Regression Analysis on Blue Book Modified Car Data 2015

Lia S.

STA 6013 Regression Analysis

12/2/2020

Problem Description

The objective of this analysis is to select the important features of a used car predict the future sales price based on the BlueBook value. The data used for this analysis is the Blue Book Modified Car dataset, year 2015. This data set has eleven predictor variables that describe the condition of the car. Mileage is the only numerical predictor variable, and the remaining are categorical values. There is a total of 79 different sub-categories options in this dataset. The predictor variables are Make, Model, Trim, Type, Cylinder, Liter Cruise, Sound and Doors. The response variable is Price. A review of the box plot, scatter plots, histogram, and distribution plots reveal similar mean, ranges, and distribution among the different sub-features within the predictors. Due to the similarities, sub-groups will be combined to minimize the number of sub-features of the predictor variables. The result of this transformation of predictor variables are a reduction in subcategories to 28 unique predictors. The loess plots and scatter plots for this data shows that there may be high variation and correlation within our data. The proposed model will be y = $b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + b_{10}x_{10} + b_{11}x_{11} + b_{12}x_{12} + b_{13}x_{13} + b_{14}x_{14} + b_{15}x_{15} + b_{16}x_{16} + b_{17}x_{17} + b_{18}x_{18} + b_{19}x_{19} + b_{21}x_{21} + b_{22}x_{22} + b_{23}x_{23} + b_{24}x_{24} + b_{25}x_{25} + b_{26}x_{26} + b_{27}x_{27} + b_{28}x_{28} + \varepsilon$ where each x_i will represent a predictor in the model and each b_i will represent the predictor's coefficient. This chart is listed in Appendix section 1. Within each predictor, the sub-feature has been set to a binary matrix and one sub-feature has been selected as the default criteria. The plan modeling method will follow a generalized regression model. The assumptions will be that the data following a normal distribution, errors are uncorrelated, and the data is normally distributed. Observation with price over $50,000 will be scrutinized as possible outliers.

The prediction equation for this the initial model : $\hat{y}$ = 49847 -0.186044x_1 – 9776.185x_2 – 1458.50x_3 -10741x_5 – 7537.0400x_6 – 855.103x_8 + 1691.5433x_9 – 7966.7573x_{11} – 2791.937x_{12} − 4744.138x_{13}– 4098.73x_{14} − 9818.43x_{16}– 4569.693x_{17} − 29.0255x_{19} − 393.686x_{23} − 941.732x_{25} − 356.30x_{27}. The categorical predictors have been set up in a binary matrix to represent 1,0 vectors. The F-value of this AVONA table is 600.95 and the P-value is <0.0001. The result from the model suggest that the predictors in this model can explain the relationship in this dataset. When using the hypothesis $H_o: B_i = 0$ verse $H_o: B_i \neq 0$ we can reject H_o in favor of $H_o: B_i \neq 0$, such that there is at least on e $B_i \neq 0$ that is explaining the variation in the data. The results from the t-value and p-value from the table C.2 show predictors Mileage, Make_1, Make_2, Model_1, Model_2, Trim_2, Type Convertible, Type Coupe, Type Hatchback, Type Sedan, Type Wagon, Cylinder 6, Cylinder 8, and Sound 0 are significant to the model. These variables explain 93.36% of the variation in the data set. The conclusion for this model is that there could be some improvement with the order of variable selection.

The diagnostic plots for this model show there is a reason to be concern with the residuals of the dataset. The results of the R Student residual of the response variable price shows several observations beyond $\pm$ 3 standard deviation. Viewing the predicted points on the Q-Q plot shows possible outliner and many influential points near the right upper end of the graph. The criteria for Influential points are: DFFITS: 0.3739, DFBETAS $\pm$ 0.7334, Cook's D at 0.0054. There are numerous observations beyond these points.

High collinearity problems are present in predictors relationship: Make 1: Liter 2, Make 1: Liter 1 and Model 3: Make 2. To reduce the number of extreme observation present, a weighted value was added to the observation based on the R Student residual value. R student residual greater than ± 3 was given a weight of 0.1 and R student residual greater than $\pm$ 2 was given a weight of 0.5. Based on the student residuals of price of price after weighting down the residual, a transformation is needed to correct the evidence of non-constant variance present in the R student residual plots. After the refit of the model, a Box Cox analysis was done, and a lambda log y was selected for the transformation of the response variable. The residuals for the response variable price are now within the $\pm$ 3 and the residuals for the predicted value has a reasonable constant variance pattern. The lasso selection method was used to select the subset of variables that meet the $\alpha = 0.05$ criteria for test of significance. The variables selected were $x_1, x_3, x_4, x_6, x_7, x_9, x_{13}, x_{11}, x_{15}, x_{27}, x_{28}$. The final prediction equation The final prediction equation is: $log(\hat{y})$= 9.95457 $-$ 0.00000805 $x_1 + 0.24334x_3 + 0.55226x_4 + 0.14223x_6 + 0.38715x_7 + 0.02953x_9 -$ $0.10023x_{13} + 0.30649x_{11} + 0.1664x_{15} - 0.34495x_{27} + 0.18026x_{28}$ where log(y) is the logarithmic value of the sales price. The recommendations for this model to improve the fit is to either collect more data or continue to weight down extreme observation.

Investigation of the Data

A manual review of the data set and output from this process can be seen in Appendix pages 1-13. A means table of the descriptive variables (Price, mileage, cruise, sound, and leather) can be seen on page 1. Sixty four percent of car price are within $11,906.951 through $31,967.13. There appears to be 1 observation of price at $70,755.47, six observations in the $60,000 range and four observation in the $50,000 range that may cause an issue with modeling the data set. The mean sales price is $21,937.04, the minimal price is $8,638.93 and the maximum price is $70,755.467. The box plots and histogram distribution also show extreme vales on the right side of the distribution. The predictor mileage mean is 19,791.304, the minimal is 266, and the maximum is 50,387. There seems to be some extreme values on the right side of the distribution for this predictor since the max value is a far distance from the mean. As expected, mileage has several extreme vales as shown in the box plot and histogram. The histogram of mileage appears to show a normally distribution of the data. The predictor Cruise is a binary indicator variable were 1 is the event of having this feature and 0 otherwise. The boxplot shows that 77.7% of the cars have cruise control. The predictor Sound is a binary indicator with 1 has having sound and 0 otherwise. The histogram shows that 69.8% of the cars have sound. This graph shows a logistic graph that favors 1. The predictor Leather is a binary indicator with 1 as having Leather interior and 0 otherwise. Statistical summary shows that 74.3% of the cars have leather. This graph shows a logistic shape that favors 1. The predictor Doors is a binary indicator regression with a 1 as having two car doors and 4 doors otherwise. boxplot shows that the means are the same between two doors and four door vehicles. The predictor Cylinder has three different features, 4, 6, 8 and will use a binary assignment when a feature is chosen. The box plot for cylinder shows that the mean value for 4 and 6 cylinder a close to the same and 8 cylinder has a higher mean. In the Litter box plot, liters 1.6, 1.8, 2.2, 3.1, 3.4, 3.5 appear to be the same mean, liters 2, 2.3, 2.5, 2.8, and 3.6 appear to be the same and liter 4.6, 5.7, and 6 to be the same when compare to price. The Make box

plot shows five different vehicle makes. Chevrolet and Pontiac show similar average sale's price. Buick Cadillac and SAAB have different means and ranges for Make. A binary indicator will be used to show when a make is selected. The Model box plot shows 16 different features. By observation, there are several mode features that have similar means and ranges. A grouping of model features may be appropriate to reduce the number of features for Model. The model feature XLR-V8 shows to be an extreme feature with the highest mean and largest range. A review of the box plot for Trim shows 26 different features. Regarding sale's price, some of the feature ranges and means are similar and could be grouped to represent the same information. The Hard-Top Conv 2D shows to be an extreme feature near $60,000 sales price. A review of the box plot for Type shows similarities between Coupe, Hatchback, Sedan, and Wagon mean value for price. Convertible appears to have a separate spread of price with the median value of $40,000. Extreme values appear in Convertible, Coupe, and Sedan. A review of the LOESS plots shows that a smoother parameter of .1 produced the best results and the lowest RSS for the data points. This is rational since the spread of the data points are widespread in the model as shows on Table pages 13 plots of Price verse Mileage. When reviewing the observation for price and since the average for price for the cars are in the $20,000 price range, the top 24 observation of price range from $40,000 to $70,755 will need to be scrutinized as possible outliers.

Specification of the Model

After a review of the numerical and categorical variables, a transformation is needed to condense features that produce the similar information. The grouping of features was created based on similar Price means and ranges using the box plots and summary statistics of the predictors. The Make of a car has been condensed to three groups:
Make
1: Buick, Chevrolet, Pontiac 2: Cadillac 3: SAAB
The predictor Trim has been condensed into three groups with similar means as follows:

Trim
1: LS Coupe 2D, LS Coupe 2D, LS Sport Coupe 2D, LS Sedan 4D, LT Hatchback 4D, Coupe 2D, GT Coupe 2D, LS Hatchback 4D, LT Sedan 4D, CX Sedan 4D, SVM Sedan 4D, SVM Hatchback 4D, LS MAXX Hback 4D, LT MAXX Hback 4D, MAXX Hback 4D, Custom Sedan 4D, LS Sport Sedan 4D, SE Sedan 4D, GT Sedan 4D, GT Sport-wagon, Sport-wagon 4D, AWD Sport-wagon 4D

2: SS Coupe 2D, Linear Wagon 4D, Coupe 2D, LT Coupe 2D, Arc Sedan 4D, GXP Sedan 4D, SLE Sedan 4D, CXL Sedan 4D, Aero Sedan 4D, Aero Wagon 4D, CXS Sedan 4D, GT Sedan 4D, GTP Sedan 4D, Limited Sedan 4D, Linear Sedan 4D, Linear Wagon 4D, Sedan 4D, Special Ed Ultra 4D

3: Aero Conv 2D, Arc Conv 2D, Arc Sedan 4D, Arc Wagon 4D, Conv 2D, Coupe 2D, DHS Sedan 4D, Hardtop Conv 2D, Linear Conv 2D

The predictor Model has been condensed into three groups with similar means as follows:

Model
1: AVEO, Vibe, Cavalier, Century, Classic, Cobalt, Grand Am, Grand Prix, Lesabre, Malibu, Monte Carlo, SunFire
2: 9-2XAWD, 9_3, 9_3 HO, 9_5, 9_5 HO, Bonneville, CTS, G6, GTO, Impala, Lacrosse, Park Avenue
3: CST-V, Corvette, Deville, STS-V6, STS-V8, XLR-V8

The predictor Liter has been condensed into two groups with similar means as follows:

Liter
1: 1.6, 1.8, 2.2, 3.1, 3.4, 3.5, 3.8
2: 2, 2.3, 2.5, 2.8, 3.6, 4.6, 5.7, 6

After the transformation of the predicted variables, the initial model will be y = $b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + b_{10}x_{10} + b_{11}x_{11} + b_{12}x_{12} + b_{13}x_{13} + b_{14}x_{14} + b_{15}x_{15} + b_{16}x_{16} + b_{17}x_{17} + b_{18}x_{18} + b_{19}x_{19} + b_{21}x_{21} + b_{22}x_{22} + b_{23}x_{23} + b_{24}x_{24} + b_{25}x_{25} + b_{26}x_{26} + b_{27}x_{27} + b_{28}x_{28} + \varepsilon$. The planned method is to use a generalized method for analysis. The model assumptions will be that the data following a normal distribution, errors are uncorrelated, and the data is normally distributed.

Estimation of the Appropriate Model

Using the initial model for the regression analysis, the resulting ANOVA table is shown in Table page 18. The results from the ANOVA table shows a sum of square of 69788417359, mean square value of 4105201021, F-Value of 600.95 and a p-value of <.0001. To assess the fit of the model, the hypothesis for the model $H_o: B_1. = \cdots B_i = 0$ versus $H_1: At\ least\ one\ B_i \neq 0$ will be used and the result is to reject $H_o: B_i = 0$ in favor of the alternative such that at least one of the coefficient is not equal to zero. Based on the resulting R^2value at .9337 and R^2_{Adj} value of .9321 the model's regressor can account for 93.37% of the variation in the data set. The parameter estimates shown on the Table on page 16 shows that the mean price for a car is $49,847. The default features selected for the model is selected as Make_3, Model_3, Trim_3, Type Wagon, Cylinder 8, four doors, no sound, and has leather interior. For each change the coefficients are decreasing the sales price of the car. The prediction equation in: $\hat{y}$ = 49847 -0.186044x_1 – 9776.185x_2 – 1458.50x_3 -10741x_5 – 7537.0400x_6 – 855.103x_8 + 1691.5433x_9 – 7966.7573x_{11} – 2791.937x_{12} − 4744.138x_{13}– 4098.73x_{14} − 9818.43x_{16}– 4569.693x_{17} − 29.0255x_{19} − 393.686x_{23} − 941.732x_{25} − 356.30x_{27}. Where each x_i is described in Table on page 18. The results from the t-value and p-value from the 16 shows predictors Mileage, Make_1, Make_2, Model_1, Model_2, Trim_2, Type Convertible, Type Coupe, Type Hatchback, Type Sedan, Type Wagon, Cylinder 6, Cylinder 8, and Sound 0 are significant to the model at an α = 0.05. A review of the Type III SS on page 22 shows that Make, and cylinder has the highest partial sum of squares. Due to this a reordering of predictor variables may be needed. In conclusion, there is evidence to suggest that the initial fitted model can explain the variation in the model at an α = 0.05. A concern in the model is the default intercept at $49,847 since mean price is $21,937. Possible leverage points could be the cause of this.

Assessment of the Chosen Prediction Equation

A review of the residual plots and normality plots is needed to further review the fit of the model. Based on the Normality plot (Q-Q plot) page 15 of table pages, about 86% of the data follow a normal distribution and the external values beyond $\pm$ 2 shows strong curvature in the data set. The result of this graph shows that model may not be able to predict data points on the low or high sales price well. This is also reflected in the predicted versus residual price graph as seen on page 19 of the table's pages. When reviewing the R Student residual for price on table page 16-17, we notice a handful observation having high standard deviation in the +2 through +6 range after sales price of $40,000. The R student residuals for Mileage shows a cone shape image were there could be a variance issue on table page 23. The other residuals show a binary plot on the 0 and 1 points with un unequal variance.

The correlation matrix on page 18 of the table pages shows the possibility of a multicorrelation issues with the following pairs: Type_convertible: Trim_3, Cylinder 6: Model 3, Liter 2:Make 1, Door 3: Type Coupe, Door 4: Type coupe, Liter 1: Make 1. The VIF is over 10 for the predictors Model_1 and Model_2 and Trim 1. My recommendation is to remove Model 1, then refit to check the VIF amount the other predictors, then check the correlation table. Model 1. Trim 1 and Make 1 was removed from the model. After correction of the correlation the refit of the model, there is still need for a transformation in price. A Box Cox analysis as seen on table page 20 shows the transform of to log(price) is needed based on the R student residual of price. A transformation and refit of the model is shown on table page 21-25. A refit of the model and check of correlation shows there is no correlation issues present. The refit of the residuals show that the variance of the errors has improved for all the predictors as seen in table pages 72. The Influential analysis using Cook's D at criteria 0.0054, DFFITS at criteria 0.254, DBEATTS at criteria 0.0733, and COVRATIO (1.0484, .9516) tests shows numerous high observations as can be seen in Table pages 109 - 110. To minimize the influence observation, a weight of 0.5 was added to the student residuals that were $\pm$ 1.5. The residual plot still show errors are normal. The ANOVA for this model has a F-value at 994.34 and a P-value< 0.0001. The R^2 is at 95.63$ R^2 predicted at 96% and the Coeff Var is at 1.05332 as seen on table pages 100. The model can account for 95.63% of the weighted variation in the data set. The new prediction equation is $\hat{y} = 10.31592 - 0.00000815x_1 + 0.26436x_3 + 0.16019x_6 + 0.40959x_7 + 0.03625x_8 - 0.4009x_{10} - 0.27786x_{24} + 0.15533x_{11} - 0.14228x_{23} - 0.1983x_{16} - 0.50579x_{17} - 0.1580718 - 0.0297x_{25} - 0.03298x_{27} - 0.0096x_{23}$

Selection of Variable

To complete the selection for variable, the lasso method was used to fit the model as listed on page 35 of table sheet. After the removal of nonsignificant predictors, the ANOVA table on page 39 has a F-value of 1039.85 and a p-value <0.0001. the R^ is 93.89%, coeff Var=1.02913 and the Root MSE is 0.10195. The VIF for the predictors are under 10 and the Eigen value ratio is 90. Both variables suggest low variance in the model. The PRESS statistic is at 94%, thus the maybe may be able to predict the weighted model at 94%. Since

the response variable price was transformed, the estimated model will be log(y) = $b_0 + b_1x_1 + b_3x_3 + b_4x_4 + b_6x_6 + b_7x_7 + b_9x_9 + b_{13}x_{13} + b_{11}x_{11} + b_{15}x_{15} + b_{27}x_{27} + b_{28}x_{28}+\in$ By the F test for H_o: $B\_1 \ldots B\_i = 0$ verses at least one B_i doesn't equal zero for fit, there is evidence to suggest that this model explains 93% of the variation in the log data of the model.

Conclusion

The final prediction equation is: $log(\hat{y})= 9.95457 - 0.00000805 \quad x_1 + 0.24334x_3 + 0.55226x_4 + 0.14223x_6 + 0.38715x_7 + 0.02953x_9 - 0.10023x_{13} + 0.30649x_{11} + 0.1664x_{15} - 0.34495x_{27} + 0.18026x_{28}$. Based on this prediction equation is adequate due to the F-test at p<0.0001, constant residual in error and variance. The signs of the coefficient are justified due to increase mileage would reduce the price, but advance feature would add value that was lost. In addition, if only the intercept was included in the model, the exponential value would be close to the mean of price in the dataset. The x_1 value is reasonably small given mileage for used cars tend to be six-digit value numbers. The recommendation that I would suggest for the model is continue weighting down or remove observations identified in the influential, DIFFTS, and DEAFBTS charts and collect more observation with high price value over $50,000 to better predict the extreme observation.

