UNIVERSIDADE FEDERAL DE SERGIPE

DECAT DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS

Modelos lineares generalizados

Disciplina ofertada pelo DECAT/UFS

Código: ESTAT0092
Nível: Graduação
Carga horária: 60h
Período: 2020.2

Professor responsável e ministrante: Luiz Henrique Dore

Aula 3: Formulação de um MLG, função de ligação canônica e casos particulares

Sumário

Metas

1. Introduzir o modelo linear generalizado, apresentando a sua formulação em termos dos componentes aleatório e sistemático e da função de ligação.
2. Apresentar a definição de função de ligação canônica.

Objetivos

- Após estudar essa aula, o aluno ou aluna será capaz de:
  1. formular o modelo linear generalizado, através da especificação dos componentes aleatório e sistemático e da função de ligação;
  2. entender o papel de cada um desses componentes no modelo;
  3. identificar a ligação canônica de uma dada distribuição.

Pré-requisitos

1. Aula 2.

Introdução

- Na aula 2, discutiu-se sobre a importância de se estudar a associação entre duas ou mais variáveis.
- Em geral, pode-se dizer que duas variáveis são associadas, ou relacionadas, quando uma variação sistemática em uma delas é acompanhada por uma variação sistemática na outra.
- O conhecimento da existência de uma associação entre duas ou mais variáveis encontra várias aplicações na prática, tais como predizer, explicar e/ou controlar o valor de uma das variáveis a partir dos valores das demais.
- A modelagem de regressão consiste na aplicação de métodos estatísticos para 1) verificar a existência de associação entre duas ou mais variáveis e 2) usar essa associação para representar uma das variáveis, denominada **variável resposta**, como função das demais, denominadas **variáveis preditoras**.

Introdução

- A representação da variável resposta como uma função das variáveis preditoras é feita por modelos matemáticos conhecidos como **modelos de regressão**.
- A modelagem de regressão é mais comumente utilizada quando se deseja obter o valor **desconhecido** de uma variável resposta a partir dos valores **conhecidos** das variáveis preditoras.
- Um modelo de regressão é composto por equações, as quais descrevem a relação funcional entre a resposta e as variáveis preditoras, e, possivelmente, um conjunto de suposições sobre os termos dessas equações.
- Um modelo de regressão pode ser formulado especificando-se três componentes: 1) uma distribuição de probabilidade para a resposta; 2) uma função η, linear nas preditoras e 3) uma relação funcional entre a média da resposta e η.

Introdução

- Essa formulação comporta, por exemplo, os modelos linear normal, de Poisson e logístico.
- O modelo linear generalizado é formulado dessa mesma maneira.
- Uma das principais características do modelo linear generalizado é que a distribuição de probabilidade da variável resposta pode ser qualquer membro da família exponencial.
- Na presente aula, é apresentada uma formulação para o modelo linear generalizado.
- O papel de cada um dos três componentes que formam o modelo é discutido.
- Por fim, são apresentados alguns casos particulares.

Tópico 4: formulação de um MLG

- Seja Y a variável resposta e sejam $x_1, \cdots, x_p$ os valores das variáveis preditoras.
- O **modelo linear generalizado** (MLG) assume que Y é uma variável aleatória tal que
  1. a distribuição de Y pertence à família exponencial;
  2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$,
  3. $g(\mu) = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$;

  onde μ é a média de Y e g é uma função monótona diferenciável.
- Modelos como o modelo linear normal, de Poisson e logístico especificam uma distribuição de probabilidade para Y e uma relação funcional entre μ e η.
- O MLG, proposto por Nelder e Wedderburn [2], estende esses três modelos, e vários outros, de duas formas.

Tópico 4: formulação de um MLG

- Por um lado, a distribuição da variável resposta deixa de ser uma distribuição específica e passa a ser qualquer membro da família exponencial.
- Por outro, a transformação aplicada à média também deixa de ser uma transformação específica e passa a ser qualquer função monótona e diferenciável g.
- Essas duas generalizações tornam bastante flexíveis os MLGs.
- Além da flexibilidade, os MLGs permitem que teorias e métodos, envolvidos nos processos de estimação, inferência e diagnóstico para diversos modelos de regressão, sejam desenvolvidos de forma unificada.
- Ou seja, em geral, as técnicas desenvolvidas para um MLG servem para todos os seus submodelos.

