

ME720 - Modelos Lineares Generalizados

Parte 8 - MLGs - Introdução

Profa. **Larissa Avila Matos**

Modelos Lineares Generalizados (MLGs)

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.

O modelo de regressão linear (modelos lineares) utiliza a linearidade para descrever a relação entre a média da variável resposta e um conjunto de variáveis explicativas, assumindo que a distribuição da variável resposta é normal. Os modelos lineares generalizados (MLGs) estendem os modelos de regressão linear para abranger distribuições de respostas não-normais e possivelmente funções não-lineares da média. Eles têm três componentes:

- *Componente aleatório,*
- *Preditor linear,*
- *Função de ligação.*

Componente aleatório: especifica a variável de resposta \mathbf{Y} e sua distribuição de probabilidade. As observações $\mathbf{y} = (y_1, \dots, y_n)$ nessa distribuição são tratadas como independentes.

Preditor linear: Para um vetor de parâmetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ e uma matriz $\mathbf{X}_{n \times p}$ do modelo que contém valores de p variáveis explicativas para as n observações, o preditor linear é $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\eta}$.

Função de ligação: Esta é uma função g aplicada a cada componente de $E(\mathbf{y}) = \boldsymbol{\mu}$ que a relaciona com o preditor linear,

$$g(E(\mathbf{Y})) = \mathbf{X}\boldsymbol{\beta}.$$

Antes dos MLGs

Os MLGs foram criados com o objetivo de reunir numa mesma família vários modelos estatísticos que eram tratados separadamente.

Em geral, nas análises de regressão, procurava-se algum tipo de transformação que levasse à normalidade, tais como a transformação de Box-Cox (1964) dada abaixo

$$z = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0, \\ \log(y), & \text{se } \lambda = 0; \end{cases}$$

em que $y > 0$ e λ é uma constante desconhecida.

Família Exponencial

Família Exponencial

A família exponencial compreende um conjunto de distribuição flexível que varia entre variáveis aleatórias contínuas e discretas. Alguns membros dessa família são:

- Gaussiana: \mathbb{R}^p
- Multinomial: *Categórico*
- Bernoulli: Binário $\{0, 1\}$
- Binomial: Contagens de sucesso/fracasso
- Gama: \mathbb{R}^+
- Poisson: \mathbb{N}^+
- Laplace: \mathbb{R}^+
- Exponencial: \mathbb{R}^+
- Beta: $\{0, 1\}$
- Dirichlet: Δ (Simplex)
- Weibull: \mathbb{R}^+
- Weishart: matrizes simétricas positivas definidas

Todas estas distribuições seguem um formato geral.

Família Exponencial

Dizemos que uma variável aleatória Y , com distribuição $f(y; \theta)$, pertence a família exponencial se sua função de densidade ou função de probabilidades pode ser escrita na forma

$$f(y; \theta) = \exp [y\theta - b(\theta) + c(y)] \underbrace{\mathcal{I}(y \in A)}_{\text{não depende de } \theta},$$

ou seja,

$$\Rightarrow Y \sim f(y; \theta) \in \text{FE}(\theta),$$

onde θ é o *parâmetro natural*.

Exemplo

Seja Y uma variável aleatória com distribuição Poisson ($Y \sim P(\mu)$), a sua função de probabilidade é dada por

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots$$

Exemplo

Seja Y uma variável aleatória com distribuição Poisson ($Y \sim P(\mu)$), a sua função de probabilidade é dada por

$$\begin{aligned} f(y; \mu) &= \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots \\ &= \exp [y \log(\mu) - \mu - \log(y!)] , \end{aligned}$$

Exemplo

Seja Y uma variável aleatória com distribuição Poisson ($Y \sim P(\mu)$), a sua função de probabilidade é dada por

$$\begin{aligned} f(y; \mu) &= \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots \\ &= \exp [y \log(\mu) - \mu - \log(y!)] , \end{aligned}$$

Temos que, $\log(\mu) = \theta \Rightarrow \mu = e^\theta$, então

$$f(y; \mu) = \exp \left[y\theta - \underbrace{\exp \theta}_{b(\theta)} - \underbrace{\log(y!)}_{c(y)} \right]$$

$\Rightarrow Y \in \text{FE}(\theta)$.

Família Exponencial com parâmetro de escala

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, cada uma com função de densidade ou função de probabilidades dada por

$$f(y_i; \theta_i, \phi) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \mathcal{I}(y_i \in A),$$

com $\phi > 0$, constante.

Dizemos que Y_i tem distribuição pertencente à família exponencial com parâmetro de escala, $Y_i \sim f(y_i; \theta_i, \phi) \in \text{FE}(\theta_i, \phi)$. Além disso,

- θ_i é o *parâmetro natural*; e
- ϕ é o *parâmetro de dispersão*.