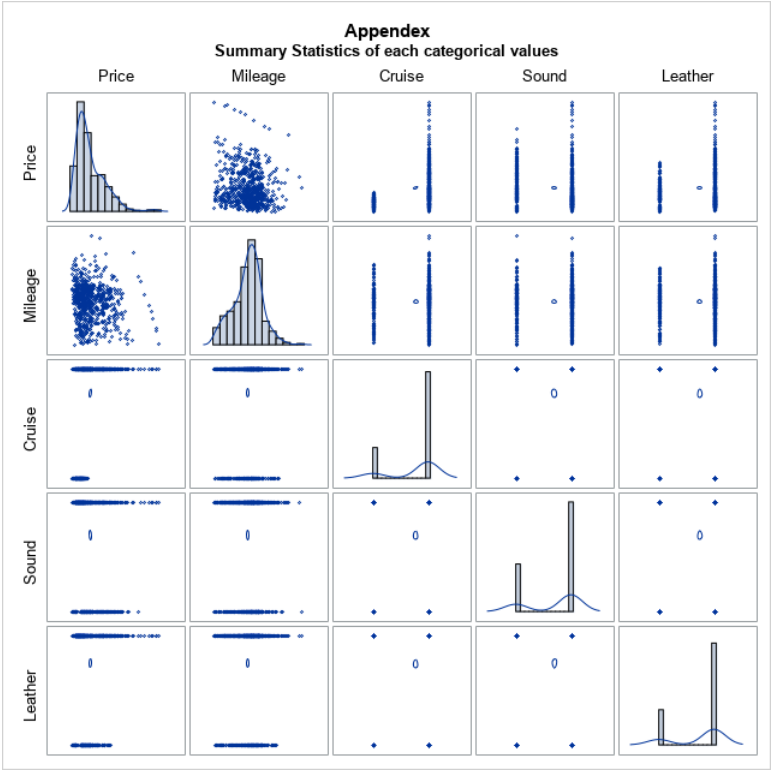
Appendix

Predictor x_i Description					
x_1	Mileage	x_{12}	Type Coupe	x_{23}	Cruise 0
x_2	Make_1	x_{13}	Type Hatchback	x_{24}	Cruise 1
x_3	Make_2	x_{14}	Type Sedan	x_{25}	Sound 0
x_4	Make_3	x_{15}	Type Wagon	x_{26}	Sound 1
x_5	Model_1	x_{16}	Cylinder 4	x_{27}	Leather 0
x_6	Model_2	x_{17}	Cylinder 6	x_{28}	Leather 1
x_7	Model_3	x_{18}	Cylinder 8		
x_8	Trim_1	x_{19}	Liter_1		
x_9	Trim_2	x_{20}	Liter_2		
x_{10}	Trim_3	x_{21}	Door 2		
x_{11}	Type Convertible	x_{22}	Door 4		

Appendex
Summary Statistics of each categorical values

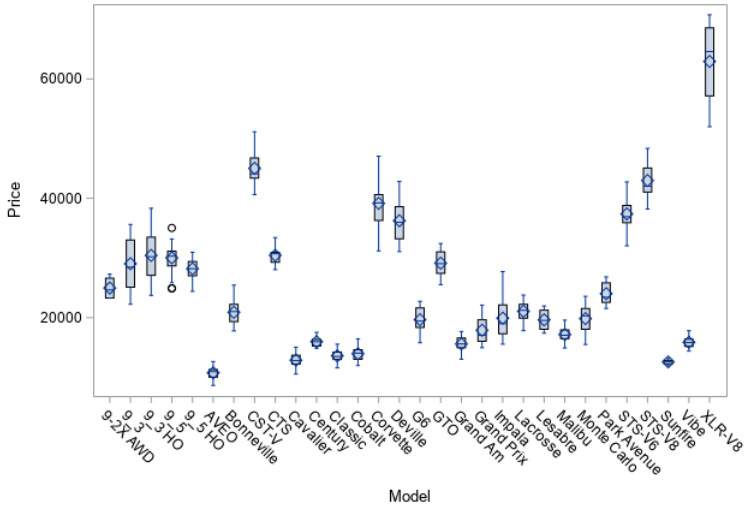
The MEANS Procedure

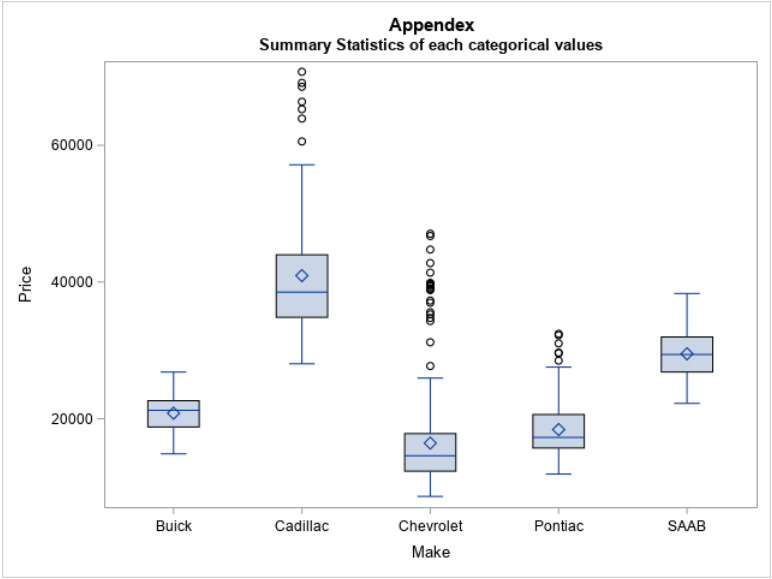
Variable	Label	N	N Miss	Mean	Std Dev	Median	Minimum	Maximum	Range
Price	Price	744	0	21937.042	10030.091	18931.944	8638.931	70755.467	62116.536
Mileage	Mileage	744	0	19791.304	8177.547	20913.500	266.000	50387.000	50121.000
Cruise	Cruise	744	0	0.777	0.417	1.000	0.000	1.000	1.000
Sound	Sound	744	0	0.698	0.460	1.000	0.000	1.000	1.000
Leather	Leather	744	0	0.743	0.437	1.000	0.000	1.000	1.000

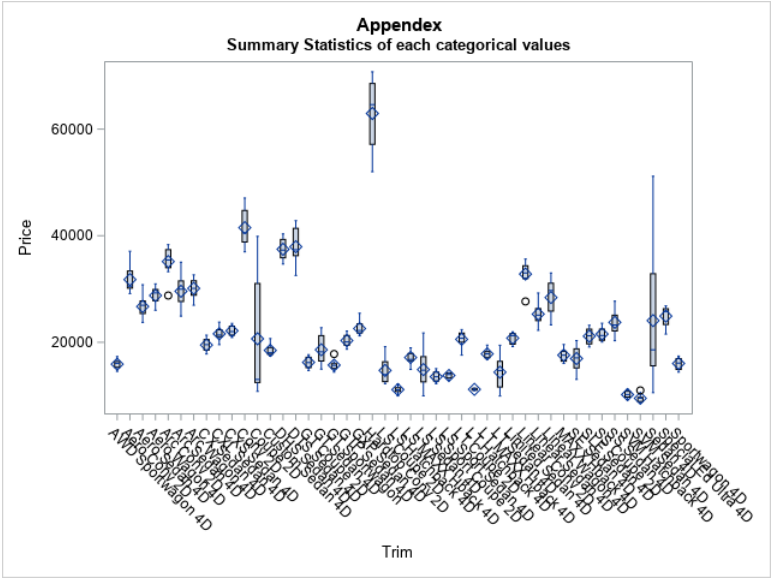

Appendex
Summary Statistics of each categorical values
Price
Mileage
Cruise
Sound
Leather
Price
Mileage
Cruise
Sound
Leather

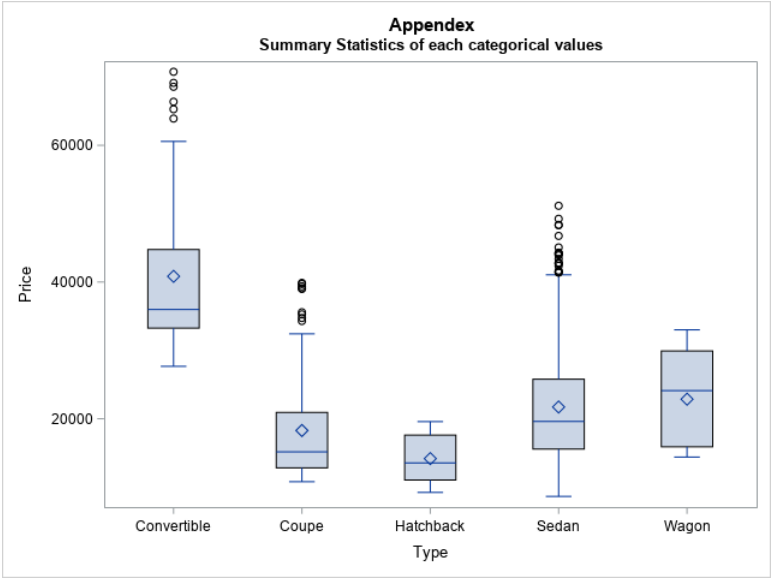
Appendex

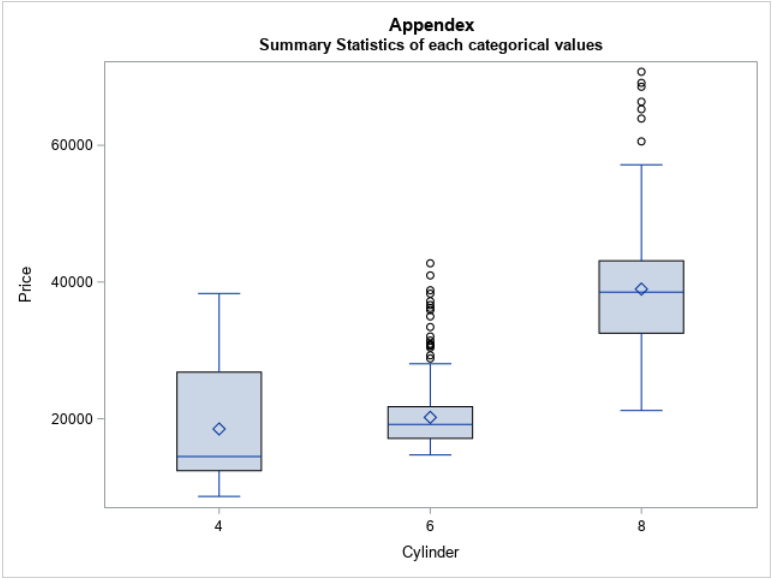
Summary Statistics of each categorical values

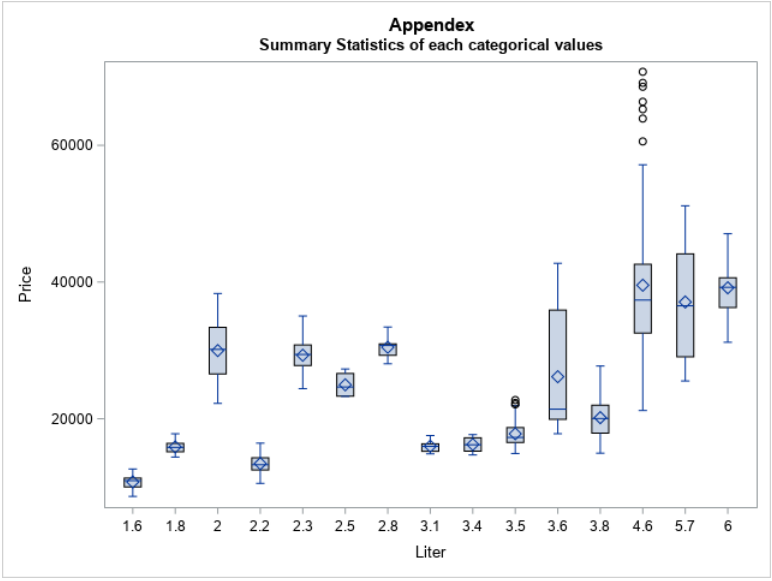


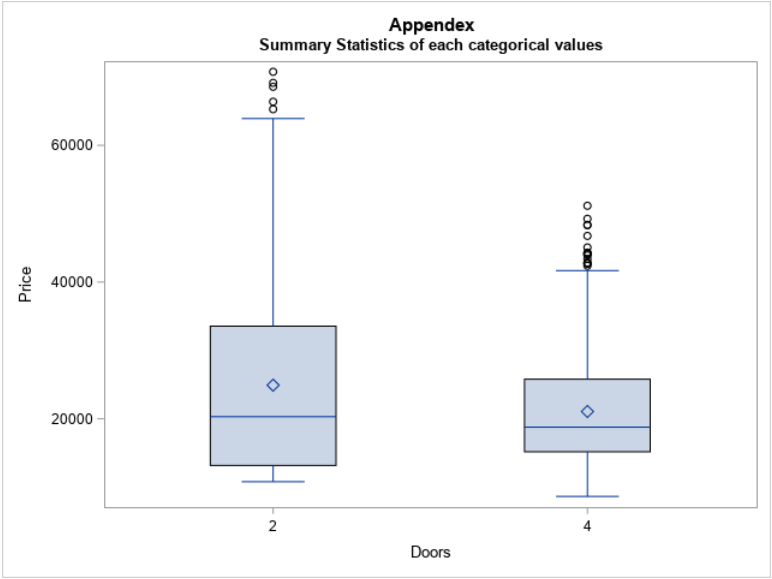

Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
Buick
Cadillac
Chevrolet
Pontiac
SAAB
Make

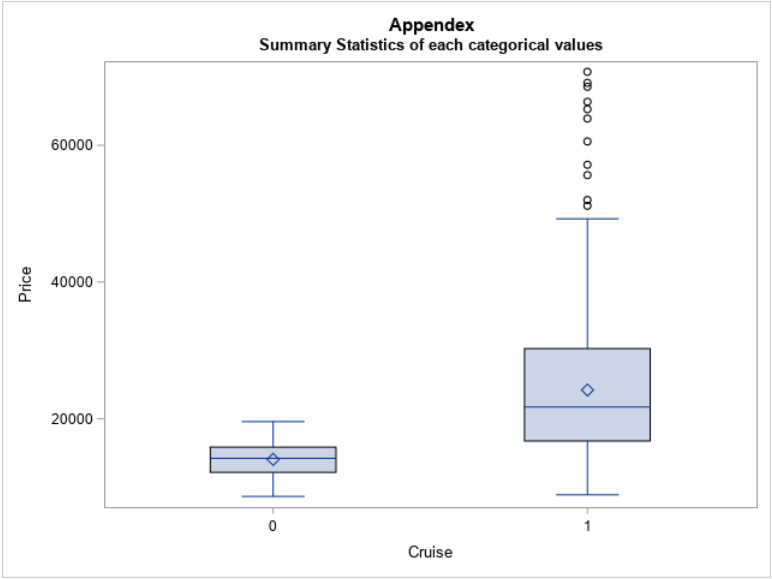

Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
Trim

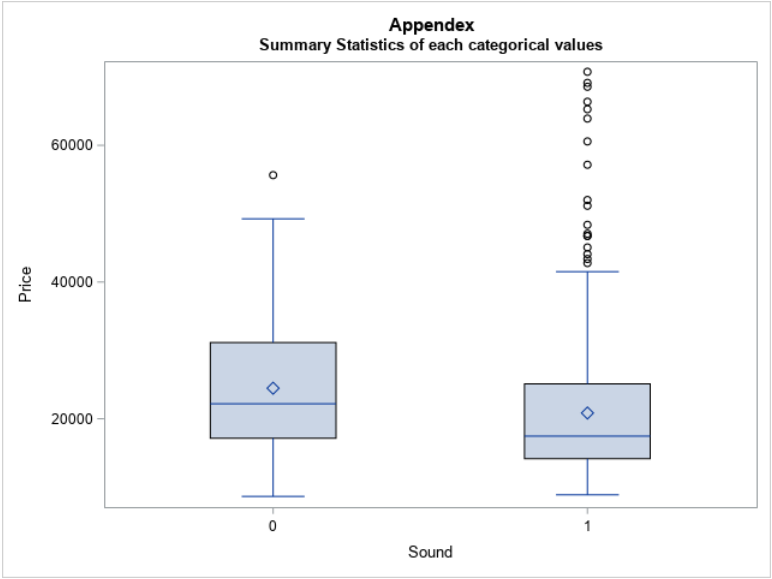

Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
Convertible
Coupe
Hatchback
Sedan
Wagon
Type

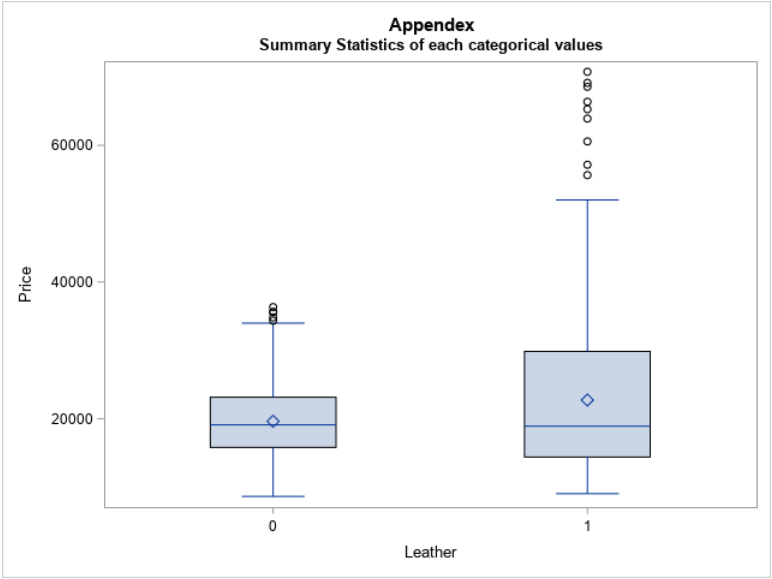

Appendex
Summary Statistics of each categorical values
60000
40000
20000
Price
4
6
8
Cylinder


Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
1.6
1.8
2
2.2
2.3
2.5
2.8
3.1
3.4
3.5
3.6
3.8
4.6
5.7
6
Liter


Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
2
4
Doors


Appendex
Summary Statistics of each categorical values
60000
40000
20000
Price
0
1
Cruise


Appendex
Summary Statistics of each categorical values
Price
60000
40000
20000
0
1
Sound


Appendex
Summary Statistics of each categorical values
60000
40000
20000
Price
0
1
Leather

Appendex
Loess plots

The LOESS Procedure
Smoothing Parameter: 0.1
Dependent Variable: Price

Fit Summary	
Fit Method	kd Tree
Blending	Linear
Number of Observations	744
Number of Fitting Points	65
kd Tree Bucket Size	14
Degree of Local Polynomials	1
Smoothing Parameter	0.10000
Points in Local Neighborhood	74
Residual Sum of Squares	70888311767

Appendex
Loess plots

The GLMSELECT Procedure
Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	69788417359	4105201021	600.95	<.0001
Error	726	4959406980	6831139		
Corrected Total	743	74747824339			

Root MSE	2613.64479
Dependent Mean	21937
R-Square	0.9337
Adj R-Sq	0.9321
AIC	12472
AICC	12473
PRESS	5255596799
SBC	11809

Parameter Estimates						
Parameter	DF	Estimate	Standardized Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	49847	0	969.680933	51.41	<.0001
Mileage	1	-0.186044	-0.151682	0.011794	-15.77	<.0001
Make_ 1	1	-9779.185048	-0.428350	676.574691	-14.45	<.0001
Make_ 2	1	-1458.050496	-0.045063	746.324132	-1.95	0.0511
Make_ 3	0	0	0	.	.	.
Model_ 1	1	-10741	-0.535184	778.710037	-13.79	<.0001
Model_ 2	1	-7537.040065	-0.359781	691.999008	-10.89	<.0001
Model_ 3	0	0	0	.	.	.
Trim_ 1	1	855.103374	0.042646	687.506849	1.24	0.2140
Trim_ 2	1	1691.543313	0.082504	648.359024	2.61	0.0093
Trim_ 3	0	0	0	.	.	.
Type Convertible	1	7966.757317	0.199003	734.086780	10.85	<.0001
Type Coupe	1	-2791.937906	-0.102448	464.646245	-6.01	<.0001
Type Hatchback	1	-4744.133800	-0.128877	527.128834	-9.00	<.0001
Type Sedan	1	-4098.731592	-0.199914	399.928521	-10.25	<.0001
Type Wagon	0	0	0	.	.	.
Cylinder 4	1	-9818.430068	-0.488389	679.581378	-14.45	<.0001
Cylinder 6	1	-4569.693380	-0.223642	608.750825	-7.51	<.0001
Cylinder 8	0	0	0	.	.	.
Liter_ 1	1	-29.025555	-0.001373	549.212461	-0.05	0.9579
Liter_ 2	0	0	0	.	.	.
Doors 2	0	0	0	.	.	.
Doors 4	0	0	0	.	.	.
Cruise 0	1	-393.686474	-0.016352	292.863530	-1.34	0.1793
Cruise 1	0	0	0	.	.	.
Sound 0	1	-941.732246	-0.043154	224.563923	-4.19	<.0001
Sound 1	0	0	0	.	.	.
Leather 0	1	-356.300280	-0.015528	240.965710	-1.48	0.1397
Leather 1	0	0	0	.	.	.