Tópico 4: formulação de um MLG

- Os três componentes do modelo são denominados **componente aleatório**, **componente sistemático** e **função de ligação**.
- O componente aleatório é a resposta Y e sua distribuição de probabilidade.
- A distribuição de Y deve pertencer à família exponencial.
- Isto é, a função de densidade de probabilidade de Y pode ser escrita como

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

  onde θ e ϕ são parâmetros e $b(\cdot)$ e $c(\cdot,\cdot)$ são funções conhecidas.
- O parâmetro θ é chamado parâmetro canônico e ϕ é o parâmetro de dispersão.

Tópico 4: formulação de um MLG

- Conforme visto na aula 2, a média e a variância de Y são

$$\mu = E(Y) = b'(\theta) \quad \text{e} \quad Var(Y) = \phi b''(\theta).$$

- Note que

$$b''(\theta) = \frac{\partial b'(\theta)}{\partial \theta} = \frac{\partial \mu}{\partial \theta}.$$

- Escrevendo, $\frac{\partial \mu}{\partial \theta}$ como uma função $V(\mu)$ de μ, obtém-se

$$Var(Y) = \phi V(\mu).$$

- A função $V(\mu)$ é denominada **função de variância**.
- A função de variância pode ser constante ou pode, de fato, depender de μ.

Tópico 4: formulação de um MLG

- Quando $V(\mu)$ é constante, a variância de Y também é. Nesse caso, a variável resposta é homoscedástica.
- Se $V(\mu)$ depende de μ, a resposta pode ser heteroscedástica, pois, de acordo com o modelo, μ pode variar com as preditoras.
- Portanto, o MLG pode se adequar às situações nas quais se observa homoscedasticidade e àquelas nas quais se observa heteroscedasticidade.
- A escolha da distribuição de Y é, em geral, baseada na natureza da própria variável resposta Y.
- Por exemplo, se Y representa uma contagem, a distribuição de Poisson é uma escolha plausível.
- Se Y é uma variável binária, a distribuição de Bernoulli é uma escolha mais adequada.

Tópico 4: formulação de um MLG

- O componente sistemático é a função η, a qual é linear nas variáveis preditoras.
- A função η é denominada **preditor linear**.
- O componente sistemático é especificado listando-se as variáveis preditoras que acredita-se estarem associadas à resposta.
- O MLG estabelece uma relação funcional entre a resposta e as variáveis preditoras, assumindo que há uma relação funcional entre a média da resposta e as variáveis preditoras.
- O terceiro componente, a função de ligação g, **lineariza** a relação funcional entre a média e as variáveis preditoras.
- De acordo com o MLG, a média da resposta é dada por

$$\mu = g^{-1}(\eta) = g^{-1}(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p).$$

Tópico 4: formulação de um MLG

- Ou seja, a função de ligação determina a forma como a média varia em termos das variáveis preditoras.
- Os valores possíveis para μ dependem da distribuição de Y.
- Se a distribuição de Y é uma Poisson, então $\mu > 0$; se é uma binomial, então $0 < \mu < 1$.
- Em geral, a função de ligação é escolhida de modo a garantir que $g^{-1}(\eta)$ seja um valor possível para μ.
- Funções de ligações comuns são a identidade $g(\mu) = \mu$, o logaritmo $g(\mu) = \ln(\mu)$ a raiz quadrada $g(\mu) = \sqrt{\mu}$ e a função logit $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$. Veja outros exemplos em [3, p. 8].
- Para mais informações sobre os componentes do MLG, veja a seção 2.3 de [1]

Tópico 4: formulação de um MLG

Exemplo 1

Conforme visto na aula 2, a distribuição normal pertence à família exponencial e sua função de densidade é

$$\begin{aligned} f(y;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\phi^{-1}\left[y\theta - b(\theta)\right] + c(y,\phi)\right\}, \quad y \in \mathbb{R}, \end{aligned}$$

onde

$$\theta = \mu,\ \ b(\theta) = \frac{\theta^2}{2},\ \ \phi = \sigma^2 \text{ e } c(y,\phi) = -\frac{1}{2}\frac{y^2}{\phi} - \frac{1}{2}\ln(2\pi\phi).$$

O modelo de regressão linear normal é um caso particular do MLG, no qual Y segue distribuição normal e a função de ligação é a função identidade $g(\mu) = \mu$.

