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 ($\boldsymbol{\mu}$) e o preditor linear $\boldsymbol{\eta}$.

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 $\boldsymbol{Y}$ e sua distribuição de probabilidade. As observações $\mathbf{y} = (\mathrm{y}_1, \ldots, \mathrm{y}_n)$ nessa distribuição são tratadas como independentes.

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

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,

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

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 $f(\text{y};\theta)$, pertence a família exponencial se sua função de densidade ou função de probabilidades pode ser escrita na forma

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

ou seja,

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

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

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

Exemplo

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

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

Exemplo

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

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

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

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

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

Família Exponencial com parâmetro de escala

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

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

com $\phi > 0$, constante.

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

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

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

Sabe-se que $\int_{-\infty}^{\infty} f(\mathrm{y}_i;\theta_i,\phi)\mathrm{dy}_i = 1$.

Função Geradora de Momentos

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

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

Então,

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

Função Cumulante

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

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

$k_1, k_2, k_3, \ldots =$ 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

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

Propriedade de invariância de cumulantes

Os cumulantes padronizados são dados por

$$\rho_j = \frac{k_j}{k_2^{1/2}}, \quad j = 2, 3, \ldots$$

- 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 $f(\text{y}_i;\theta_i,\phi)$ de outra forma.

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

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

com

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

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

Podemos mostrar sob certas condições de regularidade, que

1. $\mathbb{E}\left(\frac{\partial \ell_i}{\partial \theta_i}\right) = 0, \quad \forall_i \quad \rightarrow \quad \mathbb{E}\left(\frac{\partial \ell}{\partial \theta}\right) = 0$; e
2. $\mathbb{E}\left(\frac{\partial^2 \ell_i}{\partial \theta_i^2}\right) = -\mathbb{E}\left[\left(\frac{\partial \ell_i}{\partial \theta_i}\right)^2\right], \quad \forall_i \quad \rightarrow \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) = -\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right].$

Prova:??

Prova: Para facilitar a notação, considere $f(\mathrm{y}_i;\theta_i,\phi) = f(\mathrm{y}_i)$.

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

Prova: Para facilitar a notação, considere $f(\mathrm{y}_i; \theta_i, \phi) = f(\mathrm{y}_i)$.

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

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

Prova: Para facilitar a notação, considere $f(\mathrm{y}_i;\theta_i,\phi) = f(\mathrm{y}_i)$.

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

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

Prova: Para facilitar a notação, considere $f(\mathrm{y}_i;\theta_i,\phi) = f(\mathrm{y}_i)$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\mathrm{Var}\left(\frac{\partial \ell}{\partial \theta}\right) &= \mathrm{Var}\left(\frac{\partial \log(f(\mathrm{y}))}{\partial \theta}\right) \\
&= \mathbb{E}\left[\left(\frac{\partial \log(f(\mathrm{y}))}{\partial \theta}\right)^2\right] - \left[\underbrace{\mathbb{E}\left(\frac{\partial \log(f(\mathrm{y}))}{\partial \theta}\right)}_{0}\right]^2 \\
&= \mathbb{E}\left[\left(\frac{\partial \log(f(\mathrm{y}))}{\partial \theta}\right)^2\right] = -\mathbb{E}\left(\frac{\partial^2 \log(f(\mathrm{y}))}{\partial \theta^2}\right) \\
&= -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \theta^2}\right)
\end{aligned}
$$

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 $V_i = V(\mu_i) = \text{d}\mu_i/\text{d}\theta_i$ 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

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

e

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

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

Prova: Vimos que

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

e

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

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

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

Prova: Vimos que

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

e

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

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

$$\mathbb{E}\left(\frac{\partial \ell_i}{\partial \theta_i}\right) = \mathbb{E}\left(\frac{Y_i - b'(\theta_i)}{a(\phi)}\right) = \frac{\mathbb{E}(Y_i) - b'(\theta_i)}{a(\phi)} = 0$$

$$\Rightarrow \mathbb{E}(Y_i) = b'(\theta_i).$$

Da condição de regularidade (2.) vimos que

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

$$\begin{aligned}
\Rightarrow \quad & \text{Var}\left(\frac{Y_i - b'(\theta_i)}{a(\phi)}\right) = -\mathbb{E}\left(\frac{-b''(\theta_i)}{a(\phi)}\right) \\
\Rightarrow \quad & \frac{\text{Var}(Y_i)}{(a(\phi))^2} = \frac{b''(\theta_i)}{a(\phi)} \\
\Rightarrow \quad & \text{Var}(Y_i) = b''(\theta_i)a(\phi)
\end{aligned}$$

