Managerial Report on Vehicle Fuel Consumption

Melanie Fitzgerald
BUS610 - MBASP124

The data file "Vehicle Fuel Consumption Download Vehicle Fuel Consumption" has information regarding 774 unique vehicles from 2022. It provides model-specific fuel consumption ratings (in miles per gallon) and estimated carbon dioxide emissions (in grams per mile) for new light-duty vehicles.

1. Use methods of descriptive statistics to summarize the data.
2. Develop an estimated simple linear regression model that can be used to predict the fuel consumption rate (in MPG), given the Engine Size. Discuss your findings.
3. Develop an estimated multiple linear regression model that could be used to predict the fuel consumption rate (in MPG) using Vehicle Class, Cylinders, and Transmission as independent variables. Discuss your findings.
4. Based on the results in parts (2) and (3), do you believe another regression model may be more appropriate? Estimate this model and discuss your results.
5. What conclusions and recommendations can you derive from your analysis? What vehicles are achieving a substantially higher MPG than would be expected, given their Vehicle Class, Cylinders, and Transmission? What vehicles are achieving a substantially lower MPG than would be expected, given their Vehicle Class, Cylinders, and Transmission? What other independent variables could be included in the model?

1. Use methods of descriptive statistics to summarize the data.

The descriptive statistics summary for the quantitative data engine size, fuel consumption, CO_2 emissions, and cylinders are shown below.

	Engine Size (L)	Fuel Consumption (mpg)	CO_2 Emissions (g/m)	Cylinders
Mean	3.076	27.941	407.588	5.566
Median	2.9	27	400	6
Standard Deviation	1.335	7.658	103.864	1.958
Minimum	1.2	11	151	3
Maximum	8	71	978	16
Range	6.8	60	60	13

The correlation for the quantitative variables is shown below. The correlation data shows a very strong negative correlation between fuel consumption and CO_2 emissions and a negative correlation between fuel consumption and engine size, and fuel consumption and cylinders. This means that the two paired variables move in opposition to each other. As one variable increases, the other variable decreases and vice versa. However, there is a very strong positive relationship with engine size and cylinders and a strong positive correlation between engine size and CO_2 emissions. These variables move in the same direction so that as one increases or decreases, the other will also.

	Engine Size (L)	Fuel Consumption (mpg)	CO_2 Emissions (g/m)	Cylinders
Engine Size (L)	1.000	-0.695	0.822	0.919
Fuel Consumption (mpg)	-0.695	1.000	-0.916	-0.699
CO_2 Emissions (g/m)	0.822	-0.916	1.000	0.849
Cylinders	0.919	-0.699	0.849	1.000

The percentage of each categorical data of make, vehicle class, transmission, and fuel type to the total of 774 vehicles in the data set are shown below.

Make	
Porsche	9.95%
BMW	7.75%
Ford	6.33%
Toyota	5.56%
Audi	5.43%
Chevrolet	5.43%
Mercedes-Benz	4.39%
Jeep	3.88%
Lexus	3.62%
Hyundai	3.62%
Other	48.71%

Vehicle Class	
Compact	8.91%
Full-size	8.27%
Mid-size	15.12%
Minicompact	6.20%
Minivan	0.90%
Subcompact	10.34%
SUV: Small	25.45%
SUV: Standard	18.22%
Two-seater	6.59%

Transmission	
A	21.71%
AM	15.63%
AS	40.44%
AV	11.24%
M	10.98%

Fuel Type	
D	1.42%
X	43.02%
Z	55.56%

Transmission:
A = automatic
AM = automated manual
AS = automatic with select shift
AV = continuously variable
M = manual

Fuel type:
X = regular gasoline
Z = premium gasoline
D = diesel

2. Develop an estimated simple linear regression model that can be used to predict the fuel consumption rate (in MPG), given the Engine Size. Discuss your findings.

The standard significance level of α = 0.05 will be used for this report

Excel was used to determine an estimated linear regression model for the fuel consumption given the engine size. The summary output is below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.695
R Square	0.483
Adjusted R Square	0.482
Standard Error	5.509
Observations	774

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	21897.855	21897.855	721.535	9.738E-113
Residual	772	23429.411	30.349		
Total	773	45327.266			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	40.210	0.498	80.770	0	39.232	41.187	39.232	41.187
Engine Size (L)	-3.988	0.148	-26.861	9.738E-113	-4.280	-3.697	-4.280	-3.697

Based on the summary output, the estimated linear regression equation for $\hat{y}$ = fuel consumption given the engine size can be written as:

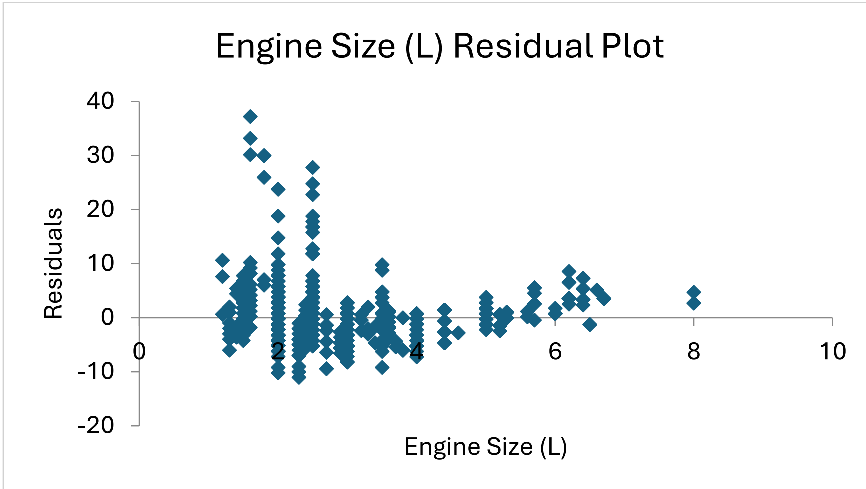
$$\hat{y} = 40.209 – 3.988(Engine\ Size)$$

or in terms of x = Engine Size(L)

