Modelos Lineares Generalizados - Componentes do Modelo

Erica Castilho Rodrigues

11 de Maio de 2017

Introdução

- Vejamos agora quais as componentes de um Modelo Linear Generalizado.
- Temos um conjunto de variáveis aleatórias independentes

$$Y_1, Y_2, \dots, Y_n$$

  cada uma delas com uma distribuição que pertence à Família Exponencial.
- Vamos supor que as variáveis estão na forma canônica.
- Vimos que em geral a densidade de distribuições pertencentes a família exponencial podem ser escritas como

$$f(y;\theta) = s(\mathbf{y})t(\theta)e^{a(\mathbf{y})b(\theta)} \, .$$

- Vimos ainda que se $a(y) = y$,
  - $b(\theta)$ é chamado paramêtro canônico da distribuição.

- Isso equivale a dizer que a variável está na forma canônica.
- Vamos considerar ainda que a distribuição depende apenas de um parâmetro θ_i.
- Os demais parâmetros são conhecidos ou são apenas parâmetros de ruído, como o σ^2.
- Nesse caso temos que a função densidade é dada por

$$f(y_i;\theta_i) = \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right]$$

- A função de densidade conjunta é dada por

$$f(y_1, y_2, \ldots, y_N; \theta_1, \ldots, \theta_N) = \prod_{i=1}^{N} \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right]$$

$$= \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right] .$$

- A variáveis são i.i.d.?

- A função de densidade conjunta é dada por

$$f(y_1, y_2, \ldots, y_N; \theta_1, \ldots, \theta_N) = \prod_{i=1}^{N} \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right]$$

$$= \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right] .$$

- A variáveis são i.i.d.? Não.
- Não são identicamente distribuídas.
- Cada variávei tem seu θ_i.
- Esse parâmetro muda de uma variável para outra.

- Não estamos interessados diretamente no parâmetro θ_i.
- Estamos interessados, geralmente, em um conjunto menor de parâmetros

$$\beta_1, \beta_2, \ldots, \beta_p$$

onde $p < N$.

- θ_i é o valor esperado de Y_i?

- Não estamos interessados diretamente no parâmetro θ_i.
- Estamos interessados, geralmente, em um conjunto menor de parâmetros

$$\beta_1, \beta_2, \dots, \beta_p$$

onde $p < N$.
- θ_i é o valor esperado de Y_i? Não necessariamente.
- Em alguns casos sim, por exemplo na Normal.
- Vamos denotar

$$E(Y_i) = \mu_i \; .$$

- Em geral, vamos ter que μ_i é função de θ_i.

- No modelo de Regressão Linear a relação entre o valor esperado d variável resposta e as explicativas é dada por

$$E(Y_i) = \mu_i = x_i^T \beta \, .$$

- Nos MLG's podemos definir

$$E(Y_i) = \mu_i = g(x_i^T \beta)$$

  onde $g(.)$ é uma função genérica e é chamada **função de ligação.**
- O termo

$$x_i^T \beta$$

  é chamado **preditor linear.**

- Um Modelo Linear Generalizado é composto por 3 componentes:
  - uma distribuição de probabilidade pertencente à família exponencial;
  - um preditor linear
    $$\eta = x_i^T \boldsymbol{\beta} \, .$$
  - uma função de ligação $g(.)$ tal que
    $$E(Y_i) = \mu_i = g(x_i^T \boldsymbol{\beta}) \, .$$

Preditor Linear

- Incorpora informação sobre as variáveis explicativas no modelo.
- Ele está relacionado com μ_i através da função de ligação

$$\mu_i = g(x_i^T \boldsymbol{\beta}) \,.$$

- É uma combinação linear de parâmetros desconhecidos

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Função de Ligação

- Fornece a relação entre o preditor linear e a média da variável.
- A escolha dessa função pode ser arbitrária.
- Veremos que para cada modelo existem escolhas mais convenientes.
- É interessante que a imagem de $g(.)$ esteja restrita aos valores onde μ_i está definido.
- Exemplo:
  - temos uma distribuição Poisson,
  - logo $\mu_i > 0$,
  - seria interessante então que

$$g(x_i^T \beta) > 0 \, .$$

- Em alguns casos não é possível garantir isso.

