Pontos de Alavanca em Regressão

Fábio Costa Andrade

Dissertação apresentada

ao

Instituto de Matemática e Estatística

da

Universidade de São Paulo

para

obtenção do grau

de

Mestre em Estatística

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Gilberto Alvarenga Paula**

São Paulo - Abril de 2004

Pontos de Alavanca em Regressão

Este exemplar corresponde à redação final
da dissertação de mestrado devidamente
corrigida e defendida por
Fábio Costa Andrade
e aprovada pela comissão julgadora.

São Paulo, Abril de 2004.

Banca examinadora:

- Prof. Dr. Gilberto Alvarenga Paula (Orientador) - IME-USP
- Prof. Dr. Francisco José de A. Cysneiros - UFPE
- Profa. Dra. Sílvia Nagib Elia - IME-USP
- Profa. Dra. Júlia Maria P. Soler - IME-USP
- Prof. Dr. Edwin M. M. Ortega - ESALQ-USP

Abstract

The aim of this work is to present some results on leverage in regression models. The basic idea behind the concept of leverage is that of evaluating the influence of each observation on its own predicted value. This influence can be well represented, when the response variable is continuous, by the derivative $\partial \hat{y}_i / \partial y_i$, that equals h_{ii} , the i th principal diagonal element of the projection matrix $\mathbf{H}$, in normal linear models. For more general regression models this derivative assumes a closed but more complex form. We discuss in the work the calculation of generalized leverage in linear and non-linear normal models as well as in exponential family non-linear models that have as a particular class the generalized linear models. Illustrative examples are presented throughout the text. We also point out that high leverage points may be influential on the parameter coefficients and can be useful to detect values of the response variable for which is more difficult to make prediction.

Agradecimentos

Agradeço, primeiramente, à Deus, e àqueles que acompanharam, cobraram e apoiaram à realização deste trabalho, principalmente a: Gilberto Alvarenga, Leila Teresinha, Maria Renata, Gisela Tunes, Tiago Costa, Michele Costa, Maria Elen, Sérgio Neto, Ana Roberta, Felipe Winter e Marcos Ruben. Agradeço à Universidade de São Paulo e ao Instituto de Matemática e Estatística.

E, também, àqueles que contribuíram ou estiveram envolvidos de alguma forma.

Fábio.

Sumário

Resumo	iii
Abstract	iv
Agradecimentos	v
1 Introdução	1
2 Pontos de alavanca em regressão normal linear	6
2.1 Introdução	6
2.2 Regressão normal linear	6
2.3 Medidas para pontos aberrantes e influentes	10
2.4 Imóveis	12
2.5 Cerejeiras	23
3 Pontos de alavanca em regressão não-linear	32
3.1 Introdução	32
3.2 Modelo aditivo-multiplicativo	33
3.3 Modelos não-lineares gerais	38
3.4 Coagulação de sangue com adição de plasma	44
3.5 Relação peso dos olhos e idade de coelhos selvagens	54
4 Modelos não-lineares de família exponencial	60
4.1 Introdução	60
4.2 Alavanca generalizada	61
4.3 Aplicações	63
4.3.1 Modelos não-lineares de família exponencial	63

<i>SUMÁRIO</i>	vii
4.3.2 Alavanca para modelos da família exponencial	66
4.4 Sobrevivência dos pacientes com leucemia	71
5 Conclusões	81
5.1 Conclusões	81
A Apêndice	83
Referências Bibliográficas	94

Capítulo 1

Introdução

No âmbito da análise de regressão podem ser encontradas várias técnicas de diagnóstico cujo objetivo principal é verificar se existem afastamentos sérios das suposições feitas para o modelo, tais como normalidade, independência e homoscedasticidade dos erros, relação funcional entre a média e os coeficientes das variáveis explicativas, e também verificar a presença de observações discrepantes.

As observações discrepantes podem, de uma forma geral, ser classificadas como aberrantes, influentes ou de alavanca (alto *leverage*), não necessariamente em uma única categoria. Observações aberrantes são aquelas mal ajustadas, caracterizadas por terem resíduo alto, afetam principalmente o intercepto do modelo. As observações influentes são aquelas que têm um peso desproporcional nas estimativas dos coeficientes. Já os pontos de alavanca são aqueles que têm uma influência desproporcional no próprio valor ajustado, os quais, em geral, estão posicionados em regiões remotas do subespaço gerado pela(s) coluna(s) da matriz modelo $\mathbf{X}$. Pontos de alavanca podem também ser influentes, porém, não é comum pontos de alavanca serem aberrantes. Na Figura 1 podemos ter uma idéia melhor sobre pontos aberrantes, influentes e de alavanca.

Figura 1.1: Comparação de pontos discrepantes.

Na Figura 1.1a podemos notar um ajuste perfeito da reta aos pontos, visto que os mesmos estão perfeitamente alinhados. Já na Figura 1.1b podemos observar o ponto 3 deslocado da reta e, claramente interferindo no valor do intercepto, o que o caracteriza como um ponto aberrante. Na Figura 1.1c temos um ponto que influencia pouco a inclinação da reta ajustada, porém, por estar mais afastado no subespaço de $\mathbf{X}$ possui um *leverage* alto. Finalmente, na Figura 1.1d temos um ponto altamente influente na estimativa da inclinação da reta ajustada, além de também ser um ponto de alavanca por estar mais afastado dos demais.

O objetivo principal deste trabalho é fazer uma revisão dos métodos de identificação de pontos de alavanca, conhecidos como *leverage* (utilizaremos a denominação alavanca), assim como seus desenvolvimentos teóricos (matriz Jacobiana de alavancas, plano tangente etc), além de avaliar seus comportamentos em alguns casos. Essa revisão foi realizada desde o trabalho de

Hoaglin e Welsch (1978), para a análise de variância e regressão até as generalizações recentes feitas por Wei, Hu e Fung (1998), para uma classe geral de estimação, que inclui entre outras: máxima verossimilhança, estimação Bayesiana e robusta.

Conforme será visto no próximo capítulo, o vetor de valores ajustados ($\hat{\mathbf{y}}$) pode ser escrito, para modelos de regressão normal linear, na forma:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y},$$

sendo $\mathbf{y}$ o vetor de valores observados da variável dependente e $\mathbf{H}$ é conhecida como matriz de projeção. Anscombe e Tukey (1963) esboçaram pela primeira vez a matriz $\mathbf{H}$ quando descreveram que os valores ajustados em uma regressão normal linear podem ser escritos como uma função linear dos valores observados,

$$\hat{y}_{ij} = \sum_{ij} h_{ij} y_{ij}$$

e usaram o exemplo da análise de variância com dois fatores.

A partir daí Hoaglin e Welsch(1978) escreveram um artigo em que explicam sobre as propriedades da matriz $\mathbf{H}$ dentro dos modelos lineares de regressão e de análise de variância. Em 1984, Emerson, Hoaglin e Kempthorne analisam a alavanca para modelos não-lineares de análise de variância com dois fatores, sendo o termo de não-linearidade uma interação entre os fatores de linha e de coluna. Mostram que, diferentemente dos modelos lineares, nos modelos não-lineares a alavanca, ou o impacto no ajuste de $\hat{y}_i$, quando de uma perturbação em y_i , depende do valor de y_i , assim como do tamanho da perturbação, além da estrutura do modelo. Levando isso em conta definiram duas novas medidas, a matriz de alavancas generalizadas e a matriz Jacobiana de alavancas, com base no conceito de que a alavanca “incorpora as alterações no valor de y observado”. Saint Laurent e Cook (1992) encontraram expressões para a matriz Jacobiana de alavancas e a alavanca generalizada em regressões não-lineares com uma única perturbação no vetor

resposta, através de uma aproximação de segunda ordem em série de Taylor do preditor linear, e compararam a alavanca generalizada e o plano tangente, como medidas de alavanca em regressões não-lineares. Também estudaram a superalavanca, caso em que a alavanca é maior do que 1.

Recentemente, Wei, Hu e Fung (1998) generalizaram a definição de pontos de alavanca para modelos bastante gerais cuja variável resposta seja contínua. Nessa generalização incluem-se outros métodos de estimação, além de máxima verossimilhança, e outros enfoques tais como o Bayesiano. Paula (1999) discute pontos de alavanca em modelos de regressão com parâmetros restritos na forma $C\beta \geq 0$, com extensões para os modelos lineares generalizados.

O trabalho se desenvolve apresentando no capítulo 2 o surgimento das medidas de identificação de pontos de alavanca. Nesse capítulo nos baseamos no trabalho de Hoaglin e Welsch (1978), que desenvolvem uma medida de alavanca por meio de uma matriz de projeção, representada por $\mathbf{H}$, para o caso de modelos de regressão lineares homoscedásticos, utilizando estimação via mínimos quadrados. Abordamos a medida $\mathbf{H}$ para o caso normal linear homoscedástico e apresentamos medidas comumente usadas para identificar pontos influentes e para identificar pontos discrepantes.

No capítulo 3 abordamos o caso dos modelos não-lineares de regressão, concentrados em dois artigos principais, Emerson, Hoaglin e Kempthorne (1984), os quais apresentam três medidas de alavanca: o plano tangente, a alavanca generalizada e a alavanca Jacobiana, para o caso específico do modelo de regressão aditivo-multiplicativo saturado. Já o trabalho de Saint Laurent e Cook (1992), dá prosseguimento ao trabalho anterior e estende as medidas de alavanca generalizada e alavanca Jacobiana para o caso não-linear geral, apresentando expressões fechadas para as duas medidas, obtidas por meio da expansão em série de Taylor de $\hat{\mathbf{y}}$. Os dois restringem-se apenas à estimação via mínimos quadrados.

Por fim, o capítulo 4 baseia-se no artigo de Wei, Hu e Fung (1998), no qual apresentam uma generalização da alavanca Jacobiana, em uma expressão fechada, com a qual pode-se estender sua aplicação para modelos lineares e não-lineares, além de estender para praticamente qualquer método de estimação. Nesse capítulo trabalhamos o método proposto pelos autores e damos uma atenção especial aos modelos da família exponencial. Em todos os capítulos apresentamos exemplos de aplicação das medidas para melhor visualização.

Capítulo 2

Pontos de alavanca em regressão normal linear

2.1 Introdução

No caso normal linear e nos outros casos, o foco do estudo da matriz de projeção do modelo (matriz $\mathbf{H}$) se concentra mais especificamente na análise de sua diagonal principal. Através da diagonal principal de $\mathbf{H}$ podemos detectar pontos de alavanca no ajuste do modelo. Uma razoável interpretação que se pode dar a alavanca, é que a mesma mede o quanto uma observação contribui com o valor ajustado de outra observação qualquer. Em geral, quando se fala em alavanca é em relação a quanto uma observação influencia o seu próprio valor ajustado. Essa medida é encontrada na diagonal principal da matriz $\mathbf{H}$.

2.2 Regressão normal linear

Visando fazer uma revisão nos conceitos básicos, considere inicialmente o modelo normal linear tradicional, o qual pode ser representado conforme

abaixo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.1)$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad , \quad \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \quad (2.2)$$

em que $\mathbf{X}$ é a matrix $n \times p$ de posto completo p de constantes conhecidas, $\mathbf{y}$ é o vetor de respostas de tamanho n , $\boldsymbol{\beta}$ é o vetor p -dimensional de parâmetros desconhecidos e $\boldsymbol{\varepsilon}$ é o vetor de erros com distribuição normal, não observáveis, de tamanho n .

Passando agora ao processo de estimação, vamos encontrar $\hat{\boldsymbol{\beta}}$ e conseqüentemente $\hat{\mathbf{y}}$, para tanto, utilizaremos mínimos quadrados como método de estimação. Obtemos então o seguinte estimador para $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Com isto podemos agora encontrar $\mathbf{H}$:

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= \mathbf{H} \mathbf{y}, \end{aligned}$$

logo,

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (2.3)$$

ainda que,

$$\hat{y}_i = \sum_{j=1}^n \mathbf{x}_j^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_j y_j,$$

o que mostra que $\hat{y}_i, \forall i = 1, 2, \dots, n$, é função de todo o vetor $\mathbf{y}$, em que $\mathbf{x}_k$ é o vetor referente à k -ésima linha da matrix $\mathbf{X}$, $\forall k = 1, 2, \dots, n$.

Figura 2.1: Descrição da projeção $\mathbf{H}\mathbf{y}$.

O uso do método de estimação de mínimos quadrados consiste em obter o vetor $\mathbf{X}\hat{\boldsymbol{\beta}}$ mais próximo, no subespaço gerado pelas colunas de $\mathbf{X}$, ao vetor $\mathbf{y}$. Conseqüentemente, denomina-se $\mathbf{H}$ matriz de projeção de $\mathbf{y}$, justo por projetar $\mathbf{y}$ no subespaço gerado pelas colunas de $\mathbf{X}$. Podemos ter uma boa idéia desse fenômeno na Figura 2.1.

A matriz $\mathbf{H}$ possui duas principais características, a primeira delas é a simetria, decorrente do fato de ser originária de uma forma quadrática. Outra bem conhecida é que $\mathbf{H}$ é idempotente, ou seja, $\mathbf{H}\mathbf{H} = \mathbf{H}$. Pelo fato de $\mathbf{H}$ ser idempotente e simétrica, ocorre que:

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

e

$$\sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 = h_{ii}. \quad (2.4)$$

Além disso, $\mathbf{H}$ é invariante sobre reparametrizações lineares não-singulares (vide, por exemplo, Cook e Weisberg, 1982).

Segundo Hoaglin e Welsch (1978), a influência do valor resposta y_i no ajuste é mais diretamente refletida na alavanca do correspondente valor ajustado $\hat{y}_i$, e isto é precisamente a informação contida em h_{ii} , o correspondente elemento da diagonal principal da matriz $\mathbf{H}$. Com isso, o foco na matriz $\mathbf{H}$ está na análise de sua diagonal principal. Cook e Weisberg (1982) fazem uma ressalva quanto ao uso de h_{ii} como medida de diagnóstico para encontrar pontos de alavanca (como sugerem Hoaglin e Welsch), sustentam que quando y_i for um ponto aberrante, $\hat{y}_i$ será altamente influenciado por y_i mesmo para um h_{ii} de valor pequeno, e utilizam a fórmula abaixo para melhor entendimento,

$$\hat{y}_i = h_{ii}y_i + \sum_{i \neq j} h_{ij}y_j.$$

Por (2.4) podemos perceber que $0 \leq h_{ii} \leq 1$, é maior ou igual a zero pois é resultado de uma soma de valores positivos (valores quadráticos), o que não pode gerar valores negativos, e menor ou igual a 1, uma vez que seu valor é resultado de uma soma em que um dos termos é seu próprio valor ao quadrado, portanto, se h_{ii} for maior do que 1 a expressão (2.4) não será mais verdadeira, visto que para $h_{ii} > 1$, h_{ii}^2 é sempre maior que h_{ii} . Um refinamento desse limite é observar que:

$$h_{ii} = \sum_{j=1}^n h_{ij}h_{ji} = \sum_{j=1}^n h_{ij}^2 \geq ch_{ii}^2,$$

o que implica $0 \leq h_{ii} \leq \frac{1}{c}$, em que c é o número de vezes que a i -ésima linha (na matriz $\mathbf{X}$) se repete.

Ainda por (2.4) também é interessante notar que para $h_{ii} = 0$ ou 1 temos que $h_{ij} = 0$, $\forall i \neq j$. Conseqüentemente, quando temos $h_{ii} = 0$, significa que pelo modelo, $\hat{y}_i$ será igual a zero qualquer que seja y_i , e para $h_{ii} = 1$ significa que o modelo sempre ajustará $\hat{y}_i = y_i$ e, portanto, o modelo estará dispondo de um parâmetro, apenas para ajustar esta observação, algo difícil de ocorrer na prática. Ficam faltando, portanto, formas para dimensionar

quais valores são altos ou baixos para a alavanca. Há na literatura algumas sugestões para balizar essa noção, a sugerida por Hoaglin e Welsch (1978) é o uso de 2 vezes a média dos valores da diagonal principal de $\mathbf{H}$, ou seja, acima de $\frac{2p}{n}$ passa-se a considerar que h_{ii} possua um valor alto. Já Velleman e Welsch (1981) sugerem o uso de $\frac{3p}{n}$. Para um modelo com intercepto e com as outras variáveis de $\mathbf{X}$ originárias de uma normal $(p-1)$ dimensional, temos que

$$\frac{(n-p)(h_{ii} - \frac{1}{n})}{(p-1)(1-h_{ii})},$$

tem distribuição F com $(p-1)$ e $(n-p)$ graus de liberdade (vide, Hoaglin e Welsch, 1978).

