FAMÍLIA EXPONENCIAL

Modelos Lineares Generalizados

Luan Fiorentin

12 de maio de 2021

Sumário

Introdução

Introdução

- **Inventário florestal** envolve a coleta de informações quali- quantitativas de variáveis da floresta.
- **Variáveis**:
  - Altura;
  - Diâmetro;
  - Presença de ataques de pragas;
  - Número de árvores por unidade amostral;
  - Entre outras.
- Fundamental reconhecer a **natureza** das variáveis.

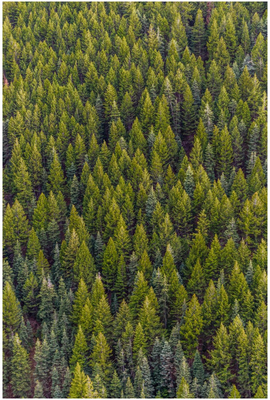

Figura 1. Floresta. Imagem de Brandon Montrone no Pexels

Introdução

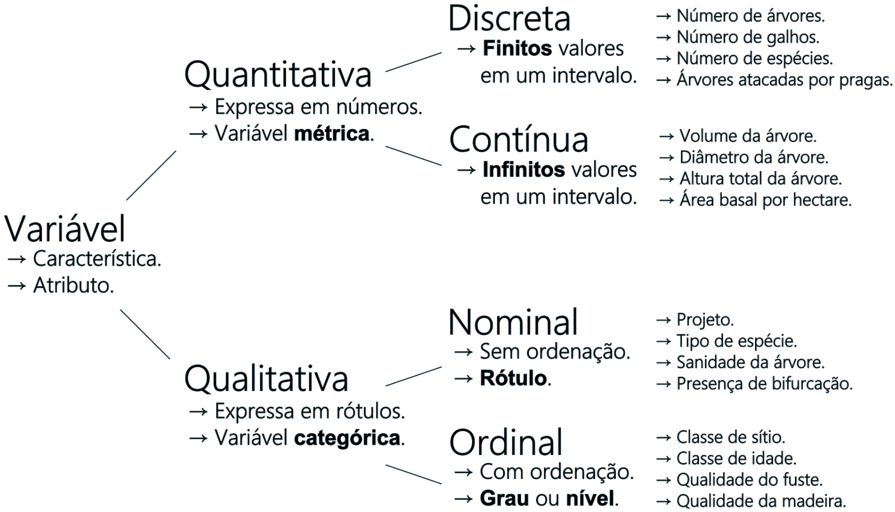

Figura 2. Principais tipos de variáveis.

Introdução

- O comportamento da resposta pode ser estudado por meio de modelos **probabilísticos**.
- Modelos disponíveis para variáveis:
  - Discreta $\rightarrow$ Poisson;
  - Contínua $\rightarrow$ Normal, Gamma;
  - Binária $\rightarrow$ Binomial, Hipergeométrica;
  - Limitada em intervalo $\rightarrow$ Beta.
- Escolha do modelo depende da **natureza** da resposta e análises **gráficas**.

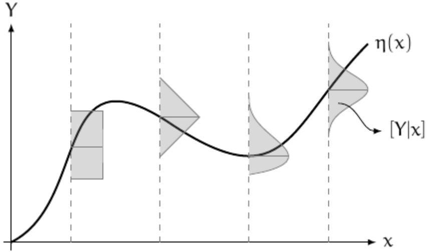

Figura 3. Modelo linear generalizado. Extraído de Walmes Zeviani no Tikz.

Família Exponencial

Família Exponencial

- O **componente aleatório** de um modelo linear generalizado consiste em uma **variável aleatória** Y, por meio de um conjunto de observações independentes $y_1, y_2, \ldots, y_n$, com distribuição pertencente à **família exponencial**.
- Assumimos que a **função de probabilidade** ou **densidade de probabilidades** de y possa ser expressa na forma:

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

em que a, b, c são funções adequadas.

Família Exponencial

- O parâmetro θ_i é chamado parâmetro natural (ou **parâmetro canônico**) e ϕ o parâmetro de dispersão da distribuição.
- Temos $a(\phi) = \phi$ ou $a_i(\phi) = \frac{\phi}{\omega_i}$, sendo ω_i um **peso** particular a cada observação.
- Para distribuições pertencentes à **família exponencial de dispersão**, expressões para $E(y_i)$ e $Var(y_i)$ são dadas por

$$E(y_i) = \mu_i = b'(\theta_i) = \frac{\partial b(\theta_i)}{\partial \theta_i},$$

$$Var(y_i) = a(\phi)b''(\theta_i) = a(\phi)\frac{\partial \mu_i}{\partial \theta_i}.$$

Família Exponencial

- A **variância** de y_i pode ser fatorada em dois componentes:
  - O primeiro ($a(\phi)$) é função de um parâmetro (ϕ) que está associado exclusivamente à **dispersão** de $y_i i$ (não à sua média).
  - O segundo, usualmente denotado por $V(\mu_i) = b''(\theta_i)$ e chamado **função de variância**, é função da média da distribuição, e exprime a relação média-variância de y.
- Cada **distribuição** pertencente à família exponencial de dispersão tem sua particular função de variância e vice-versa (unicidade).

Componentes

- Definido pela **especificação** de três componentes:
  - 1: **Componente aleatório**: conjunto de variáveis aleatórias independetes (y_i) com distribuição pertencente à família exponencial, e vetor de parâmetros $\boldsymbol{\theta}$, $\boldsymbol{\phi}$:

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

    em que a, b, c são funções adequadas.
- Membros da família exponencial: distribuições binomial, Poisson, normal, gama, normal inversa.

