GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística

Universidade Federal Fluminense

Aula 04

Modelos Lineares Generalizados (MLG)

Estrutura dos MLG estudados no curso

- **Componente aleatório**: distribuição de Y na família exponencial.

$$f_Y(y;\theta,\phi) = \exp\{\phi[y\theta - b(\theta)] + c(y,\phi)\},$$

- **Componente sistemático**: preditor linear

$$\eta = \mathbf{X}\boldsymbol{\beta},$$

onde $\mathbf{X}$ é a matriz de covariáveis e $\boldsymbol{\beta}$ o vetor de parâmetros.

- **Função de ligação**: relaciona a média $\mu = E(Y)$ com o preditor linear:

$$g(\mu) = \eta.$$

Estimação do MLG

- Desenvolveremos agora a teoria para estimação de um MLG.
- Para obter estimativas pontuais e intervalares dos parâmetros de um MLG, vamos utilizar métodos baseados na máxima verossimilhança.
- Apesar de, em alguns casos especiais, ser possível obter a expressão matemática dos estimadores, frequentemente métodos numéricos serão necessários.
- Esses métodos são iterativos e baseados no algoritmo de Newton-Raphson.

Estimação no MLG

- Considere variáveis aleatórias independentes $Y_1, \ldots, Y_n$ que seguem as propriedades de um Modelo Linear Generalizado.

- Deseja-se estimar o vetor paramétrico $\boldsymbol{\beta}$, relacionado a Y_i por:

$$E(Y_i) = \mu_i \qquad \text{e} \qquad g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$$

- Para cada Y_i, a densidade assume a forma da família exponencial:

$$f_Y(y_i; \theta_i, \phi) = \exp\{\phi\,[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\}$$

- Nesse caso, a função de log-verossimilhança é dada por:

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

onde $b(\cdot)$ e $c(\cdot, \cdot)$ são funções conhecidas.

- Propriedades importantes já vistas:

$$E(Y_i) = \mu_i = b'(\theta_i), \qquad V(Y_i) = \frac{b''(\theta_i)}{\phi}, \qquad g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i$$

Log-Verossimilhança e Função Escore no MLG

- A função de log-verossimilhança para todos os Y_i é:

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{n} \ell_i = \sum_{i=1}^{n} \left[\phi \left(y_i\theta_i - b(\theta_i) \right) + c(y_i, \phi) \right].$$

- Para obter o estimador de máxima verossimilhança de β_j, precisamos resolver:

$$U_j = \frac{\partial \ell}{\partial \beta_j} = 0,$$

onde:

$$U_j = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j}.$$

- Essa decomposição pode ser entendida como:
  1. $\frac{\partial \ell_i}{\partial \theta_i}$: contribuição direta da densidade.
  2. $\frac{\partial \theta_i}{\partial \mu_i}$: relação entre parâmetro canônico e média.
  3. $\frac{\partial \mu_i}{\partial \beta_j}$: relação entre a média e os parâmetros via função de ligação.

Derivadas Intermediárias para a Função Escore

- Do termo **(1)** temos:

$$\frac{\partial \ell_i}{\partial \theta_i} = \phi\,(y_i - b'(\theta_i)) = \phi\,(y_i - \mu_i),$$

pois, $\mu_i = E(Y_i) = b'(\theta_i)$.

- Do termo **(2)** tem-se:

$$\frac{\partial \theta_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1}$$

- Como:

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i),$$

segue que:

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} = \frac{1}{\phi V(Y_i)},$$

pois

$$V(Y_i) = \frac{b''(\theta_i)}{\phi}.$$

Função Escore e Estimação de β

- Finalmente, do termo **(3)** temos:

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij},$$

onde $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$.

- Logo, a função escore para o parâmetro β_j é dada por:

$$U_j = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(Y_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i}.$$

- O estimador de máxima verossimilhança de $\boldsymbol{\beta}$ é obtido resolvendo:

$$\mathbf{U} = (U_1, \ldots, U_p)' = \mathbf{0}.$$

- Em geral, essas equações são não lineares e devem ser resolvidas numericamente por processos iterativos, como o método de Newton-Raphson.