Frequentemente, $a(\phi) = 1$ e $c(y_i, \phi) = c(y_i)$, resultando na família exponencial natural, vista anteriormente.

Sabe-se que $\int_{-\infty}^{\infty} f(y_i; \theta_i, \phi) dy_i = 1$.

Função Geradora de Momentos

A Função Geradora de Momentos (F.G.M.) de Y_i é dada por

$$\begin{aligned}M_{Y_i}(t) &= \mathbb{E}(e^{tY_i}) = \int_A e^{tY_i} f(y_i; \theta_i, \phi) dy_i \\&= \exp \left[\frac{b(a(\phi)t + \theta_i) - b(\theta_i)}{a(\phi)} \right].\end{aligned}$$

Então,

$$\begin{aligned}\mathbb{E}(Y_i) &= \left. \frac{d}{dt} M_{Y_i}(t) \right|_{t=0} = \exp \left[\frac{b(a(\phi)t + \theta_i) - b(\theta_i)}{a(\phi)} \right] \frac{b'(a(\phi)t + \theta_i) a(\phi)}{a(\phi)} \Big|_{t=0} \\&= b'(\theta_i) \\&= \mu_i.\end{aligned}$$

Função Cumulante

Os cumulantes são definidos através da expansão em série de Taylor de $\log(\mathbb{E}(e^{tY_i}))$, ou seja

$$K_{Y_i}(t) = \log(M_{Y_i}(t)) = \sum_{j=0}^{\infty} \frac{k_j t^j}{j!},$$

k_1, k_2, k_3, \dots = cumulantes de Y_i , ou seja, são os momentos centrados na média.

Então,

$$K_{Y_i}(t) = \frac{b(a(\phi)t + \theta_i) - b(\theta_i)}{a(\phi)}.$$

Assim como na função geradora de momentos, temos que

$$\begin{aligned}k_1 &= \left. \frac{d}{dt} K_{Y_i}(t) \right|_{t=0} = \left. \frac{b'(a(\phi)t + \theta_i)a(\phi)}{a(\phi)} \right|_{t=0} \\&= b'(\theta_i) = \mu_i \\k_2 &= \mathbb{E}((Y_i - \mu_i)^2) = \text{Var}(Y) = \mathbb{E}(Y_i^2) - \mu_i^2 \\&= \left. \frac{d^2}{dt^2} K_{Y_i}(t) \right|_{t=0} = \left. b''(a(\phi)t + \theta_i)a(\phi) \right|_{t=0} \\&= b''(\theta_i)a(\phi) \\k_3 &= \mathbb{E}((Y_i - \mu_i)^3) = \mathbb{E}(Y_i^3) - 3\mathbb{E}(Y_i^2)\mu_i - 2\mu_i^3 = \left. b''(a(\phi)t + \theta_i)a(\phi) \right|_{t=0} \\&= b'''(\theta_i)a^2(\phi) \\&\vdots \\k_j &= b^{(j)}(\theta_i)[a(\phi)]^{j-1}\end{aligned}$$

Propriedade de invariância de cumulantes

Os cumulantes padronizados são dados por

$$\rho_j = \frac{k_j}{k_2^{j/2}}, \quad j = 2, 3, \dots$$

- ρ_j é invariante com respeito a locação, i.e., se $Y_i = \frac{X_i - a}{b}$, $\forall a$ e $b \neq 0$, os cumulantes de X_i e Y_i são os mesmos.

Podemos também encontrar expressões para $\mathbb{E}(Y_i)$ e $\text{Var}(Y_i)$ que usam quantidades de $f(y_i; \theta_i, \phi)$ de outra forma.

Considere $\ell_i = \log(f(y_i; \theta_i, \phi))$ a contribuição de Y_i na função de log-verossimilhança, $\ell = \sum_i \ell_i$. Então, temos que

$$\ell_i = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi),$$

com

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi)} \quad \text{e} \quad \frac{\partial^2 \ell_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)},$$

onde $b'(\theta_i)$ e $b''(\theta_i)$ são a primeira e a segunda derivada de $b(\cdot)$ avaliada em θ_i , respectivamente.

Podemos mostrar sob certas condições de regularidade, que

$$\mathbf{1} \quad \mathbb{E} \left(\frac{\partial \ell_i}{\partial \theta_i} \right) = 0, \quad \forall_i \rightarrow \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0; \text{ e}$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell_i}{\partial \theta_i^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell_i}{\partial \theta_i} \right)^2 \right], \quad \forall_i \rightarrow \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right].$$

Prova:??

