

**Universidade Federal do Paraná – Departamento de Estatística**  
**Disciplina CE225 – Modelos Lineares Generalizados**  
**Prof. Cesar Augusto Taconeli**  
**Lista de Exercícios**  
**06/09/2017**

**Exercício 1** – Mostre que a distribuição logarítmica, com função densidade de probabilidades:

$$f(y; \rho) = \rho^y \{-y \log(1 - \rho)\}, y = 1, 2, \dots; 0 < \rho < 1$$

pertence à família exponencial e represente-a na forma canônica. Identifique  $\theta$ ,  $\phi$ ,  $b(\theta)$  e  $c(y; \phi)$ . Usando as propriedades dessa família, determine a esperança de  $Y$ .

**Exercício 2** – Mostre que a distribuição geométrica, com função densidade de probabilidades:

$$f(y; \pi) = \pi(1 - \pi)^y, y = 0, 1, 2, \dots; 0 < \pi < 1$$

pertence à família exponencial e represente-a na forma canônica. Identifique  $\theta$ ,  $\phi$ ,  $b(\theta)$  e  $c(y; \phi)$ . Usando as propriedades dessa família, determine  $E(Y)$ ,  $Var(Y)$  e  $V(\mu)$ .

**Exercício 3** - Sejam  $Y_1, Y_2, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i | x_i \sim \text{Poisson}(\mu_i)$  e  $\mu_i = (\beta_0 + \beta_1 x_i)^2$ . Isso configura um modelo linear generalizado? Justifique sua resposta.

**Exercício 4** - Sejam  $Y_1, Y_2, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i | x_i \sim N(\mu_i, \sigma^2)$  e  $\mu_i = \beta_0 + \log(\beta_1 + \beta_2 x_i)$ , para  $i = 1, 2, \dots, n$ . Isso configura um modelo linear generalizado? Justifique sua resposta.

**Exercício 5** - Sejam  $Y_1, Y_2, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i | x_i \sim \text{Pareto}(\theta_i, \alpha_i)$  (segue a função densidade de probabilidade da Distribuição Pareto):

$$f_Y(y; \theta, \alpha) = \frac{\alpha \theta^\alpha}{y^{\alpha+1}}, y \geq \theta,$$

e  $\mu_i = E(Y_i | x_i) = \beta_0 + \beta_1 x_i$ , para  $i = 1, 2, \dots, n$ . Isso configura um modelo linear generalizado? Justifique sua resposta.

**Exercício 6** - Sejam  $Y_1, Y_2, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i | x_i \sim N(\mu_i, \sigma^2)$  e  $|\mu_i| = \beta_0 + \beta_1 x_i$ . Isso configura um modelo linear generalizado? Justifique sua resposta.

**Exercício 7** – Os itens a-d descrevem os objetivos de diferentes estudos. O objetivo do exercício é propor modelos lineares generalizados apropriados em cada situação. Para isso, você deve atender aos seguintes itens:

- I. Identificar a variável resposta, o conjunto de valores que ela pode assumir e a possível forma de sua distribuição;
  - II. Identificar as variáveis explicativas e suas respectivas escalas;
  - III. Postular uma distribuição de probabilidades apropriada, de acordo com as propriedades da variável resposta;
  - IV. Sugerir o preditor linear a ser considerado;
  - V. Propor uma função de ligação pertinente à distribuição de probabilidades escolhida;
  - VI. Escrever o MLG resultante em duas etapas, conforme visto em aula, especificando, num primeiro momento, a distribuição da resposta condicional às covariáveis e, posteriormente, a relação entre a média da resposta e o preditor linear.
- a) Investigar o efeito da idade, sexo, altura, consumo médio diário de calorias e gasto médio diário de energia no peso de uma pessoa;
  - b) Avaliar a mortalidade de cobaias que infectadas após exposição a uma bactéria quando cinco diferentes níveis de exposição são considerados e 50 ratos expostos a cada nível, sendo contado o número de ratos mortos após um tempo fixado de exposição;
  - c) A relação entre o número de visitas por semana ao supermercado por uma dona de casa e o número de pessoas na casa, a renda da família e a distância ao supermercado;
  - d) Avaliar a relação entre o tempo de sobrevivência de pacientes com determinado tipo de câncer segundo o sexo e tipo de tratamento (T1, T2 e T3). **Nota:** Desconfia-se que os tratamentos proporcionem resultados diferentes de acordo com o sexo.

**Exercício 8** - Na página da disciplina está disponível uma base de dados com 546 observações referentes aos preços de venda de 546 residências, no período de julho a setembro de 1987, numa certa localidade:

- `price`: preço de venda do imóvel;
- `lotsize`: dimensão da propriedade (em pés quadrados);
- `bedrooms`: número de dormitórios;
- `bathrms`: número de banheiros;
- `stories`: número de andares, excluindo o porão;
- `driveway`: 1, se a casa tem vaga de garagem e 0, caso contrário;
- `recroom`: 1, se a casa tem área de recreação e 0, caso contrário;
- `fullbase`: 1, se a casa tem porão e 0, caso contrário;
- `gaswh`: 1, se a casa tem aquecimento da água a gás e 0, caso contrário;
- `airco`: 1, se há sistema central de ar-condicionado e 0, caso contrário;
- `garagep1`: número de vagas de garagem;
- `prefarea`: 1, se localizada nas áreas preferenciais da cidade e 0, caso contrário;
- `age`: idade do imóvel (em anos);
- `dist`: distância do imóvel ao parque ou área arborizada mais próximos (em metros).
- `public`: 1, se o imóvel é de propriedade pública e 0, caso contrário.

O objetivo da análise é ajustar um modelo que explique o preço de venda em função das demais variáveis. Inicie ajustando um modelo linear e verifique suas limitações. Depois, tente o ajuste de um modelo linear generalizado com distribuição gama e função de ligação

logarítmica. Você pode tentar outras distribuições e funções de ligação também. Alguns pontos a serem considerados na sua análise:

- Apresentação do modelo ajustado na escala do preditor e da resposta;
- Interpretação dos efeitos das covariáveis, com base em seus sinais e respectivas significâncias;
- Análise de deviances e testes de hipóteses;
- Intervalos de confiança para os parâmetros do modelo;
- Predições e intervalos de confiança;
- Seleção de covariáveis;
- Diagnóstico do ajuste.

Utilize os recursos (metodológicos e computacionais) vistos em sala de aula, e outros que você julgar pertinentes.