

Métodos de Diagnóstico em Modelos Logísticos Trinomiais

José Alberto P. Silva

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: **Gilberto A. Paula**

São Paulo - junho de 2003

Resumo

Os modelos logísticos trinomiais podem ser interpretados como uma extensão natural do modelo logístico binomial para situações em que a resposta admite apenas três resultados possíveis. Introduzimos inicialmente os modelos logísticos trinomiais e discutimos em seguida alguns aspectos inferenciais, tais como estimação e testes. Medidas de qualidade do ajuste são também apresentadas. Contudo, o principal foco deste trabalho é a apresentação de métodos de diagnóstico. Mostramos que as técnicas usuais de diagnóstico desenvolvidas para o modelo logístico binomial podem ser adaptadas para o caso trinomial. O desenvolvimento de métodos diretos para o modelo logístico trinomial é mais complexo do ponto de vista computacional, embora seja sempre possível. Discutimos alguns desses métodos, dentre os quais, o desenvolvimento de resíduos, de métodos para detectar pontos de alavanca, métodos de deleção de pontos e influência local. Comparamos os métodos adaptados com alguns métodos diretos através de exemplos.

Abstract

The trinomial logistic models can be interpreted as a natural extension of the traditional binomial logistic model to situations in which the response allows only three possible results. We first introduce the trinomial logistic models and then some inferential aspects, such as estimation and hypothesis testing are discussed. Goodness-of-fit measures are also given. However, the aim of this work is the presentation of some diagnostic procedures for trinomial logistic models. We show that methods developed for binomial logistic models can be adapted straightforward for trinomial models. The development of direct methods, even though possible, in general requires more complex calculation. Some of these direct methods, such as evaluation of residuals, measures of high leverage points, deletion methods and local influence are presented. Comparisons between adapted and direct methods are made via examples with real data.

Sumário

Resumo	3
Abstract	5
1 Modelos Logísticos Trinomiais	9
1.1 Introdução	9
1.2 Modelo Logístico Trinomial	10
1.3 Função Escore	13
1.4 Matriz de Informação	14
1.5 Processo Iterativo	15
1.6 Função Desvio	16
1.7 Estatística de Hosmer e Lemeshow	18
1.8 Resultados Inferenciais	19
1.9 Exemplos	21
1.9.1 Exemplo 1	21
1.9.2 Exemplo 2	31
2 Métodos de Diagnóstico	37
2.1 Introdução	37
2.2 Pontos de Alavanca	38
2.3 Deleção de Pontos	42

2.4	Influência Local	45
2.5	Influência Local Total	49
2.6	Resíduos	50
3	Aplicações	53
3.1	Introdução	53
3.2	Exemplo 1	54
3.3	Exemplo 2	63
4	Considerações Finais	75
	Apêndice	77
	Bibliografia	95

Capítulo 1

Modelos Logísticos Trinomiais

1.1 Introdução

Regressão logística é mais frequentemente empregada para modelar o relacionamento entre uma variável com resposta dicotômica (binomial) e um conjunto de variáveis explicativas, porém com algumas modificações ela pode também ser utilizada quando a variável resposta é policotômica (multinomial). A extensão do modelo e métodos binomial para multinomial é facilmente ilustrado quando a variável resposta tem três categorias (trinomial). Generalizações para respostas com mais de três categorias é um problema mais notacional que conceitual. Em nosso trabalho consideramos a situação em que a variável resposta assume três categorias (trinomial).

Neste capítulo apresentaremos inicialmente alguns aspectos inferenciais do modelo logístico trinomial e em seguida como ilustração dois exemplos com dados reais serão apresentados. No capítulo 2 discutiremos algumas técnicas de diagnóstico que podem ser aplicadas no caso trinomial. Parte dessas técnicas são adaptações diretas das técnicas usuais da regressão logística binomial. As demais técnicas são desenvolvidas diretamente para o modelo logístico trinomial. No capítulo 3 aplicações e comparações dessas técnicas são feitas nos dois exemplos ilustrativos apresentados no capítulo 1. Conclusões e sugestões são apresentadas no último capítulo.

1.2 Modelo Logístico Trinomial

No desenvolvimento de modelos para respostas policotômicas precisamos ter conhecimento de sua escala de medida. A maioria das aplicações, inclusive o nosso foco, envolve uma variável resposta com escala nominal. Quando trabalhamos com o modelo logístico binomial a variável resposta assume os valores 0 e 1. Sem perda de generalidades vamos supor o modelo logístico trinomial. Seja Y uma variável aleatória categórica que assume os valores $Y=0,1,2$ tais que $\pi_0 = P(Y = 0)$, $\pi_1 = P(Y = 1)$, e $\pi_2 = P(Y = 2)$ sujeito à restrição $\pi_0 + \pi_1 + \pi_2 = 1$.

Para construir a função de probabilidade, é conveniente utilizar três variáveis binárias codificadas como 0 ou 1, tais que

$$\begin{aligned} Z_0 = 1, \quad Z_1 = 0, \quad Z_2 = 0 & \text{ se } Y=0, \\ Z_0 = 0, \quad Z_1 = 1, \quad Z_2 = 0 & \text{ se } Y=1 \quad \text{e} \\ Z_0 = 0, \quad Z_1 = 0, \quad Z_2 = 1 & \text{ se } Y=2. \end{aligned}$$

Note que $Z_0 + Z_1 + Z_2 = 1$. Portanto, a função de probabilidade de $\mathbf{Z} = (Z_0, Z_1, Z_2)^T$ fica dada por

$$g(z; \pi_0, \pi_1, \pi_2) = \pi_0^{z_0} \pi_1^{z_1} \pi_2^{z_2}.$$

Vamos supor agora que a variável aleatória Y é observada independentemente em n unidades experimentais, sendo que para a i -ésima unidade tem-se $P(Y_i = 0) = \pi_{0i}$, $P(Y_i = 1) = \pi_{1i}$ e $P(Y_i = 2) = \pi_{2i}$ e a variável aleatória $\mathbf{Z}_i = (Z_{0i}, Z_{1i}, Z_{2i})^T$, $i = 1, 2, \dots, n$, com as mesmas atribuições de \mathbf{Z} .

A função de probabilidade conjunta de $(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ pode ser expressa na forma

$$g(\mathbf{z}_1, \dots, \mathbf{z}_n; \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n) = \prod_{i=1}^n \pi_{0i}^{z_{0i}} \pi_{1i}^{z_{1i}} \pi_{2i}^{z_{2i}},$$

em que

$$\mathbf{z}_i = (z_{0i}, z_{1i}, z_{2i})^T \quad \text{e} \quad \boldsymbol{\pi}_i = (\pi_{0i}, \pi_{1i}, \pi_{2i})^T$$

sujeito às restrições

$$\pi_{0i} + \pi_{1i} + \pi_{2i} = 1 \quad \text{e} \quad \sum_{i=1}^n (z_{0i} + z_{1i} + z_{2i}) = n.$$

Quando trabalhamos com o modelo logístico binomial a resposta é parametrizada em termos do logito de $Y=1$ versus $Y=0$. No caso do modelo trinomial teremos dois logitos: um para $Y=1$ versus $Y=0$ e outro para $Y=2$ versus $Y=0$. Teoricamente podemos comparar quaisquer pares de logitos; porém, a extensão óbvia do caso binomial é utilizar o logito $Y=2$ versus $Y=0$ para a segunda função. Assim, o grupo codificado como $Y=0$ servirá como um valor de referência para a resposta. O logito para comparar $Y=2$ versus $Y=1$ pode ser obtido como a diferença entre os logitos $Y=2$ versus $Y=0$ e $Y=1$ versus $Y=0$.

É usual (vide, por exemplo, Hosmer e Lemeshow, 1989, Cap. 8) que a parte sistemática do modelo assuma a forma

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \eta_{1i} = \mathbf{x}_i^T \boldsymbol{\beta}_1 \quad (1.1)$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \eta_{2i} = \mathbf{x}_i^T \boldsymbol{\beta}_2, \quad (1.2)$$

em que $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ denota os valores de p variáveis explicativas observadas em cada unidade experimental, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \dots, \beta_{1p})^T$ e $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \dots, \beta_{2p})^T$ são parâmetros desconhecidos a serem estimados. Note que temos dois sub-modelos logísticos neste caso com a mesma relação funcional na parte sistemática, porém com parâmetros diferentes.

Das equações (1.1) e (1.2) mostra-se facilmente que

$$\pi_{0i} = \frac{1}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}}, \quad (1.3)$$

$$\pi_{1i} = \frac{e^{\eta_{1i}}}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}} \quad \text{e} \quad (1.4)$$

$$\pi_{2i} = \frac{e^{\eta_{2i}}}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}}. \quad (1.5)$$

Para ilustrar, considere o exemplo em que Y representa o estado nutricional de um recém-nascido e as respostas são codificadas da seguinte forma:

$$\begin{aligned} Y = 0 & \text{ se o estado nutricional está abaixo do esperado,} \\ Y = 1 & \text{ se o estado nutricional está na média esperada e} \\ Y = 2 & \text{ se o estado nutricional está acima do esperado.} \end{aligned}$$

Vamos supor ainda que sejam observadas as variáveis explicativas X_1 e X_2 tais que $X_1 = 1$ se a criança é pré-matura e $X_1 = 0$ caso contrário, enquanto X_2 representa o peso ao nascer. Pode-se propor, por exemplo, o seguinte modelo logístico:

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i}$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i},$$

em que π_{0i} , π_{1i} e π_{2i} representam, respectivamente, as probabilidades

$$P(Y_i = 0), \quad P(Y_i = 1) \quad \text{e} \quad P(Y_i = 2), \quad i = 1, \dots, n.$$

Pode-se definir algumas razões de chances. Por exemplo, a razão de chances entre prematuro e não-prematuro ter nível nutricional médio supondo o mesmo peso para as crianças ($X_2 = a$) fica dada por

$$\psi_1 = \left\{ \frac{P(Y = 1|X_1 = 1, X_2 = a)}{P(Y = 1|X_1 = 0, X_2 = a)} \right\} = \frac{e^{\beta_{10} + \beta_{11} + \beta_{12}a}}{e^{\beta_{10} + \beta_{12}a}} = e^{\beta_{11}}.$$

Similamente, a razão de chances entre prematuro e não-prematuro ter nível alto supondo o mesmo peso para os recém-nascidos fica dada por

$$\psi_2 = \left\{ \frac{P(Y = 2|X_1 = 1, X_2 = a)}{P(Y = 2|X_1 = 0, X_2 = a)} \right\} = e^{\beta_{21}}.$$

Por outro lado, se os pesos dos recém-nascidos forem diferentes ($X_2 = b$ para pré-maturo e $X_2 = a$ para não pré-maturo) as razões de chances ficam, respectivamente, dadas por

$$\psi_1 = e^{\beta_{11} + \beta_{12}(b-a)} \quad e \quad \psi_2 = e^{\beta_{21} + \beta_{22}(b-a)},$$

dependendo, nesse caso, da diferença de pesos ($b-a$). Assim, uma vantagem da utilização do modelo logístico é a facilidade de interpretação das razões de chances, as quais ficam expressas em função de coeficientes da parte sistemática do modelo.

1.3 Função Escore

Vamos obter a função escore para os parâmetros β_1 e β_2 da parte sistemática do modelo logístico trinomial.

O logaritmo da função de verossilhança de $\beta = (\beta_1^T, \beta_2^T)^T$ fica dado por

$$\begin{aligned} L(\beta) &= \log\{g(\mathbf{z}_1, \dots, \mathbf{z}_n; \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n)\} = \sum_{i=1}^n \{z_{0i} \log \pi_{0i} + z_{1i} \log \pi_{1i} + z_{2i} \log \pi_{2i}\} \\ &= \sum_{i=1}^n \{z_{1i} \eta_{1i} + z_{2i} \eta_{2i} - (z_{0i} + z_{1i} + z_{2i}) \times \log(1 + e^{\eta_{1i}} + e^{\eta_{2i}})\}. \end{aligned}$$

Diferenciando $L(\beta)$ com respeito a β_{ij} obtém-se

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_{ij}} &= \sum_{i=1}^n \left\{ z_{1i} \frac{\partial \eta_{1i}}{\partial \beta_{1j}} - \frac{(z_{0i} + z_{1i} + z_{2i})}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}} e^{\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_{1j}} \right\} = \sum_{i=1}^n \{z_{1i} x_{ij} - (z_{0i} + z_{1i} + z_{2i}) \pi_{1i} x_{ij}\} \\ &= \sum_{i=1}^n (z_{1i} x_{ij} - \pi_{1i} x_{1j}) = \sum_{i=1}^n x_{ij} (z_{1i} - \pi_{1i}). \end{aligned}$$

Em forma matricial tem-se que

$$\mathbf{U}_1(\beta) = \frac{\partial L(\beta)}{\partial \beta_1} = \mathbf{X}^T (\mathbf{z}_1 - \boldsymbol{\pi}_1),$$

em que \mathbf{X} é uma matriz $n \times p$ de linhas \mathbf{x}_i^T , $\mathbf{z}_1 = (z_{11}, \dots, z_{1n})^T$ e $\boldsymbol{\pi}_1 = (\pi_{11}, \dots, \pi_{1n})^T$.

De forma similar chega-se à função escore para β_2

$$\mathbf{U}_2(\beta) = \frac{\partial L(\beta)}{\partial \beta_2} = \mathbf{X}^T (\mathbf{z}_2 - \boldsymbol{\pi}_2),$$

em que \mathbf{X} é uma matriz $n \times p$ de linhas \mathbf{x}_i^T , $\mathbf{z}_2 = (z_{21}, \dots, z_{2n})^T$ e $\boldsymbol{\pi}_2 = (\pi_{21}, \dots, \pi_{2n})^T$.

Portanto, a função escore para $\boldsymbol{\beta}$ pode ser expressa na forma matricial

$$\mathbf{U}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\beta}) \\ \mathbf{U}_2(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T(\mathbf{z}_1 - \boldsymbol{\pi}_1) \\ \mathbf{X}^T(\mathbf{z}_2 - \boldsymbol{\pi}_2) \end{bmatrix}, \quad (1.6)$$

que é uma matriz $2p \times 1$. Ou ainda, se definirmos a matriz \mathbf{S} e os vetores \mathbf{z} e $\boldsymbol{\pi}$ abaixo

$$\mathbf{S} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \quad \text{e} \quad \boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \end{bmatrix},$$

então poderemos reescrever (1.6) na forma matricial simplificada

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{S}^T(\mathbf{z} - \boldsymbol{\pi}). \quad (1.7)$$

1.4 Matriz de Informação

Vamos agora obter a matriz de informação de Fisher para $\boldsymbol{\beta}$. Antes porém precisamos obter as matrizes Hessianas para β_{1j} e β_{2l} .

Diferenciando-se $L(\boldsymbol{\beta})$ com respeito a β_{1j} e β_{1l} obtém-se

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{1j} \partial \beta_{1l}} = - \sum_{i=1}^n \frac{e^{\eta_i} (1 + e^{\eta_{2i}})}{(1 + e^{\eta_i} + e^{\eta_{2i}})^2} x_{ij} x_{il} = - \sum_{i=1}^n \pi_{1i} (1 - \pi_{1j}) x_{ij} x_{il}.$$

De forma similar a derivada de $L(\boldsymbol{\beta})$ com respeito a β_{2j} e β_{2l} fica dada por

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{2j} \partial \beta_{2l}} = - \sum_{i=1}^n \pi_{2i} (1 - \pi_{2j}) x_{ij} x_{il}.$$

A derivada cruzada de $L(\boldsymbol{\beta})$ com respeito a β_{1j} e β_{2l} fica dada por

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_{1j} \partial \beta_{2l}} = \sum_{i=1}^n \frac{e^{\eta_i} e^{\eta_{2i}}}{(1 + e^{\eta_i} + e^{\eta_{2i}})^2} x_{ij} x_{il} = \sum_{i=1}^n \pi_{1i} \pi_{2i} x_{ij} x_{il}.$$

Na forma matricial as matrizes Hessianas assumem, respectivamente, as formas

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_1^T} = -\mathbf{X}^T \mathbf{V}_1 \mathbf{X}, \quad \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2 \partial \beta_2^T} = -\mathbf{X}^T \mathbf{V}_2 \mathbf{X} \quad \text{e} \quad \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_2^T} = -\mathbf{X}^T \mathbf{V}_3 \mathbf{X},$$

em que

$$\begin{aligned}\mathbf{V}_1 &= \text{diag} \{ \pi_{11}(1 - \pi_{11}), \dots, \pi_{1n}(1 - \pi_{1n}) \}, \\ \mathbf{V}_2 &= \text{diag} \{ \pi_{21}(1 - \pi_{21}), \dots, \pi_{2n}(1 - \pi_{2n}) \} \quad \text{e} \\ \mathbf{V}_3 &= \text{diag} \{ -\pi_{11}\pi_{21}, \dots, -\pi_{1n}\pi_{2n} \}.\end{aligned}$$

Portanto, a matriz de informação de Fisher para $\boldsymbol{\beta}$ assume a forma

$$\mathbf{K}(\boldsymbol{\beta}) = E \left\{ -\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\} = \begin{bmatrix} \mathbf{X}^T \mathbf{V}_1 \mathbf{X} & \mathbf{X}^T \mathbf{V}_3 \mathbf{X} \\ \mathbf{X}^T \mathbf{V}_3 \mathbf{X} & \mathbf{X}^T \mathbf{V}_2 \mathbf{X} \end{bmatrix}, \quad (1.8)$$

que é uma matriz não singular $2p \times 2p$. Podemos reexpressar (1.8) na forma abaixo

$$\mathbf{K}(\boldsymbol{\beta}) = \mathbf{S}^T \mathbf{V} \mathbf{S},$$

em que

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_3 \\ \mathbf{V}_3 & \mathbf{V}_2 \end{bmatrix}.$$

1.5 Processo Iterativo

Para estimar os parâmetros $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ pelo método de máxima verossimilhança deve-se fazer $\mathbf{U}_1(\boldsymbol{\beta}) = 0$ e $\mathbf{U}_2(\boldsymbol{\beta}) = 0$. Devido a concavidade da função $L(\boldsymbol{\beta})$ em $\boldsymbol{\beta}$ tem-se a garantia de que as raízes acima levam às estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_2$. Como $\mathbf{U}_1(\boldsymbol{\beta}) = 0$ e $\mathbf{U}_2(\boldsymbol{\beta}) = 0$ são equações não-lineares que em geral não levam a soluções explícitas para $\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_2$, será utilizado o método de Newton-Raphson para resolver as equações acima. Chega-se facilmente ao seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{K}^{-1}(\boldsymbol{\beta}^{(m)}) \mathbf{U}(\boldsymbol{\beta}^{(m)}),$$

que é o mesmo que

$$\begin{pmatrix} \boldsymbol{\beta}_1^{(m+1)} \\ \boldsymbol{\beta}_2^{(m+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1^{(m)} \\ \boldsymbol{\beta}_2^{(m)} \end{pmatrix} + \begin{bmatrix} \mathbf{X}^T \mathbf{V}_1^{(m)} \mathbf{X} & \mathbf{X}^T \mathbf{V}_3^{(m)} \mathbf{X} \\ \mathbf{X}^T \mathbf{V}_3^{(m)} \mathbf{X} & \mathbf{X}^T \mathbf{V}_2^{(m)} \mathbf{X} \end{bmatrix}^{-1} \times \begin{bmatrix} \mathbf{X}^T (\mathbf{z}_1 - \boldsymbol{\pi}_1^{(m)}) \\ \mathbf{X}^T (\mathbf{z}_2 - \boldsymbol{\pi}_2^{(m)}) \end{bmatrix}, \quad (1.9)$$

com $m = 0, 1, 2, \dots$

Similarmente à regressão logística binomial, pode-se iniciar o processo iterativo (1.9) em $\boldsymbol{\beta}_1^{(0)} = \boldsymbol{\beta}_2^{(0)} = 0$, que é equivalente a iniciá-lo em $\boldsymbol{\pi}_1^{(0)} = \boldsymbol{\pi}_2^{(0)} = (1/2, \dots, 1/2)^T$.

1.6 Função Desvio

Para definir a função desvio de forma análoga aos modelos lineares generalizados (vide, por exemplo, McCullagh e Nelder, 1989; Paula, 2003) considere inicialmente o caso de k grupos em que Y_{ij} , $i = 1, \dots, k$ e $j = 1, \dots, n_i$, segue distribuição trinomial com probabilidades $\pi_{0i} = P(Y_{ij} = 0)$, $\pi_{1i} = P(Y_{ij} = 1)$ e $\pi_{2i} = P(Y_{ij} = 2)$. Logo, para cada Y_{ij} pode-se definir a variável aleatória $\mathbf{Z}_{ij} = (Z_{0ij}, Z_{1ij}, Z_{2ij})^T$ com as mesmas atribuições de \mathbf{Z} definida anteriormente.

Portanto, a função de probabilidade de $(\mathbf{Z}_{11}, \dots, \mathbf{Z}_{1n_1}, \dots, \mathbf{Z}_{k1}, \dots, \mathbf{Z}_{kn_k})$ pode ser expressa na forma

$$g(\mathbf{z}_{11}, \dots, \mathbf{z}_{kn_k}; \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_k) = \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_{0i}^{z_{0ij}} \pi_{1i}^{z_{1ij}} \pi_{2i}^{z_{2ij}}, \quad (1.10)$$

em que $\mathbf{z}_{ij} = (z_{0ij}, z_{1ij}, z_{2ij})^T$ e $\boldsymbol{\pi}_i = (\pi_{0i}, \pi_{1i}, \pi_{2i})^T$ com as restrições $\pi_{0i} + \pi_{1i} + \pi_{2i} = 1$ e $\sum_{j=1}^{n_i} (z_{0ij} + z_{1ij} + z_{2ij}) = n_i$. Tem-se ainda que $n = n_1 + n_2 + \dots + n_k$.

Chamando $a_{0i} = \sum_{j=1}^{n_i} z_{0ij}$, $a_{1i} = \sum_{j=1}^{n_i} z_{1ij}$ e $a_{2i} = \sum_{j=1}^{n_i} z_{2ij}$, a função de probabilidade (1.10) fica dada por

$$g(\mathbf{z}_{11}, \dots, \mathbf{z}_{kn_k}; \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_k) = \prod_{i=1}^k \pi_{0i}^{a_{0i}} \pi_{1i}^{a_{1i}} \pi_{2i}^{a_{2i}}. \quad (1.11)$$

O logaritmo da função de verossimilhança de $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^T, \dots, \boldsymbol{\pi}_k^T)^T$ fica dado por

$$L(\boldsymbol{\pi}) = \log\{g(\mathbf{z}_{11}, \dots, \mathbf{z}_{kn_k}; \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_k)\} = \sum_{i=1}^k \{a_{0i} \log \pi_{0i} + a_{1i} \log \pi_{1i} + a_{2i} \log \pi_{2i}\}.$$

Derivando $L(\boldsymbol{\pi})$ com respeito a π_{0i} , π_{1i} , π_{2i} e igualando essas derivadas a zero, chega-se às estimativas de máxima verossimilhança saturadas, dadas por

$$\tilde{\pi}_{0i} = \frac{a_{0i}}{n_i}, \quad \tilde{\pi}_{1i} = \frac{a_{1i}}{n_i} \quad e \quad \tilde{\pi}_{2i} = \frac{a_{2i}}{n_i}.$$

Note que $a_{0i} + a_{1i} + a_{2i} = n_i$ e $\tilde{\pi}_{0i} + \tilde{\pi}_{1i} + \tilde{\pi}_{2i} = 1$. Considere a parte sistemática do modelo dada por (1.1) e (1.2) cujas estimativas de máxima verossimilhança para π_{0i} , π_{1i} e π_{2i} serão, respectivamente, dadas por $\hat{\pi}_{0i}$, $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$. A função desvio que pode ser interpretada como uma distância entre o logaritmo da função de verossimilhança do modelo saturado e o logaritmo da função de verossimilhança do modelo sob investigação fica dada por

$$D(\hat{\boldsymbol{\pi}}; \mathbf{z}) = 2\{L(\tilde{\boldsymbol{\pi}}) - L(\hat{\boldsymbol{\pi}})\} = 2 \sum_{i=1}^k \{a_{0i} \times \log(\tilde{\pi}_{0i}/\hat{\pi}_{0i}) + a_{1i} \times \log(\tilde{\pi}_{1i}/\hat{\pi}_{1i}) + a_{2i} \times \log(\tilde{\pi}_{2i}/\hat{\pi}_{2i})\}.$$

Se ocorrer

$$(a_{0i} = n_i, a_{1i} = 0, a_{2i} = 0), \quad (a_{0i} = 0, a_{1i} = n_i, a_{2i} = 0) \quad \text{ou} \quad (a_{0i} = 0, a_{1i} = 0, a_{2i} = n_i)$$

então o i -ésimo termo de $D(\hat{\boldsymbol{\pi}}; \mathbf{z})$ fica, respectivamente, dado por $-2n_i \log(\hat{\pi}_{0i})$, $-2n_i \log(\hat{\pi}_{1i})$ ou $-2n_i \log(\hat{\pi}_{2i})$.