Prova: Para facilitar a notação, considere $f(y_i; \theta_i, \phi) = f(y_i)$.

$$\mathbf{1} \quad \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0$$

Prova: Para facilitar a notação, considere $f(y_i; \theta_i, \phi) = f(y_i)$.

$$\mathbf{1} \quad \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0$$

$$\mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = \int_A \frac{\partial \ell}{\partial \theta} f(y) dy = \int_A \frac{\partial \log(f(y))}{\partial \theta} f(y) dy$$

Prova: Para facilitar a notação, considere $f(y_i; \theta_i, \phi) = f(y_i)$.

$$\mathbf{1} \quad \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) &= \int_A \frac{\partial \ell}{\partial \theta} f(y) dy = \int_A \frac{\partial \log(f(y))}{\partial \theta} f(y) dy \\ &= \int_A \frac{\partial f(y)}{\partial \theta} \frac{1}{f(y)} f(y) dy = \int_A \frac{\partial f(y)}{\partial \theta} dy \end{aligned}$$

Prova: Para facilitar a notação, considere $f(y_i; \theta_i, \phi) = f(y_i)$.

$$\textcolor{red}{1} \quad \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) &= \int_A \frac{\partial \ell}{\partial \theta} f(y) dy = \int_A \frac{\partial \log(f(y))}{\partial \theta} f(y) dy \\ &= \int_A \frac{\partial f(y)}{\partial \theta} \frac{1}{f(y)} f(y) dy = \int_A \frac{\partial f(y)}{\partial \theta} dy \\ &= \frac{\partial}{\partial \theta} \underbrace{\int_A f(y) dy}_1 = \frac{\partial}{\partial \theta} 1 = 0 \end{aligned}$$

$$\Rightarrow \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0 \quad \square$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(y) dy = \int_A \frac{\partial^2 \log(f(y))}{\partial \theta^2} f(y) dy$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(\mathbf{y}) d\mathbf{y} = \int_A \frac{\partial^2 \log(f(\mathbf{y}))}{\partial \theta^2} f(\mathbf{y}) d\mathbf{y} \\ &= \int_A \left[\frac{\partial^2 f(\mathbf{y})}{\partial \theta^2} \frac{1}{f(\mathbf{y})} - \left(\frac{\partial \log(f(\mathbf{y}))}{\partial \theta} \right)^2 \right] f(\mathbf{y}) d\mathbf{y} \end{aligned}$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(y) dy = \int_A \frac{\partial^2 \log(f(y))}{\partial \theta^2} f(y) dy \\ &= \int_A \left[\frac{\partial^2 f(y)}{\partial \theta^2} \frac{1}{f(y)} - \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] f(y) dy \\ &= \int_A \frac{\partial^2 f(y)}{\partial \theta^2} dy - \int_A \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 f(y) dy \end{aligned}$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(y) dy = \int_A \frac{\partial^2 \log(f(y))}{\partial \theta^2} f(y) dy \\ &= \int_A \left[\frac{\partial^2 f(y)}{\partial \theta^2} \frac{1}{f(y)} - \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] f(y) dy \\ &= \int_A \frac{\partial^2 f(y)}{\partial \theta^2} dy - \int_A \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 f(y) dy \\ &= \underbrace{\frac{\partial^2}{\partial \theta^2} \int_A f(y) dy}_1 - \mathbb{E} \left[\left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] \end{aligned}$$

$$\mathbf{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(y) dy = \int_A \frac{\partial^2 \log(f(y))}{\partial \theta^2} f(y) dy \\ &= \int_A \left[\frac{\partial^2 f(y)}{\partial \theta^2} \frac{1}{f(y)} - \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] f(y) dy \\ &= \int_A \frac{\partial^2 f(y)}{\partial \theta^2} dy - \int_A \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 f(y) dy \\ &= \underbrace{\frac{\partial^2}{\partial \theta^2} \int_A f(y) dy}_1 - \mathbb{E} \left[\left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] \\ &= \frac{\partial}{\partial \theta} 1 - \mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right] = \end{aligned}$$

$$\boxed{2} \quad \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$

$$\begin{aligned}
 \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= \int_A \frac{\partial^2 \ell}{\partial \theta^2} f(y) dy = \int_A \frac{\partial^2 \log(f(y))}{\partial \theta^2} f(y) dy \\
 &= \int_A \left[\frac{\partial^2 f(y)}{\partial \theta^2} \frac{1}{f(y)} - \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] f(y) dy \\
 &= \int_A \frac{\partial^2 f(y)}{\partial \theta^2} dy - \int_A \left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 f(y) dy \\
 &= \frac{\partial^2}{\partial \theta^2} \underbrace{\int_A f(y) dy}_1 - \mathbb{E} \left[\left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] \\
 &= \frac{\partial}{\partial \theta} 1 - \mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right] = -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right] \\
 \\
 \Rightarrow \mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= -\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right] \quad \square
 \end{aligned}$$