Appendex
Loess plots

The GLM Procedure

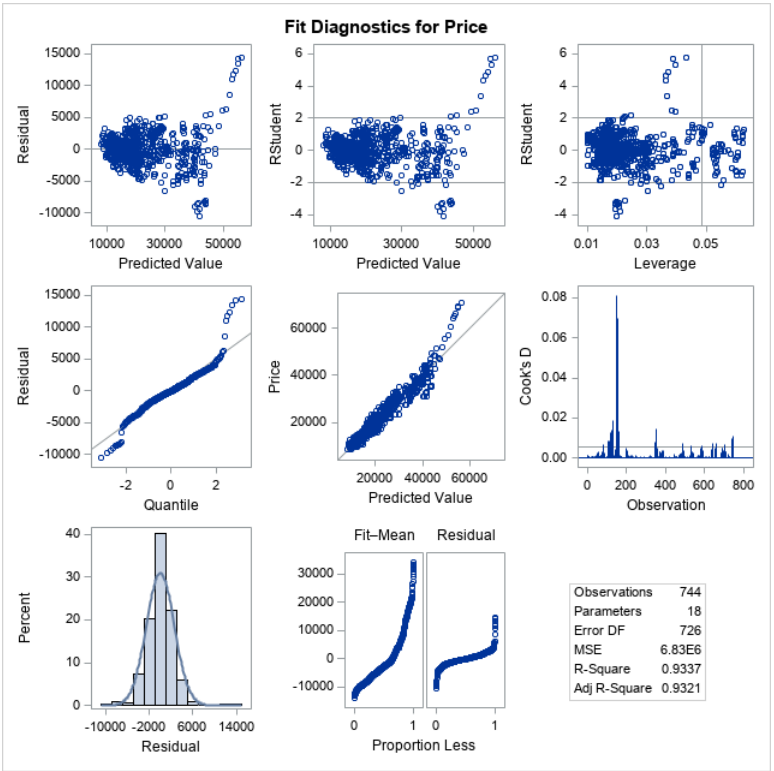
Dependent Variable: Price Price

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	69788417359	4105201021	600.95	<.0001
Error	726	4959406980	6831139		
Corrected Total	743	74747824339			

R-Square	Coeff Var	Root MSE	Price Mean
0.933651	11.91430	2613.645	21937.04

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Mileage	1	1426386946	1426386946	208.81	<.0001
Make_	2	45688196442	22844098221	3344.11	<.0001
Model_	2	15189967724	7594983862	1111.82	<.0001
Trim_	2	2574389990	1287194995	188.43	<.0001
Type	4	1937476006	484369002	70.91	<.0001
Cylinder	2	2806429546	1403214773	205.41	<.0001
Liter_	1	449066	449066	0.07	0.7977
Doors	0	0	.	.	.
Cruise	1	7416239	7416239	1.09	0.2978
Sound	1	142770065	142770065	20.90	<.0001
Leather	1	14935335	14935335	2.19	0.1397

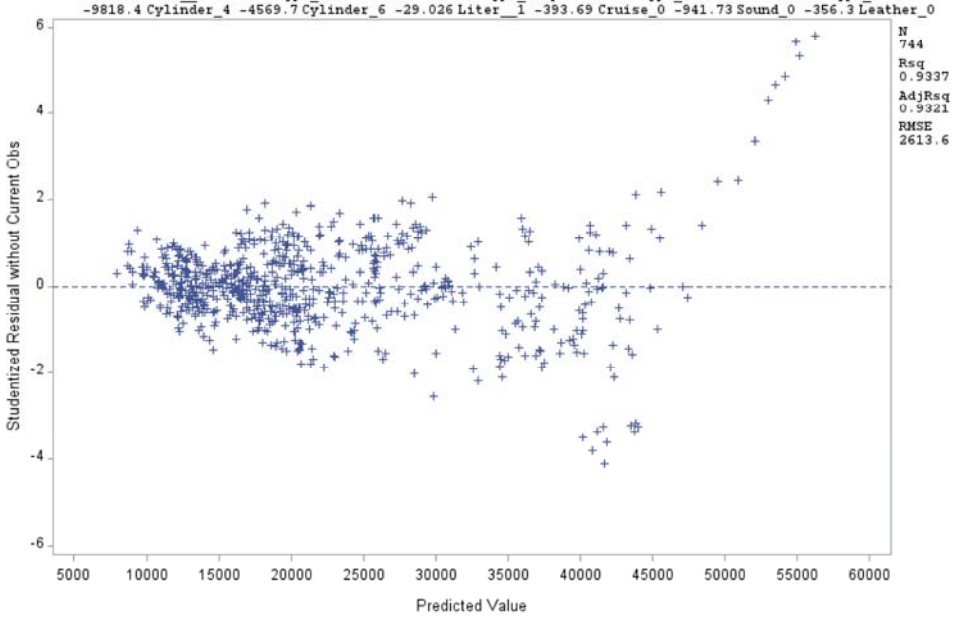
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mileage	1	1699834857	1699834857	248.84	<.0001
Make_	2	2080888460	1040444230	152.31	<.0001
Model_	2	1356656685	678328342	99.30	<.0001
Trim_	2	109253461	54626730	8.00	0.0004
Type	3	1693954016	564651339	82.66	<.0001
Cylinder	2	2488100722	1244050361	182.11	<.0001
Liter_	1	19080	19080	0.00	0.9579
Doors	0	0	.	.	.
Cruise	1	12344217	12344217	1.81	0.1793
Sound	1	120134581	120134581	17.59	<.0001
Leather	1	14935335	14935335	2.19	0.1397



Appendex

Loess plots

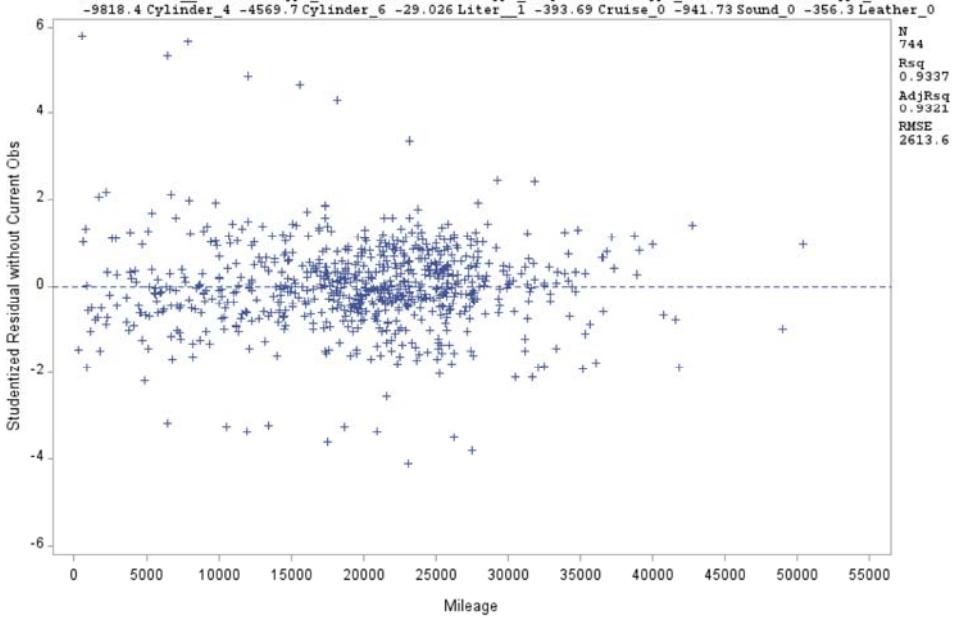
Price = 49847 -0.186 Mileage -9779.2 Make__1 -1458.1 Make__2 -10741 Model__1 -7537 Model__2 +855.1 Trim__1 +1691.5 Trim__2 +7966.8 Type_Convertible -2791.9 Type_Coupe -4744.1 Type_Hatchback -4098.7 Type_Sedan -9818.4 Cylinder_4 -4569.7 Cylinder_6 -29.026 Liter__1 -393.69 Cruise_0 -941.73 Sound_0 -356.3 Leather_0



Appendex

Loess plots

Price = 49847 -0.186 Mileage -9779.2 Make__1 -1458.1 Make__2 -10741 Model__1 -7537 Model__2 +855.1 Trim__1 +1691.5 Trim__2 +7966.8 Type_Convertible -2791.9 Type_Coupe -4744.1 Type_Hatchback -4098.7 Type_Sedan -9818.4 Cylinder_4 -4569.7 Cylinder_6 -29.026 Liter__1 -393.69 Cruise_0 -941.73 Sound_0 -356.3 Leather_0



Appendex
Loess plots

The CORR Procedure

28 Variables:	Mileage Make__1 Make__2 Make__3 Model__1 Model__2 Model__3 Trim__1 Trim__2 Trim__3 Type_Convertible Type_Coupe Type_Hatchback Type_Sedan Type_Wagon Cylinder_4 Cylinder_6 Cylinder_8 Liter__1 Liter__2 Doors_2 Doors_4 Cruise_0 Cruise_1 Sound_0 Sound_1 Leather_0 Leather_1

Simple Statistics

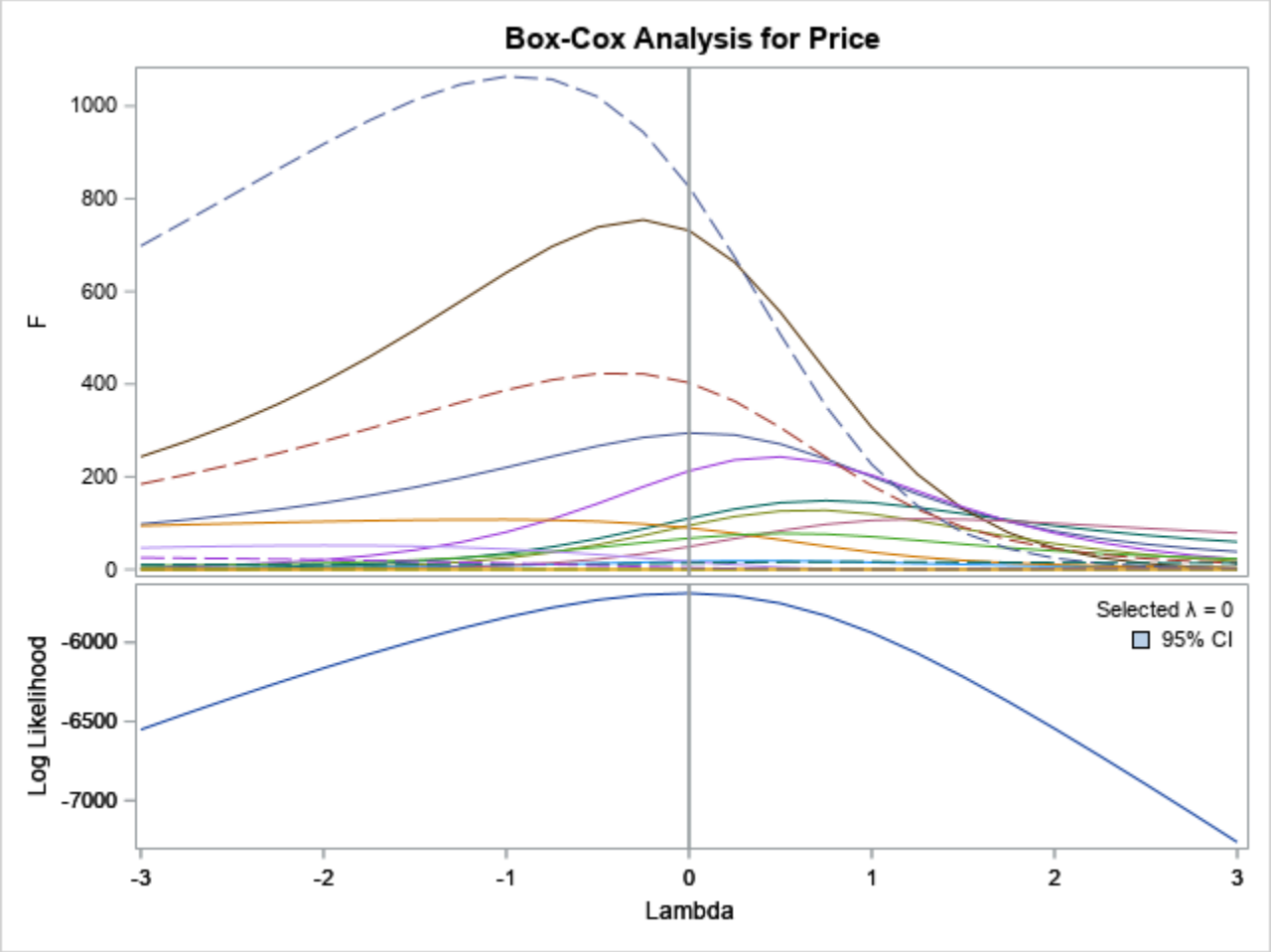
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Mileage	744	19791	8178	14724730	266.00000	50387	Mileage
Make__1	744	0.73925	0.43934	550.00000	0	1.00000	Make_ 1
Make__2	744	0.10753	0.30999	80.00000	0	1.00000	Make_ 2
Make__3	744	0.15323	0.36045	114.00000	0	1.00000	Make_ 3
Model__1	744	0.52419	0.49975	390.00000	0	1.00000	Model_ 1
Model__2	744	0.35484	0.47879	264.00000	0	1.00000	Model_ 2
Model__3	744	0.12097	0.32631	90.00000	0	1.00000	Model_ 3
Trim__1	744	0.51075	0.50022	380.00000	0	1.00000	Trim_ 1
Trim__2	744	0.39516	0.48921	294.00000	0	1.00000	Trim_ 2
Trim__3	744	0.09409	0.29214	70.00000	0	1.00000	Trim_ 3
Type_Convertible	744	0.06720	0.25054	50.00000	0	1.00000	Type Convertible
Type_Coupe	744	0.16129	0.36805	120.00000	0	1.00000	Type Coupe
Type_Hatchback	744	0.08065	0.27247	60.00000	0	1.00000	Type Hatchback
Type_Sedan	744	0.60484	0.48921	450.00000	0	1.00000	Type Sedan
Type_Wagon	744	0.08602	0.28058	64.00000	0	1.00000	Type Wagon
Cylinder_4	744	0.46237	0.49892	344.00000	0	1.00000	Cylinder 4
Cylinder_6	744	0.40323	0.49088	300.00000	0	1.00000	Cylinder 6
Cylinder_8	744	0.13441	0.34132	100.00000	0	1.00000	Cylinder 8
Liter__1	744	0.65860	0.47450	490.00000	0	1.00000	Liter_ 1
Liter__2	744	0.34140	0.47450	254.00000	0	1.00000	Liter_ 2
Doors_2	744	0.22849	0.42015	170.00000	0	1.00000	Doors 2
Doors_4	744	0.77151	0.42015	574.00000	0	1.00000	Doors 4
Cruise_0	744	0.22312	0.41662	166.00000	0	1.00000	Cruise 0
Cruise_1	744	0.77688	0.41662	578.00000	0	1.00000	Cruise 1
Sound_0	744	0.30242	0.45961	225.00000	0	1.00000	Sound 0
Sound_1	744	0.69758	0.45961	519.00000	0	1.00000	Sound 1
Leather_0	744	0.25672	0.43712	191.00000	0	1.00000	Leather 0
Leather_1	744	0.74328	0.43712	553.00000	0	1.00000	Leather 1

Pearson Correlation Coefficients, N = 744
Prob > |r| under H0: Rho=0

	Mileage	Make__1	Make__2	Make__3	Model__1	Model__2	Model__3	Trim__1	Trim__2	Trim__3	Type_Convertible	Type_Coupe	Type_Hatchback	Type_Sedan	Type_Wagon	Cylind
Mileage Mileage	1.00000	-0.02363 0.5198	-0.03749 0.3071	0.06105 0.0961	-0.03230 0.3790	0.05627 0.1252	-0.03310 0.3673	-0.03463 0.3455	0.00905 0.8053	0.04414 0.2291	0.03000 0.4139	-0.00787 0.8302	-0.02538 0.4893	-0.01238 0.7360	0.02977 0.4174	0.0 0.
Make__1 Make_ 1	-0.02363 0.5198	1.00000	-0.58444 <.0001	-0.71625 <.0001	0.62338 <.0001	-0.35294 <.0001	-0.43685 <.0001	0.54558 <.0001	-0.29643 <.0001	-0.43776 <.0001	-0.32967 <.0001	0.26045 <.0001	0.17590 <.0001	-0.01666 0.6500	-0.18901 <.0001	-0.1 <.
Make__2 Make_ 2	-0.03749 0.3071	-0.58444 <.0001	1.00000	-0.14765 <.0001	-0.36433 <.0001	-0.16674 <.0001	0.80263 <.0001	-0.26786 <.0001	0.16318 <.0001	0.18537 <.0001	0.08012 0.0289	-0.15222 <.0001	-0.10280 0.0050	0.19181 <.0001	-0.10649 0.0036	-0.3 <.
Make__3 Make_ 3	0.06105 0.0961	-0.71625 <.0001	-0.14765 <.0001	1.00000	-0.44649 <.0001	0.57359 <.0001	-0.15780 <.0001	-0.43463 <.0001	0.22097 <.0001	0.37416 <.0001	0.33292 <.0001	-0.18654 <.0001	-0.12599 0.0006	-0.14465 <.0001	0.32196 <.0001	0.4 <.
Model__1 Model_ 1	-0.03230 0.3790	0.62338 <.0001	-0.36433 <.0001	-0.44649 <.0001	1.00000	-0.77842 <.0001	-0.38937 <.0001	0.48889 <.0001	-0.29789 <.0001	-0.33826 <.0001	-0.28173 <.0001	0.27145 <.0001	0.28217 <.0001	-0.19756 <.0001	-0.03406 0.3536	0.2 <.
Model__2 Model_ 2	0.05627 0.1252	-0.35294 <.0001	-0.16674 <.0001	0.57359 <.0001	-0.77842 <.0001	1.00000	-0.27511 <.0001	-0.36437 <.0001	0.28545 <.0001	0.14588 <.0001	0.13753 0.0002	-0.24884 <.0001	-0.21965 <.0001	0.17424 <.0001	0.11311 0.0020	-0.0 0.
Model__3 Model_ 3	-0.03310 0.3673	-0.43685 <.0001	0.80263 <.0001	-0.15780 <.0001	-0.38937 <.0001	-0.27511 <.0001	1.00000	-0.21412 <.0001	0.03740 0.3084	0.30400 <.0001	0.22968 <.0001	-0.05061 0.1679	-0.10987 0.0027	0.04691 0.2012	-0.11381 0.0019	-0.3 <.
Trim__1 Trim_ 1	-0.03463 0.3455	0.54558 <.0001	-0.26786 <.0001	-0.43463 <.0001	0.48889 <.0001	-0.36437 <.0001	-0.21412 <.0001	1.00000	-0.82586 <.0001	-0.32928 <.0001	-0.27425 <.0001	0.28299 <.0001	0.28987 <.0001	-0.21911 <.0001	-0.02578 0.4826	0.1 0.
Trim__2 Trim_ 2	0.00905 0.8053	-0.29643 <.0001	0.16318 <.0001	0.22097 <.0001	-0.29789 <.0001	0.28545 <.0001	0.03740 0.3084	-0.82586 <.0001	1.00000	-0.26049 <.0001	-0.21696 <.0001	-0.20496 <.0001	-0.23939 <.0001	0.40590 <.0001	-0.01265 0.7305	-0.1 <.
Trim__3 Trim_ 3	0.04414 0.2291	-0.43776 <.0001	0.18537 <.0001	0.37416 <.0001	-0.33826 <.0001	0.14588 <.0001	0.30400 <.0001	-0.32928 <.0001	-0.26049 <.0001	1.00000	0.83289 <.0001	-0.14132 0.0001	-0.09545 0.0092	-0.30453 <.0001	0.06532 0.0750	0.0 0.
Type_Convertible Type Convertible	0.03000 0.4139	-0.32967 <.0001	0.08012 0.0289	0.33292 <.0001	-0.28173 <.0001	0.13753 0.0002	0.22968 <.0001	-0.27425 <.0001	-0.21696 <.0001	0.83289 <.0001	1.00000	-0.11771 0.0013	-0.07950 0.0301	-0.33208 <.0001	-0.08235 0.0247	0.0 0.
Type_Coupe Type Coupe	-0.00787 0.8302	0.26045 <.0001	-0.15222 <.0001	-0.18654 <.0001	0.27145 <.0001	-0.24884 <.0001	-0.05061 0.1679	0.28299 <.0001	-0.20496 <.0001	-0.14132 0.0001	-0.11771 0.0013	1.00000	-0.12988 0.0004	-0.54254 <.0001	-0.13453 0.0002	0.0 0.
Type_Hatchback Type Hatchback	-0.02538 0.4893	0.17590 <.0001	-0.10280 0.0050	-0.12599 0.0006	0.28217 <.0001	-0.21965 <.0001	-0.10987 0.0027	0.28987 <.0001	-0.23939 <.0001	-0.09545 0.0092	-0.07950 0.0301	-0.12988 0.0004	1.00000	-0.36642 <.0001	-0.09086 0.0132	0.0 0.
Type_Sedan Type Sedan	-0.01238 0.7360	-0.01666 0.6500	0.19181 <.0001	-0.14465 <.0001	-0.19756 <.0001	0.17424 <.0001	0.04691 0.2012	-0.21911 <.0001	0.40590 <.0001	-0.30453 <.0001	-0.33208 <.0001	-0.54254 <.0001	-0.36642 <.0001	1.00000	-0.37955 <.0001	-0.2 <.
Type_Wagon Type Wagon	0.02977 0.4174	-0.18901 <.0001	-0.10649 0.0036	0.32196 <.0001	-0.03406 0.3536	0.11311 0.0020	-0.11381 0.0019	-0.02578 0.4826	-0.01265 0.7305	0.06532 0.0750	-0.08235 0.0247	-0.13453 0.0002	-0.09086 0.0132	-0.37955 <.0001	1.00000	0.3 <.
Cylinder_4 Cylinder 4	0.02001 0.5858	-0.14921 <.0001	-0.32189 <.0001	0.45870 <.0001	0.26816 <.0001	-0.04544 0.2157	-0.34402 <.0001	0.13105 0.0003	-0.17610 <.0001	0.07050 0.0546	0.07410 0.0433	0.03310 0.3673	0.02236 0.5426	-0.26504 <.0001	0.33082 <.0001	1.0
Cylinder_6 Cylinder 6	-0.01309 0.7214	0.36337 <.0001	-0.10842 0.0031	-0.34966 <.0001	0.01504 0.6821	0.13485 0.0002	-0.22091 <.0001	-0.01768 0.6302	0.17627 <.0001	-0.26490 <.0001	-0.22064 <.0001	-0.06248 0.0886	0.05843 0.1113	0.27209 <.0001	-0.25218 <.0001	-0.7 <.

