada em $\widehat{\boldsymbol{\beta}}$.

iii) A estatística Escore que é dada por:

$$Es = \boldsymbol{U}^T(\boldsymbol{\beta}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}(\boldsymbol{U}(\boldsymbol{\beta}_0))$$

em que $\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$ é a matriz de informação avaliada em $\boldsymbol{\beta}_0$

Essas três estatísticas são assintoticamente equivalentes e sob H_0 tem-se que:

$$\Lambda, W \text{ e } Es \sim \chi^2_p$$

2.1.3.3 Bondade-de-ajuste

A bondade-de-ajuste refere-se ao teste estatístico aplicado na obtenção do modelo final, visando-se aferir se este é o mais indicado.

i) **Estatística *Deviance***: O processo de ajuste de um modelo consiste em propor ao mesmo um pequeno número de parâmetros, de tal forma que resuma toda a informação da amostra. Dado um conjunto de n observações, um modelo de até n parâmetros pode ser ajustado,

sendo denominado modelo saturado, sendo que este indica toda variação ao componente sistemático e reproduzindo exatamente os dados. Por outro lado, o modelo mais simples tem somente um parâmetro, β_0, sendo denominado modelo nulo, e indicando toda variação ao componente aleatório. Na prática, o modelo nulo é, em geral, muito simples e o modelo saturado não é informativo, uma vez que não resume os dados, somente os reproduzindo. No entanto, o modelo saturado serve como base para medir a discrepância de um modelo intermediário de p parâmetros.

Existem muitas estatísticas para medir esta discrepância, das quais a mais utilizada está baseada na função de verossimilhança, proposta por Nelder e Wedderburn (1972), com o nome de *deviance*. Os autores comparam o valor da função de verossimilhança para o modelo proposto com $p+1$ parâmetros ($L(\hat{\beta}_0, \ldots, \hat{\beta}_p)$) ao seu valor no modelo saturado ($L(y_1, \ldots, y_n)$). Para esta comparação é conveniente utilizar menos duas vezes o logaritmo do quociente destes máximos. Assim, a *deviance* é definida como:

$$D = -2 \ln \left[\frac{L(\hat{\beta}_0, \ldots, \hat{\beta}_p)}{L(y_1, \ldots, y_n)} \right]$$

equação na qual verifica-se a utilização de um teste de razão de verossimilhança generalizado.

No modelo de regressão logística, considerado o modelo com as proporções estimadas $\hat{\pi}_i$, temos que a *deviance* pode ser escrita como:

$$\begin{aligned} D &= -2 \sum_{i=1}^{n} \left[y_i - \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i) - y_i \ln(y_i) + (1 - y_i) \ln(1 - y_i) \right] \\ &= -2 \sum_{i=1}^{n} \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \\ &= 2 \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{\hat{\pi}_i} \right) + (1 - y_i) \ln \left(\frac{1 - y_i}{1 - \hat{\pi}_i} \right) \right] \end{aligned}$$

A *deviance* D sempre é positiva e quanto menor, melhor é o ajuste do modelo.

Considerando-se o caso de réplicas, tem-se $K \leq n$ em que $k = 1, 2, \ldots, K$ e que para cada x_k existem m_k elementos amostrais, isto é

- m_1 elementos na amostra com $X = x_1$
- m_2 elementos na amostra com $X = x_2$
- m_K elementos na amostra com $X = x_K$

sendo $\sum_{k=1}^{K} m_k = n$ (FARHAT,2003).

Na regressão logística, a probabilidade estimada de um evento é dada por:

$$\hat{\pi}_k = m_k \left[\frac{\exp \hat{g}(\boldsymbol{x}_k)}{1 + \exp \hat{g}(\boldsymbol{x}_k)} \right],$$

sendo que o número calculado de eventos associados a esta covariável padrão é calculado por:

$$\hat{y}_k = m_k \hat{\pi}_k = m_k \left[\frac{\exp \hat{g}(\boldsymbol{x}_k)}{1 + \exp \hat{g}(\boldsymbol{x}_k)} \right]$$

em que $\hat{g}(\boldsymbol{x}_k)$ é o *logit* estimado.

Para um conjunto de k valores das variáveis explicativas, o componente da *deviance* é definido por:

$$d(y_k, \hat{\pi}_k) = \pm \left\{ 2 \left[y_k \ln \left(\frac{y_k}{m_k \hat{\pi_k}} \right) + (m_k - y_k) \ln \left(\frac{(m_k - y_k)}{m_k (1 - \hat{\pi_k})} \right) \right] \right\}^{1/2}$$

em que o sinal é o mesmo de $(y_k - m_k \hat{\pi_k})$.

Para o modelo com $m_k = 1$ e $y_k = 0$ a *deviance* é dada por:

$$d(y_k, \hat{\pi}_k) = -\sqrt{2 |\ln(1 - \hat{\pi_k})|},$$

e a deviance quando $m_k = 1$ e $y_k = 1$ é definida por:

$$d(y_k, \hat{\pi}_k) = \sqrt{2 |\ln(\hat{\pi_k})|}.$$

Em resumo, a estatística baseada no resíduo da *deviance* é dada por:

$$D = \sum_{k=1}^{L} d(y_k, \hat{\pi}_k)^2.$$

A distribuição assintótica da *deviance* é dada por Collet (1991):

$$D \sim X^2_{(n-p)}$$

em que p é o número de parâmetros estimados no modelo.

ii) **Estatística X^2 de *Pearson***: Na regressão linear, o resíduo para cada elemento amostral é definido como a diferença entre os valores observados e os valores estimados, isto é:

$$r_k = y_k - \hat{y_k}. \tag{15}$$

Na regressão linear a variância dos erros não depende da média condicional $E(Y_k|\boldsymbol{x}_k)$, entretanto na regressão logística a variância dos erros é uma função da média condicional:

$$Var(Y_k|\boldsymbol{x}_k) = m_k E(Y_k|\boldsymbol{x}_k)[1 - E(Y_k|\boldsymbol{x}_k)] = m_k\pi_k(1-\pi_k).$$

Dividindo-se o resíduo definido na equação (15) pelo desvio padrão, tem-se o resíduo de *Pearson*:

$$rp(y_k, \hat{\pi}_k) = \frac{y_k - m_k\hat{\pi}_k}{\sqrt{m_k\hat{\pi}_k(1-\hat{\pi}_k)}}$$

A estatística X^2 de *Pearson* é definida por:

$$X^2 = \sum_{k=1}^{K} rp(y_k, \hat{\pi}_k)^2.$$

Esta estatística possui distribuição assintótica χ^2_{n-p}. Em geral, a diferença entre os valores observados da *deviance* e a estatística X^2 de *Pearson* não tem importância na prática.

Quando o método de máxima verossimilhança é utilizado para estimar os parâmetros, a *deviance* é uma medida de bondade-de-ajuste. Assim é preferível utilizar a *deviance* antes da estatística X^2 de *Pearson*, uma vez que ao se comparar modelos, esta pode ser utilizada para avaliar a importância do termo adicional (McCULLAGH e NELDER, 1989).

2.2 Análise de Resíduos e Diagnósticos

Quando se está ajustando um modelo a um conjunto de dados, é imprescindível que as estimativas obtidas a partir do modelo proposto sejam resistentes a pequenas perturbações, tanto no modelo como nos dados. Se o modelo ajustado não apresentar uma boa descrição dos dados que foram observados, o mesmo pode conduzir a inferências errôneas.

Assim, é importante que se faça um estudo sobre a robustez dos resultados obtidos, quanto aos vários aspectos que envolvem a formulação do modelo e as estimativas de seus parâmetros, ou seja, que se faça uma análise de resíduos e diagnósticos.

A análise de resíduos e diagnóstico é utilizada para detectar problemas, tais como:

- presença de observações discrepantes (pontos aberrantes);
- inadequação das pressuposições para os erros aleatórios ou para as médias;

- colinearidade entre as colunas da matriz do modelo;
- forma funcional do modelo inadequada;
- presença de observações influentes.

Pregibon (1981) propõe medidas de resíduos e diagnósticos para regressão logística, as definindo como estatísticas de influência. Estas estatísticas são as mesmas utilizadas pelo software **SAS** no procedimento **PROC LOGISTIC** com a opção **INFLUENCE**, basicamente as estatísticas de influência definem quanto a eliminação de uma observação em particular pode influenciar no ajuste do modelo. As medidas geralmente utilizadas para os resíduos e diagnósticos são sequencialmente abordadas.

2.2.1 Diagonal da matriz H(*leverage*)

Os elementos da matriz H são utilizados para detectar pontos extremos no espaço designado. Esses pontos exercem um papel importante no ajuste final dos parâmetros de um modelo estatístico, ou seja, sua eliminação pode implicar mudanças substancias dentro de uma análise estatística.

No modelo de regressão linear clássica, a matriz **H** é definida por:

$$\mathbf{H} \;=\; \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T,$$

conhecida como matriz de projeção da solução de mínimos quadrados ou matriz *hat*.

Como nos modelos de regressão logística, a $Var(\varepsilon_i) = \pi_i(1-\pi_i)$ não é constante, sendo utilizada a definição de mínimos quadrados ponderados, definindo a matriz de projeção para o modelo logístico como:

$$\mathbf{H} \;=\; \boldsymbol{Q}^{1/2}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Q}^{1/2},$$

o que sugere a utilização dos elementos da diagonal principal de H para detectar a presença de pontos de alavanca nesse modelo de regressão. Hosmer e Lemeshow (1989) mostram, contudo, que o uso da diagonal principal da matriz de projeção H deve ser feito com algum cuidado em regressão logística e que as interpretações são diferentes daquelas do caso normal linear. Dessa forma, a diagonal da matriz $\hat{H}$ é dada por:

$$\hat{h_{ii}} \;=\; \hat{\pi}_i(1-\hat{\pi}_i)(\boldsymbol{x}_i^T)[\boldsymbol{I}(\hat{\beta})]^{-1}(\boldsymbol{x}_i); \quad i = 1, 2, \ldots, n.$$

2.2.2 Resíduo de *Pearson*

O resíduo de *Pearson* auxilia na classificação de uma observação que pode ser considerado como *outliers*. O resíduo para cada elemento amostral é definido como a diferença entre os valores observados e os valores preditos, conhecido como resíduo ordinário e definido por:

$$r_i = y_i - \hat{\pi}_i$$

Devido ao efeito da escala de medição, este tipo de resíduo não é útil para detectar *outliers*. Assim sendo, é necessário transformar este resíduo para eliminar o efeito de medição da variável resposta e da preditora.

Na regressão logística, o resíduo de *Pearson* transformado é definido por:

$$(\mathrm{rp})_\mathrm{i} = \frac{\mathrm{y_i} - \hat{\pi}_\mathrm{i}}{\sqrt{\hat{\pi}_\mathrm{i}(1-\hat{\pi}_\mathrm{i})}}; \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n}, \tag{16}$$

sendo que, no caso desses valores serem pequenos, há indicação de que o modelo está bem ajustado. Os resíduos de *Pearson* são componentes da estatística qui-quadrado de *Pearson*.

2.2.3 Resíduo de *Deviance*

Os resíduos de *Deviance* são componentes da *Deviance*, sendo utilizados para detectar os erros no ajuste do modelo, medem a discrepância entre o modelo saturado e o modelo restrito em relação as observações y_i. O resultado da *deviance* é uma estatística de bondade-de-ajuste, para cada indivíduo $(i = 1, 2, \ldots, n)$ baseada no logaritmo da função de verossimilhança, definida por:

$$d_i = \begin{cases} -\sqrt{-2\ln(1-\hat{\pi}_i)} & \text{se } y_i = 0 \\ \pm\sqrt{2\left[y_i \ln\left(\dfrac{y_i}{\hat{\pi}_i}\right) + (-y_i)\ln\left(\dfrac{1-y_i}{(1-\hat{\pi}_i)}\right)\right]} & \text{se } 0 < y_i < 1 \\ \sqrt{-2\ln(\hat{\pi}_i)} & \text{se } y_i = 1 \end{cases}$$

Pregibon (1981), o definiu por desvio residual utilizando o contexto dos MLGs (Modelos Lineares Generalizados) e demonstrou que, se existe uma transformação que normalize a distribuição dos resíduos, então as raízes quadradas das componentes do desvio são resíduos que exibem as mesmas propriedades induzidas por esta transformação (CORDEIRO; NETO, 2004).

2.2.4 C e CBar

São diagnósticos baseados no intervalo de confiança, que fornecem medidas da influência das observações individuais sob β, e possuem a mesma idéia da Distância de Cook na teoria de regressão linear. Pregibon (1981) utilizando aproximações demonstra que essa medida pode ser escrita como:

$$C_i = \frac{(\mathrm{rp}_i)^2 h_{ii}}{(1-h_{ii})^2}; \quad i = 1, 2, \ldots, n.$$

Christensen (1997) define uma nova medida $\overline{C}_i$, também chamada por $CBar$, em termos da medida C_i, a qual é difinida como:

$$\overline{C}_i = \frac{(\mathrm{rp}_i)^2 h_{ii}}{(1-h_{ii})}; \quad i = 1, 2, \ldots, n.$$

2.2.5 DIFCHISQ

Esta medida é útil para detectar as observações mal ajustadas, ou seja, observações que contribuam pesadamente na diferença entre os dados e os valores preditos.

Usando aproximações lineares e a estatística qui-quadrado de *Pearson*, a medida DIFCHISQ pode ser definida como:

$$DIFCHISQ_i = \frac{\overline{C}_i}{h_{ii}} = \frac{(\mathrm{rp_i})^2}{1-h_{ii}}; \quad i = 1, 2, \ldots, n.$$

2.2.6 DIFDEV

De forma similar, a DIFCHISQ é utilizada para detectar observações que são influentes na estimação do ajuste do modelo de regressão logística. Baseada no resíduo da *deviance*, é definida por:

$$DIFDEV_i = d_i^2 + \overline{C}_i = d_i^2 + \frac{(\mathrm{rp_i})^2}{h_{ii}(1-h_{ii})}; \quad i = 1, 2, \ldots, n.$$

Estas estatísticas de diagnóstico são conceitualmete interessantes, pois permitem identificar as covariáveis que são pobremente ajustadas (grandes valores de $DIFDEV_i$ e/ou $DIFDEV_i$) e aquelas que têm grande influência nas estimativas dos parâmetros. Depois de identificar esses elementos, pode-se decidir sobre a sua permanência ou não na análise.

Ao contrário da regressão linear, as estatísticas de diagnóstico para modelos de regressão logística não são normalmente distribuídas, portanto, faz-se necessário confiar nas avali-

ações feitas através de gráficos, na experiência e no conjunto de dados. Os gráficos para o diagnóstico são de grande utilidade para detectar pontos influentes no modelo de regressão logística.

2.2.7 Superdispersão no modelo

Quando o modelo de regressão logística é utilizado para analisar um conjunto de dados, assume-se que a transformação logística das probabilidades da resposta depende linearmente de um conjunto de variáveis explicativas e que o número de sucessos segue uma distribuição Bernoulli. Se o modelo linear logístico ajustado for satisfatório, deve reproduzir adequadamente as probabilidades de resposta observadas e modelar, de uma maneira apropriada, a variação dos dados. Como anteriormente mencionado ao se ajustar um modelo a n proporções Bernoulli, a *deviance* terá uma distribuição assintótica $X^2_{(n-p)}$ em que p é o número de parâmetros desconhecidos. Conhecendo-se que o valor esperado para uma variável $X^2_{(n-p)}$ é $(n-p)$, infere-se que a *deviance* de um modelo bem ajustado deve ser aproximadamente igual a seus graus de liberdade ou equivalente a *deviance* média que deverá estar próximo de um.

Quando a *deviance* média é muito maior que um, é um indício de que algumas suposições feitas não estão sendo satisfeitas, o que é causado, principalmente, pelo: componente sistemático inadequado de alguma maneira; ou existe um ou mais valores discrepantes, ou a suposição de variabilidade Bernoulli não é válida. Considerando-se que a parte sistemática do modelo está correto, mas a *deviance* média é muito maior que um, então pode-se afirmar que a suposição da variabilidade Bernoulli não é válida e que os dados exibem superdispersão, isto é, quando a variância amostral $Var(y_i)$ excede a variância nominal $\pi_i(1-\pi_i)$, variância esperada conforme o modelo probabilístico estabelecido. Outra causa para o problema de superdispersão pode ser devido a uma correlação entre as respostas binárias.

No entanto, deve-se ter cuidado ao imaginar que diferentes causas estão provocando a superdispersão, e em geral não é simples de determinar a verdadeira causa. Assim, por exemplo McCullagh e Nelder (1989) observaram que a superdispersão está quase sempre presente em dados reais e sobretudo em dados discretos. Mais detalhes de superdispersão são abordados por Collet (1991); Hinde e Demétrio (1998) e Paula (2004).

2.2.7.1 Detecção da superdispersão

Existem muitas formas de detectar o problema de superdispersão, estas sendo abordada por Dean (1992) e Lu (1999). Neste trabalho enfatizar-se-á a deteção de superdispersão mediante o gráfico de envelopes. O afastamento dos resíduos observados não somente da média, como também dos envelopes estaria indicando a presença de superdispersão; esta é uma alternativa gráfica porém eficiente. Atkinson (1985) propôs adicionar um gráfico de envelope tal que sob o modelo proposto, os pontos correspondentes aos dados observados tem uma determinada probabilidade de cair dentro dos mesmos.

Hinde e Demétrio (1997) e Paula (2005) descrevem detalhadamente como construir o gráfico de envelopes. Num gráfico semi-normal, são representados os valores absolutos de alguma medida de diagnóstico, digamos td_i ordenados do menor ao maior $td_{(i)}$ e seus correspondentes valores esperados das estatísticas da normal padrão

$$\Phi^{-1}\left(\frac{i+n-1/8}{2n+1/2}\right); \quad i=1,2,\ldots,n.$$

Para construir o gráfico de envelope, simulam-se k amostras com o mesmo número de observações que o conjunto de dados originais (n), utilizando o modelo ajustado, isto é, os parâmetros ajustados correspondentes a cada observação e a distribuição suposta para a componente aleatória. Para cada amostra ajusta-se o modelo, calcula-se o valor absoluto da estatística de interesses e ordena-se do menor para o maior, obtendo-se k conjuntos de valores ordenados. Com os k valores obtidos na primeira posição, calcula-se o máximo, o mínimo e a média; devendo-se proceder da mesma forma com os valores obtidos na segunda posição e assim sucessivamente até a n-ésima posição. Depois plotam-se os mínimos, máximos e médias junto aos valores de $td_{(i)k}$ correspondentes aos dados observados versus as estatísticas de ordem esperadas da normal padrão. Assim, espera-se ter um gráfico que represente os dados conjuntamente com o valor esperado e a banda de confiança que resulta para o modelo ajustado. Por isso, uma vez que o modelo seja considerado adequado, espera-se que a banda de confiança seja um envelope que contenha os dados.

O algorítmo para construir o gráfico normal dos resíduos com os envelopes é o seguinte:

1. Definir a matriz das covariáveis $\boldsymbol{X}$ de ordem $n \times (p+1)$;

2. Calcular a diagonal da matriz $\mathbf{H}$;

3. Calcular uma das medidas de diagnóstico, neste caso, tendo sido escolhido o resíduo de "deviance"definido por d_i;

4. Obter $td_i = d_i/\sqrt{1-h_{ii}}$ em que $i = 1, \ldots, n$;

5. Gerar n observações com distribuição $U(0,1)$; calcular a diferença entre os valores simulados e $\hat{\pi}_i$, armazenando em $\boldsymbol{y}^T = (y_1, \ldots, y_n)$;

6. Ajustar um novo modelo $\boldsymbol{y}$ contra $\boldsymbol{X}$, e deste calcular os resíduos td_i;

7. Repetir os passos 5 e 6, K vezes, assim ter-se-ão os resíduos gerados td_{ik} em que $i = 1, \ldots, n$ e $k = 1, \ldots, K$;

8. Ordenar de forma crescente os n grupos dos resíduos td_{ik}, gerando os valores $td_{(i)k}$;

9. Calcular os limites inferiores $td_{(i)I} = \underbrace{\min}_{1<k<K}(td_{(i)k})$, os limites superiores $td_{(i)S} = \underbrace{\max}_{1<k<K}(td_{(i)k})$ e a média $td_{(i)M} = \underbrace{\text{media}}_{1<k<K}(td_{(i)k})$;

10. Plotar estes valores contra os valores esperados das estatísticas de ordem normal padrão z_i, dada por:

$$z_i \cong \Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$$

em que $\Phi(.)$ é a função acumulada da $N(0,1)$ e i representa a posição ocupada pelo valor absoluto ordenado do resíduo. No pacote estatístico **R**, o gráfico normal de probabilidades pode ser construído com o comando **qqnorm**.

Geralmente, utiliza-se $K = 19$, pois com este número a probabilidade de que o resíduo absoluto maior dos dados originais caia fora dos limites do envelope é de 5% (COLLET, 1991). Este resultado também pode ser utilizado para mostrar se uma observação é discrepante ou não.

Hinde e Demétrio (1998) discutem o uso destes gráficos no contexto de modelagem de dados com superdispersão, concluindo que é perfeitamente válido considerar esta técnica gráfica para avaliar a adequação do modelo com superdispersão. Uma vez estabelecido que um conjunto de dados apresenta superdispersão, Hinde e Demétrio (1998) categorizam os procedimentos em 2 grupos:

i) Assumir um modelo com duas etapas, ou seja, assumir uma distribuição base e uma distribuição para o parâmetro da distribuição base;

ii) Assumir uma forma mais geral para a função variância, possivelmente, incluindo parâmetros adicionais.

2.3 Influência Local

Ajustando um modelo a um conjunto de dados, deseja-se que as estimativas obtidas a partir do modelo proposto sejam resistentes a pequenas perturbações no modelo ou nas observações.

Enquanto a análise de resíduos estuda os problemas com o modelo ajustado, como presença de observações aberrantes e afastamentos sérios das suposições para a distribuição do erro, uma análise de influência é feita assumindo o modelo como correto, e estuda-se a robustez das conclusões a perturbações nos dados ou no modelo. Uma observação se diz influente quando produz alterações desproporcionais nos resultados da análise ao ser omitida no ajuste do modelo ou submetida a uma pequena perturbação.

Na análise de diagnóstico, considera-se que o modelo postulado é o modelo correto, e comparam-se as estimativas obtidas através desse modelo com as estimativas decorrentes de uma pequena perturbação.

Cook (1986) desenvolveu alguns procedimentos de Diagnóstico de Influência Local. Essa metodologia é extensamente discutida por vários pesquisadores para a Regressão Linear, Regressão Não-Linear, Modelos Lineares Generalizados e modelos de Análise de Sobrevivência. Hossain (2003) analisa os procedimentos de Diagnóstico para os modelos de regressão logística.

Existem na literatura numerosos trabalhos de aplicação da metodologia de Cook (1986), por exemplo, Galea; Bolfarine e Vilca-Labra (2002), Ortega; Bolfarine e Paula (2003) e Hossain (2003).

2.3.1 Metodologia de Influência Local

Dado um conjunto de observações, seja $l(\boldsymbol{\beta})$ o logaritmo da função de verossimilhança correspondente ao modelo postulado, sendo que $\boldsymbol{\beta}$ é um vetor $(p+1) \times 1$ de parâmetros desconhecidos. Perturbações podem ser introduzidas no modelo através de um vetor $\boldsymbol{w}^T = (w_0, w_1, ..., w_n)$ pertencente a um subconjunto aberto Ω de $\mathbb{R}^n$. Geralmente, $\boldsymbol{w}$ pode refletir qualquer esquema

de perturbação bem definida, por exemplo, $\boldsymbol{w}$ pode ser usado para introduzir uma menor modificação nas variáveis explicativas ou para perturbar a matriz de covariância nos erros, no modelo de regressão linear. (GALEA; PAULA; BOLFARINE, 1997).

Supondo que o esquema de perturbação esteja definido, denotado por $l(\boldsymbol{\beta}|\boldsymbol{w})$ como logaritmo da função de verossimilhança perturbada, o vetor $\boldsymbol{w}$ expressa um esquema de pesos, existindo um ponto w_0, em que $l(\boldsymbol{\beta}|\boldsymbol{w}_0) = l(\boldsymbol{\beta})$. Dado que $\widehat{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança obtido através de $l(\boldsymbol{\beta})$ e $\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}}$ é o estimador de máxima verossimilhança obtido através de $l(\boldsymbol{\beta}|\boldsymbol{w})$, o objetivo é comparar $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}}$, quando $\boldsymbol{w}$ varia em Ω. Cook (1986) sugere que a comparação entre $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}}$ seja feita através do ajuste pela verossimilhança $LD(\boldsymbol{w})$, expresso da seguinte maneira:

$$LD(\boldsymbol{w}) = 2[l(\widehat{\boldsymbol{\beta}}) - l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}})]. \tag{17}$$

Dessa forma, $LD(\boldsymbol{w})$ contém informação essencial sobre a influência do esquema de perturbação.

