

Assignment #5 — Due: Wednesday, December 16, 2020, by 11:50 p.m.

Use R to answer the questions. Your answers should include some selected, relevant outputs or graphs but not necessarily all of them. Download the data set, `Gasoline.txt` from the blackboard. Definition of the variables is given below.

Variable	Definition
Y	Miles/gallon
X_1	Displacement (cubic inches)
X_2	Horsepower (feet/pound)
X_3	Torque (feet/pound)
X_4	Compression ratio
X_5	Rear axle ratio
X_6	Carburetor (barrels)
X_7	Number of transmission speeds
X_8	Overall length (inches)
X_9	Width (inches)
X_{10}	Weight (pounds)
X_{11}	Type of transmission (1 = automatic; 0 = manual)

- [10 points] Compute the correlation matrix of the predictor variables X_1, \dots, X_{11} and the corresponding pairwise scatter plots. Identify any evidence of multicollinearity.
- [10 points] Compute the eigenvalues. How many principle components are required if we want to retain at least 85% of the information in the covariates (X)?
- [10 points] Compute the variance inflation factor (VIF) for each of the predictor variables. Which predictors are affected by the presence of multicollinearity?
- [10 points] Calculate two principle components (Z_1 , and Z_2) based on the correlation matrix of X . Then, show the principle components for the first three observations. (These are called PC scores.) **Do not use `> pcr()` function.**
- [10 points] Select a model using a forward selection method based on BIC. Write the fitted regression equation. (Use `> step()` function to answer this question. Specifically, your coding must include X_6 and exclude X_{11} during the selection procedure.)
- [10 points] Use best subset selection method to choose the 'best' model using an adjusted R^2 . Use `> regsubsets()` function to answer this question. Specifically, your coding must include X_6 and exclude X_{11} during the selection procedure. Justify your answer by showing a plot. (Use a default value for `nbest` option.) Which variables are contained in your best model?
- [10 points] Covariance matrix \mathbf{M} is given as below.

$$\mathbf{M} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Derive the two principle components (Z_1 and Z_2) by hand calculation. (Answers using R will get zero point. Do not type but write using hand then scan or take a photo your hand-written answer. Then combine it with other answers to make a one pdf file.)