
Aula - Modelos Lineares Generalizados

- Nelder e Wedderburn (1972), propuseram os Modelos Lineares Generalizados (MLGs), que são uma extensão dos modelos normais lineares.
- A idéia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta (μ) e o preditor linear η .

-
- A ligação entre a média e o preditor linear não é necessariamente a identidade, podendo assumir qualquer forma monótona não-linear.
 - Nelder e Wedderburn propuseram também um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio que tem sido largamente utilizado na avaliação da qualidade do ajuste dos MLGs, bem como no desenvolvimento de resíduos e medidas de diagnóstico.
 - Estudos abrangentes sobre MLGs são encontrados nos livros de McCullagh e Nelder (1989), Cordeiro (1986), Dobson (1990) e Paula (2004).

Definição

Consideremos n variáveis aleatórias independentes y_1, \dots, y_n , cada uma com função densidade (ou de probabilidade) na família exponencial da forma

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

onde $b(\cdot)$ e $c(\cdot)$ são funções conhecidas. Para o modelo em (1) valem as seguintes relações:

$$E(y_i) = \mu_i = b'(\theta_i), \quad \text{Var}(y_i) = \phi^{-1}V_i, \quad i = 1, \dots, n$$

sendo ϕ^{-1} o parâmetro de dispersão e $V = d\mu/d\theta$ a função de variância (caracteriza a distribuição).

Tabela 1: Características de algumas distribuições da família exponencial

Distribuição	Normal	Poisson	Binomial
Notação	$N(\mu, \phi^{-1})$	$P(\mu)$	$B(n, \mu)$
Suporte de y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)n}{n}$
$c(y, \phi)$	$-\frac{1}{2}(\phi y^2 + \log \frac{2\pi}{\phi})$	$-\log(y)!$	$\binom{\log n}{ny}$
$b(\theta)$	$\theta^2/2$	e^θ	$\log(1 + e^\theta)$
$\mu = E(y)$	θ	e^θ	$e^\theta / 1 + e^\theta$
$V(\mu)$	1	μ	$\mu(1 - \mu)$

No modelo binomial, a variável aleatória corresponde à proporção de sucessos em n ensaios de Bernoulli e $\phi = n$.

Fonte : McCullagh e Nelder (1989; Tabela 2.1, pg. 30).

Tabela 2: Características de algumas distribuições da família exponencial

Distribuição	Gama	Normal Inversa
Notação	$G(\mu, \phi)$	$N^-(\mu, \phi)$
Suporte de y	$(0, \infty)$	$(0, \infty)$
$c(y, \phi)$	$(\phi - 1) \log(y\phi) + \log \phi - \log \Gamma(\phi)$	$\frac{1}{2}(\log \frac{\phi}{2\pi y^3} - \frac{\phi}{y})$
$b(\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$\mu = E(y)$	$-1/\theta$	$-(-2\theta)^{1/2}$
$V(\mu)$	μ^2	μ^3

A parametrização do modelo gama é tal que a sua variância é dada por μ^2/ϕ .

Fonte : McCullagh e Nelder (1989; Tabela 2.1, pg. 30).

Modelo Linear Generalizado

Os MLGs são definidos por (1) e pela componente sistemática

$$g(\mu_i) = \eta_i, \quad i = 1, \dots, n \quad (2)$$

onde $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação,

- $\eta = X\beta$,
- $\beta = (\beta_1, \dots, \beta_p)^\top$ ($p < n$),
- $X = (x_1, \dots, x_p)$,
- $x_j = (x_{1j}, \dots, x_{nj})^\top$.

Consideremos um MLG definido por (1) e (2). O logaritmo da função de verossimilhança como função de β pode ser expresso na forma

$$L(\beta; y) = \sum_{i=1}^n \phi[y_i \theta_i - b(\theta_i) + c(y_i)] + a(y_i, \phi).$$

• $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$ para $i = 1, \dots, n$

$$L(\beta; y) = \sum_{i=1}^n \phi\left[y_i \sum_{j=1}^p x_{ij} \beta_j - b\left(\sum_{j=1}^p x_{ij} \beta_j\right)\right] + \phi \sum_{i=1}^n c(y_i) + \sum_{i=1}^n a(y_i, \phi)$$

que podemos reexpressar na forma

$$L(\beta; y) = \sum_{j=1}^p S_j \beta_j - \phi \sum_{i=1}^n b\left(\sum_{j=1}^p x_{ij} \beta_j\right) + \phi \sum_{i=1}^n c(y_i) + \sum_{i=1}^n a(y_i, \phi),$$

onde $S_j = \phi \sum_{i=1}^n y_i x_{ij}$.

- $S = (S_1, \dots, S_p)$ é suficiente minimal (Dudewicz e Mishra, 1988, Cap. 8) para o vetor $\beta = (\beta_1, \dots, \beta_p)^\top$.
- As ligações que fornecem estatísticas suficientes são denominadas canônicas.

Ligações canônicas

Para os modelos normal, Poisson, binomial, gama e normal inverso as ligações canônicas são dadas por

$$\eta = \mu, \quad \eta = \log \mu, \quad \eta = \log \left[\frac{\mu}{1 - \mu} \right], \quad \eta = \mu^{-1} \text{ e } \eta = \mu^{-2}.$$

- garantem a concavidade de $L(\beta; y)$, isto é, garantem a unicidade da estimativa de máxima verossimilhança de β , quando essa existe e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente.
- Para ligações não-canônicas Wedderburn (1976) discute condições à existência da concavidade $L(\beta; y)$.

Outras ligações

- *Potência*: $\eta = \mu^\lambda$, onde λ é um número real;
- *probit*: $\eta = \Phi^{-1}(\mu)$;
- *Logística*: $\eta = \log[\mu/(1 - \mu)]$;
- *Complemento log-log*: $\eta = \log[-\log(1 - \mu)]$;
- *Logaritmo*: $\eta = \log \mu$.
- *Box-Cox*: $\eta = (\mu^\lambda - 1)/\lambda$;
- *Aranda-Ordaz*: $\eta = \log\{(1 - \mu)^{-\alpha} - 1/\alpha\}$.