A idéia de Cook (1986) é estudar o comportamento da função $LD(\boldsymbol{w})$ numa vizinhança $\boldsymbol{w}_0$, que é o ponto em que as duas verossimilhanças são iguais. Para isso, o autor considerou a seguinte superfície geométrica:

$$\alpha(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{w} \\ LD(\boldsymbol{w}) \end{pmatrix},$$

que é denominada de gráfico de influência. A idéia principal do autor, foi de analisar como $\alpha(\boldsymbol{w})$ desvia-se de seu plano tangente em $\boldsymbol{w}_0$, preocupando-se com o comportamento da função $LD(\boldsymbol{w})$ em torno de $\boldsymbol{w}_0$. O procedimento consiste em selecionar uma direção unitária $\boldsymbol{d}$, e, então, considerar o gráfico de $LD(w_0 + a\boldsymbol{d})$ em função de a, em que $a \in \mathbb{R}$. Esse gráfico é chamado de linha projetada. Desde que $LD(w_0) = 0$, $LD(w_0 + a\boldsymbol{d})$ tem um mínimo local em $a = 0$. Cada linha projetada pode ser caracterizada pela curvatura normal C_d em torno de $a = 0$. Sugere-se considerar a direção $\boldsymbol{d}_{max}$ correspondente à maior curvatura $C_{\boldsymbol{d}_{max}}$. O gráfico de $\boldsymbol{d}_{max}$ revela os elementos que sob pequenas perturbações, exercem notável influência sobre $LD(\boldsymbol{w})$.

Cook (1986) mostra que a curvatura normal na direção $\boldsymbol{d}$ pode ser expressa da seguinte forma:

$$C_d = 2|\boldsymbol{d}^T \boldsymbol{F} \boldsymbol{d}|, \tag{18}$$

sendo que $\boldsymbol{F} = \boldsymbol{\Delta}^T \boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{-1} \boldsymbol{\Delta}$, $\boldsymbol{I}(\widehat{\boldsymbol{\beta}})$ é a matriz de informação observada sob o modelo postulado e

$\boldsymbol{\Delta}$ é a matriz $(p+1) \times n$ definida por:

$$\boldsymbol{\Delta} = \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial \boldsymbol{\beta} \partial \boldsymbol{w}^T} \tag{19}$$

e avaliados em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ e $\boldsymbol{w} = \boldsymbol{w}_0$.

O resultado na equação (19) pode ser utilizado para avaliar a influência que o esquema de perturbações considerado exerce sobre os componentes do modelo, tais como estimativas dos parâmetros e outros resultados da análise estatística. Segundo Cook (1986), a direção que produz a maior mudança local na estimativa dos parâmetros é dada por $\boldsymbol{d}_{max}$, que corresponde ao autovetor associado ao maior autovalor de $\boldsymbol{\Delta}^T \boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{-1} \boldsymbol{\Delta}$. O vetor $\boldsymbol{d}_{max}$ é utilizado para identificar as observações que podem estar controlando propriedades importantes na análise dos dados.

2.3.2 Esquemas de Perturbação

Os métodos de diagnóstico para dados perturbados utilizados são: casos ponderados, perturbação, perturbação de uma covariável e perturbação de um subconjunto de covariáveis.

2.3.2.1 Caso Ponderado

Para avaliar a influência das perturbações de casos, o logaritmo da função de verossimilhança perturbada é definida por:

$$l(\boldsymbol{\beta}|\boldsymbol{w}) = \sum_{i=1}^{n} w_i \left[\boldsymbol{y}_i \boldsymbol{x}_i^T \boldsymbol{\beta} - \ln(1 + \exp(x_i^T \boldsymbol{\beta})) \right] \tag{20}$$

Para esse esquema de perturbação, o vetor correspondente à não perturbação é o vetor n-dimensional $\boldsymbol{w}_0 = (1, 1, \ldots, 1)^T$. Nesse caso, a i-ésima linha da matriz $\boldsymbol{\Delta}$ é dada por

$$\boldsymbol{\Delta}_i^T = \left[\frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial \beta_0 \partial w_i}, \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial \beta_1 \partial w_i}, \ldots, \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial \beta_p \partial w_i} \right]$$

Assim, os elementos da i-ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ e $\boldsymbol{w} = \boldsymbol{w}_0$, para todo $j = 1, 2, \ldots, p$ podem ser expressos da seguinte maneira:

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial \beta_j \partial w_i} &= \left[y_i x_{ij} - \frac{\exp\{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}\}} x_{ij} \right] \\ &= (y_i - \pi_i) x_{ij} \end{aligned}$$

2.3.2.2 Variáveis Explanatórias

Assim como realizado por Thomas e Cook (1990) e Hossain (2003), modificou-se a t-ésima coluna da matriz de dados $\boldsymbol{X}$, adicionando um vetor $\boldsymbol{w}$ de pequenas perturbações multiplicadas por um fator de escala v. Neste caso, a perturbação é da forma:

$$x_{it} \longrightarrow x_{it} + vw_i, \quad i = 1, \ldots, n,$$

sendo que v está atribuindo um peso para cada elemento da perturbação w_i. Como peso utilizou-se a estimativa do desvio padrão da variável X_t. Nesse caso, o logaritmo da função de verossimilhança perturbada é dado por:

$$l(\boldsymbol{\beta}|\boldsymbol{w}) = \sum_{i=1}^{n} w_i \left[\boldsymbol{y}_i \boldsymbol{x}_i^{T*}\boldsymbol{\beta} - \ln(1 + \exp\left(\boldsymbol{x}_i^{T*}\boldsymbol{\beta}\right))\right], \tag{21}$$

sendo que,

$$\boldsymbol{x}_i^{T*}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_t(x_{it} + vw_i) + \ldots + \beta_p x_{ip}$$

Assim, os elementos da i-ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ e $\boldsymbol{w} = \boldsymbol{w}_0$, para todo $j = 0, 1, 2, \ldots, p$ podem ser expressos da seguinte maneira:

$$\frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j \partial w_i} = \begin{cases} \left[(y_i - \widehat{\pi}_i) - \widehat{\pi}_i(1 - \widehat{\pi}_i)\widehat{\beta}_t x_{ij}\right] v & \text{para } j = t \\ \\ -\widehat{\pi}_i(1 - \widehat{\pi}_i) x_{ij} \widehat{\beta}_t v & \text{para } j \neq t \end{cases}$$

Para exemplificar os resultados anteriores, perturba-se-á a primeira covariável ($t = 1$). Portanto, $\boldsymbol{X}^*$ terá a seguinte forma:

$$\boldsymbol{X}^*_{n\times(p+1)} = \begin{bmatrix} x_{10} & x_{11} + w_{11}v & \ldots & x_{1j} & \ldots & x_{1p} \\ x_{20} & x_{21} + w_{21}v & \ldots & x_{2j} & \ldots & x_{2p} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_{i0} & x_{i1} + w_{i1}v & \ldots & x_{ij} & \ldots & x_{ip} \\ \vdots & \vdots & \ldots & \vdots & \ddots & \vdots \\ x_{n0} & x_{n1} + w_{n1}v & \ldots & x_{nj} & \ldots & x_{np} \end{bmatrix}$$

e a i-ésima linha da matriz de perturbação $\boldsymbol{X}^*$ é dada por:

$$\boldsymbol{x}_i^{T*} = [x_{i0}, x_{i1} + w_{i1}v_1, x_{i2}, \ldots, x_{ij}, \ldots, x_{ip}]$$

Seja $\boldsymbol{\beta}^T = [\beta_0, \beta_1, \ldots, \beta_p]$, então:

$$\boldsymbol{x}_i^{T*}\boldsymbol{\beta} = [\beta_0 x_{i0} + (x_{i1} + w_{i1}v)\beta_1 + x_{i2}\beta_2 + \ldots + x_{ij}\beta_j + \ldots + x_{ip}\beta_p]$$

O logaritmo da função de verossimilhança perturbada é dado conforme a equação (21), sendo que do cálculo das derivadas obtém-se:

$$\frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j \partial w_i} = \begin{cases} \left[(y_i - \widehat{\pi}_i) - \widehat{\pi}_i(1-\widehat{\pi}_i)\widehat{\beta}_1 x_{ij}\right]v & \text{para } j = 1 \\ -\widehat{\pi}_i(1-\widehat{\pi}_i)x_{ij}\widehat{\beta}_1 v & \text{para } j \neq 1 \end{cases}$$

Como a primeira covariável é a que está sendo perturbada, então a curvatura $\boldsymbol{\Delta}$ será:

$$\boldsymbol{\Delta} = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_1\partial w_{11}} & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_1\partial w_{21}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_1\partial w_{i1}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_1\partial w_{n1}} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j\partial w_{11}} & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j\partial w_{21}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j\partial w_{i1}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_j\partial w_{n1}} \\ \vdots & \vdots & \ldots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_p\partial w_{11}} & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_p\partial w_{21}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_p\partial w_{i1}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_p\partial w_{n1}} \end{bmatrix}$$

2.3.3 Influência Local Total

Lesaffre e Verbeke (1998) sugeriram avaliar a direção do i-ésimo indivíduo, que é dada pelo vetor $\boldsymbol{d}_i = (0, \ldots, 1, \ldots, 0)$, sendo que o i-ésimo elemento é um. Nesse caso, a curvatura normal chamada de influência local total do i-ésimo indivíduo, é dada por

$$C_i = 2|\boldsymbol{\Delta}_i^T[\boldsymbol{I}(\boldsymbol{\beta})]^{-1}\boldsymbol{\Delta}_i|, \tag{22}$$

sendo que, sugere-se estudar o gráfico de C_i contra a ordem das observações.

2.3.4 Particionando o vetor de parâmetros

Cook (1986) propõe o uso da metodologia, em caso específico, quando há interesse somente em parte do conjunto de parâmetros para o modelo de regressão linear. Hossain (2003)

estende esta metodologia para o modelo de regressão logística.

Neste caso, considera-se que o vetor de parâmetros $\boldsymbol{\beta}$ pode ser particionado na seguinte forma: $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, admitindo-se que o interesse deste trabalho está particularmente em $\boldsymbol{\beta}_1$. Neste caso, a superfície admitida será

$$\alpha_s(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{w} \\ LD(\boldsymbol{w})_s \end{pmatrix},$$

em que $LD(\boldsymbol{w})_s$ é a função de afastamento da verossimilhança definida por:

$$LD(\boldsymbol{w}) = 2[l(\widehat{\boldsymbol{\beta}}) - l(\widehat{\boldsymbol{\beta}}_{1\boldsymbol{w}}, g(\widehat{\boldsymbol{\beta}}_{1\boldsymbol{w}}))],$$

sendo $\widehat{\boldsymbol{\beta}}_{1\boldsymbol{w}}$ o subvetor obtido de $\boldsymbol{\beta}_{\boldsymbol{w}}^T = (\boldsymbol{\beta}_{1\boldsymbol{w}}^T, \boldsymbol{\beta}_{2\boldsymbol{w}}^T)$ e $g(\widehat{\boldsymbol{\beta}}_{1\boldsymbol{w}})$, a função que, para cada $\boldsymbol{\beta}_1$ fixado maximixa $l(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, que representa o logaritmo da função de verossimilhança perfilada de $\boldsymbol{\beta}_1$. A curvatura normal na superfície $\alpha_s(\boldsymbol{w})$ na direção do vetor unitário $\boldsymbol{d}$ é dada por:

$$C_d = 2|\boldsymbol{d}^T \boldsymbol{\Delta}^T (\mathbf{I}(\boldsymbol{\beta})^{-1} - \mathbf{B_{22}}) \boldsymbol{\Delta} \boldsymbol{d}|,$$

sendo $\mathbf{B_{22}} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I}_{22}^{-1} \end{pmatrix}$ com $\mathbf{I}_{22}^{-1}$, submatriz de $I(\boldsymbol{\beta})$, obtida segundo a partição $\mathbf{I}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{pmatrix}$.

Considerando-se a direção do i-ésimo indivíduo para esse caso, a influência local total do i-ésimo indivíduo é dada por:

$$C_i = 2|\boldsymbol{\Delta}_i^T (\mathbf{I}(\boldsymbol{\beta})^{-1} - \mathbf{B_{22}}) \boldsymbol{\Delta}_i|$$

3 MATERIAL E MÉTODOS

3.1 Aplicação 1

3.1.1 Introdução

Os dados a serem utilizados são provenientes da Pesquisa Nacional por Amostra de Domicílios (PNAD - 2003), feita pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no estado de Mato Grosso.

O sistema de pesquisas domiciliares, implantado progressivamente no Brasil a partir de 1967, com a criação da PNAD, tem como finalidade a produção de informações básicas para o estudo do desenvolvimento socioeconômico do País.

A pesquisa sobre trabalho infantil, realizada como tema suplementar da PNAD, agregou informações importantes para caracterizar com mais profundidade se o adolescente trabalha (Y=1) ou não trabalha (Y=0), em relação aos aspectos socioeconômicos.

Levando em consideração o envolvimento de adolescentes em atividade econômica como em pequenos empreendimentos, especialmente em atividade agrícola, tem-se 244 indivíduos entre de 14 a 15 anos de idade. O trabalho das crianças e dos adolescentes pode representar um auxílio na geração de renda ou na produção destinada ao consumo domiciliar.

A seguir, são identificadas as variáveis utilizadas:

y_i: Indica se o adolescente trabalha (0=não trabalha, 1=trabalha) (**trabalha**);

x_{i1}: A investigação é feita em anos completos, na data de referência da pesquisa, com base no dia, mês e ano do nascimento da pessoa. (**idade**);

x_{i2}: Indica o número de anos de estudo da pessoa, conforme a data de referência da pesquisa (**anoestu**);

x_{i3}: Situação do domicílio, classificação da localização do domicílio em urbano ou rural, definida por lei municipal vigente na ocasião da realização do Censo Demográfico. A situação urbana abrange as áreas correspondentes às cidades (sedes municipais), às vilas (sedes distritais) ou às áreas urbanas isoladas. A situação rural abrange toda a área situada fora desses limites. Este critério é, também, utilizado na classificação da população urbana e rural (**rural**);

x_{i4}: Classifica se a pessoa é do sexo feminino ou masculino (0=homem, 1=mulher)(**mulher**);

x_{i5}: Indica a idade da mãe, sendo que esta investigação é feita conforme a característica idade (**idademae**);

x_{i6}: Renda total da mãe é o rendimento mensal de trabalho em dinheiro ou o valor do rendimento em produtos ou mercadorias do ramo que compreende a agricultura, silvicultura, pecuária, extração vegetal, pesca e piscicultura, provenientes do trabalho principal ou do trabalho secundário e dos demais trabalhos que a pessoa tem na semana de referência da pesquisa, exceto o valor da produção para consumo próprio (**rendtotmae**);

x_{i7}: Educação da mãe, indica quantos anos de estudo que esta possui, conforme a data de referência da pesquisa (**educamae**);

x_{i8}: Indica a idade do pai, sendo que a pesquisa é feita conforme a característica idade (**idadepai**);

x_{i9}: Renda total do pai, ídem rendtomae (**rendtotpai**);

x_{i10}: Educação do pai, indica quantos anos de estudo que este possui, conforme a data de referência da pesquisa (**educapai**);

x_{i11}: Indica o número de pessoas que residem na mesma unidade domiciliar (**numpes**);

x_{i12}: Rendimento mensal familiar, é a soma dos rendimentos mensais dos componentes da família, excluindo aquele das pessoas, cuja condição na família é de pensionista, empregado doméstico ou parente do empregado doméstico (**rendtotal**).

A distribuição da variável resposta é dada conforme a tabela 4:

Tabela 4 - Distribuição dos adolescentes que trabalham, segundo o desfecho deste estudo

Trabalha	Total de frequência	Porcentagem(%)
0 (não)	183	75
1 (sim)	61	25
Total	244	100

Assim, pode-se observar que, para a presente pesquisa, 25% dos adolescentes trabalham e 75% não trabalham.

Ajustando um modelo de regressão logística e testando as hipóteses

$$H_0 : \boldsymbol{\beta} = \mathbf{0}$$

$$H_1 : \boldsymbol{\beta} \neq \mathbf{0}$$

ter-se-á que as estatísticas são dadas por:

Tabela 5 - Estatísticas da Razão da verossimilhança, Escore e Wald

Estatísticas	p-Valor
Razão da verossimilhança(Λ)= 31,3291	0,0018
Escore (Es)= 30,5435	0,0023
Wald (W)= 26,5674	0,0089

Na tabela 5, pode ser claramente observada que foi rejeitada a hipótese nula, assim sendo, pelo menos uma variável é significativa para o modelo.

Na tabela 6 são apresentadas as estimativas de máxima verossimilhança, erro padrão e a estatística de Wald para testar a significância de cada variável.

Tabela 6 - Estimativas dos parâmetros

Efeito	Parâmetro	Estimativa	Erro padrão	Estatística Wald	p-Valor
Intercepto	β_0	-4,9995	4,8399	1,0670	0,3016
idade	β_1	0,3486	0,3410	1,0454	0,3066
anoestu	β_2	0,0769	0,0989	0,6038	0,4371
rural	β_3	1,0812	0,3645	8,7976	0,0030
mulher	β_4	-1,1415	0,3461	10,8800	0,0010
idademae	β_5	-0,0340	0,0297	1,3159	0,2513
rendtotmae	β_6	-0,00026	0,000546	0,2351	0,6278
educamae	β_7	-0,0228	0,0545	0,1752	0,6755
idadepai	β_8	0,00198	0,0178	0,0123	0,9116
rendtotpai	β_9	-0,00043	0,000372	1,3322	0,2484
educapai	β_{10}	-0,1075	0,0595	3,2694	0,0706
numpes	β_{11}	0,0132	0,1154	0,0132	0,9087
srendtotal	β_{12}	0,000537	0,000286	3,5202	0,0606

$Deviance = 243,090$ com 231 g.l.

Verifica-se que as variáveis rural e mulher são significativas para o modelo, e que a *deviance* está um pouco afastada do seu grau de liberdade, o que indica a necessidade de se ter maior cuidado com o ajuste.

A tabela 7 contém as razões de chances estimadas.

Através da razão de chances, (tabela 7), percebe-se que a variável rural é um fator de risco e a variável mulher um fator de proteção para a variável resposta, sendo que a chance de um indivíduo da zona rural trabalhar é 2,95 vezes maior em relação a zona urbana.

3.1.2 Medidas de resíduos e diagnóstico

Anteriormente, neste trabalho, foram apresentadas as definições de algumas das medidas de resíduos e diagnóstico utilizadas por Pregibon (1981). Para o cálculo dessas medidas foi utilizado o *software* SAS. No anexo A, encontram-se os programas e todos os valores dos resíduos mencionados. Através dos gráficos dessas medidas, foram verificados os possíveis pontos discrepantes.

Tabela 7 - Estimativas das razões de chances

		Limite de Confiança	
Efeito	Ponto estimado	Inferior	Superior
idade	1,417	0,726	2,765
anoestu	1,080	0,890	1,311
rural	**2,948**	1,443	6,023
mulher	**0,319**	0,162	0,629
idademae	0,967	0,912	1,024
rendtotmae	1,000	0,999	1,001
educamae	0,977	0,878	1,088
idadepai	1,002	0,968	1,038
rendtotpai	1,000	0,999	1,000
educapai	0,898	0,799	1,009
numpes	1,013	0,808	1,271
srendtotal	1,001	1,000	1,000

Na figura 1, correspondente ao resíduo de Pearson (rp_i), contra a ordem das observações, nota-se claramente que a observação 154 destaca-se dentre as outras; analogamente, observando o resíduo *deviance* na figura 2, percebe-se que não existe nenhum ponto discordante. Em relação a medida do *leverage* ($\hat{h}_{ii}$), figura 3, verifica-se que a observação 190 se destaca das demais. Na figura 4, correspondente a medida C, a observação 190 aparece como um possível ponto influente. A medida $Cbar$ na figura 5, aponta a observação 190 como um possível ponto influente. Também na figura 6, correspondente a medida $DIFCHISQ$, contra a ordem das observações nota-se claramente que a observação 190 destaca-se dentre as outras. Na figura 7 referente a medida $DIFDEV$ verifica-se novamente que a observação 190 pode ser considerada com um possível ponto discrepante.

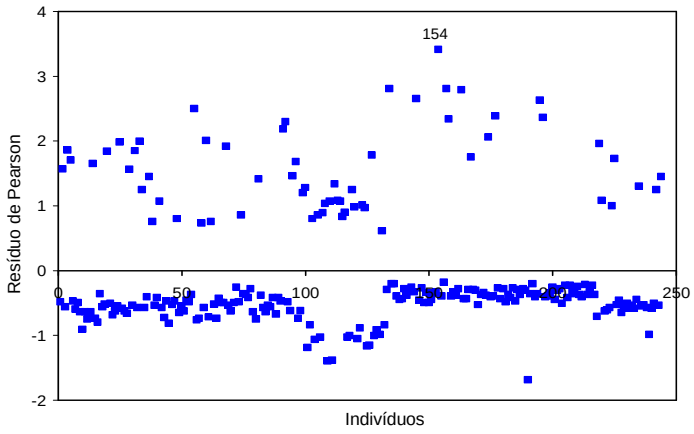

Figura 1 - Gráfico do Resíduo de Pearson

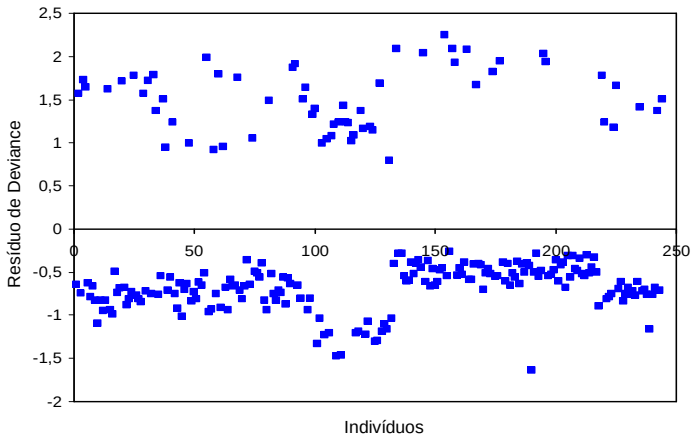

Figura 2 - Gráfico do Resíduo de Deviance

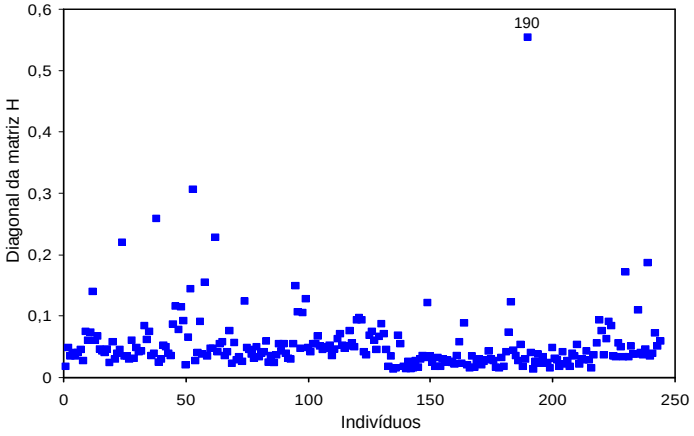

Figura 3 - Gráfico da diagonal da matriz H

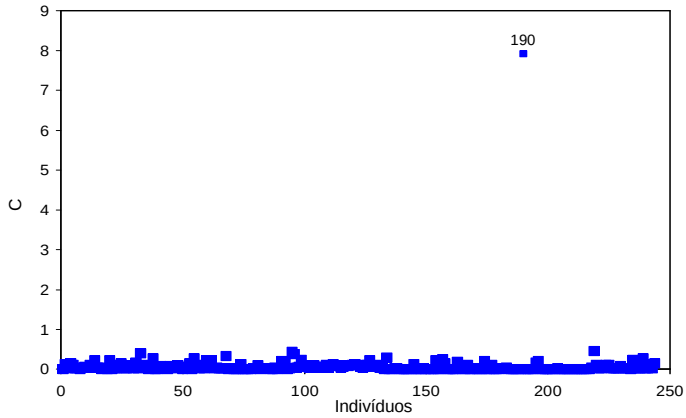

Figura 4 - Gráfico de C

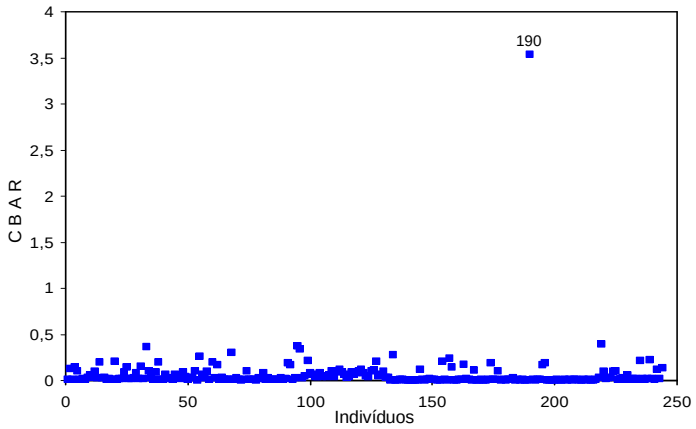

Figura 5 - Gráfico de CBAR

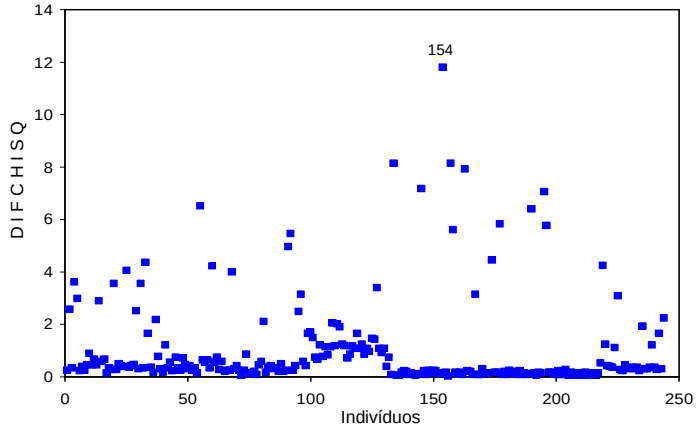

Figura 6 - Gráfico do DIFCHISQ

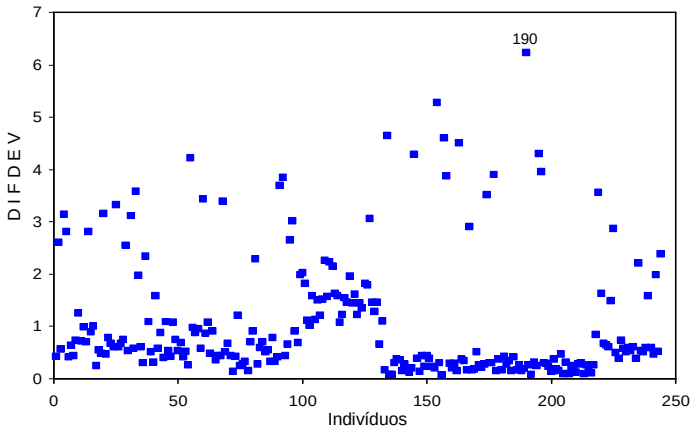

Figura 7 - Gráfico do DIFDEV

3.1.3 Influência local

Adotando o modelo de regressão logística e o esquema de perturbação de casos, temos que: $C_{d_{max}} = 3,792301$.