Componentes

- Definido pela **especificação** de três componentes:
  - 2: **Componente sistemático**: é o preditor linear do modelo, em que as variáveis são inseridas por meio de uma combinação linear de parâmetros:

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

Componentes

- Definido pela **especificação** de três componentes:
  - 3: **Função de ligação**: É uma função real, monótona e diferenciável, que conecta o componente sistemático ao aleatório:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip},$$

    em que $\mu_i = E(y_i|\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^{'})$.
  - Pelas propriedades da função de ligação, podemos escrever de forma equivalente por

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

Escolha dos Componentes

- **Componente aleatório**:
  - Definição de uma distribuição de probabilidades para a resposta.
  - Propor um modelo que tenha propriedades compatíveis à distribuição dos dados (espaço paramétrico, simetria, ...).
  - Não se tendo convicção sobre uma particular escolha, pode-se testar diferentes alternativas ou usar alguma abordagem que não exija essa especificação.
- **Preditor linear**:
  - Considerar quais variáveis explicativas devem ser utilizadas.
  - Considerar como essas variáveis serão incorporadas ao modelo.
- **Função de ligação**:
  - Finalidade de linearizar a relação entre os componentes aleatório e sistemático.
  - Deve produzir valores no espaço paramétrico (para μ_i) para qualquer valor produzido pelo preditor linear η_i.
  - Proporcionar interpretações práticas para os parâmetros de regressão.

Escolha dos Componentes

- Principais distribuições com as respectivas funções de ligação:

Ligação	$\eta = g(\mu)$	$\mu = g^{-1}(\eta)$	Ligação Canônina
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
Raíz quadrada	$\sqrt{\mu}$	η^{2}	Normal Inversa
Logito	$ln\frac{\mu}{1-\mu}$	$\frac{e^{\eta}}{1+e^{\eta}}$	Binomial
Probito	$\Phi^{-1}(\mu)$	$\Phi(\eta)$	Binomial
Log-log	$-ln[ln(\mu)]$	$exp[-exp(-\eta)]$	Binomial
Clog-log	$ln[-ln(1-\mu)]$	$1-exp[-exp(\eta)]$	Binomial

Poisson

Poisson

- $Y \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Então, $y = \{0, 1, 2, \ldots\}$.

$$f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}.$$

- Escrevendo na forma da família exponencial, temos

$$\begin{aligned} f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!} &= ln(e^{-\lambda}\lambda^{y}) - ln(y!) \\ &= exp\{yln(y) - \lambda - ln(-y!)\} \\ &= exp\{\phi[y\theta - b(\theta)] + c(y, \phi)\}, \end{aligned}$$

em que $\theta = ln(\lambda)$, $b(\theta) = exp(\theta)$, $\phi = 1$ e $c(y, \phi) = -ln(y!)$.

Binomial

Binomial

- Seja Y^* a proporção de sucessos em n ensaios de Bernoulli independentes. Então, temos $nY^* \sim \text{Binomial}(n, \mu)$, sendo $\mu \in (0,1)$. Temos que $y = \{0, 1/n, 2/n, \ldots, 1\}$ e

$$f(y) = \binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n - ny^*}.$$

Binomial

- Escrevendo na forma da família exponencial, temos

$$
\begin{aligned}
f(y) &= \binom{n}{ny^*} \mu^{ny^*}(1-\mu)^{n-ny^*} = \\
&= ln\binom{n}{ny^*} + ny^* ln\left(\frac{\mu}{1-\mu}\right) + n ln(1-\mu) \\
&= exp\left\{ ln\binom{n}{ny^*} + ny^* ln\left(\frac{\mu}{1-\mu}\right) + n ln(1-\mu) \right\} \\
&= exp\left\{ n\left[y^* ln\left(\frac{\mu}{1-\mu}\right) + ln(1-\mu)\right] + ln\binom{n}{ny^*} \right\} \\
&= exp\{\phi[y\theta - b(\theta)] + c(y,\phi)\},
\end{aligned}
$$

em que $\theta = ln\left(\frac{\mu}{1-\mu}\right)$, $b(\theta) = ln(1+e^{\theta})$, $\phi = n$, $c(y^*,\phi) = ln\binom{\phi}{\phi y^*}$. Se $n = 1$, então $Y^* \sim$ Bernoulli(μ).

Normal

Normal

- Seja $Y \sim \text{Normal}(\mu, \sigma^2)$, $y \in \{-\infty, +\infty\}$, $\mu \in \{-\infty, +\infty\}$, $\sigma^2 \in \{0, +\infty\}$ e

$$f(y) = f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}.$$

- Escrevendo na forma da família exponencial, temos

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

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

Principais Distribuições

Principais Distribuições

- Principais distribuições da família exponencial e respectivos parâmetros:

Distribuição	$b(\theta)$	θ	ϕ	$V(\mu)$
Normal	$\theta^2/2$	μ	σ^{-2}	1
Poisson	e^θ	$ln(\mu)$	1	μ
Binomial	$ln(1+e^\theta)$	$ln(\mu/(1-\mu))$	n	$\mu(1-\mu)$
Gama	$-ln(-\theta)$	$-1/\mu$	$1/(CV^2)$	μ^2
Normal Inversa	$-\sqrt{-2\theta}$	$-1/2\mu^2$	ϕ	μ^3

Considerações finais

Considerações finais

- **Família exponencial** é uma generalização de distribuições de probabilidades com características similares.
- É uma classe de modelos flexível e permite modelar variáveis de **diferentes naturezas**.

Figura 4. Fórmulas. Imagem de Nothing Ahead no Pexels