Função de Ligação Canônica

- Cada distribuição da Família Exponencial tem uma função de ligação especial.
- Essa ligação ocorre quando

$$b(\theta_i) = \eta = x_i^T \beta \, .$$

- Ou seja, igualamos o **parâmetro canônico** ao **preditor linear**.
- Veremos que ela defini uma estatística suficiente.

- Vamos substituir
$$b(\theta_i) = x_i^T \boldsymbol{\beta}$$
na função de densidade.
- Ficamos com

$$f(\mathbf{y}; \theta_1, \ldots, \theta_N) =$$

- Vamos substituir
$$b(\theta_i) = x_i^T \boldsymbol{\beta}$$
na função de densidade.
- Ficamos com

$$f(\mathbf{y}; \theta_1, \ldots, \theta_N) = \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right]$$

$$= \exp\left[\sum_{i=1}^{N} y_i x_i^T \boldsymbol{\beta} + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right]$$

$$= \exp\left[\sum_{i=1}^{N} y_i \sum_{j=1}^{p} x_{ij}\beta_j + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right]$$

$$= \exp\left[\sum_{j=1}^{p} \beta_j \sum_{i=1}^{N} y_i x_{ij} + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right]$$

- Qual estatística suficiente para estimar β_1?

- Qual estatística suficiente para estimar β_1?

$$\sum_{i=1}^{N} y_i x_{i1}$$

- E para β_j?

- Qual estatística suficiente para estimar β_1?

$$\sum_{i=1}^{N} y_i x_{i1}$$

- E para β_j?

$$\sum_{i=1}^{N} y_i x_{ij}$$

- Veremos que as ligações canônicas facilitam na estimação dos parâmetros.
- Elas tornam a função de densidade côncava.
- Portanto possuem um ponto de máximo global.
- A tabela a seguir mostra alguns exemplos de ligações canônicas

Distribuição	Normal	Binomial	Poisson	Gama	Normal Inversa
Ligação	$\theta = \eta$	$\log\left(\frac{\theta}{1-\theta}\right) = \eta$	$\log(\mu) = \eta$	$\mu^{-1} = \eta$	$\mu^{-2} = \eta$

Exemplo:

- O modelo de Regressão Linear é um caso específico de um MLG.
- Temos nesse caso que

$$Y_i \sim N(\mu_i, \sigma^2) \quad E(Y_i) = \mu_i = x_i^T \boldsymbol{\beta} \,.$$

- A função de ligação é a identidade, pois a média é igual ao preditor linear.
- Esse modelo é geralmente escrito da forma

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

onde $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$ são i.i.d. tais que $\epsilon_i \sim N(0, \sigma^2)$.

Exemplo:

- Uma clínica local está interessada em evitar epidemias de gripe.
- A clínica envia informativos aos pacientes.
- Esses informativos tem como objetivo incentivá-los a tomar uma vacina contra a doença.
- 50 pacientes foram selecionados ao acaso.
- E para cada um deles verificou-se se, de fato, tomaram a vacina.

Exemplo: (continuação)

- Defina a variável

$$Y_i = \begin{cases} 1 & \text{se o paciente toma vacina} \\ 0 & \text{se o paciente nao toma a vacina.} \end{cases}$$

- Qual a distribuição de Y_i?

Exemplo: (continuação)

- Defina a variável

$$Y_i = \begin{cases} 1 & \text{se o paciente toma vacina} \\ 0 & \text{se o paciente nao toma a vacina.} \end{cases}$$

- Qual a distribuição de Y_i?

$$Y_i \sim Bern(\theta_i)$$

onde θ_i é a probabilidade do paciente tomar a vacina.

- A probabilidade θ_i pode depender de variáveis como:
  - idade;
  - conhecimento sobre a doença;
  - sexo.