Tópico 4: formulação de um MLG

Continuação do exemplo 1

Ou seja, o modelo linear normal pode ser formulado como

1. $Y \sim N(\mu, \sigma^2)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = \mu = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Como $\mu \equiv \mu(\theta) = \theta$, a função de variância é

$$V(\mu) = \frac{\partial \mu}{\partial \theta} = 1 \quad \text{e} \quad Var(Y) = \phi V(\mu) = \phi.$$

Ou seja, a variância é constante. Portanto, de acordo com o modelo linear normal, a resposta é homoscedástica. Note que

$$\frac{\partial \mu}{\partial x_j} = \beta_j, \quad j \in \{1, 2, \cdots, p\}.$$

Tópico 4: formulação de um MLG

Continuação do exemplo 1

Portanto, o coeficiente β_i, associado à variável x_j, é a taxa de variação de μ com relação à x_i. Note também que

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_j(x_j + 1) + \cdots + \beta_p x_p = \mu + \beta_j.$$

Isto é, de acordo com o modelo linear normal, μ varia **aditivamente** com relação às variáveis preditoras. Se $\beta_j > 0$, então o aumento (redução) de uma unidade em x_j corresponde a um aumento (redução) de β_j unidades, em média, na variável resposta. Se $\beta_j < 0$, então o aumento (redução) de uma unidade em x_j corresponde a uma redução (aumento) de β_j unidades, em média, na variável resposta.

Tópico 4: formulação de um MLG

Exemplo 2

Conforme visto na aula 2, a distribuição de Poisson pertence à família exponencial e sua função de distribuição é

$$
\begin{aligned}
f(y;\mu,\sigma^2) &= \frac{\mu^y e^{-\mu}}{y!} = \\
&= \exp\left\{\phi^{-1}\cdot[y\theta - b(\theta)] + c(y,\phi)\right\}, \quad y \in \{0,1,\cdots\}
\end{aligned}
$$

onde

$$\theta = \ln(\mu),\ b(\theta) = e^{\theta},\ \phi = 1 \text{ e } c(y,\phi) = -\ln(y!)\,(2\pi\phi)\,.$$

O modelo de regressão Poisson é o caso particular do MLG, no qual Y segue distribuição de Poisson e a função de ligação é o logaritmo $g(\mu) = \ln(\mu)$.

Tópico 4: formulação de um MLG

Continuação do exemplo 2

Ou seja, o modelo de regressão Poisson pode ser escrito como pode ser escrita como

1. $Y \sim \text{Poisson}(\mu)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = \ln(\mu) = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Como $\mu \equiv \mu(\theta) = e^\theta$, a função de variância é

$$V(\mu) = \frac{\partial \mu}{\partial \theta} = e^\theta = \mu \quad \text{e} \quad Var(Y) = \phi V(\mu) = \mu.$$

Ou seja, de acordo com o modelo de regressão Poisson, a variância da resposta é igual à média. Portanto, a resposta pode ser heteroscedástica. A média da resposta é

$$\mu = e^\eta = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p} = e^{\beta_0} e^{\beta_1 x_1} \cdots e^{\beta_p x_p}.$$

Tópico 4: formulação de um MLG

Continuação do exemplo 2

Portanto,

$$\frac{\partial \mu}{\partial x_j} = \beta_j \cdot \mu.$$

Dessa forma, μ varia exponencialmente com relação a x_j, a uma taxa de β_j unidades. Note que

$$e^{\beta_0+\beta_1 x_1+\cdots+\beta_j(x_j+1)+\cdots+\beta_p x_p} = e^{\beta_j}\mu.$$

Isto é, μ varia **multiplicativamente** com relação às variáveis preditoras. Se $\beta_j > 0$, então, quando x_j aumenta (diminui) uma unidade, a resposta fica, em média, e^{β_j} vezes maior (menor). Se $\beta_j < 0$, então, quando x_j aumenta (diminui) uma unidade, a resposta fica, em média, e^{β_j} vezes menor (maior).

Tópico 4: formulação de um MLG

Exemplo 3

Conforme visto na aula 2, a distribuição de Bernoulli pertence à família exponencial e sua função de distribuição é

$$f(y;\mu) = \mu^y(1-\mu)^{1-y} =$$
$$= \exp\left\{\phi^{-1}[y\theta - b(\theta)] + c(y;\phi)\right\} = f(y;\theta,\phi), \quad y \in \{0,1\},$$

onde

$$\theta = \ln\left(\frac{\mu}{1-\mu}\right),\ b(\theta) = \ln\left(1+e^{\theta}\right),\ \phi = 1 \text{ e } c(y;\phi) = 0.$$

O modelo de regressão logística é o caso particular do MLG, no qual Y segue distribuição de Bernoulli e a função de ligação é a função logit $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$.

