Research Report: Analysis of used SUV prices for Australian SUV's brands: Toyota and Mitsubishi

Introduction

According to Straits Research report(https://straitsresearch.com/report/used-car-market) the global used car is expected to grow from USD 1933.76 billion in 2025 to reach USD 4338.97 billion by 2033, growing at a CAGR of 10.63% during the forecast period (2025-2033). The high cost of buying new vehicles is one of the reasons millennials prefer to buy used cars. Among types of cars sold SUVs are popular vehicle choice because of their driving ability in all terrains, roomy interiors making them ideal for families as well young travelers.

Our goal of the research is to find which explanatory variables impact the price of used SUVs so that we could build linear regression model to accurately predict the price of a used SUV. Our dataset consists of data on car prices for cars sold in Australia in the year 2023.

To get a representative sample of the used SUV market the data was filtered for Used cars and brands Toyota and Mitsubishi were chosen as they were the two largest by count in the dataset. The model year of the SUV was selected from 2013 to 2022 to cover cars which are relatively new to cars which are about 10 years old.

Dataset Details

The sample dataset used has information on car prices for cars sold in Australia in the year 2023. The original dataset has 16734 rows of car listings from various online platforms in Australia. The data was obtained by web scraping the various websites.
Link to original dataset:
https://www.kaggle.com/datasets/nelgiriyewithana/australian-vehicle-prices

Data Filtering and Cleanup

Following steps were performed (manually in excel) to get the relevant data:

- Filter the columns by values need for our analysis
    - BodyType = SUV
    - Car/Suv = SUV
    - UsedOrNew = USED
    - Year = 2013 to 2022
    - Remove rows with null values
- Checked count of SUV's sold by year and picked years 2013 to 2022 as they had the most cars sold and were most relevant.
- Changed format of fuel consumed form 10.4 L / 100 km to a numeric value 10.4 by removing units to make it a quantitative variable
- Checked to see the top 2 car brands by count for model year 2013 -2022 which turned out to be Toyota and Mitsubishi
- After filtering and cleanup for the above criteria we have dataset with **1161** rows of data.

Relevant Columns

The column information relevant to our analysis is shown below (after renaming):

Original Column Name	Renamed To	Description
Brand	brand	Name of the car manufacturer
Year	year	Year of manufacture or release
UsedOrNew	type	USED
Transmission	transmission	Type of transmission (manual or automatic)
DriveType	drivetype	Type of drive (front-wheel, rear-wheel, or all-wheel)
FuelType	fueltype	Type of fuel (petrol, diesel, hybrid, or electric)
FuelConsumption	fueleconomy	Fuel Economy: Fuel consumption rate (in litres per 100 km)
Kilometres	kilometers	Distance travelled by the car (in kilometres)
BodyType	bodytype	SUV
Doors	doors	Number of doors in the car
Seats	seats	Number of seats in the car
Price	price	Price of the car (in Australian dollars)

All other columns were deleted from the original dataset.

Objective:

- Use the statistical methods and formal inference to make inferences about population parameters from sample data
- Employ simple linear regression to develop a predictive model of used SUV price
- Verify underlying assumptions and perform transformation if needed and rerun regression
- Run Multiple Linear Regression with a goal to get a better model fit and identify significant explanatory variables.
- Use ANOVA and ANCOVA to analyze the effect of means of used SUV price for different groups of fuel type and the influence of different fuel types after adjusting for kilometers driven

Statistical Methods Used:

The following statistical methods are used and their results are analyzed and interpreted to arrive at conclusions, discuss limitations and suggest next steps.

- Coefficient of Correlation and Correlation Hypotheses Test
- Model 1: Simple Liner Regression
- Residual Analysis for underlying Assumptions
- Model 2: Simple Linear Regression after Transformation
- Model 3: Multiple Linear Regression
- Model 4: ANOVA
- Model 5: ANCOVA

Research Questions:

Note **: All the research questions and statistical analysis is for USED SUV prices for Australian Cars whose model years are from 2013 to 2022 and whose make is Toyota and Mitsubishi

