CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

08 de novembro, 2022

UD 05- 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**.

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 $y_1, y_2, ..., y_n$ ao qual se assume uma particular distribuição pertencente à família exponencial:

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

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 $y_1, y_2, ..., y_n$ ao qual se assume uma particular distribuição pertencente à família exponencial:

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

- Como vimos nas aulas anteriores, são membros dessa família as distribuições binomial, 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.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}.$$

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.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}.$$

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

Componentes de um modelo linear generalizado

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

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

Componentes de um modelo linear generalizado

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

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

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

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

Especificação do componente aleatório

- Definição de uma distribuição de probabilidades para a variável resposta.

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 suporte da variável (conjunto de valores com probabilidade não nula)?

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 suporte da variável (conjunto de valores com probabilidade não nula)?

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

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 suporte da variável (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 (trataremos disso adiante).

Especificação do preditor linear

- Quais variáveis explicativas devem ser consideradas?

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...

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 termos referente a 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.

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 de μ_i para qualquer valor produzido por $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$.

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 de μ_i para qualquer valor produzido por $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$.

- Apresentar propriedades estatísticas e computacionais desejadas (trataremos disso adiante);

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 de μ_i para qualquer valor produzido por $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$.

- Apresentar propriedades estatísticas e computacionais desejadas (trataremos disso adiante);

- Pode ser desejável uma função de ligação que forneça interpretações práticas para os parâmetros do modelo $\beta' s$.

Função de ligação canônica

- A função de ligação $g(\cdot)$ que associa a média ao parâmetro canônico é chamada *função de ligação canônica*:

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

Função de ligação canônica

- A função de ligação $g(\cdot)$ que associa a média ao parâmetro canônico é chamada *função de ligação canônica*:

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

- 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...

Função de ligação canônica

- A função de ligação $g(\cdot)$ que associa a média ao parâmetro canônico é chamada *função de ligação canônica*:

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

- 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 estimação dos parâmetros e ajuste do modelo, que serão discutidas na próxima UD

Especificação da função de ligação

Tabela 1: Exemplos de funções de ligação

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