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta_i) =$$

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta_i) = \theta^{y_i}(1 - \theta_i)^{1-y_i} =$$

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta_i) = \theta^{y_i}(1-\theta_i)^{1-y_i} = \exp\{y_i \log \theta_i + (1 - y_i)\log(1 - \theta_i)\}$$

$$=$$

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta_i) = \theta^{y_i}(1-\theta_i)^{1-y_i} = \exp\{y_i \log \theta_i + (1 - y_i)\log(1 - \theta_i)\}$$

$$= \theta^{y_i}(1 - \theta_i)^{1-y_i} =$$

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta_i) = \theta^{y_i}(1-\theta_i)^{1-y_i} = \exp\{y_i \log \theta_i + (1 - y_i) \log(1 - \theta_i)\}$$

$$= \theta^{y_i}(1 - \theta_i)^{1-y_i} = \exp\left\{ y_i \underbrace{\log\left(\frac{\theta_i}{1 - \theta_i}\right)}_{b(\theta_i)} + \underbrace{\log(1 - \theta_i)}_{c(\theta_i)} \right\}$$

Exemplo: (continuação)

- Vejamos primeiro como fica a ligação canônica nesse caso.
- Lembre que a ligação canônica é dada por

$$b(\theta_i) = x_i^T \beta$$

- Nesse caso temos então que

Exemplo: (continuação)

- Vejamos primeiro como fica a ligação canônica nesse caso.
- Lembre que a ligação canônica é dada por

$$b(\theta_i) = x_i^T \boldsymbol{\beta}$$

- Nesse caso temos então que

$$b(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) \Rightarrow \log\left(\frac{\theta_i}{1-\theta_i}\right) = x_i^T \boldsymbol{\beta} = \eta \, .$$

- Nesse caso temos ainda que

$$E(Y_i) = \theta_i \, .$$

Exemplo: (continuação)

- Vamos isolar θ_i para verificarmos qual é a função $g(.)$ nesse caso.
- Temos que

Exemplo: (continuação)

- Vamos isolar θ_i para verificarmos qual é a função $g(.)$ nesse caso.
- Temos que

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \eta$$

Exemplo: (continuação)

- Vamos isolar θ_i para verificarmos qual é a função $g(.)$ nesse caso.
- Temos que

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \eta$$

$$\left(\frac{\theta_i}{1-\theta_i}\right) = e^{\eta}$$

$$\theta_i = e^{\eta} - \theta_i e^{\eta} \quad \therefore \quad \theta_i(1+e^{\eta}) = e^{\eta}$$

$$\theta_i = \frac{e^{\eta}}{1+e^{\eta}}$$

Exemplo: (continuação)

- Dividindo o numerador e o denominador por e^{η} ficamos com
$$\theta_i = \frac{e^{\eta}(1/e^{\eta})}{(1+e^{\eta})(1/e^{\eta})} = \frac{1}{1+e^{-\eta}}$$
- Ou seja, a função de ligação é dada por
$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T\boldsymbol{\beta}}}$$
- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de μ_i?

Exemplo: (continuação)

- Dividindo o numerador e o denominador por e^η ficamos com

$$\theta_i = \frac{e^\eta(1/e^\eta)}{(1+e^\eta)(1/e^\eta)} = \frac{1}{1+e^{-\eta}}$$

- Ou seja, a função de ligação é dada por

$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T\boldsymbol{\beta}}}$$

- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de μ_i? Não.
- Em que intervalo μ_i está definido?

Exemplo: (continuação)

- Dividindo o numerador e o denominador por e^{η} ficamos com

$$\theta_i = \frac{e^{\eta}(1/e^{\eta})}{(1+e^{\eta})(1/e^{\eta})} = \frac{1}{1+e^{-\eta}}$$

- Ou seja, a função de ligação é dada por

$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T\boldsymbol{\beta}}}$$

- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de μ_i? Não.
- Em que intervalo μ_i está definido? $[0, 1]$.
- Observe que a função

$$\frac{1}{1+e^{-\eta}}$$

está sempre dentro do intervalo [0,1].