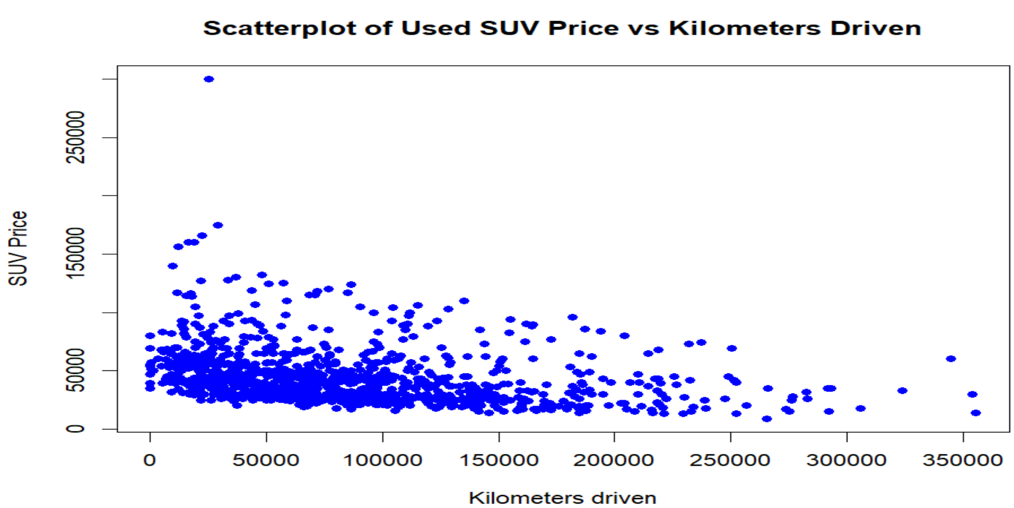
1. Is there a linear relationship between kilometers driven at the time of sale and price of a used SUV?
2. Can we quantify the linear relationship between kilometers driven and average used SUV price and use it to predict average used SUV price via simple linear regression?
3. Can we predict the average price of used SUV knowing the kilometers driven, fuel consumed in liters per 100 km and the brand of the SUV? Are we able to better predict the price using a combination of variables versus just kilometers driven? If yes which explanatory variables are significant in predicting average SUV price?
4. Does the population mean of used SUV price vary by fuel type (diesel, hybrid, premium and unleaded)?
5. After accounting for fuel economy will the SUV's having different fuel types still have a mean difference in used SUV price? Could the difference in fuel economy across fuel types be the associated with the mean differences in used SUV prices found across fuel types?

Analysis and Results

Q1) Is there a linear relationship between kilometers driven at the time of sale and price of a used SUV?

Check to see if there is a linear relationship between kilometers driven and price of the SUV by plotting a scatter plot and find the coefficient of correlation coefficient r. Interpret the strength and direction of the relationship

Using R code, we find the **correlation coefficient r** between kilometers driven and used SUV price is **r = -0.280134**



Inference of r from scatterplot and value of -0.28
The correlation coefficient r = -0.28 is quite closer to 0 than to 1 which indicates a weak negative correlation between kilometers driven and used SUV price, and hence a weak linear relationship. As $r < 0$ there is a negative association between the variables: kilometers driven and used SUV price. Since it is a negative association as the kilometers driven at time of sale increases the used SUV price decreases.

Output of running cor.test() function in R between kilometers and used SUV price at confidence level of 0.95 is shown below:

95% Confidence Interval	**-0.3323107**	**-0.2262477**

t = -9.9347	**df = 1159**	**p-value < 2.2e-16**

Inference to see there is a linear relationship between kilometers driven and price of the SUV in the population ρ (population correlation coefficient) using sample correlation coefficient r.

- From the above data we see that the p-value < 2.2e-16 is much less than 0.05 and is statistically significant.
- Hence, we can reject the null hypothesis that population correlation coefficient ρ between kilometers and used SUV price is equal to zero for alpha = 0.05.
- There is significant evidence of a linear relationship between kilometers and used SUV price in the underlying population.

Confidence Interval of ρ (population correlation coefficient)
From the data table above we can say with 95% confidence that the population correlation coefficient ρ for kilometers and used SUV price is between: -0.3323 to -0.2262

Q2) Can we quantify the linear relationship between kilometers driven and used SUV price and use it to predict SUV price via simple linear regression?

