HOMEWORK #6

Read:

- Chapter 3 in 'Regression Analysis by Example'.
- R tutorial on 'Multiple Linear Regression'

R Assignment:

Solve the following questions using R. Hand in your R code and output file for Questions 1-3, as well as answers to all questions.

1. A group of researchers are interested in developing a model that describes the gas mileage (in mpg) of sports utility vehicles. They decide to use engine size (in cubic cm), horsepower and weight of the car (in pounds) as explanatory variables. From a random sample of 11 SUVs they obtain the following data:

Engine	HP	Weight	MPG
3471	260	4420	23
2979	225	4586	21
4195	275	4787	20
4701	235	4379	19
3471	240	4439	22
3960	195	3786	21
4701	235	3786	20
4701	265	3786	19
3311	230	3860	24
4664	235	5390	17
4605	302	4834	19

   (a) Find the least-squares regression equation using MPG as the response variable and engine size, weight and horse power as explanatory variables. Write down the regression equation.
   (b) Explain in context what the coefficient corresponding to horse power means.
   (c) Conduct the F-test for the overall fit of the regression. Comment on the results.
   (d) Test each of the individual regression coefficients. Do the results indicate that any of the explanatory variables should be removed from the model?
   (e) Determine the regression model with the explanatory variable(s) identified in part (d) removed.
   (f) Going back to the original model containing all three explanatory variables, construct a 99% confidence interval for the mean gas mileage for SUVs with Engine = 2,000, HP = 250 and Weight = 4,000.
   (g) Construct a 99% prediction interval for the mileage of a particular SUV with Engine = 2,000, HP = 250 and Weight = 4,000.

2. A paper company is interested in making its operations more efficient. They collect data on the total manufacturing cost per month (in dollars), the total production of paper per month (in tons), the total number of machine hours per month, the total variable overhead cost per month (in thousands of dollars) and the total number of labor hours each month. The data can be found at

   www.stat.columbia.edu/~martin/W2024/Data/papercompany.txt

   (a) Fit a multiple regression using cost as the response variable, and the other four variables as explanatory variables. Write down the regression equation.
   (b) Conduct the F-test for the overall fit of the regression. What conclusions can you draw?
   (c) What proportion of the variation in cost has been explained by the regression?
   (d) Perform a partial F-test to determine whether the variables associated with overhead and labor hours can be removed from the model. Comment on the results of the test.

3. Researchers were interested in determining factors that affected the water consumption in the 48 contiguous states. For each state the researchers measured the per capita consumption of water (in gallons per day), the per capita income (in $1000), the average annual rainfall (in inches) and the average cost of 1000 gallons of water (in dollars). The data can be found on the web page:

   http://www.stat.columbia.edu/~martin/W2024/Data/Water.txt

   (a) Fit a multiple regression model using water consumption as the response variable and the other three variables as explanatory variables.
   (b) What proportion of the variation in water consumption has been explained by the regression model?
   (c) Conduct an F-test for the overall fit of the regression model. Comment on the results.
   (d) Test each of the individual regression coefficients. Do the results indicate that any of the explanatory variables can be removed from the model?