Cylinder_8 **Cylinder 8**	-0.01041 0.7767	-0.30448 <.0001	0.62645 <.0001	-0.16763 <.0001	-0.41361 <.0001	-0.12752 0.0005	0.82056 <.0001	-0.16613 <.0001	0.00390 0.9154	0.27793 <.0001	0.20900 <.0001	0.04147 0.2586	-0.11671 0.0014	-0.00390 0.9154	-0.12089 0.0010	-0.3 <.
Liter__1 **Liter_ 1**	-0.02904 0.4290	0.82490 <.0001	-0.48211 <.0001	-0.59083 <.0001	0.75570 <.0001	-0.43763 <.0001	-0.51524 <.0001	0.45211 <.0001	-0.19498 <.0001	-0.44761 <.0001	-0.37281 <.0001	0.16159 <.0001	0.21324 <.0001	0.02104 0.5666	-0.12283 0.0008	0.0 0.
Liter__2 **Liter_ 2**	0.02904 0.4290	-0.82490 <.0001	0.48211 <.0001	0.59083 <.0001	-0.75570 <.0001	0.43763 <.0001	0.51524 <.0001	-0.45211 <.0001	0.19498 <.0001	0.44761 <.0001	0.37281 <.0001	-0.16159 <.0001	-0.21324 <.0001	-0.02104 0.5666	0.12283 0.0008	-0.0 0.
Doors_2 **Doors 2**	0.01099 0.7647	0.03156 0.3901	-0.08556 0.0196	0.03512 0.3388	0.06979 0.0571	-0.13597 0.0002	0.09263 0.0115	0.08435 0.0214	-0.30892 <.0001	0.37287 <.0001	0.49322 <.0001	0.80580 <.0001	-0.16118 <.0001	-0.67329 <.0001	-0.16696 <.0001	0.0 0.
Doors_4 **Doors 4**	-0.01099 0.7647	-0.03156 0.3901	0.08556 0.0196	-0.03512 0.3388	-0.06979 0.0571	0.13597 0.0002	-0.09263 0.0115	-0.08435 0.0214	0.30892 <.0001	-0.37287 <.0001	-0.49322 <.0001	-0.80580 <.0001	0.16118 <.0001	0.67329 <.0001	0.16696 <.0001	-0.0 0.
Cruise_0 **Cruise 0**	-0.05734 0.1181	0.31828 <.0001	-0.18602 <.0001	-0.22797 <.0001	0.51057 <.0001	-0.39744 <.0001	-0.19880 <.0001	0.39534 <.0001	-0.30110 <.0001	-0.17271 <.0001	-0.14385 <.0001	0.01954 0.5947	0.30368 <.0001	-0.14794 <.0001	0.06586 0.0726	0.3 <.
Cruise_1 **Cruise 1**	0.05734 0.1181	-0.31828 <.0001	0.18602 <.0001	0.22797 <.0001	-0.51057 <.0001	0.39744 <.0001	0.19880 <.0001	-0.39534 <.0001	0.30110 <.0001	0.17271 <.0001	0.14385 <.0001	-0.01954 0.5947	-0.30368 <.0001	0.14794 <.0001	-0.06586 0.0726	-0.3 <.
Sound_0 **Sound 0**	0.03458 0.3463	-0.16883 <.0001	0.11153 0.0023	0.10987 0.0027	-0.24577 <.0001	0.16001 <.0001	0.14163 0.0001	-0.12246 0.0008	0.05440 0.1382	0.11858 0.0012	0.05703 0.1202	-0.13757 0.0002	-0.06604 0.0718	0.01743 0.6351	0.16328 <.0001	-0.0 0.
Sound_1 **Sound 1**	-0.03458 0.3463	0.16883 <.0001	-0.11153 0.0023	-0.10987 0.0027	0.24577 <.0001	-0.16001 <.0001	-0.14163 0.0001	0.12246 0.0008	-0.05440 0.1382	-0.11858 0.0012	-0.05703 0.1202	0.13757 0.0002	0.06604 0.0718	-0.01743 0.6351	-0.16328 <.0001	0.0 0.
Leather_0 **Leather 0**	-0.01638 0.6555	0.13178 0.0003	-0.20399 <.0001	0.01481 0.6867	0.01158 0.7526	0.13650 0.0002	-0.21802 <.0001	-0.00956 0.7945	0.04106 0.2633	-0.05239 0.1534	0.00202 0.9562	-0.09877 0.0070	-0.08366 0.0225	0.10999 0.0027	0.01723 0.6390	-0.0 0.
Leather_1 **Leather 1**	0.01638 0.6555	-0.13178 0.0003	0.20399 <.0001	-0.01481 0.6867	-0.01158 0.7526	-0.13650 0.0002	0.21802 <.0001	0.00956 0.7945	-0.04106 0.2633	0.05239 0.1534	-0.00202 0.9562	0.09877 0.0070	0.08366 0.0225	-0.10999 0.0027	-0.01723 0.6390	0.0 0.

The TRANSREG Procedure



Warning: Less than full rank model. The coefficients are not unique.

The coefficient for Identity(Make__1) was set to zero due to a linear dependency among the transformed variables.	
Identity(Make__1) =	1.0000 - Identity(Make__2) - Identity(Make__3)

The coefficient for Identity(Cylinder_4) was set to zero due to a linear dependency among the transformed variables.	
Identity(Cylinder_4) =	1.0000 - Identity(Cylinder_6) - Identity(Cylinder_8)

The coefficient for Identity(Sound_1) was set to zero due to a linear dependency among the transformed variables.	
Identity(Sound_1) =	1.0000 - Identity(Sound_0)

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: log_Price

Number of Observations Read	744
Number of Observations Used	744

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	119.52499	7.47031	778.50	<.0001
Error	727	6.97611	0.00960		
Corrected Total	743	126.50109			

Root MSE	0.09796	R-Square	0.9449
Dependent Mean	9.90663	Adj R-Sq	0.9436
Coeff Var	0.98881		

Note: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

Note: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

Type_Wagon =	Intercept - Type_Hatchback - Type_Convertible - Type_Coupe - Type_Sedan
Cylinder_8 =	Intercept - Cylinder_4 - Cylinder_6
Sound_1 =	Intercept - Sound_0
Leather_1 =	Intercept - Leather_0
Cruise_1 =	Intercept - Cruise_0

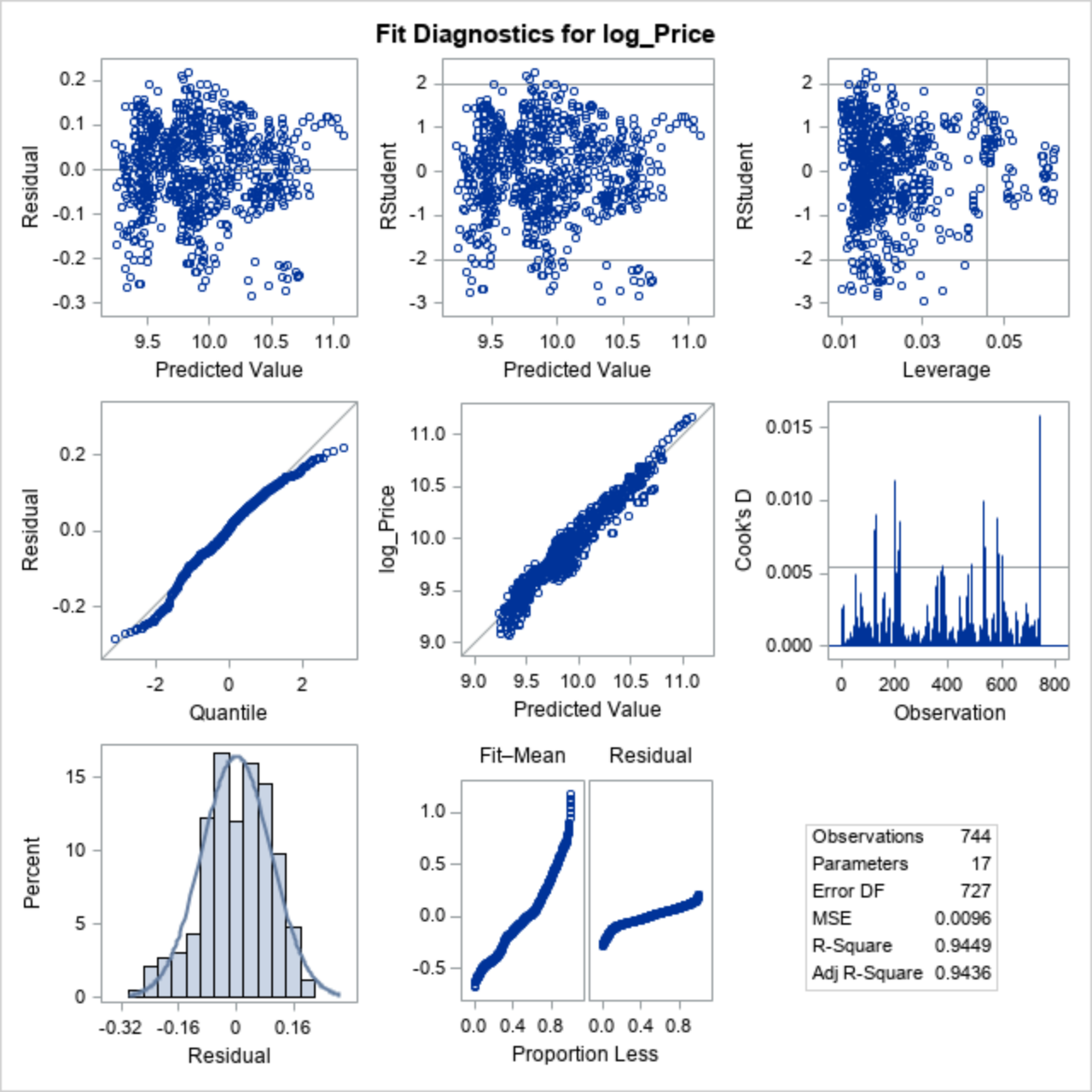
Parameter Estimates

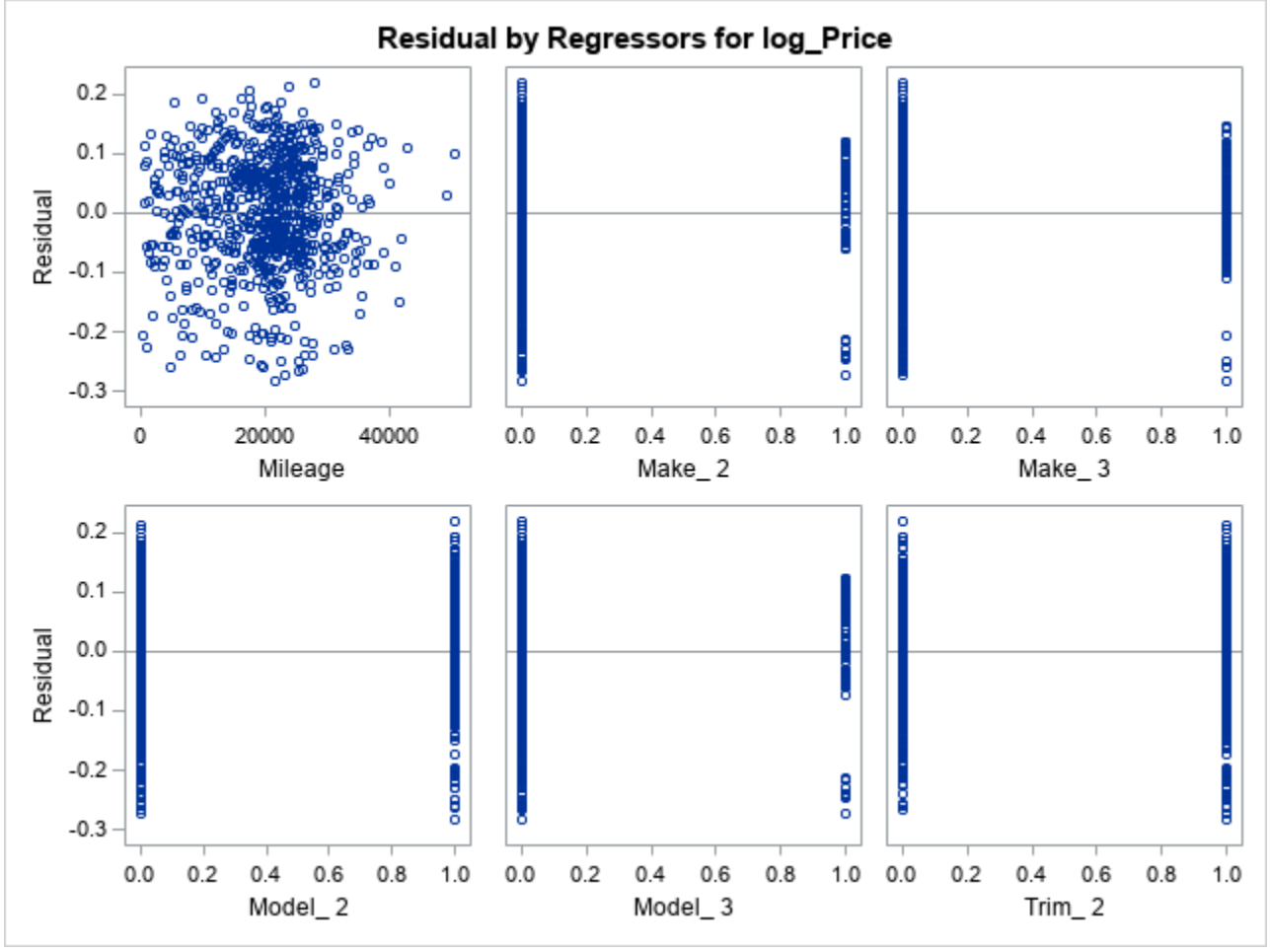
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	B	10.30637	0.02879	358.04	<.0001	0
Mileage	Mileage	1	-0.00000805	4.420281E-7	-18.21	<.0001	1.01171
Make__2	Make_ 2	1	0.25573	0.02091	12.23	<.0001	3.25291
Make__3	Make_ 3	1	0.54248	0.01894	28.64	<.0001	3.60943
Model__2	Model_ 2	1	0.16105	0.01253	12.86	<.0001	2.78565
Model__3	Model_ 3	1	0.40442	0.02879	14.05	<.0001	6.83127
Trim__2	Trim_ 2	1	0.03496	0.00980	3.57	0.0004	1.77809
Trim__3	Trim_ 3	1	-0.02526	0.02560	-0.99	0.3241	4.32935
Type_Hatchback	Type Hatchback	B	-0.27511	0.01974	-13.93	<.0001	2.24106
Type_Convertible	Type Convertible	B	0.15235	0.02746	5.55	<.0001	3.66381
Type_Coupe	Type Coupe	B	-0.13945	0.01741	-8.01	<.0001	3.18006

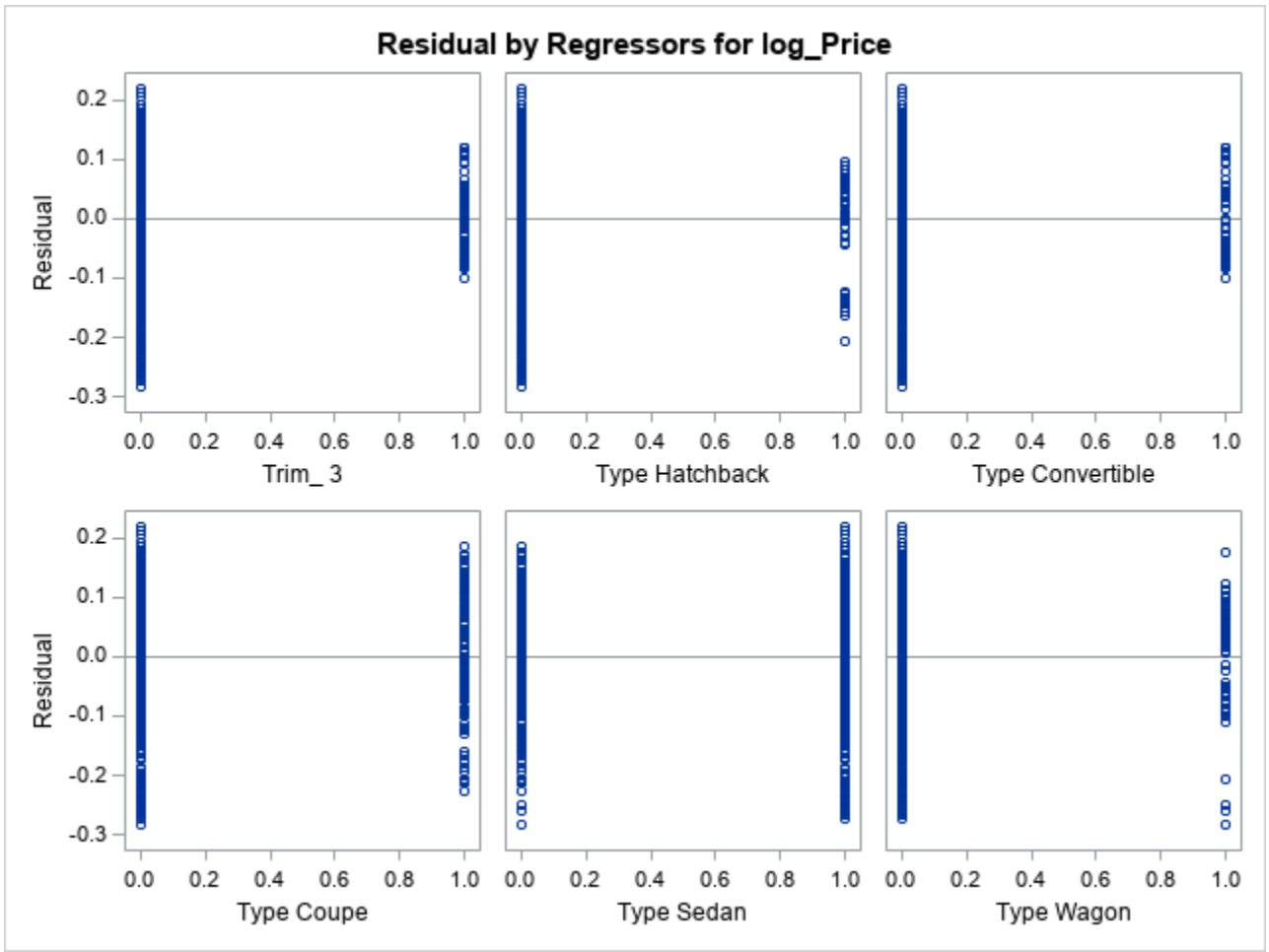
Type_Sedan	Type Sedan	B	-0.19568	0.01498	-13.06	<.0001	4.16099
Type_Wagon	Type Wagon	0	0	.	.	.	.
Cylinder_4	Cylinder 4	B	-0.50338	0.02266	-22.21	<.0001	9.89838
Cylinder_6	Cylinder 6	B	-0.15586	0.01996	-7.81	<.0001	7.43052
Cylinder_8	Cylinder 8	0	0	.	.	.	.
Sound_0	Sound 0	B	-0.03113	0.00842	-3.70	0.0002	1.15828
Sound_1	Sound 1	0	0	.	.	.	.
Leather_0	Leather 0	B	-0.02891	0.00902	-3.20	0.0014	1.20433
Leather_1	Leather 1	0	0	.	.	.	.
Cruise_0	Cruise 0	B	-0.01479	0.01098	-1.35	0.1783	1.61885
Cruise_1	Cruise 1	0	0	.	.	.	.