Assim, na figura 8, é apresentado o gráfico do autovetor correspondente a $C_{d_{max}}$ e na figura 9, a influência local total do i-ésimo indivíduo. Nota-se que a observação 190 é a que mais se destaca das demais, pois é a que apresenta a maior renda familiar total no conjunto de dados.

Quando perturba-se individualmente cada uma das covariáveis, verifica-se nos gráficos do autovetor correspondente e nos gráficos da influência local total do i-ésimo indivíduo, que a observação 190 pode ser considerada um possível ponto influente.

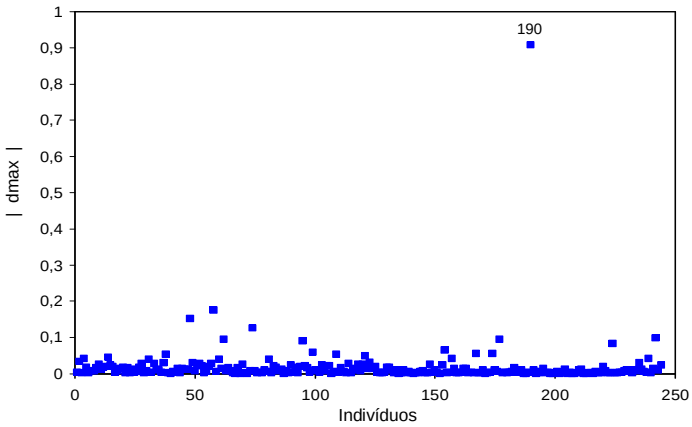

Figura 8 - Gráfico de influência - ponderação de casos

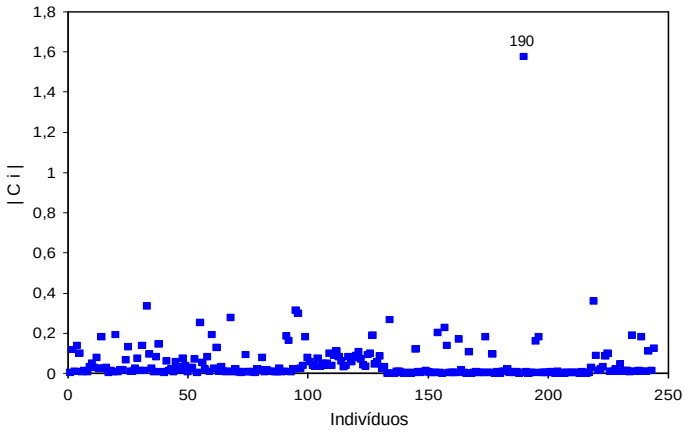

Figura 9 - Gráfico de influência local do i-ésimo indivíduo

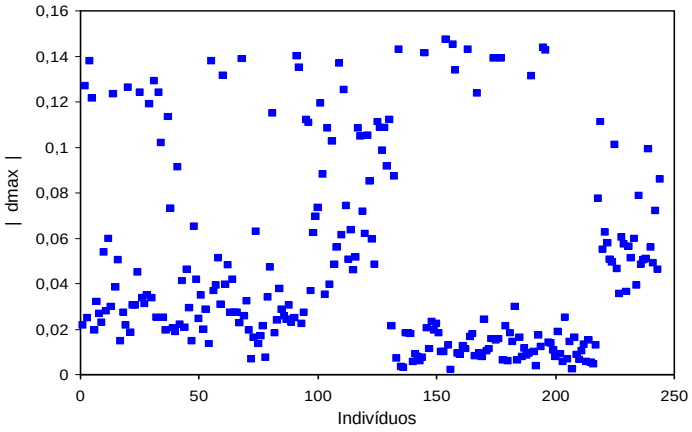

Figura 10 - Gráfico de influência - perturbação da covariável Rural

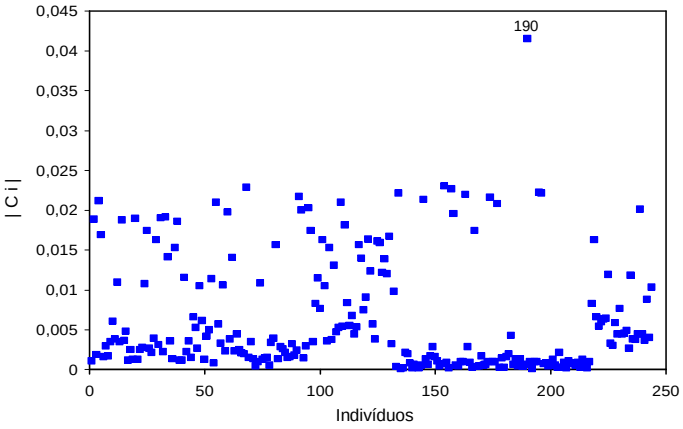

Figura 11 - Gráfico de influência local do i-ésimo indivíduo da covariável Rural

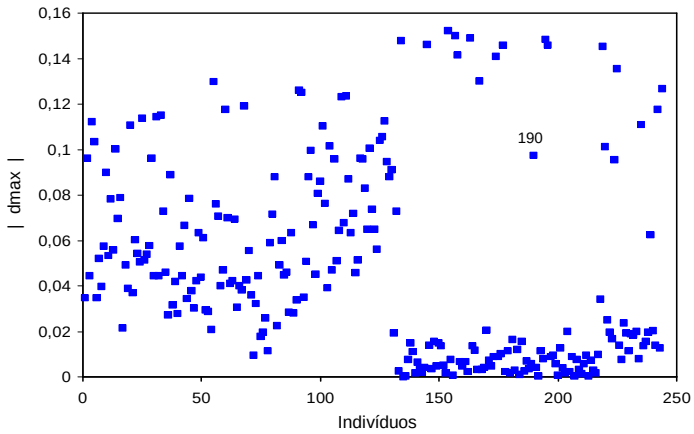

Figura 12 - Gráfico de influência - perturbação da covariável Mulher

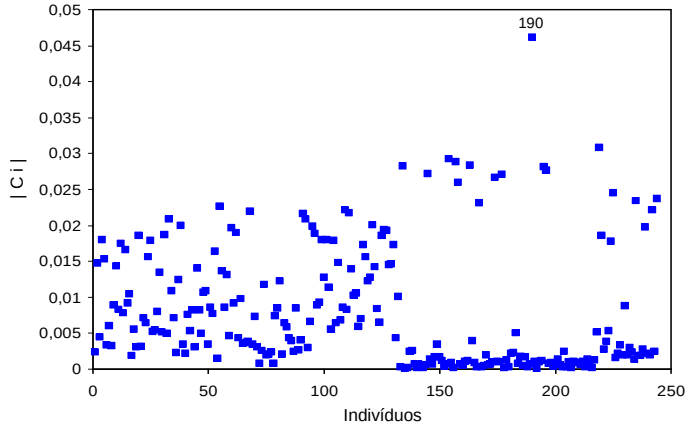

Figura 13 - Gráfico de influência local do i-ésimo indivíduo da covariável Mulher

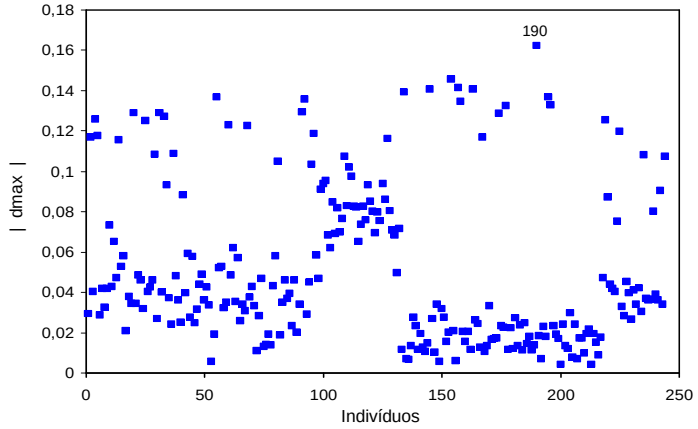

Figura 14 - Gráfico de influência - perturbação da covariável Rendtot-pai

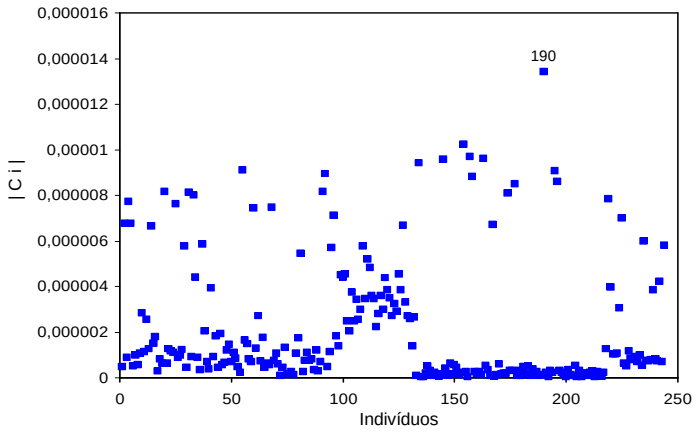

Figura 15 - Gráfico de influência local do i-ésimo indivíduo da covariável Rendtot-pai

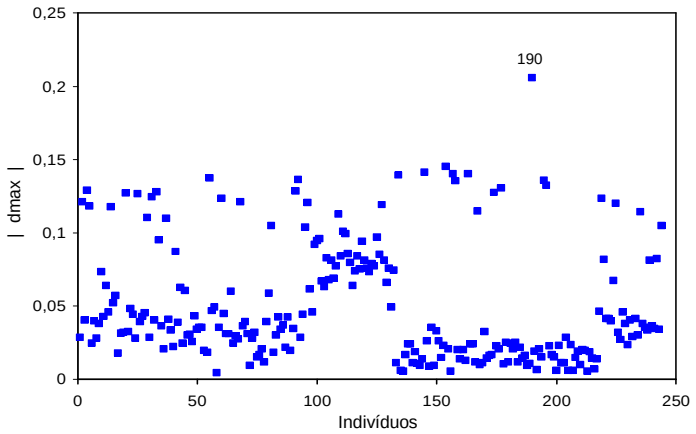

Figura 16 - Gráfico de influência - perturbação da covariável Rendtotal

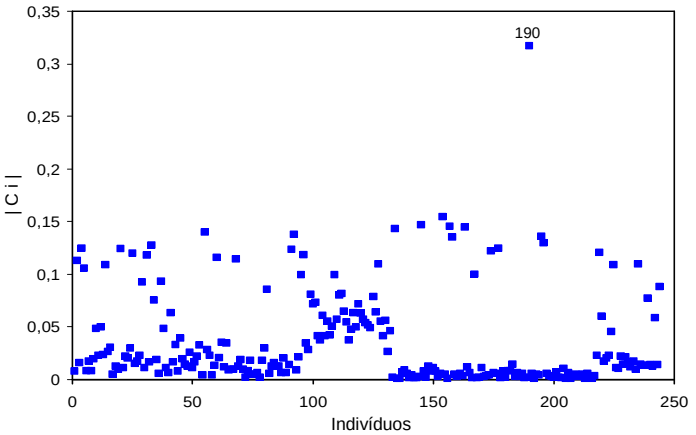

Figura 17 - Gráfico de influência local do i-ésimo indivíduo da covariável Rendtotal

3.1.4 Gráfico de envelopes

Observa-se que na Figura 18, todos os pontos caem dentro da banda de confiança, apesar de haver uma pequena separação em dois grupos e que o indivíduo 190 e 154 aparecem distante dos demais.

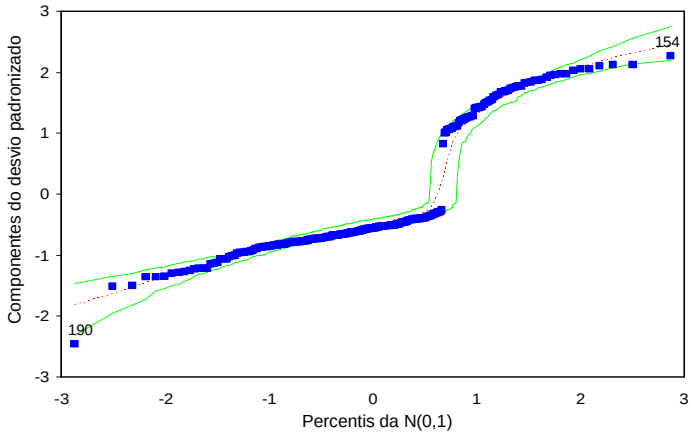

Figura 18 - Gráfico de envelopes para a componente do desvio

3.1.5 Reanálise dos dados

Para reanálise dos dados, são retirados os possíveis pontos discrepantes 154 e 190. Os resultados da reanálise são apresentados na tabela 8.

Tabela 8 - Estatísticas da Razão da verossimilhança, Escore e Wald

Estatísticas	p-Valor
Razão da verossimilhança (Λ)= 39,6637	<,0001
Escore (Es)= 39,2521	<,0001
Wald (W)= 29,90624	0,0029

Observando a tabela 8, verifica-se claramente uma significância maior para rejeitar a hipótese nula.

Na tabela 9, são apresentadas as estimativas de máxima verossimilhança.

Tabela 9 - Estimativas dos parâmetros

Efeito	Parâmetro	Estimativa	Erro padrão	Estatística Wald	p-Valor
Intercepto	β_0	-4,1195	4,9388	0,6957	0,4042
idade	β_1	0,3041	0,3466	0,7695	0,3804
anoestu	β_2	0,0765	0,1004	0,5800	0,4463
rural	β_3	1,2704	0,3781	11,2886	0,0008
mulher	β_4	-1,1592	0,3560	10,6052	0,0011
idademae	β_5	-0,0382	0,0305	1,5626	0,2113
rendtotmae	β_6	-0,00119	0,000751	2,5298	0,1117
educamae	β_7	-0,0227	0,0561	0,1632	0,6862
idadepai	β_8	0,00251	0,0184	0,0186	0,8916
rendtotpai	β_9	-0,00142	0,000611	5,3749	0,0204
educapai	β_{10}	-0,0976	0,0613	2,5396	0,1110
numpes	β_{11}	-0,0163	0,1193	0,0187	0,8912
srendtotal	β_{12}	0,00149	0,000553	7,2749	0,0070

$Deviance = 231,40$ com 229 g.l.

Nesta tabela pode-se claramente verificar que além das variáveis rural e mulher serem significativas para o modelo, também as variáveis rendtotpai e srendtotal passaram a ser significativas. Também, verifica-se significativa em relação a *deviance*, indicando um melhor ajuste do modelo.

Tabela 10 - Estimativas das razões de chances

		Limite de Confiança	
Efeito	Ponto estimado	Inferior	Superior
idade	1.355	0.687	2.674
anoestu	1.079	0.887	1.314
rural	**3.562**	1.698	7.475
mulher	**0.314**	0.156	0.630
idademae	0.963	0.907	1,022
rendtotmae	0.999	0.997	1,000
educamae	0.978	0,876	1.091
idadepai	0.997	0.962	1.034
rendtotpai	0.999	0.997	1,000
educapai	0.907	0.804	1.023
numpes	0.984	0.779	1.243
srendtotal	1,001	1,000	1,003

Através da observação da tabela 10 referente a estimativa da razão de chances, percebe-se que na variável rural o fator de risco teve um aumento e a variável mulher continua sendo um fator de proteção em relação a variável resposta. Observa-se também que a chance de um indivíduo da zona rural trabalhar é 3,5 vezes maior em relação a zona urbana.

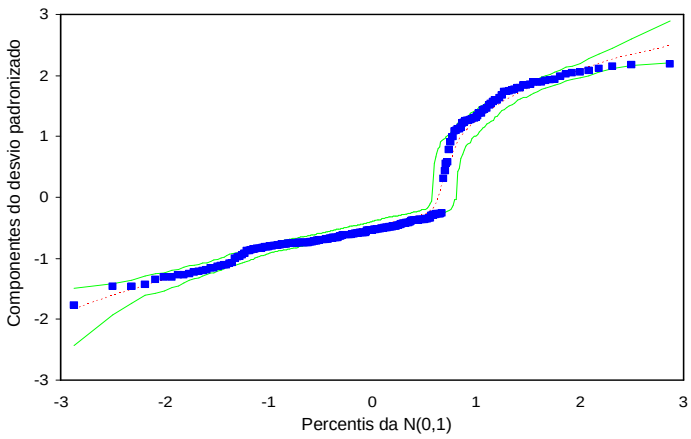

Figura 19 - Gráfico de envelopes para a componente do desvio

Em relação ao gráfico de envelopes, nota-se uma melhor distribuição das observações dentro da banda de confiança, sugerindo assim que o ajuste do modelo é melhor.

3.2 Aplicação 2

3.2.1 Introdução

Os dados utilizados foram cedidos por Paula Roberta Mendes e coletados em clínicas veterinárias da cidade de Lavras, estado de Minas Gerais. Segundo Mendes, as fichas de atendimento foram previamente avaliadas, registrando 176 animais, porém devido a observações incompletas, foram consideradas neste trabalho, 151 observações. Nesta aplicação vamos ajustar um modelo de regressão logística para prever a probabilidade de óbito de cães acometidos por gastroenterite hemorrágica.

A Gastroenterite Hemorrágica é uma patologia canina de aparecimento súbito. Os sintomas clínicos mais significantes deste tipo de gastroenterite são vómitos e/ou diarréia podendo conter sangue. O sangue pode apresentar-se sob duas formas, sendo em natureza (vermelho vivo) ou digerido (vermelho escuro a acastanhado). Pode ter etiologia viral, bacteriana ou parasitária. Além disso, sabe-se que fatores importantes associados devem ser considerados, como idade, raça, porte (peso), estresse ambiental e condições climáticas (COSTA, 1997). O diagnóstico é feito por exclusão de partes, tendo primeiramente que ser consideradas outras causas e patologias de diarréia com sangue, ou seja úlceras, trauma, tumores ou obstruções gastrointestinais, corpos estranhos, doenças infecciosas e desordens de coagulação. Para avaliação destas outras causas podem ser necessários testes laboratoriais como por exemplo: Hemograma completo, urianálise, radiografias, provas de coagulação e endoscopia ao aparelho gastrointestinal.

As variáveis utilizadas nesta aplicação foram:

y_i: Condição final do animal após o tratamento. (0 =não morreu, 1=morreu) (**obito**);

x_{i1}: Sexo do animal (0=fêmea, 1=macho) (**sexo**);

x_{i2}: Idade do animal contabilizada a cada seis meses, (1=cães com menos de seis meses, 2=cães com sete à doze meses, e assim sucessivamente) (**idade**);

x_{i3}: Quantidade de dias que o animal ficou internado (**diaria**);

x_{i4}: Número de vezes que o animal foi consultado na clínica (**atendime**).

Tabela 11 - Distribuição dos animais após o tratamento conforme o desfecho deste estudo

Óbito	Total de frequência	Porcentagem(%)
0 (não)	108	71,52
1 (sim)	43	28,48
Total	151	100

Na análise exploratória dos dados pode-se perceber, segundo a tabela 11, que a variável resposta **obito**, é a condição final do animal após o tratamento, sendo codificada como: $1 = \text{sim} \quad \text{e} \quad 0 = \text{não}$. Dos resultados obtidos, tem-se que dos 151 animais, 43 foram ao óbito, ou seja, 28,48%.

Ajustando um modelo de regressão logística e testando as hipóteses

$$H_0 : \boldsymbol{\beta} = \mathbf{0}$$

$$H_1 : \boldsymbol{\beta} \neq \mathbf{0}$$

tem-se que as estatísticas são dadas pelos resultados apresentados na tabela 12.

Tabela 12 - Estatísticas da Razão da verossimilhança, Escore e Wald

Estatísticas	p-Valor
Razão da verossimilhança(Λ)= 8,2790	0,1025
Escore (Es)= 8,0931	0,0882
Wald (W)= 7,2565	0,1229

Da tabela 12 pode-se inferir que não foi rejeitada a hipótese nula, considerando um nível de 5% de significância, isto é, nenhuma variável é significativa para o modelo proposto. Entretanto, decidiu-se pela continuidade das análises.

As estimativas de máxima verossimilhança são observadas na tabela 13, na qual pode ser verificado que considerando um nível de 5%, nenhuma variável é significativa e mediante a *deviance* observada conclui-se que o modelo não está bem ajustado.

Na tabela 14, através das estimativas das razões de chances, percebe-se que a variável sexo é um fator de risco e a variável diária um fator de proteção em relação a variável óbito dos

Tabela 13 - Estimativas dos parâmetros

Efeito	Parâmetro	Estimativa	Erro padrão	Estatística Wald	p-Valor
Intercepto	β_0	-1,5284	0,4342	12,3874	0,0004
sexo	β_1	0,5683	0,3742	2,3063	0,1289
idade	β_2	-0,0143	0,0150	0,9193	0,3377
diaria	β_3	-0,0904	0,1186	0,58070	0,4461
atendime	β_4	0,2866	0,1563	3,3617	0,0667

$Deviance = 172,136$ com 146 g.l.

Tabela 14 - Estimativas das razões de chances

		Limite de Confiança	
Efeito	Ponto estimado	Inferior	Superior
sexo	**1,765**	0,848	3,676
idade	0,986	0,957	1,015
diaria	**0,914**	0,724	1,153
atendime	1,332	0,980	1,809

animais, sendo que a chance de um animal ser macho e vir a falecer é de 1,765. Entretanto, deve-se ter cuidado com estas interpretações, uma vez que o modelo não está bem ajustado.

3.2.2 Medidas de resíduos e diagnóstico

Através dos gráficos das medidas de resíduo e diagnóstico verifica-se os possíveis pontos discrepantes. Na figura 20, correspondente ao resíduo de Pearson (rp_i), contra a ordem das observações, verifica-se que a observação 19 destaca-se dentre as outras; porém, observando o resíduo de *deviance*, na figura 21, percebe-se que não há nenhum ponto discordante. Em relação a medida do *leverage* ($\hat{h}_{ii}$), conforme a figura 22, verifica-se que as observações 29, 51, 65 e 76 destacam-se das demais. A figura 23, correspondente a medida C, indica as observações 11, 17, 19, 23, 67 e 76 como possíveis pontos discrepantes; analogamente a medida $Cbar$, na figura 24, aponta as observações 11, 17, 19, 23, 67 e 76 como pontos discrepantes. Na figura 25, correspondente a medida $DIFCHISQ$, nota-se que as observações 11, 19 e 67 se destacam das demais. Na figura 26, referente a medida $DIFDEV$, novamente verifica-se que as observações 11, 19 e 67 podem ser

consideradas como possíveis pontos discrepantes.

