

STAT 435: Quiz 1

Due 09/26 11:59pm

For this quiz discussions are permitted however the code should be your own. All work should be done in Rmarkdown. Upload both .rmd and .pdf files to canvas.

Q1 Consider the data provided with the quiz and do the following.

- a. Import the data set in R.
- b. The data contains the variables y , $X1$, $X2$ and $X3$. The objective of this problem is to predict the response y based on $X1$, $X2$ and $X3$ and to determine which variables are significantly associated with the response. Perform a multiple regression to answer this question. Provide a prediction at $X1 = 0.25$, $X2 = 0.5$, $X3 = 0$ and compute the corresponding confidence and prediction intervals.
- c. Analyse the residuals to detect potential problems with your analysis in part (b).
- d. Propose solutions to the problems that you detect in part (c) and implement them on the data set. [*Hint:* studentized residual can be computed using the function `studres()`, cooks distance using `cooks.distance()` and the variance inflation factor using `vif()`. The function `vif()` is part of the R package `car` which you may need to install and unpack].
- e. Rerun your analysis of part (a) on the data that you obtain from part (d)
- f. Provide a prediction at $X1 = 0.25$, $X2 = 0.5$, $X3 = 0$ and compute the corresponding confidence and prediction intervals. Compare with the prediction in part (a) and comment on which you think is more believable.