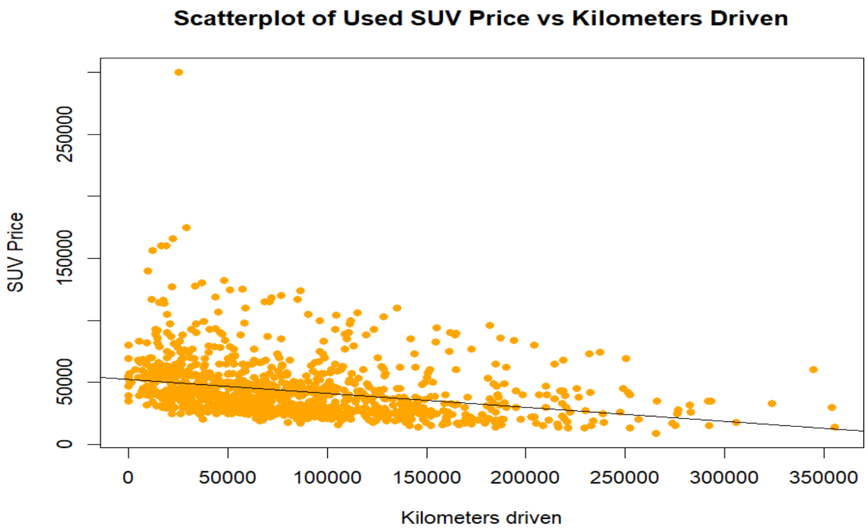
Model 1: Simple Linear Regression

We perform a Simple Linear Regression to quantify the relationship between kilometers (explanatory variable) and used SUV price(response variable) .
After Running R code to generate the SLR model and running the summary function we get the following output:

	Estimate	Std.Error	t-value	p-value	
(Intercept)	5.26E+04	1.15E+03	45.824	<2e-16	***
kilometers	-1.12E-01	1.13E-02	-9.935	<2e-16	***

Residual standard error	22480 on 1159 degrees of freedom
Multiple R-squared 0.07847	Adjusted R-squared 0.07768
F-statistic	98.7 on 1 and 1159 DF
p-value < 2.2e-16	

Equation for predicted value y hat for given input value x (kilometers) : $\breve{y} = 52600 - 0.112\,x$



Analysis and Interpretation of the results of SLR model

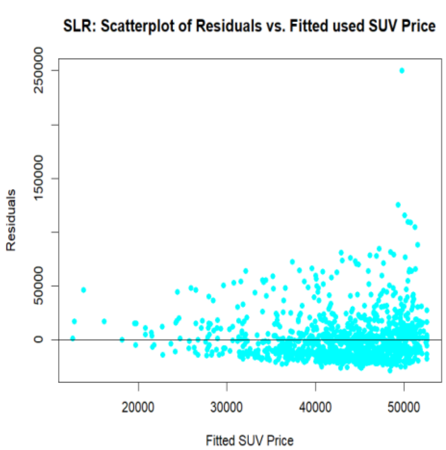
Fit of model: R-squared 0.07847
This means that 7.85% of the variability in the average used SUV price can be explained by kilometers driven. This is a very weak model.

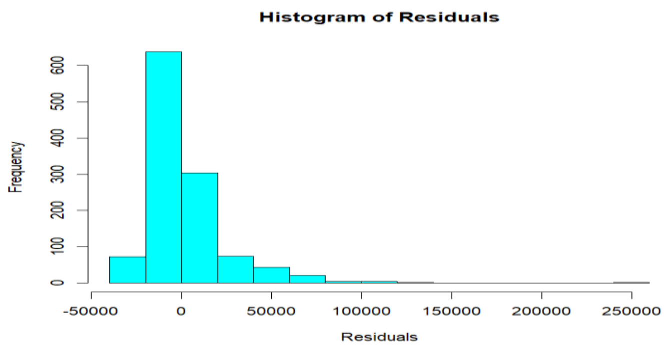
Interpreting beta0(intercept): 52600
The average SUV price for a car which has run 0 kilometers is 52,000 Australian dollars. This does not make sense in this context as used cars will have some kilometers driven already.

Interpreting beta1(slope): - 0.112
If the number of kilometers driven at the time of sale increase by 1 unit the predicted value of the average used SUV price decreases by 0.112 Australian dollars.

Inference about the population from SLR of our sample data
From the summary function we see that the beta coefficient for slope = <2e-16 is much smaller than 0.001 and is statistically significant. Hence, we can reject the null hypothesis that the population beta1 = 0. Thus, there is significant evidence of a linear relationship between kilometers driven and used SUV price at alpha =0 .01.

Analyzing Underlying Model Assumptions (Using Residual Plots and Histogram)





- **Linearity:** As the residual data points are randomly scattered with no correlation and pattern there is a linear relationship between years of kilometers and used SUV price.
- **Constant Variance:** We can see from the residual plot that the variance increase as the x value increases. Hence the assumption of constant variance is not met.
- **Normality:** The histogram of residuals seems to follow a weak normal distribution with skewness to the right. This assumption is also not met.

Conclusion (Underlying Model Assumptions)

Since the model assumptions of constant variance and normality of residuals is not met, we cannot use this model to predict for average used SUV price. The inference, prediction and interpretation of the regression equation (or correlation) are inefficient (at best) or misleading/biased/incorrect (at worst).