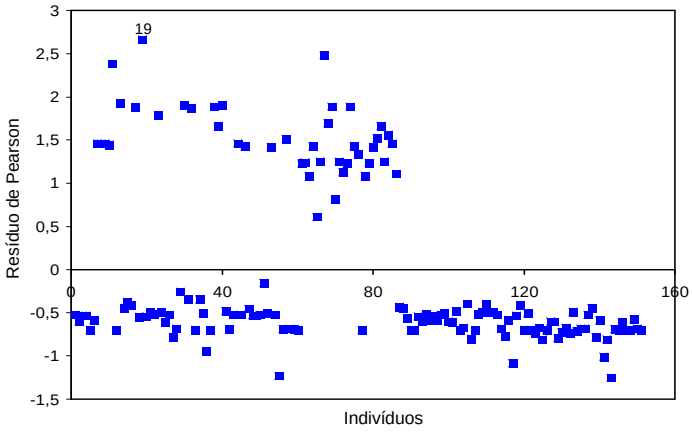

Figura 20 - Gráfico do Resíduo de Pearson

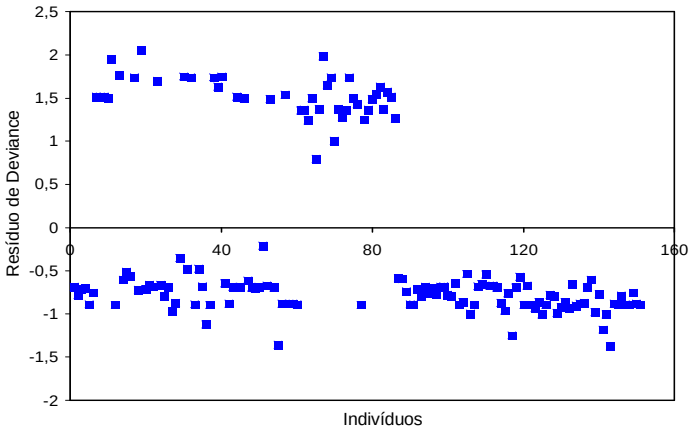

Figura 21 - Gráfico do Resíduo de Deviance

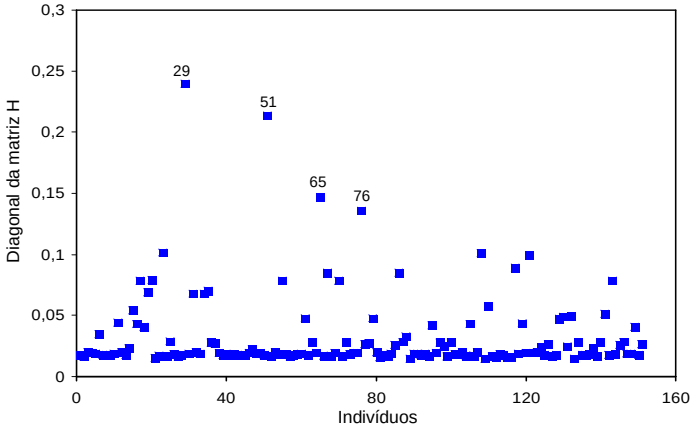

Figura 22 - Gráfico da diagonal da matriz H

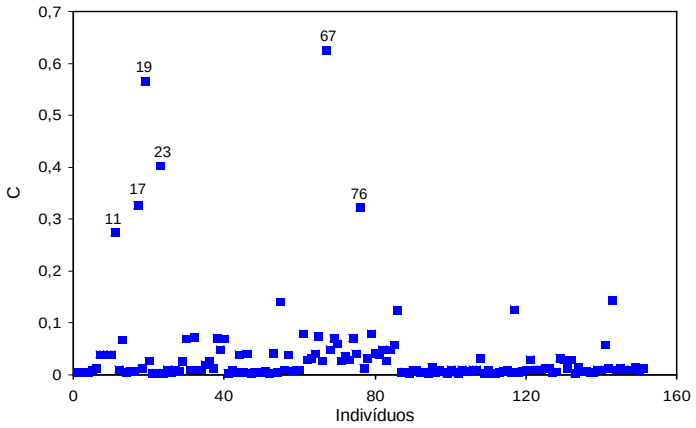

Figura 23 - Gráfico de C

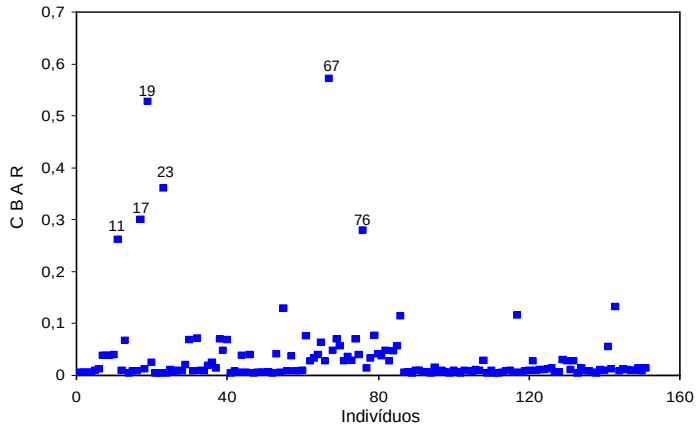

Figura 24 - Gráfico de CBAR

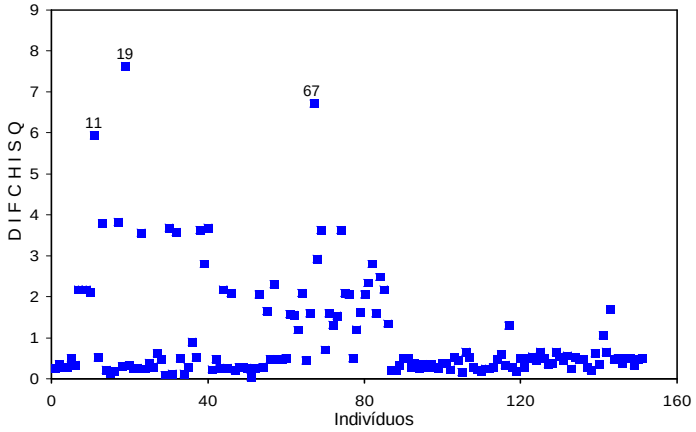

Figura 25 - Gráfico do DIFCHISQ

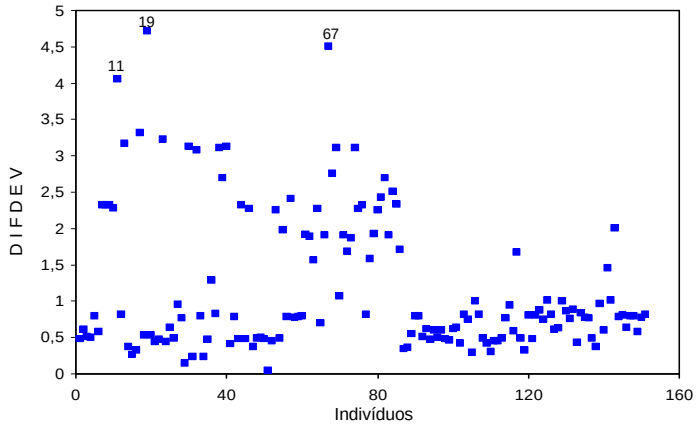

Figura 26 - Gráfico do DIFDEV

3.2.3 Influência local

Conforme o esquema de perturbação de casos, temos que: $C_{d_{max}} = 2.746262$.

Na Figura 27, é apresentado o gráfico do autovetor correspondente a $C_{d_{max}}$ e as observações 11, 17, 19 e 76 são as que se destacam das demais. Já na Figura 28, referente a influência local total as observações que se destacam são 11, 17, 19, 23, 67 e 76.

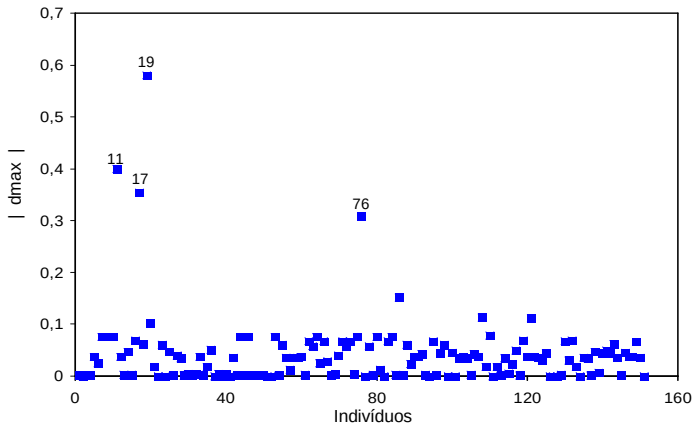

Figura 27 - Gráfico de influência - ponderação de casos

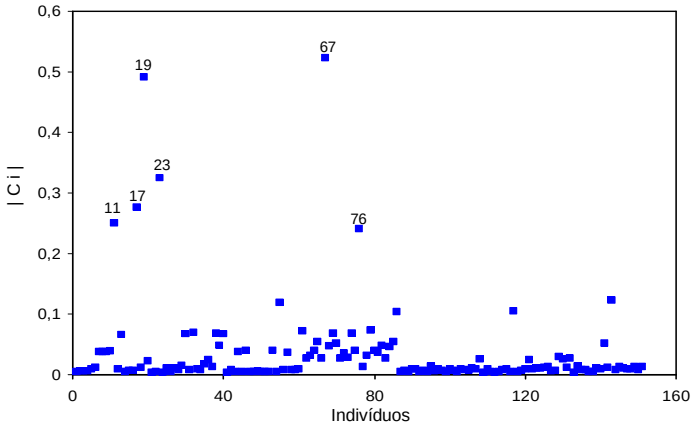

Figura 28 - Gráfico de influência local do i-ésimo indivíduo

3.2.4 Gráfico de envelopes

Nesta parte é apresentado o gráfico de envelopes. Na Figura 29, verifica-se que todos os pontos caem dentro da banda de confiança, apesar de haver uma pequena separação em dois grupos e que os indivíduos 11, 17, 19 e 76 aparecem distante dos demais.

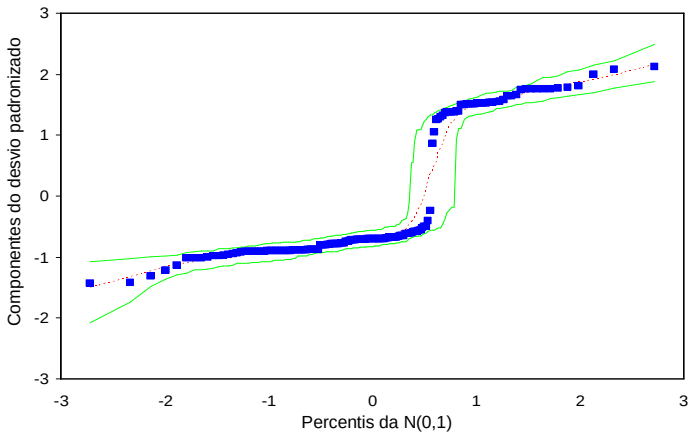

Figura 29 - Gráfico de envelopes para a componente do desvio

3.2.5 Reanálise dos dados

Para reanálise dos dados são retirados os possíveis pontos influentes 11, 17, 19 e 76. Os resultados da reanálise são apresentados na tabela 15.

Tabela 15 - Estatísticas da Razão da verossimilhança, Escore e Wald

Estatísticas	p-Valor
Razão da verossimilhança (Λ)= 19,0334	0,0008
Escore (Es)= 15,3500	0,0040
Wald (W)= 12,1444	0,0163

Observando a tabela 15, verifica-se claramente uma significância alta para rejeitar a hipótese nula, o que significa que pelo menos uma das covariáveis é significativa.

Na tabela 16 são apresentadas as estimativas de máxima verossimilhança.

Tabela 16 - Estimativas dos parâmetros

Efeito	Parâmetro	Estimativa	Erro padrão	Estatística Wald	p-Valor
Intercepto	β_0	-1,5235	0,4584	11,0481	0,0009
sexo	β_1	0,7562	0,4028	3,5250	0,0604
idade	β_2	-0,0147	0,0145	1,0273	0,3108
diaria	β_4	-0,7088	0,3055	5,3846	0,0203
atendime	β_5	0,2894	0,1729	2,8025	0,0941

$Deviance = 151,056$ com 142 g.l.

Nota-se que considerando um nível de 5% a variável diária passa ser significativa. A variável sexo e atendime passaria a ser significativo considerando um nível de 7% e 10% respectivamente. Verifica-se também, que a *deviance* diminuiu, indicando um bom ajuste do modelo.

A tabela das razões de chances estimadas é dada por:

Tabela 17 - Estimativas das razões de chances

		Limite de Confiança	
Efeito	Ponto estimado	Inferior	Superior
sexo	**2,130**	0,967	4,691
idade	0,985	0,958	1,014
diaria	**0,492**	0,270	0,896
atendime	1,336	0,952	1,874

Na tabela (17) percebe-se que a variável sexo continua sendo um fator de risco e a variável diária um fator de proteção em relação a variável óbito dos animais, sendo que a chance de um animal macho vir a falecer aumentou para 2,130.

No gráfico de envelopes nota-se uma melhor distribuição das observações dentro da banda de confiança, sugerindo ser um ajuste adequado.

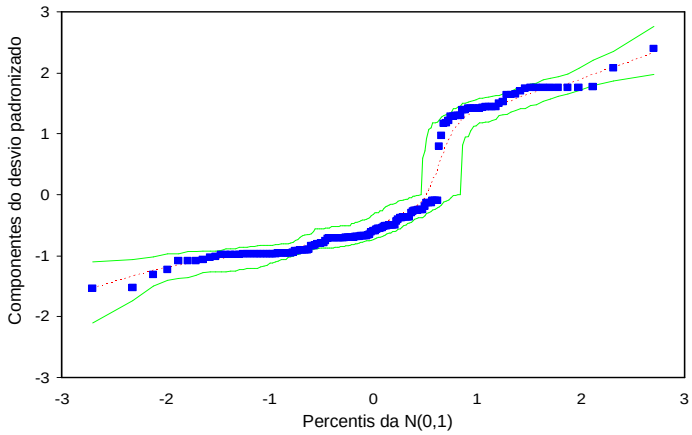

Figura 30 - Gráfico de envelopes para a componente do desvio

4 CONSIDERAÇÕES FINAIS

Neste trabalho discutiu-se a aplicação da teoria de influência local, proposta por Cook (1986), no modelo de regressão logística. Foram obtidas matrizes necessárias para a aplicação da técnica, considerando dois tipos de perturbação nos elementos dos dados e do modelo. Aplicando-se estes resultados em um conjunto de dados, obteve-se indicações de quais observações ou conjunto de observações influenciam de maneira sensível os resultados da análise. Este fato foi ilustrado através de dois conjuntos de dados reais, sendo verificado que para alguns esquemas de perturbação, a presença de algumas observações pode modificar consideravelmente os níveis de significância de certas covariáveis.

Finalmente, os resultados das aplicações indicam que o uso da técnica de influência local no modelo de regressão logística é útil na deleção de possíveis pontos influentes. Assim, a técnica de influência local pode ser considerada como uma análise complementar em relação às medidas de diagnóstico, propostas por Pregibon (1981).

4.1 Pesquisas futuras

Algumas das investigações que podem ser de interesse para ampliar e dar continuidade aos resultados obtidos são:

1. O desenvolvimento e implementação de técnicas de diagnóstico para avaliar a qualidade do ajuste dos modelos de regressão logística com efeito aleatório.

2. Um outro problema comum que ocorre em modelos de regressão é a existência de covariáveis medidas com erro, causado por, entre outros motivos, pela inexatidão da medida que pode ser resultado de uma opinião subjetiva ou de uso de instrumentos de precisão limitada, assim em uma pesquisa futura pode ser estudada uma técnica de influência local nos modelos de regressão logística com erros nas variáveis.

REFERÊNCIAS

ALLISON, P. D.; **Logistic regression using the SAS System, theory and application.** SAS Institute, 1999. 304 p.

ATKINSON, A. C.; **Plots, transformations and regression: an introduction to graphical methods of diagnostic regression analysis.** Oxford: Oxford Statistical Science Series, 1987. 280 p.

CHRISTENSEN, R.; **Log-linear models & Logistic Regression.** New York: Springer-Verlag, 1997. 500 p.

COLLET, D.; **Modelling binary data.** London: Chapman and Hall, 1991. 369 p.

COOK, R. R.; Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, London, n.48, p.133-169, 1986.

COOK, R. R.; Detection of influential observations in linear regression. **Technometrics**, Wisconsin, v.19, p.15-118, 1977.

CORDEIRO, G. M.; NETO, E. A. L.; **Modelos paramétricos.** (Livro texto de minicurso da 16º SINAPE) Caxambu-MG: ABE , 2004. 246 p.

COSTA, S. C.; **Regressão Logística aplicada na identificação de fatores de risco para doenças em animais domésticos.** 1997. 104 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1997.

COX, D. R.; HINKLEY, D. V.; **Theoretical statistics.**, London: Chapman & Hall, 1986. 174 p.

CRAMER, J. S.; **Logit models from economics and other fields.** Cambridge: Cambridge University, 2003. 184 p.

DEAN, C. B.; Testing for overdispersion in Poisson and binomial regression models. **Journal the American Statistical Association**,Alexandria, 1992, v.87, n.418, p.451-457.

DEMÉTRIO, C.G.B.; **Modelos lineares generalizados em experimentação agronômica.** Piracicaba: CALQ, Departamento Editorial, 2002. 113p.

DOBSON, A. J.; **An Introduction to generalized linear models.** London: Chapman & Hall, 2001. 225 p.

FARHAT, C. A. V.; **Análise de diagnóstico em regressão logística.** 2003. 113 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2003.

GALEA, M.; PAULA, G.A.; BOLFARINE, H.; Local influence in elliptical linear regression models. **The Statistician**, Oxford: v.46, p.71-79, 1997.

HINDE, I.; DEMÉTRIO, C.; **Overdispersion models and estimation**. (Livro texto de minicurso da 13º SINAPE), Caxambu-MG: ABE, 1998. 73 p.

HOSMER, D.W.; LEMESHOW, S.; **Applied logistic regression.** New York: John Wiley, 1989, 307 p.

HOSSAIN, M.; ISLAM, M. A.; Application of local influence to the linear logistic regression models. **Journal Statistical Science**, Dhaka: v.51, n.2, p.269-278, 2003.

KLEINBAUM, D. G.; **Logistic regression: a self-learning text.** New York: Springer-Verlac, 1994. 278 p.

LESAFFRE, E.; VERBEKE, G.; Local influence in linear mixed models. **Biometrics**, Washington: v.54, p.570-582, 1998.

LU, W.; Testing extra-binomial variations. **The Journal of Statistical Computation and Simulation.**, Virginia: v.63, n.1, p.93-103, 1999.

McCULLAGH, P.; NELDER, J. A.; **Generalized linear models.**, London: Chapman & Hall, 1989. 511 p.

MONTGOMERY, D. C.; PECK, E. A.; **Introduction to linear regression analysis.**, New York: John Wiley, 1992. 527 p.

NELDER, J. A.; WEDDERBURN, R. W.M; Generalized linear models. **Journal of the Royal Statistical Society**, London, v.135, p.370-384, 1972.

PAULA, G.A.; **Modelos de regressão com apoio computacional.** São Paulo: IME-USP, 2004. 245 p.

PREGIBON, D.; Logistic regression diagnostics. **Annals of Statistics.**, Minneapolis: v.9, p.705-724, 1981.

SILVA, G. L.; **Modelos Logísticos para dados binários.** 1992. 118 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1992.

THOMAS, W.; COOK, R. D.; Assessing influence on predictions from generalized linear models. **Technometrics**, Alexandria, v.32, p.59-65, 1990.

BIBLIOGRAFIA CONSULTADA

AGRESTI, A.; **An Introduction to Categorical Data Analysis.** New York: John Wiley, 1990. 290 p.

COX, D. R.; SNELL, E. J.; A general definition of residuals (with discussion). **Journal of the Royal Statistical Society**, v.30, p.248-275, 1968.

COX, D. R.; SNELL, E. J.; **Analysis of Binary Data.**, London: Chapman & Hall, 1989, 236 p.

CORDEIRO, G. M.; NETO, E. A. L.; **Modelos Paramétricos.** SINAPE, Caxambu-MG, 2004. 246 p.Livro texto de minicurso da 16º

CYSNEIRO, F. J.; PAULA, G.A.; GALEA, M.; **M. Modelos Simétricos Aplicados.** Livro texto de minicurso da 9ª Escola de Modelos de Regressão, São Pedro-SP, 2005. 89 p.

ESCOLA SUPERIOR DE AGRICULTURA "LUIZ DE QUEIROZ. **Normas para elaboração de dissertações e teses.** 3.ed. Piracicaba: ESALQ – Divisão de Biblioteca e Documentação, 2005. 99 p.

GALEA, M.; BOLFARINE, H.; VILCA LABRA, F.; Influence diagnostics for the structural error-in-variables model under the Student-t distribution. **Journal of Applied Statistics**, Oxford: v.29, p.1191-1204, 2002.

ORTEGA, E.M.M.; BOLFARINE, H.; PAULA, G.A.; Influence diagnostics in generalized log-gamma regression models. **Computational Statistics e Data Analysis**, New York, v.42, p.165-186, 2003.

R. **The R Foundation for Statistical Computing Version 2.0.1** Disponível em: <http://www.cran.r-project.org>. Acesso: 15 nov. 2004.

SAS Institute Inc. **SAS/STAT 9.1 User´s Guide** Cary,NC, USA: SAS Institute Inc., 2004, 5136 p.

SILVA, G. L.; **Modelos Logísticos para dados Binários.** 1992, 118 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1992.

VENABLES, B.; KUHNERT, P.; **An Introduction to R: Software for Statistical Modelling & Computing.** Piracicaba, 2005, 261 p.

ANEXOS

ANEXO A - Listagem do programa para Análise de Diagnóstico.

```
Data educa;
 input obs idade anoestu rural mulher idade_mae rendtot_mae educa_mae
       idade_pai rendtot_pai educa_pai num_pes s_rendtotal trabalha;
 cards;
1   14  7   0   0   36  400     7   39  1200    8   4   1600    0
2   15  4   0   0   37  125     2   49  400     4   8   565     1
3   14  6   0   0   32  0       3   31  419     5   4   519     0
4   14  4   0   0   37  125     2   49  400     4   8   565     1
5   14  6   0   0   36  550     11  37  1000    4   5   1750    1
6   15  7   0   0   34  0       6   36  400     11  6   400     0
7   15  4   0   0   36  0       7   32  300     3   4   300     0
8   15  7   0   0   41  480     11  46  480     8   4   960     0
9   14  4   0   0   30  60      0   56  380     3   8   440     0
10  15  4   0   0   41  240     0   45  0       0   8   980     0
11  15  1   0   0   46  0       0   51  480     0   6   720     0
12  14  8   0   0   38  1500    15  39  700     4   5   2900    0
13  14  5   0   0   30  190     4   23  300     5   8   956     0
14  14  5   0   0   36  0       11  38  300     0   5   380     1
15  14  7   0   0   36  45      1   37  600     0   5   660     0
16  15  6   0   0   34  80      5   38  700     1   4   780     0
17  15  8   0   0   37  240     11  39  480     15  4   720     0
18  15  6   0   0   36  520     6   58  240     8   6   760     0
19  14  6   0   0   44  240     4   44  300     3   4   540     0
20  15  8   0   0   33  0       8   48  2500    11  5   2500    1
21  14  5   0   0   46  120     3   52  240     4   6   720     0
22  14  4   0   0   33  0       0   37  350     0   5   350     0
23  15  4   0   0   33  0       2   36  0       4   4   0       0
24  15  8   0   0   41  3000    15  49  3000    15  5   6000    0
25  15  7   0   0   40  1140    11  45  1350    11  4   2610    1
26  14  6   0   0   36  550     11  37  1000    4   5   1750    0
27  14  6   0   0   38  0       4   38  850     1   4   850     0
28  15  6   0   0   51  0       0   50  600     0   3   600     0
29  15  7   0   0   46  0       5   43  0       3   5   360     1
30  15  8   0   0   35  400     12  38  3080    11  5   3480    0
31  14  7   0   0   37  0       6   37  2500    6   4   2500    1
32  14  6   0   0   39  240     0   32  630     4   4   970     0
33  14  7   0   0   37  2000    11  39  2000    11  4   4000    1
34  15  8   0   0   34  240     5   24  0       3   5   480     1
35  15  3   0   0   43  120     1   27  600     5   8   1110    0
36  14  7   0   0   36  1000    15  45  900     11  4   1900    0
37  15  6   0   0   43  0       7   38  600     1   5   780     1
38  14  7   0   0   24  0       0   83  240     0   7   1510    1
39  14  4   0   0   33  0       5   39  100     4   6   220     0
40  15  6   0   0   37  240     11  37  480     11  4   720     0
41  15  8   0   0   34  0       6   40  877     1   6   1197    1
42  14  4   0   0   34  180     0   24  190     4   6   520     0
43  14  4   0   0   38  240     5   43  450     0   5   1295    0
44  15  8   0   0   38  0       7   47  1000    11  5   1000    0
45  14  5   0   0   24  0       3   24  1000    0   5   1000    0
46  14  6   0   0   39  1500    4   49  4000    11  4   5500    0
47  14  2   0   0   56  240     0   63  0       0   4   480     0
48  15  6   0   0   46  0       3   44  550     2   8   3050    1
49  15  8   0   0   47  0       11  50  0       0   3   0       0
50  14  5   0   0   34  250     6   34  500     4   5   750     0
```