-
- Os MLGs são ajustados no R através da função *glm*, onde devemos especificar *formula* (a definição do modelo) e *family* (a distribuição assumida pela variável resposta com a função de ligação a ser usada). Por exemplo,
$$\text{fit.reg} = \text{glm}(y \sim 1 + x, \text{family} = \text{gaussian}).$$
 - Se a função de ligação usada for diferente do ‘default’, basta especificar a função de ligação desejada através do comando *link*. Por exemplo,
$$\text{fit.reg} = \text{glm}(y \sim 1 + x, \text{family} = \text{gaussian}(\text{link} = \text{“log”})).$$
 - O comando *summary(fit)* dá um resumo do resultado do ajuste.

Função desvio

Sem perda de generalidade, suponha que o logaritmo da função de verossimilhança seja agora definido por

$$L(\mu; y) = \sum_{i=1}^n L(\mu_i; y_i),$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = x_i^\top \beta$. Para o modelo saturado ($p = n$) a função $L(\mu; y)$ é estimada por

$$L(y; y) = \sum_{i=1}^n L(y_i; y_i).$$

Ou seja, a estimativa de máxima verossimilhança de μ_i fica nesse caso dada por $\mu_i^0 = y_i$.

Quando $p < n$, denotaremos a estimativa de $L(\mu; y)$ por $L(\hat{\mu}; y)$. Aqui, a estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ em que $\hat{\eta}_i = x_i^\top \hat{\beta}$.

A qualidade do ajuste de um MLG é avaliada através da função desvio dada por

$$D^*(y; \hat{\mu}) = \phi D(y; \hat{\mu}) = 2\{L(y; y) - L(\hat{\mu}; y)\}$$

onde

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\tilde{\theta}_i) - b(\hat{\theta}_i))],$$

denotando por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$, respectivamente, as estimativas de máxima verossimilhança de θ_i para os modelos com p parâmetros ($p < n$) e saturado ($p = n$).

Apresentaremos a seguir a função desvio dos casos especiais citados anteriormente.

● *Normal* :

$$D(y; \hat{\mu}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2,$$

que coincide com a soma de quadrados dos resíduos.

● *Binomial* :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{n_i \hat{\mu}_i} \right) + (n_i - y_i) \log \left(\frac{1 - \frac{y_i}{n_i}}{1 - \hat{\mu}_i} \right) \right].$$

Todavia, quando $y_i = 0$ ou $y_i = n_i$, o i -ésimo termo de $D(y; \hat{\mu})$ se iguala a $-2n_i \log(1 - \hat{\mu}_i)$ ou $-2n_i \log \hat{\mu}_i$,

respectivamente.

● *Gama* :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right].$$

● *Poisson* :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right].$$

● *Normal Inversa* :

$$D(y; \hat{\mu}) = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}.$$

-
- Um valor pequeno para a função desvio indica que, para um menor número de parâmetros, obtém-se um ajuste tão bom quanto o ajuste com o modelo saturado.
 - Embora seja usual comparar os valores observados da função desvio com os percentis da distribuição qui-quadrado com $n - p$ graus de liberdade, em geral, $D(y; \hat{\mu})$ não segue assintoticamente uma distribuição χ_{n-p}^2 .
 - Poisson: quando $\mu_i \longrightarrow \infty$, para todo i , tem-se $D(y; \hat{\mu}) \sim \chi_{n-p}^2$.
 - Normal: como é conhecido para σ^2 fixo, $D(y; \hat{\mu}) \sim \sigma^2 \chi_{n-p}^2$.

Lembre que $E(\chi_r^2) = r$, assim, um valor do desvio próximo de $n - p$ pode ser uma indicação de que o modelo estará bem ajustado. Geralmente, para os casos em que $D^*(y; \hat{\mu})$ depende do parâmetro de dispersão ϕ^{-1} , o seguinte resultado (Jørgensen, 1987) pode ser utilizado:

$$D^*(y; \hat{\mu}) \sim \chi_{n-p}^2, \quad \text{quando } \phi \longrightarrow \infty.$$

- Quando a dispersão é pequena, é razoável comparar os valores observados de $D^*(y; \hat{\mu})$ com os percentis da distribuição χ_{n-p}^2 .

Análise do desvio

Suponha para o vetor de parâmetros β a partição

$\beta = (\beta_1^\top, \beta_2^\top)^\top$, em que β_1 é um vetor q -dimensional enquanto β_2 tem dimensão $p - q$.

- Portanto podemos estar interessados em testar as hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ contra $H_1 : \beta_1 \neq \beta_1^{(0)}$.
- As funções desvio correspondentes aos modelos sob H_0 e H_1 são dadas por $D(y; \mu^{(0)})$ e $D(y; \hat{\mu})$, respectivamente, onde $\mu^{(0)}$ é a estimativa de máxima verossimilhança de μ sob H_0 .

Análise do desvio

A análise de desvio (*ANODEV*) é uma generalização da análise de variância para os MLGs.

- Podemos definir a seguinte estatística

$$F = \frac{\{D(y; \mu^{(0)}) - D(y; \hat{\mu})\} / q}{D(y; \hat{\mu}) / (n - p)},$$

cuja distribuição nula assintótica é $F_{q, (n-p)}$.

- Não depende de ϕ e é invariante sob reparametrização
- Pode ser obtida diretamente de funções desvio, é muito conveniente para uso prático.

-
- Através do comando *anova*, o R fornece uma tabela *ANODEV* para os ajustes colocados como objetos (ajustes de um MLG). Por exemplo, suponha que os objetos *fit.reg*, *fit1.reg* correspondam aos ajustes de um MLG com um, dois fatores, respectivamente. Então, o comando

`anova(fit.reg, fit2.reg, test = "Chi")`

fornece uma tabela comparando os três fatores.

Função escore e matriz de informação

O logaritmo da função de verossimilhança de um MLG definido por (1) e (2) pode ser expresso na forma

$$L(\beta; y) = \sum_{i=1}^n \phi[y_i \theta_i - b(\theta_i) + c(y_i)] + a(y_i, \phi).$$

A função escore total e a matriz de informação total de Fisher para o parâmetro β são dadas por

$$U(\beta) = \frac{\partial L(\beta; y)}{\partial \beta} = \phi X^\top W^{\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu),$$

e

$$K(\beta) = \mathbb{E} \left\{ - \frac{\partial^2 L(\beta; y)}{\partial \beta \partial \beta^\top} \right\} = \phi X^\top W X,$$

Estimação dos parâmetros

em que X é a matriz modelo

$$W = \text{diag}(w_1, \dots, w_n) \quad \text{com} \quad w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{V_i}$$

$$V = \text{diag}(V_1, \dots, V_n), \quad \text{com} \quad V_i = \frac{d\mu_i}{d\theta_i}.$$

Para a obtenção da estimativa de máxima verossimilhança de β utilizamos o processo iterativo de Newton-Raphson, que pode ser reescrito como um processo iterativo de mínimos quadrados reponderados dado por

$$\beta^{(m+1)} = (X^\top W^{(m)} X)^{-1} X^\top W^{(m)} z^{(m)}, \quad (3)$$

Estimação dos parâmetros

$m = 0, 1, \dots$, onde $z = \eta + W^{-\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu)$.

- Observe que z faz o papel de uma variável dependente modificada, enquanto que W é uma matriz de pesos que muda a cada passo do procedimento iterativo.
- A convergência de (3) ocorre em geral em um número finito de passos, independente dos valores iniciais utilizados (Wedderburn, 1976). É usual iniciar (3) com $\eta_i^{(0)} = g(y_i)$, para $i = 1, \dots, n$.
- $\hat{\beta} = (X^\top X)^{-1} X^\top y$.

Sob condições de regularidade (Sen e Singer, 1993, Cap. 7), mostra-se que $\hat{\beta}$ é um estimador consistente e eficiente de β e que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma^{-1}(\beta)),$$

onde

$$\Sigma(\beta) = \lim_{n \rightarrow \infty} \frac{K(\beta)}{n},$$

sendo $\Sigma(\beta)$ uma matriz positiva definida e

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N\left(0, \Sigma_{(\cdot)}^{-1}(\phi)\right), \text{ conforme } n \longrightarrow \infty,$$

em que $\sigma_{(\phi)}^2 = \lim_{n \rightarrow \infty} -n \left\{ \sum_{i=1}^n c''(y_i, \phi) \right\}^{-1}$.

A estimação do parâmetro ϕ , quando o mesmo é desconhecido, pode ser vista em Cordeiro e McCullagh (1991).

- No caso em que ϕ é desconhecido, podemos observar que os parâmetros β e ϕ são ortogonais, isto é,

$$E \left[\left(\frac{\partial L(\beta, \phi; y)}{\partial \beta} \right) \left(\frac{\partial L(\beta, \phi; y)}{\partial \phi} \right) \right] = 0.$$

Logo, os estimadores de máxima verossimilhança de ϕ e β são assintoticamente independentes.

As estimativas de máxima verossimilhança para ϕ nos casos normal e normal inverso são dadas por $\hat{\phi} = n/D(y; \hat{\mu})$.

Para o caso da distribuição gama, a estimativa de máxima verossimilhança de ϕ é dada por

$$\hat{\phi} = \frac{1 + (1 + 2\bar{D}/3)^{1/2}}{2\bar{D}}, \quad \text{onde } \bar{D} = D(y; \hat{\mu})/n.$$

$$\tilde{\phi}^{-1} = \sum_{i=1}^n \{(y_i - \tilde{\mu}_i)/\hat{\mu}_i\}^2 / (n - p).$$

Para encontrar a EMV de ϕ com o respectivo desvio padrão aproximado deve-se usar os comandos

```
library(mass)
```

```
gamma.shape(fit)
```

Testes de hipóteses

A estatística da razão de verossimilhança para o teste de H_0 pode ser escrita da seguinte forma

$$LR = 2\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1^{(0)}, \tilde{\beta}_2)\}, \quad (4)$$

onde $\hat{\beta}$ é o EMV irrestrito de β e $\tilde{\beta}_2$ é o EMV de β_2 sob H_0 , isto é, sob o modelo com componente sistemática $\eta = \eta_1^{(0)} + \eta_2$

com

$$\eta_1^{(0)} = \sum_{j=1}^q x_j \beta_{1j}^{(0)} \quad \text{e} \quad \eta_2 = \sum_{j=q+1}^p x_j \beta_{2j}.$$

- $LR = \phi\{D(y; \hat{\mu}^0) - D(y; \hat{\mu})\}$ em que $\hat{\mu}^0 = g^{-1}(\hat{\eta}^0)$ e $\hat{\eta}^0 = X\beta_1^0$.
- Assintoticamente e sob H_0 , temos que $LR \sim \chi_q^2$.

-
- Para calcular a estatística da razão de verossimilhança no R , basta fazer a diferença dos desvios correspondentes aos modelos sob H_0 e H_1 .

Objetivos principais:

1) Verificar se há afastamentos sérios das suposições feitas para o modelo.

- se há afastamento da suposição da distribuição da variável resposta;
- ausência de alguma variável explicativa ou termos (quadrático, cúbico) de variáveis incluídas no modelo;
- se há indícios de correlação entre as observações.

2) Detectar observações atípicas que destoam do conjunto.

Essas observações são classificadas em três grupos:

- **aberrantes**, mal ajustadas com resíduos altos;
- **de alavanca**, posicionadas em regiões remotas com alta influência no próprio valor ajustado;
- **influentes**, com influência desproporcional nas estimativas dos coeficientes.

Uma observação pode ser classificada em mais de um grupo.

Alavanca

Pontos de alavanca são aqueles que têm uma influência desproporcional no próprio valor ajustado. Temos que o processo iterativo dos MLGs, na convergência, é dado por

$$\hat{\beta} = (X^T \hat{W} X)^{-1} X^T \hat{W} z, \text{ com } z = \hat{\eta} + \hat{W}^{-1/2} \hat{V}^{-1/2} (y - \hat{\mu}),$$

isto é, a solução de M.Q.P. da regressão z contra X com pesos W . Então H é a matriz de projeção

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$$

Pregibon (1981) sugere usar os elementos da diagonal (h_{ii}) de H e analisar os pontos que se destacam.