Exemplo: A função de variância definida por $V(\mu) = \mu(1-\mu), \quad 0 < \mu < 1$, 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

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

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

$$f(\mathrm{y}_i;\mu) = \frac{e^{-\mu_i}\mu_i^{\mathrm{y}_i}}{\mathrm{y}_i!} = \exp\left[\mathrm{y}_i \log(\mu_i) - \mu_i - \log(\mathrm{y}_i!)\right], \quad \mathrm{y}_i = 0, 1, 2, \ldots.$$

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

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

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
- $c(y_i, \phi) = -\log(\mathrm{y}_i!)$.

Então,

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

e

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

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

Binomial

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

A função de probabilidades de $Y^\star$ fica então expressa na forma

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

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

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

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

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

- $b(\theta_i) = \log(1 + \exp(\theta_i))$,
- $a(\phi) = 1/n_i$, e
- $c(y_i, \phi) = \log \begin{pmatrix} n_i \\ n_i y_i^\star \end{pmatrix}$.

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

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

em que $-\infty < \mu_i, \mathrm{y}_i < \infty$ 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

- $b(\theta_i) = \frac{\theta_i^2}{2}$,
- $a(\phi) = \sigma^2$, e
- $c(y_i, \phi) = -\frac{1}{2}\left(\log(2\pi\sigma^2) + \frac{\mathrm{y}_i^2}{\sigma^2}\right)$.

Então,

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

e

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

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

Gama

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

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

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 $\mathrm{Var}(Y_i) = V(\mu_i)/\phi$.

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
- $c(y_i, \phi) = (\phi - 1)\log(\mathrm{y}_i) + \phi\log(\phi) - \log(\Gamma(\phi))$.

Então,

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

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

Nessa parametrização,

- se $\phi = 1$, então $Y_i \sim Exp(\mu_i)$;
- 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 σ^2, $Y_i \sim NI(\mu_i, \phi)$ e cuja função de densidade é dada por

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

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

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

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

- $b(\theta_i) = -(-2\theta_i)^{1/2}$,
- $a(\phi) = \frac{1}{\phi}$, e
- $c(y_i, \phi) = -\frac{1}{2}\left(\log(2\pi \mathrm{y}_i^3/\phi) + \frac{\phi}{\mathrm{y}_i}\right)$.

Então,

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

e

$$\mathrm{Var}(Y_i) = a(\phi)b''(\theta_i) = \frac{1}{\phi}\left(-\frac{1}{2\theta}\right)^{3/2} = \frac{(\mu_i^2)^{3/2}}{\phi} = \frac{\mu_i^3}{\phi},$$

com $V(\mu_i) = \mu_i^3$.

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	$-\sqrt{-2\theta_i}$	$-\frac{1}{2\mu_i^2}$	$\frac{1}{\phi}$	μ_i^3

Exercício

Seja Y_i uma variável aleatória com função de distribuição dada por

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

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

Funções de ligação

Função de ligação

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

Sejam $x_{i1}, x_{i2}, \cdots, x_{in}$ 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

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

iguala o parâmetro natural ao preditor linear.

Exemplo Poisson

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

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

onde

- $b(\theta_i) = \exp(\theta_i)$,
- $a(\phi) = 1$, e
- $c(y_i, \phi) = -\log(\mathrm{y}_i!)$.

