DESCRIPTIVE ANALYTICS

COURSEWORK: BNM854

STUDENT ID: 220273547

TABLE OF CONTENTS

1. SUMMARY

This report delves into a comprehensive analysis of the second-hand Land Rover Evoque market within the B4 7DR region. Conducted on behalf of Car4all, a reputable car dealership, the study aims to offer valuable insights into the factors influencing the pricing of second-hand vehicles.

The primary goal of this study was to understand the impact of these variables on the pricing of second-hand Range Rover Evoque models. Using SPSS Statistics software, I conducted thorough visualizations and statistical analyses to reveal essential insights. The study highlighted influential factors such as age, mileage, and gearbox type on car prices. Moreover, I developed a statistical price calculator to aid potential buyers and sellers in making informed decisions. These findings empower individuals to navigate the second-hand car market confidently.

2. DATA COLLECTION

Our dataset, sourced from the reputable online platform "www.autotrader.co.uk," encompasses a diverse range of second-hand cars. This dataset allowed me to derive meaningful conclusions about market trends and pricing dynamics. I collected data on several important variables, including the year of the car, gearbox type, fuel type, engine size, the number of previous owners, distance driven (mileage), and price.

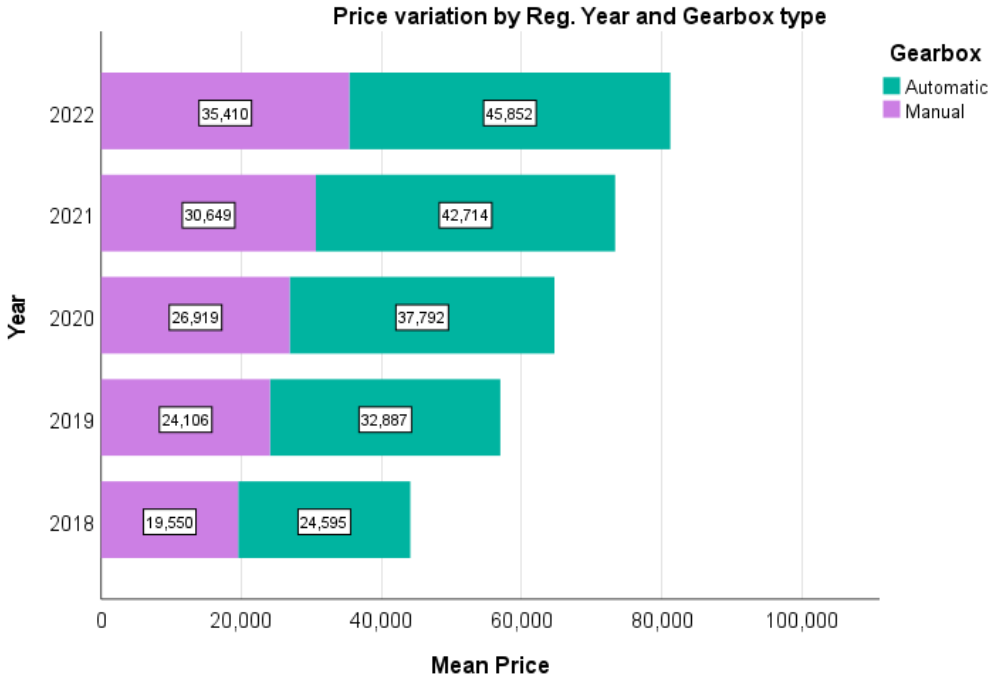
3. SAMPLING METHOD

The chosen sampling method was random sampling, which involves selecting a subset of data points from a larger dataset without any specific pattern or bias. In this case, a sample of 100 cars was selected from a dataset containing 312 cars. Random sampling was employed to ensure that each car in the dataset had an equal chance of being included in the sample, thus reducing the potential for selection bias.