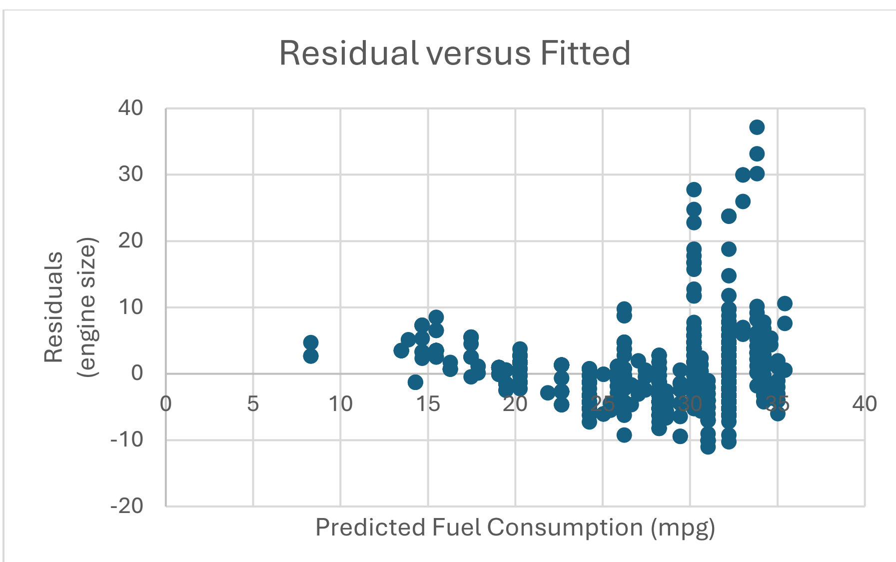
$$\hat{y} = 40.209 - 3.988x$$

The coefficient of determination R^2 shows that only 48.3% of the data's variability can be explained by the linear regression model. However, before making infererences from the estimated model, the conditions for validity of inference must be met. Regression analysis must meet certain conditions for validity of inference to justify the use of a regression model. The primary condition that must be met is for the error terms to be normally distributed with constant variance. Residual plots of the error terms can be examined to test for this condition.

The residual plot for the estimated regression model below shows variability among the residual data points and appears to follow a pattern where data points are clustered below the *0*-axis for small to mid-range sizes and increases to cluster above the *0*-axis for larger sizes. This violates the condition of a randomized, normal distribution of constant variance suggesting that the estimated linear regression model is not a good fit to make inferences about the relationship between fuel consumption and engine size.



A residual versus fitted (estimated responses) plot is also used to identify nonlinearity, error variances, and outliers. The residual versus fitted plot below for the predicted fuel consumption from the regression model is consistent with the observations with the residual plot above.



3. Develop an estimated multiple linear regression model that could be used to predict the fuel consumption rate (in MPG) using Vehicle Class, Cylinders, and Transmission as independent variables. Discuss your findings.

Excel was used to determine the coefficients to estimate a multiple linear regression model for fuel consumption given three categories: cylinder, transmission, and vehicle class. Cylinder (3, 4, 5, 6, 8, 10, 12,16) is quantitively measured, but both transmission (A, AM, AS, AV, M) and vehicle class (Compact, Full-size, Mid-size, Minicompact, Minivan, Subcompact, SUV:Small, SUV:Standard, Two-seater) are categorical data. To obtain a linear regression model, dummy variables were used to represent the categorical data. The regression summary output is below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.817
R Square	0.667
Adjusted R Square	0.661
Standard Error	4.456
Observations	774

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	30238.742	2326.057	117.162	1.8002E-171
Residual	760	15088.524	19.853		
Total	773	45327.266			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	38.818	1.002	38.732	0.000	36.851	40.786	36.851	40.786
Cylinders	-2.334	0.099	-23.677	0.000	-2.527	-2.140	-2.527	-2.140
Transmission A	0.782	0.669	1.168	0.243	-0.533	2.096	-0.533	2.096
Transmission AM	2.035	0.646	3.149	0.002	0.766	3.305	0.766	3.305
Transmission AS	0.488	0.592	0.825	0.410	-0.674	1.651	-0.674	1.651
Transmission AV	8.962	0.722	12.418	0.000	7.546	10.379	7.546	10.379
Vehicle Compact	3.133	0.889	3.522	0.000	1.387	4.879	1.387	4.879
Vehicle Full-size	2.922	0.863	3.384	0.001	1.227	4.617	1.227	4.617
Vehicle Mid-size	2.436	0.803	3.034	0.002	0.860	4.012	0.860	4.012
Vehicle Minicompact	-1.438	0.905	-1.589	0.112	-3.215	0.338	-3.215	0.338
Vehicle Minivan	1.815	1.840	0.986	0.324	-1.797	5.427	-1.797	5.427
Vehicle Subcompact	0.904	0.831	1.087	0.278	-0.729	2.536	-0.729	2.536
Vehicle SUV:Small	-0.946	0.805	-1.175	0.241	-2.526	0.635	-2.526	0.635
Vehicle SUV:Standard	-1.375	0.809	-1.699	0.090	-2.964	0.213	-2.964	0.213

From the summary output, the estimated linear regression model for $\hat{y}$ = fuel consumption can be written as:

$$\hat{y} = 38.818 - 2.334(\text{cylinders}) + 0.782(\text{Transmission A}) + 2.035(\text{Transmission AM}) + 0.488(\text{Transmission AS}) + 8.962(\text{Transmission AV}) + 3.133(\text{Vehicle Compact}) + 2.922(\text{Vehicle Full-size}) + 2.436(\text{Vehicle Mid-size}) – 1.438(\text{Vehicle Minicompact}) + 1.815(\text{Vehicle Minivan}) + 0.904(\text{Vehicle Subcompact}) – 0.946(\text{Vehicle SUV:Small}) – 1.375(\text{Vehicle SUV:Standard})$$

or in terms of x_i where i = 1-13 as defined below.

