Homework#2 (Total score = 100)

COMPSCI- 5590-0012 Econometrics of Data Science
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF MISSOURI-KANSAS CITY

Due on Wednesday, November 16, 2022 at 11:59 PM (in Canvas)

"It must be an individual submission. Any kind of copying or corroboration will be severely penalized".

The spreadsheet "hw2_car_data.csv" contains a dataset related to cars. This dataset was collected in the U.S. It contains the following variables:

Variable name	Variable meaning
foreign	1 if foreign, 0 domestic
mpg	Gas mileage (miles per gallon)
cylinders	Number of cylinders
displacement	Engine displacement (cubic inches)
hp	Horse power
weight	Vehicle weight (pounds)
acceleration	Maximum acceleration (ft/s2)
modelyr	Model year
origin	Continent of origin. 1=North America; 2=Europe; 3=Asia;
name	Vehicle name
Modyr70-82	Binary variable for model year

I. [80 points total] For the available data, estimate a linear regression model with mpg (miles per gallon) as a function of selected variables, including (but not limited to) binary variables for model year and the variable "foreign" (Think about your model!).

To earn credit, please solve the following questions using **R**:

1. (10 points) State and justify your model (i.e., explain briefly why your selected variables are important for your model).

2. (10 points) Present summary statistics (min, max, standard deviation, and selected quantiles) for the variables in your model and briefly comment.

3. (15 points) Using R, generate your model results and interpret them. Interpret the coefficients in terms of the sign, size (magnitude), and significance.

4. (10 points) Do American cars run less mileage per gallon than foreign cars? Use an appropriate statistical test (make sure to detail each step).

5. (10 points) Do the model year variables jointly have explanatory power? Use an appropriate statistical test (make sure to detail each step).

6. (20 points) Check the assumptions of the linear regression model.

7. (5 points) Comment on the model's goodness of fit.

II. [20 points] Summarize the main ideas in “Let's Take the Con Out of Econometrics,” by Edward E. Leamer (source: The American Economic Review, Vol. 73, No. 1 (Mar., 1983), pp. 31-43). Download link of this paper is provided in Canvas.

The summary should be of one-page (typed, 12 pt, Times New Roman, 1.15 spacing, with 1inch margins).

Submission guidelines:

- In the document, first, write down a question and then write your answer after that.
- In the answer script, write your name and student ID on the right-hand side of the header.
- Please type your response. *No* handwritten submission will be accepted.
- Upload a pdf file containing the results, graphs, reasoning, and interpretations.
- Upload the R file.
- Please write the file name (both pdf and R) in this format: COMPSCI-5590-hw2-your last name
- It is an *individual submission*.