51	15	8	0	0	40	0	10	61	300	4	6	300	0
52	14	0	0	0	34	1000	11	71	1050	11	6	3050	0
53	14	8	0	0	43	0	13	46	8000	11	4	8000	0
54	14	7	0	0	37	0	10	41	1400	11	4	1400	0
55	15	7	0	0	48	0	11	56	1500	11	5	1740	1
56	15	4	0	0	30	60	0	56	380	3	8	440	0
57	15	6	0	0	44	0	0	49	240	0	5	240	0
58	15	6	0	0	43	0	4	44	1500	4	5	4500	1
59	15	8	0	0	38	0	5	38	1600	8	4	1600	0
60	14	7	0	0	35	1000	5	42	850	11	5	2150	1
61	15	5	0	0	34	240	3	30	600	3	7	840	0
62	15	8	0	0	41	0	10	53	5000	4	4	7000	1
63	14	7	0	0	34	600	13	42	360	6	4	960	0
64	14	5	0	0	41	0	7	48	800	0	7	1580	0
65	15	8	0	0	48	1540	15	46	2080	15	5	4100	0
66	15	8	0	0	33	480	11	37	440	11	4	920	0
67	15	8	0	0	34	0	7	46	700	11	5	700	0
68	14	0	0	0	37	120	0	54	300	2	4	540	1
69	14	7	0	0	35	240	6	36	600	6	4	955	0
70	15	8	0	0	52	0	1	56	500	2	4	500	0
71	14	7	0	0	44	200	4	52	400	6	5	840	0
72	14	7	0	0	44	0	15	41	750	14	4	750	0
73	14	3	0	0	38	0	3	36	800	4	5	800	0
74	15	10	0	0	46	480	4	48	600	4	4	3030	1
75	14	7	0	0	46	0	10	49	3500	11	4	3500	0
76	14	7	0	0	36	700	12	38	3900	15	4	4600	0
77	14	6	0	0	34	900	15	35	3000	12	4	3900	0
78	14	6	0	0	37	240	11	39	480	15	4	720	0
79	15	4	0	0	41	0	8	44	442	2	7	682	0
80	15	6	0	0	35	150	5	34	500	3	4	900	0
81	15	7	0	0	36	240	3	42	700	7	7	1430	1
82	14	7	0	0	38	1500	11	43	2500	15	4	4000	0
83	15	7	0	0	50	0	0	56	450	4	7	450	0
84	15	7	0	0	41	360	6	49	500	4	4	860	0
85	15	5	0	0	35	240	11	40	500	8	6	1040	0
86	14	7	0	0	38	240	4	42	350	4	4	590	0
87	15	8	0	0	37	1030	15	41	1965	15	3	2995	0
88	14	6	0	0	30	50	3	35	400	3	8	558	0
89	15	7	0	0	36	813	15	49	2754	15	4	3567	0
90	15	8	0	0	45	0	11	54	900	12	5	1740	0
91	14	6	0	0	42	7	0	43	700	7	4	827	1
92	15	7	0	0	39	0	11	41	1680	11	5	1680	1
93	15	6	0	0	38	0	5	37	1150	9	4	1150	0
94	14	4	0	0	42	335	0	54	315	0	3	650	0
95	15	5	0	0	43	45	12	27	200	4	10	1335	1
96	15	7	0	0	37	2000	11	39	2000	11	4	4000	1
97	14	5	0	0	44	0	0	47	480	0	7	1230	0
98	15	2	1	0	48	0	4	33	253	8	4	253	0
99	14	6	1	0	60	255	0	56	240	0	4	495	1
100	14	7	1	0	42	0	4	42	325	6	4	325	1
101	15	5	1	0	39	0	5	43	200	2	5	450	0
102	14	7	1	0	35	300	3	36	500	8	5	800	0
103	15	7	1	0	46	0	0	48	100	0	5	100	1
104	14	9	1	0	38	0	5	44	480	4	6	630	0
105	15	4	1	0	43	50	2	41	440	0	5	490	1
106	14	7	1	0	41	25	1	42	390	2	5	415	0

107	14	5	1	0	27	0	2	34	248	3	6	248	1
108	14	7	1	0	35	0	4	43	0	4	4	0	1
109	15	8	1	0	42	349	7	46	1200	4	5	2249	0
110	14	5	1	0	33	0	4	34	150	4	6	150	1
111	15	6	1	0	34	0	1	36	150	1	6	150	0
112	14	5	1	0	35	0	11	37	280	6	5	280	1
113	15	6	1	0	35	0	11	37	280	6	5	280	1
114	14	3	1	0	44	0	0	50	240	0	4	240	1
115	14	4	1	0	37	0	0	41	150	0	8	450	1
116	14	5	1	0	36	45	0	44	420	2	8	585	1
117	15	7	1	0	32	0	2	37	800	8	4	800	0
118	14	6	1	0	37	0	8	37	900	2	6	900	0
119	14	7	1	0	34	0	5	37	430	8	5	430	1
120	14	0	1	0	36	0	0	50	980	0	7	980	1
121	15	7	1	0	53	50	0	68	410	2	6	460	0
122	14	0	1	0	41	0	0	41	400	0	4	400	0
123	14	4	1	0	44	0	0	52	300	0	5	315	1
124	15	6	1	0	41	0	3	49	240	4	5	240	1
125	14	4	1	0	35	200	0	50	300	0	3	500	0
126	15	2	1	0	31	0	1	40	360	3	7	360	0
127	14	5	1	0	40	300	11	46	500	11	5	800	1
128	15	6	1	0	37	150	4	37	720	6	4	870	0
129	14	6	1	0	30	0	6	37	200	6	8	200	0
130	14	7	1	0	34	360	11	37	2800	6	4	3160	0
131	15	8	1	0	36	0	0	50	980	0	7	980	1
132	14	4	1	0	42	0	4	47	360	4	4	640	0
133	14	8	0	1	44	0	6	46	1000	4	4	1000	0
134	14	7	0	1	41	240	11	58	600	2	5	1230	1
135	14	4	0	1	42	390	10	44	320	8	4	710	0
136	14	6	0	1	33	0	11	35	460	11	6	460	0
137	15	8	0	1	42	0	10	61	3500	4	3	3500	0
138	15	9	0	1	33	315	4	61	0	4	4	315	0
139	15	7	0	1	36	240	4	42	1100	4	5	1580	0
140	14	5	0	1	35	0	5	37	370	5	4	370	0
141	14	5	0	1	33	340	0	44	200	2	5	540	0
142	15	7	0	1	36	0	8	41	240	8	5	240	0
143	14	6	0	1	34	260	11	39	320	7	4	580	0
144	14	7	0	1	39	0	6	35	400	5	8	850	0
145	15	8	0	1	35	200	7	37	1000	6	6	1250	1
146	15	8	0	1	45	0	6	45	500	0	5	850	0
147	15	8	0	1	38	2000	15	45	2000	15	4	4000	0
148	15	8	0	1	33	304	3	37	500	5	5	1424	0
149	15	8	0	1	41	2300	15	42	6300	15	5	8600	0
150	15	5	0	1	36	240	4	38	300	0	4	900	0
151	15	8	0	1	38	240	2	48	240	3	5	720	0
152	15	5	0	1	38	0	3	36	800	4	5	800	0
153	14	6	0	1	41	0	6	38	600	7	4	1800	0
154	15	6	0	1	34	150	6	39	960	10	5	1260	1
155	15	4	0	1	33	0	3	36	350	2	4	350	0
156	14	7	0	1	40	0	3	41	800	15	6	1100	0
157	15	6	0	1	36	240	0	37	960	7	5	1300	1
158	15	8	0	1	36	550	11	37	1000	4	5	1750	1
159	14	5	0	1	37	0	0	38	240	0	6	420	0
160	14	4	0	1	34	413	8	42	300	2	5	713	0
161	15	6	0	1	47	0	0	39	200	0	3	215	0
162	14	7	0	1	46	2000	15	50	3000	15	8	6140	0

163	15	5	0	1	36	240	5	29	700	4	4	940	1
164	14	6	0	1	30	200	11	69	240	0	5	680	0
165	15	8	0	1	39	350	3	44	760	4	5	1350	0
166	15	6	0	1	37	0	7	44	600	7	4	600	0
167	15	8	0	1	38	240	5	43	450	0	5	1295	1
168	15	8	0	1	34	280	11	37	1750	11	5	2030	0
169	14	7	0	1	37	0	11	42	0	3	5	0	0
170	15	7	0	1	38	45	4	41	400	0	7	925	0
171	15	5	0	1	34	0	4	38	500	4	7	500	0
172	14	5	0	1	35	570	11	37	500	3	4	1270	0
173	15	8	0	1	37	0	11	48	1400	4	4	1400	0
174	15	4	0	1	43	0	0	58	300	0	8	980	1
175	15	5	0	1	34	240	5	24	0	3	5	480	0
176	15	8	0	1	29	240	6	35	240	6	4	510	0
177	15	6	0	1	38	0	4	42	480	4	5	1080	1
178	14	6	0	1	29	250	8	35	400	8	5	650	0
179	15	8	0	1	40	0	8	43	2000	3	4	2340	0
180	14	4	0	1	32	0	6	39	700	5	4	730	0
181	15	8	0	1	44	80	0	48	300	0	7	580	0
182	14	7	0	1	51	240	5	50	4000	4	5	4960	0
183	15	9	0	1	41	2500	15	25	2500	11	4	5600	0
184	15	0	0	1	38	0	8	45	500	5	4	740	0
185	15	7	0	1	39	0	0	40	300	0	7	300	0
186	15	8	0	1	39	0	4	41	430	11	5	910	0
187	15	7	0	1	59	0	4	49	240	0	7	700	0
188	14	7	0	1	36	1000	11	44	1000	8	6	2000	0
189	14	8	0	1	41	400	12	45	150	3	6	550	0
190	15	7	0	1	33	600	11	36	1700	11	9	8700	0
191	15	3	0	1	48	240	0	37	600	3	7	1440	0
192	14	6	0	1	39	100	11	39	300	9	4	400	0
193	15	8	0	1	30	300	8	38	300	5	4	600	0
194	15	8	0	1	43	240	7	71	240	4	4	480	0
195	14	6	0	1	39	0	4	40	240	0	4	480	1
196	14	6	0	1	36	240	0	48	150	0	4	530	1
197	14	6	0	1	36	300	7	44	300	1	5	960	0
198	15	6	0	1	41	0	8	44	442	2	7	682	0
199	15	7	0	1	41	0	4	44	300	4	4	300	0
200	15	8	0	1	42	500	10	54	4890	15	6	5390	0
201	15	8	0	1	43	300	2	43	840	1	4	1140	0
202	15	7	0	1	48	480	5	26	180	5	5	660	0
203	14	7	0	1	34	0	4	37	400	7	4	400	0
204	15	7	0	1	36	400	8	38	800	2	8	1700	0
205	15	9	0	1	43	1300	15	51	1100	15	5	2400	0
206	15	7	0	1	41	165	0	44	250	3	4	615	0
207	14	6	0	1	44	0	3	45	500	8	7	500	0
208	15	8	0	1	37	1350	15	44	700	8	3	2050	0
209	14	2	0	1	36	0	4	43	800	0	4	1040	0
210	14	8	0	1	48	600	11	24	1320	12	10	2950	0
211	14	6	0	1	35	30	4	39	800	4	7	1380	0
212	14	7	0	1	30	0	7	36	240	1	5	270	0
213	14	9	0	1	43	700	11	42	4500	15	4	5200	0
214	14	8	0	1	39	1240	6	40	900	4	3	2140	0
215	14	4	0	1	31	0	0	52	200	4	5	200	0
216	14	7	0	1	42	300	11	43	0	7	5	300	0
217	15	9	0	1	40	0	10	61	300	4	6	300	0
218	15	7	1	1	37	0	8	37	900	2	6	900	0

```
219 15  0   1   1   46  0       1   49  270     1   4   270     1
220 15  7   1   1   32  30      2   35  800     0   4   950     1
221 15  7   1   1   36  0       3   42  240     5   6   240     0
222 15  2   1   1   39  240     2   50  240     2   5   480     0
223 14  6   1   1   55  0       0   67  200     0   7   640     0
224 15  5   1   1   26  400     3   42  600     3   6   1700    1
225 14  7   1   1   37  0       4   45  500     3   6   500     1
226 14  7   1   1   37  0       4   45  600     5   6   600     0
227 14  5   1   1   48  240     0   66  240     4   5   480     0
228 14  3   1   1   30  0       3   35  240     0   6   240     0
229 15  6   1   1   37  0       5   37  240     5   5   240     0
230 14  1   1   1   38  0       0   24  220     3   13  460     0
231 14  6   1   1   34  240     5   35  200     3   5   440     0
232 15  6   1   1   42  380     4   40  200     6   8   580     0
233 15  6   1   1   36  30      5   42  300     5   4   330     0
234 14  7   1   1   42  0       5   43  320     6   4   470     0
235 15  7   1   1   42  0       4   52  200     0   10  240     1
236 14  5   1   1   30  0       2   35  488     5   5   488     0
237 14  6   1   1   33  0       8   39  300     4   5   330     0
238 14  4   1   1   36  240     4   50  300     2   8   540     0
239 15  5   1   1   35  30      4   84  240     0   8   1010    0
240 15  5   1   1   39  0       4   52  299     4   6   299     0
241 15  7   1   1   43  0       5   48  100     7   5   300     0
242 14  4   1   1   26  400     3   42  600     3   6   1700    1
243 14  5   1   1   50  90      0   54  431     0   5   521     0
244 14  6   1   1   29  200     2   27  500     2   5   700     1
;
proc print data=educa;
 run;

********* Modelo completo *********;
proc logistic data=educa descending;
model trabalha = idade anoestu rural mulher idade_mae rendtot_mae educa_mae
                 idade_pai rendtot_pai educa_pai num_pes s_rendtotal;
run;

* escolhe as quatro melhores covariáveis;
proc logistic data=educa descending;
model trabalha = idade anoestu rural mulher idade_mae rendtot_mae educa_mae
                 idade_pai rendtot_pai educa_pai num_pes s_rendtotal
                   /selection=score best=4;
run;

***********************************************;
* Calcula as medidas de diagnóstico de Pregibon  *;
***********************************************;
proc logistic data=educa descending;
 model trabalha = idade anoestu rural mulher idade_mae rendtot_mae educa_mae
                  idade_pai rendtot_pai educa_pai num_pes s_rendtotal
                                                          /influence;
    output out=graf
    reschi=resd_chi
    resdev=resd_dev
    h=hat
    c=int_c
    cbar=int_cbar
```

```
    difchisq=d_chi
    difdev=d_dev;
    run;

symbol1 i=none value=circle color=red height=.8;
symbol2 i=none value=diamond color=green height=.8;
proc gplot data=graf;
    axis2 label= (color=blue 'Observações');
    **;
    axis1 label=(angle=-90 rotate=90 color=blue 'Resíduo de Pearson');
    plot resd_chi*obs=1/frame overlay vaxis=axis1 haxis=axis2;
    run;
    axis3 label=(angle=-90 rotate=90 color=blue 'Resíduo Deviance');
    plot resd_dev*obs=2/frame overlay vaxis=axis3 haxis=axis2;
    run;
    axis4 label=(angle=-90 rotate=90 color=blue 'Diagonal da matriz H');
    plot hat*obs=1/frame overlay vaxis=axis4 haxis=axis2;
    run;
    *axis5 label=(angle=-90 rotate=90 color=blue 'Dfbeta 0');
    *plot dif_b0*novobs=2/frame overlay vaxis=axis5 haxis=axis2;
    *run;
    *axis10 label=(angle=-90 rotate=90 color=blue 'Dfbeta 1');
    *plot dif_b1*novobs=1/frame overlay vaxis=axis10 haxis=axis2;
    *run;
    *axis11 label=(angle=-90 rotate=90 color=blue 'Dfbeta 2');
    *plot dif_b2*novobs=2/frame overlay vaxis=axis11 haxis=axis2;
    *run;
    axis6 label=(angle=-90 rotate=90 color=blue 'C');
    plot int_c*obs=1/frame overlay vaxis=axis6 haxis=axis2;
    run;
    axis7 label=(angle=-90 rotate=90 color=blue 'CBAR');
    plot int_cbar*obs=2/frame overlay vaxis=axis7 haxis=axis2;
    run;
    axis8 label=(angle=-90 rotate=90 color=blue 'Delta X^2');
    plot d_chi*obs=1/frame overlay vaxis=axis8 haxis=axis2;
    run;
    axis9 label=(angle=-90 rotate=90 color=blue 'Delta Deviance');
    plot d_dev*obs=2/frame overlay vaxis=axis9 haxis=axis2;
run;

proc print data=graf;
 var resd_chi resd_dev hat int_c int_cbar d_chi d_dev;
run;

**********************;
proc logistic data=educa descending outest=betas covout;
 model trabalha = idade anoestu rural mulher idade_mae rendtot_mae educa_mae
                  idade_pai rendtot_pai educa_pai num_pes s_rendtotal
                    /selection=stepwise slentry=0.3 slstay=0.35 details lackfit;
output out=pred p=phat lower=lcl upper=ucl predprobs=(individual
crossvalidate);
run;
```