$$\hat{y} = 38.818 - 2.334x_1 + 0.782x_2 + 2.035x_3 + 0.488x_4 + 8.962x_5 + 3.133x_6 + 2.922x_7 + 2.436x_8 - 1.438x_9 + 1.815x_{10} + 0.904x_{11} - 0.946x_{12} - 1.375x_{13}$$

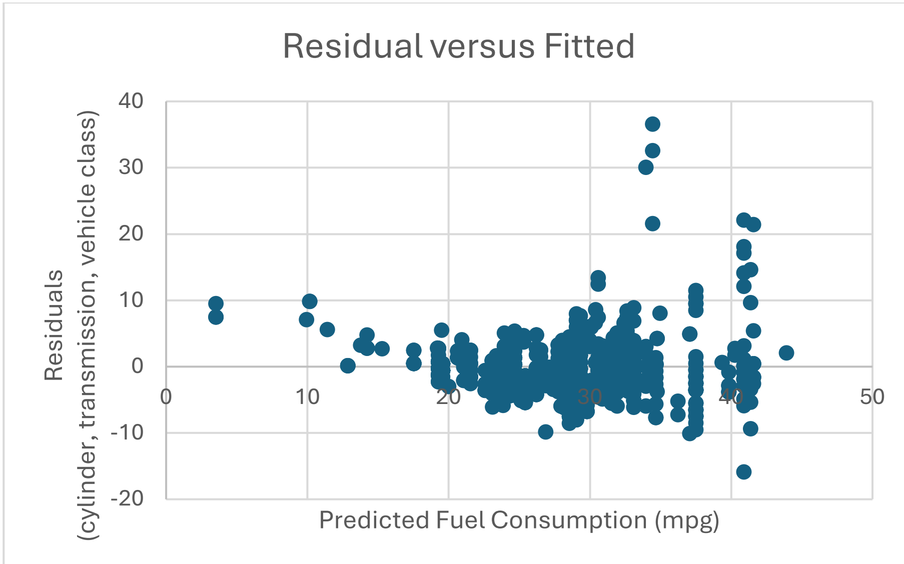
x_1 = Cylinders
x_2 = Transmission A
x_3 = Transmission AM
x_4 = Transmission AS
x_5 = Transmission AV
x_6 = Vehicle Compact
x_7 = Vehicle Full-Size
x_8 = Vehicle Mid-Size
x_9 = Vehicle Minicompact
x_{10} = Vehicle Minivan
x_{11} = Vehicle Subcompact
x_{12} = Vehicle SUV:Small
x_{13} = Vehicle SUV:Standard

Transmission M and Vehicle Two-seater were used as reference variables (not assigned a dummy variable) for their respective categories and are assumed when $x_2=x_3=x_4=x_5=0$ and $x_6=x_7=x_8=x_9=x_{10}=x_{11}=x_{12}=x_{13}=0$, respectively.

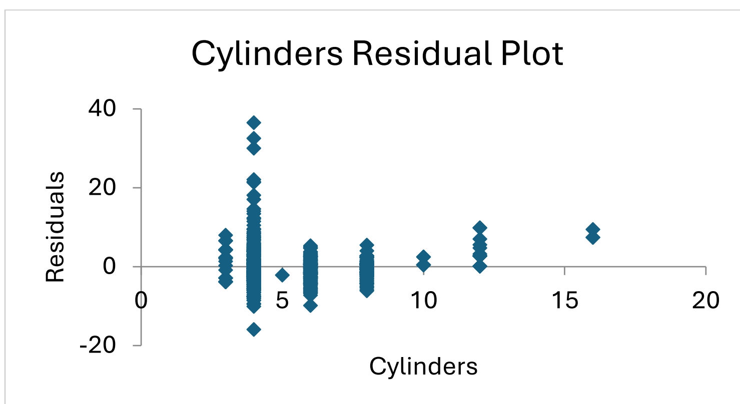
The adjusted R^2 value for the regression shows that only 66.1% of the data's variability can be explained by the model. Given the model has 13 independent variables, it would be worthwhile to test for multicollinearity, the degree of correlation between two independent variables, to ensure the validity of using the independent variables together in the regression model. One way to test for multicollinearity is to look at the correlation between the independent variables. Correlation coefficients greater than 0.7 suggest a correlation between variables. Below is the correlation matrix for the independent variables. None of the correlation coefficients for each pair of independent variables is greater than 0.7 so we can conclude that there is minimal multicollinearity among the variables.

	Cylinders	*Transmission A*	*Transmission AM*	*Transmission AS*	*Transmission AV*	*Vehicle Compact*	*Vehicle Full-size*	*Vehicle Mid-size*	*Vehicle Minicompact*	*Vehicle Minivan*	*Vehicle Subcompact*	*Vehicle SUV:Small*	*Vehicle SUV:Standard*
Cylinders	1.000												
Transmission A	0.150	1.000											
Transmission AM	0.092	-0.227	1.000										
Transmission AS	0.058	-0.434	-0.355	1.000									
Transmission AV	-0.281	-0.187	-0.153	-0.293	1.000								
Vehicle Compact	-0.172	-0.132	-0.035	0.056	0.047	1.000							
Vehicle Full-size	0.134	0.024	0.103	-0.066	0.027	-0.094	1.000						
Vehicle Mid-size	-0.052	-0.065	-0.062	-0.010	0.101	-0.132	-0.127	1.000					
Vehicle Minicompact	0.068	-0.096	0.258	-0.168	-0.092	-0.080	-0.077	-0.109	1.000				
Vehicle Minivan	-0.007	0.049	-0.041	-0.023	0.052	-0.030	-0.029	-0.040	-0.025	1.000			
Vehicle Subcompact	0.004	-0.076	0.041	0.023	-0.094	-0.106	-0.102	-0.143	-0.087	-0.032	1.000		
Vehicle SUV:Small	-0.343	0.031	-0.121	0.069	0.130	-0.183	-0.175	-0.247	-0.150	-0.056	-0.198	1.000	
Vehicle SUV:Standard	0.279	0.296	-0.203	0.082	-0.115	-0.148	-0.142	-0.199	-0.121	-0.045	-0.160	-0.276	1.000