Uma relação interessante é a seguinte:

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}. \end{aligned}$$

Com isso percebe-se que a relação entre o erro aleatório do modelo ($\boldsymbol{\epsilon}$) e os resíduos ordinários ($\mathbf{e}$) depende totalmente de $\mathbf{H}$. Conseqüentemente, para se fazer diagnósticos para as suposições sobre os erros no modelo através dos resíduos, é interessante que se analise também o comportamento de $\mathbf{H}$.

2.3 Medidas para pontos aberrantes e influentes

Na seção anterior vimos que a matriz $\mathbf{H}$ é usada para a detecção de pontos de alavanca, para a detecção de pontos aberrantes e influentes descrevemos a seguir as medidas mais comumente usadas.

Para a detecção de pontos aberrantes a medida mais usada é o resíduo

studentizado, cuja expressão é a seguinte:

$$\begin{aligned} t_i &= \frac{r_i}{s(1 - h_{ii})^{\frac{1}{2}}} \\ &= \frac{y_i - x_i^T \hat{\beta}}{s(1 - h_{ii})^{\frac{1}{2}}}, \end{aligned}$$

em que r_i é o resíduo ordinário para a observação i , enquanto $s^2 = \sum_{i=1}^n \frac{r_i^2}{n-p}$ é o erro quadrático médio. Esse resíduo studentizado, contudo, não segue uma distribuição t de Student, visto que r_i e s^2 não são independentes (vide Cook e Weisberg, 1982). Portanto, costumeiramente usa-se a medida:

$$t_i^* = \frac{r_i}{s_{(i)}(1 - h_{ii})^{\frac{1}{2}}}, \quad (2.5)$$

que segue uma distribuição t de Student t_{n-p-1} , em que $s_{(i)}^2$ é o erro quadrático médio do ajuste sem a i -ésima observação ($s_{(i)}^2 = \sum_{j=1}^{n-1} \frac{r_j^2}{n-p-1}$).

Em relação à detecção de pontos influentes a medida mais usada é conhecida como distância de Cook (Cook, 1977), e foi motivada pela seguinte região de confiança para β :

$$\frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{ps^2} \leq F_{p, (n-p)}(1 - \alpha).$$

Visto que o interesse é medir a influência do i -ésimo ponto nas estimativas, o passo natural foi retirar a i -ésima observação e recalculá-la $\hat{\beta}$, gerando a medida distância de Cook:

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T \mathbf{X}^T \mathbf{X} (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2}, \quad (2.6)$$

em que,

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{r_i}{1 - h_{ii}} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i, \quad (2.7)$$

que consegue medir a influência da i -ésima observação sobre as estimativas dos parâmetros, pois mede a distância entre $\hat{\beta}_{(i)}$ e $\hat{\beta}$ segundo a métrica

$(\mathbf{X}^T \mathbf{X})$. Portanto, o critério para identificação de um ponto influente utilizando D_i , é que será considerado como influente qualquer ponto que apresentar $D_i > F_{p,(n-p-1)}(1 - \alpha)$.

D_i também pode ser visto de outra forma:

$$\begin{aligned} D_i &= \left[\frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{s(1 - h_{ii})^{\frac{1}{2}}} \right]^2 \frac{h_{ii}}{p(1 - h_{ii})} \\ &= \frac{t_i^2 V(\hat{y}_i)}{p V(r_i)}. \end{aligned}$$

Assim, temos D_i escrito em função da combinação de duas estatísticas já disponíveis, t_i que funciona na detecção de pontos aberrantes, e a razão entre a variância do i -ésimo valor ajustado e o i -ésimo resíduo ($V(\hat{y}_i) = \mathbf{x}_i^T (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i \sigma^2 = h_{ii} \sigma^2$, $V(r_i) = \sigma^2(1 - h_{ii})$), que dá uma medida do peso de y_i no seu ajuste como, conseqüentemente, no ajuste de $\hat{\boldsymbol{\beta}}$. A combinação dessas duas estatísticas dá uma medida do impacto da observação sobre as estimativas dos parâmetros. Observa-se também que D_i cresce à medida que t_i^2 e/ou h_{ii} cresce.

2.4 Imóveis

Temos um conjunto de dados sobre 27 imóveis (vide, por exemplo, Narula e Wellington, 1977), com as informações: imposto cobrado sobre o imóvel (o qual chamamos de imposto, em US\$ 100), área do terreno (chamamos de areat, em 1000 pés quadrados), área construída (chamamos de areac, em 1000 pés quadrados), idade do imóvel (chamamos de idade, em anos) e preço de venda do imóvel (chamamos de venda, em US\$ 1000). A intenção é construir um modelo de previsão do preço de venda com base nas informações disponíveis.

O diagrama de dispersão das variáveis explicativas (imposto, areat, areac e idade) contra a variável resposta (venda), mostra que imposto e areac pos-

suem muito claramente uma relação linear com venda, já areat não apresenta uma relação linear tão forte e com a idade teremos uma relação linear mais forte se desconsiderarmos as observações 9 e 10, como se pode observar na Figura 2.2.

O *box plot* das variáveis explicativas mostra que as variáveis imposto e areac possuem dois valores atípicos (observações 9 e 10, imposto: 15,42 e 14,46, areac: 3,42 e 3,00; respectivamente) e também assimetria à direita, o que pode-se observar nas Figuras 2.3 e 2.5. Enquanto o *box plot* de areat e idade não apresentam nenhum valor que possa ser considerado atípico, mas a primeira apresenta pequena assimetria à direita, o que pode-se observar nas Figuras 2.4 e 2.6. O *box plot* da variável dependente (resposta), venda, também evidencia a presença de dois valores atípicos: observações 9{84,9} e 10{82,9}, observe a Figura 2.7.

É importante ressaltar que tal conjunto de informações traz, intrinsecamente, uma possibilidade de multicolinearidade entre as variáveis, devido a clara não interdependência entre elas, contudo, não faremos essa análise por não ser tão importante neste momento, mas poderá ser realizada em outro momento.

Buscamos ajustar um modelo normal linear supondo erros independentes e identicamente distribuídos para esse conjunto de dados. A busca por um modelo adequado para esse conjunto de dados foi feita utilizando-se o método de seleção *stepwise*, o qual nos levou a um ajuste em que alguns resultados são apresentados na expressão 2.8. Tal ajuste descarta (pela falta de significância) as variáveis área do terreno e idade do imóvel, além do intercepto, apresentando um coeficiente de determinação do modelo de $R^2 = 0,99$, mas mostra um ajuste muito bom das variáveis areac e imposto, como se pode observar também pelas Figuras 2.10 e 2.11. O modelo ajustado fica dado por (2.8),

$$\widehat{\text{venda}} = 14,267(2,293)\text{areac} + 2,324(0,474)\text{imposto}. \quad (2.8)$$

Figura 2.2: Diagrama de dispersão das variáveis do modelo.

Figura 2.3: *Box plot* de imposto. Figura 2.4: *Box plot* de área do terreno.

Figura 2.5: *Box plot* de área construída. Figura 2.6: *Box plot* de idade.

Figura 2.7: *Box plot* de venda.

A Figura 2.8 mostra que os resíduos studentizados estão dentro dos limites para uma variável aleatória com distribuição t-Student com 24 graus de liberdade ao nível de confiança de 95% (aproximadamente $(-2,06; 2,06)$), exceto a observação 10 (2,11), e, portanto, pode ser considerada um valor aberrante para o modelo ajustado nesse conjunto de dados. Na Figura 2.9 observamos o gráfico de t_i^* versus os valores ajustados, pelo qual podemos considerar que existe indícios de homoscedasticidade, dada a disposição dos pontos no gráfico. E a Figura 2.12 mostra, através de um gráfico normal de probabilidades com envelopes¹, que tais resíduos possuem uma distribuição

¹O gráfico de envelope é um conjunto de intervalos de confiança para cada resíduo studentizado (visto que a construção de intervalos de confiança para a série toda é muito complicado dado a complexidade da distribuição conjunta das estatísticas de ordem). Esses intervalos são construídos através de um processo de simulação que gera m (em geral 100) resíduos studentizados para cada um dos n resíduos encontrados no ajuste do modelo, e

Figura 2.8: Resíduos studentizados segundo o índice.

que pode ser considerada normal, visto que a distribuição dos resíduos se distribui de forma aleatória dentro do envelope gerado em torno da reta de 45 graus.

Observando as Figuras 2.13 e 2.14, notamos na primeira claramente dois pontos atípicos, porém, o maior valor é de 2,1 aproximadamente, e o valor de uma distribuição F para o valor crítico de 5% com parâmetros 2 (p) e 25 ($n - p$), possui um valor de 3,385, e portanto, nenhuma observação pode ser considerada influente segundo esse critério, já na segunda temos três observações que ultrapassam o valor dos critérios apresentados para a medida ($\frac{2p}{n}$ e $\frac{3p}{n}$), e, portanto, podem ser considerados pontos de alavanca, as observações 10, 9 e 27, por ordem crescente de grandeza.

A observação 9 é do imóvel com maior valor de venda, o maior imposto

através de estatísticas de ordem obtém-se os limites de confiança superior e inferior para cada um dos n resíduos.

Figura 2.9: Resíduos studentizados contra os valores ajustados.

Figura 2.10: Resíduos studentizados contra área construída.

Figura 2.11: Resíduos studentizados contra imposto.

Figura 2.12: Envelope dos resíduos studentizados.

Figura 2.13: Distância de Cook.

Figura 2.14: Diagonal da matriz de projeção.

Tabela 2.1: Resultados dos ajustes do modelo sem uma observação.

Observação Retirada	Variável Explicativa	Estimativa Nova	Percentual de Alteração em módulo
9	areac	14,233	0,24
	imposto	2,328	0,17
10	areac	14,080	1,31
	imposto	2,296	1,20
27	areac	9,956	30,22
	imposto	3,272	40,79

e o maior terreno; a observação 10 é referente a um imóvel que possui um dos maiores valores de venda, de imposto e de área construída; a observação 27 é de um imóvel com valores de venda e de imposto relativamente altos, porém, com uma área construída relativamente pequena, cujo valor encontra-se próximo ao valor do primeiro quartil da distribuição dos valores das áreas de construção dos imóveis, o que claramente as coloca em regiões bastante remotas de observações, contudo, a observação 27 é a que está na posição mais remota visualmente, o que corrobora o alto valor de alavanca $h_{27,27}$.

Refazendo o ajuste do modelo retirando em cada vez uma das observações 9, 10 e 27, obtemos os resultados expostos na Tabela 2.1. Esses resultados mostram uma grande alteração nas estimativas quando retira-se a observação 27, e talvez seja uma indicação de que o nível de confiança usado para verificar se a observação é influente, com a Distância de Cook, seja muito conservador.

A conclusão a que chegamos é de que a observação 10 é uma observação que não entrou muito bem ajustada no modelo, mas pode ser considerada como uma observação alavanca, já a observação 9 trata-se de uma observação puramente de alavanca, visto que não possui resíduo alto(0,093), e possui alto valor para as variáveis explicativas do modelo (portanto, esparsa no espaço gerado pelas colunas da matriz $\mathbf{X}$). Quanto à observação 27 pode ser

considerada tanto de alavanca como influente, vistos os resultados da Figura 2.14 e da Tabela 2.1.

2.5 Cerejeiras

Neste exemplo usamos um conjunto de informações sobre uma amostra de 31 cerejeiras de uma floresta da Pensilvânia, EUA, em que observou-se o diâmetro (d), a altura (h) e o volume (v) (Ryan e Joiner, 1994). Dependendo da forma da árvore existem duas possibilidades de relação do volume com a altura e o diâmetro. Quando a forma é cilíndrica a relação é

$$v = \frac{1}{4}\pi d^2 h,$$

e quando a forma é cônica a relação torna-se

$$v = \frac{1}{12}\pi d^2 h.$$

Nos dois casos podemos escrever

$$\ln(v) = a + b\ln(d) + c\ln(h).$$

em que $\ln$ é o logaritmo na base e .

As medidas do $\ln(\text{diâmetro})$, $\ln(\text{altura})$ e $\ln(\text{volume})$ não apresentaram nenhuma observação que se caracterize claramente como um ponto atípico, devido a não apresentarem observação(ões) com valor(es) muito alto(s) ou baixo(s) em relação aos demais, como se pode perceber observando as Figuras 2.15, 2.16 e 2.17.

O diagrama de dispersão (Figura 2.18) da variável $\ln(\text{volume})$ (resposta) e das variáveis $\ln(\text{diâmetro})$ e $\ln(\text{altura})$ (explicativas), mostra uma clara relação linear entre $\ln(\text{volume})$ e $\ln(\text{diâmetro})$, já a relação entre $\ln(\text{volume})$ e $\ln(\text{altura})$ é bem mais esparsa, não trazendo a idéia de uma relação linear, assim como a relação entre $\ln(\text{altura})$ e $\ln(\text{diâmetro})$.

Figura 2.15: *Box plot* do $\ln(\text{diâmetro})$.

Aplicando um modelo linear na forma $\ln(\mathbf{v}) = \alpha + \beta \ln(\mathbf{d}) + \gamma \ln(\mathbf{h}) + \epsilon$ a esse conjunto de dados, supondo $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, obtemos as estimativas contidas na Tabela 2.2 para os parâmetros.

Pela Tabela 2.2 verifica-se que todos os parâmetros do modelo são significativos com um nível de confiança de mais de 99%. O coeficiente de determinação do modelo ajustado ficou em $R^2 = 0,98$, com a estimativa de σ^2 de 0,081.

Tabela 2.2: Resultados do ajuste do modelo.

Parâmetro	Estimativa	Erro Padrão	Valor de t	$Pr(> t)$
Intercepto	-6,632	0,800	-8,292	0,000
$\ln(\text{diâmetro})$	1,983	0,075	26,432	0,000
$\ln(\text{altura})$	1,1171	0,204	5,464	0,000

Figura 2.16: *Box plot* do $\ln(\text{altura})$.

O cruzamento dos resíduos studentizados com as variáveis $\ln(\text{diâmetro})$, $\ln(\text{altura})$ e com os valores ajustados, Figuras 2.19, 2.20 e 2.21, apresentam as observações 15 e 18 fora dos limites, podendo serem consideradas como aberrantes. Além disso, não há indícios fortes de heteroscedasticidade.

O gráfico de envelope, Figura 2.22, mostra que a distribuição dos erros pode ser considerada normal, visto que não há nenhuma tendência dentro do envelope, o que reforça ainda mais a adequação de um ajuste de um modelo que supõe distribuição normal para os erros.

Para identificação de pontos de alavanca analisamos a diagonal principal da matriz de projeção $\mathbf{H}$, cujos valores estão dispostos em ordem na Figura 2.23, a qual mostra que as observações 20, 3 e 31, respectivamente, são suspeitas de serem de alavanca, por estarem próximo ou acima do critério de $\frac{2p}{n} = 0,193$ para h_{ii} , porém, abaixo do critério $\frac{3p}{n} = 0,290$.

Para identificação de ponto(s) influente(s) utilizamos a estatística D_i

Figura 2.17: *Box plot* do $\ln(\text{volume})$.

(2.6). Na Figura 2.24 identifica-se claramente que a observação 18 possui bem mais influência sobre as estimativas dos parâmetros do que as outras observações, visto que apresenta alto (em relação aos demais) valor para D_i , porém, dado que $F_{3,27}(0.95) = 2,96$, esta observação apresenta um valor bem abaixo do que seria necessário para considerá-la influente usando essa estatística.

As observações 15 e 18 estão mal ajustadas pelo modelo, apresentam valores baixos para o volume, a observação 18 possui a segunda maior altura, mas o volume é apenas mediano. As observações 3 e 31 encontram-se em posição oposta no espaço, pois a primeira o menor valor para a altura e o terceiro menor para o diâmetro, enquanto a segunda apresenta os maiores valores, com boa distância para os outros valores, para a altura e o diâmetro. A observação 20 apresenta a segunda menor altura, porém, com diâmetro mediano, o que a coloca numa posição esparsa do espaço. Portanto, as ob-

Figura 2.18: Diagrama de dispersão das variáveis do modelo.

Figura 2.19: Resíduos studentizados contra o $\ln(\text{diâmetro})$.

servações 3 e 31, por estarem nos extremos, possuem grande possibilidade de estarem em posições esparsas, contudo, a observação 20 está em uma posição que não é extrema, mas, realmente, é esparsa e pode ser considerada uma observação de alavanca.

