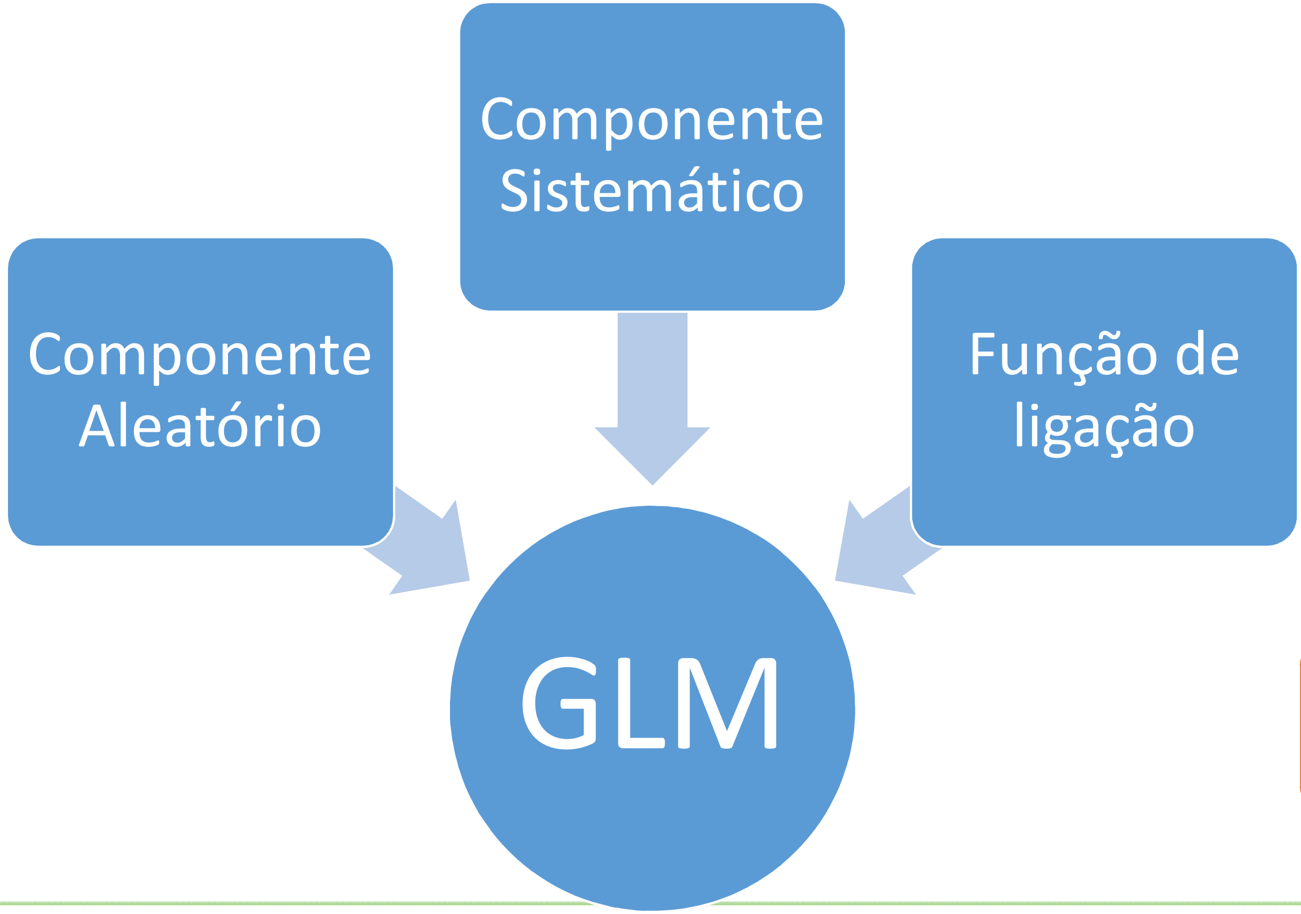
Especificação de um GLM

INTRODUÇÃO : Especificação de um GLM



Especificar os componentes do GLM

Especificação do componente Aleatório

Possíveis distribuições associadas a nossa variável resposta

Dados Contínuos

- Normal
- Gama
- Normal Inversa
- Log Normal ....

Dados Discretos

- Bernoulli
- Poisson
- Binomial
- Binomial negativa ...

Definir a distribuição de probabilidade para a variável resposta

Especificação do componente Aleatório

Para variáveis discreta:

- Proporção
- Contagem

Para variáveis Contínuas:

- Simétrica
- Assimétrica

Tomar cuidado pois estamos avaliando o comportamento da variável resposta na presença das covariáveis

Especificação da resposta - assume valores reais positivos ou tem valores negativos e positivos

No GLM temos que especificar quem é a variável resposta
Na regressão clássica especificávamos quem era nosso erro.