4. VISUALISATION

Graph I

Bar Chart Showing Price Variation with reg. Year and Gearbox Type

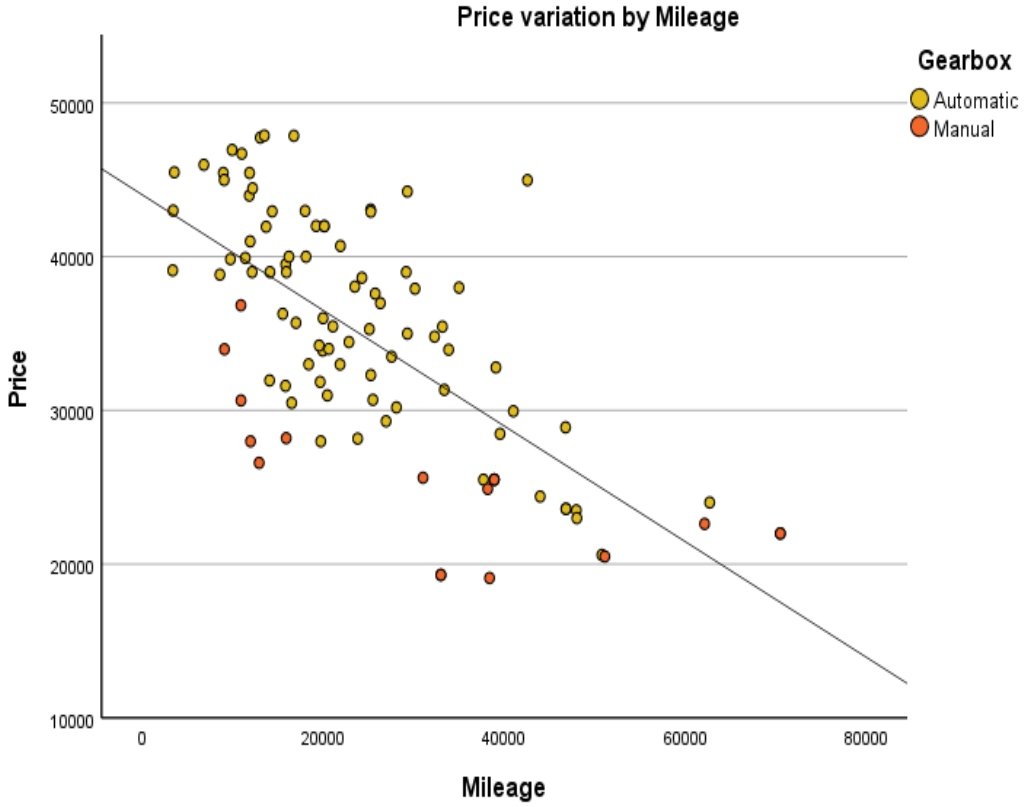


The bar chart shows a distinct trend and illustrates how prices vary depending on the gearbox type and registration year. It is clear that the average price has risen steadily over the course of the registration years, indicating a potentially sizable impact of the registration year on pricing.

The use of a bar chart is appropriate here as it facilitates a categorical comparison while effectively displaying the upward price trend over the years. The choice of colours for different gearbox types aligns with the principle of similarity, aiding viewers in distinguishing categories. This adherence to Gestalt principles enhances visual clarity, allowing viewers to easily comprehend the trend of increasing average prices over the years.