Supondo que $\frac{n_i}{n} \rightarrow \gamma_i > 0, \forall_i$, conforme $n \rightarrow \infty$, tem-se que $D(\hat{\boldsymbol{\pi}}; \mathbf{z})$ segue sob a hipótese de que o modelo sob investigação é correto, uma distribuição qui-quadrado com $(2k - 2p)$ graus de liberdade. Logo, rejeita-se o modelo postulado se $D(\hat{\boldsymbol{\pi}}; \mathbf{z}) \geq \chi_{(2k-2p)}^2(1 - \alpha)$, sendo $\chi_{(2k-2p)}^2(1 - \alpha)$ o percentil $(1 - \alpha)$ de uma qui-quadrado com $(2k - 2p)$ graus de liberdade. Esse resultado assintótico não continua valendo quando $k \rightarrow \infty$, $n_i \pi_{0i}(1 - \pi_{0i})$ e $n_i \pi_{1i}(1 - \pi_{1i})$ permanecem limitados. Neste caso pode-se utilizar a estatística de Hosmer e Lemeshow (1989) que será descrita na próxima seção.

A função escore para $\boldsymbol{\beta}$ neste caso de k grupos fica dada por

$$\mathbf{U}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\beta}) \\ \mathbf{U}_2(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T(\mathbf{a}_1 - \boldsymbol{\mu}_1) \\ \mathbf{X}^T(\mathbf{a}_2 - \boldsymbol{\mu}_2) \end{bmatrix},$$

sendo que $\mathbf{a}_1 = (a_{11}, a_{12}, \dots, a_{1k})^T$, $\mathbf{a}_2 = (a_{21}, a_{22}, \dots, a_{2k})^T$, $\boldsymbol{\mu}_1 = (n_1 \pi_{11}, n_1 \pi_{12}, \dots, n_1 \pi_{1k})^T$ e $\boldsymbol{\mu}_2 = (n_2 \pi_{21}, n_2 \pi_{22}, \dots, n_2 \pi_{2k})^T$.

A matriz de informação de Fisher para $\boldsymbol{\beta}$ assume a mesma forma dada em (1.8), porém com as mesmas matrizes \mathbf{V}_1 , \mathbf{V}_2 e \mathbf{V}_3 dadas por

$$\begin{aligned}\mathbf{V}_1 &= \text{diag} \{n_1\pi_{11}(1 - \pi_{11}), \dots, n_k\pi_{1k}(1 - \pi_{1k})\}, \\ \mathbf{V}_2 &= \text{diag} \{n_1\pi_{21}(1 - \pi_{21}), \dots, n_k\pi_{2k}(1 - \pi_{2k})\} \quad \text{e} \\ \mathbf{V}_3 &= \text{diag} \{-n_1\pi_{11}\pi_{21}, \dots, -n_k\pi_{1k}\pi_{2k}\}.\end{aligned}$$

O processo iterativo para estimar $\boldsymbol{\beta}$ assume no caso de réplicas a forma dada abaixo

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{S}^T \mathbf{V}^{(m)} \mathbf{S}]^{-1} \mathbf{S}^T (\mathbf{a} - \boldsymbol{\mu}^{(m)}),$$

em que

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_3 \\ \mathbf{V}_3 & \mathbf{V}_2 \end{bmatrix}, \quad \mathbf{a} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix},$$

com $m=0,1,2,\dots$

1.7 Estatística de Hosmer e Lemeshow

O teste qui-quadrado proposto por Hosmer e Lemeshow (1989) para avaliar a qualidade do ajuste de um modelo logístico binomial com dados não agrupados pode ser estendido com pequenas adaptações para cada um dos sub-modelos logísticos que compõe o modelo trinomial. Assim, para o sub-modelo que explica as probabilidades $P(Y_i = 1) = \pi_{1i}$ através da relação dada abaixo

$$\pi_{1i} = \frac{e^{\eta_{1i}}}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}},$$

podemos desenvolver uma estatística tipo qui-quadrado para avaliar a qualidade do ajuste.

Sejam as probabilidades ajustadas nessa categoria denotadas por $\hat{\pi}_{11}, \dots, \hat{\pi}_{1n}$. Essas probabilidades serão colocadas em ordem não decrescente, tais que

$$\hat{\pi}_{1(1)} \leq \hat{\pi}_{1(2)} \leq \dots \leq \hat{\pi}_{1(n)},$$

em que $\hat{\pi}_{1(1)}$, por exemplo, denota a menor probabilidade ajustada dessa categoria e $\hat{\pi}_{1(n)}$ a maior probabilidade ajustada. Supor ainda que os n indivíduos são agrupados em g grupos definidos conforme as probabilidades ajustadas acima, de modo que o primeiro grupo tenha n_{11} elementos e assim por diante até o último grupo com n_{1g} elementos. Para o l -ésimo grupo formado tem-se o valor observado de O_{1l} sucessos, ou seja, O_{1l} indivíduos (dentre os n_{1l} indivíduos) com resposta ($Y=1$). O valor esperado para esse grupo será dado por $n_{1l}\bar{\pi}_{1l}$, em que $\bar{\pi}_{1l}$ denota o valor médio das probabilidades estimadas do grupo. Assim, a estatística de Hosmer-Lemeshow para esse sub-modelo ajustado fica dada por

$$C_1^{HL} = \sum_{l=1}^g \frac{(O_{1l} - n_{1l}\bar{\pi}_{1l})^2}{n_{1l}\bar{\pi}_{1l}(1 - \bar{\pi}_{1l})}.$$

A sugestão é comparar C_1^{HL} com os percentis da distribuição qui-quadrado com $(g - 2)$ graus de liberdade, desde que $n_{1l}, \forall l$, não seja pequeno. O ajuste não será adequado se

$$C_1^{HL} \geq \chi_{(g-2)}^2(1 - \alpha),$$

em que $\chi_{(g-2)}^2(1 - \alpha)$ denota o percentil $(1 - \alpha)$ de uma distribuição qui-quadrado com $(g - 2)$ graus de liberdade. A sugestão é usar aproximadamente $g = 10$ grupos.

Similarmente, para o sub-modelo que explica as probabilidades $P(Y_i = 2) = \pi_{2i}$ através da relação dada abaixo

$$\pi_{2i} = \frac{e^{\eta_{2i}}}{1 + e^{\eta_{1i}} + e^{\eta_{2i}}},$$

podemos definir a estatística

$$C_2^{HL} = \sum_{l=1}^g \frac{(O_{2l} - n_{2l}\bar{\pi}_{2l})^2}{n_{2l}\bar{\pi}_{2l}(1 - \bar{\pi}_{2l})}.$$

O modelo logístico trinomial estará bem ajustado se as duas estatísticas C_1^{HL} e C_2^{HL} apresentarem valores não significativos.

1.8 Resultados Inferenciais

Supondo que $\frac{n_i}{n} \rightarrow \gamma_i > 0, \forall i$, conforme $n \rightarrow \infty$, o estimador de máxima verossimilhança $\hat{\beta}$ segue uma distribuição normal $(2p)$ -variada de média β e matriz de variância-covariância

$\hat{\mathbf{K}}^{-1}(\boldsymbol{\beta})$.

Para testar hipóteses suponha, por exemplo, que o interesse seja testar $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ contra $H_1 : \mathbf{C}\boldsymbol{\beta} \neq 0$, em que \mathbf{C} é uma matriz $r \times (2p)$ de posto completo ($r < 2p$). Denote por $\hat{\boldsymbol{\beta}}^0$ a estimativa de máxima verossimilhança sob H_0 . Assim, a estatística da razão de verossimilhanças para testar H_0 contra H_1 fica dada por

$$\begin{aligned} \xi_{RV} &= 2\{L(\hat{\boldsymbol{\pi}}) - L(\hat{\boldsymbol{\pi}}^0)\} \\ &= 2 \sum_{i=1}^k \{a_{0i} \times \log(\hat{\boldsymbol{\pi}}_{0i}/\hat{\boldsymbol{\pi}}_{0i}^0) + a_{1i} \times \log(\hat{\boldsymbol{\pi}}_{1i}/\hat{\boldsymbol{\pi}}_{1i}^0) + a_{2i} \times \log(\hat{\boldsymbol{\pi}}_{2i}/\hat{\boldsymbol{\pi}}_{2i}^0)\}, \end{aligned} \quad (1.12)$$

sendo $\hat{\boldsymbol{\pi}}^0$ a estimativa de $\boldsymbol{\pi}$ sob H_0 .

Para $\frac{n_i}{n} \rightarrow \gamma_i > 0, \forall_i$, conforme $n \rightarrow \infty$, a estatística ξ_{RV} segue sob H_0 uma distribuição qui-quadrado com $(2p - r)$ graus de liberdade. É interessante notar que (1.12) pode ser expressa como a diferença entre dois desvios

$$\xi_{RV} = D(\hat{\boldsymbol{\pi}}^0; \mathbf{z}) - D(\hat{\boldsymbol{\pi}}; \mathbf{z}).$$

Uma situação de interesse prático seria testar a homogeneidade dos coeficientes dos dois sub-modelos logísticos (1.1) e (1.2), ou seja, testar $H_0 : \beta_{1j} = \beta_{2j}$ ($j = 1, 2, \dots, p$) contra $H_1 : \beta_{1j} \neq \beta_{2j}$, para algum j . Neste caso a matriz \mathbf{C} assume a forma

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ & & \dots & & & & & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{pmatrix},$$

ou seja, \mathbf{C} é uma matriz $(r \times 2p)$ sendo $r = p$. Sob H_0 , a estatística ξ_{RV} segue uma distribuição qui-quadrado com p graus de liberdade.

1.9 Exemplos

1.9.1 Exemplo 1

Vamos considerar para uma primeira análise os dados de um estudo conduzido (vide Hosmer e Lemeshow, 1989) para avaliar fatores associados com conhecimento, atitude e comportamento de mulheres em relação a mamografia. Neste estudo temos o seguinte:

Variável resposta

Experiência: experiência com mamografia, codificada como

0: nunca realizou o exame de mamografia (referência),

1: submeteu-se ao exame de mamografia nos últimos 12 meses e

2: submeteu-se ao exame de mamografia há mais de um ano.

Variáveis explicativas

Sintoma: você não precisa de mamografia a menos que desenvolva os sintomas?

1: concordo totalmente (referência),

2: concordo,

3: discordo e

4: discordo totalmente.

Benefício: benefício percebido (soma de 5 variáveis com escala de 4 pontos cada, ou seja, assume valores de 5 a 20). Um valor baixo indica que a mulher tem forte concordância com os benefícios da mamografia.

Histórico: você tem mãe ou irmã com histórico de câncer de mama?

0: não e

1: sim.

Autoexame: alguém ensinou a você como se auto-examinar?

0: não e

1: sim.

Detectar: você acha provável que a mamografia possa detectar novos casos de câncer de mama?

1: pouco provável (referência),

2: provável e

3: muito provável.

MODELAGEM

Denote por Y_i a experiência da i -ésima mulher entrevistada ($i = 1, 2, \dots, n$), em que $n = 412$. Será assumido que Y_i segue distribuição trinomial ($Y_i=0$ nunca realizou o exame, $Y_i=1$ realizou o exame nos últimos 12 meses e $Y_i=2$ realizou o exame há mais de um ano) com probabilidades dadas por π_{0i} , π_{1i} e π_{2i} , respectivamente.

VARIÁVEL EXPLICATIVA BINÁRIA

A título de ilustração vamos propor uma modelagem considerando apenas a variável explicativa binária **Histórico**. A parte sistemática do modelo fica dada por

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_1 \times \text{Historico}_i \quad \text{e} \quad \log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_2 \times \text{Historico}_i.$$

O subscrito nas razões de chances abaixo indica que a resposta está sendo comparada ao nível de referência ($Y=0$). Isto é, a razão de chances da resposta j versus a resposta 0 para os valores da variável explicativa $X = a$ versus $X = b$ é

$$\psi_j(a, b) = \frac{P(Y = j|X = a)/P(Y = 0|X = a)}{P(Y = j|X = b)/P(Y = 0|X = b)}.$$

Tabela 1.1*Classificação cruzada experiência versus histórico familiar.*

Experiência com mamografia	Histórico	familiar	Total	$\hat{\psi}_j$
0	220	14	234	1,00
1	85	19	104	3,51
2	63	11	74	2,74
Total	368	44	412	

Usando a notação adotada, as duas razões de chances estimadas da Tabela 1.1 são:

$$\hat{\psi}_1 = \frac{19 \times 220}{85 \times 14} = 3,51 \quad \text{e} \quad \hat{\psi}_2 = \frac{11 \times 220}{63 \times 14} = 2,74.$$

As estatísticas acima informam que a chance de uma mulher com histórico de câncer na família ter se submetido ao exame nos últimos 12 meses é 3,51 vezes a de uma mulher que ainda não teve esta experiência, além disso a chance de ter realizado o exame há mais de um ano para uma mulher que já teve a experiência na família é 2,74 vezes a de uma mulher que nunca teve a experiência familiar. O resultado do ajuste do modelo trinomial aos dados (vide apêndice A) é descrito na Tabela 1.2.

Tabela 1.2*Resultados do modelo de regressão logística ajustado à variável histórico.*

Logito	Variável	Coefficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo	$\hat{\psi}_j$
1	Historico	1,2564	0,3747	11,2448	0,0008	3,51
	Intercepto	-0,9510	0,1277	55,4474	0,0001	
2	Historico	1,0093	0,4275	5,5744	0,0182	2,74
	Intercepto	-1,2505	0,1429	76,5842	0,0001	

$$-2 \times \log(\text{verossimilhança}) = 792,340$$

Valores ajustados de $\hat{\psi}_j$ podem ser também obtidos exponenciando o coeficiente angular estimado (coluna $\hat{\psi}_j$ na Tabela 1.2). As estimativas para ψ_j das duas tabelas coincidem. Isso ocorre devido à propriedade de invariância da estimativa de máxima verossimilhança. O mesmo modelo está sendo ajustado em ambos os casos com parametrizações diferentes.

Da Tabela 1.2 temos as seguintes equações estimadas:

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -0,951 + 1,256 \times \text{Historico}_i \quad \text{e} \quad \log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -1,251 + 1,009 \times \text{Historico}_i.$$

Usando os resultados da Tabela 1.2 podemos construir intervalos de confiança aproximados para os coeficientes, que exponenciando nos levam a intervalos de confiança aproximados para as razões de chances. Por exemplo, um intervalo de confiança aproximado de 95% para a razão de chances de $Y = 1$ versus $Y = 0$ fica dado por

$$\exp(1,256 \pm 1,96 \times 0,375) = [1,69; 7,32].$$

O logaritmo da razão de chances para $Y = 2$ versus $Y = 1$ pode ser obtido da diferença entre os dois coeficientes angulares estimados no modelo de regressão logística correspondente. Ilustrando com os resultados da Tabela 1.2 obtemos

$$\log \left(\frac{19 \times 63}{11 \times 85} \right) = 1,256 - 1,009 = 0,247,$$

mostrando que ao comparar mulheres que já fizeram o exame, a chance de ter realizado nos últimos 12 meses é maior (1,28 vezes) que ter realizado a mais de um ano quando a mulher já teve a experiência na família do que quando nunca teve histórico de câncer familiar.

Para uma indicação preliminar da importância da variável, vamos utilizar o teste da razão de verossimilhanças (1.12) para avaliar a significância dos coeficientes:

$$\xi_{RV} = 2\{L(\hat{\boldsymbol{\pi}}) - L(\hat{\boldsymbol{\pi}}^0)\} = 805,198 - 792,340 = 12,8581.$$

Como foi visto em (1.12), ξ_{RV} segue sob H_0 (os coeficientes são nulos) uma distribuição qui-quadrado com $(2p - r)$ graus de liberdade, quando n é grande. Para esse caso, temos 2 graus de liberdade o que produz um nível descritivo de 0,0016. Assim, do ponto de vista inferencial, o fato da mulher ter um histórico de câncer na família está fortemente associado com a decisão de submeter-se a um exame de mamografia.

VÁRIVEL EXPLICATIVA POLICOTÔMICA

Para uma variável explicativa policotômica expandimos o número de razões de chances para incluir comparações de cada nível ao nível de referência para cada logito possível. Para ilustrar, vamos considerar a variável **Detectar** que será substituída por duas variáveis binárias **Detectar₂** (=1 provável, =0 caso contrário) e **Detecta₃** (=1 muito provável, =0 caso contrário).

A parte sistemática do modelo fica dado por

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_{11} \times \text{detectar}_{2i} + \beta_{12} \times \text{detectar}_{3i}$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_{21} \times \text{detectar}_{2i} + \beta_{22} \times \text{detectar}_{3i}.$$

Tabela 1.3

Classificação cruzada experiência versus detectar.

Experiência com mamografia	Detectar			Total
	1	2	3	
0	13	77	144	234
1	1	12	91	104
2	4	16	54	74
Total	18	105	289	412

As estimativas das razões de chances obtidas da Tabela 1.3 são dadas abaixo

$$\hat{\psi}_1(2, 1) = \frac{12 \times 13}{77 \times 1} = 2,03, \quad \hat{\psi}_1(3, 1) = \frac{91 \times 13}{144 \times 1} = 8,22,$$

$$\hat{\psi}_2(2, 1) = \frac{16 \times 13}{77 \times 4} = 0,68 \quad \text{e} \quad \hat{\psi}_2(3, 1) = \frac{54 \times 13}{144 \times 4} = 1,22.$$

Por exemplo, a chance de uma mulher ter feito o exame nos últimos 12 meses é 2,03 vezes a de nunca ter feito quando a mesma considera provável que um exame possa detectar o problema frente a mulheres que acham pouco provável a detecção. A chance cresce para 8,22 vezes se a mulher acreditar que seja muito provável que um mamograma

detecte o problema. Para os casos em que os exames foram realizados há mais de um ano as chances mudam pouco. Os resultados do ajuste do modelo estão resumidos na Tabela 1.4.

Tabela 1.4

Resultados do modelo de regressão logística ajustado à variável detectar.

Logito	Variável	Coefficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo	$\hat{\psi}_j$
1	Detectar ₂	0,706	1,083	0,425	0,515	2,02
	Detectar ₃	2,106	1,046	4,051	0,044	8,22
	Intercepto	-2,565	1,038	6,109	0,013	
2	Detectar ₂	-0,393	0,634	0,383	0,536	0,68
	Detectar ₃	0,198	0,594	0,111	0,739	1,22
	Intercepto	-1,179	0,572	4,249	0,039	

$$-2 \times \log(\text{verossimilhança}) = 778,401$$

As equações estimadas são dadas abaixo

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -2,565 + 0,706 \times \text{detectar}_{2i} + 2,106 \times \text{detectar}_{3i}$$

e

$$\log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -1,179 - 0,393 \times \text{detectar}_{2i} + 0,198 \times \text{detectar}_{3i}.$$

Usaremos a razão de verossimilhanças para avaliar a significância da variável *Detectar*. O valor da estatística é $\xi_{RV} = 805,198 - 778,401 = 26,797$ com 4 graus de liberdade, produzindo um nível descritivo menor que 0,001. Assim, podemos concluir que a crença da mulher na capacidade da mamografia detectar um novo caso de câncer de mama está associado com sua decisão de fazer um exame de mamografia.

Examinando-se os coeficientes estimados e suas estatística de Wald, notamos que a associação é mais forte quando comparamos mulheres que fizeram o exame nos últimos 12 meses com mulheres que nunca se submeteram ao exame, e quando comparando respostas pouco provável a muito provável. Todos os demais coeficientes são não significantes.

ESTRATÉGIA PARA SELEÇÃO DE MODELOS

Em princípio, as estratégias e métodos para seleção de modelos com variável resposta trinomial são os mesmos do caso binomial. Por exemplo, a metodologia para seleção de variáveis *stepwise* funciona bem no caso binomial, entretanto para o caso trinomial a analogia não é direta, uma vez que trabalhamos com duas funções simultaneamente. Para o nosso caso vamos utilizar a estatística de Wald e o teste da razão de verossimilhanças para a seleção do modelo. Um outro método de seleção que poderia ser estendido para o modelo logístico trinomial é o método de Akaike que no caso consistiria em encontrar o modelo com $2p$ parâmetros tal que

$$\text{AIC} = \frac{1}{2}D_{2p} + 2p$$

seja mínimo, em que D_{2p} denote o desvio do modelo ajustado.

Na análise dos dados do estudo de mamografia existem apenas cinco variáveis independentes e 412 observações (vide Apêndice A). A variável **Sintoma** está codificada em quatro níveis, no processo de construção do modelo vamos modelá-la em três variáveis binárias, usando o primeiro nível **concordo totalmente** como referência. A variável **Detectar** está codificada em três níveis, será modelada usando duas variáveis binárias, usando o primeiro nível **pouco provável** como referência. Apesar dessas duas variáveis serem ordinais é razoável codificá-las em variáveis binárias ao invés de tratá-las como contínuas, pois dessa forma os coeficientes para as variáveis binárias podem fornecer uma interpretação direta de cada nível da variável **Detectar**. As variáveis **Historico** e **Autoexame** são binárias, o primeiro nível será usado como referência nos dois casos. Inicialmente, vamos tratar a variável **Beneficio** como se fosse contínua e linear nos logitos. Os resultados para o ajuste considerando todas as variáveis (apenas efeitos principais) são apresentados na Tabela 1.5

Tabela 1.5

Resultados do modelo de regressão logística ajustado aos dados do Apêndice A.