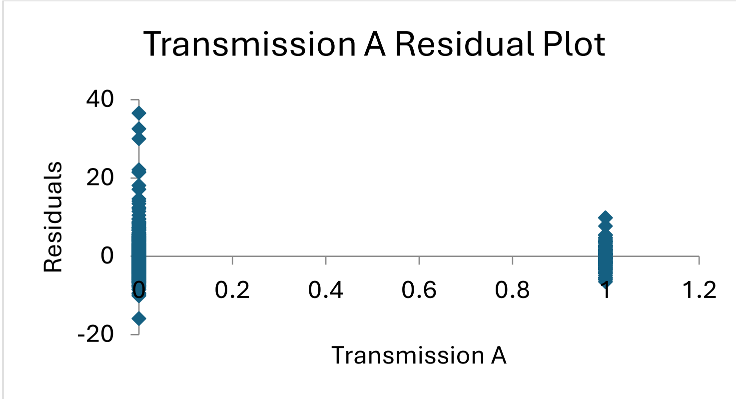
Before inferences can be made using the model, it still needs to be tested for validity of inference. The residual versus fitted plot below for the estimated linear model shows a pattern where the variances change with the predicted fuel consumption. This violates the condition for a randomized, normal distribution of constant variance suggesting the estimated multiple linear regression model may not be a good fit.

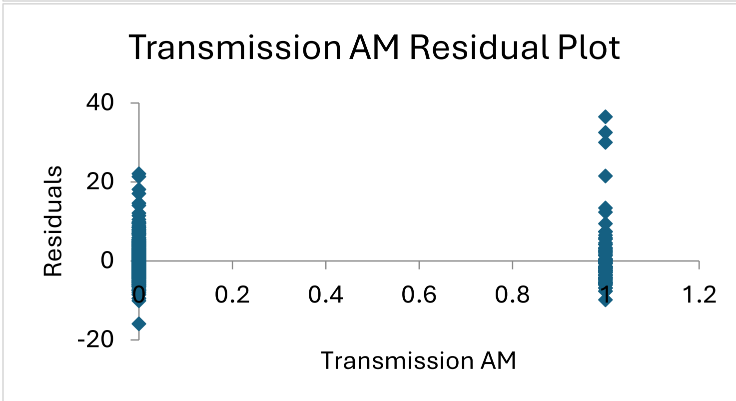


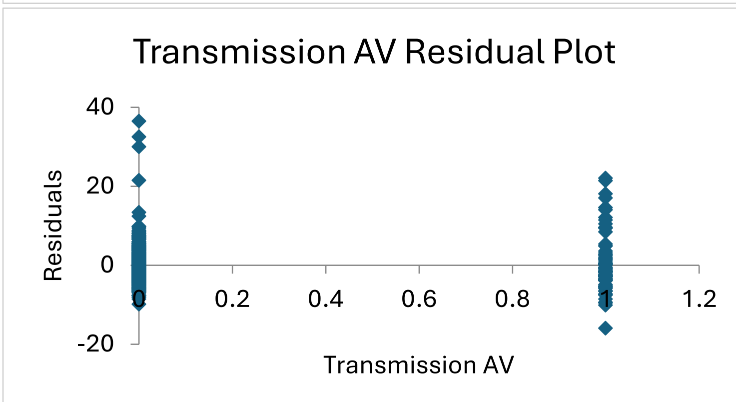
Each independent variable residual plot is also shown below. The cylinder residual plot shows a pattern where the data points cluster below the *0*-axis for smaller cylinder sizes and increases to cluster above the *0*-axis for larger sizes which violates the condition for a normal distribution of constant variance.

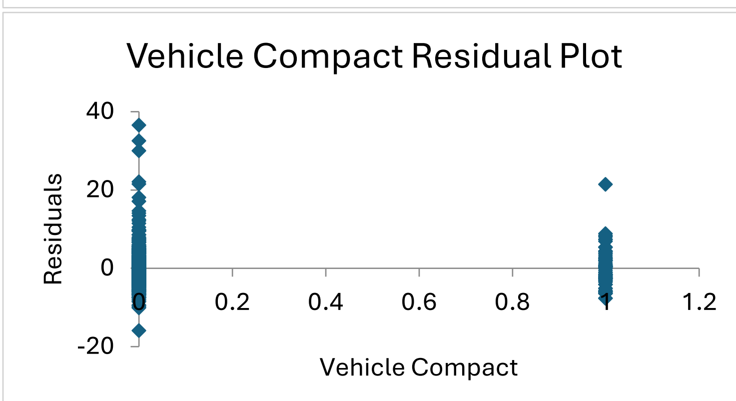


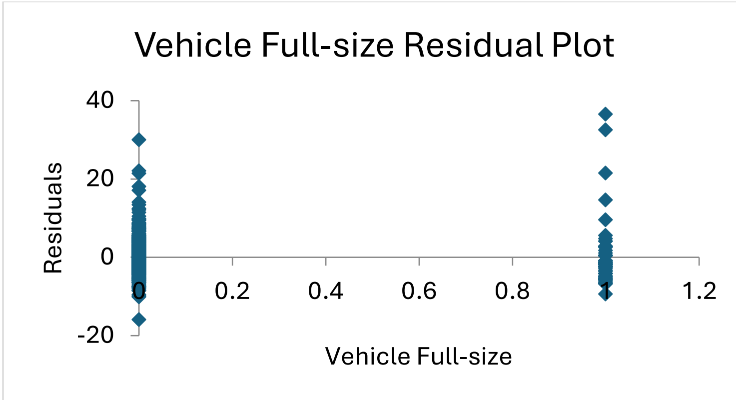
Unlike the cylinder residual plot, the ones for transmission and vehicle class are based on binary data for each categorical independent variable and observations of these residuals are to determine the similarity of spread around the *0*-axis for constant variance. The residual plots of the categorical dummy variables for transmission and vehicle class are below and several show different variances between the binary values which violate the condition for a normal distribution of constant variance.