Exemplo: (continuação)

- A ligação logística é a mais usada para dados Benoulli.
- Existem, porém, outras possibilidades.
- Probit

$$\Phi^{-1}(\theta_i) = x_i^T \boldsymbol{\beta} \Rightarrow \theta_i = \Phi(x_i^T \boldsymbol{\beta})$$

  onde $\Phi(.)$ é a função de distribuição acumulada da $N(0,1)$.
- Essa função também está definida no intervalo [0,1].

Exemplo: **(continuação)**

- Complemento log-log

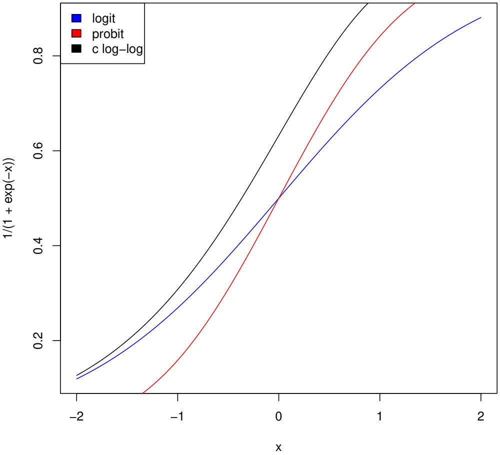
$$\log\left(-\log\left(1-\theta_i\right)\right) = x_i^T\boldsymbol{\beta}$$

$$-\log(1-\theta_i) = e^{x_i^T\boldsymbol{\beta}} \quad \therefore \quad (1-\theta_i) = \exp\left(-e^{x_i^T\boldsymbol{\beta}}\right)$$

$$\theta_i = 1 - \exp\left(-e^{x_i^T\boldsymbol{\beta}}\right)$$

- Também está definida no intervalo [0,1].
- Para valores pequenos de θ_i:
  - fica bem próxima da ligação logística.

Exemplo: **(continuação)** O gráfico a seguir compara as três funções de ligação.



Exemplo:

- Queremos modelar o número de pessoas infectadas por uma determinada doença.
- Seja Y_i o número de casos do i-ésimo grupo de pessoas.
- Suponha que a doença seja não contagiosa.
- Vamos considerar que o número de casos nas cidades são independentes entre si.
- Vamos considerar ainda que o número de casos segue uma distribuição

Exemplo:

- Queremos modelar o número de pessoas infectadas por uma determinada doença.
- Seja Y_i o número de casos do i-ésimo grupo de pessoas.
- Suponha que a doença seja não contagiosa.
- Vamos considerar que o número de casos nas cidades são independentes entre si.
- Vamos considerar ainda que o número de casos segue uma distribuição de Poisson, ou seja,

$$f(y_i, \theta_i) = \frac{\theta_i^{y} e^{-\theta_i}}{y_i!}$$

onde θ_i denota o número de casos esperados no grupo i.

Exemplo: (continuação)

- A função de densidade pode ser escrita como

Exemplo: (continuação)

- A função de densidade pode ser escrita como

$$f(y_i, \theta_i) =$$

Exemplo: (continuação)

- A função de densidade pode ser escrita como

$$f(y_i, \theta_i) = \exp\left\{ y_i \underbrace{\log(\theta_i)}_{b(\theta_i)} - \underbrace{\theta_i}_{c(\theta_i)} \underbrace{- \log(y_i!)}_{d(y_i)} \right\}.$$

- Como fica a ligação canônica nesse caso?