Graph II

Scatterplot Showing Variation of Price with Mileage

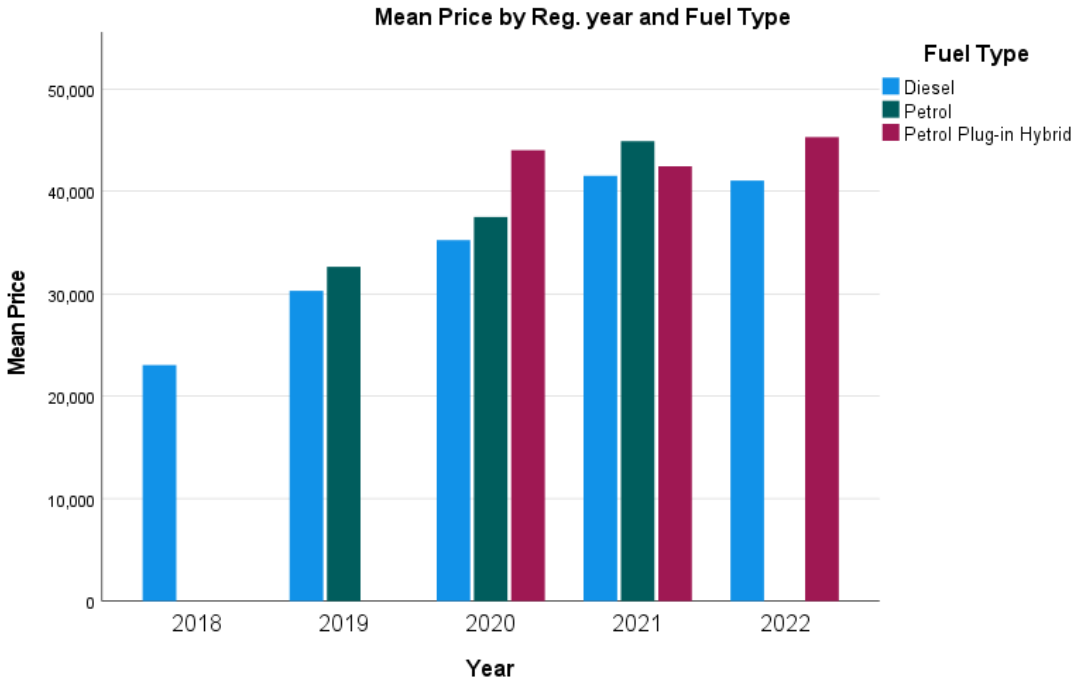


The Scatterplot illustrates the connection between the price and the mileage of second-hand Range Rover Evoque vehicles. Upon observation, a clear pattern emerges: cars with lower mileage tend to command higher prices, while cars with higher mileage generally have more affordable prices.

The choice of Scatterplot is suitable because it allows us to observe individual data points and the overall trend. The colour-coding of gearbox types follows the similarity principle, which enhances visual comparison. However, using additional labels for gearbox types improves clarity and interpretation.

Graph III

Bar chart showing mean price with reg. year and fuel type



The stacked bar chart showcases how average prices vary with fuel type and gearbox across different years of registration. Petrol and Hybrid Evoque cars have higher prices compared to diesel cars. The absence of petrol and hybrid cars in 2018 hints at evolving market trends. Hybrid cars' increasing presence since 2020 corresponds to higher average prices. The lack of petrol cars in 2022 suggests changing preferences.

The chart's colour coding efficiently differentiates gearbox types, enhancing clarity. The visualization aligns with Tufte's principles by focusing on essential information, ensuring clear communication.

5. DESCRIPTIVE STATISTICS

Descriptive Statistics is a fundamental branch of statistical analysis that aims to summarize and present the main characteristics of a dataset. It provides a concise and organized way to understand the distribution, central tendency, and variability of the data, making complex information more manageable and interpretable.

5.1. Standard Deviation

Standard deviation quantifies the amount of dispersion or spread in a dataset. It shows how individual data points differ from the mean. A larger standard deviation indicates greater variability, while a smaller one suggests less variability within the data.

5.2. Skewness

Skewness is a measure of the asymmetry of the data distribution. It indicates whether the data is skewed to the left (negatively skewed) or to the right (positively skewed) from the central tendency. A symmetrical distribution has a skewness close to zero.

5.3. Kurtosis

Kurtosis measures the shape of the distribution's tails and peak in comparison to a normal distribution. High kurtosis indicates heavy tails and a peaked distribution, while low kurtosis implies lighter tails and a flatter distribution.