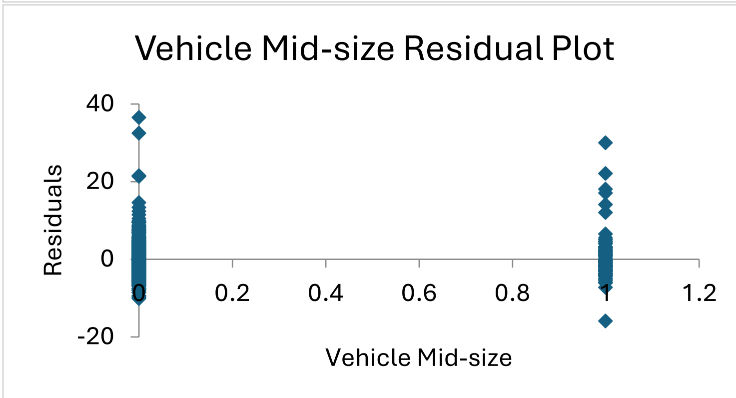

Transmission A Residual Plot
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
1.2
Residuals
Transmission A

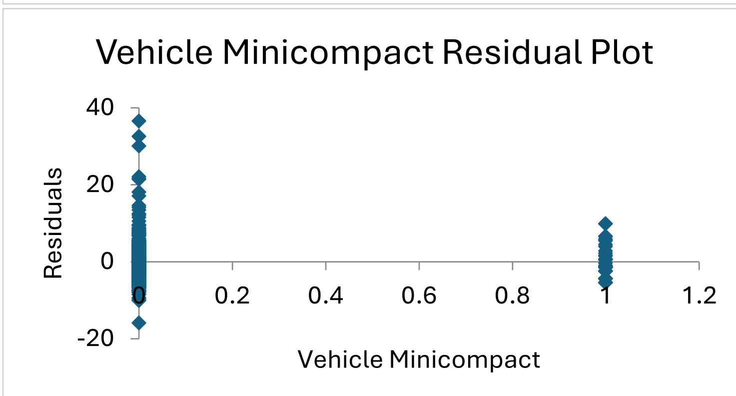

Transmission AM Residual Plot
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
1.2
Residuals
Transmission AM

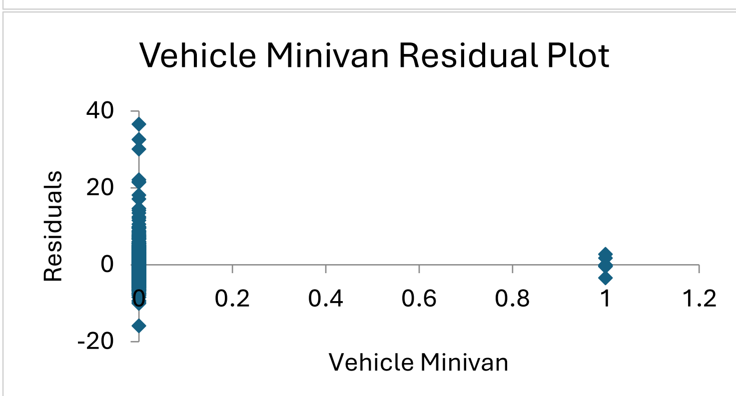

Transmission AV Residual Plot
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
1.2
Residuals
Transmission AV

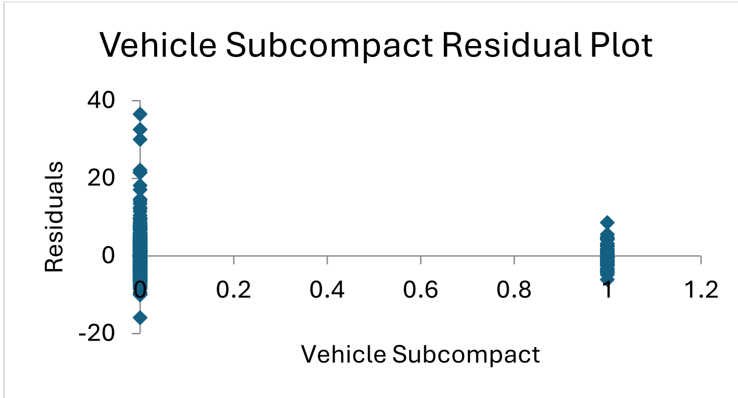

Vehicle Compact Residual Plot
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
1.2
Residuals
Vehicle Compact

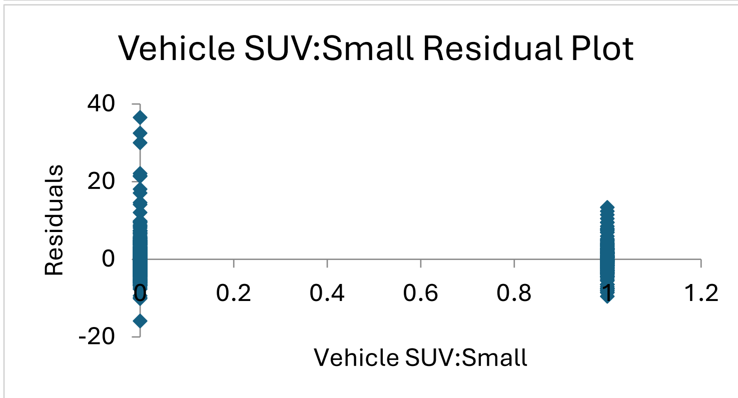

Vehicle Full-size Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle Full-size

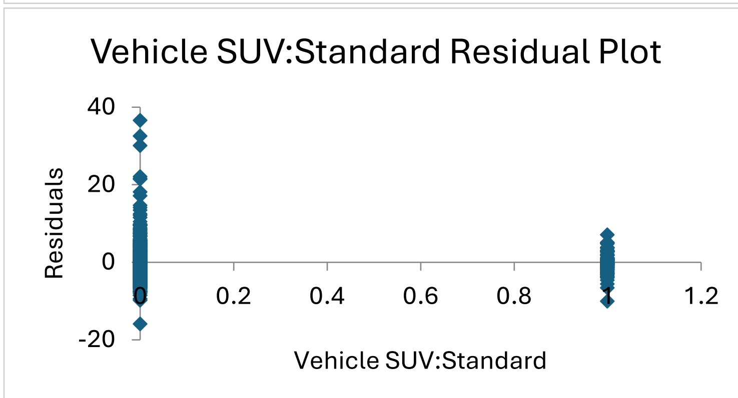

Vehicle Mid-size Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle Mid-size


Vehicle Minicompact Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle Minicompact