$\text{Var} \left(\frac{\partial \ell}{\partial \theta} \right)$

Podemos também calcular $\text{Var} \left(\frac{\partial \ell}{\partial \theta} \right)$, por (2.) temos que

$$\begin{aligned}\text{Var} \left(\frac{\partial \ell}{\partial \theta} \right) &= \text{Var} \left(\frac{\partial \log(f(y))}{\partial \theta} \right) \\&= \mathbb{E} \left[\left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] - \underbrace{\left[\mathbb{E} \left(\frac{\partial \log(f(y))}{\partial \theta} \right) \right]^2}_0 \\&= \mathbb{E} \left[\left(\frac{\partial \log(f(y))}{\partial \theta} \right)^2 \right] = -\mathbb{E} \left(\frac{\partial^2 \log(f(y))}{\partial \theta^2} \right) \\&= -\mathbb{E} \left(\frac{\partial^2 \ell}{\partial \theta^2} \right)\end{aligned}$$

Propriedades

Se $Y_i \sim FE(\theta_i, \phi)$, então

1 $\mathbb{E}(Y_i) = \mu_i = b'(\theta_i)$; e

2 $\text{Var}(Y_i) = a(\phi)V(\mu_i)$, em que $V_i = V(\mu_i) = d\mu_i/d\theta_i$ é a função de variância.

$\phi > 0$ é o parâmetro de dispersão (precisão), pois quanto menor o valor da sua função, menor será a variância.

Prova:??

A função de variância desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição. Isto é, dada a função de variância, tem-se uma classe de distribuições correspondentes, e vice-versa.

Prova: Vimos que

$$\ell_i = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi),$$

e

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi)}, \quad \frac{\partial^2 \ell_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}.$$

Então, da condição de regularidade (1.), temos

Prova: Vimos que

$$\ell_i = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi),$$

e

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi)}, \quad \frac{\partial^2 \ell_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}.$$

Então, da condição de regularidade (1.), temos

$$\mathbb{E} \left(\frac{\partial \ell_i}{\partial \theta_i} \right) =$$

Prova: Vimos que

$$\ell_i = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi),$$

e

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi)}, \quad \frac{\partial^2 \ell_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}.$$

Então, da condição de regularidade (1.), temos

$$\begin{aligned} \mathbb{E} \left(\frac{\partial \ell_i}{\partial \theta_i} \right) &= \mathbb{E} \left(\frac{Y_i - b'(\theta_i)}{a(\phi)} \right) = \frac{\mathbb{E}(Y_i) - b'(\theta_i)}{a(\phi)} = 0 \\ &\Rightarrow \mathbb{E}(Y_i) = b'(\theta_i). \end{aligned}$$

Da condição de regularidade (2.) vimos que

$$\text{Var} \left(\frac{\partial \ell_i}{\partial \theta_i} \right) = -\mathbb{E} \left(\frac{\partial^2 \ell_i}{\partial \theta_i^2} \right)$$

$$\Rightarrow \text{Var} \left(\frac{Y_i - b'(\theta_i)}{a(\phi)} \right) = -\mathbb{E} \left(\frac{-b''(\theta_i)}{a(\phi)} \right)$$

$$\Rightarrow \frac{\text{Var}(Y_i)}{(a(\phi))^2} = \frac{b''(\theta_i)}{a(\phi)}$$

$$\Rightarrow \text{Var}(Y_i) = b''(\theta_i)a(\phi)$$

Exemplo: A função de variância definida por $V(\mu) = \mu(1 - \mu)$, $0 < \mu < 1$, caracteriza a classe de distribuições binomiais com probabilidades de sucesso μ ou $1 - \mu$.

Uma propriedade interessante envolvendo a distribuição de Y e a função de variância é a seguinte

$$\frac{(Y - \mu)}{\sqrt{a(\phi)}} \xrightarrow{d} N(0, V(\mu)), \quad \text{quando } a(\phi) \rightarrow \infty.$$

Ou seja, para ϕ grande Y segue distribuição aproximadamente normal de média μ e variância $a(\phi)V(\mu)$. Esse tipo de abordagem assintótica, diferente da usual em que n é grande, foi introduzida por Jørgensen (1987).