Solution: Applying Transformation

Since the variance of the response variable (used SUV price) is increasing as kilometers is increasing or we can apply the natural log ( ) to the response variable (used SUV price) to stabilize the variance.

Outliers and Influence Points:

Table below shows the top 6 rows with highest Cook's distance. We see that the cooks distance for all are less than 0.1 and hence their influence is minimum. This includes the outlier we see at kilometers = 344782 and kilometers = 25483 in the scatterplot of used SUV prices vs kilometers. Since their Cook's distance is < 0.1 no further investigation is needed.

ID	cooks distance	fitted values(ln(price))	kilometers	cd_fitted_price	cd_price
215	0.07489924	9.814768	344782	18302.04	59990
179	0.02530857	10.085341	250251	23988.77	68950
20	0.02339942	10.042522	265211	22983.26	8990
713	0.02325089	10.122874	237138	24906.25	73990
536	0.02071114	10.138101	231818	25288.41	72990
550	0.01695503	10.728688	25483	45646.78	299900

Model 2: Simple Linear Regression (Apply Transformation)

Rerunning the SLR model with natural log of used SUV price as response variable and kilometers as the predictor variable we get the following output from the summary function:

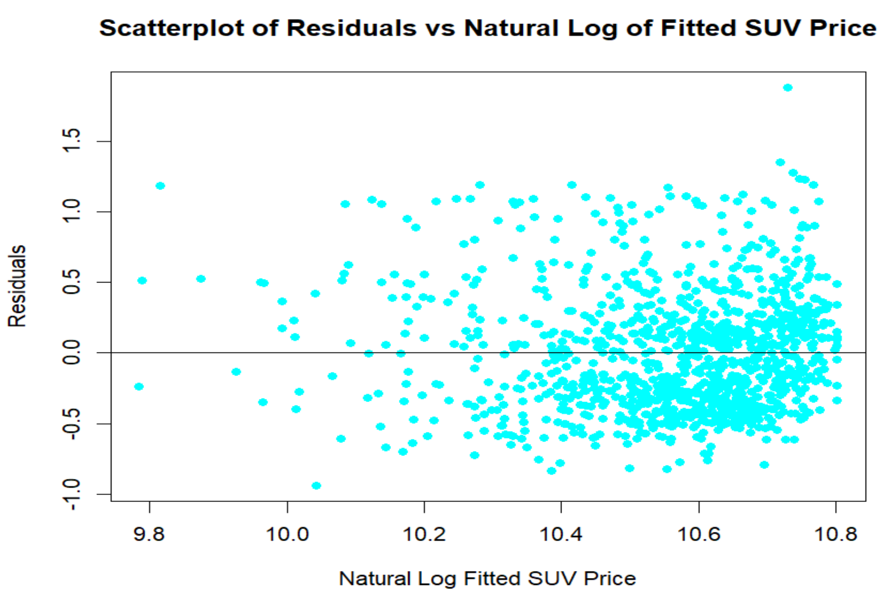
	Estimate	Std.Error	t-value	p_value	
(Intercept)	1.08E+01	2.15E-02	501.4	<2e-16	***
kilometers	-2.86E-06	2.13E-07	-13.47	<2e-16	***

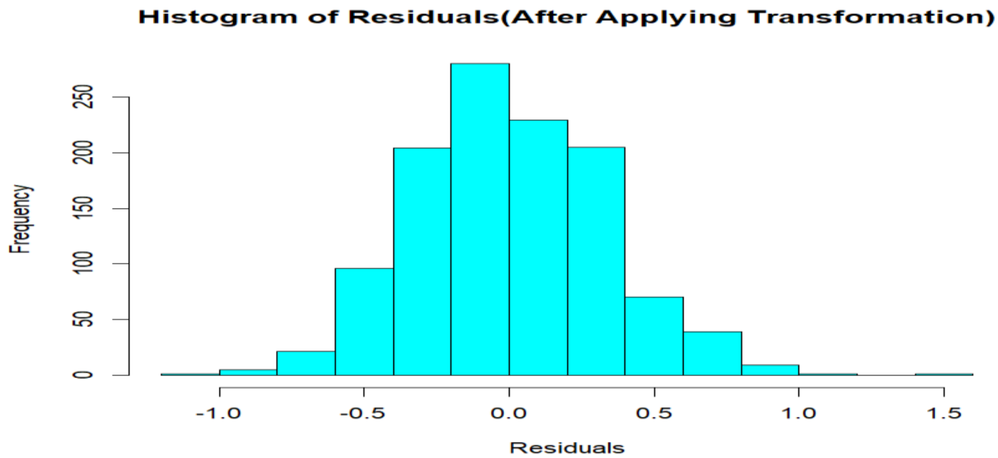
Residual standard error	0.422 on 1159 degrees of freedom
Multiple R-squared 0.1353	Adjusted R-squared 0.1346
F-statistic	181.4 on 1 and 1159 DF, p-value
p_value < 2.2e-16	