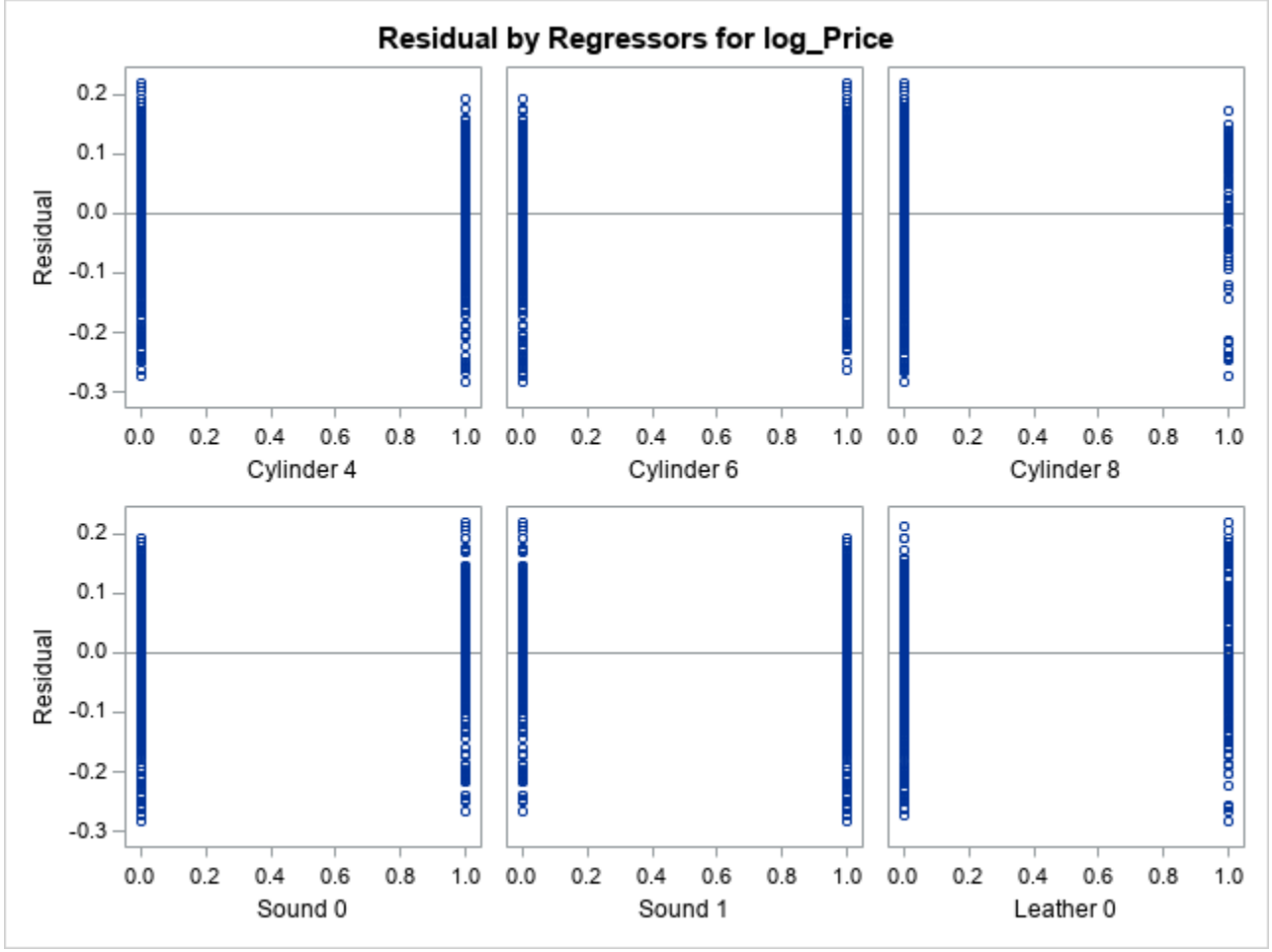
The SAS System

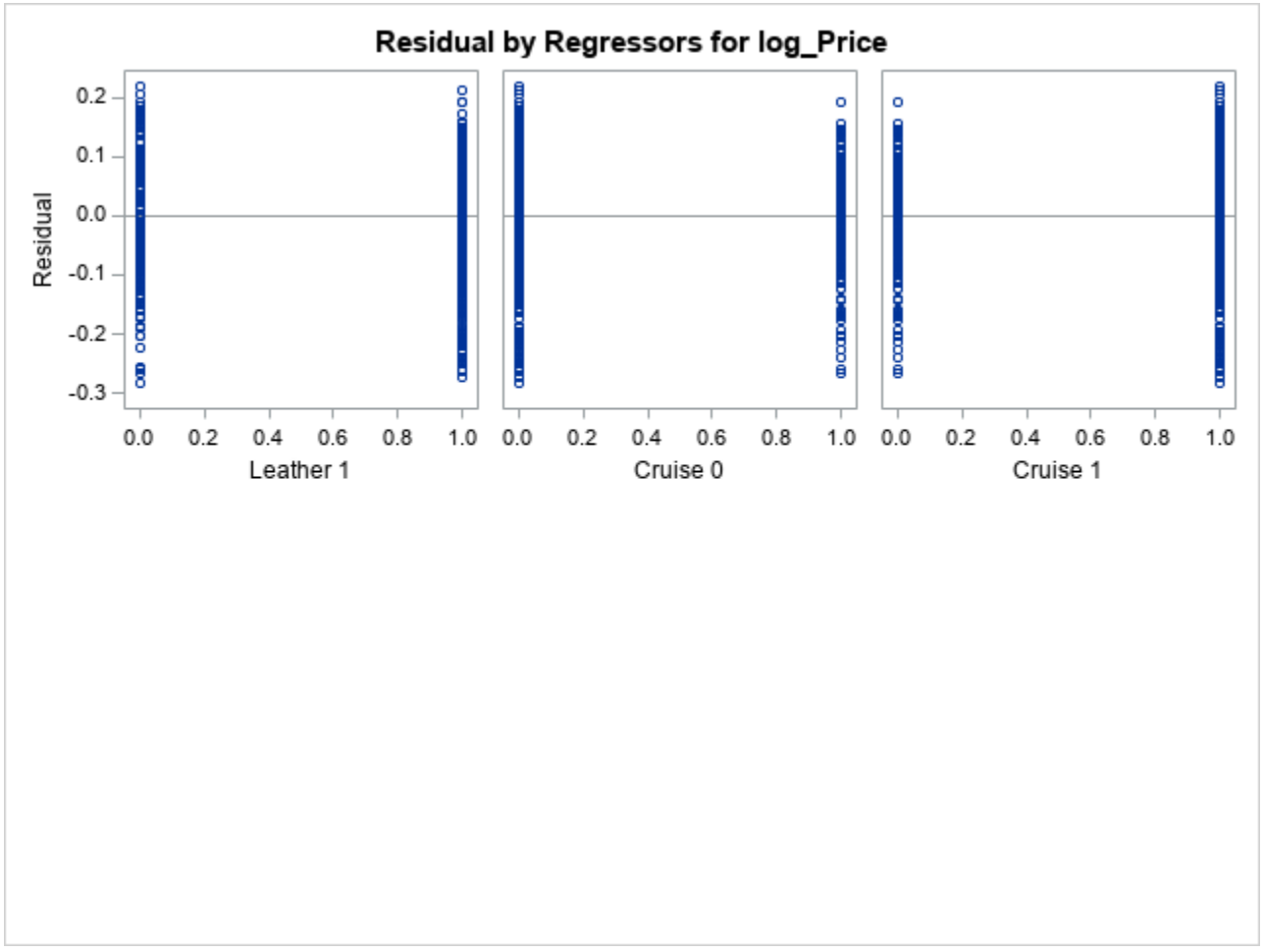
The REG Procedure
Model: MODEL1
Dependent Variable: log_Price




Residual by Regressors for log_Price
Residual
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
20000
40000
Mileage
0.0
0.2
0.4
0.6
0.8
1.0
Make_ 2
Make_ 3
Model_ 2
Model_ 3
Trim_ 2


Residual by Regressors for log_Price
Residual
0.2
0.1
0.0
-0.1
-0.2
-0.3
0.0
0.2
0.4
0.6
0.8
1.0
Trim_ 3
Type Hatchback
Type Convertible
Type Coupe
Type Sedan
Type Wagon


Residual by Regressors for log_Price
Residual
0.2
0.1
0.0
-0.1
-0.2
-0.3
0.0 0.2 0.4 0.6 0.8 1.0
Cylinder 4
Cylinder 6
Cylinder 8
Sound 0
Sound 1
Leather 0


Residual by Regressors for log_Price
Residual
0.2
0.1
0.0
-0.1
-0.2
-0.3
0.0 0.2 0.4 0.6 0.8 1.0
Leather 1
Cruise 0
Cruise 1

The SAS System

The REG Procedure
Model: MODEL1.1
Dependent Variable: log_Price

Number of Observations Read	744
Number of Observations Used	744

Weight: REWEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	114.15616	7.13476	994.34	<.0001
Error	727	5.21649	0.00718		
Corrected Total	743	119.37265			

Root MSE	0.08471	R-Square	0.9563
Dependent Mean	9.91311	Adj R-Sq	0.9553
Coeff Var	0.85450		

Note: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

Note: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

Type_Wagon =	Intercept - Type_Hatchback - Type_Convertible - Type_Coupe - Type_Sedan
Cylinder_8 =	Intercept - Cylinder_4 - Cylinder_6
Sound_1 =	Intercept - Sound_0
Leather_1 =	Intercept - Leather_0
Cruise_1 =	Intercept - Cruise_0

Parameter Estimates

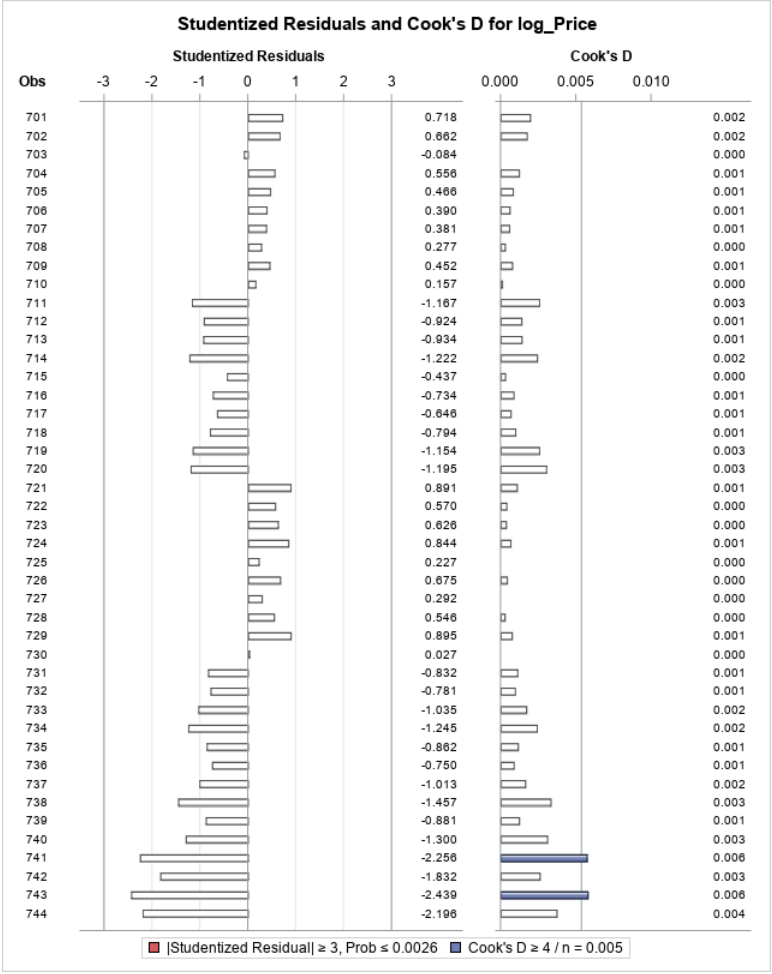
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	95% Confidence Limits	
Intercept	Intercept	B	10.31592	0.02554	403.96	<.0001	0	10.26579	10.36606
Mileage	Mileage	1	-0.00000815	3.938922E-7	-20.69	<.0001	-0.16130	-0.00000892	-0.00000738
Make__2	Make_ 2	1	0.26436	0.01825	14.49	<.0001	0.19799	0.22854	0.30018
Make__3	Make_ 3	1	0.54742	0.01686	32.46	<.0001	0.48506	0.51431	0.58053
Model__2	Model_ 2	1	0.16019	0.01128	14.20	<.0001	0.18592	0.13804	0.18234
Model__3	Model_ 3	1	0.40959	0.02529	16.20	<.0001	0.32395	0.35995	0.45923
Trim__2	Trim_ 2	1	0.03625	0.00873	4.15	<.0001	0.04267	0.01910	0.05339
Trim__3	Trim_ 3	1	-0.04009	0.02232	-1.80	0.0729	-0.02912	-0.08390	0.00373
Type_Hatchback	Type Hatchback	B	-0.27786	0.01742	-15.95	<.0001	-0.18548	-0.31206	-0.24366
Type_Convertible	Type Convertible	B	0.15533	0.02381	6.52	<.0001	0.09687	0.10859	0.20208
Type_Coupe	Type Coupe	B	-0.14228	0.01536	-9.26	<.0001	-0.12630	-0.17244	-0.11212
Type_Sedan	Type Sedan	B	-0.19830	0.01324	-14.98	<.0001	-0.23539	-0.22429	-0.17231
Type_Wagon	Type Wagon	0	0	.	.	.	.	.	.
Cylinder_4	Cylinder 4	B	-0.50579	0.02016	-25.09	<.0001	-0.61120	-0.54537	-0.46620
Cylinder_6	Cylinder 6	B	-0.15807	0.01775	-8.90	<.0001	-0.18701	-0.19292	-0.12321
Cylinder_8	Cylinder 8	0	0	.	.	.	.	.	.
Sound_0	Sound 0	B	-0.02970	0.00754	-3.94	<.0001	-0.03298	-0.04450	-0.01491
Sound_1	Sound 1	0	0	.	.	.	.	.	.
Leather_0	Leather 0	B	-0.03298	0.00813	-4.06	<.0001	-0.03455	-0.04894	-0.01703
Leather_1	Leather 1	0	0	.	.	.	.	.	.
Cruise_0	Cruise 0	B	-0.00960	0.00986	-0.97	0.3306	-0.00970	-0.02896	0.00976
Cruise_1	Cruise 1	0	0	.	.	.	.	.	.

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variation												
			Intercept	Mileage	Make__2	Make__3	Model__2	Model__3	Trim__2	Trim__3	Type_Hatchback	Type_Convertible	Type_Coupe	Type_Sedan	Type_
1	8.34108	1.00000	1.65663E-14	0.00176	0.00069147	0.00082152	0.00149	0.00033770	0.00235	0.00048471	9.97788E-15	2.81084E-14	1.20147E-14	1.74388E-14	1.85
2	2.73345	1.74685	2.01939E-15	0.00023693	0.01837	0.00027796	0.00080269	0.01149	0.00041785	0.00671	9.42162E-14	4.20748E-13	9.85204E-15	2.18971E-15	6.259
3	2.18554	1.95358	2.02933E-15	0.00010751	0.00555	0.02324	0.00527	0.00158	0.00140	0.01503	9.98678E-14	1.18853E-12	5.92586E-14	3.76271E-14	4.369
4	1.86010	2.11759	2.88565E-15	0.00027162	0.00008244	0.00073950	0.01086	0.00030858	0.01502	0.00161	8.07383E-13	1.18282E-13	2.36083E-13	5.85901E-14	5.967
5	1.23494	2.59889	1.09418E-15	0.00011294	0.00653	0.00434	0.00003398	0.00103	0.01472	0.01493	8.05875E-13	2.5382E-12	9.29302E-15	1.56241E-15	2.797
6	1.14315	2.70121	1.41691E-16	0.00000192	0.00334	0.00173	0.00113	0.00067431	0.00213	0.00017865	2.53701E-12	9.08631E-15	1.62732E-12	1.41919E-15	6.050
7	1.02058	2.85883	3.08835E-15	0.00047622	0.00440	0.00815	0.00360	0.00012074	0.01272	0.00008742	5.73181E-13	2.43888E-14	2.70872E-12	1.00977E-13	4.039
8	0.77881	3.27261	5.3695E-16	0.00000925	0.00356	0.00380	0.01073	0.00097075	0.00003557	0.00076819	4.89564E-12	2.19306E-13	3.27933E-13	2.18436E-13	2.696
9	0.63664	3.61964	3.85876E-17	3.310634E-7	0.00595	0.00374	0.00079908	0.00159	0.00328	0.00218	1.12732E-13	3.48216E-14	6.0234E-14	1.53552E-13	9.096
10	0.56857	3.83016	2.28141E-15	0.00074198	0.00725	0.03254	0.00089211	0.00380	0.02378	0.01414	4.99713E-12	5.05076E-13	1.24366E-13	6.02266E-14	1.098
11	0.37231	4.73325	1.36398E-14	0.00483	0.06096	0.02135	0.01883	0.00083327	0.33835	0.00223	3.60276E-14	1.79943E-12	1.18346E-12	3.79219E-13	1.268

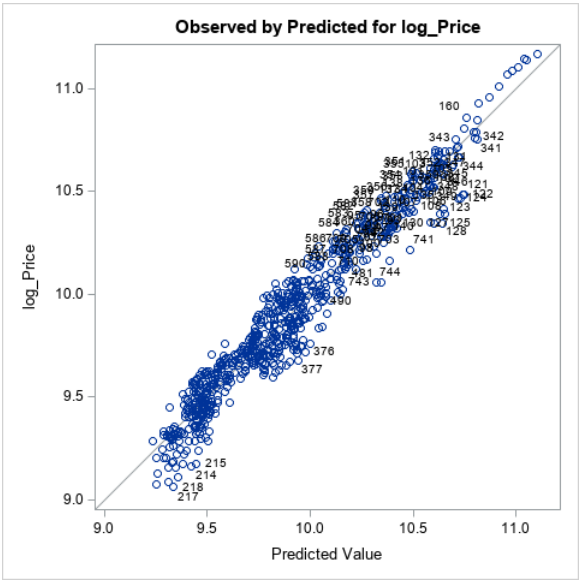
12	0.35022	4.88020	5.95131E-15	0.00237	0.04810	0.00718	0.17127	5.379142E-8	0.00893	0.00353	4.72989E-15	1.90927E-14	1.51062E-13	5.4405E-14	3.142
13	0.28832	5.37869	2.79467E-15	0.00245	0.33491	0.04106	0.11701	0.00007233	0.29944	0.00104	4.00198E-14	3.64989E-12	8.27997E-13	7.09538E-16	4.866
14	0.16139	7.18905	1.76823E-17	0.00238	0.21152	0.46889	0.18355	0.12741	0.07487	0.01808	2.90274E-13	9.21797E-12	3.62034E-16	2.55632E-13	2.942
15	0.12876	8.04868	4.33748E-15	0.00319	0.00169	0.03089	0.00604	0.09941	0.12924	0.83605	6.02761E-13	3.39786E-11	3.50606E-13	1.42508E-13	1.492
16	0.12522	8.16159	1.97316E-13	0.96340	0.02313	0.00249	0.00133	0.01439	0.00047243	0.00036171	2.04502E-13	9.28964E-14	1.34799E-13	2.21928E-13	2.576
17	0.07091	10.84580	9.66555E-14	0.01766	0.26397	0.34876	0.46637	0.73598	0.07283	0.08260	4.05118E-13	4.69082E-12	5.44451E-14	1.25538E-13	1.538
18	1E-12	2888092	1.00000	0	0	0	0	0	0	0	0.04000	0.04000	0.04000	0.04000	0
19	1E-12	2888092	0	0	0	0	0	0	0	0	0.06000	0.06000	0.06000	0.06000	0
20	1E-12	2888092	0	0	0	0	0	0	0	0	0.10000	0.10000	0.10000	0.10000	0
21	1E-12	2888092	0	0	0	0	0	0	0	0	0.20000	0.20000	0.20000	0.20000	0
22	1E-12	2888092	0	0	0	0	0	0	0	0	0.60000	0.60000	0.60000	0.60000	0

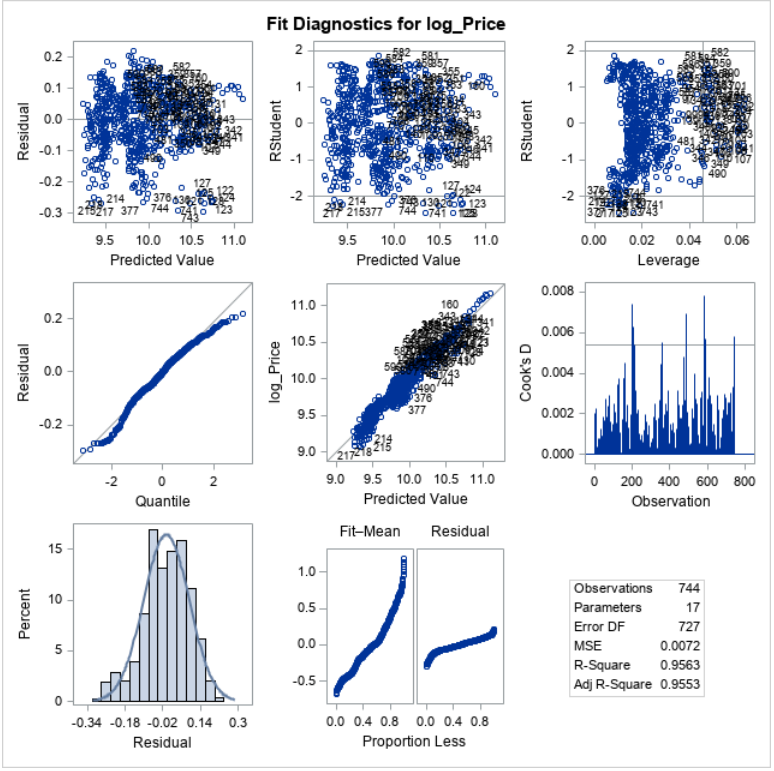
Note: Singularities or near singularities caused grossly large variance calculations. To provide diagnostics, eigenvalues are inflated to a minimum of 1e-12.

