



Projeto computacional

Back

Question 1

Not yet answered

Marked out of 2.0

🚩 Flag question

O ficheiro [winequality-white-q5.csv](#) contém um conjunto de dados baseado nos dados descritos em:

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. 2009. *Modeling wine preferences by data mining from physicochemical properties*. In *Decision Support Systems*, Elsevier, 47(4):547-553. ISSN: 0167-9236.

Neste estudo foram analisadas 4898 amostras de vinho verde branco, com o objetivo de estudar como é que as preferências dos peritos em provas de vinhos se relacionam com as caraterísticas físico-químicas dos vinhos. No ficheiro, as variáveis 1 a 11 referem-se às caraterísticas físico-químicas dos vinhos. A variável 12 representa a avaliação feita por especialistas em vinhos: qualidade do vinho entre 1 (má) e 5 (excelente).

Com recurso ao pacote [ggplot2](#), produza um gráfico que represente, através de diagramas de caixa (*box plots*) paralelos, como é que a raiz quadrada da variável **citric.acid** se relaciona com a variável **quality**. No gráfico, deverá realçar as possíveis observações discordantes (*outliers*) e, de alguma forma, tentar contrariar a sua sobreposição.

Por simplicidade, mantenha todo o texto da figura em Inglês.

Submeta um ficheiro em formato PDF com uma única página A4, que inclua:

- O código em R, incluindo os comandos para leitura dos dados do ficheiro.
- O gráfico produzido.

Maximum file size: 200 MB, maximum number of files: 1

