

CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

21 de novembro, 2017

Aula 15 - Regressão para dados contínuos assimétricos

- Algumas aplicações de modelos de regressão para dados positivos com distribuição contínua (e potencialmente assimétrica):
 - Análise do tempo de vida de pacientes segundo idade, sexo, tipo de tratamento, estágio da doença. . .
 - Resistência de moldes de alumínio em função da força aplicada, da composição do molde. . .
 - Valor do aluguel de imóveis em função do estado de conservação, localização, número de cômodos. . .
 - Renda segundo escolaridade, idade, sexo, . . .

Introdução

- Diversos modelos probabilísticos podem ser usados para a regressão de dados positivos com distribuição assimétrica contínua, dentre os quais:

- * Weibull;
- * Pareto;
- * Log-normal;
- * Gama;
- * Normal inversa...

- As distribuições Gama e Normal inversa pertencem à família exponencial e são contempladas pela teoria dos modelos lineares generalizados.

Distribuição Gama

- Uma variável aleatória y tem distribuição Gama de média μ e coeficiente de variação $1/\sqrt{\nu}$ se sua função densidade de probabilidade é dada por:

$$f(y; \mu, \phi) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu y}{\mu}\right)^{\nu} \exp\left(-\frac{\nu y}{\mu}\right) y^{-1}, \quad y > 0; \mu > 0; \nu > 0, \quad (1)$$

em que $\Gamma(\nu) = \int_0^{\infty} t^{\nu-1} e^{-t} dt$.

- Para a distribuição Gama, $Var(y) = \frac{\mu^2}{\nu}$, tal que $\phi = \nu^{-1}$ é o parâmetro de dispersão.
- Importante notar que, para a distribuição Gama, embora a variância dependa da média, **o coeficiente de variação é constante** (não varia conforme a média).

Distribuição Gama

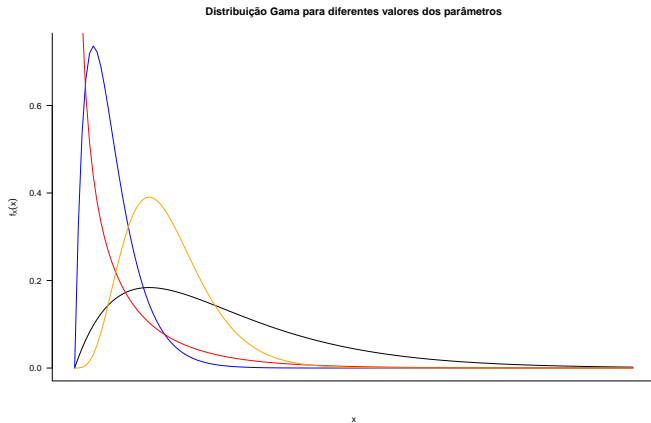


Figura 1: Distribuição Gama para diferentes valores dos parâmetros.

Distribuição Gama

- O modelo Gama pode ser usado na análise de dados contínuos assimétricos (ou simétricos tal que a relação variância-média seja quadrática).
- À medida que ϕ aumenta, a distribuição de y se aproxima de uma normal com média μ e variância $\mu^2\phi$.
- Aplicações da distribuição Gama, assim como para outras distribuições contínuas assimétricas, frequentemente apresentam **dados censurados**, que são observações *parcialmente disponíveis*.
- Como exemplo de dado censurado, podemos ter o tempo de vida de um paciente para o qual só sabemos que sobreviveu por mais de dez anos, ou a força de ruptura de um cabo de aço, que sabemos apenas que é maior que $100\text{kg}/\text{m}^2$.

- A regressão Gama é definida pela especificação da distribuição Gama para o componente aleatório e alguma função de ligação pertinente:

$$\begin{aligned}y_i | \mathbf{x}_i &\sim \text{Gama}(\mu_i, \phi) \\g(\mu_i) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.\end{aligned}\tag{2}$$

- Como alternativas de funções de ligação:
 - $g(\mu_i) = \mu_i^{-1}$ - inversa (ligação canônica);
 - $g(\mu_i) = \log(\mu_i)$ - logarítmica (efeitos multiplicativos);
 - $g(\mu_i) = \mu_i$ - identidade (efeitos aditivos).

- Para o modelo Gama, temos um parâmetro de dispersão (ϕ) a ser estimado (usualmente pelo estimador baseado na estatística X^2 - método dos momentos);
- Devido à estimação do parâmetro de dispersão, usamos a distribuição t_{n-p} ao invés da distribuição normal na construção de testes de hipóteses e intervalos de confiança, e a distribuição F ao invés da χ^2 na análise de deviances.