Figura 2.20: Resíduos studentizados contra o $\ln(\text{altura})$.

Figura 2.21: Resíduos studentizados contra os valores ajustados.

Figura 2.22: Envelope dos resíduos studentizados.

Figura 2.23: Diagonal da matriz de projeção.

Figura 2.24: Distância de Cook.

Capítulo 3

Pontos de alavanca em regressão não-linear

3.1 Introdução

O desenvolvimento de medidas de alavanca para o caso não-linear iniciou-se no artigo de Emerson, Hoaglin e Kempthorne (1984), que se basearam na definição de alavanca comumente usada no caso linear, para pensar em medidas que fossem capazes de representá-la de forma apropriada para o caso não-linear. Justificam o desenvolvimento do estudo devido à pouca atenção, à época, do estudo de alavanca para modelos não-lineares, diferentemente do caso linear.

A primeira observação que fazem é que, devido à não linearidade, o impacto no ajuste de uma perturbação depende do valor da observação assim como do tamanho da perturbação. No caso linear, como vimos, isto não acontece. Dessa forma, fica claro que há risco na aplicação do mesmo método utilizado no caso linear para o caso não-linear, visto que o mesmo não leva em consideração o tamanho da observação assim como o tamanho da perturbação.

3.2 Modelo aditivo-multiplicativo

No que concerne ao estudo do diagnóstico de pontos de alavanca em modelos não-lineares, Emerson, Hoaglin e Kempthorne (1984) iniciaram este estudo utilizando um modelo de efeitos de linha e de coluna cuja parte não-linear é representada por um termo de interação multiplicativo, o qual denominam modelo com dois fatores aditivo-multiplicativo, com apenas uma observação por célula, ou seja,

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\gamma_i\delta_j + \epsilon_{ij} \quad , \quad (3.1)$$

$$i = 1, \dots, r \quad e \quad j = 1, \dots, c.$$

O método de estimação é parecido com o caso linear (em que não há o termo de interação), porém possui mais alguns detalhes. As restrições usadas são as tradicionais $\sum \alpha_i = \sum \beta_j = 0$, mais os termos $\sum \gamma_i = \sum \delta_i = 0$, $\sum \gamma_i^2 = \sum \delta_j^2 = 1$, e para garantir unicidade impõe-se a condição de que λ seja não negativo e que o primeiro elemento não zero de $\gamma = (\gamma_1, \dots, \gamma_r)^T$, seja positivo. Utilizaram um método (ver seção 2.2 do artigo), que para encontrar as estimativas dos parâmetros do modelo (3.1), obtém, primeiramente, as estimativas dos fatores através do modelo aditivo,

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad , \quad (3.2)$$

$$i = 1, \dots, r \quad e \quad j = 1, \dots, c,$$

sendo o termo $\lambda\gamma_i\delta_j$ estimado através da decomposição do valor singular de $\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}}$ ($\mathbf{Y}$ é matriz de valores observados com dimensões $r \times c$ e $\hat{\mathbf{Y}}$ é a matriz de valores ajustados do modelo (3.2)). Portanto, chegam à estimativa dos parâmetros do modelo (3.1) pelo seguinte ajuste:

$$\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} + \mathbf{su}\mathbf{v}^T, \quad (3.3)$$

em que $\tilde{\mathbf{Y}}$ é a matriz das estimativas de (3.1), e $\mathbf{su}\mathbf{v}^T$ é a matriz das estimativas dos termos de interação $\lambda\gamma_i\delta_j$, no qual s é o maior autovalor, $\mathbf{u}$ e $\mathbf{v}^T$ são os autovetores ortogonais da matriz de resíduos ($\mathbf{R}$), e formam a matriz das estimativas dos termos de interação $\lambda\gamma_i\delta_j$.

No artigo, os autores desenvolvem essas duas medidas para o caso especial em que o modelo aditivo-multiplicativo se ajusta perfeitamente aos dados (modelo saturado), ou seja,

$$\begin{aligned} y_{ij} &= m + g_i + d_j + su_i v_j, \\ i &= 1, \dots, r \quad e \quad j = 1, \dots, c, \end{aligned}$$

em que,

$$m = \hat{\mu} = \bar{y}_{..},$$

$$g_i = \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..},$$

$$d_j = \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..},$$

$\bar{y}_{i.}$: é a média de $\mathbf{y}$ em relação ao fator j ,

$\bar{y}_{.j}$: é a média de $\mathbf{y}$ em relação ao fator i ,

$\bar{y}_{..}$: é a média de $\mathbf{y}$ em relação aos fatores i e j simultaneamente,

$su_i v_j$: é obtido através da decomposição do valor singular de $\mathbf{R}$

($\mathbf{R} = s_1 u_1 v_1 + s_2 u_2 v_2 + \dots + s_r u_r v_r$, em que r é o posto da matriz $\mathbf{R}$),

sendo fruto do primeiro termo dessa decomposição ($\mathbf{su}\mathbf{v}^T = s_1 \mathbf{u}_1 \mathbf{v}_1^T$).

A grande contribuição dos autores é a definição de duas novas medidas para avaliar as alterações nos valores ajustados, quando de uma perturbação em um valor observado. São definidas como *generalized leverage* (matriz de alavancas generalizadas) e *Jacobian leverage* (matriz Jacobiana de alavan-

cas), respectivamente:

$$\mathbf{G}(p; k, l) = \frac{\tilde{\mathbf{Y}}_{k,l;p} - \tilde{\mathbf{Y}}}{p}, \quad (3.4)$$

$$\mathbf{J}(k, l) = \lim_{p \rightarrow 0} \mathbf{G}(p; k, l) = \left(\frac{\partial \tilde{y}_{ij}}{\partial y_{kl}} \right), \quad (3.5)$$

$$k, i = 1, \dots, r \quad e \quad l, j = 1, \dots, c,$$

no qual o termo,

$$\begin{aligned} \hat{\mathbf{Y}}_{k,l;p} &= \hat{\mathbf{Y}} + p\mathbf{H}(k, l) \\ k &= 1, \dots, r \quad e \quad l = 1, \dots, c. \end{aligned}$$

Devido ao modelo aditivo com apenas uma observação por célula ser balanceado, a matriz $\mathbf{H}(k, l)$ é a mesma para $k = 1, \dots, r$ e $l = 1, \dots, c$, com a seguinte estrutura,

$$\begin{aligned} h_{kl} &= \frac{r + c - 1}{rc}, \\ h_{kj} &= \frac{r - 1}{rc}, \forall j \neq l, \\ h_{il} &= \frac{c - 1}{rc}, \forall i \neq k, \\ h_{ij} &= \frac{-1}{rc}, \forall i \neq k, j \neq l, \end{aligned}$$

já para o caso não-linear os autores apenas mencionam que o valor dependerá da perturbação (p) na célula e o próprio valor da célula perturbada. A partir daí os autores introduzem as duas novas medidas considerando-as aproximações promissoras para medir alavanca.

A primeira medida (3.4) visa mensurar o impacto nos valores ajustados no caso em que p não é pequeno, e a segunda (3.5) para o caso em que p é infinitesimal.

Por meio da tabela a seguir podemos comparar os valores da matriz Jacobiana de alavancas para o modelo aditivo com dois fatores, e para o modelo aditivo-multiplicativo (para perturbação na célula (k, l) nos dois casos):

Caso	Modelo Aditivo	Modelo Aditivo-Multiplicativo
mesma célula ($i = k, j = l$)	$\frac{r+c-1}{rc}$	$\frac{r+c-1}{rc} + \frac{c-1}{c}u_k^2 + \frac{r-1}{r}v_l^2 - u_k^2v_l^2$
mesma linha ($i = k, j \neq l$)	$\frac{r-1}{rc}$	$\frac{r-1}{rc} + \frac{-1}{c}u_k^2 + \frac{r-1}{r}v_jv_l - u_k^2v_jv_l$
mesma coluna ($i \neq k, j = l$)	$\frac{c-1}{rc}$	$\frac{c-1}{rc} + \frac{c-1}{c}u_iu_k + \frac{-1}{r}v_l^2 - u_iu_kv_l^2$
outras células ($i \neq k, j \neq l$)	$\frac{-1}{rc}$	$\frac{-1}{rc} + \frac{-1}{c}u_iu_k + \frac{-1}{r}v_jv_l - u_iu_kv_jv_l$

Podemos observar pela tabela que as fórmulas para o modelo aditivo-multiplicativo contêm os termos, respectivamente, do modelo apenas aditivo. Outro fato é que as fórmulas no modelo aditivo-multiplicativo possuem também a interferência dos valores de $\mathbf{u}$ e $\mathbf{v}$, o que mostra que a alavanca Jacobiana é influenciada pelo próprio valor de y_{ij} , ou mais especificamente, pela parte não-linear no ajuste.

Os autores constataram que o valor da alavanca generalizada é maior para as células que sofrem a perturbação do que para outras células, porém, isso não quer dizer que para as outras células não possa acontecer valores altos. O comportamento da alavanca generalizada em função da perturbação não é constante e é diferente para cada célula perturbada, conforme pode ser visto na Figura 3.1, resultado do experimento realizado pelos autores variando p e observando o comportamento de $\mathbf{G}(p; k, l)$, para as células (1;6) e (4;4), no estudo realizado na Seção 5 do artigo. Observou-se a ocorrência de valores maiores do que 1 (*superalavanca*) na alavanca generalizada das células perturbadas.

Assim, como a alavanca generalizada e a matriz Jacobiana de alavancas, o maior autovalor (s_1) da matriz perturbada ($\hat{\mathbf{Y}}(p; k, l)$), também sofre alterações diferenciadas por célula perturbada e não é constante em relação à p . Os autores apresentam um gráfico (Figura 3.2) do maior autovalor versus a perturbação nas células (1;6) e (4;4), no qual podemos perceber a possibilidade de drásticas alterações nas estimativas do modelo e, conseqüentemente,

Figure 1. Plots of leverage for the perturbed cell against p in one 7×6 table with pure multiplicative structure. The solid curve corresponds to perturbations in cell (1, 6), and the dashed curve corresponds to perturbations in cell (4, 4).

Figura 3.1: Gráfico da alavanca generalizada contra a perturbação.

nos valores ajustados à medida que a perturbação aumenta em módulo.

Chegaram à conclusão que a matriz Jacobiana de alavancas não fornece nem o máximo nem o mínimo exatos que a alavanca generalizada pode alcançar, mas é uma aproximação muito boa do mínimo, o que, muito provavelmente, será substancialmente superior ao valor obtido para o modelo apenas aditivo.

Concluem o artigo observando que os resultados mostram que uma observação anômala pode afetar fortemente o ajuste. Sustentam o uso da matriz Jacobiana de alavancas em detrimento da matriz de alavancas generalizada devido ao seu fácil cômputo e interpretação. Finalizam, considerando que em situações práticas não há como saber qual parte de alguma(s) observação(ões) é a perturbação e qual não é, o que inviabiliza a abordagem de medição de alavanca por meio de perturbação.

Figure 2. Perturbed singular value s_1 versus p for one 7×6 table with pure multiplicative structure. The solid curve corresponds to perturbations in cell (1, 6), and the dashed curve corresponds to perturbations in cell (4, 4).

Figura 3.2: Gráfico do valor singular contra a perturbação.

3.3 Modelos não-lineares gerais

Vimos na seção anterior um estudo focalizando mensuração da alavanca em modelos não-lineares aditivo-multiplicativo, agora apresentamos outro estudo mais abrangente, pois amplia a mensuração da alavanca para modelos não-lineares gerais, ou seja, com qualquer parametrização não-linear.

Saint Laurent e Cook (1992) usam a definição de outro autor (ver seção 1) para alavanca em relação à i -ésima observação: “Independente da natureza da resposta contínua, alavanca pode ser definida como a magnitude da derivada do i -ésimo valor predito com respeito à i -ésima resposta”. O que corresponde ao conceito da matriz Jacobiana de alavancas se trocarmos a i -ésima pela j -ésima resposta. Verificam que, para o caso não-linear, as restrições $0 \leq h_{ii} \leq 1$ não valem, e realmente os valores preditos podem crescer com maior velocidade do que os valores observados.

Saint Laurent e Cook (1992) utilizam um modelo não-linear na forma

(3.6), em que $\eta_i(\boldsymbol{\beta})$ é uma função não-linear em $\boldsymbol{\beta}$ e conhecida, ϵ_i 's são erros mutuamente independentes, com média zero e variância σ^2 , para $i = 1, 2, \dots, n$. Note que $\mu_i(\boldsymbol{\beta}) = E(y_i) = \eta_i(\boldsymbol{\beta})$, desde que $E(\epsilon_i) = 0$.

$$y_i = \eta_i(\boldsymbol{\beta}) + \epsilon_i. \quad (3.6)$$

No caso não-linear, o espaço de estimação, tanto para os valores preditos como para os parâmetros de locação, não é mais um plano, ele é curvo, o que claramente complica a construção de medidas de alavanca. Porém, por analogia com o que existia na época para o caso linear (matriz de projeção ortogonal $\mathbf{H}$), os autores introduziram a idéia de se medir alavanca através do plano tangente ao ponto, $\mathbf{H}$ agora chamada *matriz alavanca do plano tangente*. Para entendermos melhor sobre a construção e o conceito dessa medida, vem o trecho a seguir. A medida plano tangente é construída como se segue: denominando E o espaço de estimação para o modelo não-linear, $E = \{\boldsymbol{\mu}(\boldsymbol{\beta}); \boldsymbol{\beta} \in \mathbb{R}^p\}$, e seja P o plano tangente em E no ponto $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$, temos (suposição) que na vizinhança deste ponto a projeção de $\mathbf{y}$ em P fornece uma boa aproximação da projeção de $\mathbf{y}$ em E . Dessa forma, os autores utilizam a matriz de projeção ortogonal de $\mathbf{y}$ em P como medida de alavanca nas vizinhanças de $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$. E essa matriz, denotada $\hat{\mathbf{H}}$, pode ser escrita na forma abaixo. Quando $\hat{\mathbf{D}}_\beta$ é de posto completo ($\mathbf{D}_\beta = (\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \beta^T})$ e $\hat{\mathbf{D}}_\beta = \mathbf{D}_{\hat{\boldsymbol{\beta}}}$), ocorre que:

$$\hat{\mathbf{H}} = \hat{\mathbf{D}}_\beta (\hat{\mathbf{D}}_\beta^T \hat{\mathbf{D}}_\beta)^{-1} \hat{\mathbf{D}}_\beta^T, \quad (3.7)$$

o que obviamente nos lembra a matriz de alavancas para o caso linear (2.3).

Anteriormente, Belsley, Kuh e Welsch (1980) e Cook e Weisberg (1982), já propunham $\hat{\mathbf{H}}$ como medida de alavanca para o caso não-linear, o que ocorreu no esforço de criarem uma medida capaz de mensurar a influência de uma observação sobre uma regressão não-linear, usaram então a mesma lógica utilizada no caso linear, eles consideraram que na vizinhança de $\boldsymbol{\beta}$ a função

fosse aproximadamente linear, e que, portanto, a estimativa da medida de influência no plano tangente estaria bem próxima do valor do parâmetro real.

Voltando a Saint Laurent e Cook (1992), eles abordam o trabalho de Emerson, Hoaglin e Kempthorne (1984), visto na seção anterior, utilizando suas definições para a matriz de alavancas generalizadas (3.4) e a matriz Jacobiana de alavancas (3.5), e encontram expressões para estas medidas (supondo uma regressão não-linear suficientemente suave), no sentido de poderem comparar as duas medidas com o plano tangente. Alertam que a eficiência do plano tangente, como medida de alavanca, dependerá da aproximação de P ao espaço de estimação E de $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$.