ANEXO B - Listagem do programa para Influência Local.

```
#####################################################################
##                      Influência Local                           ##
##                                                                 ##
##                    Regressão Logística                          ##
#####################################################################
Dados<-read.table("A:/educa.txt",header=TRUE)
attach(Dados)
Dados
ajuste01<-glm(trabalha~idade+anoestu+rural+mulher+idade_mae+rendtot_mae+educa_mae+idade_pai
            +rendtot_pai+educa_pai+num_pes+s_rendtotal,data=Dados,family=binomial(logit))
summary(ajuste01)
anova(ajuste01)
####################################
obs<-1:244
uns<-c(rep(1,244))
# monta a matriz X
X <-cbind(uns,idade,anoestu,rural,mulher,idade_mae,rendtot_mae,educa_mae,idade_pai,
         rendtot_pai,educa_pai,num_pes,s_rendtotal)
# variável resposta ==>vetor Y
Y<-(Dados[,15])
beta<-ajuste01$coef
Xbeta<-X%*%beta
expXbeta<-exp(Xbeta)
#### probabilidade de sucesso e fracasso ####
PI <-(expXbeta/(1+expXbeta))
IPI<-(uns-PI)
# montando a matriz Q
Q<-(PI*IPI)
vetorQ<-c(Q)
MatrizQ<-diag(vetorQ)
Ibeta<-t(X)%*%MatrizQ%*%X
# pela inversa de Ibeta calculamos a matriz de variancia-covariância
InvIbeta<-solve(Ibeta)
YPI<-(Y-PI)

#####################################################################
##                       Caso Ponderado                            ##
#####################################################################
matYPI<-c(rep(YPI,13))
matdelta<-matYPI*X
matdelta
#####construindo a matriz H (não é matriz leverage)
H <- matdelta%*%InvIbeta%*%t(matdelta)
##### calcula o autovalor e autovetor de H
autH <- eigen(H)
##### separa os autovetores
autovetor<-autH$vectors
autovetor
autvet<-c(autovetor[,1])
autvet
#### Curvatura
curv<-2*abs(t(autvet)%*%H%*%autvet)
curv
#####desenhado grafico
plot(obs,abs(autvet),xlab="Observações",ylab="",col=14,pch=16)
```

```
title("Grafico Caso Ponderado")

plot(obs,diag(H),col=4,pch=16)
title("Grafico i-ésimo individuo H")

######################################################################
##              Pertubando as Covariáveis                           ##
######################################################################
X0 <- uns
X1 <- idade
X2 <- anoestu
X3 <- rural
X4 <- mulher
X5 <- idade_mae
X6 <- rendtot_mae
X7 <- educa_mae
X8 <- idade_pai
X9 <- rendtot_pai
X10<- educa_pai
X11<- num_pes
X12<- s_rendtotal
######################################################################
##          Perturbando a covariável  idade (X1)                    ##
######################################################################
beta1 <- c(rep(0.3486707822,244))
sq01 <- sqrt(var(idade))
v1 <- c(rep( sq01,244))
a1 <- ((Y-PI)-(PI*IPI*beta1*X1))*v1

a0 <- (-IPI*PI*v1*beta1*X0)
a2 <- (-IPI*PI*v1*beta1*X2)
a3 <- (-IPI*PI*v1*beta1*X3)
a4 <- (-IPI*PI*v1*beta1*X4)
a5 <- (-IPI*PI*v1*beta1*X5)
a6 <- (-IPI*PI*v1*beta1*X6)
a7 <- (-IPI*PI*v1*beta1*X7)
a8 <- (-IPI*PI*v1*beta1*X8)
a9 <- (-IPI*PI*v1*beta1*X9)
a10 <- (-IPI*PI*v1*beta1*X10)
a11 <- (-IPI*PI*v1*beta1*X11)
a12 <- (-IPI*PI*v1*beta1*X12)
delta1<-cbind(a0,a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12)
tdelta1<- t(delta1)
H1 <- t(tdelta1)%*%InvIbeta%*%tdelta1
H1
autH1 <- eigen(H1)
autH1
autovetor1 <-autH1$vectors
autovetor1
autvet1 <- c(autovetor1[,1])
autvet1
#####desenhado grafico
plot(obs,abs(autvet1), xlab="Observações",ylab="",col=3, pch=16)
title("Grafico perturbando Idade")

curv1<-2*abs(t(autvet1)%*%H1%*%autvet1)
```

```
curv1

plot(obs,diag(H1), col=2, pch=16)
title("Grafico i-ésimo individuo H1")
#####################################################################
##            Perturbando a covariável  anoestu (X2)          ##
#####################################################################
beta2 <- c(rep(0.0768801,244))
sq02 <- sqrt(var(anoestu))
v2 <-c(rep( sq02,244))
b2 <- ((Y-PI)-(PI*IPI*beta2*X2))*v2

b0 <- (-IPI*PI*v2*beta2*X0)
b1 <- (-IPI*PI*v2*beta2*X1)
b3 <- (-IPI*PI*v2*beta2*X3)
b4 <- (-IPI*PI*v2*beta2*X4)
b5 <- (-IPI*PI*v2*beta2*X5)
b6 <- (-IPI*PI*v2*beta2*X6)
b7 <- (-IPI*PI*v2*beta2*X7)
b8 <- (-IPI*PI*v2*beta2*X8)
b9 <- (-IPI*PI*v2*beta2*X9)
b10 <- (-IPI*PI*v2*beta2*X10)
b11 <- (-IPI*PI*v2*beta2*X11)
b12 <- (-IPI*PI*v2*beta2*X12)
delta2 <-cbind(b0,b1,b2,b3,b4,b5,b6,b7,b8,b9,b10,b11,b12)
tdelta2 <-t(delta2)
H2 <- t(tdelta2)%*%InvIbeta%*%tdelta2
H2
autH2 <- eigen(H2)
autH2
autovetor2 <- autH2$vectors
autovetor2
autvet2 <- c(autovetor2[,1])
autvet2
#####desenhado grafico
plot(obs,abs(autvet2), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Anoestu")

curv2 <- 2*abs(t(autvet2)%*%H2%*%autvet2)
curv2

plot(obs,diag(H2), col=2, pch=16)
title("Grafico i-ésimo individuo H2")
#####################################################################
##            Perturbando a covariável  rural (X3)            ##
#####################################################################
beta3 <- c(rep(1.0812234,244))
sq03 <- sqrt(var(rural))
v3 <-c(rep( sq03,244))
c3 <- ((Y-PI)-(PI*IPI*beta3*X3))*v3

c0 <- (-IPI*PI*v3*beta3*X0)
c1 <- (-IPI*PI*v3*beta3*X1)
c2 <- (-IPI*PI*v3*beta3*X2)
c4 <- (-IPI*PI*v3*beta3*X4)
c5 <- (-IPI*PI*v3*beta3*X5)
```

```
c6 <- (-IPI*PI*v3*beta3*X6)
c7 <- (-IPI*PI*v3*beta3*X7)
c8 <- (-IPI*PI*v3*beta3*X8)
c9 <- (-IPI*PI*v3*beta3*X9)
c10 <- (-IPI*PI*v3*beta3*X10)
c11 <- (-IPI*PI*v3*beta3*X11)
c12 <- (-IPI*PI*v3*beta3*X12)
delta3 <- cbind(c0,c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,c11,c12)
tdelta3 <- t(delta3)
H3 <- t(tdelta3)%*%InvIbeta%*%tdelta3
H3
autH3 <- eigen(H3)
autH3
autovetor3 <- autH3$vectors
autovetor3
autvet3 <- c(autovetor3[,1])
autvet3
#####desenhado grafico
plot(obs,abs(autvet3), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Rural")

curv3 <- 2*abs(t(autvet3)%*%H3%*%autvet3)
curv3

plot(obs,diag(H3), col=2, pch=16)
title("Grafico i-ésimo individuo H3")
#####################################################################
##            Perturbando a covariável  mulher (X4)           ##
#####################################################################
beta4 <- c(rep(-1.1416033,244))
sq04 <- sqrt(var(mulher))
v4 <- c(rep( sq04,244))
d4 <- ((Y-PI)-(PI*IPI*beta4*X4))*v4

d0 <- (-IPI*PI*v4*beta4*X0)
d1 <- (-IPI*PI*v4*beta4*X1)
d2 <- (-IPI*PI*v4*beta4*X2)
d3 <- (-IPI*PI*v4*beta4*X3)
d5 <- (-IPI*PI*v4*beta4*X5)
d6 <- (-IPI*PI*v4*beta4*X6)
d7 <- (-IPI*PI*v4*beta4*X7)
d8 <- (-IPI*PI*v4*beta4*X8)
d9 <- (-IPI*PI*v4*beta4*X9)
d10 <- (-IPI*PI*v4*beta4*X10)
d11 <- (-IPI*PI*v4*beta4*X11)
d12 <- (-IPI*PI*v4*beta4*X12)
delta4 <-cbind(d0,d1,d2,d3,d4,d5,d6,d7,d8,d9,d10,d11,d12)
tdelta4 <- t(delta4)
H4 <- t(tdelta4)%*%InvIbeta%*%tdelta4
H4
autH4 <- eigen(H4)
autH4
autovetor4 <- autH4$vectors
autovetor4
autvet4 <- c(autovetor4[,1])
autvet4
```

```
#####desenhado grafico
plot(obs,abs(autvet4), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Mulher")

curv4 <- 2*abs(t(autvet4)%*%H4%*%autvet4)
curv4

plot(obs,diag(H4), col=2, pch=16)
title("Grafico i-ésimo individuo H4")
#################################################################
##           Perturbando a covariável  idade_mae (X5)          ##
#################################################################
beta5 <- c(rep(-0.0340206,244))
sq05 <- sqrt(var(idade_mae))
v5 <- c(rep( sq05,244))
e5 <- ((Y-PI)-(PI*IPI*beta5*X5))*v5

e0 <- (-IPI*PI*v5*beta5*X0)
e1 <- (-IPI*PI*v5*beta5*X1)
e2 <- (-IPI*PI*v5*beta5*X2)
e3 <- (-IPI*PI*v5*beta5*X3)
e4 <- (-IPI*PI*v5*beta5*X4)
e6 <- (-IPI*PI*v5*beta5*X6)
e7 <- (-IPI*PI*v5*beta5*X7)
e8 <- (-IPI*PI*v5*beta5*X8)
e9 <- (-IPI*PI*v5*beta5*X9)
e10 <- (-IPI*PI*v5*beta5*X10)
e11 <- (-IPI*PI*v5*beta5*X11)
e12 <- (-IPI*PI*v5*beta5*X12)
delta5 <-cbind(e0,e1,e2,e3,e4,e5,e6,e7,e8,e9,e10,e11,e12)
tdelta5 <-t(delta5)
H5 <- t(tdelta5)%*%InvIbeta%*%tdelta5
H5
autH5 <- eigen(H5)
autH5
autovetor5 <- autH5$vectors
autovetor5
autvet5 <- c(autovetor5[,1])
autvet5
#####desenhado grafico
plot(obs,abs(autvet5), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Idade_mãe")

curv5 <- 2*abs(t(autvet5)%*%H5%*%autvet5)
curv5

plot(obs,diag(H5), col=2, pch=16)
title("Grafico i-ésimo individuo H5")
#################################################################
##        Perturbando a covariável  rendtot_mae (X6)           ##
#################################################################
beta6 <- c(rep( -0.0002647,244))
sq06 <- sqrt(var(rendtot_mae))
v6 <- c(rep( sq06,244))
f6 <- ((Y-PI)-(PI*IPI*beta6*X6))*v6
```

```
f0 <- (-IPI*PI*v6*beta6*X0)
f1 <- (-IPI*PI*v6*beta6*X1)
f2 <- (-IPI*PI*v6*beta6*X2)
f3 <- (-IPI*PI*v6*beta6*X3)
f4 <- (-IPI*PI*v6*beta6*X4)
f5 <- (-IPI*PI*v6*beta6*X5)
f7 <- (-IPI*PI*v6*beta6*X7)
f8 <- (-IPI*PI*v6*beta6*X8)
f9 <- (-IPI*PI*v6*beta6*X9)
f10 <- (-IPI*PI*v6*beta6*X10)
f11 <- (-IPI*PI*v6*beta6*X11)
f12 <- (-IPI*PI*v6*beta6*X12)
delta6 <- cbind(f0,f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12)
tdelta6 <- t(delta6)
H6 <- t(tdelta6)%*%InvIbeta%*%tdelta6
H6
autH6 <- eigen(H6)
autH6
autovetor6 <- autH6$vectors
autovetor6
autvet6 <- c(autovetor6[,1])
autvet6
#####desenhado grafico
plot(obs,abs(autvet6), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Renda_mãe")

curv6 <- 2*abs(t(autvet6)%*%H6%*%autvet6)
curv6

plot(obs,diag(H6), col=2, pch=16)
title("Grafico i-ésimo individuo H6")
#####################################################################
##          Perturbando a covariável  educa_mae (X7)          ##
#####################################################################
beta7<- c(rep(-0.0228282,244))
sq07<-sqrt(var(educa_mae))
v7<-c(rep( sq07,244))
g7 <- ((Y-PI)-(PI*IPI*beta7*X7))*v7

g0 <- (-IPI*PI*v7*beta7*X0)
g1 <- (-IPI*PI*v7*beta7*X1)
g2 <- (-IPI*PI*v7*beta7*X2)
g3 <- (-IPI*PI*v7*beta7*X3)
g4 <- (-IPI*PI*v7*beta7*X4)
g5 <- (-IPI*PI*v7*beta7*X5)
g6 <- (-IPI*PI*v7*beta7*X6)
g8 <- (-IPI*PI*v7*beta7*X8)
g9 <- (-IPI*PI*v7*beta7*X9)
g10 <- (-IPI*PI*v7*beta7*X10)
g11 <- (-IPI*PI*v7*beta7*X11)
g12 <- (-IPI*PI*v7*beta7*X12)
delta7 <- cbind(g0,g1,g2,g3,g4,g5,g6,g7,g8,g9,g10,g11,g12)
tdelta7 <-t(delta7)
H7 <- t(tdelta7)%*%InvIbeta%*%tdelta7
H7
autH7 <- eigen(H7)
```

```
autH7
autovetor7 <- autH7$vectors
autovetor7
autvet7 <- c(autovetor7[,1])
autvet7
#####desenhado grafico
plot(obs,abs(autvet7), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando educa_mãe")

curv7 <- 2*abs(t(autvet7)%*%H7%*%autvet7)
curv7

plot(obs,diag(H7), col=2, pch=16)
title("Grafico i-ésimo individuo H7")
################################################################
##          Perturbando a covariável  idade_pai (X8)          ##
################################################################
beta8 <- c(rep( 0.0019784,244))
sq08 <- sqrt(var(idade_pai))
v8 <- c(rep( sq08,244))
h8 <- ((Y-PI)-(PI*IPI*beta8*X8))*v8

h0 <- (-IPI*PI*v8*beta8*X0)
h1 <- (-IPI*PI*v8*beta8*X1)
h2 <- (-IPI*PI*v8*beta8*X2)
h3 <- (-IPI*PI*v8*beta8*X3)
h4 <- (-IPI*PI*v8*beta8*X4)
h5 <- (-IPI*PI*v8*beta8*X5)
h6 <- (-IPI*PI*v8*beta8*X6)
h7 <- (-IPI*PI*v8*beta8*X7)
h9 <- (-IPI*PI*v8*beta8*X9)
h10 <- (-IPI*PI*v8*beta8*X10)
h11 <- (-IPI*PI*v8*beta8*X11)
h12 <- (-IPI*PI*v8*beta8*X12)
delta8 <- cbind(h0,h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11,h12)
tdelta8 <- t(delta8)
H8 <- t(tdelta8)%*%InvIbeta%*%tdelta8
H8
autH8 <- eigen(H8)
autH8
autovetor8 <- autH8$vectors
autovetor8
autvet8 <- c(autovetor8[,1])
autvet8
#####desenhado grafico
plot(obs,abs(autvet8), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando idade_pai")

curv8 <- 2*abs(t(autvet8)%*%H8%*%autvet8)
curv8

plot(obs,diag(H8), col=2, pch=16)
title("Grafico i-ésimo individuo H8")
################################################################
##         Perturbando a covariável  rendtot_pai (X9)         ##
################################################################
```

```
beta9 <- c(rep( -0.0004296,244))
sq09 <- sqrt(var(rendtot_pai))
v9 <- c(rep( sq08,244))
i9 <- ((Y-PI)-(PI*IPI*beta9*X9))*v9

i0 <- (-IPI*PI*v9*beta9*X0)
i1 <- (-IPI*PI*v9*beta9*X1)
i2 <- (-IPI*PI*v9*beta9*X2)
i3 <- (-IPI*PI*v9*beta9*X3)
i4 <- (-IPI*PI*v9*beta9*X4)
i5 <- (-IPI*PI*v9*beta9*X5)
i6 <- (-IPI*PI*v9*beta9*X6)
i7 <- (-IPI*PI*v9*beta9*X7)
i8 <- (-IPI*PI*v9*beta9*X8)
i10 <- (-IPI*PI*v9*beta9*X10)
i11 <- (-IPI*PI*v9*beta9*X11)
i12 <- (-IPI*PI*v9*beta9*X12)
delta9 <- cbind(i0,i1,i2,i3,i4,i5,i6,i7,i8,i9,i10,i11,i12)
tdelta9 <- t(delta9)
H9 <- t(tdelta9)%*%InvIbeta%*%tdelta9
H9
autH9 <- eigen(H9)
autH9
autovetor9 <- autH9$vectors
autovetor9
autvet9 <- c(autovetor9[,1])
autvet9
#####desenhado grafico
plot(obs,abs(autvet9), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando renda_pai")

curv9 <- 2*abs(t(autvet9)%*%H9%*%autvet9)
curv9

plot(obs,diag(H9), col=2, pch=16)
title("Grafico i-ésimo individuo H9")
#####################################################################
##          Perturbando a covariável  educa_pai (X10)         ##
#####################################################################
beta10 <- c(rep(-0.1075582,244))
sq10 <- sqrt(var(educa_pai))
v10 <- c(rep( sq10,244))
j10<- ((Y-PI)-(PI*IPI*beta10*X10))*v10

j0 <- (-IPI*PI*v10*beta10*X0)
j1 <- (-IPI*PI*v10*beta10*X1)
j2 <- (-IPI*PI*v10*beta10*X2)
j3 <- (-IPI*PI*v10*beta10*X3)
j4 <- (-IPI*PI*v10*beta10*X4)
j5 <- (-IPI*PI*v10*beta10*X5)
j6 <- (-IPI*PI*v10*beta10*X6)
j7 <- (-IPI*PI*v10*beta10*X7)
j8 <- (-IPI*PI*v10*beta10*X8)
j9 <- (-IPI*PI*v10*beta10*X9)
j11 <-(-IPI*PI*v10*beta10*X11)
j12 <- (-IPI*PI*v10*beta10*X12)
```

```
delta10 <- cbind(j0,j1,j2,j3,j4,j5,j6,j7,j8,j9,j10,j11,j12)
tdelta10 <-t(delta10)
H10 <- t(tdelta10)%*%InvIbeta%*%tdelta10
H10
autH10 <- eigen(H10)
autH10
autovetor10 <- autH10$vectors
autovetor10
autvet10 <- c(autovetor10[,1])
autvet10
#####desenhado grafico
plot(obs,abs(autvet10), xlab="Observações", ylab="",col=3,pch=16)
title("Grafico perturbando educa_pai")

curv10 <- 2*abs(t(autvet10)%*%H10%*%autvet10)
curv10

plot(obs,diag(H10), col=2, pch=16)
title("Grafico i-ésimo indíviduo H10")
#####################################################################
##           Perturbando a covariável  num_pes (X11)          ##
#####################################################################
beta11 <- c(rep( 0.0132351,244))
sq11 <- sqrt(var(num_pes))
v11 <- c(rep( sq11,244))
l11 <- ((Y-PI)-(PI*IPI*beta11*X11))*v11

l0 <- (-IPI*PI*v11*beta11*X0)
l1 <- (-IPI*PI*v11*beta11*X1)
l2 <- (-IPI*PI*v11*beta11*X2)
l3 <- (-IPI*PI*v11*beta11*X3)
l4 <- (-IPI*PI*v11*beta11*X4)
l5 <- (-IPI*PI*v11*beta11*X5)
l6 <- (-IPI*PI*v11*beta11*X6)
l7 <- (-IPI*PI*v11*beta11*X7)
l8 <- (-IPI*PI*v11*beta11*X8)
l9 <- (-IPI*PI*v11*beta11*X9)
l10 <- (-IPI*PI*v11*beta11*X10)
l12 <- (-IPI*PI*v10*beta11*X12)
delta11 <- cbind(l0,l1,l2,l3,l4,l5,l6,l7,l8,l9,l10,l11,l12)
tdelta11 <- t(delta11)
H11 <- t(tdelta11)%*%InvIbeta%*%tdelta11
H11
autH11 <- eigen(H11)
autH11
autovetor11 <- autH11$vectors
autovetor11
autvet11<-c(autovetor11[,1])
autvet11
#####desenhado
grafico plot(obs,abs(autvet11), xlab="Observações", ylab="",col=3,pch=16)
title("Grafico perturbando num_pessoa")

curv11 <- 2*abs(t(autvet11)%*%H11%*%autvet11)
curv11
```

```
plot(obs,diag(H11), col=2, pch=16)
title("Grafico i-ésimo individuo H11")
#####################################################################
##          Perturbando a covariável  s_rentotal (X12)        ##
#####################################################################
beta12 <- c(rep(0.0005375,244))
sq12 <- sqrt(var(s_rendtotal))
v12 <- c(rep( sq12,244))
m12 <- ((Y-PI)-(PI*IPI*beta12*X12))*v12

m0 <- (-IPI*PI*v12*beta12*X0)
m1 <- (-IPI*PI*v12*beta12*X1)
m2 <- (-IPI*PI*v12*beta12*X2)
m3 <- (-IPI*PI*v12*beta12*X3)
m4 <- (-IPI*PI*v12*beta12*X4)
m5 <- (-IPI*PI*v12*beta12*X5)
m6 <- (-IPI*PI*v12*beta12*X6)
m7 <- (-IPI*PI*v12*beta12*X7)
m8 <- (-IPI*PI*v12*beta12*X8)
m9 <- (-IPI*PI*v12*beta12*X9)
m10 <- (-IPI*PI*v12*beta12*X10)
m11 <- (-IPI*PI*v12*beta12*X11)
delta12 <- cbind(m0,m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12)
tdelta12 <- t(delta12)
H12 <- t(tdelta12)%*%InvIbeta%*%tdelta12
H12
autH12 <- eigen(H12)
autH12
autovetor12 <- autH12$vectors
autovetor12
autvet12 <- c(autovetor12[,1])
autvet12
#####desenhado
grafico plot(obs,abs(autvet12), xlab="Observações", ylab="",col=3,pch=16)
title("Grafico perturbando s_rendtotal")

curv12 <- 2*abs(t(autvet12)%*%H12%*%autvet12)
curv12

plot(obs,diag(H12), col=2, pch=16)
title("Grafico i-ésimo indivíduo H12")
#####################################################################
##                Construindo o gráfico de envelope           ##
#####################################################################
X<- model.matrix(ajuste01)
n <- nrow(X)
p <- ncol(X)
w <- ajuste01$weights
W <- diag(w)
MatrizH <- solve(t(X)%*%W%*%X)
MatrizH

MatrizH <- sqrt(W)%*%X%*%MatrizH%*%t(X)%*%sqrt(W)
h <- diag(MatrizH)
ts <- resid(ajuste01,type="pearson")/sqrt(1-h)
td <- resid(ajuste01,type="deviance")/sqrt(1-h)
```

```
e <- matrix(0,n,100)
#
for(i in 1:100){
    dif <- runif(n) - fitted(ajuste01)
    dif[ dif >= 0] <- 0
    dif[ dif < 0]  <- 1
    nresp <- dif
    fit <- glm(nresp~X, family=binomial)
    w <- fit$weights
    w
    W <- diag(w)
    W
    MatrizH <- solve(t(X)%*%W%*%X)
    MatrizH
    MatrizH <- sqrt(W)%*%X%*%MatrizH%*%t(X)%*%sqrt(W)
    h <- diag(MatrizH)
    e[,i] <- sort(resid(fit, type="deviance")/sqrt(1-h))}
#
e1 <- numeric(n)
e2 <- numeric(n)
#
for (i in 1:n){
    e0 <- sort(e[i,])
    e1[i]<- e0[5]
    e2[i]<- e0[95]}
#
med <- apply(e,1,mean)
faixa <- range(td,e1,e2)
#
par(mfrow=c(1,1)) par(pty="s")
qqnorm(td, xlab="Percentis da N(0,1)", ylab="Componente do Desvio Padronizado",
ylim=faixa,col="blue",pch=16)
par(new=T)
qqnorm(e1, axes=F,xlab="", ylab="",type="l", col="green", ylim=faixa,lty=1)
par(new=T)
qqnorm(e2,axes=F, xlab="", ,ylab="",type="l",col="green", ylim=faixa,lty=1)
par(new=T)
qqnorm(med, axes=F,xlab="", ylab="", type="l",col="red", ylim=faixa,lty=2) par(new=T)
#########################################################
```

ANEXO C - Listagem das medidas de diagnóstico.

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est
1	-0.48267	-0.64719	0.01827	0.00442	0.00434	0.23731	0.42319	0.18895
2	1.56492	1.57358	0.04861	0.13151	0.12512	2.57410	2.60128	0.28994
3	-0.56576	-0.74524	0.03371	0.01156	0.01117	0.33125	0.56655	0.24247
4	1.86291	1.73060	0.03968	0.14931	0.14338	3.61380	3.13835	0.22369
5	1.69788	1.64715	0.03461	0.10706	0.10335	2.98616	2.81647	0.25755
6	-0.46887	-0.63042	0.03983	0.00950	0.00912	0.22895	0.40655	0.18022
7	-0.60013	-0.78434	0.04439	0.01751	0.01673	0.37688	0.63192	0.26479
8	-0.49370	-0.66049	0.02671	0.00687	0.00669	0.25043	0.44294	0.19597
9	-0.64388	-0.83286	0.07454	0.03608	0.03339	0.44797	0.72705	0.29307
10	-0.90508	-1.09396	0.05966	0.05527	0.05197	0.87114	1.24873	0.45030
11	-0.63854	-0.82702	0.07352	0.03493	0.03236	0.44010	0.71633	0.28964
12	-0.75390	-0.94872	0.13955	0.10713	0.09218	0.66054	0.99224	0.36239
13	-0.63824	-0.82669	0.06027	0.02780	0.02613	0.43348	0.70955	0.28945
14	1.64339	1.61773	0.06729	0.20891	0.19485	2.89557	2.81191	0.27022
15	-0.73943	-0.93398	0.04664	0.02806	0.02675	0.57350	0.89906	0.35348
16	-0.79380	-0.98859	0.04262	0.02930	0.02805	0.65818	1.00537	0.38655
17	-0.36226	-0.49660	0.03957	0.00563	0.00541	0.13664	0.25202	0.11601
18	-0.55736	-0.73556	0.04606	0.01573	0.01500	0.32565	0.55605	0.23702
19	-0.51414	-0.68491	0.02379	0.00660	0.00644	0.27078	0.47554	0.20907
20	1.83352	1.71633	0.05763	0.21816	0.20559	3.56740	3.15137	0.22926
21	-0.50571	-0.67488	0.02934	0.00796	0.00773	0.26348	0.46319	0.20366
22	-0.68494	-0.87713	0.03897	0.01979	0.01902	0.48817	0.78838	0.31933
23	-0.62592	-0.81312	0.04438	0.01904	0.01819	0.40997	0.67936	0.28150
24	-0.54868	-0.72549	0.21979	0.10870	0.08481	0.38585	0.61114	0.23139
25	1.97770	1.78412	0.03410	0.14294	0.13807	4.04936	3.32114	0.20361
26	-0.58897	-0.77174	0.03461	0.01288	0.01244	0.35932	0.60802	0.25755
27	-0.62835	-0.81581	0.02988	0.01253	0.01216	0.40699	0.67770	0.28307
28	-0.65687	-0.84700	0.05988	0.02923	0.02748	0.45896	0.74490	0.30142
29	1.55971	1.57057	0.03125	0.08101	0.07848	2.51118	2.54517	0.29132
30	-0.54337	-0.71931	0.04859	0.01585	0.01508	0.31033	0.53249	0.22795
31	1.84566	1.72225	0.04124	0.15282	0.14652	3.55298	3.11266	0.22694
32	-0.57376	-0.75443	0.04225	0.01516	0.01452	0.34373	0.58368	0.24767
33	1.99582	1.79226	0.08404	0.39899	0.36546	4.34875	3.57764	0.20067
34	1.24643	1.36931	0.06141	0.10830	0.10165	1.65524	1.97665	0.39160
35	-0.57793	-0.75919	0.07418	0.02891	0.02676	0.36077	0.60314	0.25038
36	-0.40274	-0.54829	0.03500	0.00610	0.00588	0.16808	0.30651	0.13956
37	1.44504	1.50171	0.03928	0.08887	0.08538	2.17351	2.34051	0.32382
38	0.75274	0.94754	0.25840	0.26623	0.19743	0.76405	1.09528	0.63832
39	-0.53622	-0.71095	0.02477	0.00749	0.00730	0.29483	0.51275	0.22332
40	-0.41513	-0.56390	0.03005	0.00550	0.00534	0.17767	0.32332	0.14700
41	1.06781	1.23362	0.05227	0.06636	0.06289	1.20310	1.58470	0.46724
42	-0.57090	-0.75115	0.04905	0.01768	0.01681	0.34274	0.58103	0.24581
43	-0.73059	-0.92490	0.03905	0.02257	0.02169	0.55545	0.87713	0.34801
44	-0.46683	-0.62793	0.03504	0.00820	0.00791	0.22584	0.40221	0.17893
45	-0.81960	-1.01377	0.08641	0.06954	0.06353	0.73528	1.09127	0.40182
46	-0.53048	-0.70421	0.11545	0.04152	0.03673	0.31814	0.53265	0.21961
47	-0.47282	-0.63523	0.07833	0.02062	0.01900	0.24256	0.42252	0.18271
48	0.80317	0.99778	0.11437	0.09406	0.08330	0.72838	1.07887	0.60787
49	-0.64838	-0.83778	0.09255	0.04725	0.04287	0.46327	0.74475	0.29597
50	-0.54808	-0.72480	0.02013	0.00630	0.00617	0.30656	0.53150	0.23100
51	-0.62440	-0.81144	0.06542	0.02920	0.02729	0.41717	0.68573	0.28051
52	-0.45698	-0.61587	0.14422	0.04112	0.03519	0.24402	0.41449	0.17275

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est
53	-0.48700	-0.65242	0.30635	0.15101	0.10475	0.34192	0.53040	0.19170
54	-0.37428	-0.51206	0.02665	0.00394	0.00384	0.14392	0.26605	0.12287
55	2.49826	1.98988	0.04054	0.27487	0.26373	6.50503	4.22333	0.13810
56	-0.76648	-0.96141	0.09153	0.06515	0.05919	0.64668	0.98350	0.37007
57	-0.73463	-0.92906	0.03731	0.02173	0.02092	0.56060	0.88407	0.35051
58	0.72832	0.92256	0.15452	0.11467	0.09695	0.62740	0.94807	0.65340
59	-0.57075	-0.75097	0.03423	0.01195	0.01154	0.33730	0.57550	0.24571
60	2.00727	1.79736	0.04732	0.21007	0.20013	4.22924	3.43062	0.19884
61	-0.72201	-0.91604	0.04727	0.02715	0.02587	0.54717	0.86500	0.34267
62	0.75435	0.94918	0.22778	0.21736	0.16785	0.73690	1.06879	0.63733
63	-0.51149	-0.68176	0.04171	0.01188	0.01139	0.27301	0.47619	0.20737
64	-0.74234	-0.93696	0.05487	0.03385	0.03199	0.58307	0.90989	0.35529
65	-0.43084	-0.58356	0.05732	0.01197	0.01129	0.19691	0.35182	0.15656
66	-0.49475	-0.66175	0.03522	0.00926	0.00894	0.25372	0.44685	0.19665
67	-0.49118	-0.65746	0.04163	0.01093	0.01048	0.25174	0.44273	0.19436
68	1.91869	1.75704	0.07567	0.32602	0.30135	3.98274	3.38854	0.21361
69	-0.53864	-0.71378	0.02248	0.00683	0.00667	0.29680	0.51616	0.22488
70	-0.62363	-0.81059	0.05696	0.02491	0.02349	0.41241	0.68055	0.28002
71	-0.49205	-0.65851	0.02890	0.00742	0.00721	0.24932	0.44084	0.19492
72	-0.25788	-0.35885	0.03336	0.00237	0.00230	0.06880	0.13107	0.06236
73	-0.48291	-0.64748	0.02628	0.00646	0.00629	0.23950	0.42552	0.18910
74	0.85913	1.05147	0.12420	0.11952	0.10467	0.84277	1.21026	0.57534
75	-0.36256	-0.49699	0.04894	0.00711	0.00676	0.13821	0.25376	0.11618
76	-0.37672	-0.51518	0.04515	0.00703	0.00671	0.14863	0.27213	0.12428
77	-0.41567	-0.56458	0.03761	0.00702	0.00675	0.17953	0.32550	0.14732
78	-0.28179	-0.39093	0.03025	0.00255	0.00248	0.08188	0.15530	0.07357
79	-0.63801	-0.82643	0.04999	0.02255	0.02142	0.42847	0.70441	0.28929
80	-0.74565	-0.94033	0.03282	0.01951	0.01887	0.57486	0.90309	0.35732
81	1.41765	1.48449	0.03908	0.08506	0.08174	2.09147	2.28544	0.33226
82	-0.38275	-0.52289	0.04120	0.00657	0.00630	0.15279	0.27971	0.12778
83	-0.57374	-0.75440	0.05866	0.02179	0.02051	0.34969	0.58963	0.24765
84	-0.64019	-0.82883	0.02460	0.01060	0.01034	0.42018	0.69729	0.29070
85	-0.53559	-0.71021	0.03536	0.01090	0.01052	0.29737	0.51492	0.22291
86	-0.56112	-0.73990	0.02416	0.00799	0.00780	0.32265	0.55524	0.23946
87	-0.41568	-0.56460	0.03657	0.00681	0.00656	0.17935	0.32533	0.14733
88	-0.67729	-0.86898	0.05398	0.02767	0.02617	0.48490	0.78130	0.31447
89	-0.41822	-0.56779	0.04261	0.00813	0.00778	0.18269	0.33017	0.14887
90	-0.47093	-0.63294	0.05448	0.01352	0.01278	0.23456	0.41339	0.18152
91	2.18062	1.87087	0.03878	0.19957	0.19183	4.94692	3.69200	0.17376
92	2.29823	1.91712	0.03100	0.17439	0.16898	5.45086	3.84432	0.15919
93	-0.48835	-0.65404	0.02902	0.00734	0.00713	0.24561	0.43490	0.19256
94	-0.61624	-0.80237	0.05429	0.02305	0.02180	0.40155	0.66560	0.27523
95	1.45306	1.50670	0.14902	0.43450	0.36975	2.48113	2.63988	0.32140
96	1.67657	1.63577	0.10662	0.37552	0.33548	3.14638	3.01121	0.26241
97	-0.74424	-0.93889	0.04721	0.02880	0.02744	0.58133	0.90896	0.35645
98	-0.61436	-0.80028	0.10487	0.04940	0.04422	0.42166	0.68467	0.27402
99	1.19621	1.33287	0.12764	0.24000	0.20936	1.64029	1.98591	0.41137
100	1.28004	1.39299	0.04791	0.08660	0.08245	1.72097	2.02288	0.37900
101	-1.18918	-1.32766	0.04114	0.06328	0.06068	1.47482	1.82337	0.58577
102	-0.84119	-1.03449	0.05477	0.04338	0.04100	0.74860	1.11118	0.41438
103	0.79607	0.99082	0.05435	0.03852	0.03642	0.67015	1.01815	0.61210
104	-1.05975	-1.22708	0.06783	0.08767	0.08172	1.20479	1.58745	0.52898
105	0.85284	1.04554	0.05115	0.04133	0.03921	0.76655	1.13237	0.57893
106	-1.03352	-1.20554	0.04514	0.05288	0.05049	1.11866	1.50382	0.51648