Casos particulares FE

Poisson

Seja Y_i uma variável aleatória com distribuição Poisson ($Y \sim P(\mu_i)$), a função de probabilidades fica dada por

$$f(y_i; \mu) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \exp [y_i \log(\mu_i) - \mu_i - \log(y_i!)] , \quad y_i = 0, 1, 2, \dots$$

Fazendo $\log(\mu_i) = \theta_i$, temos

$$f(y_i; \mu) = \exp \left[y_i \theta_i - \underbrace{\exp \theta_i}_{b(\theta_i)} - \underbrace{\log(y_i!)}_{c(y_i, \phi)} \right] ,$$

onde θ_i é o parâmetro natural e $a(\phi) = 1$.

Portanto, essa distribuição pertence a família exponencial com parâmetro de escala, com

- $b(\theta_i) = \exp(\theta_i),$

- $a(\phi) = 1,$ e

- $c(y_i, \phi) = -\log(y_i!).$

Então,

$$\mathbb{E}(Y_i) = b'(\theta_i) = \exp(\theta_i) = \mu_i,$$

e

$$\text{Var}(Y_i) = a(\phi)b''(\theta_i) = \exp(\theta_i) = \mu_i.$$

Segue portanto que $V(\mu_i) = \mu_i.$

Binomial

Seja Y_i^* a proporção de sucessos em n_i ensaios independentes, cada um com probabilidade de ocorrência μ_i . Assumimos que $n_i Y_i^* \sim \text{Bin}(n_i, \mu_i)$, nesse caso temos que $\mathbb{E}(Y_i^*) = \mu_i$ não depende de n_i .

A função de probabilidades de Y^* fica então expressa na forma

$$\begin{aligned} f(y_i^*; \mu_i, n_i) &= \binom{n_i}{n_i y_i^*} \mu_i^{n_i y_i^*} (1 - \mu_i)^{n_i - n_i y_i^*}, \quad y_i^* = 0, \frac{1}{n_i}, \frac{2}{n_i}, \dots, 1; \\ &= \exp \left[n_i \log(1 - \mu_i) + n_i y_i^* \log \left(\frac{\mu_i}{1 - \mu_i} \right) + \log \left(\binom{n_i}{n_i y_i^*} \right) \right]. \end{aligned}$$

Fazendo $\log(\frac{\mu_i}{1 - \mu_i}) = \theta_i$, temos

$$f(y_i^*; \mu_i, n_i) = \exp \left[\frac{y_i^* \theta_i - \log(1 + \exp(\theta_i))}{1/n_i} + \log \left(\binom{n_i}{n_i y_i^*} \right) \right],$$

onde θ_i é o parâmetro natural, chamado de *logito*.

Portanto, essa distribuição pertence a família exponencial com parâmetro de escala, com

$$\blacksquare b(\theta_i) = \log(1 + \exp(\theta_i)),$$

$$\blacksquare a(\phi) = 1/n_i, \text{ e}$$

$$\blacksquare c(y_i, \phi) = \log \begin{pmatrix} n_i \\ n_i y_i^* \end{pmatrix}.$$

Então,

$$\mathbb{E}(Y_i) = b'(\theta_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)} = \mu_i,$$

e

$$\text{Var}(Y_i) = a(\phi)b''(\theta_i) = \frac{\exp(\theta_i)}{n_i[1 + \exp(\theta_i)]^2} = \frac{\mu_i(1 - \mu_i)}{n_i}.$$

Segue portanto que $V(\mu_i) = \mu_i(1 - \mu_i)$.

Normal

Seja Y_i uma variável aleatória com distribuição normal de média μ_i e variância σ^2 , $Y_i \sim N(\mu_i, \sigma^2)$. A função densidade de Y é expressa na forma

$$\begin{aligned} f(y_i; \mu_i, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \mu_i)^2 \right] \\ &= \exp \left[\frac{\mu_i y_i - \frac{\mu_i^2}{2}}{\sigma^2} - \frac{1}{2} \left(\log(2\pi\sigma^2) + \frac{y_i^2}{\sigma^2} \right) \right] \\ &= \exp \left[\frac{\theta_i y_i - \frac{\theta_i^2}{2}}{\sigma^2} - \frac{1}{2} \left(\log(2\pi\sigma^2) + \frac{y_i^2}{\sigma^2} \right) \right], \end{aligned}$$

em que $-\infty < \mu_i, y_i < \infty$ e $\sigma^2 > 0$. Além disso, $\theta_i = \mu_i$ é o parâmetro natural.

Portanto, essa distribuição pertence a família exponencial com parâmetro de escala, com

$$\blacksquare b(\theta_i) = \frac{\theta_i^2}{2},$$

$$\blacksquare a(\phi) = \sigma^2, \text{ e}$$

$$\blacksquare c(y_i, \phi) = -\frac{1}{2} \left(\log(2\pi\sigma^2) + \frac{y_i^2}{\sigma^2} \right).$$

Então,

$$\mathbb{E}(Y_i) = b'(\theta_i) = \theta_i = \mu_i;$$

e

$$\text{Var}(Y_i) = a(\phi)b''(\theta_i) = \sigma^2.$$

Segue portanto que $V(\mu_i) = 1$.

