

**Departamento de Estatística - ICEx - Universidade Federal de Minas Gerais**  
**Lista de Exercícios 2: EST079 - Modelos Lineares Generalizados**

---

Observação:

- As listas de exercícios de MLG não valem pontos.  
O estudante deve fazer os exercícios apenas como forma de aprendizagem e preparação para as avaliações.
  - Algumas questões podem ser resolvidas através do R, outras questões devem ser feitas sem auxílio computacional.  
O enunciado do exercício avisará sobre isso.
  - Se houver algum banco de dados disponibilizado para esta lista, saiba que ele foi salvo em um computador Linux.  
Desta forma, pode haver desconfiguração ao abrir o arquivo no “bloco de notas” do Windows. Use o comando `read.table` normalmente no R para ler os dados. A `data.frame` criada estará no formato correto.
- 

**[Questão 1]** Para um indivíduo  $i$ , considere a variável aleatória  $Y_i \sim \text{Bernoulli}(\theta)$ , tal que  $0 \leq \theta \leq 1$  é a probabilidade do evento sucesso  $Y_i = 1$ . Um estudo foi desenvolvido com este tipo de variável aleatória sendo medida independentemente para cada indivíduo participante. A seguinte amostra aleatória foi observada:

$y_1 = 1, y_2 = 0, y_3 = 0, y_4 = 1, y_5 = 0,$   
 $y_6 = 0, y_7 = 0, y_8 = 1, y_9 = 0, y_{10} = 0,$   
 $y_{11} = 0, y_{12} = 1, y_{13} = 0, y_{14} = 0, y_{15} = 1.$

Responda os itens a seguir:

- (a) Escreva a expressão da função de verossimilhança e a expressão da função de log-verossimilhança relacionadas a este problema. Sua resposta deverá contemplar os valores amostrais observados para os 15 indivíduos. Atenção! é obrigatório substituir o resultado amostral indicado acima na expressão das duas funções solicitadas.  
Não use o R para responder.
- (b) Use o R para fazer o gráfico da função de verossimilhança escrita na letra (a). Você deve plotar “valores de  $\theta$  versus função de verossimilhança” ( $\theta$  deve estar no eixo horizontal do gráfico). Considere os valores de  $\theta$  definidos pelo comando `theta = seq(0, 1, 0.01)`. Qual valor de  $\theta$  está associado ao ponto de máximo da curva obtida no gráfico?
- (c) Use o R para fazer o gráfico da função de log-verossimilhança escrita na letra (a). Você deve plotar “valores de  $\theta$  versus função de log-verossimilhança” ( $\theta$  deve estar no eixo horizontal do gráfico). Considere os valores de  $\theta$  definidos pelo comando `theta = seq(0, 1, 0.01)`. Qual valor de  $\theta$  está associado ao ponto de máximo da curva obtida no gráfico?
- (d) Aplique a função `optim` do R (conforme ensinado nas aulas) para obter a estimativa de máxima verossimilhança de  $\theta$ . Você deverá maximizar a função de log-verossimilhança e considerar a amostra apresentada no enunciado desta questão (use o chute inicial  $\theta^{(0)} = 0.5$ ). Dentro da função `optim`, selecione o método L-BFGS-B para realizar a otimização numérica (especifique limite inferior = 0.0001 e superior = 0.9999 para o L-BFGS-B). Mostre seu script e indique claramente a estimativa final de  $\theta$ .

**[Questão 2]** Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória tal que  $Y_i \sim \text{Poisson}(\theta_i)$  com  $\theta_i > 0$ . Desejamos analisar o impacto do regressor  $X_{1i}$  sobre a resposta  $Y_i$ . Considere o preditor linear  $\eta_i = \beta_0 + \beta_1 X_{1i}$ . Denotando  $\beta = (\beta_0, \beta_1)^\top$  e  $X_i = (1, X_{1i})^\top$ , podemos também escrever  $\eta_i = X_i^\top \beta$ . No caso Poisson, lembre que a função de ligação é estabelecida por  $\theta_i = e^{\eta_i}$ . A seguinte amostra deverá ser utilizada para resolver esta questão:

| $i$      | 1     | 2    | 3     | 4     | 5    | 6    | 7   | 8    | 9     | 10   | 11    | 12    | 13    | 14   | 15 | 16  | 17    | 18   | 19    | 20   | 21   | 22   | 23    | 24   | 25    |
|----------|-------|------|-------|-------|------|------|-----|------|-------|------|-------|-------|-------|------|----|-----|-------|------|-------|------|------|------|-------|------|-------|
| $Y_i$    | 7     | 2    | 24    | 33    | 2    | 5    | 2   | 5    | 8     | 8    | 16    | 20    | 11    | 4    | 37 | 1   | 8     | 8    | 5     | 7    | 5    | 4    | 4     | 2    | 26    |
| $X_{1i}$ | -0.27 | 0.57 | -0.74 | -0.94 | 0.64 | 0.86 | 0.5 | 0.12 | -0.36 | 0.12 | -0.55 | -0.81 | -0.21 | 0.53 | -1 | 0.4 | -0.07 | 0.15 | -0.13 | 0.22 | 0.96 | 0.45 | -0.06 | 0.97 | -0.71 |