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est
107	0.88990	1.08008	0.04913	0.04303	0.04091	0.83283	1.20748	0.55806
108	1.03753	1.20886	0.04700	0.05571	0.05309	1.12956	1.51443	0.48159
109	-1.39462	-1.46976	0.05139	0.11108	0.10537	2.05032	2.26556	0.66044
110	1.06960	1.23506	0.03503	0.04304	0.04154	1.18557	1.56692	0.46641
111	-1.38539	-1.46379	0.04622	0.09751	0.09300	2.01231	2.23569	0.65745
112	1.33025	1.42737	0.06268	0.12624	0.11833	1.88788	2.15570	0.36107
113	1.07533	1.23968	0.07111	0.09529	0.08852	1.24484	1.62533	0.46375
114	1.06509	1.23142	0.05227	0.06602	0.06257	1.19700	1.57897	0.46851
115	0.82881	1.02266	0.04775	0.03618	0.03445	0.72138	1.08027	0.59279
116	0.89496	1.08473	0.05098	0.04534	0.04303	0.84398	1.21966	0.55526
117	-1.03402	-1.20595	0.07566	0.09468	0.08751	1.15671	1.54183	0.51672
118	-1.00849	-1.18460	0.05584	0.06371	0.06016	1.07722	1.46344	0.50423
119	1.24913	1.37123	0.04976	0.08598	0.08170	1.64203	1.96197	0.39057
120	0.98268	1.16261	0.09298	0.10914	0.09899	1.06465	1.45064	0.50874
121	-1.05277	-1.22139	0.09694	0.13174	0.11897	1.22731	1.61077	0.52569
122	-0.88018	-1.07111	0.09386	0.08856	0.08024	0.85496	1.22752	0.43653
123	1.00879	1.18486	0.04115	0.04555	0.04367	1.06134	1.44756	0.49562
124	0.96668	1.14877	0.03691	0.03718	0.03581	0.97028	1.35547	0.51694
125	-1.16455	-1.30923	0.06794	0.10606	0.09885	1.45503	1.81293	0.57558
126	-1.14753	-1.29630	0.07492	0.11529	0.10665	1.42348	1.78705	0.56838
127	1.78171	1.69056	0.05974	0.21452	0.20171	3.37620	3.05969	0.23955
128	-1.01089	-1.18662	0.04538	0.05089	0.04858	1.07047	1.45664	0.50541
129	-0.91459	-1.10259	0.06650	0.06383	0.05959	0.89607	1.27529	0.45548
130	-0.98706	-1.16637	0.08770	0.10266	0.09366	1.06795	1.45408	0.49349
131	0.60696	0.79201	0.07121	0.03041	0.02824	0.39664	0.65553	0.73078
132	-0.84147	-1.03476	0.04455	0.03455	0.03301	0.74109	1.10375	0.41455
133	-0.29268	-0.40543	0.01819	0.00162	0.00159	0.08725	0.16596	0.07890
134	2.80444	2.08906	0.03382	0.28491	0.27528	8.14014	4.63947	0.11280
135	-0.20298	-0.28416	0.01358	0.00058	0.00057	0.04177	0.08131	0.03957
136	-0.20624	-0.28863	0.01560	0.00068	0.00067	0.04321	0.08398	0.04080
137	-0.39737	-0.54150	0.06892	0.01255	0.01169	0.16959	0.30491	0.13637
138	-0.44834	-0.60525	0.05379	0.01208	0.01143	0.21243	0.37775	0.16737
139	-0.43663	-0.59075	0.01744	0.00344	0.00338	0.19403	0.35238	0.16012
140	-0.27910	-0.38732	0.01439	0.00115	0.00114	0.07903	0.15116	0.07227
141	-0.37792	-0.51672	0.02559	0.00385	0.00375	0.14657	0.27076	0.12497
142	-0.29112	-0.40336	0.01473	0.00129	0.00127	0.08602	0.16397	0.07813
143	-0.25614	-0.35650	0.01488	0.00101	0.00099	0.06660	0.12808	0.06157
144	-0.32247	-0.44482	0.02679	0.00294	0.00286	0.10685	0.20072	0.09419
145	2.65178	2.04129	0.01696	0.12343	0.12134	7.15330	4.28820	0.12450
146	-0.45675	-0.61560	0.03006	0.00667	0.00646	0.21509	0.38542	0.17261
147	-0.26691	-0.37099	0.03549	0.00272	0.00262	0.07386	0.14025	0.06650
148	-0.49260	-0.65916	0.03230	0.00837	0.00810	0.25075	0.44260	0.19527
149	-0.33444	-0.46049	0.12113	0.01754	0.01542	0.12727	0.22747	0.10060
150	-0.49068	-0.65685	0.03390	0.00874	0.00845	0.24921	0.43990	0.19404
151	-0.45470	-0.61308	0.02535	0.00552	0.00538	0.2121	0.38124	0.17133
152	-0.35082	-0.48181	0.01743	0.00222	0.00218	0.1253	0.23432	0.10959
153	-0.32533	-0.44856	0.03171	0.00358	0.00347	0.1093	0.20467	0.09571
154	3.40595	2.25110	0.01750	0.21024	0.20657	11.8070	5.27403	0.07936
155	-0.39683	-0.54081	0.02894	0.00483	0.00469	0.1622	0.29717	0.13605
156	-0.18670	-0.26178	0.02490	0.00091	0.00089	0.0357	0.06942	0.03368
157	2.80918	2.09050	0.02891	0.24195	0.23495	8.1264	4.60514	0.11247
158	2.33717	1.93184	0.02559	0.14722	0.14345	5.6058	3.87545	0.15474
159	-0.39506	-0.53857	0.02386	0.00391	0.00381	0.1599	0.29387	0.13500
160	-0.33073	-0.45564	0.02109	0.00241	0.00236	0.1117	0.20996	0.09860

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est
161	-0.38619	-0.52729	0.03566	0.00572	0.00552	0.1547	0.28355	0.12979
162	-0.27861	-0.38667	0.05765	0.00504	0.00475	0.0824	0.15426	0.07203
163	2.78019	2.08168	0.02215	0.17908	0.17511	7.9046	4.50852	0.11455
164	-0.43635	-0.59041	0.08817	0.02019	0.01841	0.2088	0.36699	0.15995
165	-0.43552	-0.58938	0.02066	0.00409	0.00400	0.1937	0.35137	0.15944
166	-0.29864	-0.41336	0.01479	0.00136	0.00134	0.0905	0.17220	0.08189
167	1.74471	1.67167	0.03447	0.11256	0.10868	3.1527	2.90316	0.24728
168	-0.28898	-0.40051	0.01650	0.00142	0.00140	0.0849	0.16181	0.07707
169	-0.30039	-0.41568	0.02312	0.00219	0.00214	0.0924	0.17492	0.08277
170	-0.52982	-0.70343	0.03002	0.00896	0.00869	0.2894	0.50351	0.21918
171	-0.37089	-0.50772	0.02011	0.00288	0.00282	0.1404	0.26060	0.12093
172	-0.33332	-0.45902	0.02602	0.00305	0.00297	0.1141	0.21367	0.09999
173	-0.37953	-0.51878	0.02812	0.00429	0.00417	0.1482	0.27330	0.12590
174	2.06325	1.82185	0.04280	0.19885	0.19034	4.4474	3.50947	0.19022
175	-0.40400	-0.54989	0.02983	0.00517	0.00502	0.1682	0.30739	0.14031
176	-0.39935	-0.54400	0.02775	0.00468	0.00455	0.1640	0.30049	0.13754
177	2.38829	1.95074	0.01701	0.10039	0.09869	5.8026	3.90407	0.14917
178	-0.27502	-0.38187	0.01545	0.00121	0.00119	0.0768	0.14701	0.07032
179	-0.44350	-0.59927	0.03203	0.00672	0.00651	0.2032	0.36563	0.16436
180	-0.28731	-0.39829	0.01775	0.00152	0.00149	0.0840	0.16013	0.07625
181	-0.48572	-0.65088	0.04151	0.01066	0.01022	0.2461	0.43386	0.19089
182	-0.37677	-0.51525	0.07306	0.01207	0.01119	0.1531	0.27667	0.12431
183	-0.41405	-0.56255	0.12309	0.02744	0.02406	0.1955	0.34052	0.14635
184	-0.27255	-0.37856	0.04384	0.00356	0.00341	0.0777	0.14671	0.06915
185	-0.47327	-0.63579	0.03542	0.00853	0.00822	0.2322	0.41246	0.18300
186	-0.29434	-0.40764	0.02718	0.00249	0.00242	0.0891	0.16859	0.07973
187	-0.36618	-0.50166	0.05328	0.00797	0.00755	0.1416	0.25921	0.11824
188	-0.28490	-0.39508	0.01771	0.00149	0.00146	0.0826	0.15755	0.07508
189	-0.30992	-0.42828	0.02914	0.00297	0.00288	0.0989	0.18631	0.08763
190	-1.68748	-1.64161	0.55383	7.92234	3.53471	6.3823	6.22960	0.74010
191	-0.36456	-0.49957	0.04037	0.00583	0.00559	0.1385	0.25516	0.11731
192	-0.20649	-0.28898	0.01346	0.00059	0.00058	0.0432	0.08409	0.04089
193	-0.40752	-0.55433	0.02566	0.00449	0.00437	0.1704	0.31166	0.14242
194	-0.35617	-0.48874	0.03793	0.00520	0.00500	0.1319	0.24387	0.11258
195	2.62500	2.03257	0.02301	0.16607	0.16225	7.0529	4.29361	0.12673
196	2.36196	1.94106	0.03251	0.19378	0.18748	5.7664	3.95521	0.15200
197	-0.40048	-0.54543	0.02228	0.00374	0.00366	0.1640	0.30115	0.13821
198	-0.38935	-0.53132	0.02395	0.00381	0.00372	0.1553	0.28602	0.13164
199	-0.34690	-0.47672	0.01521	0.00189	0.00186	0.1222	0.22912	0.10741
200	-0.25711	-0.35780	0.04858	0.00355	0.00338	0.0695	0.13140	0.06201
201	-0.44883	-0.60585	0.03117	0.00669	0.00648	0.20793	0.37354	0.16767
202	-0.30265	-0.41866	0.02883	0.00280	0.00272	0.09431	0.17800	0.08391
203	-0.27892	-0.38709	0.01785	0.00144	0.00141	0.07921	0.15125	0.07218
204	-0.50896	-0.67875	0.04083	0.01150	0.01103	0.27007	0.47172	0.20574
205	-0.22337	-0.31206	0.02109	0.00110	0.00107	0.05097	0.09846	0.04752
206	-0.41240	-0.56047	0.02710	0.00487	0.00474	0.17481	0.31887	0.14535
207	-0.22436	-0.31340	0.01795	0.00094	0.00092	0.05126	0.09914	0.04792
208	-0.33493	-0.46112	0.03846	0.00467	0.00449	0.11666	0.21712	0.10086
209	-0.35536	-0.48769	0.03473	0.00471	0.00454	0.13082	0.24238	0.11212
210	-0.24903	-0.34690	0.05341	0.00370	0.00350	0.06552	0.12384	0.05840
211	-0.37694	-0.51547	0.02109	0.00313	0.00306	0.14515	0.26878	0.12441
212	-0.40038	-0.54531	0.02909	0.00495	0.00480	0.16511	0.30216	0.13816
213	-0.21408	-0.29936	0.02882	0.00140	0.00136	0.04719	0.09098	0.04382
214	-0.37064	-0.50739	0.04252	0.00637	0.00610	0.14347	0.26354	0.12078

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est
215	-0.32515	-0.44832	0.02853	0.00320	0.00311	0.10882	0.20410	0.09561
216	-0.23203	-0.32385	0.01497	0.00083	0.00082	0.05466	0.10569	0.05109
217	-0.36668	-0.50230	0.03670	0.00532	0.00512	0.13958	0.25743	0.11852
218	-0.70501	-0.89832	0.05556	0.03096	0.02924	0.52628	0.83622	0.33202
219	1.96088	1.77649	0.09365	0.43832	0.39727	4.24231	3.55318	0.20640
220	1.07548	1.23980	0.07617	0.10324	0.09537	1.25203	1.63249	0.46368
221	-0.62660	-0.81387	0.03681	0.01558	0.01501	0.40763	0.67739	0.28193
222	-0.60415	-0.78887	0.06235	0.02588	0.02427	0.38927	0.64658	0.26740
223	-0.57532	-0.75621	0.09138	0.03664	0.03329	0.36429	0.60515	0.24868
224	1.00467	1.18137	0.08357	0.10044	0.09205	1.10140	1.48767	0.49767
225	1.72570	1.66180	0.03393	0.10826	0.10459	3.08262	2.86616	0.25138
226	-0.52321	-0.69565	0.03317	0.00971	0.00939	0.28314	0.49331	0.21491
227	-0.45618	-0.61490	0.05496	0.01281	0.01210	0.22021	0.39021	0.17226
228	-0.64951	-0.83901	0.04924	0.02298	0.02185	0.44371	0.72579	0.29670
229	-0.57276	-0.75328	0.03317	0.01164	0.01126	0.33931	0.57868	0.24702
230	-0.51034	-0.68039	0.17088	0.06474	0.05368	0.31412	0.51660	0.20663
231	-0.58017	-0.76174	0.03366	0.01213	0.01172	0.34832	0.59197	0.25183
232	-0.54215	-0.71789	0.05121	0.01672	0.01587	0.30980	0.53124	0.22716
233	-0.58589	-0.76825	0.03786	0.01404	0.01351	0.35677	0.60371	0.25554
234	-0.45474	-0.61313	0.03752	0.00838	0.00806	0.21485	0.38399	0.17136
235	1.30616	1.41102	0.10962	0.23590	0.21004	1.91609	2.20102	0.36954
236	-0.54589	-0.72225	0.03999	0.01293	0.01241	0.31041	0.53406	0.22958
237	-0.53219	-0.70623	0.03645	0.01112	0.01071	0.29394	0.50947	0.22072
238	-0.57692	-0.75804	0.04562	0.01667	0.01591	0.34875	0.59053	0.24972
239	-0.98792	-1.16711	0.18654	0.27514	0.22382	1.19980	1.58595	0.49392
240	-0.58280	-0.76474	0.03453	0.01258	0.01215	0.35180	0.59697	0.25354
241	-0.51101	-0.68119	0.03947	0.01117	0.01073	0.27186	0.47475	0.20706
242	1.24284	1.36674	0.07197	0.12907	0.11978	1.66443	1.98776	0.39298
243	-0.53483	-0.70932	0.05056	0.01604	0.01523	0.30127	0.51836	0.22242
244	1.44567	1.50210	0.05953	0.14066	0.13229	2.22225	2.38861	0.32363

ANEXO D - Listagem do programa para Análise de Diagnóstico.

```
data geh;
input obs sexo idade pelagem diaria obito atendime novobs;
cards;
1   0   5   1   0   0   1   1
2   0   6   5   0   0   2   2
3   0   1   0   0   0   1   3
5   0   2   10  0   0   1   4
6   1   4   12  0   0   1   5
7   1   30  5   0   0   1   6
9   1   6   0   0   1   1   7
11  1   6   11  0   1   1   8
12  1   6   3   0   1   1   9
13  1   4   11  0   1   1   10
14  0   3   1   5   1   1   11
15  1   2   1   0   0   1   12
17  0   5   1   0   1   1   13
18  0   6   8   3   0   1   14
19  0   51  1   0   0   1   15
20  0   5   1   5   0   1   16
21  1   3   1   6   1   1   17
22  1   13  1   4   0   1   18
23  0   6   11  7   1   1   19
26  1   2   2   6   0   1   20
28  0   5   0   1   0   1   21
29  0   7   14  0   0   1   22
30  1   54  8   0   1   2   23
31  0   11  11  0   0   1   24
32  1   3   11  3   0   1   25
33  0   4   5   0   0   1   26
34  1   8   14  0   0   2   27
35  1   7   7   0   0   1   28
36  1   144 14  0   0   1   29
37  0   3   1   0   1   1   30
38  0   60  4   0   0   1   31
39  0   1   11  0   1   1   32
40  1   4   6   0   0   1   33
41  0   60  5   0   0   1   34
42  1   48  8   0   0   1   35
43  1   2   8   0   0   3   36
44  0   2   8   0   0   3   37
45  0   2   8   0   1   1   38
46  0   4   1   0   1   2   39
47  0   3   4   0   1   1   40
49  0   16  4   0   0   1   41
50  1   5   1   0   0   1   42
51  0   5   12  0   0   1   43
52  1   6   1   0   1   1   44
53  0   5   15  0   0   1   45
54  1   3   7   0   1   1   46
55  0   24  5   0   0   1   47
56  0   3   1   0   0   1   48
58  0   2   7   0   0   1   49
60  0   5   11  0   0   1   50
61  1   216 14  0   0   1   51
62  0   9   10  0   0   1   52
```

63	1	2	0	0	1	1	53
64	0	4	12	0	0	1	54
66	1	4	7	0	0	5	55
70	1	5	1	0	0	1	56
71	1	4	12	1	1	1	57
72	1	6	5	0	0	1	58
73	1	5	1	0	0	1	59
74	1	4	12	0	0	1	60
76	0	2	0	0	1	4	61
77	1	3	1	0	1	2	62
78	1	3	16	0	1	3	63
79	1	3	0	0	1	1	64
80	1	4	16	0	1	7	65
81	1	4	1	0	1	2	66
83	0	60	4	0	1	2	67
84	0	7	9	0	1	2	68
86	0	2	1	0	1	1	69
87	1	4	18	0	1	5	70
88	1	4	8	0	1	2	71
90	1	10	3	0	1	3	72
91	1	2	14	0	1	2	73
92	0	2	10	0	1	1	74
93	1	3	0	0	1	1	75
94	0	2	12	5	1	5	76
95	0	3	12	0	0	3	77
96	1	4	18	0	1	3	78
99	0	3	9	0	1	4	79
100	1	2	4	0	1	1	80
101	1	5	7	1	1	1	81
102	0	4	7	0	1	2	82
103	1	4	1	0	1	2	83
104	1	15	14	0	1	1	84
105	0	6	1	0	1	3	85
106	1	2	7	4	1	4	86
107	0	31	12	0	0	1	87
108	0	2	3	4	0	1	88
109	0	8	12	1	0	2	89
110	1	4	7	0	0	1	90
111	1	4	7	0	0	1	91
112	0	8	4	2	0	2	92
113	1	18	19	1	0	1	93
114	0	6	13	0	0	1	94
115	1	3	4	4	0	1	95
116	0	2	8	0	0	1	96
117	1	9	12	3	0	1	97
118	0	6	4	3	0	2	98
119	0	8	13	0	0	1	99
121	1	6	3	3	0	1	100
122	0	2	14	0	0	2	101
123	0	3	1	2	0	1	102
124	1	2	7	0	0	1	103
125	1	9	11	0	0	1	104
126	0	43	10	0	0	1	105
127	1	4	8	0	0	2	106
128	1	2	11	0	0	1	107
129	1	3	0	7	0	1	108

```
130 0   8   7   1   0   1   109
131 0   3   8   6   0   1   110
132 0   10  1   0   0   1   111
133 0   4   9   1   0   1   112
134 0   4   1   0   0   1   113
135 1   7   4   0   0   1   114
136 1   3   3   1   0   2   115
137 0   3   8   1   0   2   116
138 1   9   0   2   0   5   117
139 0   3   4   0   0   1   118
140 0   4   3   5   0   1   119
141 1   3   0   0   0   1   120
142 1   5   7   7   0   1   121
143 1   3   10  0   0   1   122
144 1   16  4   0   0   2   123
145 0   3   7   1   0   3   124
146 1   3   1   0   0   2   125
147 0   3   4   0   0   3   126
148 0   6   1   0   0   2   127
149 0   3   14  0   0   2   128
151 0   6   15  0   0   4   129
152 0   6   10  2   0   4   130
153 0   2   1   1   0   3   131
154 0   4   16  2   0   4   132
155 0   7   0   1   0   1   133
156 0   1   3   0   0   3   134
157 1   6   14  0   0   1   135
158 1   7   4   0   0   1   136
159 0   4   3   0   0   1   137
160 0   6   14  3   0   1   138
161 1   1   10  1   0   2   139
162 1   8   8   3   0   1   140
163 1   12  14  0   0   4   141
164 1   3   5   0   0   2   142
165 1   2   4   0   0   5   143
166 1   5   7   0   0   1   144
167 0   4   0   0   0   3   145
168 1   4   3   3   0   1   146
170 1   4   1   0   0   1   147
171 1   4   16  0   0   1   148
173 1   5   4   4   0   1   149
174 1   6   14  0   0   1   150
176 0   3   4   0   0   3   151
run;

proc print data=geh;
run;

* calcula as medidas de diagnostico de Pregibon;
proc logistic data=geh descending;
    model obito=sexo idade diaria atendime/influence iplots;
    output out=graf
    reschi=resd_chi
    resdev=resd_dev
    h=hat
    c=int_c
```

```
    cbar=int_cbar
    difchisq=d_chi
    difdev=d_dev
    predicted=predito
    xbeta=logit;
    run;

symbol1 i=none value=star color=red height=.75;
symbol2 i=none value=star color=green height=.75;
proc gplot data=graf;
    axis2 label= (color=blue 'Observações');
    **;
    axis1 label=(angle=-90 rotate=90 color=blue 'Resíduo de Pearson');
    plot resd_chi*novobs=1/frame overlay vaxis=axis1 haxis=axis2;
    run;
    axis3 label=(angle=-90 rotate=90 color=blue 'Resíduo Deviance');
    plot resd_dev*novobs=2/frame overlay vaxis=axis3 haxis=axis2;
    run;
    axis4 label=(angle=-90 rotate=90 color=blue 'Diagonal da matriz H');
    plot hat*novobs=1/frame overlay vaxis=axis4 haxis=axis2;
    run;
    axis5 label=(angle=-90 rotate=90 color=blue 'C');
    plot int_c*novobs=1/frame overlay vaxis=axis6 haxis=axis2;
    run;
    axis6 label=(angle=-90 rotate=90 color=blue 'CBAR');
    plot int_cbar*novobs=2/frame overlay vaxis=axis7 haxis=axis2;
    run;
    axis7 label=(angle=-90 rotate=90 color=blue 'Delta X^2');
    plot d_chi*novobs=1/frame overlay vaxis=axis8 haxis=axis2;
    run;
    axis8 label=(angle=-90 rotate=90 color=blue 'Delta Deviance');
    plot d_dev*novobs=2/frame overlay vaxis=axis9 haxis=axis2;
run;
proc print data=graf;
var resd_chi resd_dev hat int_c int_cbar d_chi d_dev predito logit;
run;
```

