

**Disciplina:** Modelos Lineares Generalizados

**Curso:** Graduação em Estatística

**Código:** EST0011

**Semestre:** 2023.2

**Prof.** Frederico Machado Almeida

## LISTA DE EXERCÍCIOS #2

### Observações:

- Questões para entregar: TODAS
- Demais questões são apenas para estudar.
- Prazo de entrega: 04/10/2023
- A lista pode ser feita em duplas.

**Q1.** Considere os dados sobre a qualidade do vinho tinto, apresentados no ficheiro [Q01-data.txt](#). Ajuste o modelo de regressão linear múltipla, e faça uma análise completa desses dados. Que conclusões você tira dessa análise? (use 5% de significância durante as análises). As variáveis de interesse são:

$y$ : classificação de qualidade (20 no máximo)

$x_1$ : variedade de vinho (0 — Cabernet Sauvignon, 1 — Shiraz)

$x_2$ : nível de pH

$x_3$ : SO<sub>2</sub> total (ppm)

$x_4$ : densidade de cor

$x_5$ : cor de vinho

$x_6$ : cor de pigmento polimérico

$x_7$ : cor de antocianina

$x_8$ : antocianinas totais (g/L)

$x_9$ : grau de ionização das antocianinas (porcentagem)

$x_{10}$ : antocianinas ionizadas (porcentagem).

Com base no conjunto de dados, é possível observar que algumas covariáveis estão fortemente correlacionadas. Assim, pede-se:

- Proponha algum método para resolver o problema da multicolinearidade no conjunto de dados.
- Usando algum método de seleção de variáveis, obtenha o modelo final para o conjunto de dados.
- Apresente a tabela da Análise de Variância para testar a significância global dos coeficientes do modelo final. Apresente as hipóteses de teste, e conclua.
- Com base no modelo obtido no item anterior, faça uma análise de resíduos e conclua.

**Q02.** Uma equipe de pesquisadores de saúde mental deseja comparar três métodos de tratamento da depressão grave (A, B e C=referência). Eles também gostariam de estudar a relação entre idade e eficácia do tratamento, bem como a interação (se houver) entre idade e tratamento. Cada elemento da amostra aleatória simples de 36 pacientes, foi selecionado aleatoriamente para receber o tratamento A, B ou C. Os dados obtidos podem ser encontrados no ficheiro [Q02-data.txt](#). A variável dependente  $y$  é a eficácia do tratamento; as variáveis independentes são: a idade do paciente no aniversário mais próximo, e o tipo de tratamento administrado (use 1% de significância durante as análises).

- (a) Ajuste o modelo de regressão linear e interprete os resultados obtidos.
- (b) Obtenha a tabela ANOVA para o modelo obtido no item (a) e interprete os resultados.
- (c) Considere a possibilidade de incluir a interação entre as variáveis independentes, i.e., assumo que o modelo a ser ajustado tem a seguinte formulação:  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{Ai} + \beta_3 x_{Bi} + \beta_4 x_{1i} x_{Ai} + \beta_5 x_{1i} x_{Bi} + \epsilon_i$ , com a suposição de que  $\epsilon_i \sim N(0, \sigma^2)$ . Com base no modelo anterior,
  - (i) Liste todos os possíveis submodelos que podem ser obtidos usando o modelo apresentado anteriormente.
  - (ii) Interprete os coeficientes de regressão associados aos fatores de interação.
  - (iii) Apresente a tabela anova para testar as seguintes hipóteses,  $H_0 : \beta_1 = \beta_4 = \beta_5 = 0$  contra  $H_1 : \exists \beta_j \neq 0$ , com  $j = 1, 4, 5$ .
  - (iv) Faça uma análise completa dos resíduos do modelo.