Saint Laurent e Cook (1992) definem a matriz de alavancas generalizadas e a matriz Jacobiana de alavancas, considerando um escalar p como parâmetro de perturbação, adicionado ao m -ésimo elemento de $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, na direção $\mathbf{f}_m$, tal que,

$$\mathbf{y}_{m,p} = \mathbf{y} + p\mathbf{f}_m, \quad (3.8)$$

no qual $\mathbf{f}_m$ é a m -ésima base do $\mathbb{R}^n$, ou seja, um vetor de zeros de tamanho n , no qual a posição m possui o valor 1. O valor predito para o vetor resposta sob a perturbação dada em (3.8) fica então dada por $\hat{\mathbf{y}}_{m,p} = \boldsymbol{\mu}(\hat{\boldsymbol{\beta}}_{fm}(p))$, sendo $\hat{\boldsymbol{\beta}}_{fm}(p)$ a estimativa de $\boldsymbol{\beta}$ do modelo perturbado. A matriz de alavancas generalizadas é definida por,

$$\mathbf{G}(p; \mathbf{f}) = \frac{1}{p}(\hat{\mathbf{y}}_{m,p} - \hat{\mathbf{y}}), \quad (3.9)$$

e a matriz Jacobiana de alavancas é definida por,

$$\hat{\mathbf{J}} = \lim_{p \rightarrow 0} \mathbf{G}(p; \mathbf{f}). \quad (3.10)$$

Interpretam a matriz Jacobiana de alavancas como uma medida da magnitude da derivada de cada valor ajustado em relação ao m -ésimo valor resposta. E a matriz de alavancas generalizadas é a exata perturbação no valor

resposta em relação à perturbação ocorrida. Portanto, para a matriz Jacobiana de alavancas, temos uma medida de alavanca que relaciona-se com a perturbação do m -ésimo valor resposta e para a matriz de alavancas generalizadas, temos uma medida de alavanca que diz respeito ao valor da perturbação. Logo, a matriz Jacobiana de alavancas é mais geral, porém, é em relação à perturbação de um valor resposta.

Encontram então a seguinte expressão para a alavanca generalizada para uma direção arbitrária $\mathbf{f}$, cujo desenvolvimento encontra-se no Apêndice,

$$\mathbf{G}(p; \mathbf{f}) = \widehat{\mathbf{D}}_{\beta} \Delta_f^1 + \frac{1}{2} p \{ \widehat{\mathbf{D}}_{\beta} \Delta_f^2 + (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta\beta} \Delta_f^1 \}, \quad (3.11)$$

em que,

$$\begin{aligned} \Delta_f^k &= \left. \frac{d^k}{dp^k} \widehat{\beta}_f(p) \right|_{p=0}, \text{ supondo que } \widehat{\beta}_f(p) \text{ é no mínimo} \\ &\quad \text{duas vezes diferenciável,} \\ \Delta_f^1 &= \widehat{\mathbf{A}}^{-1} \widehat{\mathbf{D}}_{\beta}^T \mathbf{f}, \\ \Delta_f^2 &= 2 \widehat{\mathbf{A}}^{-1} \left[\widehat{\mathbf{D}}_{\beta\beta}^T \right] \left[\mathbf{f} - \widehat{\mathbf{D}}_{\beta} \Delta_f^1 \right] \Delta_f^1 \\ &\quad - \widehat{\mathbf{A}}^{-1} \left\{ \left[\widehat{\mathbf{D}}_{\beta}^T \right] \left[\widehat{\mathbf{D}}_{\beta\beta} \right] \Delta_f^1 - \left[\widehat{\mathbf{D}}_{\beta\beta\beta}^T \right] [\widehat{\epsilon}] \left[\Delta_f^1 \right] \right\} \Delta_f^1, \\ \widehat{\mathbf{D}}_{\beta\beta} &= \left. \frac{\partial^2 \mu_i(\beta)}{\partial \beta^T \partial \beta} \right|_{\beta=\widehat{\beta}}, \\ \widehat{\mathbf{A}} &= \mathbf{A}(\widehat{\beta}) = \widehat{\mathbf{D}}_{\beta}^T \widehat{\mathbf{D}}_{\beta} - [\widehat{\epsilon}^T] [\widehat{\mathbf{D}}_{\beta\beta}], \\ \widehat{\mathbf{D}}_{\beta\beta\beta} &= \left(\frac{\partial^3 \mu_i(\beta)}{\partial \beta_j \partial \beta_k \partial \beta_m} \right) \Big|_{\beta=\widehat{\beta}}, \\ [.] [.] &\text{ representa multiplicação de arrays com mais de duas dimensões} \\ &\quad \text{(vide, Bates e Watts, 1980).} \end{aligned}$$

No Apêndice apresentamos a prova para a expressão (3.11).

Com a expressão da alavanca generalizada (3.11), podemos encontrar a

expressão para a matriz Jacobiana de alavancas,

$$\hat{\mathbf{J}} = \hat{\mathbf{D}}_{\beta} \{ \hat{\mathbf{D}}_{\beta}^T \hat{\mathbf{D}}_{\beta} - [\hat{\boldsymbol{\epsilon}}^T] [\hat{\mathbf{D}}_{\beta\beta}] \}^{-1} \hat{\mathbf{D}}_{\beta}^T, \quad (3.12)$$

e por conseguinte a expressão da matriz Jacobiana de alavancas na direção $\mathbf{f}$, fica dada por,

$$\mathbf{J}(\mathbf{f}) = \lim_{p \rightarrow 0} \mathbf{G}(p; \mathbf{f}) = \hat{\mathbf{J}}\mathbf{f}. \quad (3.13)$$

É simples de observar que para o caso normal linear ($\boldsymbol{\eta}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$), ocorre que $\mathbf{J}$ apresenta o mesmo resultado de $\mathbf{H}$, logo, $\mathbf{G}(p; \mathbf{f}) = \mathbf{J}\mathbf{f} = \mathbf{H}\mathbf{f}$.

Apesar de em modelos não-lineares a parametrização do modelo não ser única, as matrizes $\mathbf{J}$ e $\mathbf{H}$ são invariantes à reparametrizações do modelo (vide, Saint Laurent e Cook, 1992), visto que as medidas de alavanca lidam com os valores observados e ajustados, não dependendo da parametrização do modelo.

Os autores apontam como o pior problema no uso de $\hat{\mathbf{J}}$, a sua interpretação, visto que $\mathbf{H}$ no caso linear é uma matriz de projeção que já possui suas propriedades conhecidas, o mesmo não ocorre com $\mathbf{J}$. Um dos principais problemas encontrados pelos autores é a ocorrência do que chamam de *superleverage* (traduzimos por superalavanca), a qual ocorre quando a taxa de alteração em $\hat{y}_i$ é maior do que a taxa de alteração em y_i , o que gera $\hat{j}_{ii} > 1$, pois a restrição $\hat{j}_{ii} \geq 0$ é válida, mas a restrição de que $\hat{j}_{ii} \leq 1$ não é válida como é válida para o plano tangente.

Os autores também ressaltam que as alterações nos valores ajustados dependem não somente do valor da perturbação, como também do valor da observação que sofreu perturbação, diferentemente do caso linear em que as alterações ocorrem apenas em função do valor da perturbação na observação.

Após a aplicação de $\mathbf{J}$, verificam que $\mathbf{J}$ pode ser muito diferente de $\mathbf{H}$, e sugerem o uso do plano tangente ($\mathbf{H}$) ao invés de $\mathbf{J}$, devido à facilidade de cálculo e interpretação de $\mathbf{H}$ em relação a $\mathbf{J}$. Quanto à diferença entre as

duas matrizes, relacionam à diferença entre as matrizes que formam o produto interno entre as colunas de $\mathbf{D}_\beta$, $(\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1}$ para $\mathbf{H}$ e $(\mathbf{D}_\beta^T \mathbf{D}_\beta - [\epsilon^T][\mathbf{D}_\beta])^{-1}$ para $\mathbf{J}$, que correspondem à inversa da matriz de informação esperada e à inversa da matriz de informação observada, respectivamente, quando ϵ é um vetor aleatório normal com $\text{Var}(\epsilon) = \sigma^2 \mathbf{I}_n$ e σ^2 é conhecido. Dessa forma, o termo referente a $\mathbf{J}$ leva em conta a curvatura (não-linearidade) intrínseca do modelo, o qual, segundo Battes e Watts (1980), será grande conforme a não-linearidade intrínseca do modelo for grande.

Utilizaram dois modelos, aplicados ao mesmo conjunto de dados, para compararem $\mathbf{J}$ e $\mathbf{H}$, o primeiro modelo foi proposto por Bacon e Watts (1971) (vide, Saint Laurent e Cook, 1992), na forma,

$$\eta_i(\beta) = \beta_1 + \beta_2(x_i - \beta_4) + \beta_3(x_i - \beta_4)\tanh\left(\frac{x_i - \beta_4}{\beta_5}\right),$$

e o segundo foi proposto por Ratkowsky (1983, p. 122) (vide, Saint Laurent e Cook, 1992), na forma,

$$\eta_i(\gamma) = \gamma_1 + \gamma_2(x_i - \gamma_4) + \gamma_3((x_i - \gamma_4)^2 + \gamma_5)^{1/2}.$$

O conjunto de dados utilizado pelos autores, para avaliação desses dois modelos vistos acima, refere-se ao logaritmo da taxa de queda de água em gm/cm-seg num canal inclinado, variável independente, e ao logaritmo do nível da coluna de água em centímetros, variável dependente. A aplicação de $\mathbf{J}$ e $\mathbf{H}$, apresenta valores bastante próximos, exceto para três observações em que, nos dois casos, apresentam grandes diferenças entre as duas medidas, inclusive com sinais diferentes de um modelo para outro.

Há a dificuldade em se criar um critério para o que seja um alto valor para a alavanca Jacobiana, devido à matriz Jacobiana de alavancas não ser uma matriz de projeção. Logo, a definição do que seria um alto valor não pode seguir os mesmos critérios já usados para a matriz de projeção, idempotente, que já possui um critério definido para o que seja um alto valor. Observou-se

Tabela 3.1: Tempo para coagulação e percentual de plasma.

Observação	Tempo de Coagulação	Percentual de Plasma
1	5	118
2	10	58
3	15	42
4	20	35
5	30	27
6	40	25
7	60	21
8	80	19
9	100	18

que os pontos com superalavanca apresentam excessiva influência no ajuste do modelo, porém, ao contrário do que se observa para o caso linear, esses pontos não apresentam resíduo pequeno e, portanto, são pontos mal ajustados pelo modelo.

3.4 Coagulação de sangue com adição de plasma

Realizou-se um experimento com 9 casos (McCullagh e Nelder, 1989), no qual adicionava-se um percentual controlado de plasma a uma quantidade de sangue observando-se o tempo de coagulação. O conjunto das observações pode ser visto na Tabela 3.1 e, para melhor visualização da disposição no espaço, pode-se observar a Figura 3.3.

Ajustamos um modelo com preditor linear, um parâmetro para o intercepto e outro para o percentual de plasma, supondo distribuição gama para a variável dependente (tempo de coagulação_{*i*} $\sim G(\mu_i; \phi)$), com função de ligação inversa (ligação canônica). Tal ajuste gerou os resultados da Tabela

Figura 3.3: Tempo de coagulação contra o percentual de plasma.

Tabela 3.2: Resultados do ajuste do modelo.

Parâmetro	Estimativa	Erro Padrão	Valor de t	$Pr(> t)$
Intercepto	0,009	0,003	3,232	0,014
Percentual de plasma	0,000	0,000	5,466	0,001

3.2, com um desvio de 9,078 com 7 graus de liberdade ($\hat{\phi} = 19,184(8,966)$), o que é um ajuste significativo.

Para os ajustes de MLGs usa-se, comumente, uma aproximação de (2.3), neste caso temos que a função escore e a matriz de informação de Fisher dadas por (3.14) e (3.16) respectivamente. No caso temos que a função de ligação canônica gera $\mathbf{W} = \mathbf{V}$, daí obtemos os termos (3.15) e (3.17),

respectivamente,

$$\frac{\partial \mathbf{L}(\beta; \mathbf{y})}{\partial \beta} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{1/2} (\mathbf{y} - \mu), \quad (3.14)$$

$$\frac{\partial \mathbf{L}(\beta; \mathbf{y})}{\partial \beta} = \phi \mathbf{X}^T \mathbf{V} (\mathbf{y} - \mu), \quad (3.15)$$

$$\mathbf{E} \left\{ - \frac{\partial^2 \mathbf{L}(\beta; \mathbf{y})}{\partial \beta \partial \beta^T} \right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X}, \quad (3.16)$$

$$\mathbf{E} \left\{ - \frac{\partial^2 \mathbf{L}(\beta; \mathbf{y})}{\partial \beta \partial \beta^T} \right\} = \phi \mathbf{X}^T \mathbf{V} \mathbf{X}, \quad (3.17)$$

em que,

$$\mu : \mathbf{E}(\mathbf{y}),$$

$$\phi : \text{parâmetro de dispersão},$$

$$\mathbf{V} = \text{diag}\{v_1, \dots, v_n\} : \text{função de variância},$$

$$\mathbf{W} = \text{diag}\{w_1, \dots, w_n\} : \text{matriz de pesos, } w_i = \frac{(d\mu/d\eta)_i^2}{V_i}.$$

A matriz de alavancas generalizadas para modelos lineares generalizados será calculada no próximo capítulo, a qual coincide com a matriz $\widehat{\mathbf{H}}^*$, definida abaixo, apenas para ligação canônica, usualmente usada nesta classe

$$\widehat{\mathbf{H}}^* = \widehat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \widehat{\mathbf{W}}^{1/2}. \quad (3.18)$$

Similarmente, faz-se necessário a utilização de outro tipo de resíduo, pois, os resíduos studentizados, criados com base na definição para a regressão normal linear, não possuem as mesmas propriedades para o caso dos modelos lineares generalizados (MLGs) (ver, Paula, 2003). Logo, as medidas utilizadas até agora para identificar pontos aberrantes e pontos influentes, não valem mais. Com isso, para o caso dos MLGs, utilizamos a estatística mais comumente usada para identificação de pontos aberrantes, uma estatística baseada em um resíduo derivado da função desvio, cuja expressão padronizada (vide, Paula, 2003, p. 59) é a dada abaixo, com sinal de $y_i - \hat{\mu}_i$.

Através de simulações Williams (vide, Paula, 2003, p. 59), observou-se que a distribuição de t_{Di} é a mais próxima da normalidade, entre os resíduos que foram propostos na literatura

$$t_{Di} = \frac{d^*(y_i; \hat{\mu}_i)}{(1 - \hat{h}_{ii})^{1/2}} = \frac{\phi^{1/2} d(y_i; \hat{\mu}_i)}{(1 - \hat{h}_{ii})^{1/2}}, \quad (3.19)$$

em que, $d(y_i; \hat{\mu}_i) = \pm \sqrt{2} \{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\hat{\theta}_i^0))\}^{1/2}$ e θ é o parâmetro canônico.

Para identificação de pontos influentes, no caso de MLGs, usamos uma outra estatística (LD_i), chamada de distância de Cook aproximada, uma aproximação do afastamento da verossimilhança (*likelihood displacement*) (vide, Cook e Weisberg, 1982), na forma a seguir

$$LD_i \cong \left(\frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right) t_{Si}^2 \quad (3.20)$$

em que,

$$t_{Si} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\hat{V}_i^{1/2}(1 - \hat{h}_{ii})^{1/2}}.$$

Voltando à análise do modelo, a Figura 3.4 mostra uma disposição comportada dos resíduos(t_{Di}) em torno de zero, porém, temos que levar em conta que o pequeno número de observações dificulta a análise. E, segundo os critérios já adotados, podemos considerar a observação 1 como aberrante. A observação 1 apresenta-se como um ponto aberrante, com valor 118, enquanto o terceiro quartil é de 42, para a variável tempo de coagulação.

Devemos ressaltar que para os elementos da diagonal da matriz (3.18) e das matrizes (3.7) e (3.12), os critérios $2p/n$ e $3p/n$ já não podem ser usados como foram para a matriz de projeção (2.3), contudo, podemos pensar em usar $2\text{traço}/n$ e $3\text{traço}/n$. Para esse ajuste, os elementos da diagonal principal da matriz (3.18) podem ser vistos na Figura 3.5, e podemos identificar

Figura 3.4: Resíduo t_{Di} contra os valores ajustados. Figura 3.5: Matriz de alavancas $\hat{H}^*$.

Figura 3.6: Gráfico de envelope.

que a observação 1 apresenta-se como um ponto de alavanca no ajuste, visto que destaca-se em relação aos demais.

O gráfico de envelope (Figura 3.6) apresenta uma observação fora das bandas de confiança e duas outras observações no limite, o que indica que há problemas com as suposições do modelo.

O ajuste de um modelo inserindo a variável porcentagem de plasma ao quadrado, devido ao que se vê na Figura 3.4, também não se mostra um bom ajuste.

O ajuste de um modelo com um número pequeno de observações pode ser bastante afetado quando uma dessas observações é de alavanca, dessa forma, procuramos ajustar um modelo sem a observação identificada como de alavanca. O novo ajuste apresenta um desvio de 8,021 com 6 graus de liberdade ($\hat{\phi} = 63,349(31,592)$), e com as estimativas da Tabela 3.3. A alteração no intercepto foi de 85,8%, no parâmetro do percentual de plasma

Tabela 3.3: Resultados do ajuste do modelo sem a observação 1.