Matriz Informação de Fisher

- Sabe-se que a matriz de variâncias e covariâncias dos U_j's é dada pela matriz informação de Fisher, com elementos:

$$\mathcal{I}_{jk} = E(U_j U_k) = E\left(\frac{\partial \ell}{\partial \beta_j} \cdot \frac{\partial \ell}{\partial \beta_k}\right) = -E\left(\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right)$$

- Nos modelos lineares generalizados temos:

$$\mathcal{I}_{jk} = -E\left(\sum_{i=1}^{n} \frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_k}\right) = E\left(\sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} \cdot \frac{\partial \ell_i}{\partial \beta_k}\right) = \sum_{i=1}^{n} E\left(\frac{\partial \ell_i}{\partial \beta_j} \cdot \frac{\partial \ell_i}{\partial \beta_k}\right)$$

Matriz Informação de Fisher

- Substituindo a expressão de U_j, temos:

$$E\left(\frac{\partial \ell_i}{\partial \beta_j} \cdot \frac{\partial \ell_i}{\partial \beta_k}\right) = E\left[\frac{(Y_i - \mu_i)}{V(Y_i)}\, x_{ij}\, \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{(Y_i - \mu_i)}{V(Y_i)}\, x_{ik}\, \frac{\partial \mu_i}{\partial \eta_i}\right]$$

- Como $E\left[(Y_i - \mu_i)^2\right] = V(Y_i)$, resulta que:

$$E\left(\frac{\partial \ell_i}{\partial \beta_j} \cdot \frac{\partial \ell_i}{\partial \beta_k}\right) = x_{ij}\, x_{ik}\, \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{V(Y_i)}$$

- Logo, os elementos da matriz informação de Fisher são dados por:

$$\mathcal{I}_{jk} = \sum_{i=1}^{n} x_{ij}\, x_{ik}\, \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{V(Y_i)}$$

Matriz Informação de Fisher: Forma Matricial

- A matriz informação de Fisher pode ser escrita de forma matricial como:

$$\mathcal{I} = \mathbf{X}'\mathbf{W}\mathbf{X}$$

onde $\mathbf{X}$ é a matriz de covariáveis $n \times p$ e $\mathbf{W}$ é uma matriz diagonal $n \times n$ com elementos

$$w_{ii} = \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{V(Y_i)}.$$

- Essa representação é fundamental para o cálculo iterativo dos estimadores de máxima verossimilhança, como no algoritmo IRLS (Iteratively Reweighted Least Squares).

Método de Newton-Raphson

- O método de Newton-Raphson é um algoritmo iterativo utilizado para encontrar uma solução numérica da equação

$$f(x) = 0.$$

- O objetivo é encontrar o valor de x no qual a função f corta o eixo x.

- A cada iteração, o método aproxima a solução usando a derivada da função:

$$x^{(m+1)} = x^{(m)} - \frac{f(x^{(m)})}{f'(x^{(m)})}.$$

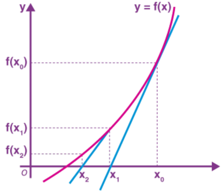

Figure: Método de Newton-Raphson para encontrar a solução de $f(x) = 0$.

Método de Newton-Raphson para EMV

- Consideremos o estimador de máxima verossimilhança (EMV) de um parâmetro qualquer ψ. A função de interesse é a função escore:

$$U = \frac{\partial \ell(\psi; y)}{\partial \psi}$$

- O objetivo é encontrar ψ tal que $U = 0$.
- O método de Newton-Raphson fornece a iteração:

$$\hat{\psi}^{(m)} = \hat{\psi}^{(m-1)} - \frac{U^{(m-1)}}{U'^{(m-1)}}$$

- Para a estimação por máxima verossimilhança, é comum aproximar U' pelo seu valor esperado $E(U')$.
- Como $\mathcal{I} = E(-U')$, a atualização iterativa, chamada de método do escore, é:

$$\hat{\psi}^{(m)} = \hat{\psi}^{(m-1)} + \frac{U^{(m-1)}}{\mathcal{I}^{(m-1)}} \tag{1}$$

- Essa abordagem é equivalente ao Newton-Raphson, mas usando a informação esperada em vez da derivada observada do escore.