Seja Y_i uma variável aleatória com distribuição gama de média μ_i e coeficiente de variação $\phi^{1/2}$, denotamos $Y_i \sim G(\mu_i, \phi)$. A função densidade de Y_i é dada por

$$\begin{aligned}f(y_i; \mu_i, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi y_i}{\mu_i} \right)^\phi \exp \left(-\frac{\phi y_i}{\mu_i} \right) y_i^{-1} \\&= \exp \left[\phi \left(-\frac{y_i}{\mu_i} - \log(\mu_i) \right) - \log(\Gamma(\phi)) + \phi \log(\phi y_i) - \log(y_i) \right] \\&= \exp \left[\frac{-\frac{y_i}{\mu_i} - \log(\mu_i)}{1/\phi} + (\phi - 1) \log(y_i) + \phi \log(\phi) - \log(\Gamma(\phi)) \right] \\&= \exp \left[\frac{-y_i \theta_i + \log(-\theta_i)}{1/\phi} + (\phi - 1) \log(y_i) + \phi \log(\phi) - \log(\Gamma(\phi)) \right]\end{aligned}$$

Obs. A escolha da parametrização, consistiu em escrever a fdp de Y_i em termos de $\mu_i = \mathbb{E}(Y_i)$ e do parâmetro de precisão ϕ , de modo que $cv(Y_i) = dp(Y_i)/\mathbb{E}(Y_i) = \phi^{-1/2}$, o que implica que $\text{Var}(Y_i) = V(\mu_i)/\phi$.

Portanto, essa distribuição pertence a família exponencial com parâmetro de escala, onde $\theta_i = -\frac{1}{\mu_i}$ é o parâmetro natural. Além disso,

- $b(\theta_i) = -\log(-\theta_i),$

- $a(\phi) = \frac{1}{\phi},$ e

- $c(y_i, \phi) = (\phi - 1) \log(y_i) + \phi \log(\phi) - \log(\Gamma(\phi)).$

Então,

$$\mathbb{E}(Y_i) = b'(\theta_i) = \mu_i \quad \text{e} \quad \text{Var}(Y_i) = a(\phi)b''(\theta_i) = \frac{\mu_i^2}{\phi},$$

com $V(\mu_i) = \mu_i^2$.

Nessa parametrização,

- se $\phi = 1$, então $Y_i \sim \text{Exp}(\mu_i);$

- se $\phi = k/2$ e $\mu_i = k$, então $Y_i \sim \chi^2(k).$

Normal Inversa

Seja Y_i uma variável aleatória com distribuição normal inversa de média μ_i e parâmetro de precisão ϕ , $Y_i \sim NI(\mu_i, \phi)$ e cuja função de densidade é dada por

$$\begin{aligned} f(y_i; \mu_i, \phi) &= \frac{\phi^{1/2}}{\sqrt{2\pi y_i^3}} \exp \left[-\frac{\phi(y_i - \mu_i)^2}{2\mu_i^2 y_i} \right] \\ &= \exp \left[\frac{\left(-\frac{y_i}{2\mu_i^2} + \frac{1}{\mu_i} \right)}{1/\phi} - \frac{1}{2} \left(\log(2\pi y_i^3/\phi) + \frac{\phi}{y_i} \right) \right], \end{aligned}$$

em que $y_i > 0$ e $\mu_i > 0$. Fazendo $\theta_i = -\frac{1}{2\mu_i}$, temos

$$f(y_i; \mu_i, \phi) = \exp \left[\frac{y_i \theta_i + (-2\theta_i)^{1/2}}{1/\phi} - \frac{1}{2} \left(\log(2\pi y_i^3/\phi) + \frac{\phi}{y_i} \right) \right].$$

Portanto, essa distribuição pertence a família exponencial com parâmetro de escala, com

$$\blacksquare b(\theta_i) = -(-2\theta_i)^{1/2},$$

$$\blacksquare a(\phi) = \frac{1}{\phi}, \text{ e}$$

$$\blacksquare c(y_i, \phi) = -\frac{1}{2} \left(\log(2\pi y_i^3 / \phi) + \frac{\phi}{y_i} \right).$$

Então,

$$\mathbb{E}(Y_i) = b'(\theta_i) = \frac{1}{\sqrt{-2\theta_i}} = \sqrt{\mu_i}$$

e

$$\text{Var}(Y_i) = a(\phi)b''(\theta_i) = \frac{1}{\phi} \frac{1}{2\sqrt{2}(-\theta_i)^{3/2}} = \frac{\mu_i^{3/2}}{\phi},$$

com $V(\mu_i) = \mu_i^{3/2}$.