Logito	Variável	Coefficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo
1	Sintoma ₂	0,1098	0,9226	0,0142	0,9052
	Sintoma ₃	1,9242	0,7774	6,1258	0,0133
	Sintoma ₄	2,4565	0,7752	10,0422	0,0015
	Beneficio	-0,2194	0,0755	8,4444	0,0037
	Historico	1,3662	0,4375	9,7512	0,0018
	Autoexame	1,2916	0,5299	5,9416	0,0148
	Detectar ₂	0,0161	1,1615	0,0002	0,9889
	Detectar ₃	0,9032	1,1264	0,6429	0,4226
	Intercepto	-2,9973	1,5389	3,7937	0,0514
2	Sintoma ₂	-0,2901	0,6441	0,2029	0,6524
	Sintoma ₃	0,8173	0,5398	2,2925	0,1300
	Sintoma ₄	1,1322	0,5477	4,2739	0,0387
	Beneficio	-0,1482	0,0764	3,7662	0,0523
	Historico	1,0654	0,4594	5,3786	0,0204
	Autoexame	1,0521	0,5150	4,1739	0,0411
	Detectar ₂	-0,9244	0,7138	1,6774	0,1953
	Detectar ₃	-0,6906	0,6871	1,0100	0,3149
	Intercepto	-0,9861	1,1118	0,7865	0,3751

$$-2 \times \log(\text{verossimilhança}) = 693,902$$

De acordo com a estatística de Wald, os dois coeficientes estimados para a variável binária **Sintoma₂**, que estima o logaritmo da razão de chances para resposta **concordo** versus o valor de referência **concordo totalmente**, sugere que essas duas categorias são similares. O sinal e magnitude dos coeficientes estimados para as variáveis **Sintoma₃** e **Sintoma₄** sugere que as respostas **discordo** e **discordo totalmente** diferem da referência **concordo totalmente** e são similares dentro de cada uma das duas funções logito. Este tipo de estrutura nos coeficientes estimados sugere que devemos dicotomizar **Sintoma** em apenas dois níveis, 0=**concordo totalmente** ou **concordo** e 1=**discordo** ou **discordo totalmente**, e em seguida reajustar o modelo. Essa nova variável dicotômica será rotulada como **SintomaD**. O novo ajuste considerando essa nova codificação é resumido na Tabela 1.6.

Tabela 1.6*Resultados do modelo de regressão logística ajustado aos dados do Apêndice A.*

Logito	Variável	Coefficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo
1	SintomaD	2,0950	0,4573	20,9842	0,0001
	Beneficio	-0,2510	0,0729	11,8454	0,0006
	Historico	1,2933	0,4335	8,8990	0,0029
	Autoexame	1,2439	0,5263	5,5863	0,0181
	Detectar ₂	0,0893	1,1606	0,0059	0,9386
	Detectar ₃	0,9719	1,1259	0,7451	0,3880
	Intercepto	-2,7024	1,4341	3,5510	0,0595
2	SintomaD	1,1214	0,3572	9,8552	0,0017
	Beneficio	-0,1681	0,0742	5,1366	0,0234
	Historico	1,0140	0,4538	4,9932	0,0254
	Autoexame	1,0286	0,5140	4,0049	0,0454
	Detectar ₂	-0,9022	0,7146	1,5936	0,2068
	Detectar ₃	-0,6698	0,6876	0,9490	0,3300
	Intercepto	-0,9987	1,0720	0,8680	0,3515

$-2 \times \log(\text{verossimilhança})=697,496$

Resta agora avaliar a contribuição ao modelo da variável **Detectar**. Na Tabela 1.6 observamos que nenhuma das estatísticas de Wald para as variáveis binárias **Detectar** são significativas marginalmente, porém acreditava-se que no total, a variável ainda poderia ser significativa e também uma **confundidora** para alguma outra variável. Para avaliar sua significância e seu papel como **confundidora** ajustamos o modelo excluindo a variável **Detectar**. O valor da estatística do teste da razão de verossimilhanças para a significância da variável **Detectar** é 8,54 com 4 graus de liberdade, produzindo um nível descritivo de 0,074. Nota-se Tabela 1.7 nas variáveis restantes no modelo uma sensível mudança no coeficiente estimado da variável **SintomaD** no primeiro logito. A mudança foi de cerca de 8% no valor do coeficiente estimado, indicando que a variável **Detectar** exerce um papel de **confundidora** da associação da variável **SintomaD** com a variável resposta. Devido a esse papel e sua proximidade à significância estatística, decidimos mantê-la no modelo.

As estimativas para os modelos sem a variável *Detectar* estão resumidas na Tabela 1.7.

Tabela 1.7

Resultado do modelo de regressão excluindo a variável Detectar.

Logito	Variável	Coefficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo
1	SintomaD	2,2302	0,4519	24,3538	0,0001
	Beneficio	-0,2825	0,0713	15,6819	0,0001
	Historico	1,2966	0,4293	9,1223	0,0025
	Autoexame	1,2209	0,5210	5,4910	0,0191
	Intercepto	-1,7885	0,8470	4,4584	0,0347
2	SintomaD	1,1531	0,3514	10,7697	0,0010
	Beneficio	-0,1578	0,0712	4,9144	0,0266
	Historico	1,0613	0,4527	5,4969	0,0191
	Autoexame	0,9604	0,5072	3,5853	0,0583
	Intercepto	-1,7421	0,8087	4,6410	0,0312

$$-2 \times \log(\text{verossimilhança}) = 706,038$$

O próximo passo no desenvolvimento do modelo é avaliar a necessidade de incluir os termos de interação de 1ª ordem no modelo. Neste modelo cada par de variável cria uma interação biologicamente plausível.

Nenhuma das 10 possíveis interações contribuíram significativamente para o modelo. Assim, as equações estimadas para o modelo selecionado, são as seguintes:

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -2,7024 + 2,0950 \times \text{SintomaD}_i - 0,2510 \times \text{Beneficio}_i + \\ 1,2933 \times \text{Historico}_i + 1,2439 \times \text{Autoexame}_i + 0,0893 \times \text{Detectar}_{2i} + \\ 0,9719 \times \text{Detectar}_{3i}$$

e

$$\log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -0,9987 + 1,1214 \times \text{SintomaD}_i - 0,1681 \times \text{Beneficio}_i + \\ 1,0140 \times \text{Historico}_i + 1,0286 \times \text{Autoexame}_i - 0,9022 \times \text{Detectar}_{2i} - \\ 0,6698 \times \text{Detectar}_{3i}.$$

1.9.2 Exemplo 2

Este estudo foi realizado (vide Forthofer and Lehnen, 1981 e Apêndice B) para investigar a associação entre resultado no teste respiratório, hábito de fumar e idade. A amostra consistiu de 2219 trabalhadores da indústria de Houston, Texas, durante os anos de 1974 e 1975. Neste estudo são consideradas as seguintes variáveis.

Variável resposta

Resultado: resultado obtido no teste respiratório:

- 1: normal (referência),
- 2: indefinido e
- 3: anormal.

Variáveis explicativas

Status: condição atual do pesquisado:

- 1: nunca fumou (referência),
- 2: ex-fumante e
- 3: fumante.

Idade: idade do pesquisado:

- 1: menos de 40 anos e
- 2: 40 a 59 anos.

MODELAGEM

Denote por Y_i o resultado do i -ésimo exame respiratório ($i = 1, 2, \dots, n$), em que $n = 2219$. Novamente assumiremos que Y_i segue distribuição trinomial ($Y_i=1$ normal, $Y_i=2$ indefinido e $Y_i=3$ anormal) com probabilidades dadas por π_{0i} , π_{1i} e π_{2i} , respectivamente.

Vamos iniciar nossa modelagem considerando primeiramente a variável explicativa **idade**. A parte sistemática do modelo fica dada por

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_1 \times \text{idade}_i$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_2 \times \text{idade}_i.$$

Os coeficientes e razões de chances estimadas estão dadas na Tabela 1.9

Tabela 1.9

Resultados do modelo de regressão logística ajustado à variável idade.

Logito	Variável	Coeficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo	$\hat{\psi}_j$
1	Idade	0,6199	0,1685	13,5315	0,0002	1,86
	Intercepto	-3,3673	0,2504	180,7722	0,0000	
2	Idade	1,4447	0,2820	26,2446	0,0000	4,24
	Intercepto	-5,6802	0,4738	143,7485	0,0000	

$$-2 \times \text{Log}(\text{Verossimilhança}) = 1.613,298$$

As razões de chances acima informam que a chance de um indivíduo com idade entre 40 e 59 anos ter o resultado de seu teste respiratório anormal é 4,2 vezes a de um resultado normal. Além disso, a chance de ter o resultado indefinido para o seu teste respiratório é 1,9 vezes a de um resultado normal.

A razão de chances para $Y = 3$ versus $Y = 2$ pode ser obtida da razão entre as duas razões de chances estimadas no modelo de regressão logística correspondente. Da Tabela 1.9 obtemos $\hat{\psi}_3 = \hat{\psi}_2 / \hat{\psi}_1 = 4,241 / 1,859 = 2,28$. Mostrando que ao compararmos pessoas com resultados não normais no teste respiratório, a chance de ter resultado anormal é maior (2,28 vezes) do que ter um resultado indefinido quando o pesquisado faz parte do grupo na faixa etária de 40 a 59 anos.

Para uma indicação preliminar da importância da variável, vamos avaliar a significância dos coeficientes. O teste da razão de verossimilhanças nos fornece um nível descritivo

menor que 0,0001, dando indicação de que esta variável é significativa quando avaliada isoladamente.

Vamos agora considerar a variável explicativa **status**. A parte sistemática do modelo fica dada por

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_1 \times \mathbf{status}_{2_i} + \beta_2 \times \mathbf{status}_{3_i}$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_1 \times \mathbf{status}_{2_i} + \beta_2 \times \mathbf{status}_{3_i}.$$

A Tabela 1.10 exibe os coeficientes e razões de chances estimadas

Tabela 1.10

Resultados do modelo de regressão logística ajustado à variável status.

Logito	Variável	Coeficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo	$\hat{\psi}$
1	status ₂	0,9093	0,2552	12,6912	0,0004	2,48
	status ₃	0,8747	0,2132	16,8364	0,0001	2,40
	Intercepto	-3,1740	0,1833	299,7644	0,0001	
2	status ₂	1,1446	0,4972	5,3003	0,0213	3,14
	status ₃	1,4677	0,4142	12,5536	0,0004	4,34
	Intercepto	-4,6621	0,3797	150,7219	0,0001	

$$-2 \times \text{Log}(\text{Verossimilhança}) = 1.614,763$$

Analisando as razões de chances podemos afirmar que a chance de um indivíduo ter o resultado de seu teste respiratório anormal é 4,3 vezes a de um resultado normal se o mesmo for fumante. Esta chance decresce para 3,1 se o indivíduo for ex-fumante. Avaliando a chance de ter o resultado do teste indefinido contra normal, podemos afirmar que é praticamente indiferente para um ex-fumante ou fumante, sendo que a chance é aproximadamente 150% maior. Da Tabela 1.10 obtemos:

$$\hat{\psi}_5 = \hat{\psi}_{22}/\hat{\psi}_{12} = 3,141/2,483 = 1,27 \quad \text{e} \quad \hat{\psi}_6 = \hat{\psi}_{23}/\hat{\psi}_{13} = 4,339/2,398 = 1,81.$$

As estatísticas acima mostram que ao compararmos pessoas com resultados não normais no teste respiratório verifica-se que, a chance do resultado ser anormal é 81% maior se o indivíduo é fumante e 27% maior se é um ex-fumante.

Para verificar a importância da variável, vamos avaliar a significância dos coeficientes. O teste da razão de verossimilhanças nos fornece um nível descritivo menor que 0,0001, indicando que a variável **status** está associada ao resultado no teste respiratório.

Consideramos agora as duas variáveis explicativas conjuntamente, usando apenas os efeitos principais, uma vez que a interação **idade** \times **status** mostrou-se não significativa. A parte sistemática do modelo fica dada por

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \beta_{10} + \beta_1 \times \text{idade}_i + \beta_2 \times \text{status}_{2i} + \beta_3 \times \text{status}_{3i}$$

e

$$\log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \beta_{20} + \beta_1 \times \text{idade}_i + \beta_2 \times \text{status}_{2i} + \beta_3 \times \text{status}_{3i}.$$

As estimativas para o modelo acima são apresentados na Tabela 1.11.

Tabela 1.11

Resultados do modelo de regressão logística ajustado aos dados do apêndice B.

Logito	Variável	Coeficiente Estimado	Erro padrão Estimado	Estatística de Wald	Nível Descritivo
1	Idade	0,5380	0,1713	9,8597	0,0017
	status ₂	0,7881	0,2588	9,2723	0,0023
	status ₃	0,8319	0,2140	15,1194	0,0001
	Intercepto	-3,8508	0,2894	177,0264	0,0001
2	Idade	1,3786	0,2855	23,3130	0,0001
	status ₂	0,8221	0,5031	2,6702	0,1022
	status ₃	1,3463	0,4164	10,4549	0,0012
	Intercepto	-6,5231	0,5745	128,9185	0,0001

$$-2 \times \text{Log}(\text{Verossimilhança}) = 1.582,751$$

Ao considerar as duas variáveis explicativas temos que o teste de razão de verossimilhanças produz um valor de 68,8545 com 6 graus de liberdade, o que fornece um nível descritivo

menor que 0,0001. Portanto, podemos afirmar que essas duas variáveis são altamente significativas, ou seja, tanto idade como status de fumo estão associadas ao resultado de um teste respiratório.

Assim, os sub-modelos ajustados ficam dados por:

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -3,8508 + 0,5380 \times \text{idade}_i + 0,7881 \times \text{status}_{2_i} + 0,8319 \times \text{status}_{3_i}$$

e

$$\log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -6,5231 + 1,3786 \times \text{idade}_i + 0,8221 \times \text{status}_{2_i} + 1,3463 \times \text{status}_{3_i}.$$

Nota-se pelos coeficientes estimados acima que a variável binária **idade** é mais importante nas estimativas da probabilidade de π_2 de resultado anormal do que na estimativa da probabilidade π_1 de resultado indiferente. O mesmo ocorre para a variável binária **status₃** enquanto a outra variável binária **status₂** tem a mesma importância na estimativa das duas probabilidades.

Capítulo 2

Métodos de Diagnóstico

2.1 Introdução

Uma etapa importante na análise de um ajuste de regressão é a verificação de possíveis afastamentos das suposições feitas para o modelo, especialmente para a parte aleatória e para a parte sistemática do modelo, bem como a existência de observações extremas com alguma interferência desproporcional nos resultados do ajuste. Tal etapa, conhecida como análise de diagnóstico, tem longa data, e iniciou-se com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta. Uma referência importante nesse tópico é o artigo de Cox e Snell (1968) em que é apresentada uma forma bastante geral de definir resíduos, usada até os dias atuais.

Belsley, Kuh e Welsch (1980) e Cook e Weisberg (1982) discutem a padronização de resíduos para o caso normal linear. Pregibon (1981) propõe o componente do desvio como resíduo na classe dos modelos lineares generalizados (MLGs) e sugere uma padronização que é mais tarde comprovada matematicamente por McCullagh (1987) que usa as aproximações propostas por Cox e Snell (1968). Nesse mesmo trabalho McCullagh apresenta uma outra forma de padronização para o componente do desvio em que procura corrigir os efeitos de assimetria e curtose. Atkinson (1981) propõe a construção por simulação

de Monte Carlo de uma banda de confiança para os resíduos de regressão normal linear, a qual denominou envelopes, e que permite uma melhor comparação entre os resíduos e os percentis da distribuição normal padrão. Williams (1984, 1987) discute, com base em estudos de simulação de Monte Carlo, a aproximação da forma padronizada proposta por Pregibon (1981) encontrando fortes evidências de concordância entre a distribuição empírica do componente do desvio padronizado e a distribuição normal padrão para vários MLGs. Williams (1987) também discute a construção de envelopes em MLGs. Davison e Gigli (1989) estendem a proposta de Cox e Snell (1968) e definem uma forma geral de padronização para o componente do desvio para distribuições contínuas, mesmo quando a função de distribuição acumulada não é expressa em forma fechada. Fahrmeir e Tutz (1994) estendem o trabalho de McCullagh (1987) para modelos mais gerais. Paula (1995) apresenta uma forma padronizada para o componente do desvio em MLGs com parâmetros restritos na forma de desigualdades lineares $\mathbf{C}\boldsymbol{\beta} \geq 0$ e verifica através de simulação, forte concordância na maioria dos modelos estudados entre a distribuição empírica do resíduo padronizado e a distribuição normal padrão, generalizando os resultados de Williams para parâmetros restritos. De Souza e Paula (2002) usam o método proposto por Davison e Gigli (1989) a fim de obterem uma forma padronizada para o componente do desvio em modelos de regressão von Mises, os quais têm sido aplicados na análise de dados circulares. A construção de envelopes com o resíduo proposto é também discutida no trabalho.

Neste capítulo serão apresentadas algumas técnicas usuais de diagnóstico que têm sido largamente aplicadas no modelo logístico binomial e que podem ser diretamente estendidas para o modelo logístico trinomial.

2.2 Pontos de Alavanca

Um tópico importante de análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Durante a década de 70 surgiram várias propostas relacionadas com a influência

das observações nas estimativas dos coeficientes do modelo normal linear. O estudo da diagonal principal da matriz de projeção $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ apresentada por Hoaglin e Welsch (1978), em que \mathbf{X} denota a matriz modelo, motivou a definição de pontos de alavanca que receberam este nome por terem peso desproporcional no próprio valor ajustado. Esses pontos em geral são remotos no subespaço gerado pelas colunas da matriz \mathbf{X} , ou seja, têm um perfil diferente dos demais pontos no que tange aos valores das variáveis explicativas. Dependendo da localização, tais pontos podem exercer forte influência nas estimativas dos coeficientes de regressão. Extensões da definição de pontos de alavanca para modelos normais não-lineares são dadas em St. Laurent e Cook (1992). 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 enfoque Bayesiano. Paula (1999) discute pontos de alavanca em modelos de regressão com parâmetros restritos na forma $\mathbf{C}\boldsymbol{\beta} \geq 0$, com extensões para os MLGs. Em modelos de regressão em que a resposta é contínua pontos de alavanca são aqueles cuja derivada $\frac{\partial \hat{y}_i}{\partial y_i}$ assume um valor desproporcional, isto é, uma perturbação infinitesimal em y_i causa uma grande variação em \hat{y}_i . Expressões gerais para essa derivada podem ser encontradas em Wei, Hu e Fung (1998). Todavia, quando a resposta é discreta, como é o caso da distribuição trinomial, não faz sentido o cálculo da derivada. Nesses casos, outros conceitos de pontos de alavanca são utilizados. Em particular, no caso da regressão logística binomial usa-se os elementos da diagonal principal da matriz de projeção correspondente ao modelo linear com solução de mínimos quadrados dada por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}\mathbf{z}, \quad (2.1)$$

em que $\hat{\mathbf{V}} = \text{diag}\{n_1\pi_1(1 - \pi_1), \dots, n_k\pi_k(1 - \pi_k)\}$, \mathbf{X} é a matriz modelo, $\mathbf{z}=(z_1, \dots, z_k)^T$ com $z_i = \hat{n}_1 + (y_i - n_i\hat{\pi}_i)/n_i\hat{\pi}_i(1 - \hat{\pi}_i)$. A estimativa (2.1) coincide com a solução de

máxima verossimilhança do modelo logístico binomial com parte sistemática dada por

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Essa matriz de projeção assume a forma

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

cujo i -ésimo elemento da diagonal principal fica dado por

$$h_{ii} = n_i \pi_i (1 - \pi_i) \mathbf{x}_i^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{x}_i.$$

Logo, \hat{h}_{ii} que detecta pontos de alavanca no modelo linear correspondente é sugerido também para o modelo logístico binomial, embora não tenha a mesma interpretação nesse modelo. Como a função $\pi_i(1 - \pi_i)$ é quadrática alguns cuidados devem ser tomados no uso de h_{ii} . Hosmer e Lemeshow (1989, Cap. 5) mostram através de estudos numéricos que h_{ii} deve ser pequeno quando π_i está próximo de zero ou um, deve ser alto para valores de π_i entre 0,1 e 0,3 ou entre 0,7 e 0,9 e deve ser moderado ou pequeno para valores de π_i entre 0,3 e 0,7. Assim, observações com valores para \hat{h}_{ii} maiores do que essas tendências esperadas devem ser olhados com mais atenção. O gráfico de \hat{h}_{ii} contra $\hat{\pi}_i$ deve ser usado por ser mais informativo do que o gráfico de \hat{h}_{ii} contra a ordem das observações.

No caso do modelo logístico trinomial tem-se dois conjuntos de valores para \hat{h}_{ii} , os quais serão denotados por $\hat{h}_{ii}^{(1)}$ e $\hat{h}_{ii}^{(2)}$. O primeiro desses conjuntos será formado pelos elementos da diagonal principal de

$$\hat{\mathbf{H}}^{(1)} = \hat{\mathbf{V}}_1^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{V}}_1 \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}_1^{1/2},$$

e assume a forma

$$\hat{h}_{ii}^{(1)} = n_{1i} \hat{\pi}_{1i} (1 - \hat{\pi}_{1i}) \mathbf{x}_i^T (\mathbf{X}^T \hat{\mathbf{V}}_1 \mathbf{X})^{-1} \mathbf{x}_i,$$

em que

$$\mathbf{V}_1 = \text{diag}\{n_{1i} \pi_{1i} (1 - \pi_{1i}), \dots, n_{ki} \pi_{ki} (1 - \pi_{ki})\}.$$

O outro conjunto deve ser formado pelos elementos

$$\hat{h}_{ii}^{(2)} = n_{2i}\hat{\pi}_{2i}(1 - \hat{\pi}_{2i})\mathbf{x}_i^T(\mathbf{X}^T\hat{\mathbf{V}}_1\mathbf{X})^{-1}\mathbf{x}_i$$

da diagonal principal da matriz

$$\hat{\mathbf{H}}^{(2)} = \hat{\mathbf{V}}_2^{1/2}\mathbf{X}(\mathbf{X}^T\hat{\mathbf{V}}_2\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}_2^{1/2}.$$

Gráficos de $\hat{h}_{ii}^{(1)}$ e $\hat{h}_{ii}^{(2)}$ contra os valores ajustados $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente, devem ser construídos. Aqueles pontos que se destacarem nos gráficos devem ser olhados com mais atenção. Note que a mesma observação pode ser detectada nos dois gráficos.

Podemos para o modelo logístico trinomial fazer uma analogia com o modelo logístico binomial e tentar encontrar uma matriz de projeção com base na expressão da convergência do processo iterativo para $\hat{\boldsymbol{\beta}}$. Note que na convergência, supondo o caso de réplicas, tem-se o seguinte:

$$\hat{\boldsymbol{\beta}} = (\mathbf{S}^T\hat{\mathbf{V}}\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{V}}\boldsymbol{\xi},$$

em que

$$\boldsymbol{\xi} = \mathbf{S}\hat{\boldsymbol{\beta}} + \hat{\mathbf{V}}^{-1}(\mathbf{a} - \boldsymbol{\mu}).$$

Assim, podemos propor a matriz de projeção

$$\mathbf{M} = \mathbf{S}(\mathbf{S}^T\mathbf{V}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{V}.$$

Note que essa matriz \mathbf{M} tem dimensão $(2n \times 2n)$ o que significa dizer que teremos para cada observação duas medidas de alavanca, uma para cada sub-modelo logístico. Para a i -ésima observação teremos as seguintes medidas:

$$\hat{m}_{ii}^{(1)} = \mathbf{s}_{0i}^T(\mathbf{S}^T\hat{\mathbf{V}}\mathbf{S})^{-1}\hat{\mathbf{V}}\mathbf{s}_{1i} \quad \text{e} \quad \hat{m}_{ii}^{(2)} = \mathbf{s}_{0i}^T(\mathbf{S}^T\hat{\mathbf{V}}\mathbf{S})^{-1}\hat{\mathbf{V}}\mathbf{s}_{2i}$$

em que

$$\mathbf{s}_{0i}^T = [\mathbf{x}_i^T, \mathbf{x}_i^T], \quad \mathbf{s}_{1i}^T = [\mathbf{x}_i^T\hat{\pi}_{1i}(1 - \hat{\pi}_{1i}), -\mathbf{x}_i^T\hat{\pi}_{1i}\hat{\pi}_{2i}] \quad \text{e} \quad \mathbf{s}_{2i}^T = [-\mathbf{x}_i^T\hat{\pi}_{1i}\pi_{2i}, \mathbf{x}_i^T\hat{\pi}_{2i}(1 - \hat{\pi}_{2i})].$$

Obviamente que os gráficos de $\hat{m}_{ii}^{(1)}$ e $\hat{m}_{ii}^{(2)}$ contra os valores ajustados $\hat{\pi}_i$ podem revelar observações influentes nas estimativas $\hat{\beta}_1$ e $\hat{\beta}_2$, respectivamente. Contudo, como pouco se conhece a respeito do comportamento dessas medidas em relação aos valores ajustados, fica difícil estabelecer critérios para usá-las. É muito razoável esperar que $\hat{m}_{ii}^{(1)}$ e $\hat{m}_{ii}^{(2)}$ sejam mais informativas que $\hat{h}_{ii}^{(1)}$ e $\hat{h}_{ii}^{(2)}$ uma vez que esses últimos são construídos com base nos modelos logísticos marginais enquanto os dois primeiros são construídos para o modelo logístico trinomial.