ANEXO E - Listagem do programa para Influência Local.

```
#################################################################
##                     Influência Local                        ##
##                                                             ##
##                   Regressão Logística                       ##
#################################################################
Dados<-read.table("A:/gastro.txt",header=TRUE)
attach(Dados)
Dados
ajuste02<-glm(obito~sexo+idade+diaria+atendime,data=Dados,family= binomial(logit))
summary(ajuste02)
anova(ajuste02)
###################################
obs<-1:151
uns<-c(rep(1,151))
# monta a matriz X
X<-cbind(uns,sexo,idade,pelagem,diaria,atendime)
# variavel obito ==>vetor Y
Y<-(Dados[,6])
beta<-ajuste02$coef
Xbeta<-X%*%beta
expXbeta<-exp(Xbeta)
#### probabilidade de sucesso e fracasso ####
PI<-(expXbeta/( 1 + expXbeta))
IPI<-(uns-PI)
# montando a matriz Q
Q<-(PI*IPI)
vetorQ<-c(Q)
MatrizQ<-diag(vetorQ)
Ibeta<-t(X)%*%MatrizQ%*%X
## pela inversa de Ibeta calculamos a matriz de variancia-covariacia
InvIbeta<-solve(Ibeta)
YPI<-(Y-PI)
#################################################################
##                      Caso Ponderado                         ##
#################################################################
matYPI<-c(rep(YPI,5))
matdelta<-matYPI*X
##### construindo a matriz H (não é matriz levarege)
H<-matdelta%*%InvIbeta%*%t(matdelta)
##### calcula o autovalor e autovetor de H
autH <- eigen(H)
##### separa os autovetores
autovetor<-autH$vectors
autvet<-c(autovetor[,1])
#### Curvatura
curv<-2* abs(t(autvet)%*%H%*%autvet)
curv

#####desenhado grafico
plot(obs,abs(autvet), xlab="Observações", ylab="",col=14, pch=16)
title("Grafico Caso Ponderado")
plot(obs,diag(H), col=4, pch=16)
title("Grafico i-ésimo individuo H")
#################################################################
##                 Perturbando as covariaveis                  ##
```

```
#####################################################################
X0<-uns
X1<-sexo
X2<-idade
X3<-diaria
X4<-atendime
#####################################################################
##               Perturbando a covariável sexo (X1)           ##
#####################################################################
beta1<-c(rep(0.58386068,151))
sq01<-sqrt(var(sexo))
v1<-c(rep( sq01,151))
X1<-sexo
a1<-((Y-PI)-(PI*IPI*beta1*X1))*v1
a0<- (-IPI*PI*v1*beta1*X0)
a2<- (-IPI*PI*v1*beta1*X2)
a3<- (-IPI*PI*v1*beta1*X3)
a4<- (-IPI*PI*v1*beta1*X4)
delta1<-cbind(a0,a1,a2,a3,a4)
tdelta1<-t(delta1)
H1 <- t(tdelta1)%*%InvIbeta%*%tdelta1
H1
autH1 <- eigen(H1)
autH1
autovetor1<-autH1$vectors
autovetor1
autvet1<-c(autovetor1[,1])
autvet1

#####desenhado grafico
plot(obs,abs(autvet1), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Idade")
curv1<-2* abs(t(autvet1)%*%H1%*%autvet1)
curv1
plot(obs,diag(H1), col=2, pch=16)
title("Grafico i-ésimo individuo H1")
#####################################################################
##               Perturbando a covariável idade (X2)           ##
#####################################################################
beta2<-c(rep(-0.01319756,151))
sq02<-sqrt(var(idade))
v2<-c(rep(sq02,151))
b2<- ((Y-PI)-(PI*IPI*beta2*X2))*v2

b0<- (-IPI*PI*v2*beta2*X0)
b1<- (-IPI*PI*v2*beta2*X1)
b3<- (-IPI*PI*v2*beta2*X3)
b4<- (-IPI*PI*v2*beta2*X4)
delta2<-cbind(b0,b1,b2,b3,b4)
tdelta2<-t(delta2)
H2 <- t(tdelta2)%*%InvIbeta%*%tdelta2
H2
autH2 <- eigen(H2)
autH2
autovetor2<-autH2$vectors
autovetor2
```

```
autvet2<-c(autovetor2[,1])
autvet2
#####desenhado grafico
plot(obs,abs(autvet2), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando Idade")
curv2<-2* abs(t(autvet2)%*%H2%*%autvet2)
curv2
plot(obs,diag(H2), col=2, pch=16)
title("Grafico i-ésimo individuo H2")
#####################################################################
##             Perturbando a covariável diaria (X3)            ##
#####################################################################
beta3<- c(rep(-0.09980164,151))
sq03<-sqrt(var(diaria))
v3<-c(rep( sq03,151))
c3<- ((Y-PI)-PI*(IPI)*beta3*X3)*v3

c0<- (-IPI*PI*v3*beta3*X0)
c1<- (-IPI*PI*v3*beta3*X1)
c2<- (-IPI*PI*v3*beta3*X2)
c4<- (-IPI*PI*v3*beta3*X4)
delta3<-cbind(c0,c1,c2,c3,c4)
tdelta3<-t(delta3)
H3 <- t(tdelta3)%*%InvIbeta%*%tdelta3
H3
autH3 <- eigen(H3)
autH3
autovetor3<-autH3$vectors
autovetor3
autvet3<-c(autovetor3[,1])
autvet3
#####desenhado grafico
plot(obs,abs(autvet3), xlab="Observações", ylab="",col=3, pch=16)
title("Grafico perturbando diaria")
curv3<-2* abs(t(autvet3)%*%H3%*%autvet3)
curv3
plot(obs,diag(H3), col=2, pch=16)
title("Grafico i-ésimo individuo H3")
#####################################################################
##           Perturbando a covariável atendimento (X4)           ##
#####################################################################
beta4<-c(rep( 0.31771649,151))
sq04<- sqrt(var(atendime))
v4<-c(rep(sq04,151))
d4<- ((Y-PI)-PI*(IPI)*beta4*X4)*v4

d0<-(-IPI*PI*v4*beta4*X0)
d1<-(-IPI*PI*v4*beta4*X1)
d2<-(-IPI*PI*v4*beta4*X2)
d3<-(-IPI*PI*v4*beta4*X3)
delta4<-cbind(d0,d1,d2,d3,d4)
tdelta4<-t(delta4)
H4 <- t(tdelta4)%*%InvIbeta%*%tdelta4
H4
autH4 <- eigen(H4)
autH4
```

```
autovetor4<-autH4$vectors
autovetor4
autvet4<-c(autovetor4[,1])
autvet4
#####desenhado grafico
plot(obs,abs(autvet4),xlab="Observações", ylab="", col=3, pch=18)
title("Grafico perturbando atendimento")
curv4<-2* abs(t(autvet4)%*%H4%*%autvet4)
curv4
plot(obs,diag(H4), col=2, pch=16)
title("Grafico i-ésimo individuo H4")

#####################################################################
##              Construindo o gráfico de envelope                  ##
#####################################################################
X<-model.matrix(ajuste02)
n<-nrow(X)
p<-ncol(X)
w<-ajuste02$weights
W<-diag(w)
MatrizH<-solve(t(X)%*%W%*%X)
MatrizH
MatrizH <- sqrt(W)%*%X%*%MatrizH%*%t(X)%*%sqrt(W)
h <- diag(MatrizH)
ts <- resid(ajuste02,type="pearson")/sqrt(1-h)
td <- resid(ajuste02,type="deviance")/sqrt(1-h)
e <- matrix(0,n,100)
#
for(i in 1:100){
    dif <- runif(n) - fitted(ajuste01)
    dif[ dif >= 0] <- 0
    dif[ dif < 0]  <- 1
    nresp <- dif
    fit <- glm(nresp~X, family=binomial)
    w <- fit$weights
    w
    W <- diag(w)
    W
    MatrizH <- solve(t(X)%*%W%*%X)
    MatrizH
    MatrizH <- sqrt(W)%*%X%*%MatrizH%*%t(X)%*%sqrt(W)
    h <- diag(MatrizH)
    e[,i] <- sort(resid(fit, type="deviance")/sqrt(1-h))}
#
e1 <- numeric(n)
e2 <- numeric(n)
#
for (i in 1:n){
    e0 <- sort(e[i,])
    e1[i]<- e0[5]
    e2[i]<- e0[95]}
#
med <- apply(e,1,mean)
faixa <- range(td,e1,e2)
#
par(mfrow=c(1,1)) par(pty="s")
```

```
qqnorm(td, xlab="Percentis da N(0,1)", ylab="Componente do Desvio Padronizado",
ylim=faixa,col="blue",pch=16)
par(new=T)
qqnorm(e1, axes=F,xlab="", ylab="",type="l", col="green", ylim=faixa,lty=1)
par(new=T)
qqnorm(e2,axes=F, xlab="", ,ylab="",type="l",col="green", ylim=faixa,lty=1)
par(new=T)
qqnorm(med, axes=F,xlab="", ylab="", type="l",col="red", ylim=faixa,lty=2) par(new=T)
```

ANEXO F - Listagem das medidas de diagnóstico.

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	predito	logit
1	-0.51854	-0.69012	0.01769	0.00493	0.00484	0.27372	0.48111	0.21190	-1.31349
2	-0.59416	-0.77761	0.01678	0.00613	0.00603	0.35905	0.61071	0.26091	-1.04123
3	-0.53363	-0.70792	0.01996	0.00592	0.00580	0.29056	0.50695	0.22165	-1.25610
4	-0.52982	-0.70344	0.01926	0.00562	0.00551	0.28622	0.50034	0.21918	-1.27045
5	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
6	-0.57583	-0.75679	0.03483	0.01240	0.01196	0.34354	0.58469	0.24901	-1.10389
7	1.46194	1.51218	0.01794	0.03975	0.03904	2.17630	2.32574	0.31875	-0.75953
8	1.46194	1.51218	0.01794	0.03975	0.03904	2.17630	2.32574	0.31875	-0.75953
9	1.46194	1.51218	0.01794	0.03975	0.03904	2.17630	2.32574	0.31875	-0.75953
10	1.44111	1.49926	0.01878	0.04052	0.03976	2.11656	2.28754	0.32501	-0.73083
11	2.38305	1.94882	0.04404	0.27370	0.26165	5.94056	4.05956	0.14972	-1.73676
12	-0.70394	-0.89720	0.02003	0.01034	0.01013	0.50566	0.81509	0.33134	-0.70213
13	1.92851	1.76160	0.01769	0.06818	0.06697	3.78611	3.17022	0.21190	-1.31349
14	-0.44955	-0.60674	0.02304	0.00488	0.00477	0.20686	0.37290	0.16812	-1.59902
15	-0.37278	-0.51014	0.05452	0.00848	0.00801	0.14698	0.26825	0.12201	-1.97353
16	-0.41365	-0.56205	0.04307	0.00805	0.00770	0.17881	0.32360	0.14611	-1.76546
17	1.87652	1.73713	0.07862	0.32609	0.30046	3.82179	3.31807	0.22117	-1.25884
18	-0.54293	-0.71880	0.04086	0.01309	0.01256	0.30733	0.52923	0.22767	-1.22154
19	2.66525	2.04563	0.06912	0.56658	0.52742	7.63096	4.71203	0.12340	-1.96059
20	-0.53674	-0.71156	0.07936	0.02697	0.02483	0.31292	0.53115	0.22366	-1.24449
21	-0.49562	-0.66280	0.01544	0.00391	0.00385	0.24949	0.44315	0.19720	-1.40388
22	-0.51115	-0.68135	0.01704	0.00461	0.00453	0.26580	0.46877	0.20715	-1.34219
23	1.78751	1.69348	0.10158	0.40210	0.36125	3.55645	3.22914	0.23837	-1.16165
24	-0.49669	-0.66408	0.01662	0.00424	0.00417	0.25087	0.44517	0.19788	-1.39958
25	-0.61028	-0.79573	0.02876	0.01136	0.01103	0.38348	0.64422	0.27137	-0.98766
26	-0.52227	-0.69454	0.01813	0.00513	0.00504	0.27780	0.48742	0.21431	-1.29914
27	-0.77817	-0.97311	0.01662	0.01041	0.01024	0.61579	0.95717	0.37716	-0.50161
28	-0.67913	-0.87095	0.01766	0.00844	0.00829	0.46952	0.76684	0.31564	-0.77387
29	-0.25415	-0.35382	0.23948	0.02674	0.02034	0.08493	0.14553	0.06067	-2.73964
30	1.90103	1.74876	0.01866	0.07001	0.06871	3.68263	3.12686	0.21674	-1.28479
31	-0.34947	-0.48006	0.06776	0.00952	0.00888	0.13101	0.23934	0.10884	-2.10267
32	1.87395	1.73590	0.01996	0.07296	0.07151	3.58319	3.08485	0.22165	-1.25610
33	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
34	-0.34947	-0.48006	0.06776	0.00952	0.00888	0.13101	0.23934	0.10884	-2.10267
35	-0.50607	-0.67530	0.07033	0.02084	0.01937	0.27548	0.47540	0.20389	-1.36217
36	-0.93758	-1.12318	0.02794	0.02599	0.02527	0.90432	1.28680	0.46782	-0.12891
37	-0.70567	-0.89901	0.02717	0.01430	0.01391	0.51187	0.82213	0.33243	-0.69722
38	1.88744	1.74233	0.01926	0.07135	0.06998	3.63241	3.10569	0.21918	-1.27045
39	1.65908	1.62631	0.01743	0.04969	0.04882	2.80138	2.69370	0.26648	-1.01253
40	1.90103	1.74876	0.01866	0.07001	0.06871	3.68263	3.12686	0.21674	-1.28479
41	-0.47919	-0.64297	0.01763	0.00419	0.00412	0.23374	0.41753	0.18674	-1.47133
42	-0.68895	-0.88139	0.01831	0.00902	0.00885	0.48350	0.78570	0.32187	-0.74518
43	-0.51854	-0.69012	0.01769	0.00493	0.00484	0.27372	0.48111	0.21190	-1.31349
44	1.46194	1.51218	0.01794	0.03975	0.03904	2.17630	2.32574	0.31875	-0.75953
45	-0.51854	-0.69012	0.01769	0.00493	0.00484	0.27372	0.48111	0.21190	-1.31349
46	1.43081	1.49280	0.01936	0.04121	0.04041	2.08762	2.26887	0.32817	-0.71648
47	-0.45246	-0.61032	0.02221	0.00476	0.00465	0.20937	0.37714	0.16993	-1.58612
48	-0.52603	-0.69898	0.01866	0.00536	0.00526	0.28197	0.49383	0.21674	-1.28479
49	-0.52982	-0.70344	0.01926	0.00562	0.00551	0.28622	0.50034	0.21918	-1.27045
50	-0.51854	-0.69012	0.01769	0.00493	0.00484	0.27372	0.48111	0.21190	-1.31349
51	-0.15162	-0.21321	0.21391	0.00796	0.00626	0.02924	0.05171	0.02247	-3.77275
52	-0.50387	-0.67267	0.01669	0.00438	0.00431	0.25819	0.45680	0.20248	-1.37089
53	1.42058	1.48634	0.02003	0.04210	0.04126	2.05930	2.25048	0.33134	-0.70213

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est	logit
54	-0.52227	-0.69454	0.01813	0.00513	0.00504	0.27780	0.48742	0.21431	-1.29914
55	-1.23098	-1.35823	0.07870	0.14049	0.12944	1.64474	1.97423	0.60243	0.41562
56	-0.68895	-0.88139	0.01831	0.00902	0.00885	0.48350	0.78570	0.32187	-0.74518
57	1.50774	1.53999	0.01634	0.03839	0.03777	2.31104	2.40935	0.30550	-0.82122
58	-0.68402	-0.87616	0.01794	0.00870	0.00855	0.47644	0.77620	0.31875	-0.75953
59	-0.68895	-0.88139	0.01831	0.00902	0.00885	0.48350	0.78570	0.32187	-0.74518
60	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
61	1.22790	1.35601	0.04811	0.08006	0.07621	1.58394	1.91497	0.39877	-0.41061
62	1.23978	1.36455	0.01785	0.02844	0.02793	1.56499	1.88994	0.39416	-0.42987
63	1.07426	1.23882	0.02750	0.03356	0.03263	1.18666	1.56731	0.46425	-0.14326
64	1.43081	1.49280	0.01936	0.04121	0.04041	2.08762	2.26887	0.32817	-0.71648
65	0.60992	0.79533	0.14689	0.07508	0.06405	0.43606	0.69660	0.72886	0.98884
66	1.24871	1.37093	0.01739	0.02809	0.02760	1.58687	1.90704	0.39074	-0.44422
67	2.47943	1.98332	0.08510	0.62502	0.57183	6.71940	4.50538	0.13991	-1.81606
68	1.69518	1.64572	0.01660	0.04931	0.04849	2.92213	2.75688	0.25816	-1.05558
69	1.88744	1.74233	0.01926	0.07135	0.06998	3.63241	3.10569	0.21918	-1.27045
70	0.81236	1.00675	0.07870	0.06119	0.05637	0.71630	1.06992	0.60243	0.41562
71	1.24871	1.37093	0.01739	0.02809	0.02760	1.58687	1.90704	0.39074	-0.44422
72	1.12958	1.28250	0.02747	0.03705	0.03604	1.31199	1.68084	0.43938	-0.24370
73	1.23092	1.35819	0.01842	0.02896	0.02843	1.54358	1.87310	0.39759	-0.41552
74	1.88744	1.74233	0.01926	0.07135	0.06998	3.63241	3.10569	0.21918	-1.27045
75	1.43081	1.49280	0.01936	0.04121	0.04041	2.08762	2.26887	0.32817	-0.71648
76	1.33373	1.42971	0.13568	0.32308	0.27924	2.05809	2.32332	0.35986	-0.57597
77	-0.70062	-0.89371	0.02666	0.01382	0.01345	0.50432	0.81217	0.32925	-0.71157
78	1.08199	1.24503	0.02717	0.03361	0.03270	1.20340	1.58280	0.46068	-0.15760
79	1.23674	1.36237	0.04768	0.08042	0.07658	1.60611	1.93265	0.39533	-0.42496
80	1.42058	1.48634	0.02003	0.04210	0.04126	2.05930	2.25048	0.33134	-0.70213
81	1.51859	1.54647	0.01595	0.03799	0.03739	2.34352	2.42894	0.30247	-0.83557
82	1.65908	1.62631	0.01743	0.04969	0.04882	2.80138	2.69370	0.26648	-1.01253
83	1.24871	1.37093	0.01739	0.02809	0.02760	1.58687	1.90704	0.39074	-0.44422
84	1.55945	1.57042	0.01884	0.04759	0.04669	2.47857	2.51291	0.29139	-0.88866
85	1.45835	1.50997	0.02574	0.05768	0.05620	2.18300	2.33622	0.31982	-0.75462
86	1.10731	1.26512	0.08520	0.12483	0.11419	1.34033	1.71473	0.44921	-0.20387
87	-0.43030	-0.58288	0.02864	0.00562	0.00546	0.19061	0.34521	0.15623	-1.68656
88	-0.44219	-0.59765	0.03336	0.00698	0.00675	0.20228	0.36393	0.16355	-1.63202
89	-0.55981	-0.73839	0.01464	0.00473	0.00466	0.31804	0.54988	0.23861	-1.16032
90	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
91	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
92	-0.53507	-0.70961	0.01776	0.00527	0.00518	0.29148	0.50872	0.22258	-1.25071
93	-0.59986	-0.78405	0.01910	0.00714	0.00701	0.36684	0.62174	0.26462	-1.02210
94	-0.51483	-0.68573	0.01732	0.00476	0.00467	0.26972	0.47489	0.20952	-1.32784
95	-0.58332	-0.76533	0.04185	0.01551	0.01486	0.35512	0.60059	0.25387	-1.07805
96	-0.52982	-0.70344	0.01926	0.00562	0.00551	0.28622	0.50034	0.21918	-1.27045
97	-0.58457	-0.76675	0.02752	0.00994	0.00967	0.35139	0.59758	0.25469	-1.07375
98	-0.51882	-0.69045	0.02527	0.00716	0.00698	0.27615	0.48370	0.21208	-1.31241
99	-0.50749	-0.67700	0.01683	0.00448	0.00441	0.26196	0.46274	0.20480	-1.35654
100	-0.59729	-0.78115	0.02775	0.01047	0.01018	0.36694	0.62038	0.26295	-1.03071
101	-0.61145	-0.79704	0.01843	0.00715	0.00702	0.38090	0.64229	0.27213	-0.98383
102	-0.48057	-0.64464	0.01827	0.00438	0.00430	0.23524	0.41986	0.18762	-1.46558
103	-0.70394	-0.89720	0.02003	0.01034	0.01013	0.50566	0.81509	0.33134	-0.70213
104	-0.66946	-0.86059	0.01741	0.00808	0.00794	0.45611	0.74855	0.30948	-0.80257
105	-0.39480	-0.53824	0.04326	0.00737	0.00705	0.16292	0.29675	0.13485	-1.85874
106	-0.80083	-0.99549	0.01739	0.01155	0.01135	0.65268	1.00236	0.39074	-0.44422
107	-0.70394	-0.89720	0.02003	0.01034	0.01013	0.50566	0.81509	0.33134	-0.70213

Obs	resd_chi	resd_dev	hat	int_c	int_cbar	d_chi	d_dev	pi_est	logit
108	-0.50935	-0.67922	0.10081	0.03235	0.02909	0.28852	0.49042	0.20600	-1.34923
109	-0.48507	-0.65009	0.01478	0.00358	0.00353	0.23882	0.42614	0.19047	-1.44693
110	-0.40109	-0.54621	0.05701	0.01031	0.00973	0.17060	0.30807	0.13858	-1.82715
111	-0.50027	-0.66836	0.01662	0.00430	0.00423	0.25450	0.45094	0.20017	-1.38523
112	-0.49919	-0.66708	0.01580	0.00406	0.00400	0.25319	0.44899	0.19948	-1.38954
113	-0.52227	-0.69454	0.01813	0.00513	0.00504	0.27780	0.48742	0.21431	-1.29914
114	-0.67913	-0.87095	0.01766	0.00844	0.00829	0.46952	0.76684	0.31564	-0.77387
115	-0.77095	-0.96589	0.01570	0.00963	0.00948	0.60385	0.94243	0.37279	-0.52026
116	-0.58025	-0.76184	0.01567	0.00544	0.00536	0.34206	0.58576	0.25189	-1.08858
117	-1.08496	-1.24741	0.08915	0.12650	0.11522	1.29236	1.67124	0.54068	0.16309
118	-0.52603	-0.69898	0.01866	0.00536	0.00526	0.28197	0.49383	0.21674	-1.28479
119	-0.41663	-0.56579	0.04353	0.00826	0.00790	0.18148	0.32802	0.14791	-1.75111
120	-0.69891	-0.89191	0.01936	0.00983	0.00964	0.49811	0.80514	0.32817	-0.71648
121	-0.50210	-0.67056	0.09921	0.03082	0.02777	0.27987	0.47741	0.20134	-1.37793
122	-0.69891	-0.89191	0.01936	0.00983	0.00964	0.49811	0.80514	0.32817	-0.71648
123	-0.73477	-0.92920	0.01996	0.01122	0.01100	0.55088	0.87441	0.35060	-0.61640
124	-0.66966	-0.86080	0.02444	0.01152	0.01123	0.45968	0.75222	0.30961	-0.80196
125	-0.80659	-1.00114	0.01785	0.01204	0.01182	0.66242	1.01410	0.39416	-0.42987
126	-0.70062	-0.89371	0.02666	0.01382	0.01345	0.50432	0.81217	0.32925	-0.71157
127	-0.59416	-0.77761	0.01678	0.00613	0.00603	0.35905	0.61071	0.26091	-1.04123
128	-0.60708	-0.79215	0.01788	0.00683	0.00671	0.37526	0.63421	0.26930	-0.99818
129	-0.79136	-0.98618	0.04701	0.03242	0.03089	0.65714	1.00345	0.38509	-0.46800
130	-0.72296	-0.91703	0.04926	0.02848	0.02708	0.54976	0.86802	0.34326	-0.64879
131	-0.67448	-0.86598	0.02487	0.01190	0.01160	0.46653	0.76151	0.31268	-0.78762
132	-0.73341	-0.92781	0.04944	0.02943	0.02798	0.56587	0.88880	0.34976	-0.62009
133	-0.48856	-0.65430	0.01493	0.00367	0.00362	0.24231	0.43173	0.19270	-1.43258
134	-0.71075	-0.90432	0.02778	0.01485	0.01443	0.51960	0.83224	0.33562	-0.68287
135	-0.68402	-0.87616	0.01794	0.00870	0.00855	0.47644	0.77620	0.31875	-0.75953
136	-0.67913	-0.87095	0.01766	0.00844	0.00829	0.46952	0.76684	0.31564	-0.77387
137	-0.52227	-0.69454	0.01813	0.00513	0.00504	0.27780	0.48742	0.21431	-1.29914
138	-0.44955	-0.60674	0.02304	0.00488	0.00477	0.20686	0.37290	0.16812	-1.59902
139	-0.78209	-0.97701	0.01678	0.01061	0.01044	0.62211	0.96498	0.37953	-0.49156
140	-0.58878	-0.77153	0.02751	0.01008	0.00981	0.35647	0.60507	0.25742	-1.05940
141	-1.00713	-1.18345	0.05121	0.05770	0.05475	1.06907	1.45531	0.50355	0.01422
142	-0.80659	-1.00114	0.01785	0.01204	0.01182	0.66242	1.01410	0.39416	-0.42987
143	-1.24877	-1.37097	0.07844	0.14403	0.13274	1.69216	2.01230	0.60929	0.44432
144	-0.68895	-0.88139	0.01831	0.00902	0.00885	0.48350	0.78570	0.32187	-0.74518
145	-0.69561	-0.88844	0.02626	0.01340	0.01305	0.49693	0.80237	0.32609	-0.72592
146	-0.60592	-0.79085	0.02834	0.01102	0.01071	0.37785	0.63615	0.26855	-1.00201
147	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
148	-0.69391	-0.88664	0.01878	0.00939	0.00922	0.49073	0.79534	0.32501	-0.73083
149	-0.57501	-0.75585	0.04102	0.01475	0.01414	0.34477	0.58545	0.24848	-1.10675
150	-0.68402	-0.87616	0.01794	0.00870	0.00855	0.47644	0.77620	0.31875	-0.75953
151	-0.70062	-0.89371	0.026665	0.013816	0.013448	0.50432	0.81217	0.32925	-0.71157

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo