USP – ICMC – SME0823 - Modelos Lineares Generalizados

1ª prova – 1º/2014 – 5/5/2014

S O L U Ç Ã O

1. Descreva os três componentes de um modelo linear generalizado.
   Solução. Um modelo linear generalizado é formado por três componentes. O componente aleatório especifica a distribuição da variável resposta Y como sendo da família exponencial uniparamétrica. O componente sistemático é representado pelo preditor linear $\eta = \boldsymbol{x}^\top \boldsymbol{\beta}$, em que $\boldsymbol{x}$ e $\boldsymbol{\beta}$ denotam, respectivamente, o vetor de variáveis explicativas e o vetor de coeficientes. O terceiro componente é a função de ligação $g(\cdot)$ que relaciona a média da variável resposta $\mu = E(Y)$ com o preditor linear por meio de $g(\mu) = \eta$.

2. (a) Apresente uma distribuição que não seja da família exponencial.
   Solução. A distribuição Uniforme$(0, \theta)$, $\theta > 0$, com função densidade

$$f(y; \theta) = \frac{1}{\theta} I_{(0,\theta)}$$

   não é da família exponencial, pois o conjunto $A = (0, \theta)$ depende de θ.

   (b) A distribuição logarítmica tem função massa de probabilidade dada por

$$f(y; \rho) = -\frac{\rho^y}{y \, \log(1-\rho)} \, I_{\{1,2,\ldots\}}(y), \;\; 0 < \rho < 1. \tag{1}$$

   Apresentando o parâmetro canônico θ, o conjunto A e as funções $a(\phi)$, $b(\theta)$ e $c(y; \phi)$, mostre que (1) pertence à família exponencial.
   Solução. A expressão pode ser escrita como

$$\begin{aligned} f(y; \rho) &= \exp\{y \log(\rho) - \log\big(-y \log(1-\rho)\big)\} I_{\{1,2,\ldots\}}(y) \\ &= \exp\{[y \log(\rho) - \log\big(-\log(1-\rho)\big)] - \log(y)\} I_{\{1,2,\ldots\}}(y). \end{aligned}$$

   Desta forma, o parâmetro canônico é $\theta = \log(\rho)$, $A = \{1, 2, \ldots\}$ (que não depende de θ), $a(\phi) = 1$ e

$$f(y; \theta) = \exp\{[y\theta - \log\big(-\log(1-e^\theta)\big)] - \log(y)\} I_A(y),$$

   de modo que $b(\theta) = \log\big(-\log(1-e^\theta)\big)$, $c(y; \phi) = c(y) = -\log(y)$ e assim provamos que a distribuição é da família exponencial.

   (c) Calcule a média μ da distribuição como função do parâmetro canônico θ.
   Solução. No item 2b obtivemos $b(\theta)$. Calculamos

$$E(Y) = \mu = b'(\theta) = -\frac{e^\theta}{(1-e^\theta)\log(1-e^\theta)}.$$

   (d) Apresente a equação que deve ser resolvida a fim de obter a função de ligação canônica.
   Solução. A função de ligação canônica é tal que $g(\mu) = \theta$. Como $\mu = b'(\theta)$, segue que $\theta = b'^{-1}(\mu)$, de forma que $g(\cdot)$ pode ser encontrada resolvendo

$$-\frac{e^\theta}{(1-e^\theta)\log(1-e^\theta)} = \mu.$$

   <u>Obs.</u> Prove que a equação acima tem solução única.

3. Em um experimento foi registrado o número de insetos mortos após cinco horas de exposição ao gás carbônico em oito diferentes concentrações. Um modelo para dados de contagem com resposta binomial foi ajustado utilizando o logaritmo natural da concentração (`logdose`). Alguns resultados do ajuste com a função de ligação logito são mostrados abaixo.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.740      5.182  -11.72   <2e-16
logdose       34.286      2.913   11.77   <2e-16
---

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.116  on 6  degrees of freedom
```

(a) Argumente de duas formas distintas que o efeito da variável explicativa `logdose` é significativo a um nível de 5%.
Solução. De acordo com os resultados acima, o valor-p para o teste de $\mathrm{H}_0 : \beta_{\text{logdose}} = 0$ contra $\mathrm{H}_1 : \beta_{\text{logdose}} \neq 0$ é $2{,}0\times 10^{-16}$ ($< 0,05$). Também temos que a diferença entre as desviâncias dos dois modelos é `Null deviance` $-$ `Residual deviance` $=$ $284,202 - 11,116 = 273{,}086$, para um grau de liberdade, que representa um valor elevado (de fato, valor-p $= 2{,}4\times 10^{-61}$). (Obs. Aplique a desigualdade de Markov a fim de obter um limite superior para o valor-p.).

(b) Com base no valor da desviância (`Residual deviance`) apresente uma estimativa do parâmetro de dispersão e comente sobre a adequação do modelo ajustado.
Solução. A desviância do modelo ajustado tem seis graus de liberdade. Assim, uma estimativa do parâmetro de dispersão é dada por $\widehat{\phi} = 11,116/6 = 1,85 > 1$. Para a família binomial temos $\phi = 1$. Há indicação de que um modelo que permita maior dispersão pode levar a um melhor ajuste.

(c) Efetue o teste de $\mathrm{H}_0 : \beta_{\text{logdose}} = 30$ contra $\mathrm{H}_1 : \beta_{\text{logdose}} \neq 30$.
Solução. Uma estatística de teste é dada por

$$Z = \frac{\widehat{\beta}_{\text{logdose}} - 30}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_{\text{logdose}})}}.$$

Sob H_0, $Z \sim \mathrm{N}(0,1)$, aproximadamente. Substituindo pelos resultados (p. 1) obtemos $z = (34,286 - 30)/2,913 = 1,47$. Como $|z| < 1,96$, não rejeitamos H_0 a um nível de 5% (Obs. Obtenha o valor-p correspondente. Obtenha o valor-p utilizando a distribuição t_6. Por que t_6?).

(d) Uma quantidade de interesse é a concentração de gás correspondente à morte de 50% dos insetos, denotada por DL_{50}. Apresente a expressão de DL_{50}. Com base nos resultados acima, apresente uma estimativa para DL_{50}.
Solução. Para o modelo binomial com ligação logito temos

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_1 + \beta_{\text{logdose}}\, x,$$

em que π denota a probabilidade do evento de interesse (morte do inseto) e x é a variável explicativa (`logdose`). Quando $\pi = 0,5$, temos $x = \log(\mathrm{DL}_{50})$ e substituindo na equação acima resulta em

$$\log\left(\frac{0,5}{1-0,5}\right) = 0 = \beta_1 + \beta_{\text{logdose}} \log(\mathrm{DL}_{50}) \;\Rightarrow\; \mathrm{DL}_{50} = \exp\left(-\frac{\beta_1}{\beta_{\text{logdose}}}\right).$$

Substituindo pelos resultados (p. 1) obtemos a estimativa $\widehat{\mathrm{DL}}_{50} = \exp(60,740\,/34,286)$ $= 5,880$.