

Departamento de Estatística - ICEx - Universidade Federal de Minas Gerais
Lista de Exercícios 1: 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.

[Questão 1] Estamos estudando uma técnica de purificação de água baseada em membranas de filtração. Existem dois tipos de membranas no mercado (tipo “A” e tipo “B”). Desejamos examinar a relação entre: tipo de membrana, pressão exercida na membrana e o fluxo através da membrana. A variável resposta é o fluxo. A tabela abaixo mostra 20 resultados do experimento:

Experimento	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y (fluxo)	2.38	0.62	2.86	4.53	3.74	2.03	5.53	5.72	6.00	3.53	4.25	0.03	1.15	3.04	1.35	4.74	4.80	5.90	2.79	2.51
X ₁ (pressão)	4.65	3.87	5.92	5.67	4.53	4.24	4.73	2.19	4.91	4.83	7.08	4.92	3.94	3.68	4.20	4.93	5.19	6.98	4.42	4.22
X ₂ (tipo)	A	B	B	B	A	B	A	A	A	A	B	B	B	A	B	A	A	A	B	A

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

- Construa a matriz de regressores X . Mostre o resultado!
- Aplique o teste Shapiro-Wilk para a variável resposta fluxo.
Escreva as hipóteses, nível de significância e a conclusão do teste.
- Dizem que a matriz $X^T X$ é: simétrica, tem na diagonal principal as somas dos quadrados dos elementos das colunas de X , fora da diagonal principal aparecem a soma dos produtos cruzados das colunas de X e ela é inversível. Use o R para mostrar que isso tudo é válido para X sendo a matriz obtida em (a).
- Obtenha as estimativas de mínimos quadrados dos coeficientes do modelo de regressão desta aplicação. Use a fórmula do estimador de mínimos quadrados explicada em aula e, depois disso, aplique a função `lm` do R para verificar se o resultado é o mesmo.
- Estime o parâmetro σ^2 presente no modelo de regressão (use um estimador não viciado).
- Use a função `lm` do R para ajustar o modelo de regressão (exiba o resultado obtido via `summary`). Quais coeficientes são significativos? Escreva claramente os valores-p sendo analisados.
- Interprete o valor do(s) coeficiente(s) significativo(s) em (f).
- Interprete o valor do R^2_{ajustado} .
- Use o resultado em (a) para calcular a matriz H de projeção (matriz chapéu) deste problema. Use o comando `round` do *software* R para exibir a sua matriz H com apenas 2 casas decimais. Verifique se H é simétrica e idempotente para as potências $p = 2$ e $p = 3$.

- (j) Faça os gráficos “resíduos contra valores ajustados \hat{Y} ” e “resíduos contra a covariável X_1 ”. Interprete os dois gráficos.
- (k) Explique o que são pontos influentes e pontos alavanca em um ajuste de regressão. Avalie se há pontos influentes ou alavancas no banco de dados relacionado a este exercício. Apenas uma ou as duas covariáveis são responsáveis pelas alavancas?
- (l) Calcule o resíduo studentizado internamente para cada elemento amostral. Mostre cada passo da conta e os resíduos obtidos.

[Questão 2] Em cada item a seguir, escreva a distribuição mencionada usando a “Notação 1” da família exponencial apresentada nos slides das aulas. Identifique claramente os termos $a(Y)$, $b(\theta)$, $c(\theta)$ e $d(Y)$. Indique se foi possível obter o formato canônico da família exponencial. Mostre suas contas! Não use o R para responder esta questão. Não olhe os slides das aulas para obter a solução.

- (a) $Y \sim \text{Poisson}(\theta)$, sendo $\theta > 0$ a média da distribuição.
- (b) $Y \sim \text{Binomial}(m, \theta)$, sendo $0 < \theta < 1$ e m é o número conhecido de ensaios Bernoulli independentes.
- (c) $Y \sim \text{Binomial-Negativa}(\kappa, \theta)$, com $0 < \theta < 1$. A variável Y representa o número de sucessos até que $\kappa > 0$ falhas ocorram (suponha κ conhecido). A função massa de probabilidade é dada por:

$$f(y) = \binom{y + \kappa - 1}{y} \theta^y (1 - \theta)^\kappa \quad \text{para } y = 0, 1, 2, \dots$$

- (d) $Y \sim \text{Rayleigh}(\theta)$, sendo $\theta > 0$. A densidade é dada por:

$$f(y) = \frac{y}{\theta^2} \exp \left\{ -\frac{y^2}{2\theta^2} \right\} \quad \text{para } y > 0.$$

- (e) $Y \sim \text{Exponencial}(\theta)$, com média $\theta > 0$. A densidade é dada por:

$$f(y) = \frac{1}{\theta} \exp \left\{ -\frac{1}{\theta} y \right\} \quad \text{para } y > 0.$$

- (f) $Y \sim \text{Gama}(\theta, \alpha)$, sendo $\theta > 0$ a média e $\alpha > 0$ o parâmetro de forma (conhecido). A densidade é dada por:

$$f(y) = \frac{(\alpha/\theta)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp \left\{ -\frac{\alpha}{\theta} y \right\} \quad \text{para } y > 0.$$

- (g) $Y \sim \text{Normal}(\theta, \sigma^2)$, com variância $\sigma^2 > 0$ conhecida e média $\theta \in \mathbb{R}$. A densidade é dada por:

$$f(y) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - \theta)^2 \right\} \quad \text{para } -\infty < y < \infty.$$

- (h) $Y \sim \text{Normal-Inversa}(\theta, \tau)$, com média $\theta > 0$ e parâmetro de forma conhecido $\tau > 0$. A densidade é:

$$f(y) = \left(\frac{\tau}{2\pi y^3} \right)^{1/2} \exp \left\{ -\frac{\tau(y - \theta)^2}{2\theta^2 y} \right\} \quad \text{para } y > 0.$$

[Questão 3] Para cada item da Questão 2, calcule $E[a(Y)]$ e $Var[a(Y)]$ utilizando as formulações dadas nas Propriedades 1 e 2 da família exponencial (essas duas propriedades foram explicadas nas aulas).

[Questão 4] Crie um script do R para reproduzir a figura comparativa (comportamento de três funções de ligação do caso Bernoulli) mostrada no slide 18 do arquivo `MLG_A03.pdf`, o qual foi usado nas aulas e está disponível no Moodle. Mostre o script e o gráfico produzido.