Principais distribuições pertencentes à família exponencial.

Distribuição(Y_i)	$b(\theta_i)$	θ_i	$a(\phi)$	$V(\mu_i)$
Poisson	$\exp(\theta_i)$	$\log(\mu_i)$	1	μ_i
Binomial	$\log(1 + \exp(\theta_i))$	$\log(\frac{\mu_i}{1-\mu_i})$	$\frac{1}{n_i}$	$\mu_i(1 - \mu_i)$
Normal	$\frac{\theta_i^2}{2}$	μ_i	σ^2	1
Gama	$-\log(-\theta_i)$	$-\frac{1}{\mu_i}$	$\frac{1}{\phi}$	μ_i^2
Normal Inversa	$-\sqrt{-2\theta_i}$	$-\frac{1}{2\mu_i^2}$	$\frac{1}{\phi}$	$\mu_i^{3/2}$

Exercício

Seja Y_i uma variável aleatória com função de distribuição dada por

$$f(y_i; \lambda) = \frac{y_i}{\lambda^2} \exp \left[-\frac{y_i^2}{2\lambda^2} \right] \mathcal{I}(y_i > 0).$$

- Mostre que Y_i pertence a família exponencial.
- Encontre $\mathbb{E}(Y_i^2)$ e $\text{Var}(Y_i^2)$.

Funções de ligação

Função de ligação

A *função de ligação* de um MLG conecta a componente aleatória e o preditor linear.

Sejam $x_{i1}, x_{i2}, \dots, x_{in}$ covariáveis (constantes conhecidas).

Um MLG formula que um preditor linear $\eta_i = \sum_{j=1}^p \beta_j x_{ij}$ está relacionado com μ_i por

$$\eta_i = g(\mu_i),$$

para uma função de ligação $g(\cdot)$.

Ou equivalentemente, a função resposta g^{-1} mapeia os valores do preditor linear para a média.

Função de ligação canônica

A função de ligação g que transforma a média μ_i para o parâmetro natural θ_i na expressão da família exponencial com parâmetro de escala é chamada de ligação canônica, ou seja, $g(\mu_i) = \theta_i$.

Para isso, a relação direta

$$\theta_i = \sum_{j=1}^p \beta_j x_{ij}$$

igual a o parâmetro natural ao preditor linear.

Exemplo Poisson

Seja Y_i uma variável aleatória com distribuição Poisson ($Y \sim P(\mu_i)$), então

$$f(y_i; \mu) = \exp \left[y_i \theta_i - \underbrace{\exp \theta_i}_{b(\theta_i)} - \underbrace{\log(y_i!)}_{c(y_i, \phi)} \right],$$

onde

- $b(\theta_i) = \exp(\theta_i)$,
- $a(\phi) = 1$, e
- $c(y_i, \phi) = -\log(y_i!)$.

Temos que,

$$\mu_i = b'(\theta_i) = \exp(\theta_i) \quad \Rightarrow \quad \theta_i = \underbrace{\log(\mu_i)}_{\text{função de ligação canônica}}$$

Exercício

- Encontrar a função de ligação canônica de Y_i , onde Y_i é uma variável aleatória com distribuição Binomial.

As ligações canônicas mais comuns são:

Distribuição:

- Poisson
- Binomial
- Normal
- Gama
- Normal Inversa

Ligação:

$$\eta_i = \log(\mu_i)$$

$$\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right) \text{ (logito)}$$

$$\eta_i = \mu_i$$

$$\eta_i = \mu_i^{-1}$$

$$\eta_i = \mu_i^{-2}$$

Uma das vantagens de usarmos ligações canônicas é que as mesmas garantem a concavidade de ℓ e conseqüentemente muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a concavidade de ℓ garante a unicidade da estimativa de máxima verossimilhança de θ_i , quando essa existe.

Ligação probito

Seja μ_i a proporção de sucessos de uma distribuição binomial. A ligação probito é definida por

$$\Phi^{-1}(\mu_i) = \eta_i,$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

De modo equivalente, $\mu_i = \Phi(\eta_i)$.