Exemplo: (continuação)

- A função de densidade pode ser escrita como

$$f(y_i, \theta_i) = \exp\left\{ y_i \underbrace{\log(\theta_i)}_{b(\theta_i)} - \underbrace{\theta_i}_{c(\theta_i)} \underbrace{- \log(y_i!)}_{d(y_i)} \right\}.$$

- Como fica a ligação canônica nesse caso?

$$b(\theta_i) = \log(\theta_i)$$

  vimos que a ligação canônica é dada por

Exemplo: (continuação)

- A função de densidade pode ser escrita como

$$f(y_i, \theta_i) = \exp\left\{ y_i \underbrace{\log(\theta_i)}_{b(\theta_i)} - \underbrace{\theta_i}_{c(\theta_i)} \underbrace{- \log(y_i!)}_{d(y_i)} \right\}.$$

- Como fica a ligação canônica nesse caso?

$$b(\theta_i) = \log(\theta_i)$$

vimos que a ligação canônica é dada por

$$b(\theta_i) = \eta = x_i^T \boldsymbol{\beta}$$

portanto ficamos com

$$\log(\theta_i) = x_i^T \boldsymbol{\beta} \text{ ou } \theta_i = e^{x_i^T \boldsymbol{\beta}}$$

Exemplo: (continuação)

- O número esperado de casos pode depender de variáveis como:

Exemplo: (continuação)

- O número esperado de casos pode depender de variáveis como:
  - tamanho da população;
  - características da população, como faixa etária, sexo, história média.
- Uma das variáveis que devemos considerar aqui é o tamanho da população.
- Talvez um local tenha mais casos apenas porque sua população é maior.
- Essa variável entra no modelo sem um β associado.
- Ela é chamada *offset*.
- Veremos mais detalhes sobre isso adiante.

- Modelamos na verdade é a proporção de pessoas da população que são infectadas.
- Essa proporção é modelada por

$$e^{(x_i^T \boldsymbol{\beta})}$$

onde as variáveis x incluem caracterísiticas da população.

- Dessa maneira, o número de casos esperados na população é dado por

$$\theta_i = ne^{(x_i^T \boldsymbol{\beta})}$$

onde
  - n é o tamanho da população sob risco;
  - $e^{(x_i^T \boldsymbol{\beta})}$ é a taxa de casos da doença.
- A taxa de casos vai depender das características da população.

Exemplo: (continuação)

- Suponha que queremos verificar como o a incidência da doença depende da idade.
- Pegamos grupos de faixas etárias distintas

$$30-34 \quad 35-39 \quad 40-44 \quad 45-49 \quad 50-54 \quad 55-59 \quad 60-64 \quad 6$$

- Para cada um desses grupos contamos o número de casos da doença.
- A taxa da doença em cada grupo é estimada por

$$\frac{\text{número de casos no grupo}}{\text{tamanho do grupo}} = \frac{y_i}{n_i}.$$

- Se a doença é rara esse número é muito pequeno, da ordem de 10^{-5}.

Exemplo: (continuação)

- Portanto multiplicamos o resultado por 100.000

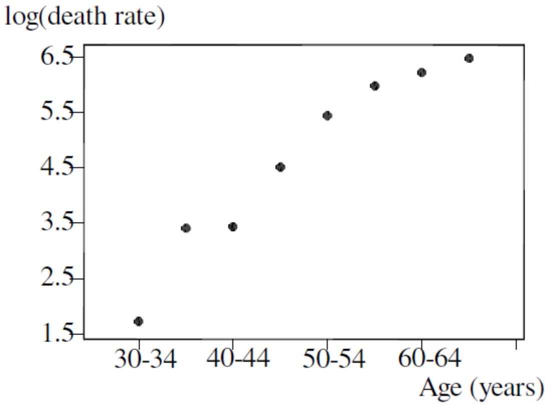
$$\frac{y_i}{n_i} \times 100.000 \, .$$

- Um valor de 5,2, por exemplo, significa que :
  - a taxa de casos da doença é de 5,2 casos por 100.000 habitantes.
- Veja a tabela abaixo.

Age group (years)	Number of deaths, y_i	Population size, n_i	Rate per 100,000 men per year, $y_i/n_i \times 10,000$
30 - 34	1	17,742	5.6
35 - 39	5	16,554	30.2
40 - 44	5	16,059	31.1
45 - 49	12	13,083	91.7
50 - 54	25	10,784	231.8
55 - 59	38	9,645	394.0
60 - 64	54	10,706	504.4
65 - 69	65	9,933	654.4

Exemplo: (continuação)

- A figura a seguir mostra como o log da taxa de casos varia com a idade.