Resíduo padronizado

Similar ao modelo normal, temos que o resíduo ordinário pode ser definido como

$$r^* = \hat{W}^{1/2}(z - \hat{\eta})$$

Se assumirmos que $Var(z) \approx \hat{W}^{-1}\phi^{-1}$ temos

$$Var(r^*) \approx \phi^{-1}(I - \hat{H}).$$

A fim de comparar os resíduos deve-se padronizá-los, definindo

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

Williams (1984) mostra via simulação que t_{S_i} em geral é assimétrica.

Resíduo Componente do desvio

Resíduo componente do desvio é mais utilizado (McCullagh, 1987, Davison e Gigli, 1989) e é dado por

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

em que

$$d(y_i; \hat{\mu}_i) = \text{sinal}(y_i - \hat{\mu}_i) \sqrt{2} \{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\hat{\theta}_i^0))\}^{1/2}.$$

William (1984) verificou que a distribuição t_{D_i} está mais próxima da normal. Assim, para grandes amostras, são considerados marginalmente aberrantes pontos tais que

$$|t_{D_i}| > 2$$

Alguns gráficos de diagnóstico envolvendo resíduos:

- gráfico de t_{D_i} contra a ordem das observações ;
- gráfico de t_{D_i} contra \hat{y}_i para detectar indícios de parâmetro de dispersão não constante, correlação entre as observações e pontos aberrantes;
- gráfico de t_{D_i} contra variáveis explicativas para detectar ausência de termos na parte sistemática;
- envelope, banda de confiança empírica para detectar afastamentos da distribuição postulada.

Influência

Pontos influentes são aqueles com influência desproporcional nas estimativas dos coeficientes, isto é, quando retirados do modelo mudam de forma substancial as estimativas ou mesmo a significância dos coeficientes.

O método mais conhecido para detectar tais pontos é o de deleção de pontos, que consiste em retirar um ponto e verificar as variações nas estimativas.

Afastamento da verossimilhança

Uma medida de influência mais conhecida para avaliar o impacto em $L(\hat{\beta})$ é o afastamento da verossimilhança (*likelihood displacement*) definida por

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

em que $\hat{\beta}_{(i)}$ é a estimativa de β com a exclusão da i -ésima observação. Então usando uma versão aproximada (Pregibon, 1982), temos

$$LD_i \approx \left(\frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right) t_{S_i}^2,$$

Recomenda-se como critério olhar com mais atenção aqueles pontos com LD_i muito maior que os demais.

Análise confirmatória

Uma vez detectados os pontos aberrantes, de alavanca e influentes deve-se ainda fazer uma análise confirmatória com os pontos mais destacados.

Essa análise consiste em retirar cada pontos dos dados, reajustar o modelo e verificar quanto mudam as estimativas.

No entanto, deve-se retirá-los apenas em último caso. Por exemplo, se nenhum método alternativo de acomodá-los no modelo for bem sucedido.

Exemplo 1 - Binomial

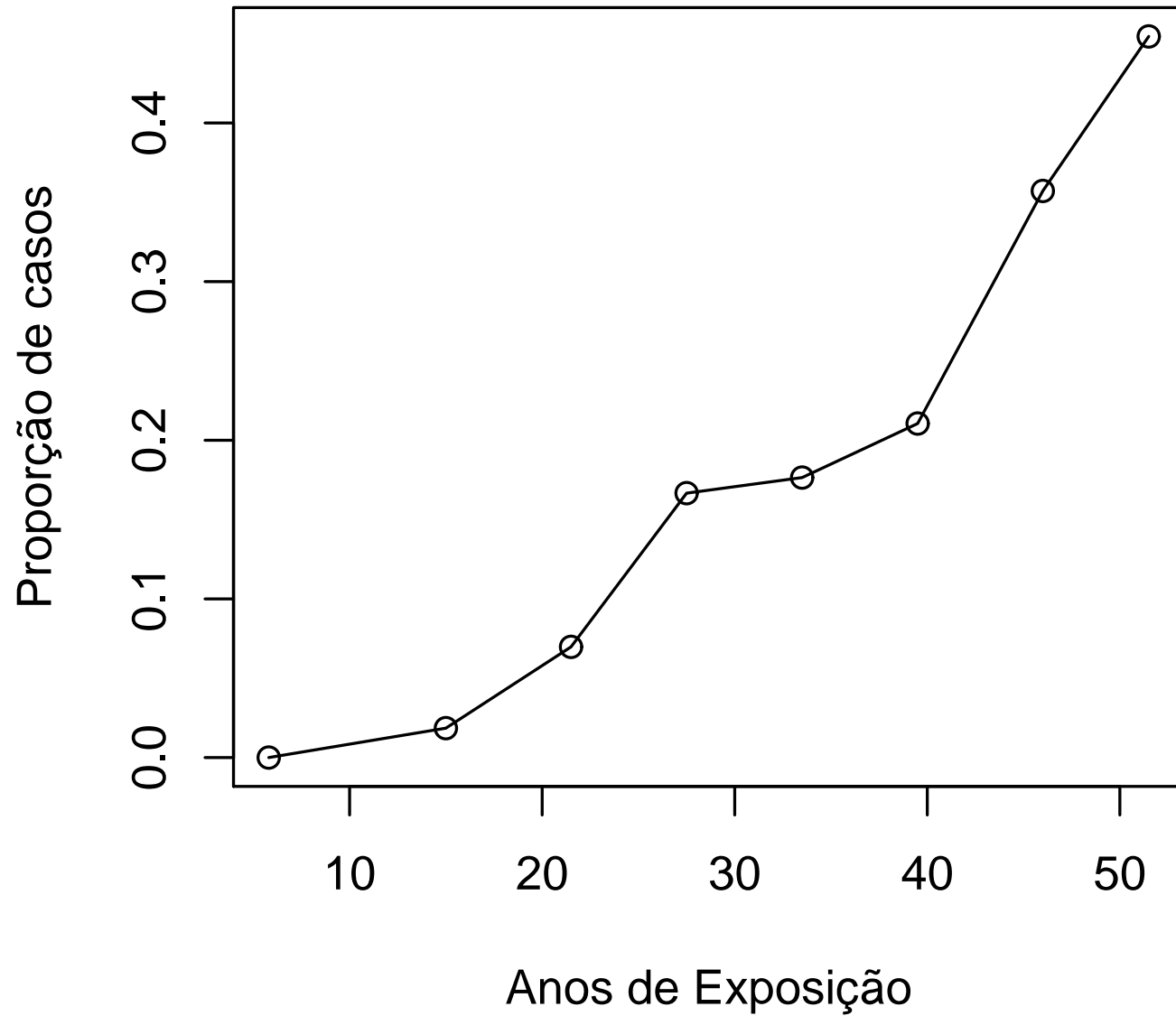
A tabela abaixo fornece dados da proporção de mineradores que apresentam sintomas sérios de doenças no pulmão e o número de anos de exposição (Biometrics, 1959).