Parâmetro	Estimativa	Erro Padrão	Valor de t	$Pr(> t)$
Intercepto	0,016	0,002	6,364	0,000
Percentual de plasma	0,000	0,000	6,492	0,000

Figura 3.7: Resíduo t_{Di} contra os valores ajustados - sem a observação 1.

foi de 2,6% e, em ϕ o percentual de mudança foi de 230,2%, um resultado bastante significativo, comprovando o peso da observação que foi retirada, na estimativa dos parâmetros.

A Figura 3.7 apresenta a mesma disposição dos resíduos do primeiro ajuste, contudo, agora a nova observação número 1 (10;58), torna-se aberrante, o que mostra que os valores iniciais para esse conjunto de dados têm grande peso no ajuste. Quanto à identificação de pontos influentes a Figura 3.8 apresenta os valores da estatística (3.20) e mostra a observação 1 como um candidato forte a ser influente, mas não conclusivo, com um valor de 3,84.

Figura 3.8: Distância de Cook aproximada LD_i - sem a observação 1.

A disposição gráfica das medidas $\hat{H}^*$, plano tangente e alavanca Jacobiana, Figuras 3.9, 3.10 e 3.11, respectivamente, apresentam diferenças quase imperceptíveis, porém, existentes como se pode observar na tabela 3.4 que apresenta os valores com várias casas decimais para evidenciar as diferenças, nelas a nova observação 1 agora também destaca-se em relação às demais.

O gráfico de envelope (Figura 3.12) já não apresenta qualquer observação fora das bandas de confiança, o que indica que as suposições do modelo estão sendo respeitadas, principalmente a suposição de erros com distribuição gama.

Podemos observar que neste caso, em que temos um experimento com observações com dosagens decrescentes, que os valores iniciais possuem grande peso no ajuste, visto que sua retirada altera bastante o ajuste além de aparecerem como fortes candidatos a serem pontos de alavanca e influentes.

Figura 3.9: Matriz de alavancas $\hat{H}^*$ - sem a observação 1.

Figura 3.10: Plano tangente - sem a observação 1.

Figura 3.11: Alavanca Jacobiana - sem a observação 1.

Tabela 3.4: Valores para o plano tangente e a alavanca Jacobiana.

Observação	$\hat{H}^*$	Plano Tangente	Alavanca Jacobiana
1	0,42240	0,42234	0,42296
2	0,27424	0,27424	0,27446
3	0,19117	0,19118	0,19122
4	0,12865	0,12866	0,12865
5	0,13181	0,13180	0,13195
6	0,20132	0,20131	0,20191
7	0,28620	0,28621	0,28725
8	0,36421	0,36426	0,36568

Figura 3.12: Gráfico de envelope - sem a observação 1.

3.5 Relação peso dos olhos e idade de coelhos selvagens

Analizamos outro conjunto de dados para ampliar a aplicação das medidas. Um experimento analisado por Dudzinski e Mykytowycz (ver Ratkowsky, 1983, p. 105) observou o peso dos olhos (y , em mg) de 71 coelhos selvagens europeus na Austrália e a idade (x , em dias) deles, com a intenção de identificar a relação entre o peso dos olhos e a idade dos coelhos, a relação proposta entre as duas informações foi expressa na forma a seguir, que ajusta um modelo para o logaritmo de y com preditor não-linear,

$$\ln(y) = \alpha - \frac{\beta}{x + \gamma} + \epsilon.$$

Pode-se observar a disposição dessas informações no espaço na Figura 3.13, que apresenta uma curva que assemelha-se ao formato de uma função lo-

Figura 3.13: Peso versus idade dos coelhos.

garítmica, no qual a idade seria o expoente e o peso seria o logaritmando. O box plot da variável peso dos olhos, Figura 3.15, apresenta um comportamento bastante simétrico e sem valores atípicos, contudo, a aplicação da função $\ln$ ao peso dos olhos faz com que a mesma, Figura 3.16, apresente assimetria à esquerda e valores atípicos, contudo, comparando-se as Figuras (3.13) e (3.14), observa-se que a aplicação do logaritmo conseguiu diminuir a variância. Já a variável idade do coelho, Figura 3.17, apresenta 5 valores atípicos e um comportamento assimétrico à direita (maior idade), ou seja, há uma concentração de coelhos jovens no experimento.

O modelo ajustado supõe uma distribuição normal para a variável logaritmo do peso dos olhos, com função de ligação identidade, o método de estimação utilizado foi o de mínimos quadrados. O ajuste do modelo gera os resultados da Tabela 3.5, os quais mostram que as estimativas dos parâmetros são significativas, apresentando estimativas bastante precisas, apresentam

Figura 3.14: Ln(peso) versus idade dos coelhos.

alto valor para a estatística t , ou, teste de Wald.

A Figura 3.18 mostra as 4 primeiras observações com valor alto para o plano tangente, as 3 primeiras com mesmo valor 0,23 e a quarta um pouco menor 0,16, visto que também possuem mesmo valor de idade. Essas 4 observações são referentes aos 4 coelhos com menor idade e estão na região do modelo em que a curvatura intrínseca é maior.

A alavanca Jacobiana, Figura 3.19, como no exemplo anterior de Coa-

Tabela 3.5: Resultados do ajuste do modelo.

Parâmetro	Estimativa	Erro Padrão	Valor de t	$Pr(> t)$
α (intercepto)	5,640	0,020	282,461	0,000
β	130,584	5,725	22,809	0,000
γ	37,603	2,323	16,187	0,000

Figura 3.15: Peso dos olhos dos coelhos.

Figura 3.16: Logaritmo do peso dos olhos dos coelhos.

Figura 3.17: Idade dos coelhos.

Figura 3.18: Plano tangente.

Figura 3.19: Alavanca Jacobiana.

gulação de Sangue com Adição de Plasma, apresenta um comportamento muito parecido ao do plano tangente, o que indica pouca influência do termo $-\hat{\epsilon}[\hat{\mathbf{D}}_{\beta\beta}]$, que representa a curvatura intrínseca do modelo, também apresenta as 4 primeiras observações com valor alto para o plano tangente, já a matriz de alavanca Jacobiana não é idempotente, contudo, dada a semelhança de valores, adotamos os mesmos critérios de avaliação. Novamente, essas 4 observações são referentes aos 4 coelhos com menor idade.

Neste exemplo podemos reforçar o peso dos valores iniciais no ajuste para os conjuntos de dados de curva de crescimento (formato logarítmico), visto que aparecem como pontos de alavanca, além da semelhança de comportamento das medidas plano tangente e alavanca Jacobiana.

Capítulo 4

Modelos não-lineares de família exponencial

4.1 Introdução

Neste capítulo discutimos a generalização de Wei, Hu e Fung (1998), para mensurar alavanca, a qual denominam alavanca generalizada (*generalized leverage*), e procuram ampliar a utilização do que é denominado matriz Jacobiana de alavancas, conforme o artigo de Saint Laurent e Cook (1992).

Wei, Hu e Fung (1998) utilizam a matriz de alavancas generalizadas como medida da “importância individual das observações”, expandindo, principalmente, para os modelos que utilizem outros métodos de estimação além de mínimos quadrados (máxima verossimilhança, estimadores Bayesianos, estimadores M etc), que possuem dados censurados ou dados correlacionados etc. Também se concentram na aplicação da matriz de alavancas generalizadas em modelos não-lineares de família exponencial, e incluem modelos não-lineares e modelos lineares generalizados como casos especiais. Portanto, apresentamos neste capítulo o que há de mais relevante na literatura em relação às medidas conhecidas para se mensurar alavanca seja no caso linear ou no caso

não-linear.

A generalização feita no artigo se baseia na definição de que $h_{ij} = \frac{\partial \hat{y}_i}{\partial y_j}$, ou seja, que a alavanca generalizada representa a taxa instantânea de alteração no i -ésimo valor predito em relação ao j -ésimo valor observado. Citam os principais textos no assunto (Hoaglin e Welsch, 1978; Cook e Weisberg, 1982; Emerson, Hoaglin e Kempthorne, 1984; St. Laurent e Cook, 1992, dentre outros), argumentando que eles elegem $h_{ii} = \frac{\partial \hat{y}_i}{\partial y_i}$ como a medida que mais diretamente reflete a influência de y_i no seu próprio valor ajustado. Dessa maneira os autores definem a alavanca generalizada na forma:

$$\mathbf{GL}(\hat{\delta}) = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^T} = \left(\frac{\partial \hat{y}_i}{\partial y_j} \right)_{n \times n}, \quad (4.1)$$

o que explicita $\mathbf{GL}_{ij}$ como a taxa instantânea de alteração no valor predito i em relação ao valor resposta j . Também sustentam que sua definição consegue ser de muito mais fácil aplicação para estimadores e modelos muito mais gerais, do que apenas mínimos quadrados em modelos lineares e não-lineares de regressão, como já tinha sido feito por alguns dos autores citados anteriormente. Analisamos essas afirmações no decorrer do capítulo.

4.2 Alavanca generalizada

Considerando que $\mathbf{y} = (y_1, \dots, y_n)^T$ seja um vetor de variáveis aleatórias com função densidade de probabilidade $p(\mathbf{y}; \delta)$, $E(\mathbf{y}) = \boldsymbol{\mu} = \boldsymbol{\mu}(\delta)$, $\hat{\delta} = \hat{\delta}(\mathbf{y})$ é um estimador qualquer de δ , e o vetor resposta predito fica expresso como $\hat{\mathbf{y}} = \boldsymbol{\mu}(\hat{\delta}(\mathbf{y}))$. Desse modo, podemos reescrever (4.1) como:

$$\mathbf{GL}(\hat{\delta}) = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^T} = \frac{\partial \hat{\boldsymbol{\mu}}}{\partial \hat{\delta}^T} \frac{\partial \hat{\delta}(\mathbf{y})}{\partial \mathbf{y}^T}. \quad (4.2)$$

A partir da expressão (4.2) Wei, Hu e Fung (1998) deduzem uma forma explícita para $\mathbf{GL}(\hat{\delta})$ com base no trabalho de Gay e Welsch (1988) utilizando

a função:

$$Q(\boldsymbol{\delta}; \mathbf{y}) = \sum_{i=1}^n \rho_i(f(\mathbf{x}_i, \boldsymbol{\beta}); \boldsymbol{\gamma}), \quad (4.3)$$

em que $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, $E(y_i) = \mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$ e $\boldsymbol{\gamma}$ denota um parâmetro de perturbação. Por exemplo, no caso de estimação via mínimos quadrados, supondo que $\boldsymbol{\gamma} = \mathbf{0}$, temos que

$$\rho_i(f(\mathbf{x}_i, \boldsymbol{\beta})) = \{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})\}^2.$$

No trabalho de Gay e Welsch (1988) $\boldsymbol{\gamma}$ é um parâmetro de perturbação apenas, ou seja, não entra no modelo, e no nosso caso muito dificilmente será usado.

Através da utilização de (4.3), supondo que é minimizada por $\widehat{\boldsymbol{\delta}}(\mathbf{y})$, que possui derivada de segunda ordem com relação a $\boldsymbol{\delta}$ e $\mathbf{y}$, e que $\widehat{\boldsymbol{\delta}}(\mathbf{y})$ existe e é único, obtém-se a seguinte forma fechada para $\mathbf{GL}(\widehat{\boldsymbol{\delta}})$:

$$\mathbf{GL}(\widehat{\boldsymbol{\delta}}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\delta}^T} \left(- \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right)^{-1} \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \mathbf{y}^T} \right\} \Big|_{\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}(\mathbf{y})}. \quad (4.4)$$

O desenvolvimento de (4.4) vem a seguir. Primeiramente deriva-se (4.3) em relação a $\boldsymbol{\delta}$, no ponto $\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}(\mathbf{y})$, minimizando a função, portanto, como segue,

$$\frac{\partial Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta}} \Big|_{\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}(\mathbf{y})} = \mathbf{0}, \quad (4.5)$$

diferenciando novamente, em ambos os lados da equação, em relação a $\mathbf{y}$, no ponto $\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}(\mathbf{y})$, encontra-se,

$$\left\{ \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \mathbf{y}^T} + \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \frac{\partial \boldsymbol{\delta}(\mathbf{y})}{\partial \mathbf{y}^T} \right\} \Big|_{\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}(\mathbf{y})} = \mathbf{0},$$

em que o primeiro componente da expressão acima refere-se à parte de (4.5) que não contém $\boldsymbol{\delta}$, e a segunda componente refere-se à parte de (4.5) que contém $\boldsymbol{\delta}$.

Conseqüentemente,

$$\frac{\partial \hat{\delta}(\mathbf{y})}{\partial \mathbf{y}^T} = \left[-\frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right]^{-1} \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \mathbf{y}^T} \Big|_{\boldsymbol{\delta}=\hat{\delta}(\mathbf{y})}, \quad (4.6)$$

e substituindo (4.6) em (4.2) obtém-se (4.4).

A alavanca generalizada no formato (4.4), traz a grande contribuição de dar liberdade ao método de estimação utilizado no modelo, como apregoaram os autores, justamente devido à liberdade de uso da função (4.3). Para o caso de estimadores de máxima verossimilhança, substituímos $\mathbf{Q}(\boldsymbol{\delta}; \mathbf{y})$ por $\mathbf{L}(\boldsymbol{\delta}; \mathbf{y})$ (o logaritmo da função de verossimilhança de $\boldsymbol{\delta}$), mantendo as suposições de que possua derivada de segunda ordem com relação a $\boldsymbol{\delta}$ e $\mathbf{y}$, que $\hat{\delta}(\mathbf{y})$ maximize $\mathbf{L}(\boldsymbol{\delta}; \mathbf{y})$, e que seja única, dessa forma (4.4) transforma-se em

$$\mathbf{GL}(\hat{\delta}) = \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\delta}^T} \left(-\frac{\partial^2 \mathbf{L}(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right)^{-1} \frac{\partial^2 \mathbf{L}(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \mathbf{y}^T} \right\} \Big|_{\boldsymbol{\delta}=\hat{\delta}(\mathbf{y})}. \quad (4.7)$$

Podemos observar que nesse formato $\mathbf{GL}(\hat{\delta})$ possui relação com a matriz de informação de Fisher observada.

Wei, Hu e Fung (1998) adotam as notações, $\mathbf{D}_{\boldsymbol{\delta}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\delta}^T}$, $\dot{\mathbf{Q}}_{\boldsymbol{\delta}} = \frac{\partial Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta}}$ e $\ddot{\mathbf{Q}}_{\boldsymbol{\delta}\boldsymbol{\delta}} = \frac{\partial^2 Q(\boldsymbol{\delta}; \mathbf{y})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T}$. Observa-se nas notações acima que ponto(s) em cima significa o número de derivações, e em baixo aparece por qual(is) parâmetro(s) foi(ram) feita(s) a(s) derivação(ões), respectivamente. Por conveniência adotamos tal notação. Reescrevendo $\mathbf{GL}(\hat{\delta})$, com essa nova notação, temos que

$$\mathbf{GL}(\hat{\delta}) = \{ \mathbf{D}_{\boldsymbol{\delta}} (-\ddot{\mathbf{L}}_{\boldsymbol{\delta}\boldsymbol{\delta}})^{-1} \ddot{\mathbf{L}}_{\boldsymbol{\delta}\mathbf{y}} \} \Big|_{\boldsymbol{\delta}=\hat{\delta}(\mathbf{y})}. \quad (4.8)$$

4.3 Aplicações

4.3.1 Modelos não-lineares de família exponencial

Nesta seção, aplicamos o conceito de Wei, Hu e Fung (1998) de alavanca generalizada para o caso das distribuições que pertencem à classe de distribuições de probabilidade conhecida, na literatura científica, como família

exponencial. Descrevemos a seguir as características dessa família de distribuições.

Suponha que $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ sejam variáveis aleatórias independentes, tais que a função densidade de y_i , $\forall i$, assume a forma

$$p(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i; \phi)], \quad i = 1, \dots, n, \quad (4.9)$$

em que θ é denominado parâmetro canônico, $E(y_i) = \mu_i = b'(\theta_i)$, $\text{Var}(y_i) = V_i\phi^{-1}$, com $V_i = \frac{\partial \mu_i}{\partial \theta_i}$ denominada função de variância, e ϕ^{-1} é o parâmetro de dispersão (ou de escala), com $\phi > 0$, $c(y_i; \phi)$ é uma função de ϕ . A função de variância é de grande importância, visto que caracteriza a classe da distribuição da variável na família exponencial.

