THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

REGRESSIONS DATA ANALYSIS PROJECT EMSE 4765 SPRING 2025

Perform the data analyses and answer the questions below in a data analysis report no later than **April 29th, 2025.** You may not work together on this assignment and you should perform your analysis and write the report on your own.

Please provide me with your electronic files uploaded on BLACKBOARD and a HARD COPY OF YOUR WRITTEN REPORT detailing your analysis steps and conclusions.

REGRESSION PROJECT

The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 8 atrributes being:

1. mpg: continuous
2. cylinders: multi-valued discrete
3. displacement: continuous
4. horsepower: continuous
5. weight: continuous
6. acceleration: continuous
7. model year: multi-valued discrete
8. origin: 1 - American, 0 - Non American

A snapshot of the data is provided below. The data set contains data for 105 cars. Denoting the "mpg" as Y please address the following questions below in your written report.

a. Find the estimated linear regression of $Log(Y)$ on an appropriate ("best") subset of predictor variables $X_1, \ldots, X_7$ using the 105 datapoints and interpret the results. Motivate your model development in your report. As part of your model development you need to motivate the use of $Log(Y)$ as the dependent variable in your regression model.

	A	B	C	D	E	F	G	H
1	mpg	cyl	displ	hp	weight	accel	year	origin
2	38	6	262	85	3015	17	2017	0
3	24	4	134	96	2702	13.5	2010	0
4	16	8	318	150	4498	14.5	2010	0
5	26.4	4	140	88	2870	18.1	2015	0
6	13	8	360	175	3821	11	2008	0
7	13	8	318	150	3755	14	2011	0
8	20.5	6	225	100	3430	17.2	2013	0
9	21	4	122	86	2226	16.5	2007	0
10	32.4	4	108	75	2350	16.8	2016	0
11	24	4	90	75	2108	15.5	2009	1

b. Perform a diagnostic analysis of the fitted model chosen in Part a.

c. Forecast the mpg of a car to be developed with the following values of the independent variables and provide a 95% confidence interval and a 95% prediction interval.

cyl	displ	hp	weight	accel	year	origin
3	100	120	2200	15	2022	1