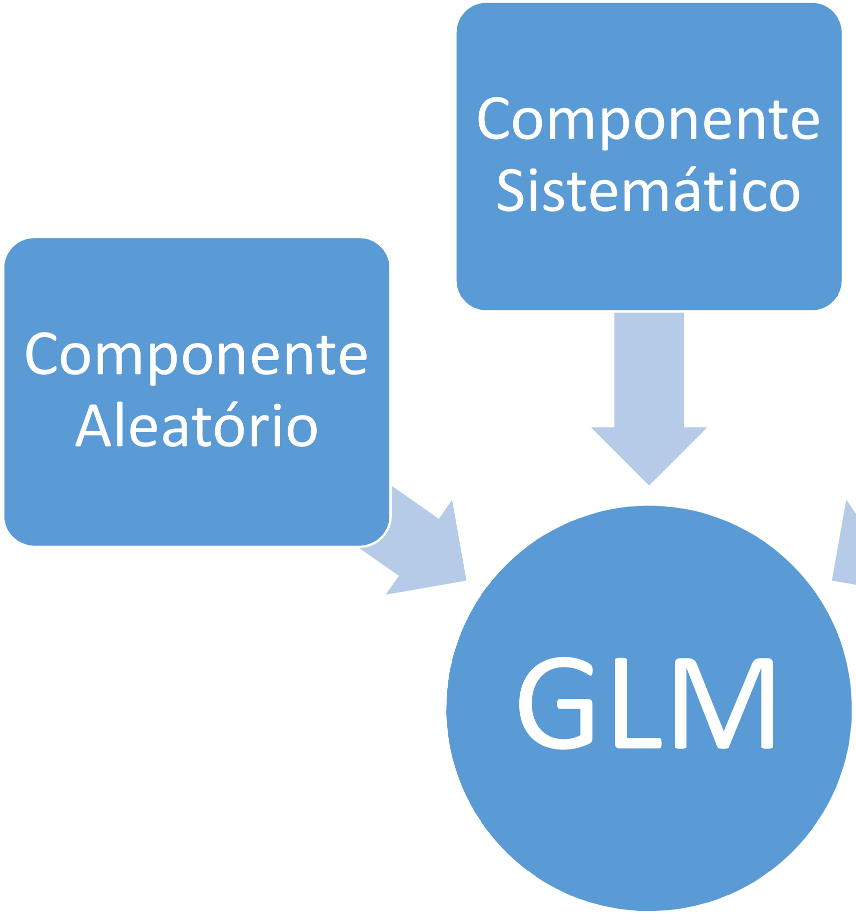
Definição de um GLM



Quem são os componentes e como conseguimos identificá-los

Um modelo linear generalizado é definido pela especificação de três componentes, o componente aleatório, sistemático e uma função de ligação.

Componentes de um GLM

Seja Y uma variável aleatória associada a um conjunto de variáveis explanatórias $X_1, \ldots, X_p$. Para uma amostra aleatória de tamanho n, (y_i, x_i) em que $x_i = \left(x_{i1}, \ldots, x_{ip}\right)^T$ é o vetor coluna de variáveis explanatórias , o GLM envolve os três componentes:

1º - Componente aleatório: Y $\sim família - exponencial(\theta_i, \phi)$

$$\mathrm{f}(x;\ \theta;\ \phi) = \exp\{\phi^{-1}[y\theta - \mathrm{b}(\theta)] + c(y, \phi)\}$$

em que $\phi > 0$ é um parâmetro de dispersão e o θ_i o parâmetro canônico

Propriedades do componente aleatório

Temos que:

E(Y_i) = μ_i = b'(θ_i)

Var(Y_i) = ϕb ''(θ_i) = ϕV(μ_i)

Em que **V($\boldsymbol{\mu_i}$) é a função de variância** que depende somente da média μ_i.

Componentes de um GLM

2º Componente sistemático: É o preditor linear do modelo, em que são inseridas as covariáveis por meio de uma combinação linear de parâmetros isto é:

$$\eta_i = \sum_{r=1}^{p} x_{ir}\beta_r = x_i^T \beta$$

Parâmetro de interesse será os $\boldsymbol{\beta}$ -> é ele que queremos estimar.

Ou

$$\eta_i = \boldsymbol{X}\beta$$

Em que **X** = $(x_1, \dots, x_n)^T$ é a matriz do modelo de β = $(\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros desconhecidos e $\boldsymbol{\eta}$ = $(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)^T$ **é o preditor linear**

$$\hat{Y}_i = \hat{\beta}_o + \hat{\beta}_1 X_i + \hat{\beta}_2 Z_i + \ldots + \hat{\beta}_n K_i$$

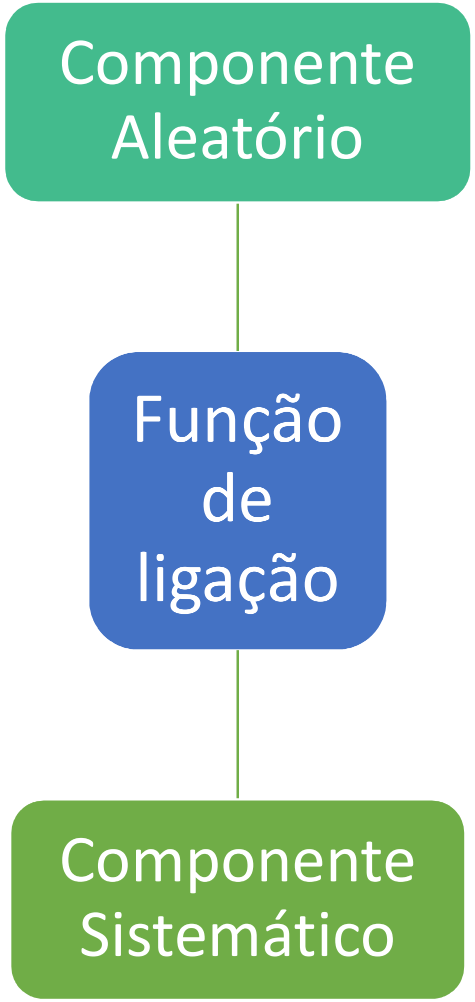
$$Y = XB$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & & k_1 \\ 1 & x_2 & & k_2 \\ 1 & x_3 & & k_3 \\ 1 & x_4 & & k_4 \\ 1 & x_5 & \ldots & . \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ n & x_n & \ldots & k_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_n \end{bmatrix}$$

3º Função de ligação: é uma função que relaciona o componente aleatório ao componente sistemático, ou seja, vincula a média ao preditor linear, isto é:

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

Sendo g(.) uma função monótona e diferenciável



Seja o valor esperado $E\left(Y_i \middle| x_{i1}, \dots, x_{ip}\right)$ para i = 1, ..., n.

Então, temos que (definição de GLM):

Função da média

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

ou

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})$$

Média em termos da inversa g(.), aplicada ao preditor linear.

PUC Minas
Virtual