Inference

- We see that the p-value is very low and there are 3 stars indicating that there is a linear relationship between used SUV price and kilometers run in the underlying population.
- We have evidence that kilometers driven is significant in predicting mean used SUV price.
- Also, a R squared of 0.1353 indicates that 13.53% of the variability in used SUV prices can be explained by kilometers run.
- R square has increased to 13.53% from 7.85% after applying the transformation but it is still a weak fit.

Analyzing Underlying Model Assumptions: SLR after transformation (Using Residual Plots and Histogram)





- **Linearity:** As the residual data points are randomly scattered with no correlation and pattern there is a linear relationship between years of kilometers and used SUV price
- **Independence:** We can assume that the data was collected in an independent manner
- **Constant Variance:** The variation is now approximately constant across all values of x in the residual plot
- **Normality:** The histogram of residuals seems to follow a much better normal distribution with a little skewness to the right after applying the transformation.

Conclusion on Underlying Model Assumptions (After Transformation) :

All the model assumptions are now met, and the model can be used even though it is still a weak fit at R squared value of around 13.53%. The small value of R squared indicates that kilometers run is one of the many variables that the used SUV price depends on.

Q3) Can we predict the average price of used SUV knowing the kilometers driven, fuel consumed in liters per 100 km and the brand of the SUV? Are we able to better predict the price using a combination of variables versus just kilometers driven? If yes which explanatory variables are significant in predicting average SUV price?

Model 3: Multiple Linear Regression

Run a multiple linear regression with natural log (used SUV price) **as the response variable and kilometers driven, fuel economy in L/100km, and brand as the explanatory variables.**
(Convert categorical variable brand into dummy variable dbrand (Values = 1 for Toyota, Values = 0 for Mitsubishi)

After fitting the regression model that predicts natural log (used SUV price) from kilometers, fuel economy and running the summary function we get the following output:

	Estimate	**Std.Error**	**t**	**p-value**	**Pr(>|t|)**
(Intercept)	1.00E+01	4.75E-02	211.23	<2e-16	***
kilometers	-3.47E-06	1.78E-07	-19.53	<2e-16	***
fueleconomy	7.31E-02	6.47E-03	11.3	<2e-16	***
prices$dbrand	4.93E-01	1.97E-02	25.07	<2e-16	***

$\boldsymbol{\beta_{kilometers}}$ = -3.467e-06 (**Coefficient of: kilometers)**
$\boldsymbol{\beta_{fueleconomy}}$ = 7.311e-02 (**Coefficient of: fuel economy))**
$\boldsymbol{\beta_{brand}}$= 4.925e-01 (**Coefficient of: brand)**
$\boldsymbol{\beta_0}$= 1.002e+01 (**intercept of the equation)**

Equation for predicted value y hat(used SUV price) :
$\breve{y} = \mathbf{10.02} - 3.467\text{e} - 06\ \boldsymbol{kilometers} + .4925\ \boldsymbol{brand} + \mathbf{0.0711}\ \boldsymbol{fueleconomy}$

Global Statistics
F-statistic: 348.3 on 3 and 1157 DF
p-value: < 2.2e-16
Multiple R-squared: 0.4746

Interpretation of MLR model

Fit of model

R-squared = 0.4746 means 47.46% of the variability in the predicted value of average used SUV prices can be explained by the set of three predictor variables of kilometers, brand and fuel economy. There has been a significant improvement in the fit of the model as we progressed from simple linear regression to multiple linear regression. It has improved the R squared from 13.53% to 47.46% and the model is now a moderately good fit.