Os modelos lineares e não-lineares generalizados da família exponencial (vide, por exemplo, Cordeiro e Paula, 1989a; Wei, Hu e Fung, 1998) são definidos por (4.9) e pela componente sistemática,

$$g(\mu_i) = \eta_i(\boldsymbol{\beta}), \quad (4.10)$$

em que $g(\cdot)$ é denominada função de ligação, podendo ser qualquer função monótona e diferenciável, e que $\eta_i(\boldsymbol{\beta}) = \eta_i(\boldsymbol{\beta}; \mathbf{x}_i)$ é uma função, que pode ser linear ou não-linear em $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, dados os valores de $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})^T$, q pode ser diferente de p . Quando $\eta_i(\boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta}$ tem-se um modelo linear generalizado, porém, quando $\eta_i(\boldsymbol{\beta})$ é uma função não-linear em $\boldsymbol{\beta}$ tem-se um modelo não-linear de família exponencial.

Os trabalhos de McCullagh e Nelder (1989) e Cordeiro e Paula (1989a), contêm uma discussão de como encontrar a estimativa de máxima verossimilhança para o vetor de parâmetros $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^T$ em modelos não-lineares de família exponencial.

Como é conhecido (vide, por exemplo, Paula, 2003), as distribuições normal, gama, normal inversa, binomial, Poisson e binomial negativa, dentre outras, pertencem à família exponencial.

Definindo $\mathbf{J}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}$, o processo iterativo *scoring* de Fisher (vide, por exemplo, Paula, 1992), para obter a estimativa de máxima verossimilhança para $\boldsymbol{\beta}$ em um modelo não-linear de família exponencial é dado por

$$\boldsymbol{\beta}^{m+1} = (\mathbf{J}^{(m)T} \mathbf{W}^{(m)} \mathbf{J}^{(m)})^{-1} \mathbf{J}^{(m)T} \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (m = 0, 1, \dots), \quad (4.11)$$

sendo $\mathbf{z}^{(m)}$ uma variável dependente modificada dada por

$$\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \boldsymbol{\tau}^{(m)} + \mathbf{W}^{-1/2(m)} \mathbf{V}^{-1/2(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}),$$

no qual $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$, tais que $w_i = (\frac{d\mu_i}{d\eta_i})^2 / V_i$, $u_i = \frac{d\eta_i}{d\mu_i}$ e $\boldsymbol{\tau} = f(\mathbf{x}; \boldsymbol{\beta}) - \mathbf{J}\boldsymbol{\beta}$ com $f(\mathbf{x}; \boldsymbol{\beta}) = (f(\mathbf{x}_1; \boldsymbol{\beta}), \dots, f(\mathbf{x}_n; \boldsymbol{\beta}))^T$. O processo iterativo (4.11) pode ser ajustado por meio de algoritmos desenvolvidos para modelos lineares generalizados tratando-se a quantidade $\boldsymbol{\tau}$ como sendo um *offset* (vide, por exemplo, Paula e Cordeiro, 1986; Cordeiro e Paula, 1989b). Vê-se, portanto, que trata-se de um método iterativo de mínimos quadrados ponderados.

Quando $\boldsymbol{\eta}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$, tem-se um modelo linear generalizado e conseqüentemente tem-se $\boldsymbol{\tau} = 0$, ficando o processo iterativo (4.11) expresso na forma simplificada

$$\boldsymbol{\beta}^{m+1} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (m = 0, 1, \dots),$$

no qual

$$\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{U}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}), \quad \text{em que} \quad \mathbf{U} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}^T}.$$

Sob condições de regularidade (vide, por exemplo, Wei, 1998) tem-se a consistência de $\hat{\boldsymbol{\beta}}$ na classe dos modelos não-lineares de família exponencial, e a normalidade assintótica. Isto é, para n grande,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N_p(0; \phi^{-1} \mathbf{K}^{-1}(\boldsymbol{\beta})),$$

sendo $\mathbf{K}(\boldsymbol{\beta}) = (\mathbf{J}^T \mathbf{W} \mathbf{J})$, e ainda

$$\hat{\phi} - \phi \sim N(0; \sigma_\phi^2),$$

sendo $\sigma_\phi^2 = \left\{ -\sum_{i=1}^n c''(y_i; \phi) \right\}^{-1}$. Em particular se $y_i \sim G(\mu_i; \phi)$, obtém-se

$$\sigma_\phi^2 = -\frac{1}{\sum_{i=1}^n c''(y_i; \phi)} = \frac{\phi}{n\phi\psi'(\phi) - 1},$$

sendo $\psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$ denominada de função digama e $\Gamma(\phi)$ é a função gama. Lembrando que para o caso gama temos: $c(y; \phi) = (\phi - 1)\ln y + \phi \ln \phi - \ln \Gamma(\phi)$.

A matriz de alavancas generalizadas pode ser obtida sempre que y_i for uma variável contínua, para satisfazer a condição de $\frac{\partial \hat{y}_i}{\partial y_i}$ ser diferenciável.

4.3.2 Alavanca para modelos da família exponencial

Agora aplicamos $\mathbf{GL}(\hat{\boldsymbol{\delta}})$ para distribuições pertencentes à família exponencial. Dessa forma, identificamos os termos $\mathbf{D}_\delta$, $\ddot{\mathbf{L}}_{\delta\delta}$ e $\ddot{\mathbf{L}}_{\delta y}$ da equação (4.8), conforme as especificações da família exponencial. A ligação canônica entre o parâmetro canônico e o preditor ($\theta = \eta$), tem a propriedade de garantir a concavidade da verossimilhança, favorecendo a obtenção de resultados assintóticos e, garantir a unicidade da estimativa de máxima verossimilhança.

Os modelos lineares generalizados (vide, por exemplo, McCullagh e Nelder, 1989) são um caso particular em que $\eta_i(\boldsymbol{\beta}; \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ (aqui $q = p$) e a função de ligação é livre dentro das condições já vistas (diferenciável e monótona).

Seja $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \phi)^T$ e, portanto, $\ddot{\mathbf{L}}_{\delta\delta}$ é uma matriz na forma

$$\ddot{\mathbf{L}}_{\delta\delta} = \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\phi} \\ \ddot{\mathbf{L}}_{\phi\beta} & \ddot{\mathbf{L}}_{\phi\phi} \end{pmatrix}.$$

Através do logaritmo da versão matricial de (4.9), $\mathbf{L}(\boldsymbol{\theta})$, obtemos as submatrizes $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\phi}$, $\ddot{\mathbf{L}}_{\phi\beta}$ e $\ddot{\mathbf{L}}_{\phi\phi}$. Podemos encontrar esses valores de duas

formas, apresentamos aqui a mais simples e no Apêndice outra mais complexa em que obtemos quatro termos na expressão de $\ddot{\mathbf{L}}_{\beta\phi}$. Primeiramente, vamos à busca de $\ddot{\mathbf{L}}_{\beta\beta}$, derivando $\mathbf{L}(\boldsymbol{\theta})$ em relação à $\boldsymbol{\beta}$,

$$\begin{aligned}\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^T} &= \phi \left\{ \mathbf{y}^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T} - \left(\frac{\partial \mathbf{b}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T} \right\} = \dot{\mathbf{L}}_{\beta} \\ \dot{\mathbf{L}}_{\beta} &= \phi \left\{ \mathbf{y}^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \mathbf{R} - \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \mathbf{R} \right\} \\ &= \phi \left\{ (\mathbf{y} - \boldsymbol{\mu})^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \mathbf{R} \right\} \\ &= \phi \left\{ \boldsymbol{\epsilon}^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}^T} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} \mathbf{R} \right\} \\ &= \phi \left\{ \boldsymbol{\epsilon}^T \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} \mathbf{R} \right\} \\ &= \phi \left\{ \boldsymbol{\epsilon}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} \right\},\end{aligned}$$

em que

$$\mathbf{D}_{\beta} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T} = \mathbf{U}^{-1} \mathbf{R} = \mathbf{V}^{\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{R}, \quad \text{com} \quad \mathbf{R} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T}.$$

Em seguida derivamos $\dot{\mathbf{L}}_{\beta}$ em relação à $\boldsymbol{\beta}$ novamente, obtendo $\ddot{\mathbf{L}}_{\beta\beta}$

$$\begin{aligned}\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \boldsymbol{\beta}} = \ddot{\mathbf{L}}_{\beta\beta} &= \phi \left\{ -\mathbf{D}_{\beta}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} + \mathbf{R}^T \left[\boldsymbol{\epsilon}^T \right] \left[\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T \partial \boldsymbol{\eta}} \right] \mathbf{R} + \right. \\ &\quad \left. \boldsymbol{\epsilon}^T \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \right] \left[\frac{\partial^2 \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right] \right\}\end{aligned}$$

em que $[.][.]$ denota multiplicação de matrizes (vide, Bates e Watts, 1980).

Os cálculos para encontrar $\ddot{\mathbf{L}}_{\phi\phi}$ são

$$\begin{aligned}\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \phi} &= \left\{ \mathbf{y}^T \boldsymbol{\theta} - \mathbf{b}(\boldsymbol{\theta}) \right\} + \dot{\mathbf{c}}(\mathbf{y}; \phi) = \dot{\mathbf{L}}_{\phi}, \\ \frac{\partial \dot{\mathbf{L}}_{\phi}}{\partial \phi} &= \ddot{\mathbf{c}}(\mathbf{y}; \phi) = \ddot{\mathbf{L}}_{\phi\phi}.\end{aligned}$$

Para completarmos a matriz $\ddot{\mathbf{L}}_{\delta\delta}$, encontramos as expressões para $\ddot{\mathbf{L}}_{\beta\phi}$ e $\ddot{\mathbf{L}}_{\phi\beta}$. Dado que

$$\dot{\mathbf{L}}_{\beta} = \phi \left\{ \boldsymbol{\epsilon}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} \right\}, \quad (4.12)$$

temos

$$\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \phi} = \left\{ \boldsymbol{\epsilon}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} \right\} = \ddot{\mathbf{L}}_{\beta\phi} = (\ddot{\mathbf{L}}_{\phi\beta})^T.$$

Porém, como para a estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$, temos $\dot{\mathbf{L}}_{\beta} \big|_{\beta=\hat{\boldsymbol{\beta}}} = 0$, implica que $\ddot{\mathbf{L}}_{\beta\phi} = (\ddot{\mathbf{L}}_{\phi\beta})^T = 0$, visto que $\dot{\mathbf{L}}_{\beta} = \phi \ddot{\mathbf{L}}_{\beta\phi}$, dessa forma, a matriz $\ddot{\mathbf{L}}_{\delta\delta} \big|_{\delta=\hat{\boldsymbol{\delta}}}$ será bloco diagonal, ou seja,

$$\ddot{\mathbf{L}}_{\delta\delta} = \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{\phi\phi} \end{pmatrix}.$$

Por fim, para completarmos a expressão de $\mathbf{GL}(\hat{\boldsymbol{\delta}})$, falta apenas $\ddot{\mathbf{L}}_{\beta y}$, dado (4.12) temos

$$\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \mathbf{y}} = \phi \mathbf{D}_{\beta}^T \mathbf{V}^{-1} = \ddot{\mathbf{L}}_{\beta y}. \quad (4.13)$$

Dado que $\mathbf{D}_{\delta} = (\mathbf{D}_{\beta}; \mathbf{0})$, logo temos que,

$$\mathbf{GL}(\hat{\boldsymbol{\delta}}) = \mathbf{GL}(\hat{\boldsymbol{\beta}}) = \left\{ \mathbf{D}_{\beta} (-\ddot{\mathbf{L}}_{\beta\beta})^{-1} (\ddot{\mathbf{L}}_{\beta y}) \right\} \bigg|_{\beta=\hat{\boldsymbol{\beta}}}. \quad (4.14)$$

Aplicando as expressões (4.12) e (4.13), à matriz Jacobiana de alavancas (4.14) fica expressa na forma,

$$\begin{aligned} \mathbf{GL}(\hat{\boldsymbol{\beta}}) &= \mathbf{D}_{\beta} \left\{ \mathbf{D}_{\beta}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} - \mathbf{R}^T \left[\boldsymbol{\epsilon}^T \right] \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T \partial \boldsymbol{\eta}} \right] \mathbf{R} \right. \\ &\quad \left. - \boldsymbol{\epsilon}^T \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T} \right] \left[\frac{\partial^2 \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right] \right\}^{-1} \mathbf{D}_{\beta}^T \mathbf{V}^{-1} \bigg|_{\beta=\hat{\boldsymbol{\beta}}}. \end{aligned} \quad (4.15)$$

Podemos dividir os modelos de regressão de família exponencial em quatro categorias, duas em relação à linearidade (linear ou não-linear) e outras duas em relação à ligação do preditor com o parâmetro canônico (ligação canônica ou ligação geral). Obtemos quatro categorias para (4.15):

(4.16) - Modelos não-lineares de família exponencial (MNFE) com ligação geral;

(4.16) - Modelos não-lineares de família exponencial (MNFE) com ligação canônica;

(4.17) - Modelos lineares de família exponencial (MLGs) com ligação geral;

(4.18) - Modelos lineares de família exponencial (MLGs) com ligação canônica.

Cada caso gera uma configuração para $\mathbf{GL}(\hat{\beta})$, como podemos observar a seguir

MNFE - ligação geral

$$\mathbf{GL}(\hat{\beta}) = \left\{ \mathbf{D}_{\beta} \left\{ \mathbf{D}_{\beta}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} - \mathbf{R}^T \left[\epsilon^T \right] \left[\frac{\partial \theta}{\partial \eta^T \partial \eta} \right] \mathbf{R} - \epsilon^T \left[\frac{\partial \theta}{\partial \eta^T} \right] \left[\frac{\partial^2 \eta}{\partial \beta^T \partial \beta} \right] \right\}^{-1} \mathbf{D}_{\beta}^T \mathbf{V}^{-1} \right\} \Big|_{\beta = \hat{\beta}(y)},$$

MNFE - ligação canônica

$$\mathbf{GL}(\hat{\beta}) = \left\{ \mathbf{VR} \left(\mathbf{R}^T \mathbf{VR} - \left[\epsilon^T \right] \left[\frac{\partial^2 \eta}{\partial \beta^T \partial \beta} \right] \right)^{-1} \mathbf{R}^T \right\} \Big|_{\beta = \hat{\beta}(y)}, \quad (4.16)$$

MLGs - ligação geral

$$\mathbf{GL}(\hat{\beta}) = \left\{ \mathbf{V}^{1/2} \mathbf{W}^{1/2} \mathbf{X} \left(\mathbf{X}^T \mathbf{W} \mathbf{X} - \mathbf{X}^T \left[\boldsymbol{\epsilon}^T \right] \left[\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T \partial \boldsymbol{\eta}} \right] \mathbf{X} \right)^{-1} \mathbf{X}^T \right\} \Big|_{\beta = \hat{\beta}(y)} \quad (4.17)$$

MLGs - ligação canônica

$$\mathbf{GL}(\hat{\beta}) = \left\{ \mathbf{V} \mathbf{X} \left(\mathbf{X}^T \mathbf{V} \mathbf{X} \right)^{-1} \mathbf{X}^T \right\} \Big|_{\beta = \hat{\beta}(y)} \quad (4.18)$$

Encontrada a forma fechada para cada caso apresentado, o próximo passo é analisar a matriz resultante para que se consiga, para cada caso, determinar quais são os novos limitantes, se existirem, para a diagonal principal dessas matrizes. Obtendo-se matrizes com estruturas cujas propriedades sejam conhecidas como já visto, por exemplo, as matrizes idempotentes, esse trabalho fica muito mais fácil, caso contrário, acredito que será necessário se analisar caso a caso. Realizado esse trabalho, a alavanca generalizada proposta por Wei, Hu e Fung torna-se uma ferramenta poderosíssima para se identificar pontos de alavancas em modelos de regressão.

Em MLGs, o comumente usado para identificação de pontos de alavanca são os elementos da diagonal principal de

$$\widehat{\mathbf{W}}^{1/2} \mathbf{X} \left(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X} \right)^{-1} \mathbf{X}^T \widehat{\mathbf{W}}^{1/2}$$

que coincide com $\mathbf{GL}(\hat{\beta})$ apenas para os casos de ligação canônica, o que faz com que $\mathbf{W} = \mathbf{V}$, e resposta contínua, por exemplo, um modelo gama com ligação recíproca.

