USP – ICMC – SME0823 - Modelos Lineares Generalizados

1ª prova – 1º/2015

S O L U Ç Ã O

1. (a) Para os MLG com esperança da variável resposta igual a μ, temos $g(\mu) = \eta$, em que $g(\cdot)$ e η denotam a função de ligação e o preditor linear, respectivamente. Se a função de ligação é selecionada de modo que $g(\mu) = \theta$, em que θ é o parâmetro canônico, dizemos que $g(\cdot)$ é a função de ligação canônica. Também sabemos que $\mu = b'(\theta)$. Para cada uma das relações abaixo, apresente a função de ligação canônica.
(i) $b(\theta) = \theta^2/2$, (ii) $b(\theta) = e^\theta$ e (iii) $b(\theta) = -\log(-\theta)$, para $\theta < 0$.
Solução. Partindo de $\mu = b'(\theta)$, invertemos e obtemos $(b')^{-1}(\mu) = \theta$, de modo que a função de ligação canônica é dada por $g = (b')^{-1}$.
(i) $b'(\theta) = \theta$ e $g(\mu) = \mu$, (ii) $b'(\theta) = e^\theta$ e $g(\mu) = \log(\mu)$ e (iii) $b'(\theta) = -1/\theta$ e $g(\mu) = -1/\theta$.

   (b) Suponha que Y_i, $i = 1, \ldots, n$, são variáveis aleatórias independentes seguindo uma distribuição Bernoulli(μ_i), em que $\mu_i = E(Y_i) = 1 - \exp\big(-\exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i})\big)$. Esta especificação corresponde a um modelo linear generalizado?
Solução. Sim. A distribuição da variável resposta é da família exponencial. A relação com as variáveis explicativas é representada pelo preditor linear $\eta = \beta_1 + \beta_2 x_2 + \beta_3 x_3$. Invertendo a função $\mu = 1 - \exp\big(-\exp(\eta)\big)$, em que $\mu \in (0, 1)$, obtemos a função de ligação $g(\mu) = \log\big(-\log(1-\mu)\big)$.

2. As situações abaixo podem ser representadas por modelos lineares generalizados. Para cada uma delas identifique a variável resposta e as variáveis explicativas, selecione uma distribuição de probabilidade adequada para a variável resposta e apresente o preditor linear bem como uma função de ligação.

   (a) O efeito da temperatura ambiente e do nível de voltagem sobre o tempo de vida de um componente eletrônico.
Solução. Variável resposta: tempo de vida de um componente eletrônico, que é positiva e contínua; variáveis explicativas: temperatura ambiente (x_2) e nível de voltagem (x_3); distribuição de probabilidade para a variável resposta: gama; preditor linear: $\beta_1 + \beta_2 x_2 + \beta_3 x_3$; função de ligação: log.

   (b) As quantidades de besouros de uma certa espécie que morrem após exposição a uma bactéria quando cinco diferentes níveis de exposição são usados e 20 besouros são utilizados em cada nível.
Solução. Variável resposta: contagem de besouros mortos de um total de 20 em teste, que é limitada; variável explicativa: exposição a uma bactéria (x_2), desdobrada em quatro variáveis indicadoras auxiliares; distribuição de probabilidade para a variável resposta: binomial; preditor linear: $\beta_1 + \beta_2 x_2$; função de ligação: logito.

   (c) A relação entre o número semanal de idas a um supermercado de uma dona de casa e o número de pessoas no domicílio, a renda familiar e distância do domicílio ao supermercado.
Solução. Variável resposta: contagem do número semanal de idas a um supermercado, com limite superior desconhecido; variáveis explicativas: número de

pessoas no domicílio (x_2), renda familiar (x_3) e distância do domicílio ao supermercado (x_4); distribuição de probabilidade para a variável resposta: Poisson; preditor linear: $\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$; função de ligação: log.

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:   7.011  on 6  degrees of freedom
```

(a) Explique os graus de liberdade (`degrees of freedom`) apresentados acima (`Null` indica o modelo nulo).
Solução. O tamanho da amostra é dado por $n = 8$ diferentes concentrações. O modelo nulo inclui um coeficiente (intercepto), de modo que restam $8 - 1 = 7$ graus de liberdade. O modelo com a variável `logdose` apresenta dois coeficientes, de modo que restam $8 - 2 = 6$ graus de liberdade.

(b) 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` $= 284,202 - 7,011 = 277{,}191$, para um grau de liberdade, que representa um valor elevado.

(c) 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 está associada a seis graus de liberdade. Assim, uma estimativa do parâmetro de dispersão é dada por $\widehat{\phi} = 7,011/6 = 1,17$, que não difere muito de 1, lembrando que para a família binomial temos $\phi = 1$. Em uma primeira avaliação, não há indicativo de sobredispersão ($\phi > 1$).

(d) Efetue o teste de $\mathrm{H}_0 : \beta_{\text{logdose}} = 29$ contra $\mathrm{H}_1 : \beta_{\text{logdose}} \neq 29$.
Solução. Uma estatística de teste é dada por

$$Z = \frac{\widehat{\beta}_{\text{logdose}} - 29}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_{\text{logdose}})}}.$$

Sob H_0, $Z \sim N(0,1)$, aproximadamente. Substituindo pelos resultados (p. 2) obtemos $z = (34,286 - 29)/2,913 = 1,81$. Como $|z| < 1,96$, não rejeitamos H_0 a um nível de 5%.
<u>Obs.</u> Obtenha o valor-p correspondente. Obtenha o valor-p utilizando a distribuição t_6. Por que t_6?

(e) Uma quantidade de interesse é a concentração de gás correspondente a uma probabilidade de morte igual a $0,5$, 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 $x = \log(DL_{50})$ e substituindo na equação acima resulta em

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

Substituindo pelos resultados (p. 2) obtemos a estimativa $\widehat{DL}_{50} = \exp(60,740\,/34,286) = 5,880$.