Inference about population parameter of average used SUV price (using Global Statistics at alpha = 0.05 level)

- Step1: Null Hypothesis
  - H0: $\boldsymbol{\beta_{kilometers}} = \beta_{brand} = \beta_{fueleconomy} = 0$
  - (Kilometers, brand and fuel economy are not significant predictors of used SUV price)
- Step1: Alternative Hypothesis
  - HA: (at least one of the slope coefficients is different than 0. Kilometers and/or brand and/or fuel economy are significant predictors of used SUV price)
- Step1: Alpha level = 0.05
- Step2: Get the global F-statistic and p_value from summary function
- Step3: Decision Rule
  - Find the appropriate F critical value for df1=3 and df2 =n-k-1 = 1157 degrees of freedom for alpha = 0.05 (Using R code – right tailed test)
  - **fcv = 2.612595**
- Step3: **Decision Rule**
  - Reject H0: If F Statistic >= **2.612595**
  - Otherwise, do not reject H0
- Step4: **Get F statistic and p value (summary function)**
  - **F-statistic: 348.3 on 3 and 1157 DF**
  - **p-value: < 2.2e-16**
- Step5: **Conclusion:**
  - **Since Global F statistic 348.3 > 2.612595 Reject H0**
  - We have significant evidence that at alpha = 0.05 that kilometers and/or brand and/or fuel economy are significant predictors of used SUV price
  - There is evidence of a significant linear association between the average used SUV price and kilometers and brand and fuel economy.
  - p value is also significantly less than alpha = 0.05 at p-value: < 2.2e-16 and is significant

Analysis of significant predictor variables

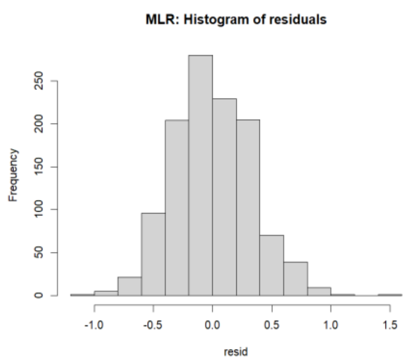
From the summary table we see that p_values of all individual predictor variables are <2e-16 which is much lower than <0.05. **Hence all the three predictors are significant in predicting the average value of used SUV price.**

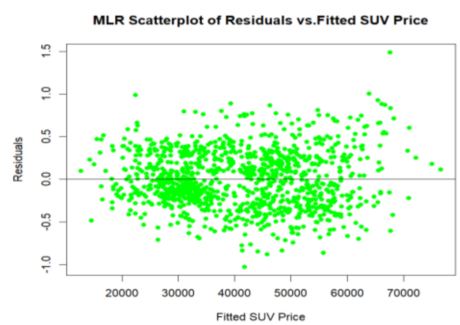
95% confidence intervals for any estimates that are significant

	95% Confidence Interval of Coefficients	
	2.50%	**97.50%**
(Intercept)	9.93E+00	1.01E+01
kilometers	-3.82E-06	-3.12E-06
fueleconomy	6.04E-02	8.58E-02
prices$dbrand	4.54E-01	5.31E-01

- **We can say with 95% confidence that the population $\beta_{kilometers}$ is between -3.82E-06 and -3.12E-06**
- **We can say with 95% confidence that the population $\beta_{fueleconomy}$ is between 6.04E-02 and 8.58E-02**
- **We can say with 95% confidence that the population β_{brand} is between 4.54E-01 and 5.31E-01**

Analyzing Underlying Model Assumptions MLR (Using Residual Plots and Histogram)





- Residual plots have constant variance and have random points scattered uniformly
- The histogram of residuals follows approximately normal distribution
- All model assumptions are met

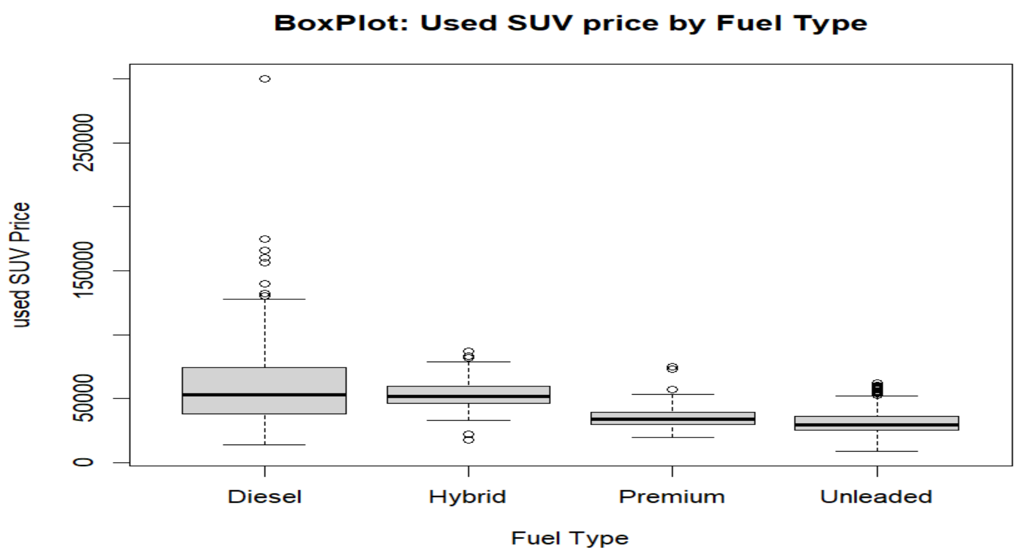
Q4) Does the population means of used SUV price vary by fuel type (diesel, hybrid, premium and unleaded)?