2.3 Deleção de Pontos

A deleção de pontos talvez seja a técnica mais conhecida para avaliar o impacto da retirada de uma observação particular nas estimativas da regressão. A distância de Cook (1977), originalmente desenvolvida para modelos normais lineares, foi rapidamente assimilada e estendida para diversas classes de modelos. Por exemplo, Moolgavkar, Lustbader e Verzon (1984) estendem a metodologia para regressão não-linear com aplicações em estudos emparelhados, Ross (1987) discute a geometria da deleção de casos em regressão não-linear, Cook, Peña e Weisberg (1988) comparam o afastamento da verossimilhança com medidas tradicionais de deleção de pontos tais como a distância de Cook e o $DFFITS_i$, esse último proposto por Belsley, Kuh e Welsch (1980) e Paula e Peres (1988) discutem a deleção de pontos em MLGs com parâmetros restritos na forma $\mathbf{C}\beta \geq 0$. Davison e Tsai (1992) e Cordeiro e Paula (1992) estendem a metodologia para modelos cuja distribuição não pertence à família exponencial de distribuições. Recentemente, Galea, Riquelme e Paula (2000) investigam a metodologia em modelos elípticos multivariados. Referências importantes nesse tópico são, dentre outras, os livros de Cook e Weisberg (1982), Atkinson (1985) e Chatterjee e Hadi (1988). Um problema que pode ocorrer na deleção individual de pontos é o que se denomina *masking effect* ou seja, deixar de detectar pontos conjuntamente discrepantes. Embora esse procedimento de deleção múltipla de pontos não seja muito popular, provavelmente em virtude do custo computacional envolvido, existem

vários procedimentos robustos para detecção de pontos discrepantes, muitos dos quais com um custo computacional relativamente baixo (ver, por exemplo, Fung, 1993; Peña e Yohai, 1999). Como em geral esses procedimentos têm sido desenvolvidos para modelos lineares, abre-se uma perspectiva de pesquisas em classes mais abrangentes.

Uma das medidas mais populares de deleção de pontos é conhecida como afastamento da verossimilhança (*likelihood displacement*) definida abaixo

$$LD_i = 2\{L(\hat{\boldsymbol{\beta}}) - L(\hat{\boldsymbol{\beta}}_{(i)})\},$$

em que $\hat{\boldsymbol{\beta}}_{(i)}$ denota a estimativa de máxima verossimilhança de $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ sem a i -ésima observação. Uma aproximação quadrática por série de Taylor leva à seguinte aproximação:

$$LD_i = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T \mathbf{K}(\hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}).$$

A fim de obter-se uma forma fechada para LD_i é usual considerar alguma aproximação para $\hat{\boldsymbol{\beta}}_{(i)}$. A aproximação de um passo proposta por Pregibon (1981) no caso de regressão logística binomial poderia ser estendida para o caso trinomial. Assim, a medida de diagnóstico fica dada por

$$LD_i = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}^1)^T \mathbf{K}(\hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}^1),$$

em que $\hat{\boldsymbol{\beta}}_{(i)}^1$ denota a aproximação de um passo do processo iterativo para estimar $\boldsymbol{\beta}$ sem a i -ésima observação e iniciando em $\hat{\boldsymbol{\beta}}$.

Uma outra possibilidade mais simples seria considerar a aproximação para o afastamento da verossimilhança para cada um dos sub-modelos logísticos que compõe o modelo logístico trinomial. Assim, teremos dois conjuntos de medidas de influência

$$LD_i^{(1)} = (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_{1(i)}^1)^T \mathbf{K}_1(\hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_{1(i)}^1) \quad (2.2)$$

e

$$\text{LD}_i^{(2)} = (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_{2(i)}^1)^T \mathbf{K}_2(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_{2(i)}^1), \quad (2.3)$$

em que

$$\mathbf{K}_1(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T \mathbf{V}_1 \mathbf{X} \quad \text{e} \quad \mathbf{K}_2(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T \mathbf{V}_2 \mathbf{X}.$$

A aproximação de um passo proposta por Pregibon (1981) para o modelo logístico binomial com parte sistemática

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta},$$

fica dada por

$$\boldsymbol{\beta}_{(i)}^1 = \hat{\boldsymbol{\beta}} - \left\{ \frac{y_i - n_i \hat{\pi}_i}{1 - \hat{h}_{ii}} \right\} (\mathbf{X}^T \hat{\mathbf{V}} \mathbf{X})^{-1} \mathbf{x}_i, \quad (2.4)$$

em que $\mathbf{V} = \text{diag}\{n_1 \pi_1 (1 - \pi_1), \dots, n_k \pi_k (1 - \pi_k)\}$. Usando (2.4) nas expressões (2.2) e (2.3) chega-se às medidas de influência

$$\text{LD}_i^{(1)} = \left\{ \frac{\hat{h}_{ii}^{(1)}}{1 - \hat{h}_{ii}^{(1)}} \right\} \{t_{S_i}^{(1)}\}^2 \quad \text{e} \quad \text{LD}_i^{(2)} = \left\{ \frac{\hat{h}_{ii}^{(2)}}{1 - \hat{h}_{ii}^{(2)}} \right\} \{t_{S_i}^{(2)}\}^2,$$

em que

$$t_{S_i}^{(1)} = \frac{(a_{1i} - n_{1i} \hat{\pi}_{1i})}{\sqrt{n_{1i} \hat{\pi}_{1i} (1 - \hat{\pi}_{1i}) (1 - \hat{h}_{ii}^{(1)})}} \quad \text{e} \quad t_{S_i}^{(2)} = \frac{(a_{2i} - n_{2i} \hat{\pi}_{2i})}{\sqrt{n_{2i} \hat{\pi}_{2i} (1 - \hat{\pi}_{2i}) (1 - \hat{h}_{ii}^{(2)})}},$$

são os respectivos resíduos de Pearson padronizados, em que a_{1i} e a_{2i} representam o número de sucessos nos níveis 1 e 2, respectivamente, do i -ésimo grupo (vide seção 1.6).

Assim, são sugeridos dois conjuntos de gráficos de diagnóstico. O gráfico de $\text{LD}_i^{(1)}$ contra $\hat{\pi}_{1i}$ e o gráfico de $\text{LD}_i^{(2)}$ contra $\hat{\pi}_{2i}$. Aquelas observações que destoarem com valores muito acima dos demais devem ser retiradas para avaliação do impacto das mesmas nas estimativas. Se a variação for muito desproporcional pode-se dizer que esses pontos são influentes em determinadas estimativas. Porém, deve-se verificar também se a eliminação desses pontos leva a mudanças inferenciais, por exemplo, deixar algum coeficiente não

significativo. A eliminação definitiva de um ponto somente deve ser feita em último caso após várias tentativas de acomodá-lo no ajuste através, por exemplo, de transformações de variáveis explicativas, inclusão ou eliminação de variáveis explicativas ou interação, ou mesmo mudanças no modelo.

2.4 Influência Local

Uma das propostas mais inovadoras na área de diagnóstico em regressão foi apresentada por Cook (1986) que propõe avaliar a influência conjunta das observações sob pequenas mudanças (perturbações) no modelo, ao invés da avaliação pela retirada individual ou conjunta de pontos. Essa metodologia, denominada influência local, teve uma grande receptividade entre os usuários e pesquisadores de regressão, havendo inúmeras publicações no assunto em que se aplica a metodologia em classes particulares de modelos ou em que se propõe extensões da técnica. Por exemplo, na classe de erros normais, Lawrence (1988) investiga a aplicação de influência local em modelos lineares com parâmetros na transformação da resposta, Beckman, Nachtsheim e Cook (1987) apresentam estudos de influência em modelos de análise de variância com efeito misto, Tsai e Wu (1992) investigam influência local em modelos auto-regressivos de 1ª ordem e modelos heterocedásticos e Paula (1993) aplica influência local em modelos lineares com restrições nos parâmetros na forma de desigualdades lineares. Saindo da classe de erros normais tem-se, por exemplo, o trabalho de Pettitt e Bin Daud (1989) que investigam influência local em modelos de Cox com riscos proporcionais, Escobar e Meeker (1992) adaptam influência local numa classe paramétrica de modelos para análise de sobrevivência, O'Hara Hines, Lawless e Carter (1992), Kim (1995) e Pan, Fang e von Resen (1997) aplicam métodos de influência local em regressão multivariada. Mais recentemente, Galea, Paula e Bolfarine (1997), Liu (2000) e Galea, Paula e Uribe-Opazo (2003) apresentam estudos de

influência local em modelos elípticos lineares, enquanto Kwan e Fung (1998) aplicam a metodologia em análise fatorial, Gu e Fung (1998) em análise de correlação canônica

e Paula (1996) em modelos próprios de dispersão. Svetliza e Paula (2003) discutem influência local em modelos de regressão com resposta binomial negativa. Esses últimos modelos têm sido muito usados para corrigir problemas de superdispersão, freqüentemente encontrados em modelos com resposta de Poisson. Ortega, Bolfarine e Paula (2003) aplicam influência local em modelos log-gama generalizados com dados censurados. Uma discussão interessante a respeito do uso de influência local é apresentada por Fung e Kwan (1997). Os autores mostram que o afastamento da verossimilhança neste caso é uma medida de influência invariante com mudanças de escala nos dados, fato que não ocorre com outras medidas de influência propostas.

Para formalizar o método de influência local nos modelos logísticos trinomiais considere $L(\boldsymbol{\beta})$ como sendo o logaritmo da função de verossimilhança do modelo postulado, em que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, com $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ denotando os parâmetros da parte sistemática dos dois sub-modelos logísticos formados. Seja $\boldsymbol{\omega}$ um vetor $q \times 1$ de perturbações, restritas a um conjunto aberto $\Omega \subset \Re^q$. As perturbações são feitas no logaritmo da verossimilhança de modo que a mesma tome a forma $L(\boldsymbol{\beta} | \boldsymbol{\omega})$. Denotando o vetor de não perturbação por $\boldsymbol{\omega}_0$, assume-se que $L(\boldsymbol{\beta} | \boldsymbol{\omega}_0) = L(\boldsymbol{\beta})$. A fim de verificar-se a influência das perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$, considera-se o afastamento da verossimilhança

$$LD(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\beta}}) - L(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})\},$$

em que $\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$ denota a estimativa de máxima verossimilhança sob o modelo $L(\boldsymbol{\beta} | \boldsymbol{\omega})$.

A idéia de influência local consiste basicamente em estudar o comportamento da função $LD(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0 = 0$. O procedimento procura selecionar uma direção unitária $\boldsymbol{\ell}$, $\|\boldsymbol{\ell}\| = 1$, e então estudar o gráfico de $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ contra a , em que $a \in \Re$. Esse gráfico é conhecido como linha projetada. Note que, desde que $LD(\boldsymbol{\omega}_0) = 0$, então $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ tem um mínimo local em $a = 0$. Cada linha projetada pode ser caracterizada por uma curvatura normal $C_{\boldsymbol{\ell}}(\boldsymbol{\beta})$ em torno de $a = 0$. Essa curvatura é interpretada como sendo o inverso do raio de melhor círculo ajustado em $a = 0$. A sugestão é considerar a direção

ℓ_{max} que corresponde à maior curvatura $C_{\ell_{max}}(\boldsymbol{\beta})$. O gráfico de ℓ_{max} contra a ordem das observações pode revelar quais observações que sob pequenas perturbações exercem uma influência desproporcional em $LD(\boldsymbol{\omega})$. Cook (1986) mostra que a curvatura normal na direção $\boldsymbol{\ell}$ toma a forma $C_{\ell}(\boldsymbol{\beta}) = 2 | \boldsymbol{\ell}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \boldsymbol{\ell} |$, em que $-\ddot{\mathbf{L}}$ é a matriz de informação observada de Fisher enquanto que $\boldsymbol{\Delta}$ é uma matriz $2p \times q$ definida no modelo logístico trinomial por

$$\boldsymbol{\Delta} = \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \omega^T} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2 \partial \omega^T} \end{bmatrix},$$

ambas avaliadas em $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $i = 1, \dots, p$ e $j = 1, \dots, q$. Portanto, $C_{\ell_{max}}$ é o maior autovalor da matriz $\mathbf{B} = -\boldsymbol{\Delta}^T (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta}$ e ℓ_{max} denota o autovetor correspondente.

O gráfico de ℓ_{max} contra a ordem das observações pode revelar quanto deve ser perturbado no modelo de modo a se obter uma variação desproporcional na estimativa $\hat{\boldsymbol{\beta}}$. Em particular se o interesse é apenas com relação ao subvetor $\boldsymbol{\beta}_1$, a curvatura normal na direção $\boldsymbol{\ell}$ fica dada por $C_{\ell}(\boldsymbol{\beta}_1) = 2 | \boldsymbol{\ell}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_1) \boldsymbol{\Delta} \boldsymbol{\ell} |$, sendo

$$\mathbf{B}_1 = \begin{bmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22} \end{bmatrix},$$

com $-\ddot{\mathbf{L}}_{22} = \mathbf{X}^T \mathbf{V}_2 \mathbf{X}$ sendo a matriz de informação de Fisher para $\boldsymbol{\beta}_2$. O gráfico do maior autovetor de $-\boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_1) \boldsymbol{\Delta}$ contra a ordem das observações pode revelar os pontos com maior influência em $\hat{\boldsymbol{\beta}}_1$. De forma similar, se o interesse é apenas com respeito ao sub-vetor $\boldsymbol{\beta}_2$, a curvatura na direção $\boldsymbol{\ell}$ fica dada por $C_{\ell}(\boldsymbol{\beta}_2) = 2 | \boldsymbol{\ell}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_2) \boldsymbol{\Delta} \boldsymbol{\ell} |$, em que

$$\mathbf{B}_2 = \begin{bmatrix} \ddot{\mathbf{L}}_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

com $-\ddot{\mathbf{L}}_{11} = \mathbf{X}^T \mathbf{V}_1 \mathbf{X}$ sendo a matriz de informação de Fisher para $\boldsymbol{\beta}_1$. A influência local das observações em $\boldsymbol{\beta}_2$ pode ser avaliada pelo gráfico de ℓ_{max} correspondente ao maior autovalor de $-\boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_1) \boldsymbol{\Delta}$ contra a ordem das observações.

Consideraremos aqui o logaritmo da função de verossimilhança perturbada como sendo

$$L(\boldsymbol{\beta} \mid \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L(z_i; \pi_i),$$

em que $0 \leq \omega_i \leq 1$ e $L(z_i; \pi_i)$ denota o logaritmo da função de verossimilhança correspondente à i -ésima observação. Note que $\omega_i = 0$ significa que o i -ésimo ponto foi eliminado e $\boldsymbol{\omega}_0 = (1, 1, \dots, 1)^T$. Mostra-se facilmente para esse tipo de perturbação que $\boldsymbol{\Delta}$ é uma matriz $2p \times n$ dada por

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{X}^T \text{diag}(\mathbf{z}_1 - \hat{\boldsymbol{\pi}}_1) \\ \mathbf{X}^T \text{diag}(\mathbf{z}_2 - \hat{\boldsymbol{\pi}}_2) \end{bmatrix}.$$

Logo, deveremos diagonalizar a matriz $\mathbf{B} = \boldsymbol{\Delta}^T \mathbf{K}^{-1}(\hat{\boldsymbol{\beta}}) \boldsymbol{\Delta}$, em que $\mathbf{K}(\boldsymbol{\beta})$ é dada em (1.8). Portanto, teremos que encontrar o maior autovalor da matriz \mathbf{B} e o correspondente autovetor unitário que será ℓ_{max} . O gráfico de ℓ_{max} contra a ordem das observações poderá revelar aqueles pontos com maior influência local em $LD(\boldsymbol{\omega})$ e conseqüentemente em $\hat{\boldsymbol{\beta}}$.

Similarmente às seções anteriores poderemos fazer uma analogia com o modelo logístico binomial considerando as expressões das curvaturas desse modelo para cada um dos submodelos que compõe o modelo logístico trinomial. Essas curvaturas serão dadas por

$$C_\ell^{(1)}(\boldsymbol{\beta}_1) = 2 \mid \boldsymbol{\ell}^T \boldsymbol{\Delta}_1^T (\ddot{\mathbf{L}}_{11})^{-1} \boldsymbol{\Delta}_1 \boldsymbol{\ell} \mid \quad \text{e} \quad C_\ell^{(2)}(\boldsymbol{\beta}_2) = 2 \mid \boldsymbol{\ell}^T \boldsymbol{\Delta}_2^T (\ddot{\mathbf{L}}_{22})^{-1} \boldsymbol{\Delta}_2 \boldsymbol{\ell} \mid,$$

em que

$$\boldsymbol{\Delta}_1^T = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\omega}^T} \Big|_{\hat{\boldsymbol{\beta}}, \boldsymbol{\omega}_0} = \mathbf{X}^T \text{diag}(\mathbf{z}_1 - \hat{\boldsymbol{\pi}}_1) \quad \text{e} \quad \boldsymbol{\Delta}_2^T = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\omega}^T} \Big|_{\hat{\boldsymbol{\beta}}, \boldsymbol{\omega}_0} = \mathbf{X}^T \text{diag}(\mathbf{z}_2 - \hat{\boldsymbol{\pi}}_2).$$

Assim, as matrizes a serem diagonalizadas ficam dadas por

$$\begin{aligned} \mathbf{B}^{(1)} &= -\boldsymbol{\Delta}_1^T \ddot{\mathbf{L}}_{11}^{-1} \boldsymbol{\Delta}_1 = \text{diag}(\mathbf{z}_1 - \hat{\boldsymbol{\pi}}_1)^T \mathbf{X} (\mathbf{X}^T \hat{\mathbf{V}}_1 \mathbf{X})^{-1} \mathbf{X}^T \text{diag}(\mathbf{z}_1 - \hat{\boldsymbol{\pi}}_1) \\ &= \text{diag}\{\mathbf{r}_1\} \hat{\mathbf{H}}^{(1)} \text{diag}\{\mathbf{r}_1\}, \end{aligned}$$

em que $\mathbf{r}_1 = (r_{11}, \dots, r_{1n})^T$ com $r_{1i} = (z_{1i} - \hat{\pi}_{1i}) / \sqrt{\hat{\pi}_{1i}(1 - \hat{\pi}_{1i})}$

e

$$\begin{aligned}\mathbf{B}^{(2)} &= -\mathbf{\Delta}_2^T \ddot{\mathbf{L}}_{22}^{-1} \mathbf{\Delta}_2 = \text{diag}(\mathbf{z}_2 - \hat{\boldsymbol{\pi}}_2)^T \mathbf{X} (\mathbf{X}^T \hat{\mathbf{V}}_2 \mathbf{X})^{-1} \mathbf{X}^T \text{diag}(\mathbf{z}_2 - \hat{\boldsymbol{\pi}}_2) \\ &= \text{diag}\{\mathbf{r}_2\} \hat{\mathbf{H}}^{(2)} \text{diag}\{\mathbf{r}_2\},\end{aligned}$$

em que $\mathbf{r}_2 = (r_{21}, \dots, r_{2n})^T$ com $r_{2i} = (z_{2i} - \hat{\pi}_{2i}) / \sqrt{\hat{\pi}_{2i}(1 - \hat{\pi}_{2i})}$.

O autovetor correspondente ao maior autovalor de $\mathbf{B}^{(1)}$ será denotado por $\boldsymbol{\ell}_{max}^{(1)}$ e $\boldsymbol{\ell}_{max}^{(2)}$ será o autovetor correspondente ao autovalor de maior valor de $\mathbf{B}^{(2)}$. Deve-se proceder os gráficos de $\boldsymbol{\ell}_{max}^{(1)}$ e $\boldsymbol{\ell}_{max}^{(2)}$ contra os valores ajustados $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente.

2.5 Influência Local Total

Lesaffre e Verbeke (1998) segerem ao invés de considerar a direção de maior curvatura, as curvaturas nas direções de cada uma das observações. A curvatura na direção da i -ésima observação será dada por

$$C_i = 2 | \boldsymbol{\ell}_i^T \mathbf{\Delta}^T (\ddot{\mathbf{L}})^{-1} \mathbf{\Delta} \boldsymbol{\ell}_i |,$$

em que $\boldsymbol{\ell}_i$ é um vetor de zeros com 1 na i -ésima posição. Mostra-se facilmente para o modelo logístico trinomial que $C_i = 2 | B_{ii} |$, em que B_{ii} é o i -ésimo elemento da diagonal principal da matriz \mathbf{B} . Note que \mathbf{B} matriz $n \times n$. É sugerido o gráfico de C_i contra a ordem das observações. Deve-se considerar aqueles pontos tais que $C_i > 2\bar{C}$, em que \bar{C} é o valor médio dos C_i 's.

Por outro lado, podemos considerar as expressões das curvaturas dos sub-modelos logísticos que compõem o modelo logístico trinomial. Essas curvaturas serão dadas por

$$C_i^{(1)} = 2 | \boldsymbol{\ell}_i^T \mathbf{\Delta}_1^T (\ddot{\mathbf{L}}_{11})^{-1} \mathbf{\Delta}_1 \boldsymbol{\ell}_i | \quad \text{e} \quad C_i^{(2)} = 2 | \boldsymbol{\ell}_i^T \mathbf{\Delta}_2^T (\ddot{\mathbf{L}}_{22})^{-1} \mathbf{\Delta}_2 \boldsymbol{\ell}_i |.$$

Verifica-se facilmente que $C_i^{(1)} = 2 | B_{ii}^{(1)} |$ e $C_i^{(2)} = 2 | B_{ii}^{(2)} |$, em que $B_{ii}^{(1)}$ e $B_{ii}^{(2)}$ denotam, respectivamente, o i -ésimo elemento da diagonal principal de $\mathbf{B}^{(1)}$ e $\mathbf{B}^{(2)}$. A sugestão é construir os gráficos de $C_i^{(1)}$ contra $\hat{\pi}_{1i}$ e $C_i^{(2)}$ contra $\hat{\pi}_{2i}$ para avaliar observações influentes em $\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_2$, respectivamente. A sugestão é olhar com atenção aqueles pontos tais que $C_i^{(1)} > 2\bar{C}^{(1)}$ e $C_i^{(2)} > 2\bar{C}^{(2)}$.

Desenvolvendo as expressões para $C_i^{(1)}$ e $C_i^{(2)}$ mostra-se facilmente que

$$C_i^{(1)} = 2r_{1i}^2 \hat{h}_{ii}^{(1)} \quad \text{e} \quad C_i^{(2)} = 2r_{2i}^2 \hat{h}_{ii}^{(2)}.$$

2.6 Resíduos

Tradicionalmente a definição mais intuitiva de resíduo é dada pela diferença $(y_i - \hat{y}_i)$, que mede de forma linear quão o valor predito se afasta do valor observado. Essa definição funciona bem no caso linear, embora seja necessário sempre obter a forma padronizada da diferença, uma vez que as variâncias não são as mesmas. Além disso, mesmo que os erros sejam independentes os resíduos são em geral correlacionados, exigindo algum cuidado na hora de interpretá-los. Nesse sentido, Atkinson (1981, 1985) propõe, quando os resíduos são confrontados com alguma distribuição postulada o uso de envelopes, que são bandas de confiança obtidas pelo método de Monte Carlo a partir do modelo ajustado para avaliarem afastamentos sérios da distribuição postulada para os resíduos e de forma indireta da distribuição postulada para os erros.