Descriptive Statistics

	N	Range	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Price	100	28781	34523.40	7818.798	61133606.909	-.186	.241	-.940	.478
Mileage	100	67034	25289.76	14503.368	210347697.52	1.027	.241	.897	.478
Previous owners	100	2	1.19	.465	.216	2.476	.241	5.621	.478
Valid N (listwise)	100								

- **Price**

The data covers 100 cars with prices that range from £19,099 to £47,880. On average; second-hand cars are priced at approximately £34,523.40, with a standard deviation of £7,818.80. This deviation indicates that prices vary around the average. A skewness value of -0.186 suggests a slight left-leaning distribution, indicating a few cars with lower prices that affect the distribution. Kurtosis value of -0.940 indicates a distribution with moderate flatness compared to a normal distribution.

- **Mileage**

The dataset includes 100 cars, with mileage ranging from 3,415 to 70,449 miles. The average mileage is about 25,289.76 miles, and the standard deviation is 14,503.37 miles, reflecting variability in mileage. A skewness value of 1.027 indicates a slight right-leaning distribution, suggesting some cars with higher mileage that influence the shape.

- **Previous Owners**

Based on 100 cars, the average count of previous owners is 1.19, with a range of 1 to 2. The Skewness value of 2.476 signifies a pronounced right-leaning distribution, indicating a majority of cars with fewer previous owners.

Price Descriptive Statistics by Fuel Type

Price

FUEL TYPE	Mean	N	Std. Deviation	Range	Variance	Kurtosis	Skewness
Diesel	32975.85	74	7934.337	28781	62953707.416	-1.019	.034
Petrol	36374.29	17	5108.401	18966	26095764.846	.088	.690
Petrol Plug-in Hybrid	43751.56	9	2155.883	6060	4647831.528	-.610	-.856
Total	34523.40	100	7818.798	28781	61133606.909	-.940	-.186

- **Diesel Cars**
  - Diesel cars have an average price of approximately £32,975.85. The prices show some variation around the average, with a standard deviation of £7,934.34.
  - The price range for diesel cars spans from £19,066 to £28,781, indicating a diverse pricing spectrum.

- **Petrol Cars**
  - Petrol cars have an average price of about £36,374.29.
  - The prices exhibit less variation compared to diesel cars, with a standard deviation of £5,108.40.
  - The price range for petrol cars ranges from £18,966 to £28,781.
- **Petrol Plug-in Hybrid Cars**
  - Petrol plug-in hybrid cars have an average price of around £43,751.56.
  - The prices within this category show relatively low variation, with a standard deviation of £2,155.88.
  - The price range for these cars spans from £6,060 to £28,781.

Variation in Second-Hand Car Prices by Registration Year

Price

YEAR	Mean	N	Std. Deviation	Range	Variance	Kurtosis	Skewness
2018	23042.92	13	3114.039	9069	9697241.910	-.934	.251
2019	30837.97	30	5015.153	17504	25151762.516	-.802	-.301
2020	36388.90	31	5466.906	20085	29887056.624	-.729	-.430
2021	42079.26	19	4369.519	17216	19092700.538	1.127	-.848
2022	42868.71	7	5246.670	13900	27527546.905	-.258	-1.185
Total	34523.40	100	7818.798	28781	61133606.909	-.940	-.186

Understanding pricing variance between registration years offers important information about market patterns. The dataset, which spans the years 2018 through 2022, shows a steadily rising trend in price movement. This pattern shows an increase in perceived worth, which may be fuelled by features and technological improvements and drive demand and market appreciation.

Price Descriptive statistics by Gearbox type

Price

GEARBOX	Mean	N	Std. Deviation	Minimum	Maximum	Range	Kurtosis	Skewness
Automatic	36539.72	82	6833.908	20599	47880	27281	-.627	-.325
Manual	25337.94	18	4954.261	19099	36840	17741	.425	.800
Total	34523.40	100	7818.798	19099	47880	28781	-.940	-.186

The average price of automatic cars in the area is £36539.72 and manual cars are £25337.94 which clearly indicates that automatic cars have much resale value and demand in the market. And most cars available in the area for sale are automatic cars.