4.4 Sobrevivência dos pacientes com leucemia

Este exemplo refere-se a um conjunto de dados sobre 33 pacientes com leucemia (Feigl e Zelen, 1965), sobre os quais verificou-se a presença de uma característica morfológica nas células brancas. Os pacientes foram então classificados como AG positivo e AG negativo, em que, intuitivamente, o primeiro grupo é daqueles que possuem a característica. Também registrou-se o tempo de sobrevivência do paciente (em semanas) após o diagnóstico da doença e o número de glóbulos brancos (WBC) no momento do diagnóstico.

Supondo que o tempo de sobrevivência após o diagnóstico segue uma distribuição gama ($\text{tempo} \sim G(\mu, \phi)$), ajustamos um MLG com função de ligação inversa para explicar o tempo de sobrevivência dados o $\ln(WBC)$ e AG (positivo=1, ou negativo=0). Um primeiro ajuste apontou o intercepto como um parâmetro não significativo e foi retirado do modelo.

A variável $\ln(WBC)$, segundo a característica AG, na Figura 4.1, mostra que não possui observações discrepantes, ao contrário de WBC que possui, Figura 4.2, além de um comportamento muito parecido com e sem a presença da característica AG, o que sugere que não há interação entre as duas variáveis. A Figura 4.3 mostra o comportamento do tempo de sobrevivência segundo a característica AG, os quais mostram-se muito diferentes, sem a presença da característica o tempo de sobrevivência mostra-se, para quase a totalidade dos indivíduos, pequeno, enquanto com a presença da característica ele mostra-se com uma amplitude e média muito maiores.

No experimento, 17 pacientes possuíam a característica AG, com tempo médio de sobrevivência de 62,5 semanas, enquanto os 16 restantes que não possuíam a característica AG, tiveram tempo médio de sobrevivência de 18 semanas, contudo, a média do número de glóbulos brancos não apresentou diferença significativa entre aqueles que possuem a característica AG e aqueles

Figura 4.1: *Box plot* de $\ln(\text{WBC})$ segundo AG.

Tabela 4.1: Resultados do ajuste do modelo.

Parâmetro	Estimativa	Erro Padrão	Valor de t
AG	-0,035	0,010	-3,595
$\ln(\text{WBC})$	0,006	0,001	4,864

que não possuem.

O ajuste do modelo gerou as estimativas contidas na Tabela 4.1 para os parâmetros desse modelo. O modelo gerou um desvio de 38,299 com 31 graus de liberdade ($\hat{\phi}=0,951(0,205)$), o que resulta num p-valor de 14,22%, indicando não rejeição do modelo sob o nível mínimo de significância de 10%.

A Figura 4.4 apresenta os resíduos com uma tendência de queda da variância conforme o aumento do valor ajustado, apresenta também uma observação (14) fora do limite $(-2;2)$, porém, bem próxima do limite, e outra (2) que se apresenta bastante espaçado no espaço. O gráfico de envelope,

Figura 4.2: *Box plot* de WBC segundo AG. Figura 4.3: *Box plot* do tempo de sobrevivência segundo AG.

Figura 4.4: Resíduo t_{Di} contra os valores ajustados.

Figura 4.5 confirma que o ajuste não foi bom visto que vários resíduos ficam fora das bandas de confiança.

Tentamos um novo ajuste usando a função de ligação logarítmica com o intuito de estabilizar a variância, o que gerou as estimativas contidas na Tabela 4.2. O modelo gerou um Desvio de 38,327 com 30 graus de liberdade, o que resulta num p -valor de 14,15%, indicando não rejeição do modelo sob o nível mínimo de significância de 10% .

Tabela 4.2: Resultados do ajuste do modelo.

Parâmetro	Estimativa	Erro Padrão	Valor de t
Intercepto	5,813	1,348	4,312
lnWBC	-0,304	0,137	-2,213
AG	1,018	0,364	2,797

Figura 4.5: Gráfico de envelope para ligação canônica.

Para o novo ajuste a Figura 4.6 apresenta os resíduos versus os valores ajustados com um comportamento bem mais próximo da aleatoriedade, apresenta também uma observação (14) fora do limite $(-2;2)$, porém, bem próxima do limite, e outra (2) que se apresenta, novamente, bastante espaçado no espaço, o que sugere que seja um ponto de alavanca. O gráfico de envelope, Figura 4.7 apresenta um comportamento melhor do que a Figura 4.5, evidenciando um melhor ajuste. A Figura 4.8 mostra a observação 33 como um candidato a ponto influente, porém, bem abaixo do que poderíamos usar como critério $\chi^2_3(90\%) = 6,25$.

Na Figura 4.9 podemos observar o comportamento da diagonal principal da matriz de projeção (3.18), na qual podemos ver que a observação 2 se destaca. Nas Figuras 4.10 e 4.11 podemos observar um comportamento parecido com a Figura 4.9, contudo, aquelas apresentam a observação 2 como ponto de alavanca com maior intensidade e em menor grau apresentam a observação

Figura 4.6: Resíduo t_{Di} contra os valores ajustados.

Figura 4.7: Gráfico de envelope para ligação logarítmica.

Figura 4.8: Distância de Cook aproximada LD_i .

21 como candidato a ponto de alavanca. É relevante notar a similaridade entre os valores do plano tangente e da alavanca Jacobiana.

Aplicamos também a alavanca generalizada para MLGs com função de ligação geral (4.17), cuja expressão utilizada seguiu as seguintes considerações:

$$\begin{aligned}
 \ln(\mu_i) &= \eta_i, \quad \text{logo} \quad \mu_i = e^{\eta_i}; \\
 \frac{\partial \mu_i}{\partial \eta_i} &= e^{\eta_i} = \mu_i; \\
 \theta_i &= \frac{-1}{\mu_i} = \frac{-1}{e^{\eta_i}} = -e^{-\eta_i}, \quad \text{logo} \quad \frac{\partial \theta_i}{\partial \eta_i} = e^{-\eta_i} = \mu_i^{-1} \quad \text{e} \\
 \frac{\partial^2 \theta_i}{\partial \eta_i^2} &= -e^{-\eta_i} = -\mu_i^{-1}.
 \end{aligned}$$

Por independência temos que

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T} = \text{diag}(\boldsymbol{\mu}) \quad \text{e} \quad \frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^T \partial \boldsymbol{\eta}} = -\text{diag}^*(\boldsymbol{\mu}),$$

Figura 4.9: Matriz de projeção $\hat{H}^*$.

Figura 4.10: Plano tangente.

Figura 4.11: Alavanca Jacobiana.

em que $\text{diag}(\boldsymbol{\mu})$ é uma matriz $n \times n$ diagonal, cuja diagonal principal é o vetor $\boldsymbol{\mu}$, a matriz $\text{diag}^*(\boldsymbol{\mu})$ é uma matriz $n \times n \times n$ na qual a i -ésima face possui apenas o valor referente à i -ésima coluna e à i -ésima linha diferente de zero, e este valor corresponde ao i -ésimo valor do vetor $\boldsymbol{\mu}$.

Com isso, a expressão 4.17 torna-se

$$\text{GL}(\hat{\boldsymbol{\beta}}) = \text{diag}(\boldsymbol{\mu})\mathbf{X}\left\{\mathbf{X}^T\mathbf{X} - \mathbf{X}^T[\boldsymbol{\epsilon}^T][-\text{diag}^*(\boldsymbol{\mu})]\mathbf{X}\right\}^{-1}\mathbf{X}^T,$$

a diagonal principal resultante dessa expressão está na Figura 4.12 e mostra que a observação 2 pode ser considerada um ponto de alavanca na regressão.

Observamos que, claramente, podemos considerar a observação (2) como de alavanca e a observação 21 como candidata, contudo, nenhuma das duas pode ser considerada influente, nem nenhuma outra. Ressaltamos também a grande similaridade entre as estatísticas plano tangente e alavanca Jacobiana. No geral, podemos considerar grande a concordância entre as estatísticas $\hat{\mathbf{H}}^*$,

Figura 4.12: Matriz $\mathbf{GL}(\hat{\beta})$.

plano tangente, alavanca Jacobiana e $\mathbf{GL}(\hat{\beta})$. A retirada das observações 2 e 21 não resultou em melhora no ajuste do modelo.

Capítulo 5

Conclusões

5.1 Conclusões

A idéia original de se medir e identificar pontos de alavanca, por meio de $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, a matriz de projeção dos valores observados sobre o espaço de estimação, foi um avanço muito importante, criou-se uma base forte em relação ao conceito de alavanca, porém, mostrou-se limitada, à medida que avançou-se na possibilidade de outros tipos de métodos de estimação, além de mínimos quadrados, e do ajuste de modelos não-lineares.

A matriz de alavanca Jacobiana (Saint Laurent e Cook, 1992) e a alavanca generalizada (Wei, Hu e Fung, 1998), surgiram como generalizações, tanto no campo dos métodos de estimação quanto na possibilidade de ajuste de modelos não-lineares, principalmente a última.

A matriz de alavanca Jacobiana apresenta o avanço de levar em conta a não-linearidade do modelo, porém, ainda para estimadores de mínimos quadrados. Contudo, ocorre que não gera uma matriz idempotente, o que dificulta a avaliação se a observação é de alavanca ou não. O plano tangente e a alavanca Jacobiana mostraram-se muito similares nos exemplos usados, contudo, o plano tangente é uma medida que apresenta risco quando do

ajuste de modelos com espaço de estimação pouco suave.

A alavanca generalizada apresenta grandes avanços, permitindo não apenas a medição em modelos não-lineares, como também o uso em praticamente todos os métodos de estimação (desde de que exista uma função de estimação fechada). Com uma expressão fechada para qualquer método de estimação e modelo, apenas simplificando-se conforme o caso, todavia, balizando-se por $\mathbf{H}$, visto que para modelos lineares usando estimadores de mínimos quadrados a alavanca Jacobiana possui a mesma expressão de $\mathbf{H}$. Alavanca generalizada torna-se uma medida praticamente universal. Uma limitação, talvez, seja a obtenção de matrizes que podem não ser idempotentes e, portanto, podem não ter uma estrutura conhecida o que dificulta na criação de limites para balizar o que sejam observações de alavanca.

Apêndice A

Apêndice

Demonstração de $G(p; \mathbf{f})$

Expandindo $\mu(\hat{\beta}_f(p))$ em torno de $p = 0$ via série de Taylor até segunda ordem, obtemos

$$\begin{aligned}
 \mu(\hat{\beta}_f(p)) &= \mu(\hat{\beta}_f(0)) + \frac{\partial \mu(\hat{\beta}_f(p))}{\partial p} \Big|_{p=0} \frac{(p-0)}{1!} + \frac{\partial^2 \mu(\hat{\beta}_f(p))}{\partial^2 p} \Big|_{p=0} \frac{(p-0)^2}{2!} \\
 &= \mu(\hat{\beta}_f(0)) + \frac{\partial \mu(\hat{\beta}_f(p))}{\partial \hat{\beta}_f(p)} \frac{\partial \hat{\beta}_f(p)}{\partial p} \Big|_{p=0} \frac{(p-0)}{1!} \\
 &\quad + \left\{ \frac{\partial \hat{\beta}_f^T(p)}{\partial p} \frac{\partial^2 \mu(\hat{\beta}_f(p))}{\partial \hat{\beta}_f(p) \partial \hat{\beta}_f^T(p)} \frac{\partial \hat{\beta}_f(p)}{\partial p} + \frac{\partial \mu(\hat{\beta}_f(p))}{\partial \hat{\beta}_f(p)} \frac{\partial^2 \hat{\beta}_f(p)}{\partial p^2} \right\} \Big|_{p=0} \frac{(p-0)^2}{2!} \\
 &= \mu(\hat{\beta}_f(0)) + \hat{\mathbf{D}}_\beta \Delta_f^1 \Big|_{p=0} p + \left\{ (\Delta_f^1)^T \hat{\mathbf{D}}_{\beta\beta} \Delta_f^1 + \hat{\mathbf{D}}_\beta \Delta_f^2 \right\} \Big|_{p=0} \frac{p^2}{2}.
 \end{aligned}$$

Com esse resultado podemos encontrar uma forma fechada para a alavanca generalizada,

$$G(p; \mathbf{f}) = \frac{\mu(\hat{\beta}_f(p)) - \mu(\hat{\beta}_f(0))}{p} = \hat{\mathbf{D}}_\beta \Delta_f^1 + \frac{p}{2} \left\{ (\Delta_f^1)^T \hat{\mathbf{D}}_{\beta\beta} \Delta_f^1 + \hat{\mathbf{D}}_\beta \Delta_f^2 \right\}.$$

Para obter os termos Δ_f^1 e Δ_f^2 derivamos a função,

$$\left\{ \mathbf{y} + p\mathbf{f} - \boldsymbol{\mu}(\boldsymbol{\beta}_f(p)) \right\}^T \left\{ \mathbf{y} + p\mathbf{f} - \boldsymbol{\mu}(\boldsymbol{\beta}_f(p)) \right\},$$

em relação a $\boldsymbol{\beta}^T$, obtendo,

$$\mathbf{K} = \left\{ \mathbf{y} + p\mathbf{f} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}}_f(p)) \right\}^T \hat{\mathbf{D}}_\beta = 0.$$

O resultado, derivamos uma vez em relação a p , igualando a zero, para encontrar Δ_f^1 , e derivamos duas vezes, em relação a p , para encontrar Δ_f^2 , conforme a seguir.

Encontrando Δ_f^1 ,

$$\begin{aligned} \frac{\partial \mathbf{K}}{\partial p} &= (\mathbf{f} - \hat{\mathbf{D}}_\beta \Delta_f^1)^T \hat{\mathbf{D}}_\beta + (\Delta_f^1)^T [\hat{\boldsymbol{\epsilon}}^T][\hat{\mathbf{D}}_{\beta\beta}] \\ &= \mathbf{f}^T \hat{\mathbf{D}}_\beta - (\Delta_f^1)^T \hat{\mathbf{D}}_\beta^T \hat{\mathbf{D}}_\beta + (\Delta_f^1)^T [\hat{\boldsymbol{\epsilon}}^T][\hat{\mathbf{D}}_{\beta\beta}] \\ &= \mathbf{f}^T \hat{\mathbf{D}}_\beta - (\Delta_f^1)^T \left\{ \hat{\mathbf{D}}_\beta^T \hat{\mathbf{D}}_\beta - [\hat{\boldsymbol{\epsilon}}^T][\hat{\mathbf{D}}_{\beta\beta}] \right\} \\ &= \mathbf{f}^T \hat{\mathbf{D}}_\beta - (\Delta_f^1)^T \hat{\mathbf{A}} = 0, \end{aligned}$$

em que

$$\hat{\mathbf{A}} = \left\{ \hat{\mathbf{D}}_\beta^T \hat{\mathbf{D}}_\beta - [\hat{\boldsymbol{\epsilon}}^T][\hat{\mathbf{D}}_{\beta\beta}] \right\}.$$

Conseqüentemente,

$$\begin{aligned} \hat{\mathbf{A}} \Delta_f^1 &= \hat{\mathbf{D}}_\beta^T \mathbf{f} \\ \Delta_f^1 &= \hat{\mathbf{A}}^{-1} \hat{\mathbf{D}}_\beta^T \mathbf{f}. \end{aligned}$$

Agora Δ_f^2 ,

$$\begin{aligned}
 \frac{\partial^2 \mathbf{K}}{\partial^2 p} &= (\Delta_f^1)^T [\mathbf{f}^T] [\widehat{\mathbf{D}}_{\beta\beta}] - (\Delta_f^2)^T \widehat{\mathbf{A}} - (\Delta_f^1)^T \left\{ (\Delta_f^1)^T [\widehat{\mathbf{D}}_{\beta\beta}^T] [\widehat{\mathbf{D}}_{\beta}] \right. \\
 &\quad \left. + (\Delta_f^1)^T [\widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] - [(\Delta_f^1)^T] [[\widehat{\epsilon}^T] [\widehat{\mathbf{D}}_{\beta\beta\beta}]] - [\mathbf{f}^T - (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] \right\} \\
 &= (\Delta_f^1)^T [\mathbf{f}^T - (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] - (\Delta_f^2)^T \widehat{\mathbf{A}} - (\Delta_f^1)^T \left\{ (\Delta_f^1)^T [\widehat{\mathbf{D}}_{\beta\beta}^T] [\widehat{\mathbf{D}}_{\beta}] \right. \\
 &\quad \left. - [(\Delta_f^1)^T] [[\widehat{\epsilon}^T] [\widehat{\mathbf{D}}_{\beta\beta\beta}]] \right\} - [\mathbf{f}^T - (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] \\
 &= 2(\Delta_f^1)^T [\mathbf{f}^T - (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] - (\Delta_f^2)^T \widehat{\mathbf{A}} \\
 &\quad - (\Delta_f^1)^T \left\{ (\Delta_f^1)^T [\widehat{\mathbf{D}}_{\beta\beta}^T] [\widehat{\mathbf{D}}_{\beta}] - [(\Delta_f^1)^T] [[\widehat{\epsilon}^T] [\widehat{\mathbf{D}}_{\beta\beta\beta}]] \right\} = 0.
 \end{aligned}$$