Especificação do componente Aleatório

Como saber que meu modelo é adequado?

Características dos dados encaixam com as propriedade do modelo de probabilidade?

Adequabilidade do modelo
Análise de resíduos.

Suponha ter dados de contagem e selecionamos a Poisson: $Var(Y) = E(Y) = \mu$.

Especificação do componente Aleatório

- Definir uma distribuição de probabilidade para a variável resposta

- A variável resposta é discreta ou contínua? Sua distribuição é simétrica?

- Deve-se propor um modelo que tenha as propriedades compatíveis à distribuição dos dados

- Não se tendo convicção sobre uma particular escolha, pode-se testar diferentes alternativas.

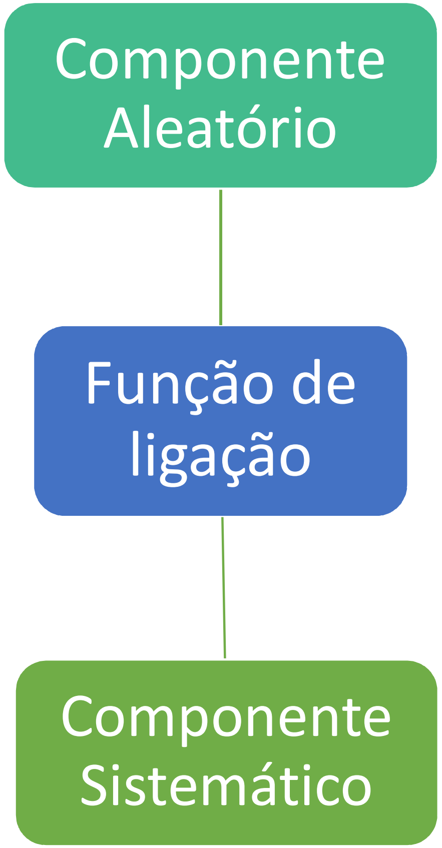
Especificação do Componente Sistemático

- Quais variáveis explicativas devem ser consideradas?
- Como essas variáveis serão incorporadas ao modelo?
- Avaliar a necessidade de escalonar, transformar, categorizar ou incluir variáveis numérica, etc.
- Avaliar a necessidade de incluir efeitos de interação.

$$\mathrm{g}(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \cdots + \beta_p x_{ip}$$

Especificação da função de ligação

- A função de ligação tem o papel **de linearizar a relação entre os componentes aleatório e sistemático do modelo.**
- Deve produzir valores no espaço paramétrico (μ) para qualquer valor produzido por
  - $\eta_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- Apresentar propriedades matemáticas e computacionais desejadas
- Proporcionar interpretações práticas para os parâmetros de regressão β's



Especificação da função de ligação

Mas como assim linearizar?

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

Por exemplo:

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

Log para linearizar

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

Papel da função de ligação: linearizar os dois componentes

- Irá produzir valores no espaço paramétrico no formato:

$$\eta_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- É difícil é saber qual a função de ligação é prioritariamente interessante para cada caso.
- Pode **existir mais de uma função de ligação** -> mais de uma função de ligação que linearize esses dois componentes.
- A função de ligação deve ser aquela na qual quando adicionarmos os dados , estimar os parâmetros e achar **os valores ajustados** , esses valores **devem estar no espaço paramétrico da média da distribuição**.
- tem certos tipos de função de ligação que são interessantes de utilizar pelas **suas propriedades matemáticas e computacionais.**

Distribuição	θ
Normal : N (μ, σ^2)	μ
Poisson: P(μ)	$\log(\mu)$
Binomial: Bin(m, π)	$\log\left(\frac{\mu}{m-\mu}\right)$
Binomial Negativa: BinN(π , k)	$\log\left(\frac{\mu}{\mu+k}\right)$
Gama : G$(\mu$, v)	$-\frac{1}{\mu}$
Normal Inversa: IG (μ, σ^2)	$-\frac{1}{2\mu^2}$

Então na Poisson temos:

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

O interesse é sempre utilizar a função de ligação canônica

Função de ligação canônica

A função de ligação g(.) que transforma a média no parâmetro canônico é a função de ligação canônica , isto é,

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

Ex.: ligação logarítmica para a distribuição Poisson; a logito(logística) para a distribuição binomial a ligação identidade para a normal e etc.

PUC Minas
Virtual