6. CONFIDENCE INTERVAL

The confidence interval indicates the level of precision in our estimate, taking into account the inherent variability in the data. This confidence interval helps us establish a reasonable price range that captures the likely variability in the market for second-hand cars of this make and model.

Confidence Interval = Mean ± (Z * (Standard Deviation / √Sample Size))

Confidence Interval = 34523.40 ± (1.96 * (7818.798 / √100))

Confidence Interval ≈ £32997.18 to £36049.62

Based on my analysis, I am 95% confident that the average second-hand car price for Range Rover Evoque falls within the range of £32,997.18 to £36,049.62.

7. HYPOTHESIS TESTING

Sample Average = £34523

National average = £35044 (*Used Land Rover Range Rover Evoque Price Guide, Average Prices, Average Mileage and Prices by Year*, n.d.)

Null Hypothesis (H0): The average price of used Evoque cars in the B4 7DR region is the same as the average price in the UK.

Alternative Hypothesis (Ha): The average price of second-hand cars in B4 7DR region is different from the average price in the UK.

One-Sample Test

Test Value = 35044

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Price	-.666	99	.254	.507	-520.600	-2072.02	1030.82

The one-sample two tailed t-test was conducted to determine whether the average price of the second-hand car in our dataset significantly differs from a test

value of £35,044, which represents the average price in the UK for Range Rover Evoque cars.

The corresponding p-value is 0.507 with an estimated t-value of -0.666 and 99 degrees of freedom. We lack sufficient data to reject the null hypothesis because the p-value exceeds the usual significance level of 0.05. **The null hypothesis is accepted**. This indicates that the average price of the used car in the B4 7DR region is similar to the average price in UK.

8. CORRELATION ANALYSIS

Correlation analysis investigates the strength and direction of correlations between variables in a dataset, in order to determine how changes in one variable may be related to changes in another, and how the independent variables (Fuel type, Gearbox, Age, Mileage etc) affect the dependent variable (Price).

Correlations

		Price	Mileage	Age	Fuel Type	Engine Size	Gearbox	Previous Owners
Price	Pearson Correlation	1	-.699**	-.774**	.132	-.373**	.553**	-.049
	Sig. (2-tailed)		<.001	<.001	.192	<.001	<.001	.627
	N	100	100	100	100	100	100	100
Mileage	Pearson Correlation	**-.699****	1	.654**	-.083	.180	-.290**	.038
	Sig. (2-tailed)	<.001		<.001	.412	.073	.003	.707
	N	100	100	100	100	100	100	100
Age	Pearson Correlation	**-.774****	.654**	1	-.235*	.379**	-.137	.090
	Sig. (2-tailed)	<.001	<.001		.019	<.001	.174	.373
	N	100	100	100	100	100	100	100
Fuel Type	Pearson Correlation	.132	-.083	-.235*	1	-.674**	-.074	-.106
	Sig. (2-tailed)	.192	.412	.019		<.001	.462	.292
	N	100	100	100	100	100	100	100
Engine Size	Pearson Correlation	**-.373****	.180	.379**	-.674**	1	-.147	.054
	Sig. (2-tailed)	<.001	.073	<.001	<.001		.143	.596
	N	100	100	100	100	100	100	100
Gearbox	Pearson Correlation	**.553****	-.290**	-.137	-.074	-.147	1	.080
	Sig. (2-tailed)	<.001	.003	.174	.462	.143		.429
	N	100	100	100	100	100	100	100
Previous Owners	Pearson Correlation	-.049	.038	.090	-.106	.054	.080	1
	Sig. (2-tailed)	.627	.707	.373	.292	.596	.429	
	N	100	100	100	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

8.1. Price and Age

There's a strong negative correlation between price and age (-0.774). This means that older cars generally tend to have lower prices.