Vehicle Minivan Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle Minivan

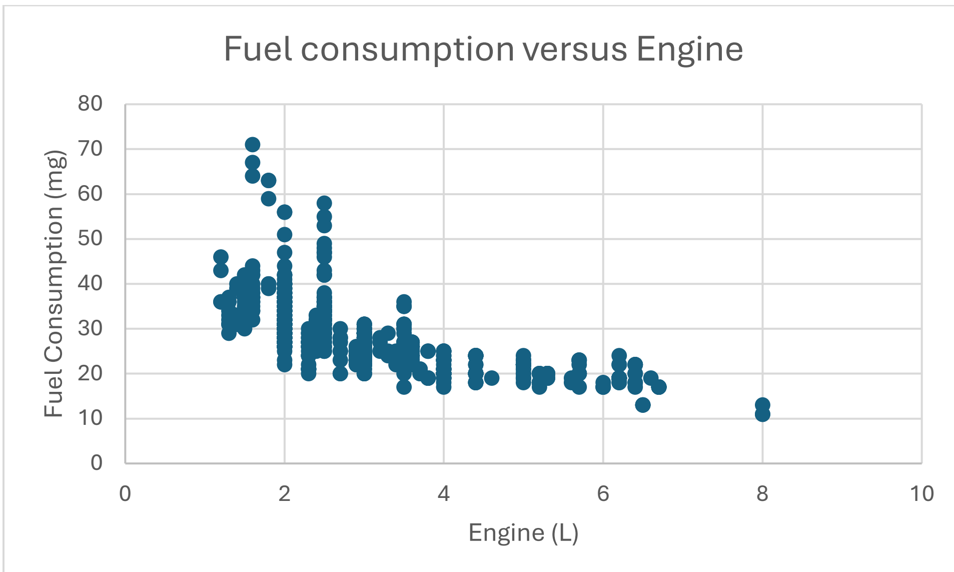

Vehicle Subcompact Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle Subcompact

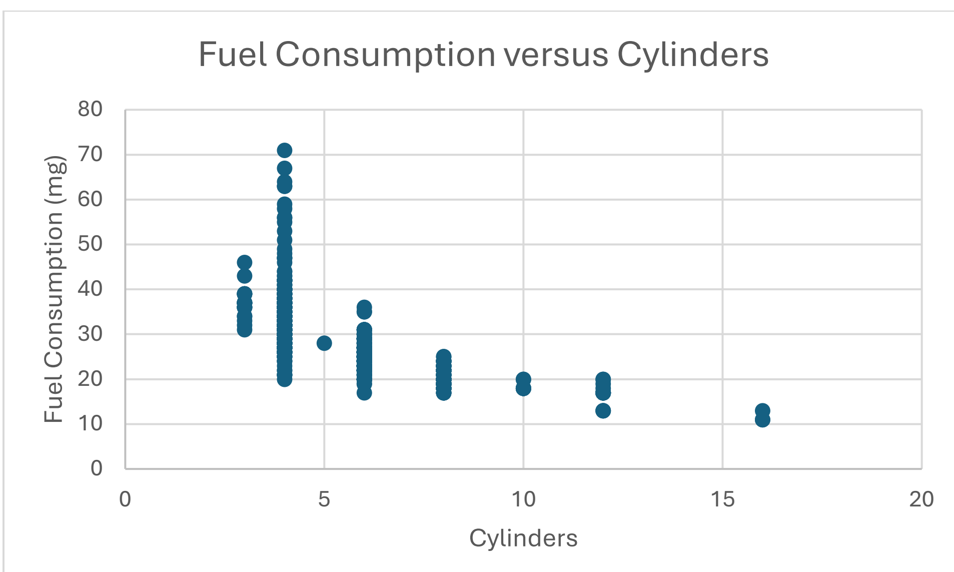

Vehicle SUV:Small Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle SUV:Small


Vehicle SUV:Standard Residual Plot
40
20
0
-20
Residuals
0
0.2
0.4
0.6
0.8
1
1.2
Vehicle SUV:Standard

4. Based on the results in parts (2) and (3), do you believe another regression model may be more appropriate? Estimate this model and discuss your results.

Based on Parts 2 and 3, a nonlinear regression may be a better fit given the patterns observed in the residual plots. To determine which nonlinear regression could be appropriate, I looked at the plots for fuel consumption against each of the quantitative data, engine and cylinders, to see if there was a pattern. Both graphs show a logarithmic trend in the data so it seems reasonable to apply a logarithmic regression to the data.





However, engine size and cylinders have a strong correlation as noted in the correlation matrix in Part 1 with a correlation coefficient of 0.919, which makes sense since cylinders are part of engine design. These two variables should not be in the same regression analysis as it would result in multicollinearity. I decided to use the same independent variables in Part 2 and apply a logarithmic regression. A logarithmic regression is a modified linear regression using the log value of chosen variables to try to linearize nonlinear data. For this logarithmic regression, the natural log of fuel consumption is taken for the dependent variable given independent variables cylinders, transmission, and vehicle class. The regression summary output is below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.871
R Square	0.758
Adjusted R Square	0.754
Standard Error	0.128
Observations	774