i	Número de Anos(x_i)	Número de casos(n_i)	Total de mineradores	Proporção de casos
1	5,8	0	98	0
2	15,0	1	54	0,0185
3	21,5	3	43	0,0698
4	27,5	8	48	0,1667
5	33,5	9	51	0,1765
6	39,5	8	38	0,2105
7	46,0	10	28	0,3571
8	51,5	5	11	0,4545

```
doenca.dat<-scan(what=list(anos=0,suc=0,total=0)
  5.8    0    98
15.0    1    54
21.5    3    43
27.5    8    48
33.5    9    51
39.5    8    38
46.0   10    28
51.5    5    11
```

```
attach(doenca.dat)
anos=doenca.dat$anos
suc=doenca.dat$suc
total=doenca.dat$total
prop=suc/total
plot(anos,prop,xlab="Anos de Exposição",
      ylab="Proporção de casos")
lines(anos,prop)
```


Diagrama de Dispersão dos dados de doenças no pulmão



A variável resposta de interesse é $Y_i \sim \text{Bin}(n_i, \pi_i)$.

- π_i é a probabilidade do mineiro ter sintomas severos da doença quando exposto a categoria i (anos) $i = 1, \dots, k$.
- n_i é o total de mineiros expostos na categoria i .
- x_i é o número de anos de exposição.

Modelo : $\text{Log} \left\{ \frac{\pi(x_i)}{1-\pi(x_i)} \right\} = \eta_i = \beta_0 + \beta_1 x_i, \quad i = 1, \dots, 8$

- Valor predito : $\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i = -4,7965 + 0,0935 x_i$
- valor ajustado : $\hat{\mu} = \hat{y}_i = \hat{\pi}(x_i) = \frac{\exp(\hat{\eta}(x_i))}{1+\exp(\hat{\eta}(x_i))}$

`predict(fit)`

`fitted(fit)`

```
Xmat=cbind(suc,total-suc)
```

```
fit=glm(Xmat~anos,family=binomial())
```

```
summary(fit)
```

Call:

```
glm(formula = Xmat ~ anos, family = binomial())
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6625	-0.5746	-0.2802	0.3237	1.4852

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.79648	0.56859	-8.436	< 2e-16
anos	0.09346	0.01543	6.059	1.37e-09

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

(Dispersion parameter for binomial family taken to

Null deviance: 56.9028 on 7 degrees of freedom

Residual deviance: 6.0508 on 6 degrees of freedom

AIC: 32.877

Number of Fisher Scoring iterations: 4

Interpretação dos parâmetros

O valor ajustado em x_i é $\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$

O valor ajustado em $x_i + 1$ é $\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1(x_i + 1)$

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1 \quad \text{chance}(x_i) : \frac{\pi(x_i)}{(1 - \pi(x_i))}$$

Razão de chances : $\psi = \frac{\text{chance}(x_i + 1)}{\text{chance}(x_i)}$

$$\hat{\eta}(x_i + 1) = \log(\psi(x_i + 1)) \quad \text{e} \quad \hat{\eta}(x_i) = \log(\psi(x_i + 1))$$

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1 = \log \left(\frac{\text{chance}(x_i + 1)}{\text{chance}(x_i)} \right)$$

$$\psi = \exp(\beta_1)$$

A razão de chances pode ser interpretado como o aumento estimado na probabilidade de sucesso associado a mudança em uma unidade no valor da variável do preditor linear.

Em geral, o aumento estimado da razão de chances associada com a mudança de d unidades na variável preditora é $\exp(d\hat{\beta}_1)$.

$$\hat{\psi} = \exp(\hat{\beta}_1) = \exp(0,0935) = 1,10$$

A cada aumento de um ano na exposição do trabalho, a chance de contrair a doença dos pulmões aumenta em 10%.

Se o tempo de exposição for de $d = 10$ anos tem-se que

$$\exp(d\hat{\beta}_1) = 2.55.$$

Exemplo 2 - Poisson

(Montgomery et al. pag 481) apresenta os resultados de um experimento conduzido para avaliar número de falhas (y) de um certo tipo de equipamento usado no processo e o tempo de instalação até apresentar defeito em meses (x).

equip	Falhas	meses	equip	Falhas	meses
1	5	18	2	3	15
3	0	11	4	1	14
5	4	23	6	0	10
7	0	5	8	1	8
9	0	7	10	0	12
11	0	3	12	1	7
13	0	2	14	7	30
15	0	9			

Modelo Poisson : $y_i \sim P(\mu_i)$ com

$$\log(\mu_i) = \beta_0 + \beta_1 x_i \quad i = 1, \dots, 15.$$

```
fab.dat=scan("c:/temp/fabrica.dat",
```

```
what=list(falha=0,mes=0))
```

```
attach(fab.dat)
```

```
y=fab.dat$falha
```

```
x=fab.dat$mes
```

```
fit=glm(y~x,family=poisson())
```

```
summary(fit)
```

Call:

```
glm(formula = y ~ x, family = poisson())
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3106	-1.0114	-0.7003	0.4031	1.8813

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.71995	0.55770	-3.084	0.00204
x	0.13065	0.02433	5.370	7.88e-08

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

(Dispersion parameter for poisson family taken t

Null deviance: 44.167 on 14 degrees of fre

Residual deviance: 14.935 on 13 degrees of fre

AIC: 38.481

Number of Fisher Scoring iterations: 5

Estamos interessados em testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$.

```
fit0=glm(y~1,family=poisson())
```

```
anova(fit0,fit,test="Chi")
```

Analysis of Deviance Table

Model 1: $y \sim 1$

Model 2: $y \sim x$

	Resid.	Df	Resid.	Dev	Df	Deviance	P(> Chi)
1		14		44.167			
2		13		14.935	1	29.233	6.419e-08

Interpretação dos parâmetros

O modelo ajustado é dado por

$$\hat{\mu}(x) = \exp(-1.71 + 0.13x)$$

em que x é o tempo de instalação. Logo se Aumentarmos de

uma unidade o tempo de exposição a variação no valor

esperado fica dado por $\frac{\hat{\mu}(x+1)}{\hat{\mu}(x)} = \exp(0,13) = 1,138$. Ou seja o

número esperado falhas aumenta aproximadamente 13,8%.