Tópico 4: formulação de um MLG

Continuação do exemplo 3

Ou seja, o modelo de regressão logística pode ser escrito como

1. $Y \sim \text{Bernoulli}(\mu)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right) = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Como $\mu \equiv \mu(\theta) = \frac{e^\theta}{e^\theta+1}$, a função de variância é

$$V(\mu) = \frac{\partial \mu}{\partial \theta} = \frac{e^\theta}{(e^\theta + 1)^2} = \mu(1-\mu) \quad \text{e} \quad Var(Y) = \phi\mu(1-\mu).$$

Ou seja, de acordo com o modelo de regressão logística, a resposta pode ser heteroscedástica. A média da resposta é

$$\mu = \frac{e^\eta}{e^\eta + 1} = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p} + 1}.$$

Tópico 4: formulação de um MLG

Continuação do exemplo 3

A interpretação dos coeficientes é feita em termos de uma quantidade denominada chance. A chance é definida como

$$odds(\mu) = \frac{\mu}{1-\mu}$$

A chance mede a possibilidade de sucesso relativamente à possibilidade de fracasso. Se $odds(\mu) = 2$, então o sucesso é duas vezes mais frequente do que o fracasso. Note que

$$\frac{\partial}{\partial x_j} odds(\mu) = \beta_j \cdot odds(\mu).$$

Ou seja, $odds(\mu)$ varia exponencialmente com relação a x_j, a uma taxa de β_j unidades.

Tópico 4: formulação de um MLG

Continuação do exemplo 3

Note também que

$$e^{\beta_0+\beta_1 x_1+\cdots+\beta_j(x_j+1)+\cdots+\beta_p x_p} = e^{\beta_j} \cdot odds(\mu).$$

Isto é, $odds(\mu)$ varia **multiplicativamente** com relação às preditoras. Se $\beta_j > 0$, então, quando x_j aumenta (diminui) uma unidade, a chance fica e^{β_j} vezes maior (menor). Se $\beta_j < 0$, então, quando x_j aumenta (diminui) uma unidade, a chance fica e^{β_j} vezes menor (maior).

Tópico 5: função de ligação canônica

- A função de ligação g é dita ser uma **ligação canônica** se

$$g(\mu) \equiv g(\mu(\theta)) = \theta.$$

- Pela formulação do MLG, se g é uma ligação canônica, então

$$\theta = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

- Isto é, Pela formulação do MLG, se g é uma ligação canônica, então **o parâmetro canônico é igual ao preditor linear**.
- O uso de ligações canônicas adiciona importantes, teóricas e práticas, vantagens ao MLG, como, por exemplo, a unicidade da estimativa de máxima verossimilhança dos coeficientes, quando esta existe [3, p. 8]

Tópico 5: função de ligação canônica

Exemplo 4

Conforme visto na aula 2, a distribuição normal pertence à família exponencial e sua função de densidade é

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\} \\ &= \exp\left\{ \phi^{-1}\left[y\theta - b(\theta) \right] + c(y, \phi) \right\}, \quad y \in \mathbb{R}, \end{aligned}$$

onde

$$\theta = \mu, \; b(\theta) = \frac{\theta^2}{2}, \; \phi = \sigma^2 \text{ e } c(y, \phi) = -\frac{1}{2}\frac{y^2}{\phi} - \frac{1}{2}\ln(2\pi\phi).$$

Portanto, a função de ligação canônica da distribuição normal é a função identidade

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

Tópico 5: função de ligação canônica

Exemplo 5

Conforme visto na aula 2, a distribuição de Poisson pertence à família exponencial e sua função de distribuição é

$$f(y;\mu,\sigma^2) = \frac{\mu^y e^{-\mu}}{y!} =$$
$$= \exp\left\{\phi^{-1} \cdot [y\theta - b(\theta)] + c(y,\phi)\right\}, \quad y \in \{0, 1, \cdots\}$$

onde

$$\theta = \ln(\mu),\ b(\theta) = e^{\theta},\ \phi = 1 \text{ e } c(y,\phi) = -\ln(y!)\,(2\pi\phi)\,.$$