Os resíduos, contudo, mais utilizados em modelos lineares generalizados são definidos a partir dos componentes da função desvio, que pode ser interpretada como foi visto na Seção 1.4 como sendo uma distância entre o logaritmo da verossimilhança do modelo saturado e o logaritmo da verossimilhança do modelo postulado. A versão padronizada (vide, por exemplo, McCullagh, 1987) é a seguinte:

$$t_{d_i} = \frac{\pm d(y_i; \hat{\mu}_i)}{(1 - \hat{h}_{ii})^{1/2}},$$

em que $d^2(y_i; \hat{\mu}_i)$ é o i -ésimo componente do desvio, \hat{h}_{ii} é o elemento da diagonal principal da matriz de projeção correspondente e o sinal atribuído é o mesmo de $(y_i - \hat{y}_i)$. Williams (1984) verificou através de estudos de simulação que a distribuição empírica está próxima da normalidade para vários MLGs. McCullagh (1987, p. 214) mostra para os MLGs que

a distribuição de probabilidade de

$$\frac{d(Y_i; \eta_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}},$$

é aproximadamente $N(0, 1)$, em que ρ_{3i} e ρ_{4i} são os coeficientes de assimetria e curtose de $\partial L(\eta_i)/\partial \eta_i$, respectivamente, e η_i é o preditor linear. É possível mostrar para os MLGs usando resultados de Cox e Snell (1968) que $E\{d(Y_i; \eta_i)\}=0$ e $\text{Var}\{d(Y_i; \eta_i)\} = 1 - h_{ii}$, em que os termos negligenciados são de $O(n^{-1})$. Esses resultados reforçam o uso da padronização acima para t_{d_i} .

No caso do modelo logístico trinomial o componente do desvio assume a forma:

$$\bullet d(y_i; \hat{\mu}_i) = \sqrt{2}\{a_{0i}\log(a_{0i}/n_i\hat{\pi}_{0i}) + a_{1i}\log(a_{1i}/n_i\hat{\pi}_{1i}) + a_{2i}\log(a_{2i}/n_i\hat{\pi}_{2i})\}$$

$$\text{se } 0 < a_{0i}, a_{1i}, a_{2i} < n_i,$$

$$\bullet d(y_i; \hat{\mu}_i) = \sqrt{-2n_i\log(\hat{\pi}_{0i})} \quad \text{se } (a_{0i} = n_i, a_{1i} = a_{2i} = 0),$$

$$\bullet d(y_i; \hat{\mu}_i) = \sqrt{-2n_i\log(\hat{\pi}_{1i})} \quad \text{se } (a_{1i} = n_i, a_{0i} = a_{2i} = 0) \text{ e}$$

$$\bullet d(y_i; \hat{\mu}_i) = \sqrt{-2n_i\log(\hat{\pi}_{2i})} \quad \text{se } (a_{2i} = n_i, a_{0i} = a_{1i} = 0).$$

Como fica difícil a atribuição de um sinal para $d(y_i; \hat{\mu}_i)$ uma vez que temos três diferenças para o i -ésimo elemento, $(a_{0i} - n_i\hat{\pi}_{0i})$, $(a_{1i} - n_i\hat{\pi}_{1i})$ e $(a_{2i} - n_i\hat{\pi}_{2i})$ deixaremos o resíduo componente do desvio sem sinal.

O principal desafio técnico neste caso é encontrar o fator de correção para a padronização do resíduo componente do desvio. Como isso envolve cálculos complicados abrangendo expansões de Cox e Snell (1968) podemos propor algum tipo de correção empírica envolvendo as quantidades $\hat{h}_{ii}^{(1)}$, $\hat{h}_{ii}^{(2)}$, $\hat{m}_{ii}^{(1)}$ e $\hat{m}_{ii}^{(2)}$ propostas na seção 2.2. Uma sugestão mais simples seria considerar como fator de correção a quantidade $\sqrt{1 - \bar{h}_{ii}}$, em que $\bar{h}_{ii} = (\hat{h}_{ii}^{(1)} + \hat{h}_{ii}^{(2)})/2$.

Assim, o resíduo componente do desvio padronizado ficaria dado por

$$t_{d_i} = \frac{d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}}$$

Resta saber empiricamente qual seria a distribuição de t_{d_i} . Contudo, o uso de envelopes conforme sugerido por Atkinson (1981) (vide também Williams, 1987) pode contornar o problema de não se saber a distribuição aproximada de t_{d_i} . O envelope consiste basicamente numa banda de confiança empírica para t_{d_i} construída através do modelo ajustado que pode ajudar na detecção de observações aberrantes segundo t_{d_i} , bem como afastamentos sérios da distribuição postulada para a resposta ou mesmo superdispersão, variâncias maiores do que aquelas postuladas para o modelo.

Um outro resíduo que poderia ser usado no caso trinomial é o resíduo de Pearson padronizado definido na seção 2.3. Nesse caso seriam calculados dois grupos de resíduos, um para cada sub-modelo logístico. Assim, seriam calculados os resíduos referentes às categorias $Y = 1$ e $Y = 2$, os quais são, respectivamente, definidos por

$$t_{S_i}^{(1)} = \frac{(a_{1i} - n_{1i}\hat{\pi}_{1i})}{\sqrt{n_{1i}\hat{\pi}_{1i}(1 - \hat{\pi}_{1i})(1 - \hat{h}_{ii}^{(1)})}} \quad \text{e} \quad t_{S_i}^{(2)} = \frac{(a_{2i} - n_{2i}\hat{\pi}_{2i})}{\sqrt{n_{2i}\hat{\pi}_{2i}(1 - \hat{\pi}_{2i})(1 - \hat{h}_{ii}^{(2)})}}$$

Como é mais razoável esperar que $t_{S_i}^{(1)}$ e $t_{S_i}^{(2)}$ sejam aproximadamente normais de média zero e variância um quando n_{1i} e n_{2i} não são pequenos, algumas técnicas usuais de diagnóstico podem ser usadas com mais segurança. Por exemplo, os gráficos de $t_{S_i}^{(1)}$ e $t_{S_i}^{(2)}$ contra os valores ajustados $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente, podem revelar quais observações são marginalmente aberrantes em cada sub-modelo. Nesse caso também poderiam ser atribuídos envelopes para cada um dos resíduos $t_{S_i}^{(1)}$ e $t_{S_i}^{(2)}$. Esses envelopes embora levem em conta resíduos de sub-modelos logísticos devem ser gerados do modelo logístico trinomial ajustado.

Capítulo 3

Aplicações

3.1 Introdução

Neste capítulo vamos aplicar as diversas técnicas de diagnóstico desenvolvidas no capítulo 2 aos exemplos vistos no capítulo 1. Iremos ainda exibir um painel comparativo onde serão confrontadas as técnicas que serão aplicadas aos sub-modelos do modelo trinomial com os diagnósticos desenvolvidos diretamente para este modelo.

Como acontece com qualquer modelo ajustado, antes de utilizá-lo e fazer inferências devemos avaliar seu ajuste final e examinar sua contribuição para cada um dos indivíduos. Na regressão logística trinomial este passo é bem mais complicado do que no caso binomial. Quando nosso modelo tem respostas binárias temos apenas um valor ajustado, a probabilidade logística da resposta é dada por $P(Y=1 \mid \mathbf{x})$. Quando a variável resposta tem três categorias teremos duas probabilidades a serem estimadas, $P(Y=1 \mid \mathbf{x})$ e $P(Y=2 \mid \mathbf{x})$.

Para esta avaliação vamos utilizar a abordagem de Hosmer e Lemeshow que recomendam que avaliemos os dois sub-modelos calculando as medidas de diagnósticos usando as regressões logísticas individuais, a seguir os resultados devem ser integrados levando em consideração os efeitos dos pontos influentes em cada modelo. Em particular, os pontos que são influentes em apenas um dos sub-modelos devem ser examinados com mais

atenção.

3.2 Exemplo 1

Este exemplo avalia fatores associados ao conhecimento, atitude e comportamento de mulheres em relação ao exame de mamografia, os ajustes finais são dados abaixo

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -2,7024 + 2,0950 \times \text{SintomaD}_i - 0,2510 \times \text{Beneficio}_i + \\ 1,2933 \times \text{Historico}_i + 1,2439 \times \text{Autoexame}_i + 0,0893 \times \text{Detectar}_{2i} + \\ 0,9719 \times \text{Detectar}_{3i}$$

e

$$\log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -0,9987 + 1,1214 \times \text{SintomaD}_i - 0,1681 \times \text{Beneficio}_i + \\ 1,0140 \times \text{Historico}_i + 1,0286 \times \text{Autoexame}_i - 0,9022 \times \text{Detectar}_{2i} - \\ 0,6698 \times \text{Detectar}_{3i}.$$

O primeiro sub-modelo compara mulheres que realizaram o exame de mamografia nos últimos 12 meses (**Experiência=1**) ao grupo de mulheres que nunca realizaram o exame (**Experiência=0**) enquanto que o segundo sub-modelo compara o grupo de mulheres que submeteram-se ao exame a mais de um ano (**Experiência=2**) ao grupo de mulheres que nunca realizaram o exame (**Experiência=0**).

As estatísticas de qualidade do ajuste de Hosmer-Lemeshow apresentaram níveis descritivos de 0,55 para o sub-modelo 1 e 0,11 para o sub-modelo 2, ou seja, C_1^{HL} e C_2^{HL} apresentam valores não significativos, indicando que ambos os modelos estão satisfatoriamente ajustados.

Ainda sob a ótica de Hosmer e Lemeshow vamos construir os gráficos de diagnóstico para os dois sub-modelos estimados.

Apresentaremos alguns gráficos de diagnóstico. Inicialmente os gráfico de $\hat{h}_{ii}^{(1)}$ e $\hat{h}_{ii}^{(2)}$ contra as probabilidades ajustadas $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente (Figuras 3.1 e 3.2). Como podemos notar nesses dois gráficos três mulheres se destacam com um valor relativamente alto para as medidas de alavanca, #83 e #12 (Figura 3.1) e #241 (Figura 3.2). Em seguida são apresentadas as medidas aproximadas dos afastamentos da verossimilhança $LD_i^{(1)}$ e $LD_i^{(2)}$ contra as probabilidades ajustadas $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente (Figuras 3.3 e 3.4). Destacam-se na primeira figura as mulheres #80 e #293 e na segunda figura as mulheres #90 e #260. Essas quatro mulheres têm em comum o fato de terem probabilidades ajustadas baixas (menores que 0,20) para π_1 e π_2 , porém tiveram respostas observadas $Y=1$ ou $Y=2$. Depois são apresentados nas Figuras 3.5 e 3.6 os gráficos de resíduos de Pearson padronizados $t_{S_i}^{(1)}$ e $t_{S_i}^{(2)}$ contra as probabilidades ajustadas $\hat{\pi}_{1i}$ e $\hat{\pi}_{2i}$, respectivamente. Podemos notar por essas figuras e principalmente pelos histogramas apresentados nas Figuras 3.7 e 3.8 uma acentuada assimetria na distribuição empírica desses resíduos, dificultando a definição de critérios para classificação de pontos aberrantes. Esse tipo de comportamento já era esperado nesse exemplo uma vez que não se tem réplicas, ou seja, $n_i = 1, \forall_i$. A distribuição de t_{S_i} se aproxima de uma normal padrão quando n_i não é pequeno, \forall_i . Partimos então para o desenvolvimento do gráfico do resíduo do componente do desvio t_{d_i} (Figura 3.11) contra a ordem das observações. Como não foi atribuído sinal para esse resíduo, olhamos apenas aqueles pontos que se destacaram com os maiores valores que correspondem às mulheres #80 e #293 que já haviam sido destacados em figuras anteriores.

Tabela 3.1*Pontos por sub-modelo para cada medida de diagnóstico.*

Gráfico	Sub-modelo	Medida de diagnóstico	Pontos a serem observados
3.1	1	Alavanca - binomial	#12 e #83
3.3	1	Alavanca - trinomial	#12 e #83
3.5	1	Afastamento da verossimilhança	#80 e #293
3.11	1	Influência local total - binomial	#83 e #234
3.2	2	Alavanca - binomial	#241
3.4	2	Alavanca - trinomial	#241, #279 e #382
3.6	2	Afastamento da verossimilhança	#90 e #260
3.8	2	Influência local - binomial	#128 e #260
3.12	2	Influência local total - binomial	#260
3.9	ambos	Influência local - trinomial	#39 e #108
3.10	ambos	Influência local total - trinomial	#80 e #293
3.13	ambos	Componente do desvio - trinomial	#80 e #293

A tabela acima mostra os pontos que foram detectados como possíveis observações influentes e/ou aberrantes, em cada um dos gráficos para cada sub-modelo avaliado.

Para auxiliar na análise vamos traçar um perfil de cada uma das mulheres:

Observação #12 - Nunca realizou o exame, acha necessário o exame mesmo antes de desenvolver os sintomas, tem histórico de câncer familiar, não sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,424$ e $\hat{\pi}_2=0,212$.

Observação #83 - Submeteu-se ao exame nos últimos 12 meses, acha necessário o exame mesmo antes de desenvolver os sintomas, tem histórico de câncer familiar, não sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,378$ e $\hat{\pi}_2=0,205$.

Observação #80 - Submeteu-se ao exame nos últimos 12 meses, não acha necessário o exame antes de desenvolver os sintomas, não tem histórico de câncer familiar, não sabe como se auto-examinar, acha provável que o exame venha a detectar a doença e concorda

fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,015$ e $\hat{\pi}_2=0,051$.

Observação #293 - Submeteu-se ao exame nos últimos 12 meses, não acha necessário o exame antes de desenvolver os sintomas, não tem histórico de câncer familiar, sabe como se auto-examinar, acha pouco provável que o exame venha a detectar a doença e concorda parcialmente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,016$ e $\hat{\pi}_2=0,158$.

Observação #234 - Submeteu-se ao exame nos últimos 12 meses, acha necessário o exame mesmo antes de desenvolver os sintomas, não tem histórico de câncer familiar, não sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,247$ e $\hat{\pi}_2=0,150$.

Observação #241 - Nunca realizou o exame, acha necessário o exame mesmo antes de desenvolver os sintomas, tem histórico de câncer familiar, não sabe como se auto-examinar, acha pouco provável que o exame venha a detectar a doença e concorda parcialmente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,078$ e $\hat{\pi}_2=0,303$.

Observação #382 - Submeteu-se ao exame a mais de um ano, acha necessário o exame mesmo antes de desenvolver os sintomas, não tem histórico de câncer familiar, sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,278$ e $\hat{\pi}_2=0,190$.

Observação #279 - Nunca realizou o exame, não acha necessário o exame antes de desenvolver os sintomas, tem histórico de câncer familiar, sabe como se auto-examinar, acha pouco provável que o exame venha a detectar a doença e concorda parcialmente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,036$ e $\hat{\pi}_2=0,298$.

Observação #90 - Submeteu-se ao exame a mais de um ano, não acha necessário o exame antes de desenvolver os sintomas, não tem histórico de câncer familiar, não sabe como se auto-examinar, acha pouco provável que o exame venha a detectar a doença e concorda parcialmente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,006$ e $\hat{\pi}_2=0,075$.

Observação #260 - Submeteu-se ao exame a mais de um ano, acha necessário o exame antes de desenvolver os sintomas, não tem histórico de câncer familiar, não sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,035$ e $\hat{\pi}_2=0,168$.

Observação #128 - Submeteu-se ao exame a mais de um ano, acha necessário o exame antes de desenvolver os sintomas, não tem histórico de câncer familiar, sabe como se auto-examinar, acha pouco provável que o exame venha a detectar a doença e concorda com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,074$ e $\hat{\pi}_2=0,308$.

Observação #39 - Nunca realizou o exame, acha necessário o exame mesmo antes de desenvolver os sintomas, tem histórico de câncer familiar, sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,639$ e $\hat{\pi}_2=0,237$.

Observação #108 - Nunca realizou o exame, acha necessário o exame mesmo antes de desenvolver os sintomas, tem histórico de câncer familiar, sabe como se auto-examinar, acha muito provável que o exame venha a detectar a doença e concorda fortemente com o benefício do exame. Probabilidades ajustadas $\hat{\pi}_1=0,639$ e $\hat{\pi}_2=0,237$.

Os pontos que se destacaram nos gráficos de alavanca serão considerados possivelmente influentes se os valores de \hat{h}_{ii} destoarem de seus valores esperados, conforme tabela abaixo:

Tabela 3.2

Pontos de alavanca por sub-modelo.

Sub-modelo	Observação	Valor de $\hat{\pi}_i$	Valor esperado de \hat{h}_{ii}
1	#12	0,4239	Pequeno para π_i entre 0,3 e 0,7
1	#83	0,3777	Pequeno para π_i entre 0,3 e 0,7
2	#241	0,3035	Pequeno para π_i entre 0,3 e 0,7

Para os três pontos (#12, #83 e #241) que se destacaram o valor de \hat{h}_{ii} esperado deve ser pequeno segundo o critério de Hosmer e Lemeshow (1989), porém os valores

encontrados podem ser considerados altos uma vez que $\bar{h} \cong 2p/n=0,034$. Assim, essas três mulheres podem ter alguma influência nos resultados dos modelos. Além desses pontos o gráfico de alavanca detectou também no sub-modelo 2 as observações #279 e #382.

As observações #80 e #293 da Figura 3.5 (sub-modelo 1) e #90 e #260 da Figura 3.6 (sub-modelo 2) parecem destoar das demais segundo LD_i . Se concluirmos que a variação que essas observações causam nas estimativas são muito desproporcionais, podemos afirmar que esses pontos são influentes. Porém, deve-se verificar ainda se a eliminação dos mesmos leva a mudanças inferenciais, deixando, por exemplo, algum coeficiente não significativo. Note ainda que esses mesmos pontos foram detectados pelo gráfico de Influência local total trinomial Figura 3.10 e pelo gráfico de resíduos de componente do desvio trinomial Figura 3.13.

Iremos apresentar a seguir três tabelas em que retiramos as observações destacadas acima individualmente e verificaremos as variações percentuais nos coeficientes estimados bem como os novos níveis descritivos. Assim, poderemos verificar se as variações ocorridas nas estimativas levam a mudanças inferenciais, ou seja, se algum coeficiente deixa de ser significativo ou passa a ser significativo. Na Tabela 3.3 apresentamos esse estudo para as observações #12, #83 e #241 que se destacaram nos gráficos de pontos de alavanca binomial. Nas Tabelas 3.4 e 3.5 são desenvolvidos estudos similares para as observações (#80 e #293) que se destacaram nos gráficos de $LD_i^{(1)}$, t_{d_i} e C e ainda para as observações (#90 e #260) que se destacaram nos gráficos de $LD_i^{(2)}$, $C_l^{(2)}$ e $C_i^{(2)}$ para os sub-modelos 1 e 2, respectivamente.

Antes de avaliarmos as variações das estimativas e níveis descritivos que serão exibidos pelas tabelas a seguir, vale lembrar (como foi visto no capítulo 1) que a variável explicativa **Detectar** foi avaliada como não-significativa pela estatística de Wald (vide Tabela 1.6). Apesar disso ela permaneceu no modelo devido a sua influência confundidora na variável **SintomaD**.

Inicialmente iremos considerar a retirada das observações #12, #83 e #241 (Tabela 3.3). Além das fortes variações nas estimativas observadas nos coeficientes da variável explicativa `Detectar`, notamos também variações nas estimativas dos coeficientes das variáveis `Historico` e `Autoexame`. Esse comportamento entretanto não implica em mudança inferencial para a variável `Historico` nem na variável `Detectar`. Já para a variável `Autoexame` notamos mudança na significância de seu coeficiente, variando o nível descritivo de $\simeq 4\%$ para $\simeq 10\%$ no sub-modelo 2. Isso significa dizer que a retirada das observações #12, #83 e #241 afeta o sub-modelo 2 tornando a variável `Autoexame` não significativa para o mesmo se considerarmos um nível de significância, por exemplo, de 5%.

Consideramos agora a retirada das observações #80 e #293. A Tabela 3.4 exhibe as variáveis `SintomaD` e `Autoexame` apresentando variações destacadas no sub-modelo 1 principalmente quando retiramos as duas observações (#80 e #293) simultaneamente dos dados. Note ainda que a retirada dessas observações pouco influencia nas estimativas para o sub-modelo 2, isso parece fazer sentido, uma vez, que esses pontos mostraram-se influentes apenas no sub-modelo 1.

Tabela 3.3

Variação nas estimativas e níveis descritivos eliminando as observações #12, #83 e #241.

Sub-modelo	Variável	Var. %	Nível descritivo
1	SintomaD	0	0,0001
	Beneficio	-1	0,0005
	Historico	5	0,0032
	Autoexame	1	0,0276
	Detectar ₂	-166	0,9597
	Detectar ₃	16	0,4707
	Intercepto	-5	0,0787
2	SintomaD	3	0,0013
	Beneficio	0	0,0236
	Historico	14	0,0145
	Autoexame	-17	0,0970
	Detectar ₂	-12	0,1603
	Detectar ₃	-16	0,2652
	Intercepto	23	0,4719

Apesar das variações desproporcionais observadas não notamos mudanças inferenciais significativas, levando-nos a não considerar a retirada desses pontos.

As Figuras 3.5, 3.10 e 3.13, apresentam as mesmas observações #80 e #293 em destaque, esse pontos foram estudados na Tabela 3.4.

Na Tabela 3.5 podemos notar que as variáveis **SintomaD**, **Historico** e **Autoexame** apresentam variações em seus coeficientes, sendo que essas variações ficam mais evidentes no sub-modelo 2. quando as duas observações (#90, #260) são retiradas simultaneamente dos dados.

Tabela 3.4*Variação nas estimativas e níveis descritivos eliminando #80, #293 e ambas.*

Sub-modelo	Variável	Var. %	Nível descritivo	Var. %	Nível descritivo	Var. %	Nível descritivo
1	SintomaD	10	0,0001	10	0,0001	21	0,0001
	Beneficio	-4	0,0011	-2	0,0005	2	0,0010
	Historico	3	0,0024	5	0,0022	8	0,0019
	Autoexame	23	0,0075	-3	0,0233	21	0,0095
	Detectar ₂	-136	0,9779	11.965	0,9547	11.809	0,9546
	Detectar ₃	-1	0,3942	1.098	0,9510	1.096	0,9504
	Intercepto	-19	0,0287	-399	0,9433	-420	0,9400
2	SintomaD	1	0,0015	0	0,0017	1	0,0015
	Beneficio	-1	0,0255	0	0,0231	1	0,0252
	Historico	2	0,0236	2	0,0232	4	0,0215
	Autoexame	3	0,0389	-2	0,0486	2	0,0417
	Detectar ₂	-2	0,1979	-14	0,2770	-12	0,2691
	Detectar ₃	0	0,3319	-19	0,4277	-20	0,4348
	Intercepto	-5	0,3279	-10	0,3049	-16	0,2813

Tabela 3.5*Variação nas estimativas e níveis descritivos eliminando #90, #260 e ambas.*

Sub-modelo	Variável	Var. %	Nível descritivo	Var. %	Nível descritivo	Var. %	Nível descritivo
1	SintomaD	1	0,0001	0	0,0001	0	0,0001
	Beneficio	0	0,0006	0	0,0006	0	0,0006
	Historico	2	0,0025	1	0,0026	4	0,0021
	Autoexame	3	0,0153	2	0,0163	5	0,0132
	Detectar ₂	167	0,8373	117	0,8676	309	0,7529
	Detectar ₃	12	0,3193	11	0,3392	28	0,2676
	Intercepto	-7	0,0438	-4	0,0505	-12	0,0351
2	SintomaD	7	0,0010	-3	0,0015	5	0,0015
	Beneficio	1	0,0290	-1	0,0226	1	0,0277
	Historico	6	0,0194	4	0,0204	11	0,0150
	Autoexame	20	0,0273	18	0,0290	46	0,0167
	Detectar ₂	56	0,5110	60	0,4733	950	0,9866
	Detectar ₃	42	0,7031	111	0,6721	130	0,8016
	Intercepto	-69	0,1420	-48	0,1939	-140	0,0580

Novamente as variações desproporcionais devidas à retirada das observações não levam a mudanças inferenciais significativas, levando-nos a não considerar a retirada destas observações.

