

# CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

21 de agosto, 2017

## Aula 5 - Modelo Linear Generalizado

# Componentes de um modelo linear generalizado

- Um modelo linear generalizado é definido pela especificação de três componentes: o **componente aleatório**, o **componente sistemático** e uma **função de ligação**.
- **Componente aleatório:** Um conjunto de variáveis aleatórias independentes  $y_1, y_2, \dots, y_n$  com distribuição pertencente à família exponencial de dispersão:

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi) \right\}, \quad (1)$$

- Como vimos nas aulas anteriores, são membros dessa família as distribuições binomia, Poisson, normal, gama, normal inversa...

# Componentes de um modelo linear generalizado

- **Componente sistemático:** preditor linear do modelo, em que são inseridas as covariáveis por meio de uma combinação linear de parâmetros.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}. \quad (2)$$

- **Função de ligação:** Função real, monótona e diferenciável, que “liga” o componente aleatório ao sistemático.

# Componentes de um modelo linear generalizado

- Seja  $\mu_i = E(y_i | x_{i1}, x_{i2}, \dots, x_{ip})$ , para  $i = 1, 2, \dots, n$ . Então:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}. \quad (3)$$

- Pelas propriedades de  $g(\cdot)$ , o modelo pode ser escrito de maneira equivalente por:

$$\mu_i = g^{-1}(\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}). \quad (4)$$

# Especificação do componente aleatório

- Definição de uma distribuição de probabilidades para a variável resposta.
- A variável resposta é discreta ou contínua? Sua distribuição é simétrica? Qual o conjunto de valores com probabilidade não nula?
- Deve-se propor um modelo que tenha propriedades compatíveis à distribuição dos dados;
- Não se tendo convicção sobre uma particular escolha, pode-se testar diferentes alternativas ou usar alguma abordagem que não exija essa especificação.

# Especificação do preditor linear

- Quais variáveis explicativas devem ser consideradas?
- Como essas variáveis serão incorporadas ao modelo? Avaliar a necessidade (conveniência) de escalonar, transformar, categorizar ou incluir potências de variáveis numéricas. . .
- Avaliar a necessidade de incluir efeitos de interação. . .

# 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 (para  $\mu_i$ ) para qualquer valor produzido por  $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$ .
- Apresentar propriedades estatísticas e computacionais desejadas (trataremos disso adiante);
- Proporcionar interpretações práticas para os parâmetros de regressão  $\beta$ 's.



# Função de ligação canônica

- A função de ligação  $g(\cdot)$  que transforma a média no parâmetro canônico é a função de ligação canônica:

$$g(\mu_i) = \theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}. \quad (5)$$

- Como exemplos de função de ligação canônicas temos a ligação logarítmica para a distribuição de Poisson; a logito para a distribuição binomial, a identidade para a normal. . .
- A ligação canônica garante algumas simplificações e propriedades desejadas no processo de ajuste do modelo, que serão discutidas na próxima aula.

# Especificação da função de ligação

**Tabela 1:** Exemplos de funções de ligação

Ligação	$\eta = g(\mu)$	$\mu = g^{-1}(\eta)$	Ligação canônica
Identidade	$\mu$	$\eta$	Normal
Logarítmica	$\ln(\mu)$	$e^\eta$	Poisson
Inversa	$\mu^{-1}$	$\eta^{-1}$	Gama
Inversa-quadrada	$\mu^{-2}$	$\eta^{-1/2}$	Normal inversa
Raiz quadrada	$\sqrt{\mu}$	$\eta^2$	
Logito	$\ln \frac{\mu}{1-\mu}$	$\frac{e^\eta}{1+e^\eta}$	Binomial
Probit	$\Phi^{-1}(\mu)$	$\Phi(\eta)$	
Log-log	$-\ln[\ln(\mu)]$	$\exp[-\exp(-\eta)]$	
Clog-log	$\ln[-\ln(1-\mu)]$	$1 - \exp[-\exp(\eta)]$	

## Exercício 1

*Para cada exemplo de motivação apresentado na aula 1, especificar um modelo linear generalizado que você julgue adequado. Determine o componente aleatório, o componente sistemático e a função de ligação. Se tiver dúvidas quanto à função de ligação, opte, neste momento, pela ligação canônica da distribuição escolhida.*

## Exercício 2

*Fazer a lista de exercícios 2, disponível na página da disciplina.*