8.2. Price and Fuel Type

There's a weak positive correlation between price and fuel type (0.132). This suggests that certain fuel types might have slightly higher prices on average.

8.3. Price and Engine Size

There's a moderate negative correlation between price and engine size (-0.373). This implies that cars with larger engine sizes might have lower average prices.

8.4. Price and Gearbox

There's a strong positive correlation between price and gearbox type (0.553). This indicates that certain gearbox types might be associated with higher prices. i.e. Automatic gearbox cars have higher prices compared to manual gearbox cars.

8.5. Price and Previous Owners

There's a very weak negative correlation between price and previous owners (-0.049). The correlation is not significant, suggesting that the number of previous owners might not strongly influence the price.

9. REGRESSION ANALYSIS

Regression analysis is a statistical technique used to examine the relationships between one or more independent variables and a dependent variable. This analysis aimed to provide insights into how variables like age, mileage, gearbox type, fuel type, previous owners and more contribute to determining the price of these cars.

The first step involved data pre-processing, including transforming categorical variables into numerical equivalents for compatibility with the analysis. After preparing the data, I proceeded to build a regression model that could predict car prices based on the chosen independent variables.

The process involved iterative steps of selecting and refining variables to achieve a parsimonious model. I used a backward elimination approach, removing variables with higher p-values to achieve a model with optimal predictive power while minimizing unnecessary complexity. This stepwise process helped identify the most influential variables and led to the most parsimonious model. The standardized coefficients provided insights into both the significance and direction of the relationships between the independent variables and car prices.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	43574.943	1351.199		32.249	<.001	40892.833	46257.052
	Age	-4140.548	399.657	-.592	-10.360	<.001	-4933.861	-3347.236
	Mileage	-.103	.032	-.190	-3.220	.002	-.166	-.039
	Gearbox	8438.029	914.740	.417	9.225	<.001	6622.284	10253.774

a. Dependent Variable: Price

- **Age**

The variable "Age" holds considerable significance in affecting the price of second-hand cars, as indicated by its low p-value (< 0.001). We see a significant negative impact of age on car pricing with a standardised coefficient of -0.592. The car price is predicted to reduce by -4140.548 units for every unit rise in the standardised "Age". This highlights the significant impact that an automobile's age has on its resale value.

- **Mileage**

The variable "Mileage" is highly significant in determining car price (p-value = 0.002). The standardised coefficient of -0.190 indicates a clear inverse relationship between price and mileage. The car price is predicted to reduce by -0.103 for every unit increase in the standardised "Mileage". This reaffirms the importance of mileage as a key factor impacting the pricing of used cars.

- **Gearbox Type**

The variable "Gearbox" also demonstrates a significant impact on price (p-value < 0.001). We discover a significant positive association between gearbox type and automobile price with a standardised coefficient of 0.417. As seen by a larger