Método do Escore no MLG

- Voltando ao problema de estimação por máxima verossimilhança no MLG.
- Para obter numericamente o EMV do vetor de parâmetros $\boldsymbol{\beta}$, vamos utilizar o método do escore, ou seja, vamos encontrar a solução da equação:

$$\mathbf{U} = \mathbf{0}$$

- Os elementos do vetor $\mathbf{U}$ são dados por:

$$U_j = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(Y_i)} \, x_{ij} \, \frac{\partial \mu_i}{\partial \eta_i}, \quad j = 1, \ldots, p$$

Método do Escore no MLG

- Para tanto, é preciso escrever a equação (1) de forma matricial:

$$\hat{\boldsymbol{\beta}}^{(m)} = \hat{\boldsymbol{\beta}}^{(m-1)} + \left[\boldsymbol{\mathcal{I}}^{(m-1)}\right]^{-1} \mathbf{U}^{(m-1)} \tag{2}$$

onde

  - $\hat{\boldsymbol{\beta}}^{(m)}$ é o vetor de estimativas dos parâmetros $\beta_1, \ldots, \beta_p$ na m-ésima iteração.
  - $\mathbf{U}^{(m-1)}$ é o vetor dos elementos dados do vetor escore, todos avaliados em $\hat{\boldsymbol{\beta}}^{(m-1)}$.
  - $\boldsymbol{\mathcal{I}}^{(m-1)}$ é a matriz informação de Fisher, com elementos $\mathcal{I}_{jk}$.

Método do Escore no MLG

- Multiplicando ambos os lados de (2) por $\boldsymbol{\mathcal{I}}^{(m-1)}$, obtemos:

$$\boldsymbol{\mathcal{I}}^{(m-1)}\,\hat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{\mathcal{I}}^{(m-1)}\,\hat{\boldsymbol{\beta}}^{(m-1)} + \mathbf{U}^{(m-1)}$$

- O lado direito desta equação é dado pelo vetor com elementos:

$$\sum_{k=1}^{p}\sum_{i=1}^{n} x_{ij}x_{ik}\,V(Y_i)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\hat{\beta}_k^{(m-1)} + \sum_{i=1}^{n}\frac{y_i-\mu_i}{V(Y_i)}\,x_{ij}\,\frac{\partial\mu_i}{\partial\eta_i}$$

  todos avaliados em $\hat{\boldsymbol{\beta}}^{(m-1)}$.

- Dessa forma, o lado direito da equação pode ser escrito matricialmente como:

$$\mathbf{X}'\mathbf{W}\mathbf{z}$$

  onde o vetor $\mathbf{z}$ tem elementos:

$$z_i = \sum_{k=1}^{p} x_{ik}\,\hat{\beta}_k^{(m-1)} + (y_i-\mu_i)\frac{\partial\eta_i}{\partial\mu_i}, \qquad i = 1,\ldots,n$$

  com μ_i e $\frac{\partial\eta_i}{\partial\mu_i}$ avaliados em $\hat{\boldsymbol{\beta}}^{(m-1)}$.

Método do Escore no MLG

- Logo, as equações iterativas para o método do escore podem ser escritas como:

$$\mathbf{X}'\mathbf{W}\mathbf{X}\,\hat{\beta}^{(m)} = \mathbf{X}'\mathbf{W}\mathbf{z}$$

  que têm a mesma forma das equações de mínimos quadrados ponderados, porém sendo resolvidas iterativamente.

- Esta última equação pode ser resolvida utilizando algum pacote estatístico que inclua a abordagem de MLG, tais como `R`, `Python` e `SPSS`.

Algoritmo do Método Numérico do Escore

1. Definir valor inicial para $\boldsymbol{\beta}$, ex.: $\boldsymbol{\beta}^{(0)} = \mathbf{0}$.
2. Encontrar $\mathbf{W}^{(m)}$ e $\mathbf{z}^{(m)}$.
3. Calcular $\hat{\boldsymbol{\beta}}^{(m+1)}$ resolvendo:

$$\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\,\hat{\boldsymbol{\beta}}^{(m+1)} = \mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)}$$

4. Comparar $\hat{\boldsymbol{\beta}}^{(m)}$ com $\hat{\boldsymbol{\beta}}^{(m+1)}$:

$$\|\hat{\boldsymbol{\beta}}^{(m)} - \hat{\boldsymbol{\beta}}^{(m+1)}\| \leq \delta$$

5. Se o resultado da comparação for verdadeiro, o método convergiu. Caso contrário, incrementar m e voltar ao Passo 2.