Portanto, a função de ligação canônica da distribuição de Poisson o logaritmo

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

Tópico 5: função de ligação canônica

Exemplo 6

Conforme visto na aula 2, a distribuição de Bernoulli pertence à família exponencial e sua função de distribuição é

$$f(y;\mu) = \mu^y(1-\mu)^{1-y} =$$
$$= \exp\left\{\phi^{-1}[y\theta - b(\theta)] + c(y;\phi)\right\} = f(y;\theta,\phi), \quad y \in \{0,1\},$$

onde

$$\theta = \ln\left(\frac{\mu}{1-\mu}\right), b(\theta) = \ln\left(1+e^{\theta}\right), \phi = 1 \text{ e } c(y;\phi) = 0.$$

Portanto, a função de ligação canônica da distribuição de Bernoulli é a função logit

$$\text{logit}(\mu) = g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right).$$

Tópico 6: outros casos particulares

Exemplo 7

Conforme visto na aula 2, a distribuição gama pertence à família exponencial e sua função de densidade é

$$
\begin{aligned}
f(y;\mu,\sigma^2) &= \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by} \\
&= \exp\left\{\phi^{-1}[\theta y - b(\theta)] + c(y,\phi)\right\}, \quad y > 0,
\end{aligned}
$$

onde

$$\theta = -\frac{b}{a}, \quad b(\theta) = -\ln(-\theta), \quad \phi = a^{-1}$$

e

$$c(y,\phi) = -\phi \ln(\phi) - \ln\Gamma(\phi) + (\phi - 1)\ln(y)$$

Tópico 6: outros casos particulares

Continuação do exemplo 7

No exemplo 14 do tópico 2, viu-se que

$$\mu \equiv \mu(\theta) = -\frac{1}{\theta}.$$

Portanto, a função de ligação canônica da distribuição gama é

$$g(\mu) = -\frac{1}{\mu}.$$

O modelo de regressão gama, com ligação canônica, pode ser formulado como

1. $Y \sim Gama(a, b)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = -\frac{1}{\mu} = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Tópico 6: outros casos particulares

Continuação do exemplo 7

A função de variância é

$$V(\mu) = \frac{\partial \mu}{\partial \theta} = \frac{1}{\theta^2} = \mu^2 \quad \text{e} \quad Var(Y) = \phi V(\mu) = \phi \mu^2.$$

Ou seja, o modelo de regressão, baseado na gama, supõe que a resposta pode ser heteroscedástica. Modelos de regressão gama são utilizados em situações nas quais a variável resposta assume apenas valores positivos [3, p. 114]. Note que

$$\frac{\partial g(\mu)}{\partial x_j} = \beta_j, \quad j \in \{1, 2, \cdots, p\}.$$

Portanto, o coeficiente β_j, associado à variável x_j, é a taxa de variação de $g(\mu) = -\frac{1}{\mu}$ com relação à x_j.

Tópico 6: outros casos particulares

Continuação do exemplo 7

Note também que

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_j(x_j + 1) + \cdots + \beta_p x_p = \mu + \beta_j.$$

Isto é, de acordo com o modelo de regressão gama, com ligação canônica, $-\frac{1}{\mu}$ varia **aditivamente** com relação às variáveis preditoras. Se $\beta_j > 0$, então o aumento (redução) de uma unidade em x_j corresponde a um aumento (redução) de β_j unidades em $-\frac{1}{\mu}$. Se $\beta_j < 0$, então o aumento (redução) de uma unidade em x_j corresponde a uma redução (aumento) de β_j unidades em em $-\frac{1}{\mu}$.

Tópico 6: outros casos particulares

Exemplo 8

O modelo de regressão gama, com ligação canônica, estabelece uma relação linear entre $-\frac{1}{\mu}$ e o preditor linear. Como o preditor linear pode assumir quaisquer valores reais, esse modelo pode atribuir valores negativos à μ, contrariando o fato de que, no caso da distribuição gama, $\mu > 0$. Para contornar esse problema, pode-se utilizar, como função de ligação, o logaritmo. O modelo de regressão gama, com ligação logarítmica, pode ser formulado como

1. $Y \sim Gama(a, b)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = \ln(\mu) = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Tópico 6: outros casos particulares

Continuação do exemplo 8

A média da resposta é

$$\mu = e^{\eta} = e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p} = e^{\beta_0} e^{\beta_1 x_1} \cdots e^{\beta_p x_p}.$$

Note que $\frac{\partial \mu}{\partial x_j} = \beta_j \cdot \mu$. Dessa forma, μ varia exponencialmente com relação a x_j, a uma taxa de β_j unidades. Agora,

$$e^{\beta_0+\beta_1 x_1+\cdots+\beta_j(x_j+1)+\cdots+\beta_p x_p} = e^{\beta_j}\mu.$$

Isto é, μ varia **multiplicativamente** com relação às variáveis preditoras. Se $\beta_j > 0$, então, quando x_j aumenta (diminui) uma unidade, a resposta fica, em média, e^{β_j} vezes maior (menor). Se $\beta_j < 0$, então, quando x_j aumenta (diminui) uma unidade, a resposta fica, em média, e^{β_j} vezes menor (maior).