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	38.904	2.993	183.054	0.000
Residual	760	12.425	0.016		
Total	773	51.329			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.736	0.029	129.909	0.000	3.680	3.793	3.680	3.793
Cylinders	-0.093	0.003	-32.779	0.000	-0.098	-0.087	-0.098	-0.087
Transmission A	0.050	0.019	2.621	0.009	0.013	0.088	0.013	0.088
Transmission AM	0.056	0.019	3.011	0.003	0.019	0.092	0.019	0.092
Transmission AS	0.035	0.017	2.088	0.037	0.002	0.069	0.002	0.069
Transmission AV	0.263	0.021	12.705	0.000	0.222	0.304	0.222	0.304
Vehicle Compact	0.098	0.026	3.855	0.000	0.048	0.148	0.048	0.148
Vehicle Full-size	0.087	0.025	3.517	0.000	0.039	0.136	0.039	0.136
Vehicle Mid-size	0.077	0.023	3.334	0.001	0.032	0.122	0.032	0.122
Vehicle Minicompact	-0.031	0.026	-1.188	0.235	-0.082	0.020	-0.082	0.020
Vehicle Minivan	0.069	0.053	1.303	0.193	-0.035	0.172	-0.035	0.172
Vehicle Subcompact	0.040	0.024	1.685	0.092	-0.007	0.087	-0.007	0.087
Vehicle SUV:Small	-0.030	0.023	-1.296	0.195	-0.075	0.015	-0.075	0.015
Vehicle SUV:Standard	-0.058	0.023	-2.480	0.013	-0.103	-0.012	-0.103	-0.012

From the summary output, the estimated regression model for $ln(\hat{y})$ = fuel consumption can be written as:

ln(ŷ) = 3.736 – 0.093(cylinders) + 0.050(Transmission A) + 0.056(Transmission AM) + 0.035(Transmission AS) + 0.263(Transmission AV) + 0.098(Vehicle Compact) + 0.087(Vehicle Full-size) + 0.077(Vehicle Mid-size) – 0.031(Vehicle Minicompact) + 0.069(Vehicle Minivan) + 0.040(Vehicle Subcompact) – 0.030(Vehicle SUV:Small) – 0.058(Vehicle SUV:Standard)

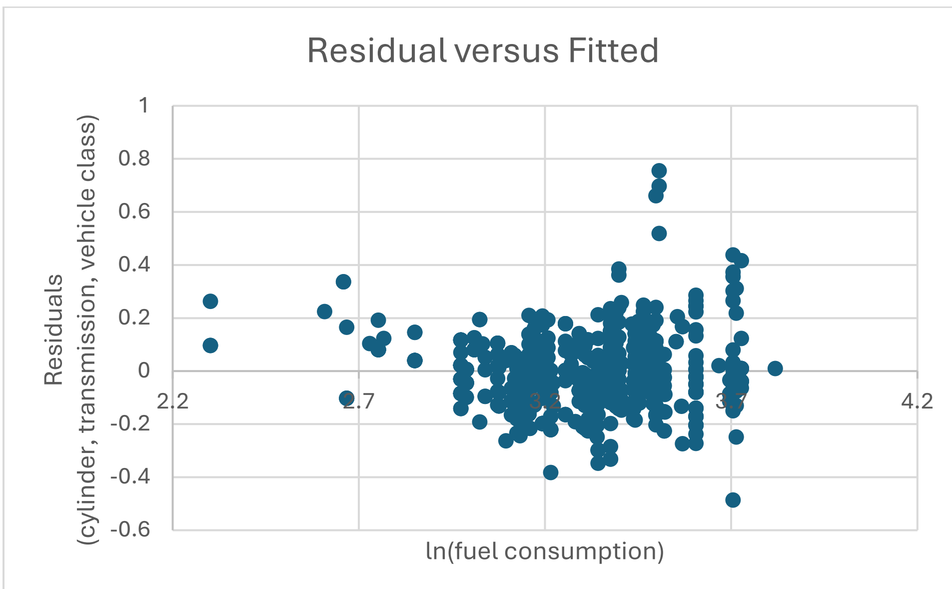
or in terms of x_i where x_i = 1-13 is defined as in Part 2:

$$ln(\hat{y}) = 3.736 - 0.093x_1 + 0.050x_2 + 0.056x_3 + 0.035x_4 + 0.263x_5 + 0.098x_6 + 0.087x_7 + 0.077x_8 - 0.031x_9 + 0.069x_{10} + 0.040x_{11} - 0.030x_{12} - 0.058x_{13}$$

The regression model can also be written in terms of $\hat{y}$ = fuel consumption by exponentiating the equation and following the product rule of exponents exp(a+b) = exp(a)*exp(b):

$$\hat{y} = exp(3.736)*exp(-0.093x_1)*exp(0.050x_2)*exp(0.056x_3)*exp(0.035x_4)*exp(0.263x_5)* exp(0.098x_6)*exp(0.087x_7)*exp(0.077x_8)*exp(-0.031x_9)*exp(0.069x_{10})*exp(0.040x_{11})* exp(-0.030x_{12})*exp(-0.058x_{13})$$

The adjusted R^2 value is 0.754 which means 75.4% of the data's variability can be explained by the regression model. This is a 9.3% increase from Part 3. As before, the conditions for validity of inference must be met before using the model to make inferences. In the residual versus fitted plot for the regression model below, the residual points look to be normally distributed around the *0*-axis, showing mostly constant variance, and no specific pattern. There are outliers but this should not affect the model significantly. This meets the condition for validity of inference which means the logarithmic regression model can be used to make inferences on the data.



The *p*-value for the model is 0.000 (unrounded 9.5759E-224) which is less than the significance level of 0.05. This means the independent variables overall are significant to the model and their ability to explain variablity in the dependent variable. Most of the *p*-values for each independent variable are less than 0.05 which means they are each statistically significant to the model. However, there are four independent variables with *p*-values greater than the significance level of 0.05 which usually indicates they have little effect to the dependent variable. However, since the sample size is fairly large and the variables are part of a category, keeping them in the equation should have little effect to the overall interpretation of the model.