Tem-se nas Figuras 3.9 e 3.10 os gráficos de $m_{ii}^{(1)}$ e $m_{ii}^{(2)}$ contra as probabilidades estimadas pelos sub-modelos 1 e 2, respectivamente. Como podemos notar, são as mesmas observações (#12, #83 e #241) destacadas pelas medidas de alavanca (Figuras 3.1 e 3.2) considerando os dois sub-modelos, as quais foram estudadas na Tabela 3.3.

Ao analisarmos as duas estratégias de detecção de pontos aberrantes e/ou influentes, verificamos que as observações em destaque são as mesmas, dando indicação de que é indiferente utilizar a abordagem de Hosmer e Lemeshow que calcula essas medidas para os dois sub-modelos isoladamente ou medidas desenvolvidas neste trabalho diretamente para o modelo logístico trinomial.

Se compararmos as Figuras 3.1 com 3.9 e ainda 3.2 com 3.10, temos a impressão de tratar-se das mesmas medidas. Entretanto, existe uma sutil diferença entre elas, porém isso não altera nossa indicação apontada pela Tabela 3.2.

A Figura 3.11 do resíduo componente do desvio t_{d_i} contra a ordem das observações, mostrou-se eficaz na detecção de pontos, embora essas mesmas observações já tivessem sido evidenciadas pela Figura 3.3 que exhibe os afastamentos da verossimilhança.

3.3 Exemplo 2

Este exemplo investiga a associação entre resultado num teste respiratório, hábito de fumar e idade. Os ajustes finais são dados abaixo

$$\log \left\{ \frac{\hat{\pi}_{1i}}{\hat{\pi}_{0i}} \right\} = -3,8508 + 0,5380 \times \text{idade}_i + 0,7881 \times \text{status}_{2_i} + 0,8319 \times \text{status}_{3_i}$$

e

$$\log \left\{ \frac{\hat{\pi}_{2i}}{\hat{\pi}_{0i}} \right\} = -6,5231 + 1,3786 \times \text{idade}_i + 0,8221 \times \text{status}_{2_i} + 1,3463 \times \text{status}_{3_i}.$$

O primeiro sub-modelo compara o grupo de trabalhadores que tiveram resultado indefinido no teste respiratório (**Resultado=1**) ao grupo de trabalhadores que tiveram resultado normal no teste respiratório (**Resultado=0**) enquanto que o segundo sub-modelo compara o grupo de trabalhadores que tiveram resultado anormal no teste respiratório (**Resultado=2**) ao grupo de trabalhadores que tiveram resultado normal no teste respiratório (**Resultado=0**).

As estatísticas de qualidade do ajuste de Hosmer-Lemeshow apresentaram níveis descritivos de 0,011 para o sub-modelo 1 e 0,037 para o sub-modelo 2, ou seja, C_1^{HL} e C_2^{HL} apresentam valores significativos para $\alpha = 5\%$, indicando que ambos os modelos não estão satisfatoriamente ajustados.

Uma alternativa seria incluir no modelo a interação **idade** \times **status** e avaliar a significância de C_1^{HL} e C_2^{HL} .

Como temos um exemplo de dados agrupados perde-se um pouco o sentido a aplicação de gráficos de diagnóstico para avaliação do impacto individual das observações nas estimativas. Aqui as observações são grupos não necessariamente do mesmo tamanho. Assim, iremos apenas desenvolver gráficos de $t_{S_i}^{(1)}$ e $t_{S_i}^{(2)}$ contra as probabilidades ajustadas $\hat{\pi}_{1_i}$ e $\hat{\pi}_{2_i}$, respectivamente, que são apresentados nas Figuras 3.12 e 3.13. Como podemos notar apesar do exemplo contar com 2.219 respondentes, temos apenas seis agrupamentos o que dificulta qualquer estratégia de eliminação de pontos, uma vez que ao eliminarmos um grupo estaremos eliminando um grande número de observações, o que deverá implicar em mudanças não apenas nos coeficientes mas também em suas significâncias. Sendo assim, preferimos manter todos os grupos de observações. As Figuras 3.12 e 3.13 apresentam os resíduos de Pearson para os dois sub-modelos. Pode-se notar a presença de resíduos com valores altos fora dos padrões usuais, confirmando a falta de ajuste apontada pelas estatísticas de Hosmer e Lemeshow.

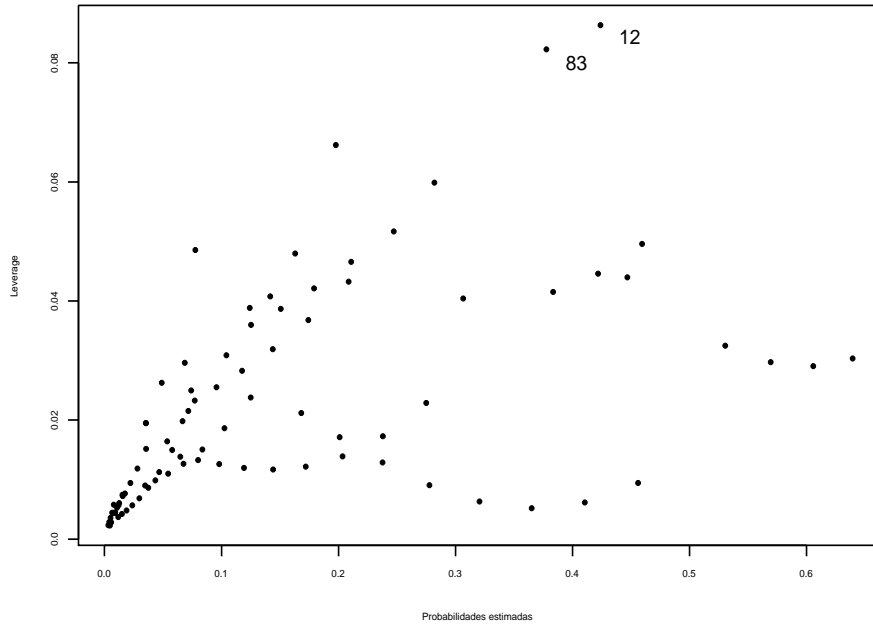


Figura 3.1: Gráfico de $\hat{h}_{ii}^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

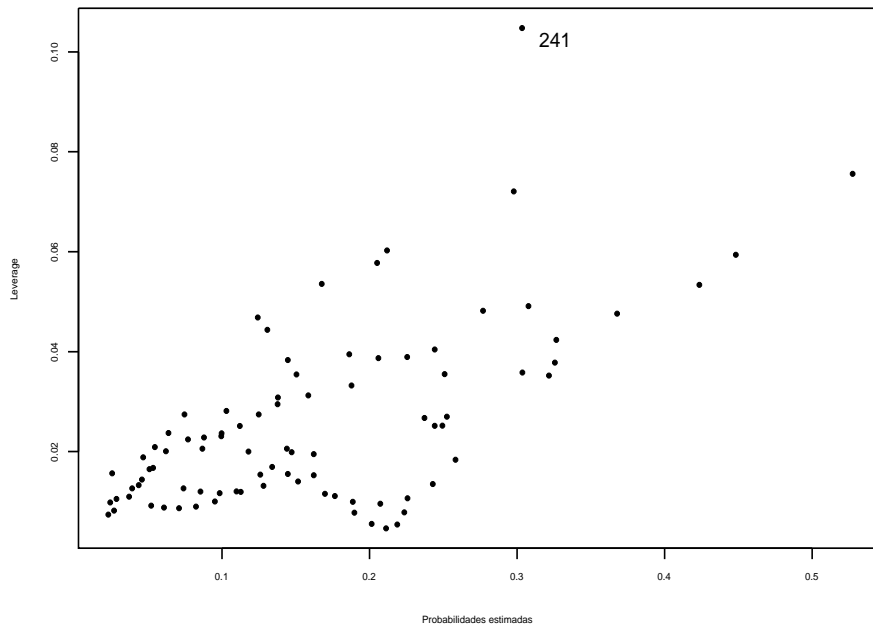


Figura 3.2: Gráfico de $\hat{h}_{ii}^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

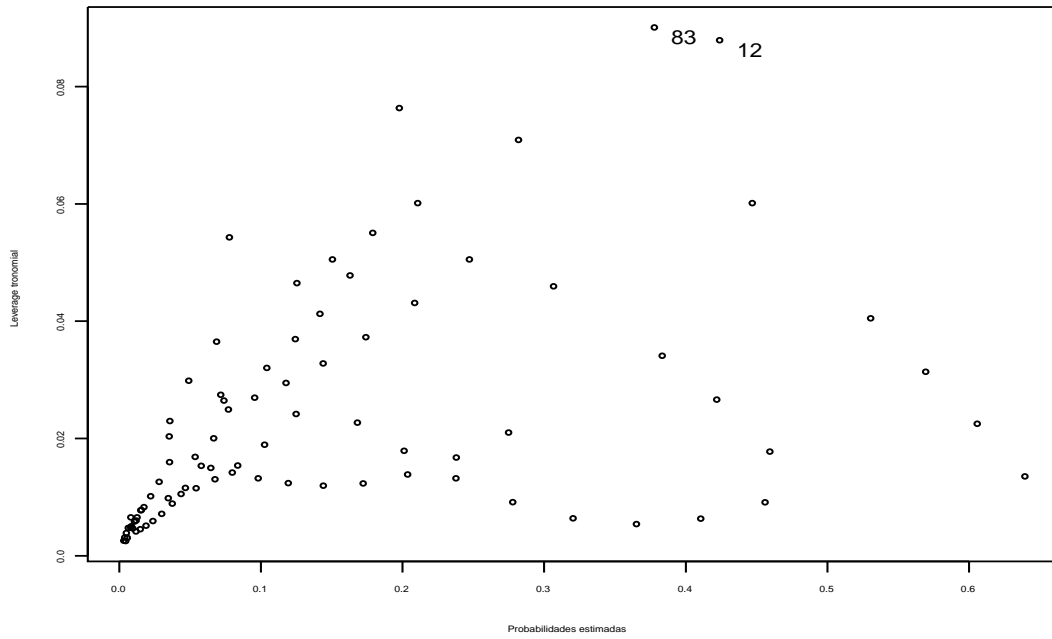


Figura 3.3: Gráfico de $m_{ii}^{(1)}$ contra as probabilidades estimadas do sub-modelo 1.

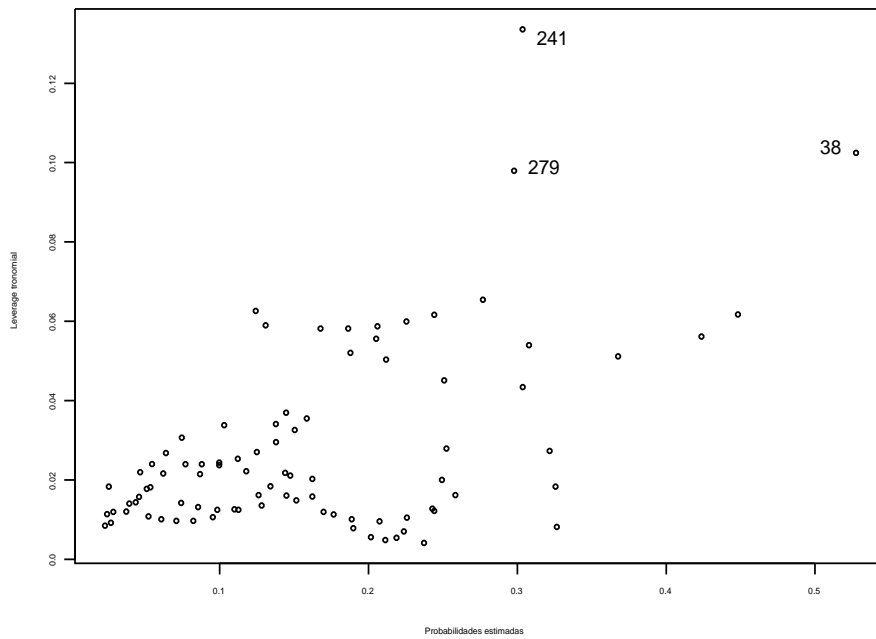


Figura 3.4: Gráfico de $m_{ii}^{(2)}$ contra as probabilidades estimadas do sub-modelo 2.

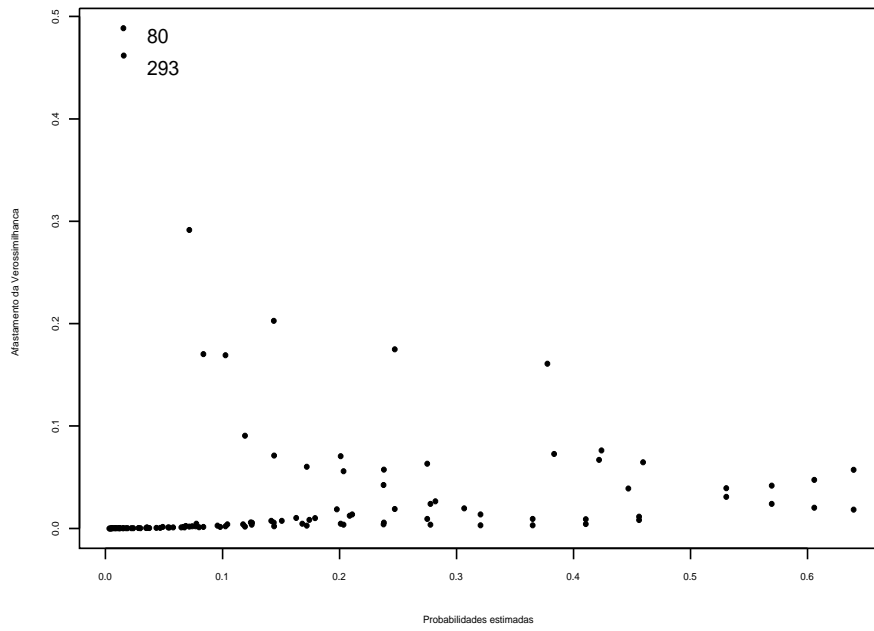


Figura 3.5: Gráfico de $LD_i^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

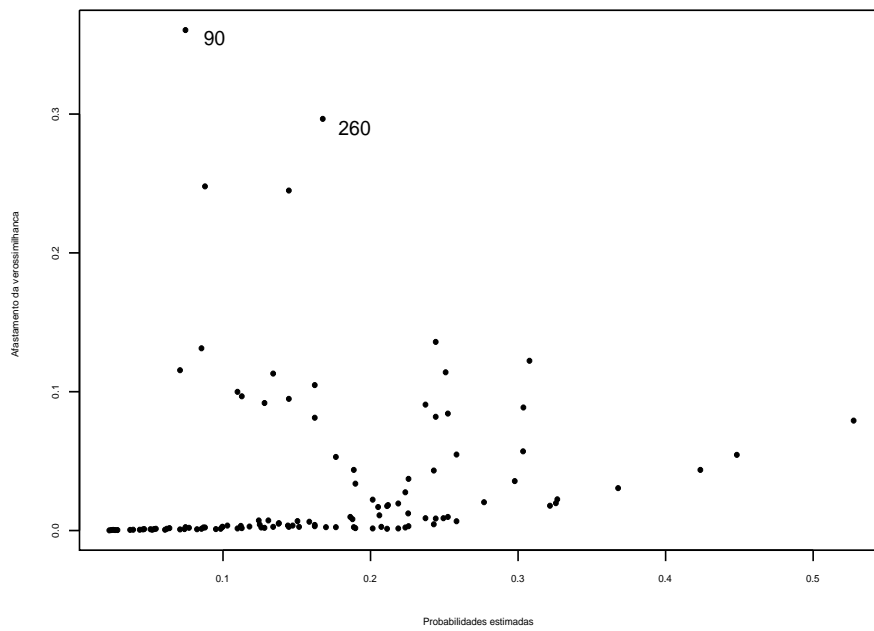


Figura 3.6: Gráfico de $LD_i^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

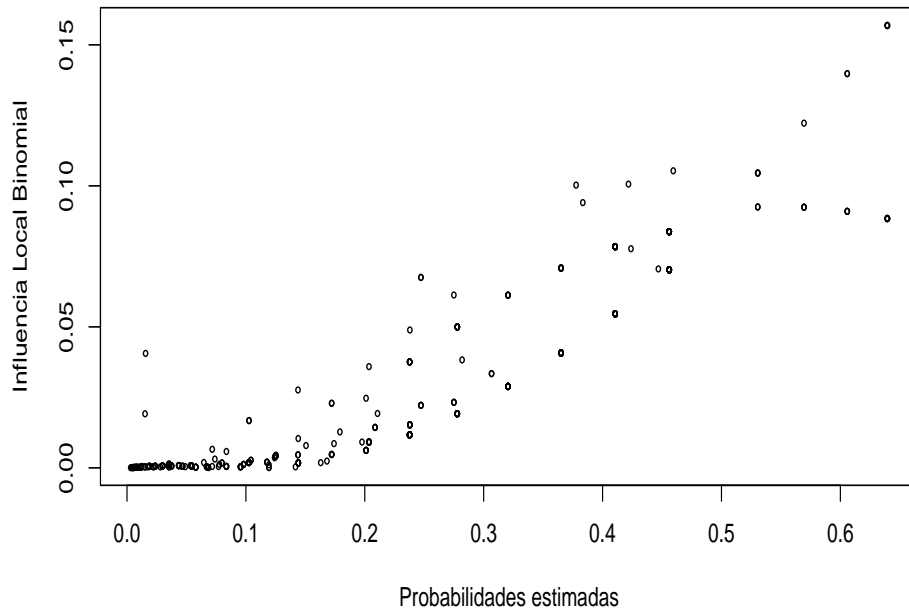


Figura 3.7: Gráfico de $C_\ell^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

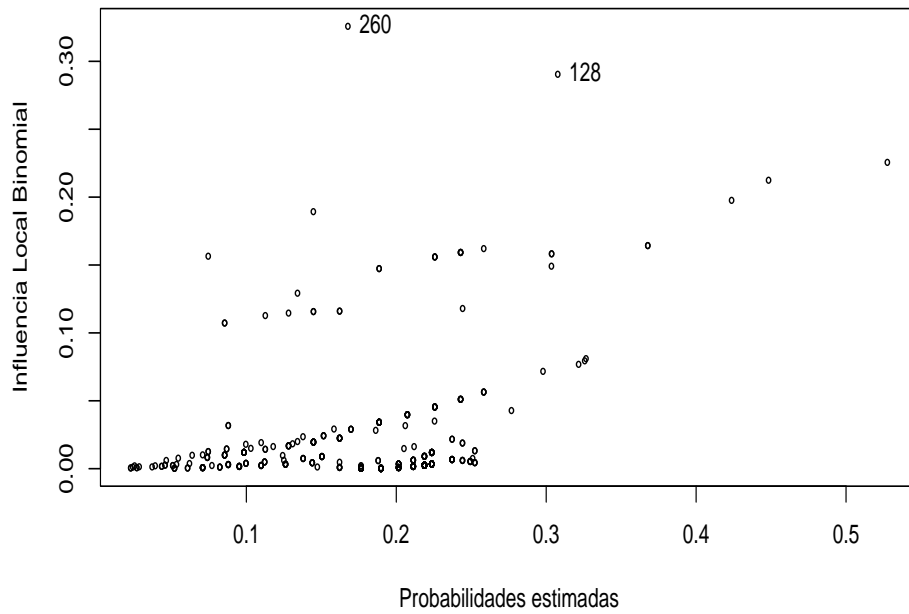


Figura 3.8: Gráfico de $C_\ell^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

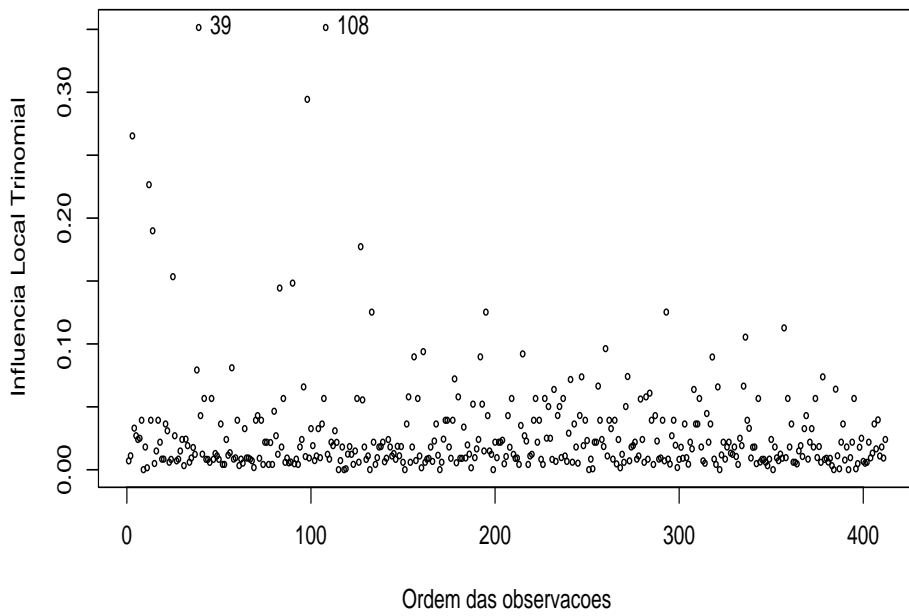


Figura 3.9: Gráfico de ℓ_{max} contra a ordem das observações.

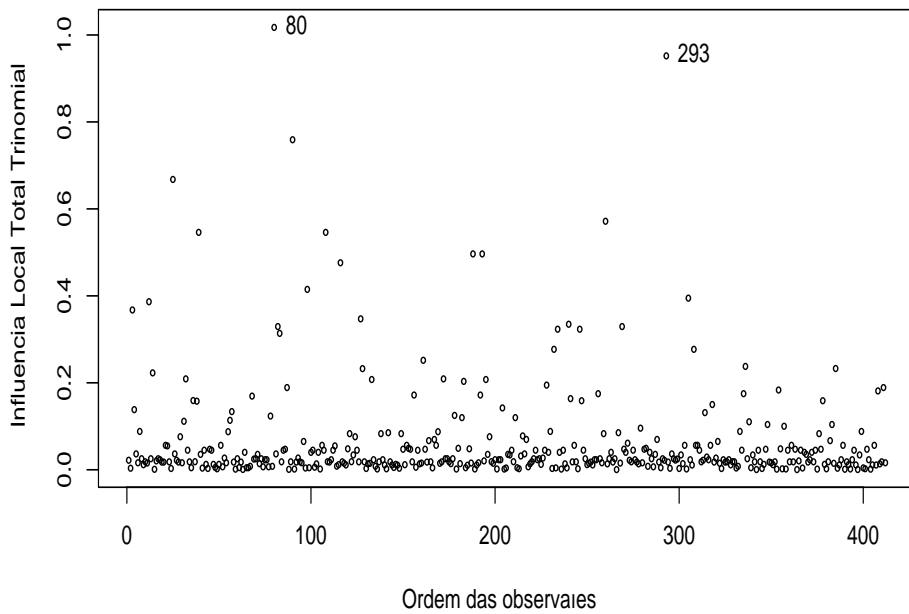


Figura 3.10: Gráfico de C_i contra a ordem das observações.