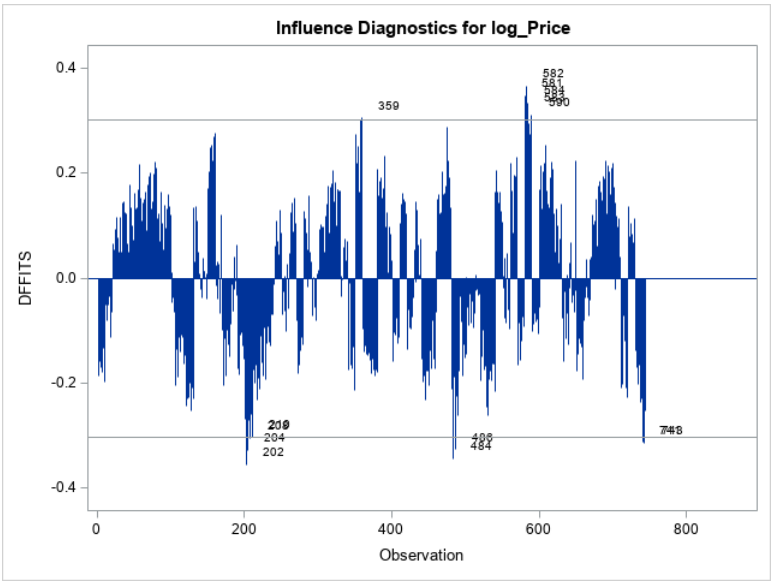


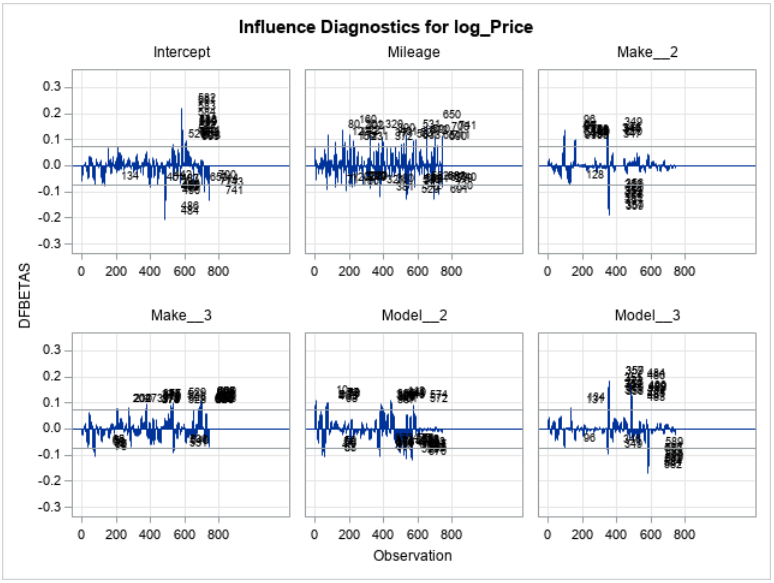
Sum of Residuals	-7.0496E-12
Sum of Squared Residuals	5.21649
Predicted Residual SS (PRESS)	5.42051

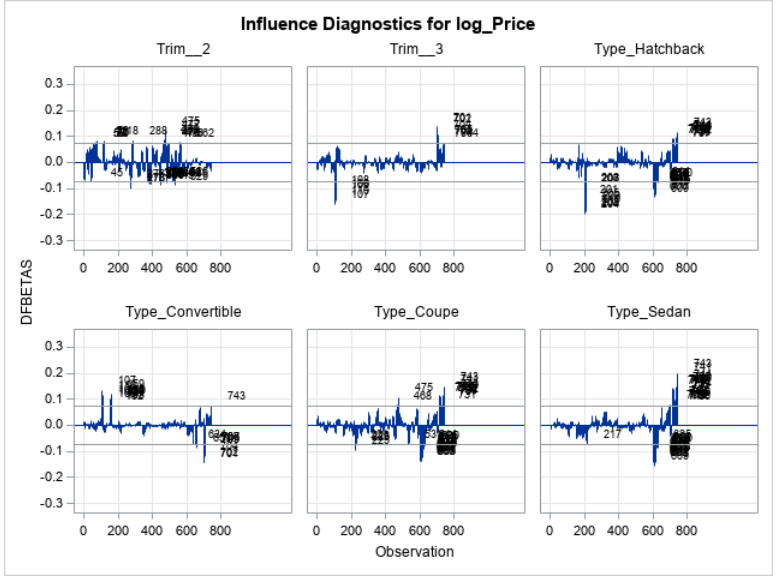
Note: The above statistics use observation weights or frequencies.

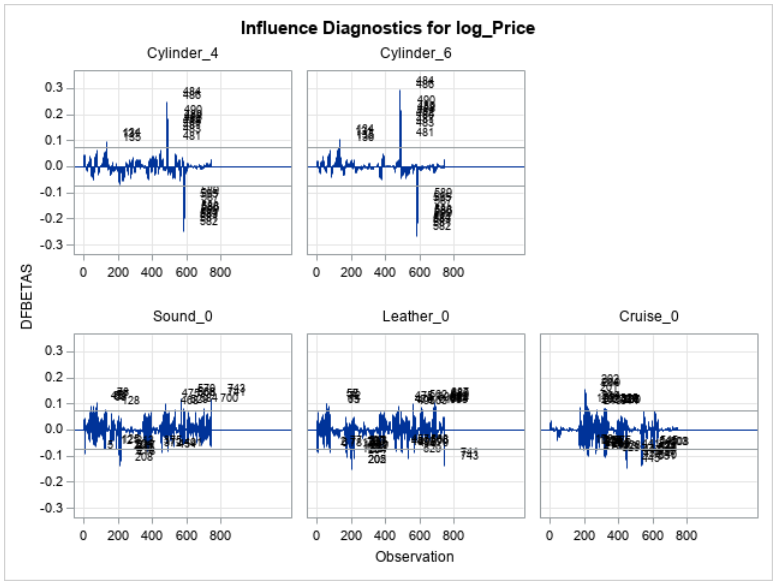


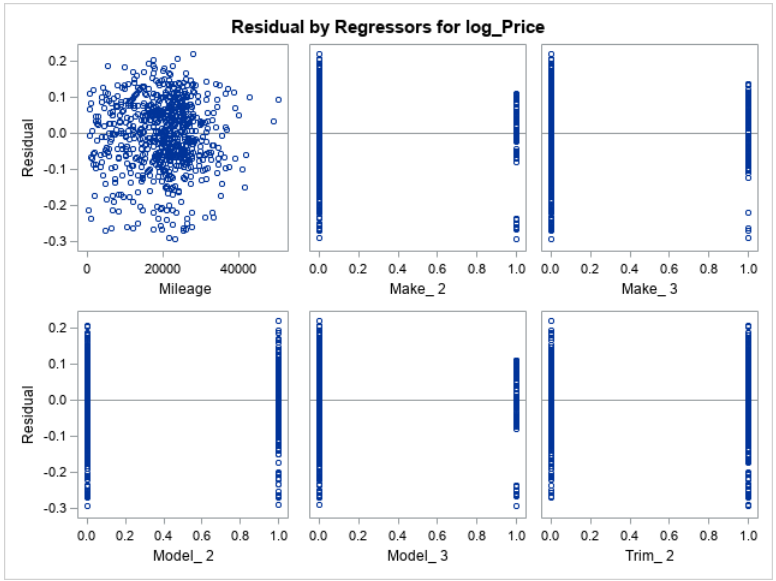

Fit Diagnostics for log_Price
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
log_Price
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 744
Parameters 17
Error DF 727
MSE 0.0072
R-Square 0.9563
Adj R-Square 0.9553

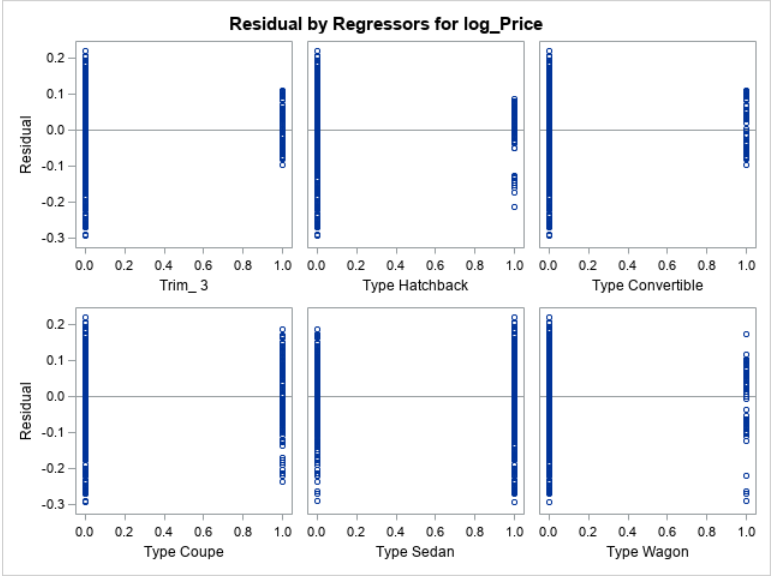

Influence Diagnostics for log_Price
DFFITS
Observation

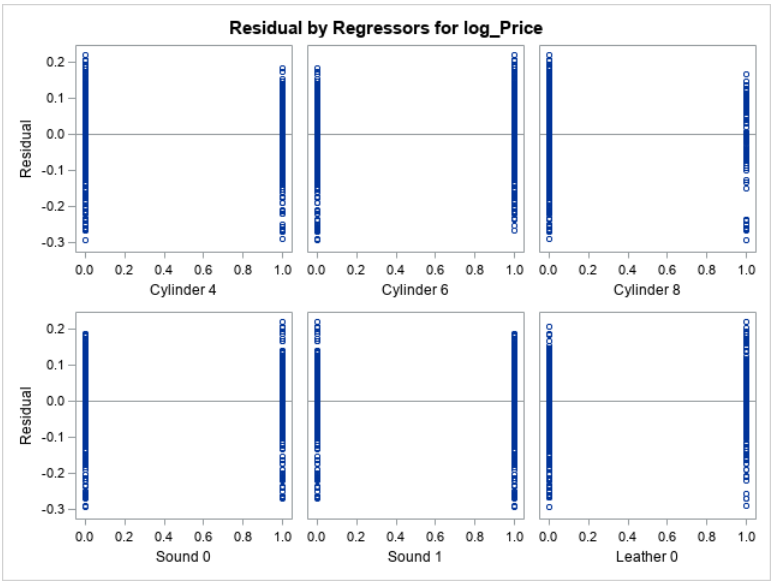

Influence Diagnostics for log_Price
Intercept
Mileage
Make__2
Make__3
Model__2
Model__3
DFBETAS
Observation

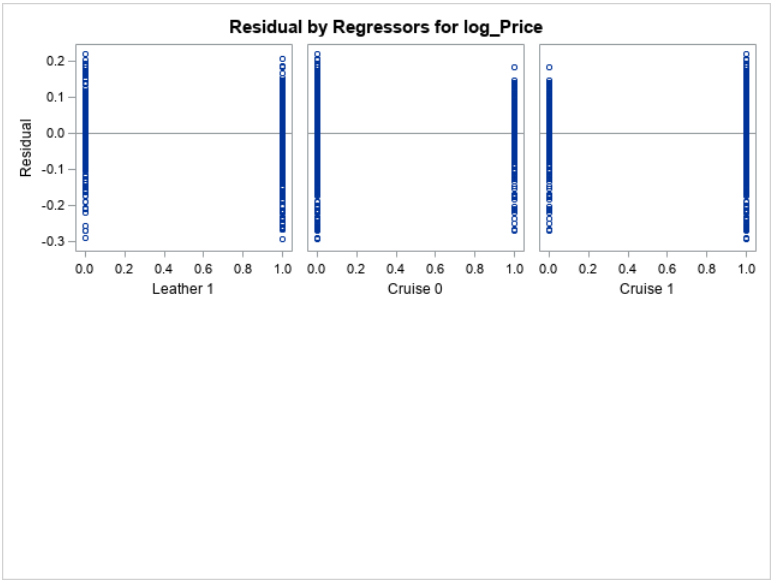

Influence Diagnostics for log_Price
Trim__2
Trim__3
Type_Hatchback
Type_Convertible
Type_Coupe
Type_Sedan
DFBETAS
Observation


Influence Diagnostics for log_Price
Cylinder_4
Cylinder_6
Sound_0
Leather_0
Cruise_0
DFBETAS
Observation

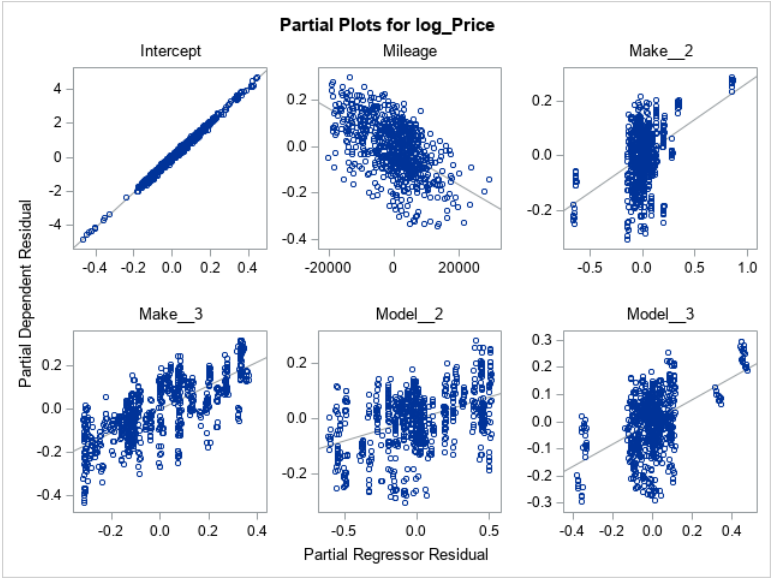

Residual by Regressors for log_Price
Residual
Mileage
Make_ 2
Make_ 3
Model_ 2
Model_ 3
Trim_ 2

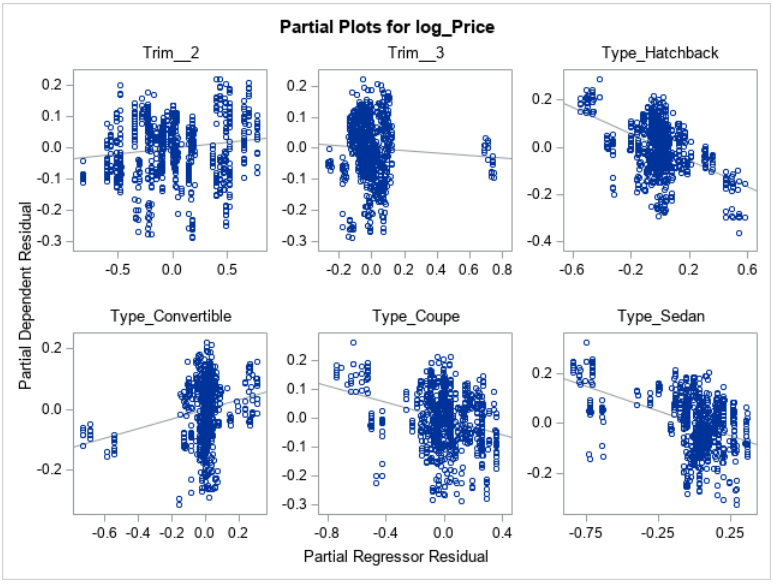

Residual by Regressors for log_Price
Residual
Trim_ 3
Type Hatchback
Type Convertible
Type Coupe
Type Sedan
Type Wagon

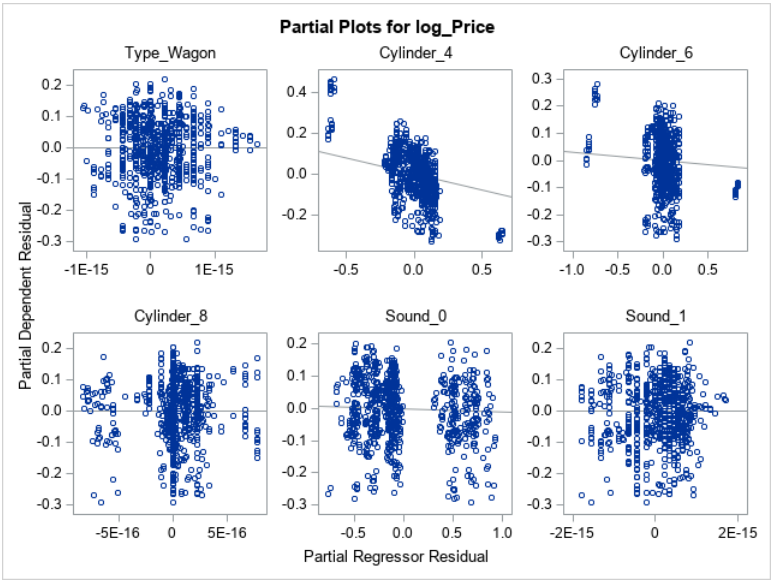

Residual by Regressors for log_Price
Residual
Cylinder 4
Cylinder 6
Cylinder 8
Sound 0
Sound 1
Leather 0


Residual by Regressors for log_Price
Residual
Leather 1
Cruise 0
Cruise 1

The REG Procedure
Model: MODEL1.1
Partial Regression Residual Plot







The SAS System

The GLMSELECT Procedure

Data Set	WORK.REG_DESIGN3
Dependent Variable	log_Price
Selection Method	LASSO
Stop Criterion	SBC
Effect Hierarchy Enforced	None

Number of Observations Read	744
Number of Observations Used	744

Class Level Information		
Class	**Levels**	**Values**
Make_	3	1 2 3
Type	5	Convertible Coupe Hatchback Sedan Wagon
Model_	3	1 2 3
Trim_	3	1 2 3
Cylinder	3	4 6 8
Doors	2	2 4
Liter_	2	1 2
Cruise	2	0 1
Sound	2	0 1
Leather	2	0 1

Dimensions	
Number of Effects	22
Number of Parameters	22

The GLMSELECT Procedure

LASSO Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Model R-Square	Adjusted R-Square	AIC	SBC
0	Intercept		1	0.0000	0.0000	-570.2118	-1311.5998
1	Model__3		2	0.2129	0.2118	-746.2843	-1483.0602
2	Cruise_0		3	0.2292	0.2272	-759.9387	-1492.1026
3	Model__2		4	0.2741	0.2712	-802.5929	-1530.1447
4	Trim__3		5	0.3163	0.3126	-845.1262	-1568.0660
5	Make__3		6	0.3822	0.3781	-918.5725	-1636.9002
6	Cylinder_8		7	0.5199	0.5160	-1104.1422	-1817.8580
7	Make__2		8	0.6226	0.6190	-1281.2117	-1990.3154
8	Cylinder_4		9	0.6692	0.6656	-1377.1749	-2081.6665
9	Type_Convertible		10	0.7999	0.7974	-1749.1512	-2449.0308
10	Mileage		11	0.8619	0.8600	-2022.9809	-2718.2484
11	Type_Hatchback		12	0.8870	0.8853	-2170.5837	-2861.2392
12	Type_Wagon		13	0.9127	0.9113	-2360.2899	-3046.3334
13	Trim__2		14	0.9174	0.9159	-2399.3450	-3080.7764
14		Trim__3	13	0.9308	0.9296*	-2532.7202*	-3218.7637*

*** Optimal Value of Criterion**

Selection stopped at a local minimum of the SBC criterion.

Stop Details

Candidate For	Effect	Candidate SBC		Compare SBC
Entry	Type_Sedan	-3217.3666	>	-3218.7637

The SAS System

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 14).