Alavanca

```
yh=fitted(fit)
```

```
X=model.matrix(fit)
```

```
w=fit$weights
```

```
W=diag(w)
```

```
H=sqrt(W)*%X*solve(t(X)*%W*X)*%t(X)
```

```
%*%sqrt(W)
```

```
h=diag(H)
```

```
plot(h,xlab="Índice")
```

```
identify(h)
```

Sugestão :

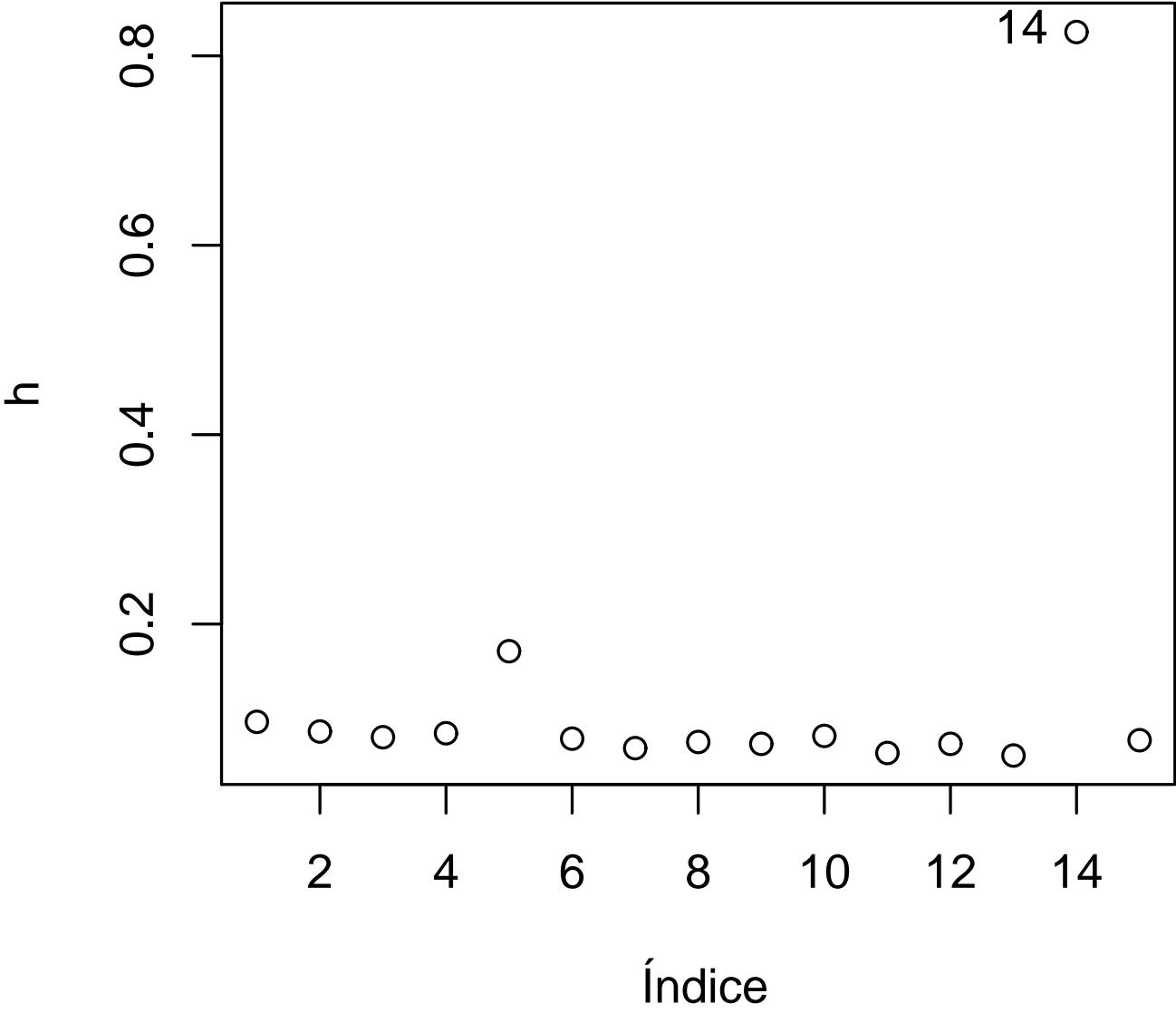
gráfico h_{ii} versus índice.

Resíduo componente do desvio versus índice.

Distância de Cook.