Model 4 – One Way ANOVA Model

Summary of used SUV Price by fuel type

	Count	Min	Q1	Median	Mean	Q3	Max	Sigma(sd)
Diesel	351	13985	37985	52480	59672.13	73999	299900	32086.32
Hybrid	161	17999	45990	51668	53332.45	59880	86850	11036.17
Premium	66	19800	29547.5	33894	35886.67	38990	74955	9830.915
Unleaded	583	8990	24999	28999	31408.73	35999	62412	9532.406

Box Plots of used SUV Price by fuel type



Observation and Analysis from summary table and boxplot

- the mean used SUV price of the premium (35886.67) and unleaded (35999) SUV are close to each other. The variability in the difference of their means is small and the variability within these two groups is similar. This suggests the underlying population may be the same for these groups. This makes sense as both run on gasoline and the only difference is the quality of gasoline
- The variability in means of used SUV price for Diesel (59672.13) an Hybrid (53332.45) is small while the within group variance of the Diesel(32086.32) group is almost 3 times higher than the hybrid group(11036.17).
- Even though the mean difference in used SUV price for hybrid-Diesel group is small the difference in within group variability suggest they belong to different underlying populations.
- Compared to the premium and unleaded fuel group the mean of used SUV price of the hybrid and diesel group are almost 5 times higher.
- The variance in SUV prices for hybrid is much closer to premium and unleaded groups than to the Diesel group

Result of one-way analysis of ANOVA at significance level of alpha = 0.05 is shown below

	Df	SumSq	Mean Sq	F-value	p-value	
prices$fueltype	3	1.96E+11	6.55E+10	172.5	<2e-16	***
Residuals	1157	4.39E+11	3.79E+08			

F-critical = 2.612595 (at alpha = 0.05(right tailed test) for 3 and 1157 degrees of freedom)

Interpretation and Inference of one way ANOVA result

Since the Global F statistic from the mode of 172.5 > 2.6125(F critical value) we can:

- Reject the null hypothesis that all underlying population means are equal by fuel type
- We have significant evidence at alpha = 0.05 that there is a difference in mean used SUV price between the different fuel types..
- The mean of used SUV prices scores for the underlying populations of different fuel types are not equal.
- **at p_value of (**<2e-16**) < 0.001** results of the model are significant

Since the results of the overall model are significant we perform pairwise comparisons using Tukey's procedure to adjust for multiple comparisons and summarize these results.

Tukey's Test Results(alpha = 0.05)

	Confidence interval at alpha = 0.05			
	Difference in mean	**Lower CI**	**Upper CI**	**p_adj**
Hybrid-Diesel	-6339.678	-11109.81	-1569.551	0.0036274
Premium-Diesel	-23785.459	-30509.09	-17061.826	0
Unleaded-Diesel	-28263.396	-31649.09	-24877.707	0
Premium-Hybrid	-17445.781	-24770.47	-10121.086	0
Unleaded-Hybrid	-21923.718	-26385.42	-17462.014	0
Unleaded-Premium	-4477.938	-10986.39	2030.512	0.2883273

From Tukey's test we can see that the adjusted p value is < 0.001 the all groups except the Unleaded-Premium fuel type pair. Hence, we have significant evidence that the underlying population mean of used SUV price for all pairs except the Unleaded-Premium groups are not equal (i.e. statistically significant).

Also, since the confidence interval of the difference in means of all group pairs except the **Unleaded-Premium** do not contain a zero all these groups are statistically significant and there is significant evidence to say that the mean of used SUV price of the underlying population pairs are not equal.

For the Unleaded-Premium the adjusted p-value is > 0.001 and hence the mean of the used SUV price is not statistically significant. We do not have significant evidence to say that the mean used SUV price of the underlying population of these two groups is not equal.

Q5) After accounting for fuel economy will the SUV's having different fuel types still have a mean difference in used SUV price? Could the difference in fuel economy across fuel types be the associated with the mean differences in used SUV prices found across fuel types?