Effects:	Intercept Mileage Make__2 Make__3 Model__2 Model__3 Trim__2 Type_Hatchback Type_Convertible Type_Wagon Cylinder_4 Cylinder_8 Cruise_0

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	117.74095	9.81175	818.75	<.0001
Error	731	8.76015	0.01198		
Corrected Total	743	126.50109			

Root MSE	0.10947
Dependent Mean	9.90663
R-Square	0.9308
Adj R-Sq	0.9296
AIC	-2532.72022
AICC	-2532.14409
SBC	-3218.76369

Parameter Estimates

Parameter	DF	Estimate	Standardized Estimate
Intercept	1	9.921804	0
Mileage	1	-0.000006165	-0.122179
Make__2	1	0.218589	0.164219
Make__3	1	0.502560	0.439012
Model__2	1	0.142799	0.165696
Model__3	1	0.402703	0.318465
Trim__2	1	0.011582	0.013731
Type_Hatchback	1	-0.067496	-0.044570
Type_Convertible	1	0.249535	0.151518
Type_Wagon	1	0.098274	0.066827
Cylinder_4	1	-0.289203	-0.349686
Cylinder_8	1	0.160720	0.132947
Cruise_0	1	-0.023755	-0.023985

Appendex
Loess plots

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Make_	3	1 2 3
Type	5	Convertible Coupe Hatchback Sedan Wagon
Model_	3	1 2 3
Trim_	3	1 2 3
Cylinder	3	4 6 8
Doors	2	2 4
Liter_	2	1 2
Cruise	2	0 1
Sound	2	0 1
Leather	2	0 1

Number of Observations Read	744
Number of Observations Used	744

Appendex
Loess plots

The GLM Procedure

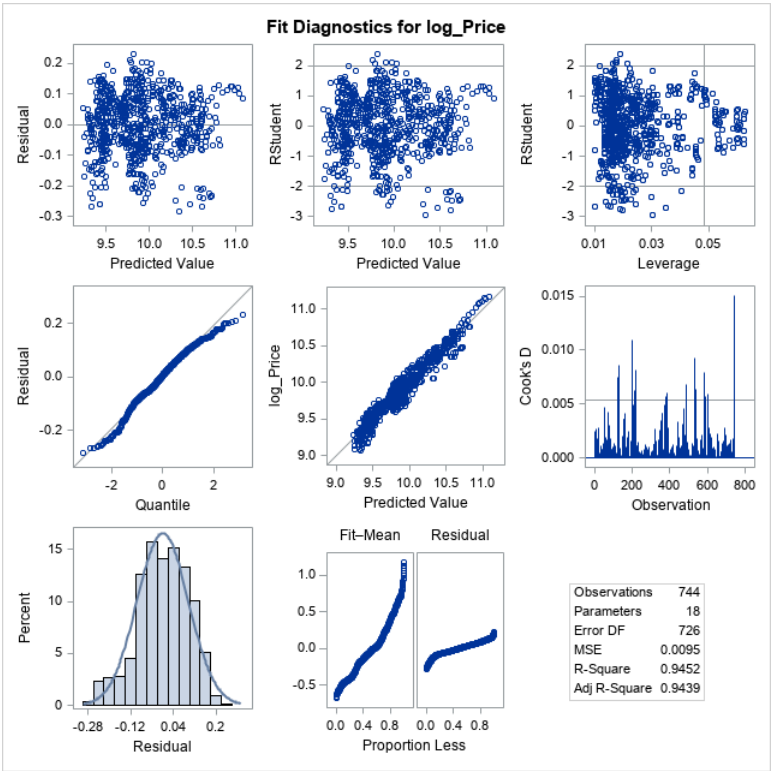
Dependent Variable: log_Price

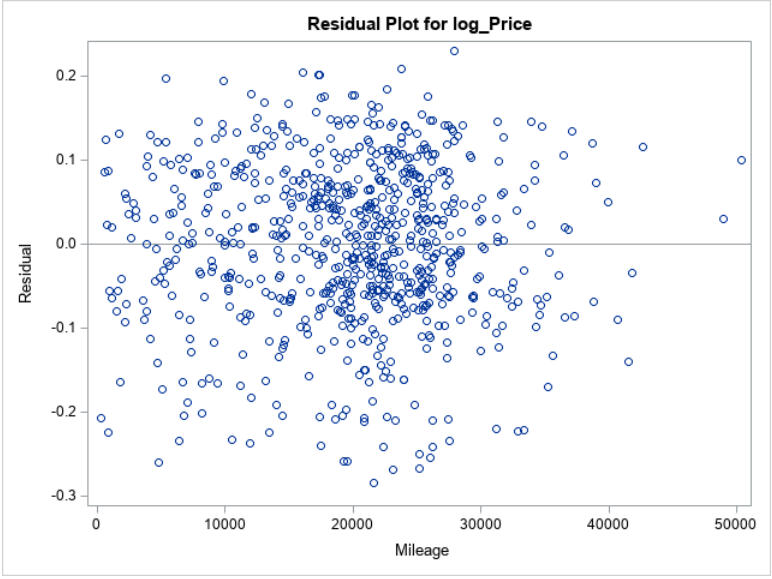
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	119.5706025	7.0335649	736.80	<.0001
Error	726	6.9304923	0.0095461		
Corrected Total	743	126.5010949			

R-Square	Coeff Var	Root MSE	log_Price Mean
0.945214	0.986252	0.097704	9.906627

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Mileage	1	2.36320800	2.36320800	247.56	<.0001
Make_	2	71.84001677	35.92000839	3762.78	<.0001
Model_	2	29.41030370	14.70515185	1540.43	<.0001
Trim_	2	3.04109530	1.52054765	159.28	<.0001
Type	4	2.03256174	0.50814043	53.23	<.0001
Cylinder	2	10.54441668	5.27220834	552.29	<.0001
Liter_	1	0.05765958	0.05765958	6.04	0.0142
Doors	0	0.00000000	.	.	.
Cruise	1	0.00735489	0.00735489	0.77	0.3804
Sound	1	0.18154546	0.18154546	19.02	<.0001
Leather	1	0.09244041	0.09244041	9.68	0.0019

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mileage	1	3.18157043	3.18157043	333.28	<.0001
Make_	2	4.00677000	2.00338500	209.86	<.0001
Model_	2	2.04204292	1.02102146	106.96	<.0001
Trim_	2	0.18774262	0.09387131	9.83	<.0001
Type	3	2.51007674	0.83669225	87.65	<.0001
Cylinder	2	9.50137902	4.75068951	497.66	<.0001
Liter_	1	0.04561539	0.04561539	4.78	0.0291
Doors	0	0.00000000	.	.	.
Cruise	1	0.01659667	0.01659667	1.74	0.1877
Sound	1	0.12837771	0.12837771	13.45	0.0003
Leather	1	0.09244041	0.09244041	9.68	0.0019




Residual Plot for log_Price
Residual
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
10000
20000
30000
40000
50000
Mileage

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: log_Price

Number of Observations Read	744
Number of Observations Used	744

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	118.89251	10.80841	1039.85	<.0001
Error	732	7.60858	0.01039		
Corrected Total	743	126.50109			

Root MSE	0.10195	R-Square	0.9399
Dependent Mean	9.90663	Adj R-Sq	0.9389
Coeff Var	1.02913		

Parameter Estimates

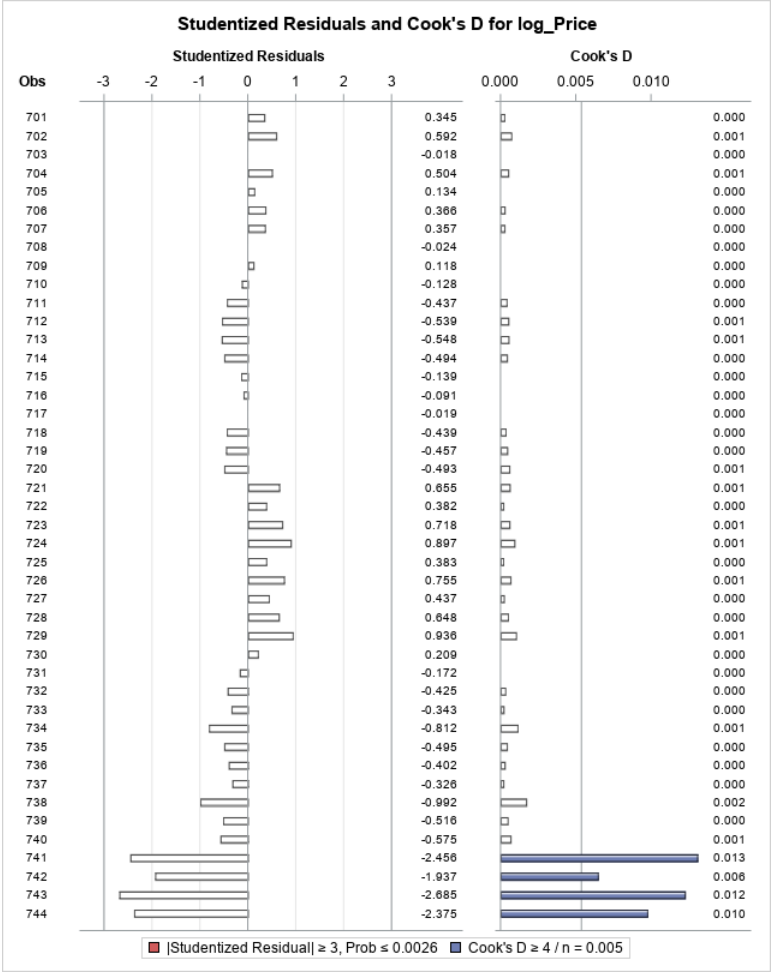
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation	95% Confidence Limits	
Intercept	Intercept	1	9.95457	0.01329	748.95	<.0001	0	0	9.92848	9.98066
Mileage	Mileage	1	-0.00000805	4.588922E-7	-17.54	<.0001	-0.15948	1.00662	-0.00000895	-0.00000715
Make__2	Make_ 2	1	0.24334	0.02124	11.45	<.0001	0.18282	3.10023	0.20164	0.28505
Make__3	Make_ 3	1	0.55226	0.01879	29.39	<.0001	0.48243	3.27963	0.51537	0.58915
Model__2	Model_ 2	1	0.14223	0.01232	11.55	<.0001	0.16504	2.48586	0.11805	0.16641
Model__3	Model_ 3	1	0.38715	0.02966	13.05	<.0001	0.30616	6.69393	0.32893	0.44537
Trim__2	Trim_ 2	1	0.02953	0.00939	3.14	0.0017	0.03501	1.50822	0.01109	0.04796
Type_Hatchback	Type Hatchback	1	-0.10023	0.01478	-6.78	<.0001	-0.06618	1.15876	-0.12924	-0.07122
Type_Convertible	Type Convertible	1	0.30649	0.01865	16.44	<.0001	0.18610	1.55987	0.26988	0.34309
Type_Wagon	Type Wagon	1	0.16640	0.01492	11.15	<.0001	0.11315	1.25311	0.13710	0.19569
Cylinder_4	Cylinder 4	1	-0.34495	0.01106	-31.19	<.0001	-0.41709	2.17637	-0.36666	-0.32323
Cylinder_8	Cylinder 8	1	0.18026	0.02017	8.94	<.0001	0.14911	3.38764	0.14066	0.21986

Covariance of Estimates

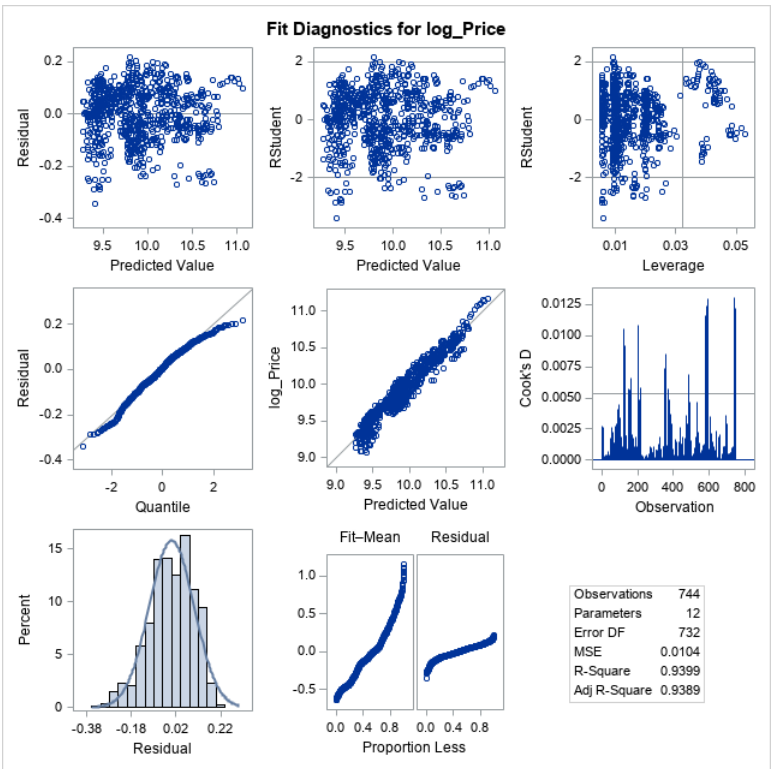
Variable	Label	Intercept	Mileage	Make__2	Make__3	Model__2	Model__3	Trim__2	Type_Hatchback	Type_Convertible	Type_Wagon	Cylinder_4	Cyli
Intercept	Intercept	0.0001766587	-4.156415E-9	0.0000135514	0.0001023625	-0.000070639	-0.000080513	-0.000044461	-0.000056705	-0.000031059	-0.000010073	-0.000081209	-8.917
Mileage	Mileage	-4.156415E-9	2.10582E-13	1.183296E-10	-2.18338E-10	-3.13714E-11	2.870657E-10	2.716252E-12	1.091347E-10	-1.30735E-10	-1.00818E-10	5.416625E-11	-2.351
Make__2	Make_ 2	0.0000135514	1.183296E-10	0.000451337	0.0000390907	-0.000035678	-0.000406215	-0.000031873	-0.000010295	0.0000407759	5.1840673E-6	-0.000011032	0.0000
Make__3	Make_ 3	0.0001023625	-2.18338E-10	0.0000390907	0.0003531387	-0.000145357	-0.000122217	-0.000059829	-0.000036993	-0.000130102	-0.000063414	-0.000122403	0.0000
Model__2	Model_ 2	-0.000070639	-3.13714E-11	-0.000035678	-0.000145357	0.0001517042	0.000151918	-1.96256E-6	0.0000381304	-2.726376E-6	2.4337997E-6	0.0000662148	-0.000
Model__3	Model_ 3	-0.000080513	2.870657E-10	-0.000406215	-0.000122217	0.000151918	0.0008794808	5.6863034E-6	0.0000396008	-0.000097034	-0.000014172	0.0000708646	-0.000
Trim__2	Trim_ 2	-0.000044461	2.716252E-12	-0.000031873	-0.000059829	-1.96256E-6	5.6863034E-6	0.0000881597	0.0000311344	0.0000652237	0.0000176227	0.0000265493	0.0000
Type_Hatchback	Type Hatchback	-0.000056705	1.091347E-10	-0.000010295	-0.000036993	0.0000381304	0.0000396008	0.0000311344	0.0002183501	0.0000271492	0.0000231617	0.000019017	3.5549
Type_Convertible	Type Convertible	-0.000031059	-1.30735E-10	0.0000407759	-0.000130102	-2.726376E-6	-0.000097034	0.0000652237	0.0000271492	0.0003476362	0.0000670453	0.0000128111	-8.291
Type_Wagon	Type Wagon	-0.000010073	-1.00818E-10	5.1840673E-6	-0.000063414	2.4337997E-6	-0.000014172	0.0000176227	0.0000231617	0.0000670453	0.000222673	-0.000022328	-6.713
Cylinder_4	Cylinder 4	-0.000081209	5.416625E-11	-0.000011032	-0.000122403	0.0000662148	0.0000708646	0.0000265493	0.000019017	0.0000128111	-0.000022328	0.0001223149	3.6506
Cylinder_8	Cylinder 8	-8.917562E-7	-2.35113E-10	0.0000500318	0.0000552539	-0.000057885	-0.000398213	0.0000110666	3.5549314E-6	-8.291756E-6	-6.713643E-7	3.6506931E-6	0.0004

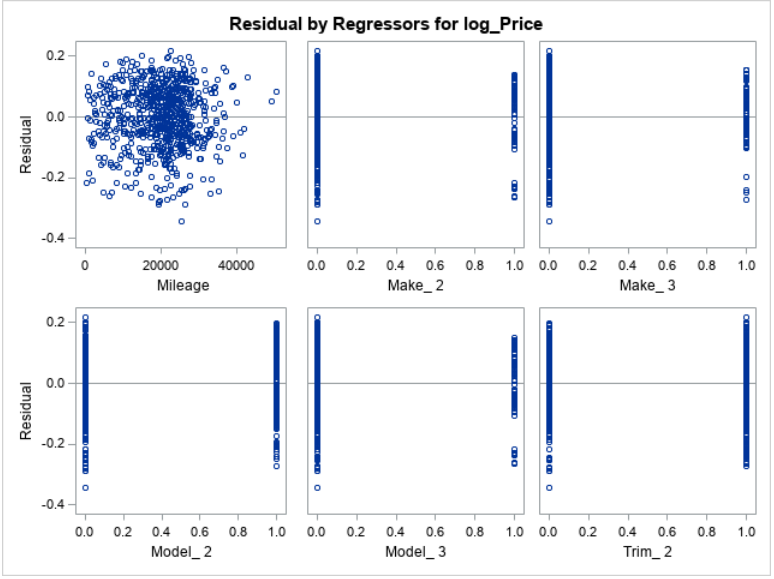
Collinearity Diagnostics

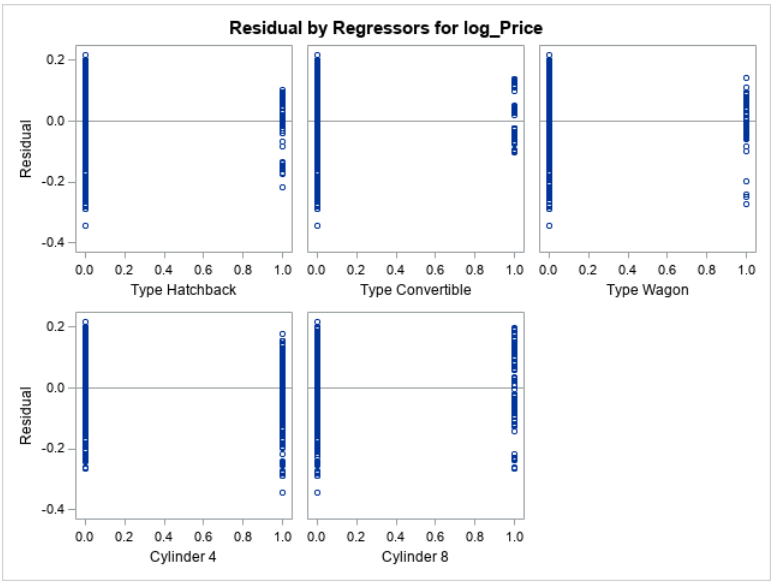
Number	Eigenvalue	Condition Index	Proportion of Variation										
			Intercept	Mileage	Make__2	Make__3	Model__2	Model__3	Trim__2	Type_Hatchback	Type_Convertible	Type_Wagon	Cylinder_4	Cylinder_8
1	4.44718	1.00000	0.00331	0.00560	0.00325	0.00499	0.00628	0.00159	0.00977	0.00130	0.00496	0.00567	0.00564	0.00321
2	2.47889	1.33941	0.00007317	0.00020103	0.02597	0.00874	0.00517	0.01428	0.00015033	0.00115	0.00106	0.01495	0.00824	0.02323
3	1.22271	1.90713	0.00385	0.00653	0.00000127	0.03134	0.01137	0.00022481	0.00027922	0.31044	0.08654	0.00425	0.00296	0.00066732
4	1.07108	2.03766	0.00001739	0.00002440	0.00414	0.00336	0.00523	0.00026018	0.09445	0.13059	0.25930	0.01318	0.00748	0.00042169
5	0.90392	2.21808	0.00057365	0.00131	0.00366	0.00003726	0.03313	0.00354	0.03632	0.00259	0.01512	0.53232	0.01638	0.00185
6	0.56797	2.79822	0.00996	0.02630	0.01422	0.06609	0.04494	0.00044149	0.00926	0.44062	0.00977	0.05758	0.07266	0.00081902
7	0.44168	3.17312	0.00447	0.01795	0.07811	0.09237	0.10406	0.00307	0.06918	0.00118	0.00884	0.15251	0.10354	0.05442
8	0.31519	3.75628	0.00045862	0.00171	0.00716	0.06551	0.05975	0.00072486	0.29028	0.00753	0.47184	0.20542	0.07600	0.13520
9	0.27363	4.03146	0.00179	0.00837	0.42809	0.01555	0.10529	0.00388	0.27797	0.01297	0.00123	0.00023177	0.00065782	0.22409
10	0.14065	5.62302	0.00018704	0.33333	0.05210	0.35504	0.15263	0.07462	0.09669	0.01948	0.10734	0.01128	0.31504	0.09083
11	0.08783	7.11566	0.06758	0.20856	0.34054	0.04213	0.14242	0.71238	0.00483	3.600523E-7	0.02124	0.00082934	0.00945	0.42999
12	0.04927	9.50063	0.90773	0.39011	0.04276	0.31484	0.32973	0.18499	0.11081	0.07216	0.01276	0.00179	0.38196	0.03528



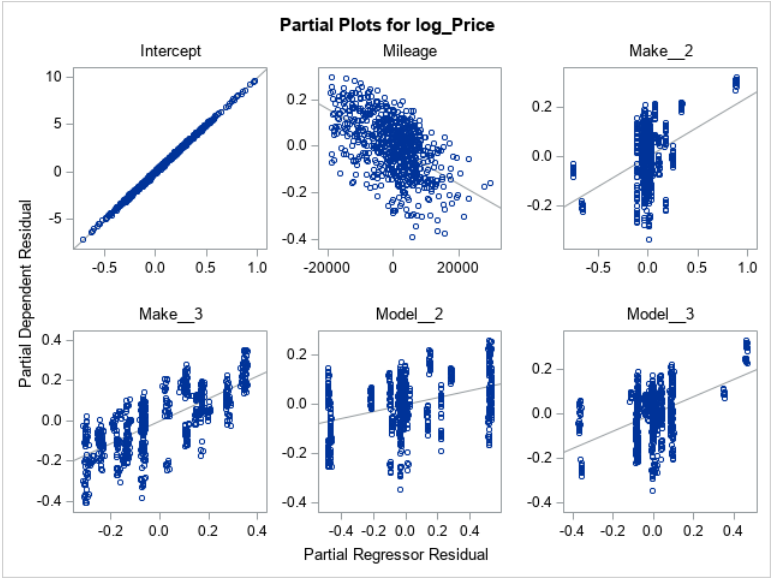
Sum of Residuals	-6.2718E-13
Sum of Squared Residuals	7.60858
Predicted Residual SS (PRESS)	7.84919