Envelope

Pontos de alavanca



resíduo padronizado

```
ro=residuals(fit,type="response")
fi=1
rd=residuals(fit,type="deviance")
td=rd*sqrt(fi/(1-h))
rp=residuals(fit,type="pearson")
rp=sqrt(fi)*rp
ts=rp/sqrt(1-h)
LD=h*(ts^2)/(1-h)
envelope.poisson(form=y~x,Fam=poisson(),k=100
,alfa=0.05)
plot(td,ylab="Componente do desvio",xlab="Índice
identify(td)
plot(LD,ylab="Distância de Cook",xlab="Índice")
identify(LD)
```

Normal Q-Q Plot

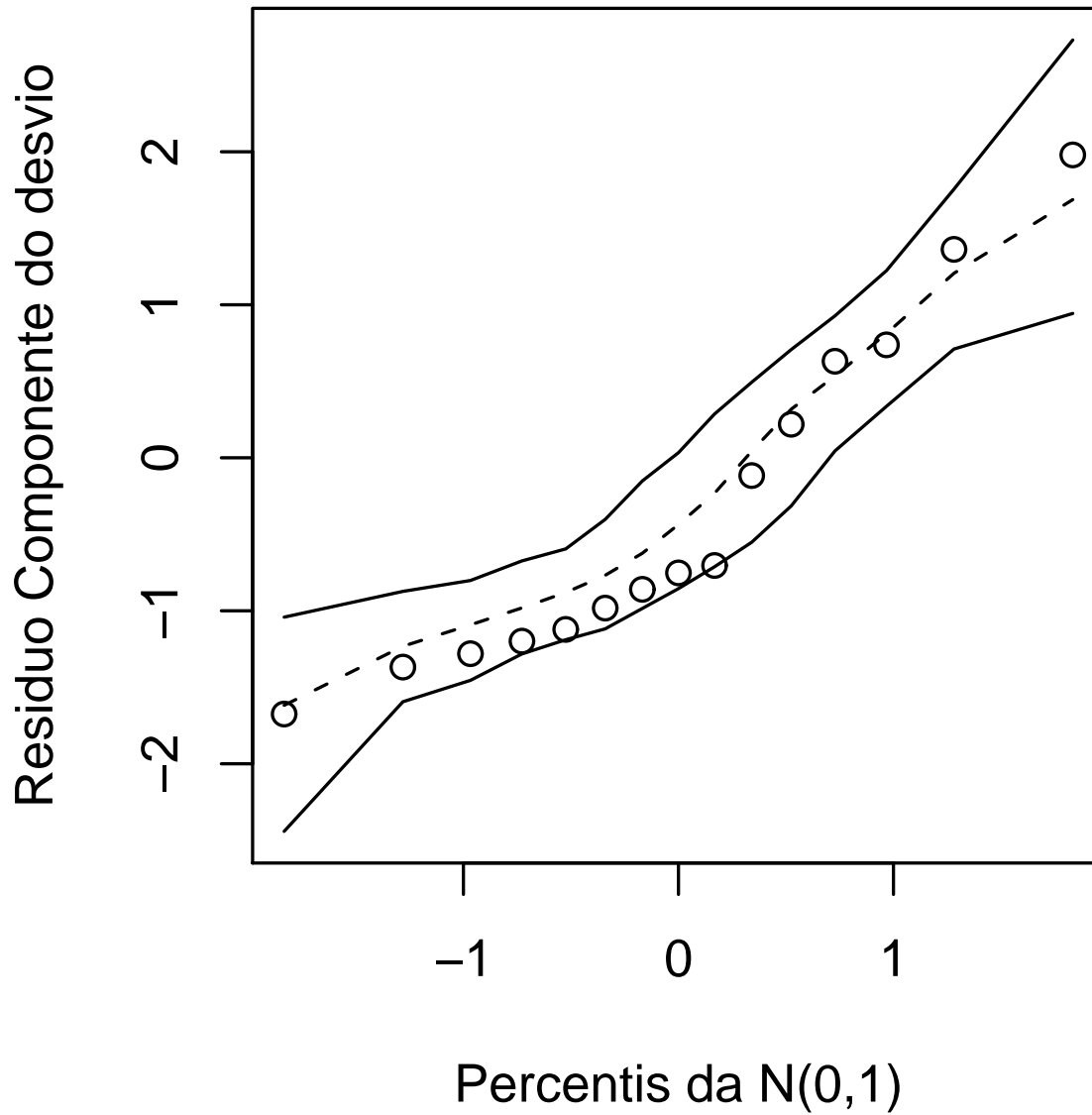
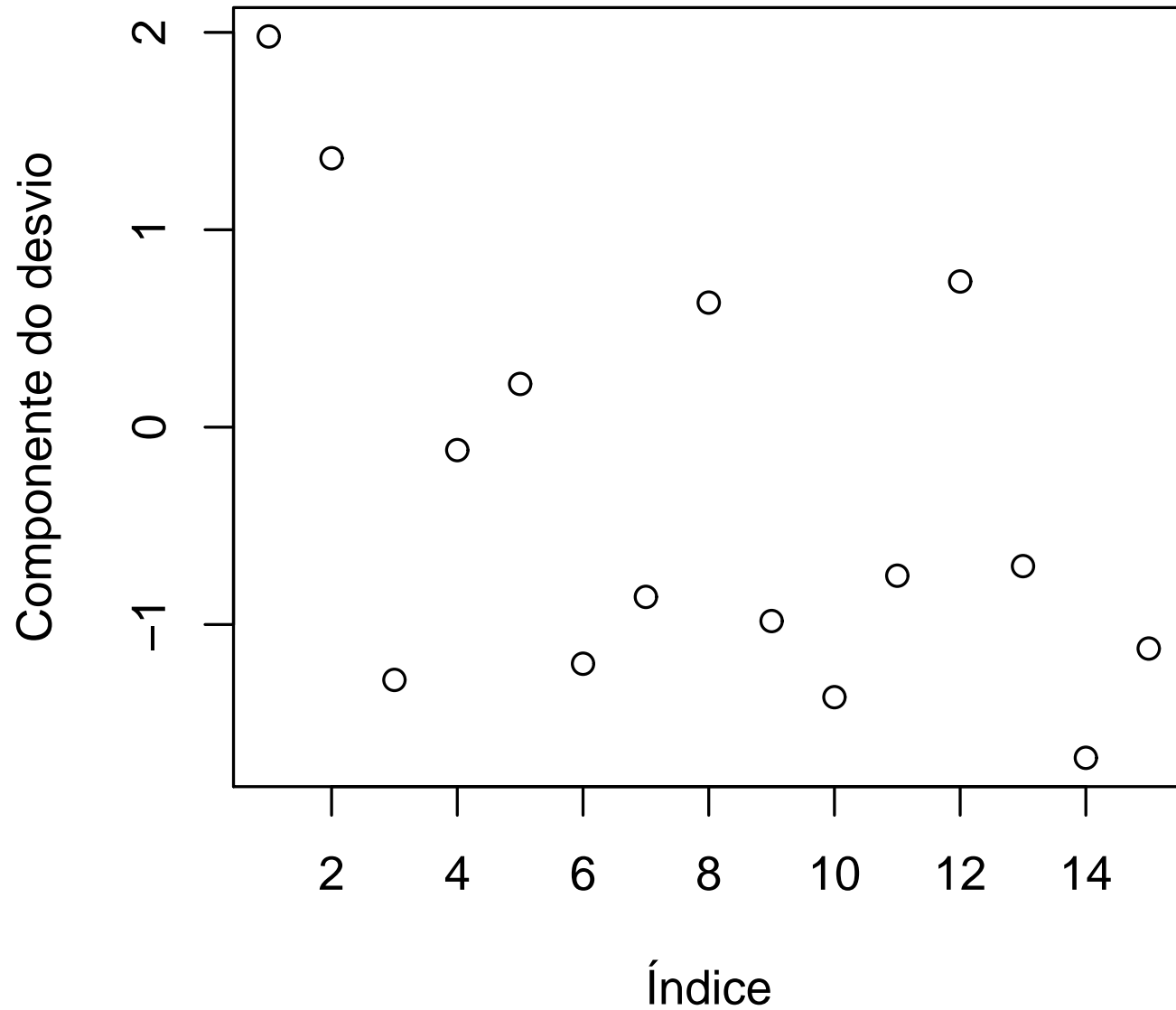
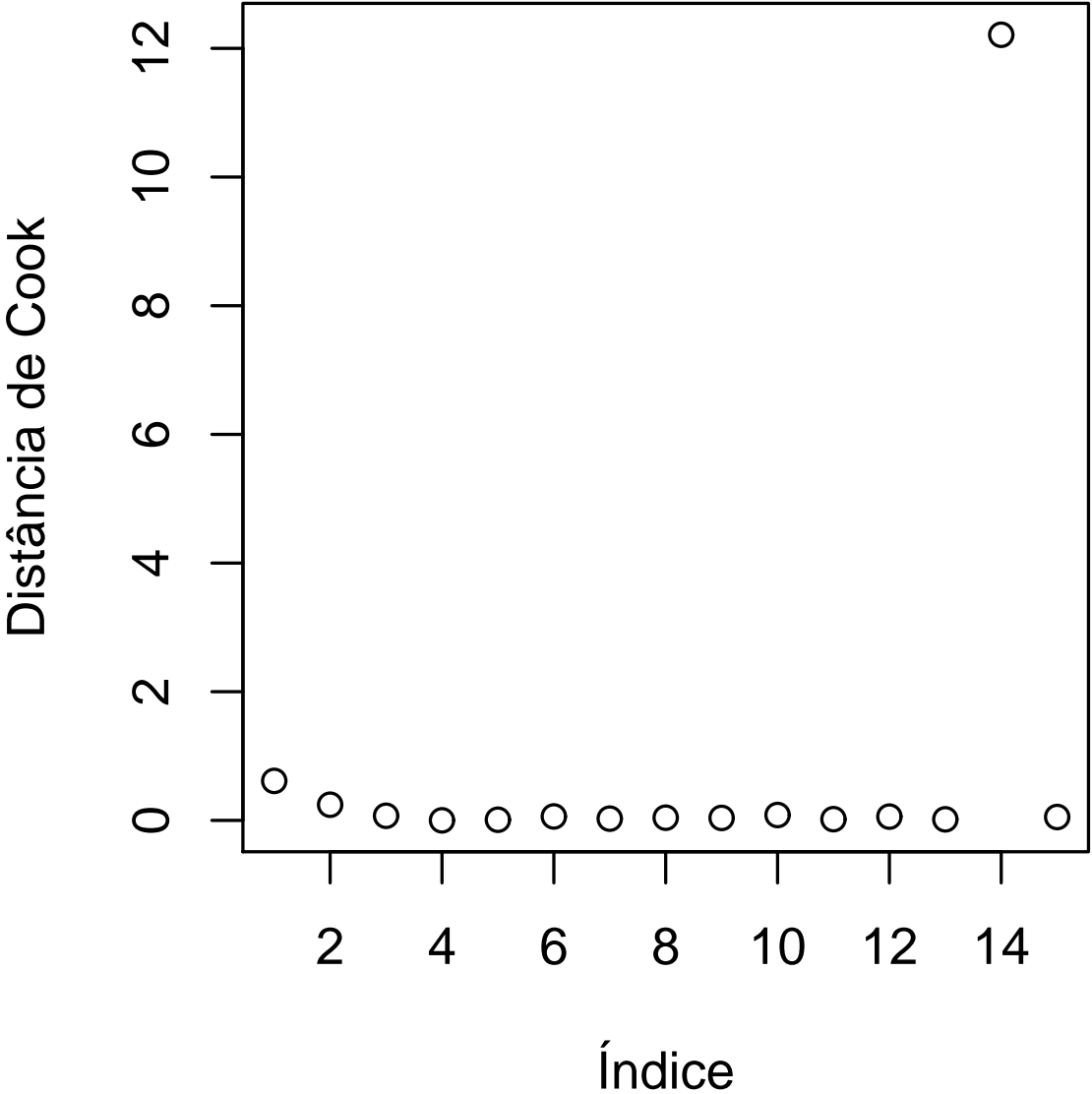


Gráfico do resíduo



Distância de Cook



Análise confirmatória

a observação 14 é influente e alta alavanca

`fit1=glm(y x,family=poisson,subset=-c(14))`

$\beta_0 = -2,45096$ e $\beta_1 = 0,18511$ com *desvio* = 12,530 com 12 g.l.
e *AIC* : 32,269.

Modelo com todas as observações

$\beta_0 = -1,71995$ e $\beta_1 = 0,13065$ com *desvio* = 14,935 com 13 g.l.
e *AIC* : 38,481

$$\%mudanca(\beta) = \left| \frac{\hat{\beta}_{(i)} - \hat{\beta}}{\hat{\beta}} \right| * 100$$

$$\%mudanca(\beta_0) = 42,50\% \quad \text{e} \quad \%mudanca(\beta_1) = 41,68\%$$