One Way ANCOVA test

While doing the one-way ANOVA test, we found that there was significant evidence of mean differences of used SUV price Hybrid-Diesel, Premium-Diesel, Unleaded-Diesel, Premium-Hybrid, Unleaded-Hybrid groups.
Generally, fuel type and fuel economy are related, and a having a different fuel in the same kind of vehicle may change mileage. For example, SUV's which run on diesel or are hybrid will have a better fuel economy. Hence, we are interested to adjust for fuel economy in our ANOVA model and then see if the difference we see between the mean SUV price across fuel types still exist.
We use one-way ANCOVA methodology to adjust for the covariate fuel economy in our model.

Result of one-way analysis of ANCOVA is shown below:

Response:	used SUV Price	ANCOVA Type 3 TEST				
	Sum Sq	Df	F-value	p-value	value	Pr(>F)
(Intercept)	3.42E+09	1	9.937	0.001662	**	
prices$fueltype	2.22E+11	3	214.845	< 2.2E-16	***	***
prices$fueleconomy	4.08E+10	1	118.574	< 2.2E-16	***	***
Residuals	3.98E+11	1156				

Inference

Looking at the table above showing the results of the ANCOVA test which accounts for both fuel economy and fuel type we see that:

- P-value of values of both fuel type and fuel economy are < 0.01 and hence are statistically significant.
- We can infer that the mean differences in the one-way ANCOVA model were due to differences in fuel type even after adjusting for fuel economy

Least Square Means after adjusting for fuel economy:

This is the average mean difference in used SUV prices given the average fuel economy is same across all fuel types.

				Confidence Level = .95	
Fuel Type	LS Means	SE	df	lower.CL	upper.CL
Diesel	64150	954	1156	62278	66021
Hybrid	47114	1400	1156	44362	49866
Premium	31930	2130	1156	27743	36117
Unleaded	30878	715	1156	29476	32280

Inference:

From the table above we see that after accounting for the covariate fuel economy, the least square means:

- for **Diesel** group is 64150: increased after adjusting for fuel economy
- for the **Hybrid** group is 47114: decreased after adjusting for fuel economy
- for the **Premium** group is 31930: decreased after adjusting for fuel economy
- for the **Unleaded** group is 30878: decreased after adjusting for fuel economy

We see that the mean used SUV price for Diesel fuel group has increased after adjusting for fuel economy while it has decreased for the other groups. We can say that if the average fuel economy of all the three fuel types were the same then the mean of each of the groups would be the least square means noted above for each group.

Conclusion

The primary goal of this study was to find which explanatory variables impacted the price of used SUV so that we could build model to accurately predict the price of a used SUV.

Our first simple regression model (used SUV price vs kilometers (r =-0.28)) gave us a R squared of 7.85% (weak fit). The residual diagnostics showed that it did not satisfy constant variance and normality of underlying model assumptions. Our transformed simple regression model using natural log used SUV price vs kilometers satisfied all underlying assumptions and gave us a better fit of 13.53%. We also checked the p-values to confirm that there was evidence that the underlying population parameters of kilometers could be linearly associated with used SUV price.

Our multiple linear regression model (response variable – natural log (SUV price) with brand, kilometers and fuel economy as the explanatory variables improved the fit dramatically and gave us a R squared of 47.46% which is a moderately good fit. Our global F_test and analysis of p-value of each predictor variable determined that the predictor variables of brand, kilometers and fuel economy were significant.

Looking at mean difference in underlying population for different fuel saw that there was a mean difference for used SUV prices among all fuel types group pairs except the unleaded-premium. Our one-way ANCOVA test adjusting for fuel economy revealed that fuel type was still a factor in accounting for mean difference in used SUV price after adjusting for fuel economy. This indicates that since fuel type is a factor influencing the used SUV price adding it to our multiple linear regression model may improve the fit of the model.

This research shows that we have achieved a model fit of 47.46% in predicting the used SUV price and my adding more variables like fuel type to the multiple regression model it is possible to improve the fit.

Limitations

The scope of the research is limited by the following:

- The applicability of this study is only for used SUV’s in Australia and is cannot be applied globally
- The reliability of the data is not a hundred percent as it has been web scraped.
- The scope can be broadened to include price of all car types like sedans, vans etc.
- The study does not account for factors like regions within Australia, number of seats in SUV, transmission type, model year etc. which could have an impact on used SUV price
- Our research considered only 2 brands of car and model years from 2013 to 2022.
- Finally, the time limitation of the research study was a factor in the limiting the scope of the research

Many of these limitations can be overcome and This analysis can further be extrapolated to other brands and other types of cars following a similar statistical method for analysis.