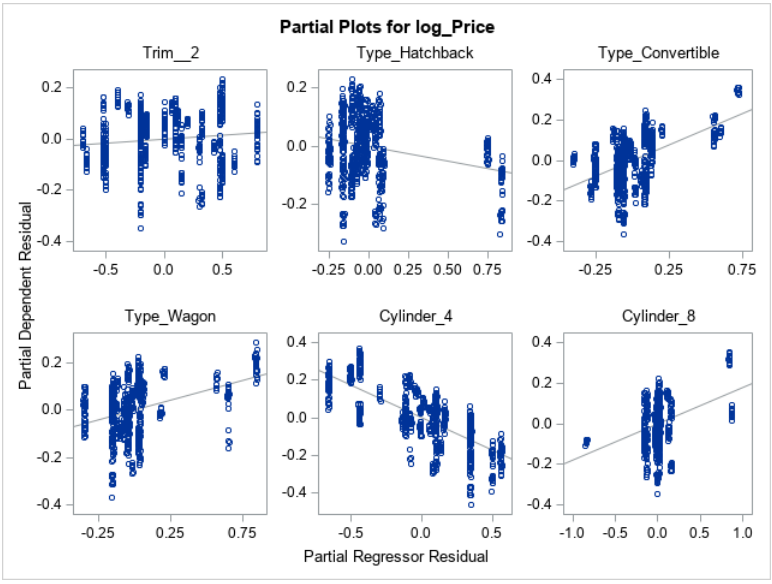



Residual by Regressors for log_Price
Residual
0.2
0.0
-0.2
-0.4
0
20000
40000
Mileage
0.0 0.2 0.4 0.6 0.8 1.0
Make_ 2
Make_ 3
Model_ 2
Model_ 3
Trim_ 2


Residual by Regressors for log_Price
Residual
0.2
0.0
-0.2
-0.4
0.0 0.2 0.4 0.6 0.8 1.0
Type Hatchback
Type Convertible
Type Wagon
Cylinder 4
Cylinder 8

The REG Procedure
Model: MODEL1
Partial Regression Residual Plot





```
*IMPORT THE DATASET;

PROC IMPORT OUT= WORK.BlueBook
        DATAFILE= "\\Client\C$\Users\ltyle\Documents\UTSA\STA
6013\PROJECT\Blue Book Modified Car Data 2015 11 11
2020.xlsX"
        DBMS=EXCEL REPLACE;
    RANGE="'Cars Lab Data$'";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;

Title "Appendex";
*Discriptive by numerical and catogorical;
title2 'Discriptive by numerical and catogorical';
proc means data=Bluebook n nmiss mean std median min max
range clm cv maxdec=3;
class Type Make Model Trim Type Cylinder Liter;
 var Price Mileage Cruise Sound Leather;
 run;
 *Discriptive means of numerical values;
title2 'Summary Statistics of each numerical values';
proc means data=Bluebook n nmiss mean std median min max
range maxdec=3;
var Price Mileage Cruise Sound Leather ;
run;
*Discriptive by Classification;
title2 'Summary Statistics of each categorical values';
proc means data=Bluebook n nmiss mean std median min max
range maxdec=3;
id Type Make Model Trim Type Cylinder Liter;
 var Price Mileage Cruise Sound Leather;
 run;

 *Scatter plots;
proc sgscatter data=Bluebook;
matrix Price Mileage Cruise Sound Leather
     / ellipse=(type=mean)
       diagonal=(histogram kernel);
run;
*Box plots for catogorial values based on Price;

proc sgplot data=BlueBook;
vbox Price / category= Model;
run;

proc sgplot data=BlueBook;
vbox Price / category= Make;
run;

proc sgplot data=BlueBook;
vbox Price / category= Trim;
run;

proc sgplot data=BlueBook;
vbox Price / category= Type;
run;

proc sgplot data=BlueBook;
vbox Price / category= Cylinder;
run;

proc sgplot data=BlueBook;
vbox Price / category= Liter;
run;

proc sgplot data=BlueBook;
vbox Price / category= Doors;
run;

proc sgplot data=BlueBook;
vbox Price / category= Cruise;
run;

proc sgplot data=BlueBook;
vbox Price / category= Sound;
run;

proc sgplot data=BlueBook;
vbox Price / category= Leather;
run;

title2'Loess plots';
proc loess data=Bluebook plot(only)=fitplot;
  model Price=mileage /smooth=0.1 0.2 0.4 0.6 0.8 1.0;
run;


*Specification of the model, transformation;

data BlueBook;
        length Model_  8;
        set BlueBook;

        select (Model);
                when ('AVEO') Model_=1;
                when ('Vibe') Model_=1;
                when ('Cavalier') Model_=1;
                when ('Century') Model_=1;
                when ('Classic') Model_=1;
                when ('Cobalt') Model_=1;
                when ('Grand Am') Model_=1;
                when ('Grand Prix') Model_=1;
                when ('Lesabre') Model_=1;
                when ('Malibu') Model_=1;
                when ('Monte Carlo') Model_=1;
                when ('Sunfire') Model_=1;
                when ('9-2XAWD') Model_=2;
                when ('9_3') Model_=2;
                when ('9_3 HO') Model_=2;
                when ('9_5') Model_=2;
                when ('9_5 HO') Model_=2;
                when ('9-2X AWD') Model_=2;
                when ('Bonneville') Model_=2;
                when ('CTS') Model_=2;
                when ('G6') Model_=2;
                when ('GTO') Model_=2;
                when ('Impala') Model_=2;
                when ('GTO') Model_=2;
                when ('Lacrosse') Model_=2;
                when ('Park Avenue') Model_=2;
                when ('CST-V') Model_=3;
                when ('Corvette') Model_=3;
                when ('Deville') Model_=3;
                when ('STS-V6') Model_=3;
                when ('STS-V8') Model_=3;
                when ('XLR-V8') Model_=3;
                otherwise Model_=Model;
        end;
```

```
run;

data BlueBook;
        length Make_ 8;
        set BlueBook;

        select (Make);
        when ('Buick') Make_=1;
        when ('Chevrolet') Make_=1;
        when ('Pontiac') Make_=1;
        when ('Cadillac') Make_=2;
        when ('SAAB') Make_=3;
        otherwise Make_=Make;
        end;
        run;


data BlueBook;
        length Trim_ 8;
        set BlueBook;
        select (Trim);
        when ('CXL Sedan 4') Trim_=1;
        when ('DTS Sedan 4') Trim_=1;
        when ('LS Coupe 2D') Trim_=1;
        when ('LS Sport Coupe 2D') Trim_=1;
        when ('LS Sedan 4D') Trim_=1;
        when ('LT Hatchback 4D') Trim_=1;
        when ('Coupe 2D') Trim_=1;
        when ('CXL Sedan 4D') Trim_=1;
        when ('DTS Sedan 4D') Trim_=1;
        when ('SS Sedan 4D') Trim_=1;
        when ('GT Coupe 2D') Trim_=1;
        when ('LS Hatchback 4D') Trim_=1;
        when ('LT Sedan 4D') Trim_=1;
        when ('CX Sedan 4D') Trim_=1;
        when ('SVM Sedan 4D') Trim_=1;
        when ('SVM Hatchback 4D') Trim_=1;
        when ('LS MAXX Hback 4D') Trim_=1;
        when ('LT MAXX Hback 4D') Trim_=1;
        when ('MAXX Hback 4D') Trim_=1;
        when ('Custom Sedan 4D') Trim_=1;
        when ('LS Sport Sedan 4D') Trim_=1;
        when ('SE Sedan 4D') Trim_=1;
        when ('GT Sedan 4D') Trim_=1;
        when ('GT Sportwagon') Trim_=1;
        when ('Sportwagon 4D') Trim_=1;
        when ('AWD Sportwagon 4D') Trim_=1;
        when ('SS Coupe 2D') Trim_=2;
        when ('Linear Wagon 4D') Trim_=2;
        when ('Coupe 2D') Trim_=2;
        when ('LT Coupe 2D') Trim_=2;
        when ('Arc Sedan 4D') Trim_=2;
        when ('GXP Sedan 4D') Trim_=2;
        when ('SLE Sedan 4D') Trim_=2;
        when ('Linear Sedan 4D') Trim_=2;
        when ('Aero Sedan 4D') Trim_=2;
        when ('Aero Wagon 4D') Trim_=2;
        when ('CXS Sedan 4D') Trim_=2;
        when ('GT Sedan 4D') Trim_=2;
        when ('GTP Sedan 4D') Trim_=2;
        when ('Limited Sedan 4D') Trim_=2;
        when ('Linear Seda') Trim_=2;
        when ('Linear Sedan 4D') Trim_=2;
        when ('Linear Wagon 4D') Trim_=2;
        when ('Sedan 4D') Trim_=2;
        when ('Special Ed Ultra 4D') Trim_=2;
        when ('Aero Conv 2D') Trim_=3;
        when ('Arc Conv 2D') Trim_=3;
  when ('Arc Sedan 4D') Trim_=3;
        when ('Arc Wagon 4D') Trim_=3;
        when ('Conv 2D') Trim_=3;
        when ('Coupe 2D') Trim_=3;
        when ('DHS Sedan 4D') Trim_=3;
        when ('Hardtop Conv 2D') Trim_=3;
        when ('Linear Conv 2D') Trim_=3;
        otherwise Trim_=Trim;
        end;
        run;
data BlueBook;
        length Liter_ 8;
        set BlueBook;
        select (Liter);
        when ('1.6') Liter_=1;
        when ('1.8') Liter_=1;
        when ('2.2') Liter_=1;
        when ('3.1') Liter_=1;
        when ('3.4') Liter_=1;
        when ('3.5') Liter_=1;
        when ('3.8') Liter_=1;
        when ('2') Liter_=2;
        when ('2.3') Liter_=2;
        when ('2.5') Liter_=2;
        when ('2.8') Liter_=2;
        when ('3.6') Liter_=2;
        when ('4.6') Liter_=2;
        when ('5.7') Liter_=2;
        when ('6') Liter_=2;
        otherwise Liter_=Liter;
        end;
        run;

*Creating the model and diagnotics;
ods graphics on;
proc glmselect data=bluebook plot=all
outdesign(addinputvars)=Work.reg_design;
        class  Make_ Type Model_ Trim_ Cylinder Doors Liter_
Cruise Sound Leather / show
                param=glm;
        model Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather
                / showpvalues stats = adjrsq stats = AIC
stats = PRESS stats= RSQUARE stb selection=none;

run;

*Goodness of fit;
proc genmod plots=all plots=predicted;
class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather / ;
        model Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather  /
                ;


run;

 *Another way to view model diagnostic plot;
 ods graphics on;
 proc glm data=bluebook plots=(DIAGNOSTICS RESIDUALS) ;
 class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather;
```

```
            model Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather  ;
            run;
             ods graphics off;

 *diagnostic plot;
 ods graphics on;
Proc reg data=Work.reg_design alpha=0.05 ;
            ods select DiagnosticsPanel ResidualPlot
ObservedByPredicted ParameterEstimates OutputStatistics
ResidualStatistics CollinDiag
                        PartialPlot RStudentByPredicted
DFFITSPlot
                        DFBETASPanel;
            model Price=&_GLSMOD /  r stb influence collin vif;

run;
quit;

*remove two predictors for VIF;
proc reg data=Work.reg_design alpha=0.05 ;
            ods select DiagnosticsPanel ResidualPlot
ObservedByPredicted ParameterEstimates OutputStatistics
ResidualStatistics CollinDiag
                        PartialPlot RStudentByPredicted
DFFITSPlot
                        DFBETASPanel;
            model Price= Mileage Make__2 Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1/  r stb influence collin vif;
run;
quit;
             ods graphics off;
*RStudent residual;
proc reg data=Work.reg_design alpha=0.05 ;
                        model Price=Mileage Make__2 Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1/ ;
            plot rstudent.*(predicted. Mileage Make__2 Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 obs.);
    plot npp.*rstudent.;
run;
quit;


*correlation matrix;
proc corr data = work.reg_design;
var &_GLSMOD;
run;

*refit of correlation with out doors;
proc glmselect data=bluebook plot=all
outdesign(addinputvars)=Work.reg_designcor;
            class  Make_ Type Model_ Trim_ Cylinder  Liter_ Cruise
Sound Leather / show
                        param=glm;
            model Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_  Cruise Sound Leather

                        ;

run;


proc corr data = work.reg_designcor;
var Mileage Make__1 Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather
Cruise;

run;
ods graphics on;
*refit of residuals;
Proc reg data=Work.reg_design alpha=0.05 ;
            ods select DiagnosticsPanel ResidualPlot
ObservedByPredicted ParameterEstimates OutputStatistics
ResidualStatistics CollinDiag
                        PartialPlot RStudentByPredicted
DFFITSPlot
                        DFBETASPanel;
            model Price=Mileage Make__1 Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather
Cruise /  r stb influence collin vif;

run;
quit;

proc glm data=bluebook plots=(DIAGNOSTICS RESIDUALS) ;
 class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather;
            model Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather  ;
            run;
             ods graphics off;
ods graphics on;
*Running Box Cox Transformation of y with variables removed
and transformed;
*using the three new tranformation variables ;
proc transreg data = Work.reg_design;
            model boxcox(price)=identity(Mileage Make__1
Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
```

```
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather
Cruise  );
run;
                ods graphics off;

*Need to inverse x;
data Work.reg_design;
                set Work.reg_design;
                Tr_Price =Price**.25;
                log_Price=log(Price);
                Tr_price5=price**.5;
run;
data Bluebook;
set Bluebook;
log_Price=log(Price);
run;

*Transformation of y to Log(y);
 *Another way to view model diagnostic plot;
 ods graphics on;
 proc glm data=bluebook plots=(DIAGNOSTICS RESIDUALS) ;
 class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather;
                model log_Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather  ;

                run;
                 ods graphics off;
*Creating the model and diagnotics;
ods graphics on;
proc glmselect data=bluebook plot=all
outdesign(addinputvars)=Work.reg_design2;
                class  Make_ Type Model_ Trim_ Cylinder Doors Liter_
Cruise Sound Leather / show
                                param=glm;
                model Log_Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather
                                 / showpvalues stats = adjrsq stats = AIC
stats = PRESS stats= RSQUARE stb selection=none;


run;

*reviewing transformation correlation;
proc corr data = work.reg_design2;
var log_price Mileage Make__1 Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather
Cruise;

run;
 *RStudent residual check on transformed y and Mileage;
proc reg data=Work.reg_design2 alpha=0.05 ;
                                model Log_Price=Mileage Make__2
Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 / vif;
                plot rstudent.*(predicted. log_price Mileage Make__2
Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 obs.);
    plot npp.*rstudent.;
run;
quit;
ods graphics off;



 ods graphics on;
 proc glm data=bluebook plots=(DIAGNOSTICS RESIDUALS) ;
 class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather;
                model log_Price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather  ;

                run;
                 ods graphics off;

 ods graphics on;
 *weight down Make_;
 proc glm data=bluebook plots=(DIAGNOSTICS (unpack)) ;
 class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather;
                model Log_price=Mileage Make_ Model_  Trim_ Type
Cylinder Liter_ Doors Cruise Sound Leather
;
                weight Make_ ;
                ID Mileage Make_ Model_  Trim_ Type Cylinder Liter_
Doors Cruise Sound Leather ;
                output out=infl COOKD=CO COVRATIO=cov DFFITS=df
H=h RSTUDENT=r RESIDUAL=re ;
                run;
                QUIT;
                 ods graphics off;
*Weight down variables;
proc reg data=reg_design2 plots=(DIAGNOSTICS (unpack)) ;

model Log_Price=Mileage Make__2 Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1/ partial r p cli clm clb stb rsquare influence
collin ;
output out=infl COOKD=CO COVRATIO=cov DFFITS=df H=h
RSTUDENT=r RESIDUAL=re student=s ;

run;
quit;
```

```
*Correct issues with obs weights;
 ods graphics on;

proc reg data=Work.reg_design2 alpha=0.05
plots(label)=(diagnostics
                        residuals dffits dfbetas
observedbypredicted);

                        model log_Price=Mileage Make__2
Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 / partial r p cli clm clb stb rsquare influence
collin;


reweight rstudent.<-1.5 or rstudent.>1.5 /weight = 0.5;
refit;

output out=reg_design3;
run;
quit;
        ods graphics off;

*Check correlineatiry;
 ods graphics on;

proc glm data=Work.reg_design3 plots(label)=(DIAGNOSTICS
(unpack));
                        class  Make_ Type Model_ Trim_ Cylinder
Doors Liter_ Cruise Sound Leather;
                        model log_Price=Mileage Make__1
Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1;

run;
quit;
        ods graphics off;

        proc reg data=Work.reg_design3 alpha=0.05 ;
                        model log_Price=Mileage Make__1
Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1/ partial r p cli clm clb stb rsquare covb influence
collin vif;
        plot rstudent.*(predicted. log_Price=Mileage Make__1
Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 obs.);
    plot npp.*rstudent.;
run;
quit;
ods graphics on;
proc reg data=Work.reg_design3  ;
                        model log_Price=Mileage Make__1
Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1 ;
run;

*selection of subset;
proc glmselect data=work.reg_design3;
        class  Make_ Type Model_ Trim_ Cylinder Doors Liter_
Cruise Sound Leather / show
                        param=glm;
        model log_Price=Mileage Make__1 Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1
                        / showpvalues stats = adjrsq stats = AIC
stats = PRESS stats= RSQUARE stb selection=LASSO;



run;

*All possible regression method;
proc reg data=work.reg_design3;
  model log_Price=Mileage Make__1 Make__2 Make__3
Model__1 Model__2
Trim__1 Trim__3
Type_Hatchback Type_Sedan Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1  /
        selection=rsquare cp mse adjrsq aic bic best=3 b;
run;

*lasso order for regressor;
proc glm data=Work.reg_design3 plots=(DIAGNOSTICS (unpack));

                        model log_Price=Model__1 Make__1
Cylinder_8 Cylinder_4 Trim__3 Trim__1
Mileage  Type_Hatchback Type_Sedan Type_Wagon Make__3
Sound_0
Model__2  ;

run;
quit;
```

```
*selection of subset;
proc glmselect data=work.reg_design3;
            class  Make_ Type Model_ Trim_ Cylinder Doors Liter_
Cruise Sound Leather / show
                        param=glm;
            model log_Price=Mileage Make__2 Make__3
 Model__2  Model__3
 Trim__2 Trim__3
Type_Hatchback Type_Convertible Type_Coupe Type_Sedan
Type_Wagon
Cylinder_4 Cylinder_6 Cylinder_8
Sound_0 Sound_1
Leather_0 Leather_1
Cruise_0 Cruise_1
                         / showpvalues stats = adjrsq stats = AIC
stats = PRESS stats= RSQUARE stb selection=LASSO;


run;

*review estimates;
proc reg data=work.reg_design3;
*class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather  ;
                        model log_Price=Mileage Make__2
Make__3 Model__2 Model__3 Trim__2 Type_Hatchback
Type_Convertible Type_Wagon Cylinder_4 Cylinder_8/partial r p
cli clm clb stb rsquare covb influence collin vif
;
run;
quit;

*review estimates;
proc genmod data=work.reg_design3;
class  Make_ Type Model_ Trim_ Cylinder Doors Liter_ Cruise
Sound Leather  ;
                        model log_Price=Mileage Make__2
Make__3 Model__2 Model__3 Trim__2 Type_Hatchback
Type_Convertible Type_Wagon Cylinder_4 Cylinder_8/
;
run;
quit;

proc glmselect data=work.infl
outdesign(addinputvars)=Work.reg_design4;
            class  Make_ Type Model_ Trim_ Cylinder   Cruise
Sound Leather / show
                        param=glm;
            model Log_Price=Mileage Make_   Type Cylinder
Cruise  Leather
                         / showpvalues stats = adjrsq stats = AIC
stats = PRESS stats= RSQUARE stb selection=none;


run;

proc glm data=Work.reg_design4 plots=(DIAGNOSTICS
RESIDUALS);
                        class  Make_ Type Model_ Trim_ Cylinder
Doors Liter_ Cruise Sound Leather;
                        model Log_price=Mileage Make_ Type
Cylinder Cruise Leather;
                        output out=infl COOKD=CO COVRATIO=cov
DFFITS=df H=h RSTUDENT=r RESIDUAL=re student=s ;

run;
quit;
proc reg data=infl;
model Log_price=Mileage Make   Cruise Leather;
plot rstudent.*(predicted. Log_price Mileage Make   Cruise
Leather obs.);
plot npp.*rstudent.;
run;
```