Temos que,

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

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:	Ligação:
Poisson	$\eta_i = \log(\mu_i)$
Binomial	$\eta_i = \log(\frac{\mu_i}{1-\mu_i})$ (logito)
Normal	$\eta_i = \mu_i$
Gama	$\eta_i = \mu_i^{-1}$
Normal Inversa	$\eta_i = \mu_i^{-2}$

Uma das vantagens de usarmos ligações canônicas é que as mesmas garantem a concavidade de ℓ e consequentemente 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

$$f(y_i) = \exp\left[\mathrm{y}_i - \exp(\mathrm{y}_i)\right],$$

em que $-\infty < \mathrm{y}_i < \infty$. Logo, a função de distribuição acumulada fica dada por

$$F(\mathrm{y}_i) = 1 - \exp\left[-\exp(\mathrm{y}_i)\right].$$

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

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

ou equivalentemente,

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

Sabemos que,

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

Além disso, note que

$$logito(\mu_i) = -logito(1-\mu_i)$$

Prova:

Sabemos que,

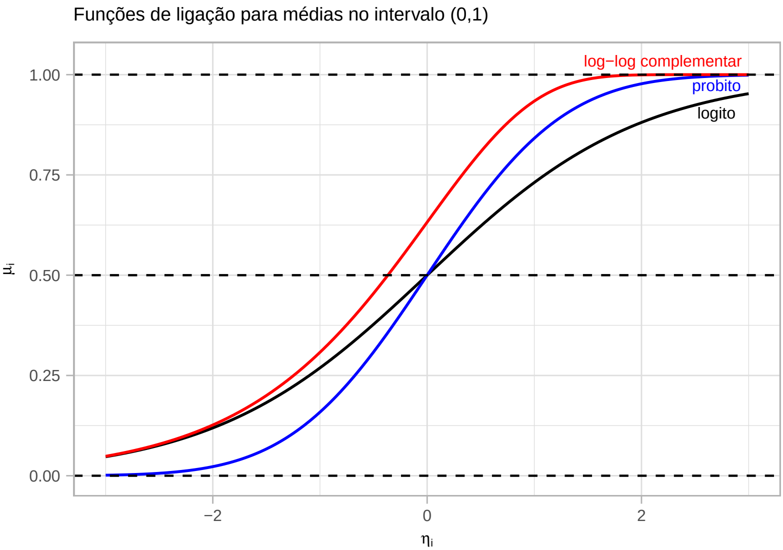
$$logito(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right).$$

Além disso, note que

$$logito(\mu_i) = -logito(1-\mu_i)$$

Prova:

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



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

Ligação de Box-Cox

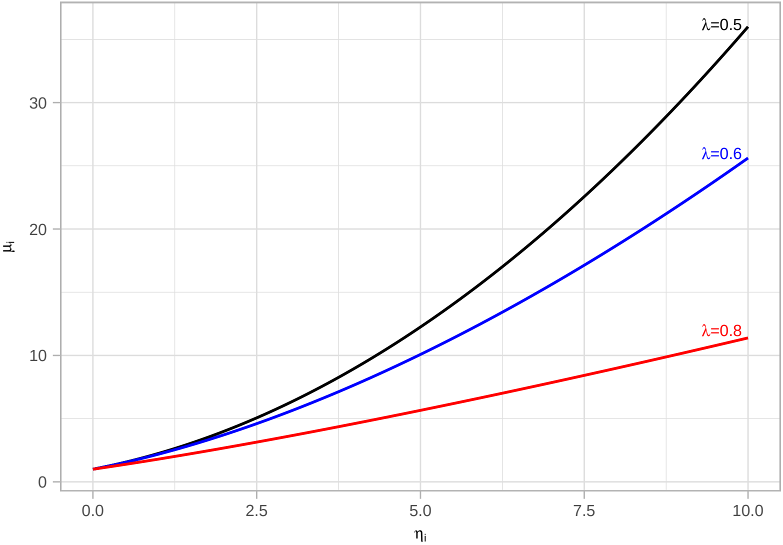
Uma classe importante de ligações, pelo menos para observações positivas, é a classe de ligações de Box-Cox definida por

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

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

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 $(0, 10)$.



Referência

- Paula, G.A. (2013). Modelos de Regressão com Apoio Computacional.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.