

STA 4320 – Homework 3

Prof. He Jiang

Name: _____

Student Number: _____

For the current assignment, Your report should be no longer than the following amount of pages:
5

You do not have to submit this instruction page.

Please compile your coding into a single PDF file.

Please submit your compiled (PDF file) report to the corresponding assignment on Gradescope.

Please label your solution. If there is only 1 question, please label as well.

1. The **Auto** dataset in the **ISLR2**¹ package contains Gas mileage, horsepower, and other information for $n = 392$ vehicles.

We would like to use multiple linear regression to investigate the relationship between the response, Y , mpg, and the independent variables, X_1 , number of cylinders, X_2 , displacement of cylinders, X_3 , horsepower, and X_4 , weight.

Grading Method

The grading of this assignment will be based on: Completion.

Submissions that are empty or did not show sufficient efforts will receive 0. Submissions that showed significant efforts will receive full amount of points.

Tasks for this Assignment

- (a) (1 point) Load the **Auto** dataset, remove all the columns not mentioned above, and print the first 6 rows (with only the 5 columns mentioned above).
- (b) (1 point) Conduct linear regression using the **lm** command, and print the summary of this regression.
- (c) (1 point) A new car has 6 cylinders, with 275 displacement, 125 horsepower, and weights 3500. Give an appropriate interval prediction for the **mpg** of this single car (at the 95% level). Please round all results to 4 decimal points.
- (d) (1 point) Is there a relationship between the predictors and the response? I.e. should the model only consist of the intercept as the independent variable, or should the model consist of all of **horsepower**, **weight**, **cylinders**, **displacement** (and intercept) as the independent variables? Please use $\alpha = 0.05$ as the significance level.
- (e) (2 points) With **horsepower**, **weight** and **cylinders** (and intercept) already in the model, should we additionally include **displacement**? Conduct an appropriate test at the $\alpha = 0.01$ significance level.
- (f) (2 points) Print the correlation matrix. Does the correlation matrix provide some insight to the result in the previous part? Please explain.
- (g) (2 points) With **horsepower** and **weight** (and intercept) already in the model, should we additionally include **cylinders** and **displacement**? Conduct an appropriate test at the $\alpha = 0.01$ significance level.

¹This is the R package corresponding to our textbook.