Nas aulas de MLG foi explicado como fica estruturado o algoritmo IWLS para esta situação. Revise as aulas e responda o itens a seguir:

- Estabeleça o chute inicial  $\hat{\beta}^{(r=0)} = (1, 1)^\top$  a ser utilizado no IWLS. Usando este chute e o resultado amostral fornecido no enunciado, construa no R (exibir o script) a matriz  $W^{(0)}$  correspondente a este problema. Mostre a submatriz  $5 \times 5$  formada pelas linhas 1 a 5 e colunas 1 a 5 da matriz  $W^{(0)}$ .
- Ainda levando em conta o chute inicial  $\hat{\beta}^{(r=0)} = (1, 1)^\top$ , calcule o vetor  $z^{(0)}$  definido no IWLS deste problema relacionado à Poisson. Mostre os valores do seu vetor  $z^{(0)}$ .
- Utilizando  $W^{(0)}$  e  $z^{(0)}$  obtidos em (a) e (b), respectivamente, calcule a estimativa  $\beta^{(1)} = (X^\top W^{(0)} X)^{-1} X^\top W^{(0)} z^{(0)}$  referente à primeira iteração do IWLS. Mostre o script R e o resultado final de  $\beta^{(1)}$ .
- Use os dados amostrais fornecidos no enunciado e aplique a função `glm` do R para estimar  $\beta_0$  e  $\beta_1$ . Use o mesmo chute inicial sugerido em (a). Mostre a saída do comando `summary` e avalie a significância dos coeficientes. Atenção! sua análise sobre a significância deve indicar claramente a estimativa do coeficiente e o valor-p sob avaliação. Note também que a questão NÃO está pedindo para você realizar vários ajustes para encontrar o melhor modelo. O objetivo aqui é apenas ver se você sabe interpretar a saída computacional solicitada.
- A saída computacional investigada em (d) mostra a deviance do modelo ajustado. Calcule esta deviance através do R. Depois disso, avalie se essa deviance é pequena, moderada ou grande. Atenção! você deve mostrar o passo a passo da conta (use a expressão da deviance para o caso Poisson explicada nas aulas). Não aplique uma função pronta do R que calcula a deviance automaticamente.
- Calcule (use o R) a estatística de Pearson generalizada para o ajuste em (d). Em seguida compare o resultado com algum limiar adequado que deve ser obtido a partir da distribuição qui-quadrado. Atenção! mostre o passo-a-passo da conta e não use função pronta do R que calcula a estatística.

**[Questão 3]** Um experimento de laboratório foi desenvolvido para estudar um tipo de larva que ataca lavouras de café. O estudo considerou 52 recipientes, os quais receberam (cada um) 10 larvas. As larvas em cada recipiente foram submetidas (ao mesmo tempo) a uma dose de um produto químico A (variável  $X_1$ ) e uma dose de um produto químico B (variável  $X_2$ ). Os dados estão disponíveis no arquivo `dados_Q3_L2_MLG.txt` disponibilizado para esta lista de exercícios. A variável resposta do estudo é o número de larvas que morreram no recipiente após 1 hora da aplicação dos dois produtos químicos. Carregue os dados no R com a função `read.table`. A *data frame* obtida terá a variável resposta na coluna 1,  $X_1$  na coluna 2 e  $X_2$  na coluna 3.

Responda os itens a seguir (você poderá usar o R para responder).

- (a) Aplique a função `glm` do R para estimar os coeficientes  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  do modelo MLG apropriado a estes dados (use obrigatoriamente o link canônico). Mostre a saída do comando `summary` e avalie a significância dos coeficientes. Atenção! sua análise sobre a significância deve indicar claramente o valor estimado do coeficiente e o valor-p sob avaliação. Note também que a questão NÃO está pedindo para você realizar vários ajustes para encontrar o melhor modelo. O objetivo aqui é apenas ver se você sabe qual é o MLG (Normal, Binomial, Bernoulli, Poisson, Gama, etc.) apropriado e como interpretar o resultado.
- (b) Interprete o impacto da covariável  $X_1$  quando seu valor é aumentado em 0.1 (zero ponto um). Atenção! explique claramente qual elemento do modelo sofrerá influência direta deste aumento em  $X_1$ .
- (c) Baseando-se no ajuste obtido em (a), construa os seguintes gráficos: “resíduos de Pearson vs. Valores ajustados” e “resíduos componente do desvio vs. valores ajustados”. Interprete os dois gráficos. Atenção! mostre os comandos que usou para obter os resíduos e os valores ajustados.
- (d) Levando em consideração o resultado do ajuste em (a), obtenha as estimativas das probabilidades de morte da larva em cada recipiente do experimento (mostre o script R). Faça o gráfico: “ $X_1$  vs. probabilidade de morte” e “ $X_2$  vs. probabilidade de morte”. Comente os resultados dos dois gráficos.