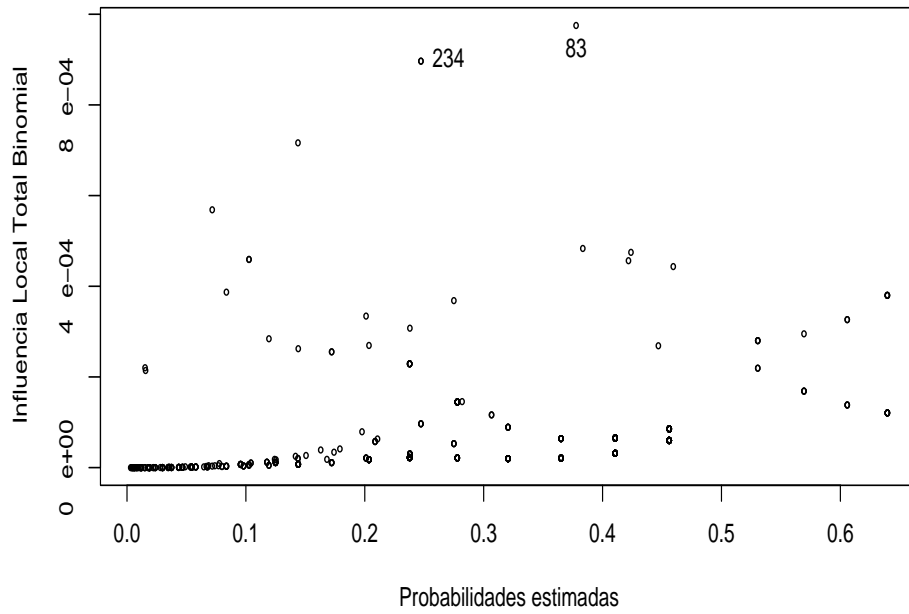


Figura 3.11: Gráfico de $C_i^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

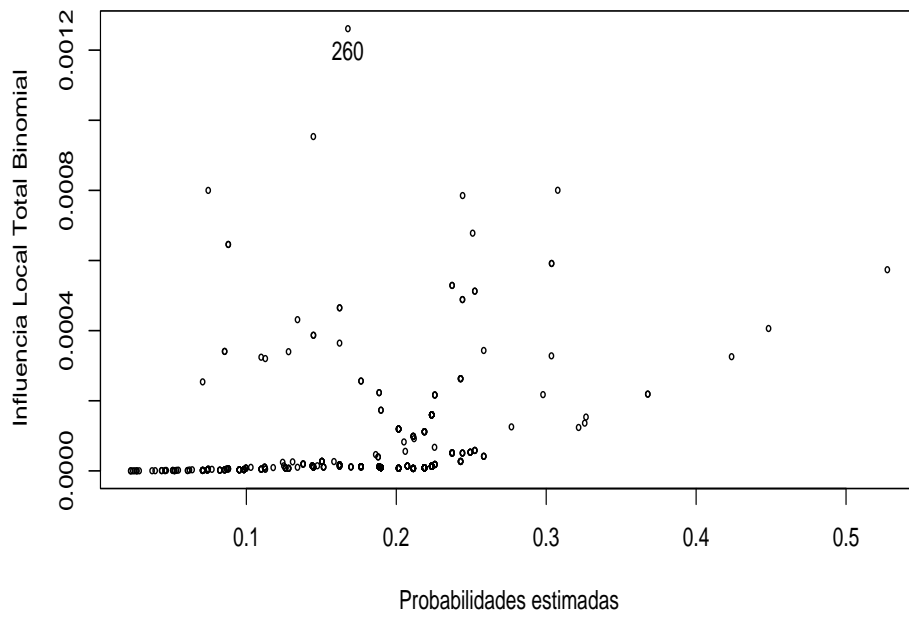


Figura 3.12: Gráfico de $C_i^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

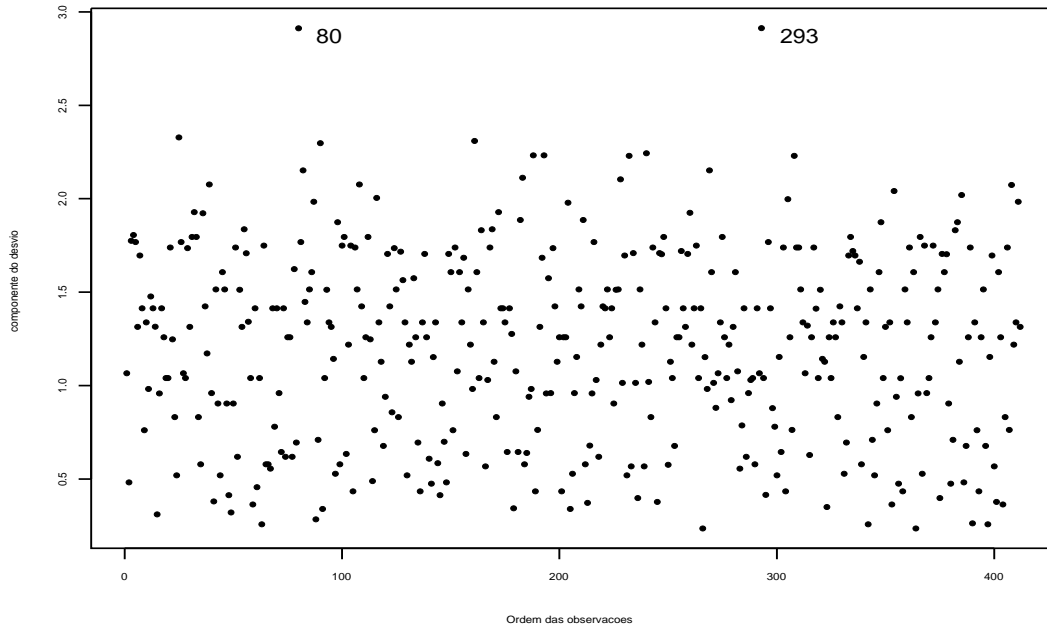


Figura 3.13: Gráfico de t_{d_i} contra a ordem das observações.

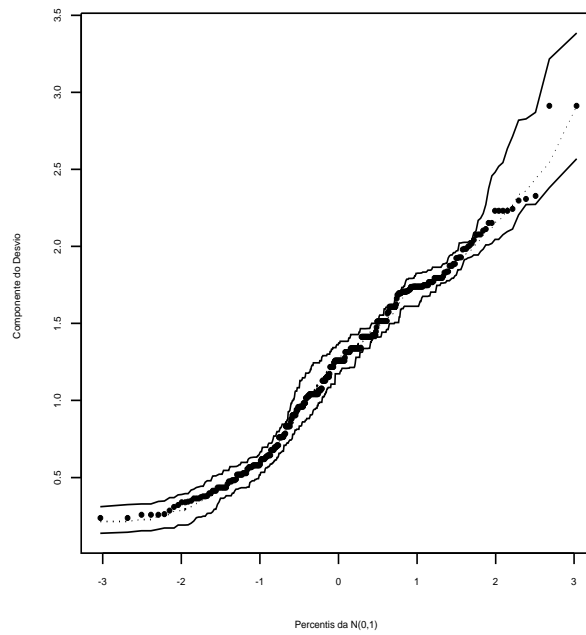


Figura 3.14: Gráfico de envelope t_{d_i} .

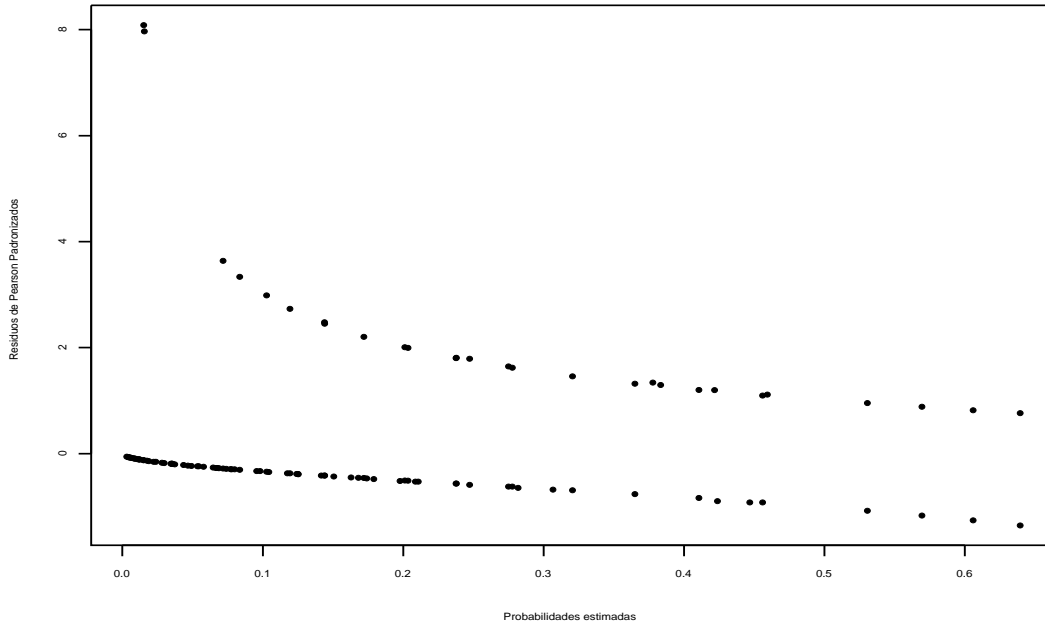


Figura 3.15: Gráfico de $t_{S_i}^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

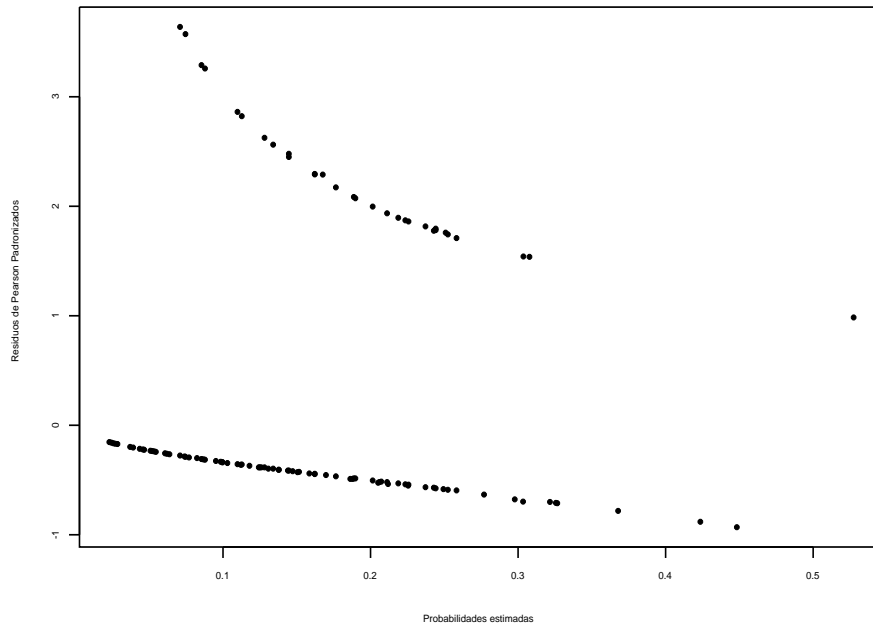


Figura 3.16: Gráfico de $t_{S_i}^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

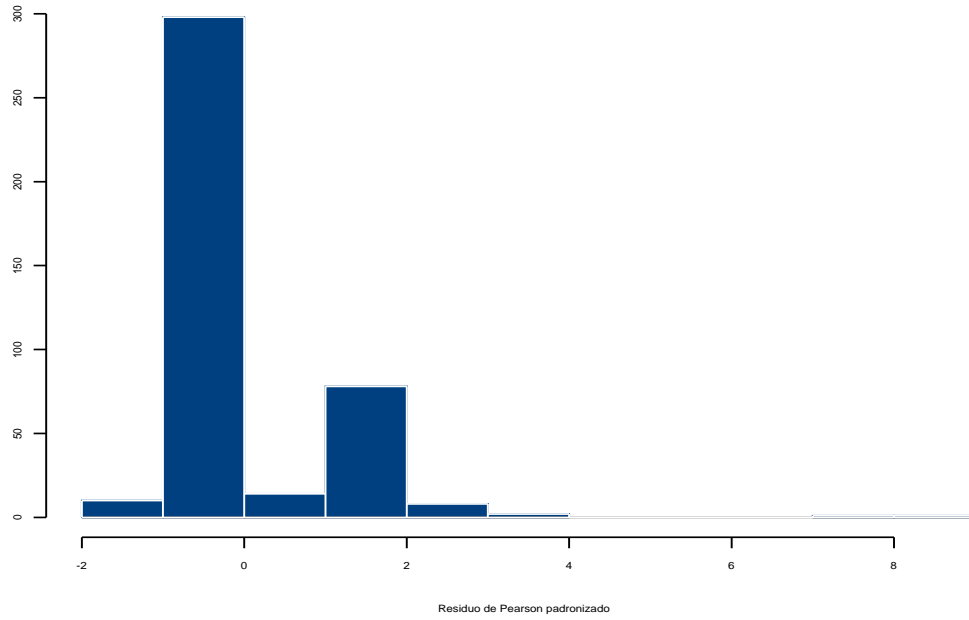


Figura 3.17: Histograma de $t_{S_i}^{(1)}$ para o sub-modelo 1.

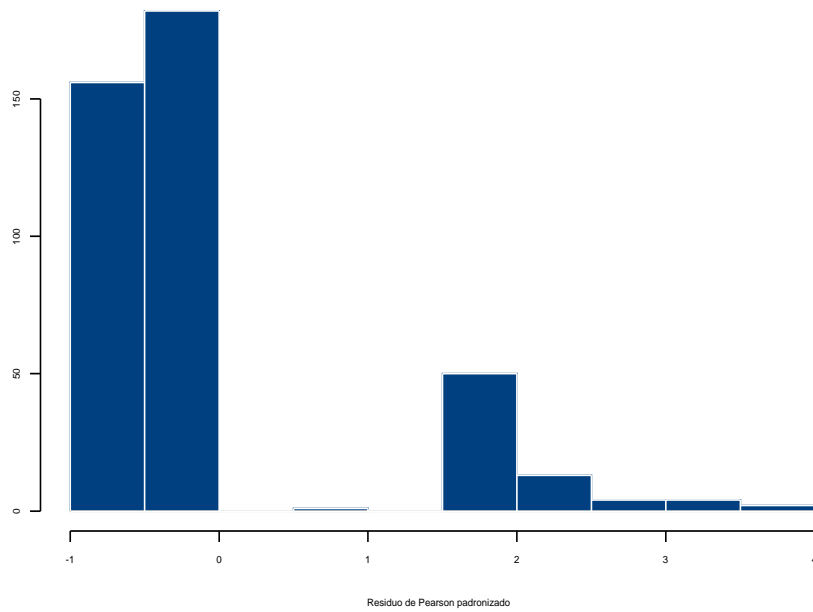


Figura 3.18: Histograma de $t_{S_i}^{(2)}$ para o sub-modelo 2.

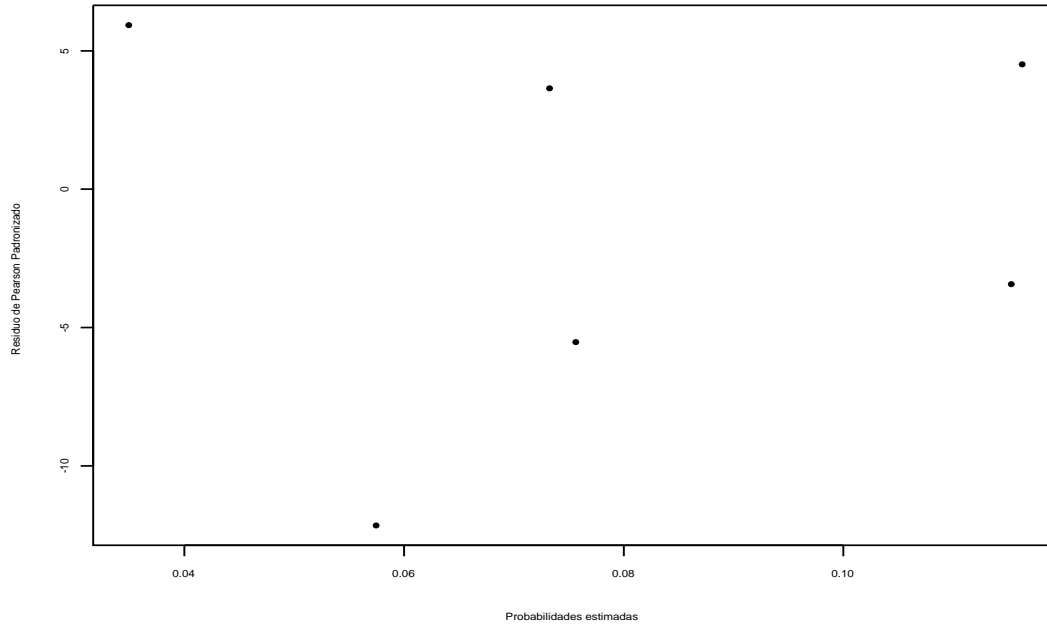


Figura 3.19: Gráfico de $t_{S_i}^{(1)}$ contra as probabilidades estimadas pelo sub-modelo 1.

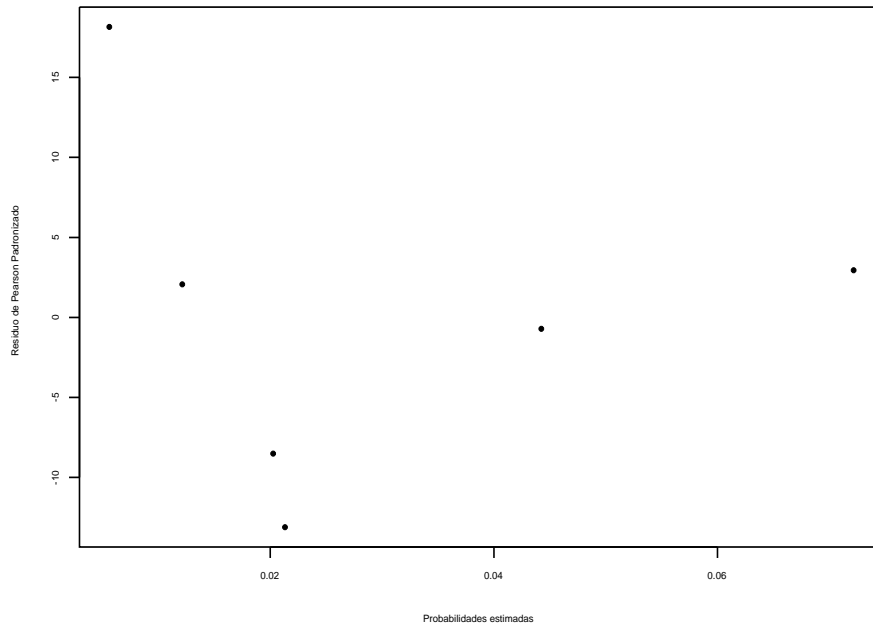


Figura 3.20: Gráfico de $t_{S_i}^{(2)}$ contra as probabilidades estimadas pelo sub-modelo 2.

Capítulo 4

Considerações Finais

Neste trabalho procuramos estudar o modelo logístico trinomial que em princípio pode ser olhado como uma extensão natural do modelo logístico binomial. Porém, como foi visto ao longo do texto e principalmente no capítulo 2 onde discutimos os métodos de diagnóstico, as extensões podem ficar bastante complexas do ponto de vista computacional, sugerindo o uso de adaptações do modelo logístico binomial. Essas adaptações mostraram-se satisfatórias, como vimos na análise dos exemplos no capítulo 3.

Contudo, o desenvolvimento de técnicas mais recentes de diagnóstico, tais como influência local e influência local total, vistas no capítulo 2, que aparentemente não parecem tão complicadas como as técnicas de deleção e alavanca, abrem uma perspectiva de trabalhos futuros com aplicações diretas dessas novas técnicas em modelos logísticos trinomiais em substituição às técnicas adaptadas do modelo logístico binomial.