Tópico 6: outros casos particulares

Exemplo 9

Conforme visto na aula 2, a distribuição binomial pertence à família exponencial e sua função de probabilidade é

$$f(y;\pi,m) = \binom{m}{y}\pi^y(1-\pi)^{m-y} =$$
$$= \exp\left\{\phi^{-1}[y\theta - b(\theta)] + c(y,\phi)\right\} = f(y;\theta,\phi),$$

onde $y \in \{0,1,\cdots,m\}$,

$$\theta = \ln\left(\frac{\pi}{1-\pi}\right), \quad b(\theta) = m\ln\left(1+e^{\theta}\right), \quad \phi = 1$$

e

$$c(y;\phi) = \ln\binom{m}{y}.$$

Tópico 6: outros casos particulares

Continuação do exemplo 9

No exemplo 13 da aula 2, viu-se que

$$E(Y) = \mu \equiv \mu(\theta) = m\frac{e^{\theta}}{e^{\theta}+1} = m\pi.$$

Portanto, a ligação canônica da distribuição binomial é

$$g(\mu) = \ln\left(\frac{\mu/m}{1-\mu/m}\right) = \ln\left(\frac{\mu}{m-\mu}\right).$$

O modelo de regressão binomial, com ligação canônica, pode ser formulado como

1. $Y \sim Binomial(\pi, m)$;
2. $\eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
3. $g(\mu) = \ln\left(\frac{\mu}{m-\mu}\right) = \eta(x_1, \cdots, x_p; \beta_0, \cdots, \beta_p)$.

Tópico 6: outros casos particulares

Continuação do exemplo 9

A função de variância é

$$V(\mu) = \frac{\partial \mu}{\partial \theta} = m\frac{e^{\theta}}{(e^{\theta}+1)^2} = \mu(1-\mu/m) \text{ e } Var(Y) = \mu(1-\mu/m).$$

Ou seja, o modelo de regressão, baseado na binomial, supõe que a resposta pode ser heteroscedástica. De acordo com o modelo de regressão binomial, com ligação canônica,

$$\mu = m\frac{e^{\eta}}{e^{\eta}+1} = m\frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}+1}.$$

Os coeficientes podem ser interpretados em termos da chance, definida como $odds(\mu/m) = \frac{\mu}{m-\mu}$ (ver exemplo 3).

Tópico 6: outros casos particulares

Continuação do exemplo 9

A chance expressa o número médio de sucessos relativamente ao número médio de fracassos. Se $odds(\mu/m) = 2$, por exemplo, então o número médio de sucessos é duas vezes maior do que o número médio de fracassos. Note que

$$\frac{\partial}{\partial x_j} odds(\mu/m) = \beta_j \times odds(\mu/m).$$

Ou seja, $odds(\mu/m)$ varia exponencialmente com relação a x_j, a uma taxa de β_j unidades. Note também que

$$e^{\beta_0+\beta_1 x_1+\cdots+\beta_j(x_j+1)+\cdots+\beta_p x_p} = e^{\beta_j} \times odds(\mu/m).$$

Tópico 6: outros casos particulares

Continuação do exemplo 9

Isto é, $odds(\mu/m)$ varia **multiplicativamente** com relação às preditoras. Se $\beta_j > 0$, então, quando x_j aumenta (diminui) uma unidade, a chance fica e^{β_j} vezes maior (menor). Se $\beta_j < 0$, então, quando x_j aumenta (diminui) uma unidade, a chance fica e^{β_j} vezes menor (maior).

Referências I

[1] G. M. Cordeiro e C. G. B. Demétrio, *Modelos lineares generalizados e extensões*, 2013, disponível em https://docs.ufpr.br/~taconeli/CE22518/LivClarice.pdf.

[2] J. A. Nelder and R. W. M. Wedderburn, *Generalized linear models*, Journal of the Royal Statistical Society. Series A (General) **135** (1972), no. 3, 370–384.

[3] G. A. Paula, *Modelos lineares generalizados com apoio computacional*, 2013, disponível em https://www.ime.usp.br/~giapaula/texto_2013.pdf.