**Instructions:**

Load the dataset named Auto (in the ISLR package) into R.

Background Information

The dataset Auto contains the fuel efficiency and other characteristics of 392 cars that were produced between 1970 and 1982. The variables have the following definitions:

mpg: The fuel efficiency of the car, measured in miles per gallon,

cylinders: The number of cylinders in the engine of the car,

displacement: The displacement of the car’s engine, measured in cubic inches,

horsepower: The power of the car’s engine, measured in horsepower,

weight: The weight of the car, measured in pounds,

acceleration: The quarter mile drag time, measured in seconds,

year: The last two digits of the car’s model year,

origin: The origin region of the car. 1 stands for U.S., 2 stands for Europe, 3 stands for Japan,

name: The make and model of the car.

Question 1

Use the summary function to find the average value of the following variables: mpg, cylinders, horsepower, weight, acceleration, year

23.40 mpg, 4 cylinders, 104.5 horsepower, 2804 weight, 15.50 acceleration, and 73 year

22.75 mpg, 5.472 cylinders, 194.4 horsepower, 2978 weight, 15.50 acceleration, and 76 year

23.45 mpg, 5.472 cylinders, 200 horsepower, 3000 weight, 15.54 acceleration, and 75.98 year

23.45 mpg, 5.472 cylinders, 104.5 horsepower, 2978 weight, 15.54 acceleration, and 75.98 year

22.75 mpg, 4.000 cylinders, 93.5 horsepower, 2804 weight, 15.50 acceleration, and 76 year

Question 2Use the correlation function to find the two numeric variables (mpg, cylinders, horsepower, weight, acceleration, year) with the maximum negative correlation.

mpg, cylinders

horsepower, cylinders

weight, year

weight, mpg

mpg, horsepower

Question 3 - Part A

Please estimate a linear regression model (using the lm function) with mpg as the dependent variable and horsepower as the independent variables. What is this model’s R-squared value?

0.8087

0.8062

0.6059

0.9045

0.6049

Question 3 - Part B

Please estimate a linear regression model (using the lm function) with mpg as the dependent variable and horsepower as the independent variables. What is this model’s adjusted R-squared value?

0.8087

0.8062

0.6059

0.9045

0.6049

Question 3 - Part C

Based on the model in Question 3 - Part A, what are the values of Total Sum of Squares (SST), Sum of Squared Errors (SSE), and Sum of Squared Regression (SSR)?

23819, 9385.9, 14433.1

14433.1, 5047.2, 9385.9

14433.1, 9385.9, 5047.2

23819, 14433.1, 9385.9

19480.3, 14433.1, 5047.2

Question 3 - Part D

Based on the model in Question 3 - Part A, use the predict.lm function in R to answer this question: What is the predicted value of mpg for a car that has a horsepower of 101?

26.47996

23.99354

23.85633

24.17526

25.47669

Question 3 - Part E

Based on the model in Question 3 - Part A, use the predict.lm function in R to answer this question: What is the 95% prediction interval of mpg for a car that has a horsepower of 101?

14.33611,33.65098

19.6714,33.28852

12.94096,35.04612

11.2787,36.70838

15.89468,32.0924

Question 3 - Part F

For the linear model in Question 3 - Part A, what is the estimated horsepower coefficient?

39.935861

0.717499

-0.157845

0.006446

-24.49

Question 3 - Part G

For the linear model in Question 3 - Part A, is the estimated horsepower coefficient statistically significant at 5%?

Yes

No

Depends on VIF

Need more information

Depends on other variables

Question 3 - Part H

For the linear model in Question 3 - Part A, what is the F-statistic of the overall model?

567.1

599.7

599.1

567.2

390

Question 4 - Part A

Please estimate a linear regression model (using the lm function) with mpg as the dependent variable and horsepower, cylinders, weight, acceleration, and year as the independent variables. What is this model’s R-squared value?

0.8087

0.8062

0.6059

0.9045

0.6049

Question 4 - Part B

Please estimate a linear regression model (using the lm function) with mpg as the dependent variable and horsepower, cylinders, weight, acceleration, and year as the independent variables. What is this model’s adjusted R-squared value?

0.8087

0.8062

0.6059

0.9045

0.6049

Question 4 - Part C

Using the model in Question 4 - Part A, please use the “Predict” tab to make prediction on the mpg for a car that has a a horsepower of 104.5, weight of 3000, 4 cylinders, a model year of 1980, and an acceleration time of 16 seconds.

14.33611

33.65098

26.47996

26.4906

22.89093

Question 4 - Part D

For the model estimated in Question 4 - Part A, list all the predictors that appear to have a relationship to mpg that are statistically significant at 5%?

weight, year, acceleration

weight, acceleration

acceleration, cylinders

acceleration, year

weight, year

Question 4 - Part E

In Question 4 - Part A, we used mpg as dependent variable and horsepower, cylinders, weight, acceleration, and year as explanatory variables to run a regression. Which of the independent variables has the highest Variance Inflation Factor (VIF)?

cylinders

weight

acceleration

horsepower

year

Question 4 - Part F

Do the VIF values in Question 4 - Part E indicate the presence of multicollinearity?

Yes

No

VIF values measure variance, and are not related to multicollinearity in the model

More info is required

Yes, but VIF values are not related to multicollinearity