- A relação entre o log da taxa e a idade parece linear.
- Quanto maior a idade maior o número de casos.

Exemplo: (continuação)

- A função de ligação canônica parece adequada nesse caso?

Exemplo: (continuação)

- A função de ligação canônica parece adequada nesse caso?
- Sim.
- A relação entre a taxa de incidência da doença e a idade parece de fato exponencial.

Exemplo:

- Estamos analisando a descendência entre línguas.
- O Grego, por exemplo, é descendente do grego antigo.
- Suponha que duas línguas são seperadas por um tempo t.
- A proximidade entre duas línguas é dada pelo número de palavras iguais com o mesmo sentido.
- Por exemplo, “amigo” tem o mesmo sentido no português do Brasil ou de Portugal.

Exemplo: (continuação)

- A probabilidade de terem a mesma palavra representando o mesmo significado é dada por

$$e^{-\theta t}$$

  quanto mais distantes, menor a probabilidade de terem palavras em comum.
- O parâmetro θ controla a velocidade com que essa probabilidade decai para zero.
- Se o valor de θ é pequeno:
  - mesmo duas línguas muito distantes no tempo tendem a ser parecidas.

Exemplo: (continuação)

- Vamos denotar por por Y_i a variável indicadora que diz se duas línguas tem palavras em comum ou não.
- Portanto

$$Y_i = \begin{cases} 1 & \text{se as línguas usam a mesma palavra para} \\ & \text{representar o i-ésim signficado} \\ 0 & \text{caso contrário.} \end{cases}$$

- Temos então que

$$P(Y_i = 1) = e^{-\theta t}$$

$$P(Y_i = 0) = 1 - e^{-\theta t}$$

- Esse é um caso especial de qual distribuição?

Exemplo: (continuação)

- Vamos denotar por por Y_i a variável indicadora que diz se duas línguas tem palavras em comum ou não.
- Portanto

$$Y_i = \begin{cases} 1 & \text{se as línguas usam a mesma palavra para} \\ & \text{representar o i-ésim signficado} \\ 0 & \text{caso contrário.} \end{cases}$$

- Temos então que

$$P(Y_i = 1) = e^{-\theta t}$$

$$P(Y_i = 0) = 1 - e^{-\theta t}$$

- Esse é um caso especial de qual distribuição? Bernoulli.
- A diferença é que o parâmetro de interesse não é a probabilidade de sucesso.
- Estamos interessados no parâmetro θ.

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta) =$$

Exemplo: (continuação)

- A função densidade é dada por

$$f(y_i, \theta) = (e^{-\theta t})^{y_i}(1 - e^{-\theta t})^{1-y_i}\ .$$

- Pode ser reescrita como

$$f(y_i, \theta) =$$

Exemplo: **(continuação)**

- A função densidade é dada por

$$f(y_i, \theta) = (e^{-\theta t})^{y_i}(1 - e^{-\theta t})^{1-y_i} \, .$$

- Pode ser reescrita como

$$f(y_i, \theta) = \exp\left\{ y_i \log(e^{-\theta t}) + (1 - y_i) \log(1 - e^{-\theta t}) \right\}$$

$$= \exp\left\{ y_i \log\left(\frac{e^{-\theta t}}{1 - e^{-\theta t}} \right) + \log(1 - e^{-\theta t}) \right\}$$

- Como fica a licação canônica nesse caso?

Exemplo: (continuação)

- Temos que

$$b(\theta) = \log\left(\frac{e^{-\theta t}}{1 - e^{-\theta t}}\right)$$

e portanto a ligação canônica é dada por

Exemplo: (continuação)

- Temos que

$$b(\theta) = \log\left(\frac{e^{-\theta t}}{1 - e^{-\theta t}}\right)$$

e portanto a ligação canônica é dada por

$$\log\left(\frac{e^{-\theta t}}{1 - e^{-\theta t}}\right) = x_i^T \boldsymbol{\beta} \,.$$