Apêndice

Apêndice A

Tabela 1

Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
1	0	3	7	0	1	2
2	0	2	11	0	1	3
3	0	3	8	1	1	3
4	1	3	11	0	1	3
5	2	4	7	0	1	3
6	0	3	7	0	1	3
7	2	4	6	0	1	2
8	0	4	6	0	1	3
9	0	2	6	0	1	3
10	1	4	6	0	1	3
11	0	4	8	0	1	2
12	0	3	6	1	0	3
13	0	4	6	0	1	3
14	0	1	5	1	1	3
15	0	2	8	0	0	2
16	0	4	11	0	1	3
17	0	3	6	0	1	3
18	1	4	5	0	1	3
19	0	3	10	0	1	3
20	0	3	10	0	1	3
21	2	3	5	0	1	3
22	0	4	5	0	1	2
23	0	1	5	0	1	3
24	0	2	7 ⁸	0	1	2
25	1	2	9	1	1	2
26	2	4	7	0	1	3
27	0	4	7	0	1	2
28	0	3	10	0	1	3
29	2	3	7	0	1	2
30	0	3	7	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
31	1	4	7	0	1	2
32	2	2	5	0	1	3
33	2	3	8	0	1	3
34	0	1	5	0	1	3
35	0	2	9	0	1	3
36	2	3	11	0	1	3
37	1	3	7	0	1	3
38	2	3	9	1	1	1
39	0	3	5	1	1	3
40	1	3	5	1	1	3
41	0	2	10	0	0	1
42	0	4	5	0	1	3
43	0	3	9	0	1	2
44	0	2	8	0	1	2
45	1	3	9	0	1	3
46	0	3	5	0	1	3
47	0	3	9	0	1	2
48	0	2	9	0	0	1
49	0	1	12	0	0	1
50	0	4	9	0	1	2
51	2	4	5	0	1	3
52	0	2	6	0	1	2
53	1	4	8	0	1	3
54	0	4	7	0	1	3
55	2	3	9	0	1	2
56	1	4	6	0	1	2
57	1	4	6	1	1	2
58	0	3	10	0	1	3
59	0	2	12	0	1	2
60	0	4	6	0	1	3
61	0	1	6	0	0	3
62	0	3	10	0	1	3
63	0	2	12	0	0	3
64	2	4	6	0	1	3
65	0	2	9	0	1	3
66	0	2	9	0	1	3
67	0	3	9	0	0	2
68	0	4	6	0	1	1
69	0	3	8	0	0	3
70	0	3	6	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
71	1	4	5	1	1	3
72	0	3	10	0	0	3
73	0	4	6	0	1	3
74	0	1	6	0	1	2
75	1	4	5	0	1	3
76	1	4	5	0	1	3
77	0	2	6	0	1	2
78	1	3	5	0	1	2
79	0	2	7	0	1	3
80	1	1	6	0	0	2
81	2	4	7	0	1	3
82	1	1	6	0	1	3
83	1	4	7	1	0	3
84	1	4	6	0	1	3
85	0	4	5	0	1	3
86	1	3	9	0	1	3
87	2	2	6	0	1	3
88	0	2	11	0	0	3
89	0	4	9	0	0	3
90	2	1	9	0	0	1
91	0	2	7	0	0	2
92	0	3	10	0	1	3
93	1	4	8	0	1	3
94	1	3	6	0	1	3
95	0	3	7	0	1	3
96	1	3	8	1	1	3
97	0	2	10	0	1	3
98	0	3	7	1	1	3
99	0	1	9	0	1	3
100	2	4	6	0	1	3
101	2	4	8	0	1	3
102	0	2	8	0	1	3
103	0	4	8	0	1	3
104	2	4	6	0	1	3
105	0	2	10	0	1	2
106	2	4	5	0	1	3
107	0	4	5	0	1	3
108	0	3	5	1	1	3
109	1	3	7	0	1	3
110	0	3	10	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico Familiar	Autoexame	Detectar
	mamografia	Sintoma				
111	1	3	5	0	1	3
112	2	3	8	0	1	3
113	0	3	5	0	1	2
114	0	1	13	0	1	1
115	0	2	6	0	1	3
116	2	4	5	0	0	2
117	1	3	6	0	1	3
118	0	4	9	0	1	3
119	0	1	5	0	1	2
120	0	4	6	0	0	3
121	1	4	10	0	1	3
122	1	4	7	0	1	3
123	0	4	7	0	0	3
124	2	3	7	0	1	2
125	0	4	5	0	1	3
126	0	1	5	0	1	3
127	2	2	7	1	1	3
128	2	3	11	0	1	1
129	1	4	6	0	1	3
130	0	1	8	0	1	2
131	0	3	8	0	1	3
132	0	3	9	0	1	3
133	2	3	9	1	1	2
134	1	4	5	0	1	3
135	0	2	7	0	1	3
136	0	1	10	0	1	2
137	1	3	6	0	1	3
138	1	3	10	0	1	3
139	1	3	5	0	1	3
140	0	4	8	0	0	2
141	0	1	9	0	1	2
142	0	3	9	0	1	1
143	1	4	6	0	1	3
144	0	4	11	0	0	3
145	0	2	9	0	0	1
146	0	3	9	0	1	2
147	0	4	12	0	1	2
148	0	1	11	0	1	3
149	1	3	10	0	1	3
150	1	3	9	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
151	0	2	6	0	1	3
152	2	3	5	0	1	3
153	1	3	7	1	1	3
154	1	3	9	0	1	3
155	1	4	6	0	1	3
156	2	3	8	1	1	3
157	0	1	8	0	1	3
158	0	4	5	0	1	3
159	0	3	8	0	1	3
160	0	4	8	0	1	2
161	2	1	10	0	1	2
162	1	4	9	0	1	3
163	0	4	10	0	1	3
164	2	4	9	0	1	3
165	1	4	6	0	1	3
166	0	1	7	0	1	2
167	0	3	5	0	0	3
168	2	3	5	0	1	3
169	2	4	9	0	1	2
170	0	3	9	0	1	3
171	0	1	5	0	1	3
172	2	1	5	0	1	3
173	0	4	6	0	1	3
174	0	3	6	0	1	3
175	1	4	6	0	1	3
176	0	3	10	0	0	3
177	0	4	6	0	1	3
178	1	3	5	1	1	2
179	0	2	9	0	0	3
180	1	3	7	1	1	3
181	0	3	10	0	0	3
182	1	3	8	0	1	2
183	2	2	7	0	1	2
184	0	2	9	0	1	3
185	0	3	13	0	1	2
186	0	4	6	0	0	3
187	0	3	8	0	1	2
188	2	3	10	0	0	3
189	0	2	10	0	1	2
190	0	4	11	0	1	2

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
191	0	4	7	0	1	3
192	2	3	8	1	1	3
193	2	4	10	0	0	3
194	0	3	11	0	1	3
195	2	3	9	1	1	2
196	1	4	5	1	1	3
197	2	3	7	0	1	2
198	1	4	7	0	1	3
199	0	3	9	0	1	3
200	1	3	5	0	1	3
201	0	2	10	0	1	2
202	1	4	5	0	1	3
203	1	3	5	0	1	3
204	1	3	9	0	1	2
205	0	2	7	0	0	2
206	0	1	10	0	1	3
207	1	4	5	1	1	3
208	0	4	6	0	1	2
209	0	3	5	0	1	3
210	1	3	7	0	1	3
211	1	3	8	0	1	2
212	0	2	9	0	1	3
213	0	1	6	0	0	2
214	0	2	13	1	1	3
215	0	2	9	1	1	3
216	2	4	7	0	1	3
217	0	4	5	0	0	3
218	0	1	6	0	1	2
219	0	3	8	0	1	3
220	1	4	7	0	1	3
221	0	3	6	0	1	3
222	0	3	5	0	1	3
223	1	4	5	0	1	3
224	0	3	6	0	1	3
225	0	4	9	0	1	2
226	1	3	8	0	1	3
227	0	3	5	0	1	3
228	2	2	8	0	1	3
229	1	4	6	1	1	3
230	2	4	6	0	1	2

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
231	0	2	8	0	1	2
232	2	1	10	0	1	3
233	0	2	7	0	1	2
234	1	4	5	0	0	3
235	1	3	6	1	1	3
236	0	2	11	0	1	2
237	0	4	5	0	1	3
238	0	4	8	0	1	3
239	0	2	7	0	1	2
240	1	2	7	0	1	3
241	0	3	11	1	0	1
242	0	1	5	0	1	3
243	2	4	5	0	1	3
244	1	4	6	0	1	3
245	0	1	8	0	0	3
246	1	4	5	0	0	3
247	2	3	6	1	1	3
248	2	4	8	0	1	3
249	0	3	6	0	1	3
250	0	1	11	0	1	1
251	0	3	9	0	1	3
252	0	3	10	0	1	3
253	0	1	5	0	1	2
254	1	3	5	0	1	3
255	1	4	5	0	1	3
256	2	4	5	1	1	3
257	0	4	6	0	1	3
258	0	4	7	0	1	3
259	1	3	10	0	1	3
260	2	3	10	0	0	1
261	0	3	8	0	1	3
262	0	4	6	0	1	3
263	2	4	6	0	1	3
264	0	3	10	0	1	3
265	0	4	6	0	1	3
266	0	2	11	0	0	2
267	0	3	9	0	1	1
268	0	3	8	0	1	2
269	1	2	6	0	1	3
270	1	3	9	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
271	1	3	6	1	1	3
272	0	2	10	1	1	3
273	0	3	7	0	1	2
274	1	3	6	0	1	3
275	2	3	8	0	1	3
276	1	4	5	0	1	3
277	0	3	10	0	1	3
278	0	4	8	0	1	3
279	0	2	11	1	1	1
280	0	4	7	0	1	3
281	1	3	9	0	1	3
282	1	4	7	1	1	3
283	0	3	9	0	0	2
284	0	2	9	1	1	2
285	0	3	6	0	1	3
286	0	2	6	0	1	2
287	1	4	5	1	1	3
288	0	3	5	0	0	3
289	0	4	10	0	1	3
290	0	2	9	0	1	3
291	0	3	6	0	1	3
292	0	3	7	0	1	2
293	1	2	10	0	1	1
294	0	3	10	0	1	3
295	0	1	7	0	0	3
296	2	4	7	0	1	3
297	0	4	6	0	1	3
298	0	3	12	0	1	3
299	0	3	8	0	0	3
300	0	2	8	0	1	2
301	0	4	6	0	1	2
302	0	3	10	0	0	3
303	2	3	5	0	1	3
304	0	2	10	0	1	2
305	1	4	8	0	0	3
306	1	4	5	0	1	3
307	0	4	11	0	1	2
308	2	2	10	0	1	3
309	2	3	5	0	1	3
310	2	3	5	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
311	0	3	5	0	1	3
312	1	4	6	0	1	3
313	0	3	7	0	1	2
314	0	3	7	0	1	1
315	0	2	6	1	0	2
316	1	4	5	0	1	3
317	2	4	5	0	1	3
318	1	4	7	1	1	2
319	0	3	10	0	1	3
320	1	4	8	0	1	3
321	1	4	8	1	1	3
322	0	3	9	0	1	3
323	0	2	11	0	0	1
324	1	4	5	0	1	3
325	0	3	10	0	1	3
326	1	4	6	0	1	3
327	1	4	5	0	1	3
328	0	3	10	0	1	2
329	1	3	7	0	1	3
330	1	4	6	0	1	3
331	0	2	10	0	1	3
332	0	2	7	0	1	3
333	2	4	6	0	1	2
334	2	4	8	0	1	3
335	2	4	5	1	1	3
336	2	3	10	1	1	3
337	0	3	6	0	1	3
338	2	4	5	0	1	2
339	0	1	9	0	1	3
340	0	4	6	0	1	2
341	1	3	6	0	1	3
342	0	2	10	0	0	2
343	0	3	5	0	1	3
344	0	3	9	0	0	3
345	0	1	8	0	1	2
346	0	3	9	0	1	2
347	1	4	9	0	1	3
348	2	3	10	0	1	3
349	0	4	10	0	1	3
350	0	3	7	0	1	3

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico		
	mamografia	Sintoma		Familiar	Autoexame	Detectar
351	0	2	6	0	1	3
352	1	3	6	0	1	3
353	0	2	12	0	1	2
354	2	1	7	0	1	3
355	0	4	6	0	0	3
356	0	1	9	0	1	2
357	0	2	8	1	1	3
358	0	2	10	0	1	2
359	0	4	5	0	1	3
360	1	3	6	0	1	3
361	2	3	5	0	1	3
362	0	2	5	0	1	3
363	1	4	9	0	1	3
364	0	1	11	0	0	2
365	0	3	11	0	1	3
366	2	4	8	0	1	3
367	0	1	10	0	1	3
368	2	4	6	0	1	3
369	1	4	5	1	1	3
370	0	4	10	0	1	3
371	1	4	5	0	1	3
372	2	4	6	0	1	3
373	1	4	6	0	1	3
374	0	3	5	0	1	3
375	0	2	11	0	1	2
376	1	3	10	0	1	3
377	1	4	9	0	1	3
378	2	3	6	1	1	3
379	0	3	9	0	1	2
380	0	2	9	0	1	2
381	0	3	9	0	0	3
382	2	3	9	0	1	3
383	2	3	10	0	1	3
384	0	3	9	0	1	3
385	2	3	12	0	1	2
386	0	1	11	0	1	3
387	0	1	5	0	1	2
388	1	4	5	0	1	3
389	2	4	5	0	1	3
390	0	3	17	0	0	2

Continua...

Tabela 1
Dados do Exemplo 1 descrito na seção 1.9.1.

Observação	Experiência		Benefício percebido	Hitórico Familiar	Autoexame	Detectar
	mamografia	Sintoma				
391	1	3	6	0	1	3
392	0	2	6	0	1	3
393	0	2	10	0	1	2
394	1	3	5	0	1	3
395	0	3	5	0	1	3
396	0	1	5	0	1	2
397	0	2	10	0	0	2
398	0	4	6	0	1	2
399	2	3	6	0	1	2
400	0	2	7	0	1	2
401	0	2	8	0	0	3
402	1	3	9	0	1	3
403	1	4	5	0	1	3
404	0	2	12	0	1	2
405	0	3	10	0	1	2
406	2	3	5	0	1	3
407	0	3	11	0	1	2
408	1	3	10	0	1	2
409	0	4	8	0	1	3
410	1	4	6	0	1	3
411	2	2	6	0	1	3
412	0	4	7	0	1	3

Tabela 2
Dados do Exemplo 2 descrito na seção 1.9.2.

Idade em anos	Status	Teste			Total
		Normal	Indefinido	Anormal	
menos de 40	Nunca fumou	577	27	7	611
menos de 40	Ex-fumante	192	20	3	215
menos de 40	Fumante	682	46	11	739
40 - 59	Nunca fumou	164	4	0	168
40 - 59	Ex-fumante	145	15	7	167
40 - 59	Fumante	245	47	27	319
	Total	2005	159	55	2219

Apêndice B

Neste apêndice apresentamos os programas em S-plus desenvolvidos para geração dos gráficos do exemplo 1 exibidos no trabalho.

Os sub-modelos foram gerados pelo programa SAS, uma vez que o S-plus não funciona para para modelos trinomias.

Passo 1.

Geração do sub-modelos e estimação das probabilidades - feito no SAS

Passo 2.

Gerão dos gráficos. De posse das estimativas das probabilidades estimadas para os dois sub-modelos, partimos para geração dos gráficos.

Passo 3.

Leitura dos dados.

Matriz planejamento, respostas e probabilidades estimadas

`graficos.dat=scan("data.dat",`

`what=list(um=0,x1=0,x2=0,x3=0,x4=0,x5=0,y0=0,y1=0,y2=0,pi0=0,pi1=0,pi2=0))`


```
attach(graficos.dat)
X=cbind(um,x1,x2,x3,x4,x5)
```

Grafico de Alavanca - binomial

```
v1=pi1*(1-pi1)
w1=diag(v1)
H1=sqrt(w1)%*%X%*%solve(t(X)%*%w1%*%X)%*%t(X)%*%sqrt(w1)
hii1=diag(H1)
```

Gráfico de Alavanca - trinomial

```
zero=matrix(c(0),nrow=412,ncol=6)
s1=cbind(X,zero)
s2=cbind(zero,X)
sgrande=rbind(s1,s2)
v1=pi1*(1-pi1)
v2=pi2*(1-pi2)
v3=-pi1*p12
v1d=diag(v1)
v2d=diag(v2)
v3d=diag(v3)
vint1=cbind(v1d,v3d)
vint2=cbind(v3d,v2d)
vgrande=rbind(vint1,vint2)
mgrande=sgrande%*%
solve(t(sgrande)%*%vgrande%*%sgrande)%*%t(sgrande)%*%vgrande
mii=diag(mgrande)
mii1=array(,412)
```

```
k=1
while(k<=412)
mii1[k]=mii[k]
k=k+1
```

Gráfico Afastamento da verossimilhança - binomial

```
res1=y1-pi1
corrp1=sqrt(pi1*(1-pi1)*(1-hii1))
tS1=res1/corrp1
LDi1=(hii1/(1-hii1))*(tS1*tS1)
```

Gráfico Influencia local - trinomial

```
v1=pi1*(1-pi1)
v2=pi2*(1-pi2)
v3=-pi1*pi2
v1d=diag(v1)
v2d=diag(v2)
v3d=diag(v3)
vint1=cbind(v1d,v3d)
vint2=cbind(v3d,v2d)
vzona=rbind(vint1,vint2)
zero=matrix(c(0),nrow=412,ncol=6)
s1=cbind(X,zero)
s2=cbind(zero,X)
szona=rbind(s1,s2)
da=diag(y1-pi1)
db=diag(y2-pi2)
```

```

d1=t(X)%*%da
d2=t(X)%*%db
delta=rbind(d1,d2)
B=t(delta)%*%solve(t(szona)%*%vzona)%*%szona)%*%delta
l=eigen(B)vectors
ll=Mod(l)
lmax=array(,412)
k=1
while(k<=412)
lmax[k]=ll[k,1]
k=k+1

```

Gráfico Influencia local - binomial

```

w1_i=diag(v1)
H1_i=sqrt(w1)%*%X)%*%solve(t(X)%*%w1)%*%X)%*%t(X)%*%sqrt(w1)
r1=((y1-pi1)/sqrt(pi1)%*%(1-pi1)))
B1=diag(r1)%*%H1)%*%diag(r1)
B11=eigen(B1)vectors
B111=Mod(B11)
C11=array(,412)
k=1
while(k<=412)
C11[k]=B111[k,1]
k=k+1

```

Gráfico Influencia local total - trinomial

```

C=array(,412)

```

```

k=1
while(k≤412) C[k]=2*Mod(B[k,k])
k=k+1
Cbarra=2*mean(C)

```

Gráfico Influencia local total - binomial

```

v1=pi1*(1-pi1)
v2=pi2*(1-pi2)
v3=-pi1*p12
w1j=diag(v1)
H1j=sqrt(w1)%*%X%*%solve(t(X)%*%w1%*%X)%*%t(X)%*%sqrt(w1) h1j=diag(H1)
w2j=diag(v2)
r1=((y1-pi1)/sqrt(pi1%*%(1-pi1)))
Ci1=array(,412)
k=1
while(k≤412)
Ci1[k]=2*(r1[k]%*%r1[k]%*%h1[k])
k=k+1
C1barra=2*mean(C1)

```

Gráfico Componente do desvio - trinomial

```

v1=pi1*(1-pi1)
v2=pi2*(1-pi2)
v3=-pi1*p12
w1=diag(v1)
H1=sqrt(w1)%*%X%*%solve(t(X)%*%w1%*%X)%*%t(X)%*%sqrt(w1)
h1=diag(H1)

```

```

w2=diag(v2)
H2=sqrt(w2)%*%X%*%solve(t(X)%*%w2%*%X)%*%t(X)%*%sqrt(w2)
h2=diag(H2)
lp0=sqrt(-2*log(pi0))
lp1=sqrt(-2*log(pi1))
lp2=sqrt(-2*log(pi2))
d0=lp0*y0
d1=lp1*y1
d2=lp2*y2
d=d0+d1+d2
hmedio=sqrt(1-(h1+h2)/2)
td=sort(d/hmedio)

```

Gráfico de Envelope - trinomial

O gráfico de foi gerado para os resíduos componente do desvio por simulação, após gerar 20 amostras aleatórias de Y_{ij} , $i = 0, 1, 2$ e $j = 1, \dots, 412$, estimamos as estimativas de π_1 e π_2 para as 20 amostras no SAS, a seguir transportamos essas estimativas para o S-plus. A partir desse ponto é só proceder como se fosse gerar 20 gráficos componente do desvio, após gerar os 20 vetores poderemos gerar o envelope como segue.

```

tdi=cbind(td1,td2,...,td20)
TD=t(tdi)
v=array(,412)
ve=array(,412)
vet=array(,412)
Med=array(,412)
k=1
PQ=array(,412)

```

```
TQ=array(,412)
while(k≤412)
l=1
while(l≤20)
v[l]=TD[l,k]
l=l+1
ve=sort(v)
Med[k]=median(ve)
PQ[k]=ve[2]
TQ[k]=ve[19]
k=k+1
faixa=range(td,PQ,TQ) - td é o vetor de componente desvio original
par(pty="s")
qqnorm(td,xlab="Percentis da N(0,1)",ylab="Componente do Desvio",ylim=faixa)
par(new=2)
qqnorm(PQ,axes=F,xlab=,ylab=,main=,type="l",ylim=faixa,lty=1)
par(new=2)
qqnorm(TQ,axes=F,xlab=,ylab=,main=,type="l",ylim=faixa,lty=1)
par(new=2)
qqnorm(Med,axes=F,xlab=,ylab=,main=,type="l",ylim=faixa,lty=2)
```

Bibliografia

- Agresti A. (1990). *Categorical Data Analysis*. John Wiley, New York.
- Atkinson, A. C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika* **68**, 13-20.
- Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford Statistical Science Series, Oxford.
- Beckman, R. J.; Nachtsheim, C. J. e Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics* **29**, 413-426.
- Belsley, D. A.; Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics*. John Wiley, New York.
- Chatterjee, S. e Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. New York: Wiley.
- Chatterjee, S. e Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 379- 416.
- Collett, D. (1991). *Modelling Binary Data*. Chapman and Hall, London.
- Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics* **19**, 15-18.

- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* **48**, 133-169.
- Cook, R. D.; Peña, D. e Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics, Theory and Methods* **17**, 623-640
- Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Cordeiro, G. M. e Paula, G. A. (1992). Estimation, large-sample parametric tests and diagnostics for non-exponential family nonlinear models. *Communications in Statistics, Simula* **21**, 149-172.
- Cox, D. R. e Snell, E. J. (1968). A General definition of residuals (with discussion). *Journal of the Royal Statistical Society B* **30**, 248-275.
- Cox, D. R. e Snell, E. J. (1989). *The Analysis of Binary Data, 2nd Edition*. Chapman and Hall, London.
- Davison, A. C. e Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika* **76**, 211-221.
- de Souza, F. A. M. e Paula, G. A. (2002). Assessing local influence in mixed von Mises regression models. *Australian and New Zealand Journal of Statistics* **44**, 345-356.
- Davison, A.C. e Tsai, C-L. (1992). Regression model diagnostics. *International Statistical Review* **60**, 337-353.
- Escobar, L. A. e Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507-528.

- Fahrmeir, L. e Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer, New York.
- Forthofer, R. N. e Lehnen, R. G. (1981). Public program analysis, A new categoric data approach. *Lifetime Learning Publications* **16**, 21.
- Fung, W. K. (1993). Unmasking outliers and leverage points: a confirmation. *Journal of the American Statistical Association* **88**, 515-519.
- Fung, W. K. e Kwan, C. W. (1997). A note on local influence based on normal curvature. *Journal of the Royal Statistical Society* **B 59**, 839-843.
- Galea, M; Paula, G. A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71-79.
- Galea, M; Paula, G. A. e Uribe-Opazo, M. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers* **44**, 23-45.
- Galea, M.; Riquelme, M. e Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics* **14**, 167-184.
- Gu, H. e Fung, W. K. (1998). Assessing local influence in canonical correlation analysis. *Annals of the Institute of Statistical Mathematics* **50**, 755-772.
- Kwan, C. W. e Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Applications to factor analysis. *Psychometrika* **63**, 35-46.
- Lesaffre, F. E. e Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**, 579-582.
- Hoaglin, D. C. e Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician* **32**, 17-22.

- Hosmer, D. W. e Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley, New York.
- Kim, M. G. (1995). Local influence in multivariate regression. *Communications in Statistics, Theory and Methods* **20**, 1271-1278.
- Lawrence, A. J. (1988). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association* **84**, 124-141.
- Liu, S. Z. (2000). On local influence for elliptical linear models. *Statistical Papers* **41**, 211-224.
- McCullagh, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models, 2nd. Edition*. Chapman and Hall, London.
- Moolgavkar, S. H., Lustbader, E. D. e Venzon, D. J. (1984). A geometric approach to non-linear regression diagnostics with application to matched case-control studies. *Annals of Statistics* **12**, 816-826.
- O'Hara Hines, J.; Lawless, J. F. e Carter, E. M. (1992). Diagnostics for a cumulative multinomial generalized linear model with application to grouped toxicological mortality data. *Journal of the American Statistical Association* **87**, 1059-1069.
- Ortega, E. M. M.; Bolfarine, H. e Paula, G. A. (2003). Influence diagnostic in generalized log-gamma regression models. *Computational Statistics and Data Analysis* **42**, 165-186.
- Pan, J. X.; Fang, K. T. e von Rosen, D. (1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference* **62**, 263-278.

- Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* **16**, 63-79.
- Paula, G. A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation* **51**, 315-352.
- Paula, G. A. (1996). Influence diagnostic in proper dispersion models. *Australian Journal of Statistics* **38**, 307-316.
- 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, 300 p., IME-USP.
- Paula, G. A. e Peres, C. A. (1988). Diagnostics for GLMs with linear inequality parameter constraints. *Communications in Statistics, Theory and Methods* **17**, 4205-4219.
- Peña, D. e Yohai, V. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association* **94**, 434-445.
- Pettitt, A. N. e Bin Daud, I. (1989). Case-weight measures of influence for proportional hazards regression. *Applied Statistics* **38**, 51-67.
- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics* **9**, 705-724.
- Ross, W. H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *Canadian Journal of Statistics* **15**, 91-103.
- Silva, G. L. (1992). *Modelos Logísticos para Dados Binários*. Dissertação de Mestrado, IME-USP.
- St. Laurent, R. T. e Cook, R. D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, **87**, 985-990.

- Svetliza, C. F. e Paula, G. A. (2003). Diagnostics in nonlinear negative binomial models. *Communications in Statistics, Theory and Methods* **32**, 1227-1250.
- Tsai, C. H. e Wu, X. (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics and Probability Letters* **14**, 247-252.
- Wei, B.C., Hu, Y.Q. e Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25-37.
- Williams, D. A. (1984). Residuals in generalized linear models. *In: Proceedings of the 12th. International Biometrics Conference*, Tokyo, pp. 59-68.
- Williams, D. A. (1987). Generalized linear model diagnostic using the deviance and single case deletion. *Applied Statistics* **36**, 181-191.