

Classwork 09

Instruções:

1. Este Classwork (CW) poderá ser resolvido em duplas;
2. O CW precisará conter o(s) **nome(s) completo(s)** e a(s) **matrícula(s)** do(s) discente(s);
3. **Serão considerados apenas** os Classworks resolvidos por discentes presentes;
4. O CW deverá ser enviado via **Google Classroom até as 07h51 do dia 22/06/2026** em **um único arquivo** e em **PDF** (*Portable Document Format*);
5. **Não serão considerados** Classworks com péssima qualidade visual ou em desacordo com as instruções anteriores.

Exercício. Considere o conjunto de dados utilizado por Fawcett and Higginson (2012), que investiga a relação entre o número de citações recebidas por artigos de biologia evolutiva e a quantidade de equações presentes nesses artigos. Ajuste um modelo de regressão linear para explicar número total de citações recebidas, excluindo autocitação (**OtherCites**), utilizando como potenciais variáveis preditoras:

- Densidade de equações por página (**TotPerPage**);
- Número total de páginas (**Pages**);
- Periódico onde o artigo foi publicado (**Journal**).

Como modelo inicial, considere a seguinte relação:

$$\text{OtherCites}_i = \beta_1 + \beta_2 \text{TotPerPage}_i + \beta_3 \text{Pages}_i + \beta_4 \text{Journal}_i + \beta_5 \text{TotPerPage}_i \times \text{Pages}_i + \varepsilon_i,$$

para $i = 1, \dots, 649$.

Tarefas.

1. Realize uma análise exploratória dos dados. Apresente gráficos e medidas descritivas que permitam compreender o comportamento das variáveis. Discuta possíveis padrões, relações entre as variáveis e a presença de valores atípicos.
2. Ajuste o modelo completo e avalie a presença de multicolinearidade entre as variáveis explicativas por meio de medidas e diagnósticos apropriados (por exemplo, matriz de correlações e fatores de inflação da variância). Discuta os resultados obtidos e, caso necessário, proponha estratégias para lidar com esse problema.
3. Avalie a significância estatística dos coeficientes do modelo completo (considere, por exemplo, um nível de significância de 5%). Caso existam variáveis não significativas, proponha um modelo reduzido, justificando a exclusão dessas variáveis.
4. Verifique a adequação do modelo ajustado por meio da análise de resíduos e de diagnósticos apropriados. Discuta a presença de possíveis problemas, como observações influentes ou padrões sistemáticos nos resíduos. Quando apropriado, proponha uma transformação da variável resposta que possa melhorar o ajuste do modelo.
5. Interprete os resultados do modelo final, incluindo os coeficientes estimados. Discuta o efeito de cada variável explicativa sobre a variável resposta, mantendo as demais variáveis constantes.

Referências

Fawcett, T. W. and Higginson, A. D. (2012). Heavy use of equations impedes communication among biologists. *Proceedings of the National Academy of Sciences*, 109(29):11735–11739.