Interpreting the coefficients is different for a logarithmic model versus a linear one. The coefficients represent a multiplicative change to the natural log of the dependent variable unlike an additive change in a linear model. For the categorical data, the coefficients represent an expected change in the natural log of the dependent variable relative to the reference variable. We can exponentiate each coefficient β_i

and convert to a percentage and make the following conclusions about fuel consumption (mpg) for each independent variable while holding the other variables constant:

	Coefficient β	exp(β)	exp(β) - 1	Percentage
Cylinders	-0.093	0.911	-0.089	-8.85%
Transmission A	0.050	1.052	0.052	5.16%
Transmission AM	0.056	1.057	0.057	5.74%
Transmission AS	0.035	1.036	0.036	3.61%
Transmission AV	0.263	1.301	0.301	30.10%
Vehicle Compact	0.098	1.103	0.103	10.34%
Vehicle Full-size	0.087	1.091	0.091	9.11%
Vehicle Mid-size	0.077	1.080	0.080	7.98%
Vehicle Minicompact	-0.031	0.970	-0.030	-3.04%
Vehicle Minivan	0.069	1.071	0.071	7.12%
Vehicle Subcompact	0.040	1.041	0.041	4.10%
Vehicle SUV:Small	-0.030	0.971	-0.029	-2.95%
Vehicle SUV:Standard	-0.058	0.944	-0.056	-5.60%

Fuel consumption will be 8.85% lower with a cylinder increase of 1 unit.

On average, fuel consumption will be 5.16% higher with Transmission A than Transmission M.

On average, fuel consumption will be 5.74% higher with Transmission AM then Transmission M.

On average, fuel consumption will be 3.61% higher with Transmission AS than Transmission M.

On average, fuel consumption will be 30.10% higher with Transmission AV than Transmission M.

On average, fuel consumption will be 10.34% higher for Vehicle Class Compact than Vehicle Class Two-seater.

On average, fuel consumption will be 9.11% higher for Vehicle Class Full-size than Vehicle Class Two-seater.

On average, fuel consumption will be 7.98% higher for Vehicle Class Mid-size than Vehicle Class Two-seater.

On average, fuel consumption will be 3.04% lower for Vehicle Class Minicompact than Vehicle Class Two-seater.

On average, fuel consumption will be 7.12% higher for Vehicle Class Minivan than Vehicle Class Two-seater.

On average, fuel consumption will be 4.10% higher for Vehicle Class Subcompact than Vehicle Class Two-seater.

On average, fuel consumption will be 2.95% lower for Vehicle Class SUV:Small than Vehicle Class Two-seater.

On average, fuel consumption will be 5.60% lower for Vehicle Class SUV:Standard than Vehicle Class Two-seater.

5. What conclusions and recommendations can you derive from your analysis? What vehicles are achieving a substantially higher MPG than would be expected, given their Vehicle Class, Cylinders, and Transmission? What vehicles are achieving a substantially lower MPG than would be expected, given their Vehicle Class, Cylinders, and Transmission? What other independent variables could be included in the model?

Based on the analysis, the vehicles that were predicted to have the highest fuel consumption (greater than 40 mpg) all have AV transmission, cylinder size 3-4, engine size 2 or smaller, with varied vehicle class. The vehicles that have the worst fuel consumption (less than 18 mpg) have either AM, AS, or A transmission, cylinder size 12 or larger, engine size greater than 5, and varied vehicle class. Both the mean and median fuel consumption is 28 mpg with standard deviation of 6 mpg and mode of 29 mpg.

Calculating the percent change from the actual to the predicted fuel consumption, the vehicles that have a greater than 30% *increase* percent change are shown below:

Model Year	Make	Model	Engine Size (L)	Cylinders	Transmission	Vehicle Class	Fuel Consumption (mpg)	Predicted Fuel Consumption (mpg)	Percent change (predicted-original)/original
2022	Subaru	WRX AWD	2.4	4	AV	Mid-size	25	41	63%
2022	Ford	GT	3.5	6	AM	Two-seater	17	25	50%
2022	Ford	Bronco Badlands 4WD	2.3	4	M	SUV: Small	20	28	40%
2022	Ford	Bronco Badlands 4WD	2.3	4	AS	SUV: Small	21	29	39%
2022	Ford	Bronco Sasquatch 4WD	2.3	4	AS	SUV: Small	21	29	39%
2022	Ford	Bronco Black Diamond 4WD	2.3	4	M	SUV: Small	21	28	34%
2022	Ford	Bronco Sasquatch 4WD	2.3	4	M	SUV: Small	21	28	34%
2022	Ford	Bronco Black Diamond 4WD	2.3	4	AS	SUV: Small	22	29	32%
2022	Subaru	Ascent AWD	2.4	4	AV	SUV: Standard	27	36	32%
2022	Subaru	Outback Wilderness AWD	2.4	4	AV	SUV: Small	28	37	31%

The vehicles that have a greater than 30% *decrease* percent change are shown below.

Model Year	Make	Model	Engine Size (L)	Cylinders	Transmission	Vehicle Class	Fuel Consumption (mpg)	Predicted Fuel Consumption (mpg)	Percent change (predicted-original)/original
2022	Hyundai	IONIQ Blue	1.6	4	AM	Full-size	71	33	-53%
2022	Hyundai	IONIQ	1.6	4	AM	Full-size	67	33	-50%
2022	Hyundai	Elantra Hybrid Blue	1.6	4	AM	Mid-size	64	33	-48%
2022	Hyundai	Sonata Hybrid	2	4	AM	Full-size	56	33	-40%
2022	Toyota	Prius	1.8	4	AV	Mid-size	63	41	-35%
2022	Toyota	Corolla Hybrid	1.8	4	AV	Compact	63	42	-34%
2022	Hyundai	Tucson Hybrid	1.6	4	AM	SUV: Small	44	30	-32%
2022	Toyota	Prius AWD	1.8	4	AV	Mid-size	59	41	-31%
2022	Kia	Sorento Hybrid AWD	1.6	4	AM	SUV: Small	43	30	-31%

From the given and predicted data, a compact vehicle with AV transmission with the smallest engine and cylinder sizes is recommended for optimum fuel consumption. Fuel consumption does not have a linear relationship with a single variable, but instead a nonlinear relationship with multiple parameters. The estimated logarithmic linear regression model took into account cylinder size, transmission type, and vehicle class. However, these are not the only factors that could affect fuel consumption. Some other variables to consider in the model which are not in the original data set are vehicle speed, type of driving (city versus highway), and vehicle weight.