Logo

$$\begin{aligned}
 (\Delta_f^2)^T \widehat{\mathbf{A}} &= 2(\Delta_f^1)^T [\mathbf{f}^T - (\Delta_f^1)^T \widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] \\
 &\quad - (\Delta_f^1)^T \left\{ (\Delta_f^1)^T [\widehat{\mathbf{D}}_{\beta\beta}^T] [\widehat{\mathbf{D}}_{\beta}] - [(\Delta_f^1)^T] [[\widehat{\epsilon}^T] [\widehat{\mathbf{D}}_{\beta\beta\beta}]] \right\} \\
 &\quad \therefore \\
 \widehat{\mathbf{A}} \Delta_f^2 &= 2[\widehat{\mathbf{D}}_{\beta\beta}^T] [\mathbf{f} - \widehat{\mathbf{D}}_{\beta} \Delta_f^1] \Delta_f^1 \\
 &\quad - \left\{ [\widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] \Delta_f^1 - [[\widehat{\mathbf{D}}_{\beta\beta\beta}^T] [\widehat{\epsilon}]] [\Delta_f^1] \right\} \Delta_f^1 \\
 &\quad \therefore \\
 \Delta_f^2 &= 2\widehat{\mathbf{A}}^{-1} [\widehat{\mathbf{D}}_{\beta\beta}^T] [\mathbf{f} - \widehat{\mathbf{D}}_{\beta} \Delta_f^1] \Delta_f^1 \\
 &\quad - \widehat{\mathbf{A}}^{-1} \left\{ [\widehat{\mathbf{D}}_{\beta}^T] [\widehat{\mathbf{D}}_{\beta\beta}] \Delta_f^1 - [[\widehat{\mathbf{D}}_{\beta\beta\beta}^T] [\widehat{\epsilon}]] [\Delta_f^1] \right\} \Delta_f^1.
 \end{aligned}$$

Com Δ_f^1 e Δ_f^2 completamos a demonstração.

Expressão mais complexa de $\mathbf{GL}(\beta)$

Primeiramente, vamos à busca de $\ddot{\mathbf{L}}_{\beta\beta}$, derivando $\mathbf{L}(\theta)$ em relação à β ,

$$\begin{aligned}\frac{\partial \mathbf{L}(\theta)}{\partial \beta^T} &= \phi \left\{ \mathbf{y}^T \frac{\partial \theta}{\partial \mu^T} \frac{\partial \mu}{\partial \eta^T} \frac{\partial \eta}{\partial \beta^T} - \left(\frac{\partial \mathbf{b}(\theta)}{\partial \theta} \right)^T \frac{\partial \theta}{\partial \mu^T} \frac{\partial \mu}{\partial \eta^T} \frac{\partial \eta}{\partial \beta^T} \right\} = \dot{\mathbf{L}}_{\beta} \\ \dot{\mathbf{L}}_{\beta} &= \phi \left\{ \mathbf{y}^T \mathbf{V}^{-1} \mathbf{U}^{-1} \mathbf{R} - \mu^T \mathbf{V}^{-1} \mathbf{U}^{-1} \mathbf{R} \right\} \\ &= \phi \left\{ \mathbf{y}^T \mathbf{V}^{-\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{R} - \mu^T \mathbf{V}^{-\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{R} \right\} \\ &= \phi \left\{ (\mathbf{y} - \mu)^T \mathbf{V}^{-\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{R} \right\} \\ &= \phi \left\{ \epsilon^T \mathbf{V}^{-1} \mathbf{D}_{\beta} \right\}.\end{aligned}$$

Em seguida derivamos $\dot{\mathbf{L}}_{\beta}$ em relação à β novamente, obtendo $\ddot{\mathbf{L}}_{\beta\beta}$

$$\begin{aligned}\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \beta} &= \phi \left\{ -\mathbf{R}^T \mathbf{W}^{1/2} \mathbf{V}^{1/2} \mathbf{V}^{-1/2} \mathbf{W}^{1/2} \mathbf{R} + \mathbf{D}_{\beta}^T [\epsilon^T] \left[\frac{\partial^2 \theta}{\partial \mu^T \partial \mu} \right] \mathbf{D}_{\beta} \right. \\ &\quad \left. + \mathbf{R}^T [\epsilon^T] \left[\mathbf{V}^{-1} \frac{\partial^2 \mu}{\partial \eta^T \partial \eta} \right] \mathbf{R} + [\epsilon^T \mathbf{V}^{-1} \mathbf{U}^{-1}] \left[\frac{\partial^2 \eta}{\partial \beta^T \partial \beta} \right] \right\} \\ &= \phi \left\{ -\mathbf{R}^T \mathbf{W} \mathbf{R} + \mathbf{D}_{\beta}^T [\epsilon^T] \left[\frac{\partial^2 \theta}{\partial \mu^T \partial \mu} \right] \mathbf{D}_{\beta}^T \right. \\ &\quad \left. + \mathbf{R}^T [\epsilon^T] \left[\mathbf{V}^{-1} \frac{\partial^2 \mu}{\partial \eta^T \partial \eta} \right] \mathbf{R} + [\epsilon^T \mathbf{V}^{-1} \mathbf{U}^{-1}] \left[\frac{\partial^2 \eta}{\partial \beta^T \partial \beta} \right] \right\},\end{aligned}$$

ou de outra forma

$$\begin{aligned}\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \beta} &= \phi \left\{ -\mathbf{D}_{\beta}^T \mathbf{V}^{-1} \mathbf{D}_{\beta} + \mathbf{D}_{\beta}^T [\epsilon^T] \left[\frac{\partial^2 \theta}{\partial \mu^T \partial \mu} \right] \mathbf{D}_{\beta} + \mathbf{R}^T [\epsilon^T] \left[\mathbf{V}^{-1} \frac{\partial^2 \mu}{\partial \eta^T \partial \eta} \right] \mathbf{R} \right. \\ &\quad \left. + [\epsilon^T \mathbf{V}^{-1} \mathbf{U}^{-1}] \left[\frac{\partial \mathbf{R}}{\partial \beta} \right] \right\} = \ddot{\mathbf{L}}_{\beta\beta}.\end{aligned}\tag{A.1}$$

Por fim, encontramos o termo $\dot{\mathbf{L}}_{\beta}$, a seguir

$$\frac{\partial \dot{\mathbf{L}}_{\beta}}{\partial \mathbf{y}} = \phi \mathbf{D}_{\beta}^T \mathbf{V}^{-1} = \ddot{\mathbf{L}}_{\beta y}.\tag{A.2}$$

Podemos, então, montar a expressão para $\mathbf{GL}(\hat{\beta})$, juntando os termos $\mathbf{D}_\beta$, (A.1) e (A.2), como se segue

$$\begin{aligned} \mathbf{GL}(\hat{\beta}) = & \mathbf{D}_\beta \left\{ -\mathbf{D}_\beta^T \mathbf{V}^{-1} \mathbf{D}_\beta + \mathbf{D}_\beta^T [\boldsymbol{\epsilon}^T] \left[\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T \partial \boldsymbol{\mu}} \right] \mathbf{D}_\beta + \mathbf{R}^T [\boldsymbol{\epsilon}^T] \left[\mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T \partial \boldsymbol{\eta}} \right] \mathbf{R} \right. \\ & \left. + [\boldsymbol{\epsilon}^T \mathbf{V}^{-1} \mathbf{U}^{-1}] \left[\frac{\partial^2 \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right] \right\}^{-1} \mathbf{D}_\beta^T \mathbf{V}^{-1} \Big|_{\beta=\hat{\beta}(y)}. \end{aligned}$$

Conjuntos de dados dos Exemplos

Tabela A.1: Conjunto de Observações para o exemplo 2.4 - Imóveis.

Ordem	Imposto do terreno	Área do terreno	Área construída	Idade do imóvel	Valor de venda do imóvel
1	4,918	3,472	0,998	42	25,9
2	5,021	3,531	1,500	62	29,5
3	4,543	2,275	1,175	40	27,9
4	4,557	4,050	1,232	54	25,9
5	5,060	4,455	1,121	42	29,9
6	3,891	4,455	0,988	56	29,9
7	5,898	5,850	1,240	51	30,9
8	5,604	9,520	1,501	32	28,9
9	15,420	9,800	3,420	42	84,9
10	14,460	12,800	3,000	14	82,9
11	5,828	6,435	1,225	32	35,9
12	5,300	4,988	1,552	30	31,5
13	6,271	5,520	0,975	30	31,0
14	5,959	6,666	1,121	32	30,9
15	5,050	5,000	1,020	46	30,0
16	8,246	5,150	1,664	50	36,9
17	6,697	6,902	1,488	22	41,9
18	7,784	7,102	1,376	17	40,5
19	9,038	7,800	1,500	23	43,9
20	5,989	5,520	1,256	40	37,5
21	7,542	4,000	1,690	22	37,9
22	8,795	9,890	1,820	50	44,5
23	6,093	6,727	1,652	44	37,9
24	8,361	9,150	1,777	48	38,9
25	8,140	8,000	1,504	3	36,9
26	9,142	7,326	1,831	31	45,8
27	12,000	5,000	1,200	30	41,0

Tabela A.2: Conjunto de Observações para o exemplo 2.5 - Cerejeiras.

Ordem	Diâmetro da árvore	Altura da árvore	Volume da árvore
1	8,3	70	10,3
2	8,6	65	10,3
3	8,8	63	10,2
4	10,5	72	16,4
5	10,7	81	18,8
6	10,8	83	19,7
7	11,0	66	15,6
8	11,0	75	18,2
9	11,1	80	22,6
10	11,2	75	19,9
11	11,3	79	24,2
12	11,4	76	21,0
13	11,4	76	21,4
14	11,7	69	21,3
15	12,0	75	19,1
16	12,9	74	22,2
17	12,9	85	33,8
18	13,3	86	27,4
19	13,7	71	25,7
20	13,8	64	24,9
21	14,0	78	34,5
22	14,2	80	31,7
23	14,5	74	36,3
24	16,0	72	38,3
25	16,3	77	42,6
26	17,3	81	55,4
27	17,5	82	55,7
28	17,9	80	58,3
29	18,0	80	51,5
30	18,0	80	51,0
31	20,6	87	77,0

Tabela A.3: Conjunto de Observações para o exemplo 3.5 - Relação peso dos olhos e idade de coelhos selvagens.

Ordem	Idade do coelho	Peso dos olhos
1	15	21,66
2	15	22,75
3	15	22,30
4	18	31,25
5	28	44,79
6	29	40,55
7	37	50,25
8	37	46,88
9	44	52,03
10	50	63,47
11	50	61,13
12	60	81,00
13	61	73,09
14	64	79,09
15	65	79,51
16	65	65,31
17	72	71,90
18	75	86,10
19	75	94,60
20	82	92,50
21	85	105,00
22	91	101,70
23	91	102,90
24	97	110,00

Ordem	Idade do coelho	Peso dos olhos
25	98	104,30
26	125	134,90
27	142	130,68
28	142	140,58
29	147	155,30
30	147	152,20
31	150	144,50
32	159	142,15
33	165	139,81
34	183	153,22
35	192	145,72
36	195	161,10
37	218	174,18
38	218	173,03
39	219	173,54
40	224	178,86
41	225	177,68
42	227	173,73
43	232	159,98
44	232	161,29
45	237	187,07
46	246	176,13
47	258	183,40
48	276	186,26

Ordem	Idade do coelho	Peso dos olhos
49	285	189,66
50	300	186,09
51	301	186,70
52	305	186,80
53	312	195,10
54	317	216,41
55	338	203,23
56	347	188,38
57	354	189,70
58	357	195,31
59	375	202,63
60	394	224,82
61	513	203,30
62	535	209,70
63	554	233,90
64	591	234,70
65	648	244,30
66	660	231,00
67	705	242,40
68	723	230,77
69	756	242,57
70	768	232,12
71	860	246,70

Tabela A.4: Conjunto de Observações para o exemplo 4.4 - Sobrevida de pacientes com leucemia.

Ordem	Característica AG	Tempo	Número de glóbulos brancos
1	1	65	2300
2	1	156	750
3	1	100	4300
4	1	134	2600
5	1	16	6000
6	1	108	10500
7	1	121	10000
8	1	4	17000
9	1	39	5400
10	1	143	7000
11	1	56	9400
12	1	26	32000
13	1	22	35000
14	1	1	100000
15	1	1	100000
16	1	5	52000
17	1	65	100000
18	0	56	4400
19	0	65	3000
20	0	17	4000
21	0	7	1500
22	0	16	9000
23	0	22	5300
24	0	3	10000
25	0	4	19000
26	0	2	27000
27	0	3	28000
28	0	8	31000
29	0	4	26000
30	0	3	21000
31	0	30	79000
32	0	4	100000

Referências Bibliográficas

- Anscombe, F. J. e Tukey, J. W. (1963). The Examination and Analysis of Residuals. *Technometrics* **5**, 141-159.
- Bacon, D. W. e Watts, D. G. (1971). Estimating the Transition Between Two Intersecting Straight Lines. *Biometrika* **58**, 525-534.
- Bates, D. M. e Watts, D. G. (1980). Relative Curvature Measures of Nonlinearity (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **42**, 1-25.
- Belsley, D. A., Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics*. New York: John Wiley.
- Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics* **19**, 15-18.
- Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Cordeiro, G. M. e Paula, G. A. (1989a). Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models. *Biometrika* **76**, 93-100.
- Cordeiro, G. M. e Paula, G. A. (1989b). *Modelos de Regressão para a Análise de Dados Univariados*. 17o. Colóquio Brasileiro de Matemática.

- Dudzinski, M. L. e Mykytowycz, R. (1961). The Eye Lens as an Indicator of Age in The Wild Rabbit in Australia. *CSIRO. WILDL. RES.* **6**, 156-159.
- Emerson, J. D.; Hoaglin, D. C. e Kempthorne, P. J. (1984). Leverage in Least Squares Additive-Plus-Multiplicative Fits for Two-Way Tables. *Journal of the American Statistical Association* **79**, 329-335.
- Feigl, P. e Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* **21**, 826-838.
- Gay, D. M. e Welsch, R. E. (1988). Maximum Likelihood and Quasi-Likelihood for Nonlinear Exponential Family Regression Models. *Journal of the American Statistical Association* **83**, 990-998.
- Hoaglin, D. C. e Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician* **32**, 17-22.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models, 2nd. Edition*. Chapman and Hall.
- Narula, S. C. e Wellington, J. F. (1977). Prediction, linear regression and minimum sum of relative erros. *Technometrics* **19**, 185-190.
- Paula, G. A. (1992). Bias correction for exponential family nonlinear models. *Journal of Statistical Computation and Simulation* **40**, 43-54.
- Paula, G. A. (1999). Leverage in Inequality-Constrained Regression Models. *The Statistician* **48**, 529-538.
- Paula, G. A. (2003). *Modelos de Regressão com Apoio Computacional*. Versão Preliminar, IME-USP, 300p.

- Paula, G. A. e Cordeiro, G. M. (1986). Alguns modelos não-lineares via o Glim. *Atas do VII Simpósio Nacional de Probabilidade e Estatística*, UNICAMP, São Paulo, pp. 204-217.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modelling*. Marcel Dekker, New York.
- Ryan, B. F. e Joiner, B. L. (1994). *Minitab Handbook, Third Edition*. Duxbury Press.
- Saint Laurent, R. T. e Cook, R. D. (1992). Leverage and Superleverage in Nonlinear Regression. *Journal of the American Statistical Association* **87**, 985-990.
- Velleman, P. F. e Welsch, R. (1981). Efficient Computing of Regression Diagnostics. *The American Statistician* **35**, 234-242.
- Wei, B. C. (1998). *Exponential Family Nonlinear Models*. Lecture Notes in Statistics Vol. 130. Springer, New York.
- Wei, B.C., Hu, Y.Q., e Fung, W.K. (1998). Generalized Leverage and its Applications. *Scandinavian Journal of Statistics* **25**, 25-37.