

STAT 53533: R Assignments 02
Upload your answers, **as a single file**, in Blackboard by
11:59 PM on Wednesday, November 20, 2024

A. Separate your answer in two parts (within the same file) as follows:

1. Part 1 must contain complete answers for all questions **without** any R code. It must have all numbers/values/diagrams as asked in the questions and should be accompanied by necessary explanations.

2. Part 2 contains only the R code and nothing else. **Any output/numbers/explanations written inside the R Code will NOT be graded.**

3. **You must type all your answers and explanations for submission, as a professional report.** Scanned copy of any **handwritten answers/explanations will not be graded.**

Consider the attached dataset, extracted from pls package in R, on octane number and spectral features (indicated by column names ending with "nm") for a set of gasoline samples. Our goal is to use spectral features as covariates to predict octane number of a gasoline sample. Now, answer the following questions. [1+3+4+2 = 10 Points]

First, **exclude the 4th observation from your dataset and treat it as your test data point. That means, to answer Parts (a), (b) and (c) below, you will use all observations except the 4th observation as your training dataset.**

(a) Justify, using the values of n (for training dataset) and p , why usual OLS regression cannot be used.

(b) Run PCR allowing 10% estimated loss of variability. Report the following output:

- (i) number of principal components we need and corresponding ELOV value.
- (ii) estimated regression coefficients for PCR predictors.
- (iii) value of R^2 .

(c) Now, run a PLS1 regression using same number of predictors as used for PCR in Part (b). Report the following output:

- (i) Report, for each step, the maximum covariance attained at that step.
- (ii) Report, for each step, the value of R^2 after that step.
- (iii) After completion of PLS1 algorithm, write the equation of fitted line using numeric values of estimated coefficients.
- (iv) The estimated regression coefficient of in the simple linear regression of $x_{(0)45}$ on the first PLS predictor is = _____. The estimated coefficient of $x_{(1)260}$ in the expression of second PLS predictor is = _____.

(d) Now, use the test data point (the observation you removed in the beginning) and fill in the following table.

Its true Octane Number	Its predicted Octane Number using PCR	Its predicted Octane Number using PLS1 regression