Ligação complemento log-log

A distribuição do valor extremo (logaritmo da exponencial) tem função densidade dada por

$$f(y_i) = \exp [y_i - \exp(y_i)] ,$$

em que $-\infty < y_i < \infty$. Logo, a função de distribuição acumulada fica dada por

$$F(y_i) = 1 - \exp [- \exp(y_i)] .$$

Assim, o modelo binomial com ligação complemento log-log é definido tal que

$$\mu_i = 1 - \exp[-\exp(\eta_i)],$$

ou equivalentemente,

$$\eta_i = \log[-\log(1 - \mu_i)],$$

Sabemos que,

$$\textit{logito}(\mu_i) = \log \left(\frac{\mu_i}{1 - \mu_i} \right).$$

Além disso, note que

$$\textit{logito}(\mu_i) = -\textit{logito}(1 - \mu_i)$$

Prova:

Sabemos que,

$$\text{logito}(\mu_i) = \log \left(\frac{\mu_i}{1 - \mu_i} \right).$$

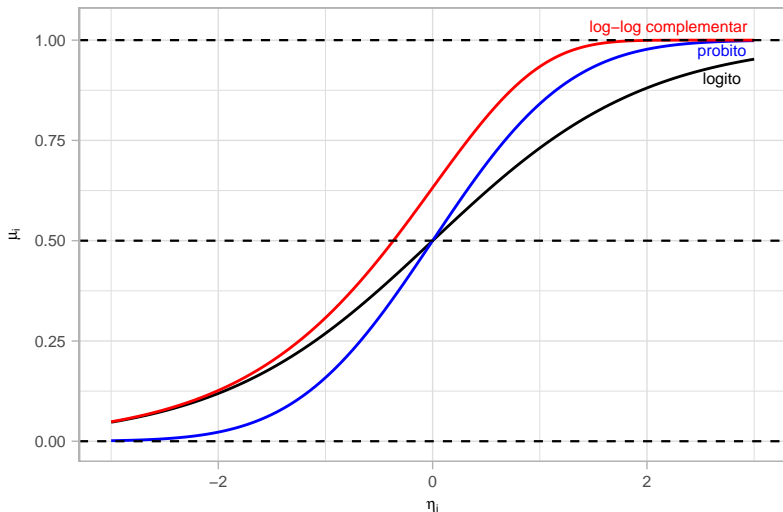
Além disso, note que

$$\text{logito}(\mu_i) = -\text{logito}(1 - \mu_i)$$

Prova:

$$\text{logito}(\mu_i) = \log \left(\frac{\mu_i}{1 - \mu_i} \right) = -\log \left(\frac{1 - \mu_i}{\mu_i} \right) = -\text{logito}(1 - \mu_i)$$

Funções de ligação para médias no intervalo (0,1)



Essas funções de ligações (probit e logito) são simétricas em torno de zero. As funções probit e logito não são apropriadas para dados que não tem simetria.

Ligação de Box-Cox

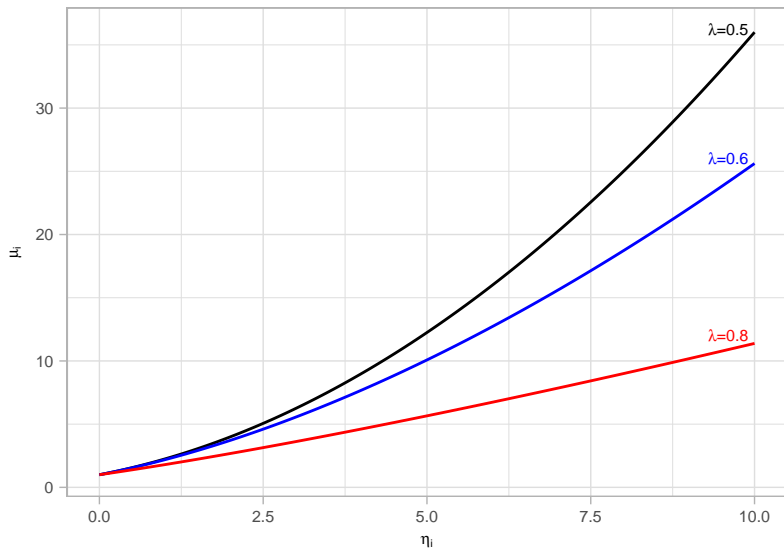
Uma classe importante de ligações, pelo menos para observações positivas, é a classe de ligações de Box-Cox definida por

$$\eta_i = \frac{(\mu_i^\lambda - 1)}{\lambda},$$

para $\lambda \neq 0$ e $\eta_i = \log(\mu_i)$ para $\lambda \rightarrow 0$.

A ideia agora é aplicarmos a transformação de Box-Cox, vista anteriormente, na média da variável resposta ao invés de transformarmos a própria variável resposta.

Comportamento de μ_i para alguns valores de λ e para η_i variando no intervalo $(0, 10)$.



- [Notas](#) de aula do Prof. Gilberto de Paula.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.