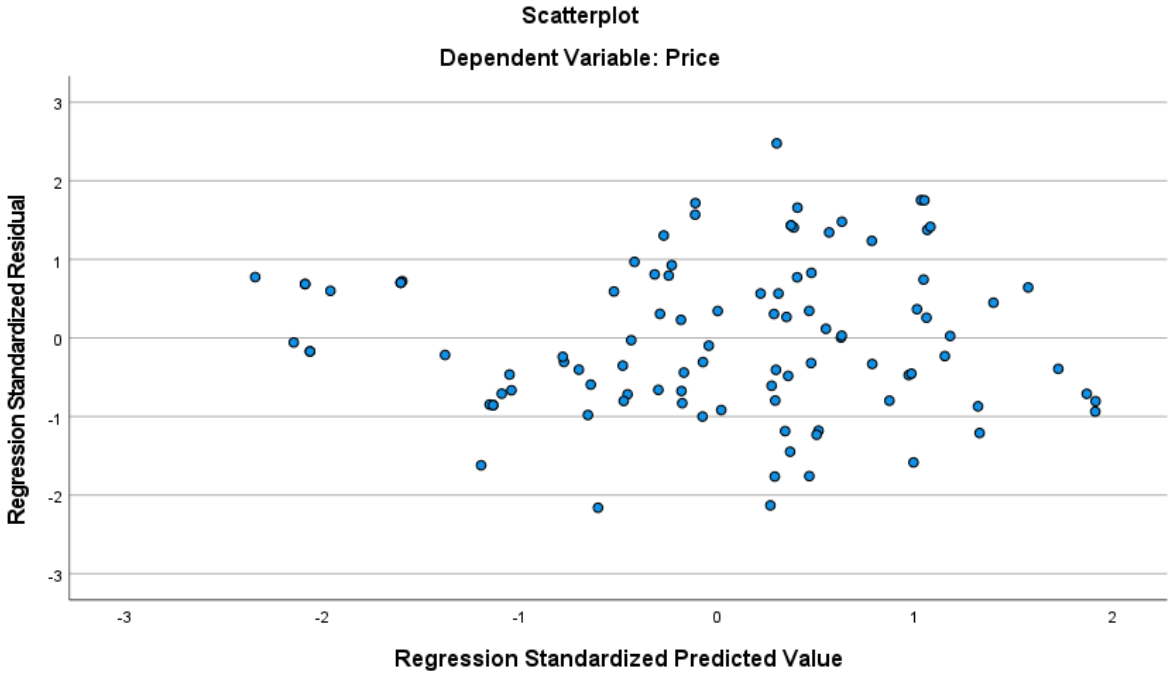
coefficient, automatic gearbox cars typically cost more than their manual equivalents. This highlights the significant impact gearbox type has on pricing choices.

10. RESIDUAL ANALYSIS

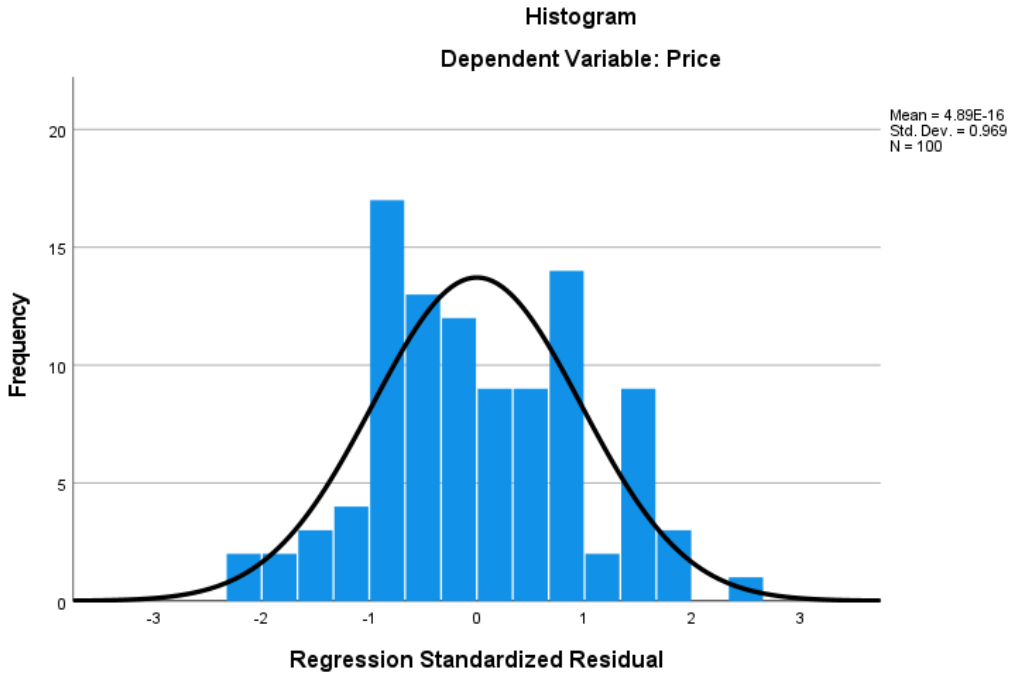
Residual analysis was performed to assess the validity of assumptions underlying the multiple linear regression models. The analysis encompassed four critical assumptions:

a. Average of residuals is zero: The assumption is valid, fig. (a) which shows that the scatter diagram contains roughly equal numbers of residuals above and below zero.
b. Residuals are independent: There isn't any clear pattern in the scatter diagram, therefore the assumption is valid.
c. The standard deviation of the residuals is the same for all the values of estimated dependent variable. The assumption is valid since there isn't any clear pattern of increase or decrease in residuals in the diagram as the estimated values of dependent variable increases.
d. There is no multi-co linearity in the model.

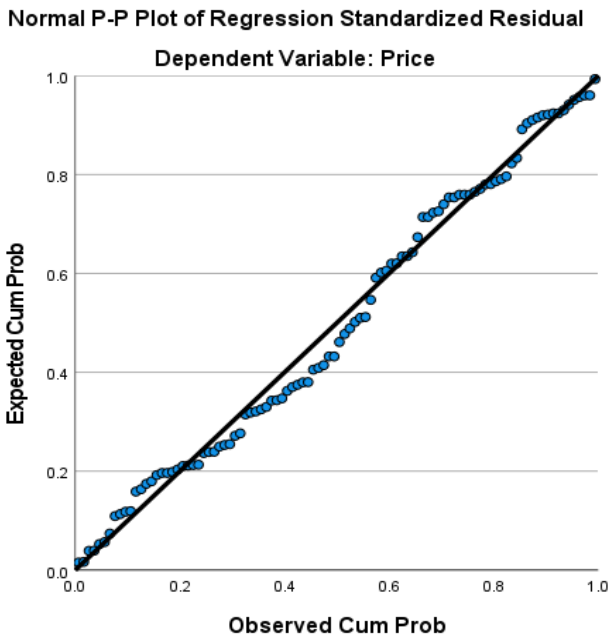
(**Figure A)**



e. Residuals are normally distributed: It is evident from fig(b) Histogram and fig(c)Normal P-P plot that the assumption is valid. The histogram is normally distributed and the normal P-P plot is also



(**Figure B)**



(**Figure C)**

Since all assumptions are valid, the model is **adequate**.

11. GOODNESS OF FIT

Goodness of Fit measures the extent to which a regression model accurately explains the variation in the dependent variable, often indicated by metrics like the coefficient of determination (R-squared) and the F-test's significance level.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.906[a]	.822	.816	3353.972

a. Predictors: (Constant), Gearbox, Age, Mileage

b. Dependent Variable: Price

The results demonstrate that the chosen model is highly significant ($p < 0.001$), with adjusted R-squared value of **0.816**. This adjusted R-squared value indicates that around **81.6%** of the variance in car prices could be explained by the included variables.

12. DERIVED STATISTICAL MODEL

The different elements that affect a car's pricing are mathematically represented in this equation. For instance, the price of the car drops by about £4140.55 when its age increases by one year. In a similar vein, the cost drops by £0.103 for each unit of increased mileage. The price is also impacted by the gearbox type, with automatic Gearbox types seeing an increase of £8438.03.

$$Price = 43574.943 – 4140.548 * Age – 0.103 * Mileage + 8438.029 * Gearbox$$

12.1. Example Usage

Suppose we have a 3-year-old car with a mileage of 30,000 miles and equipped with an automatic gearbox. By plugging in these values into the model:

Price = 43574.943 – 4140.548 * 3 – 0.103 * 30000 + 8438.029 * 1

Price ≈ £31994.00

The model predicts that the estimated price of this car would be approximately $31994.00. This derived statistical model provides a valuable tool for estimating car prices based on specific characteristics, aiding both buyers and sellers in making informed decisions.

REFERENCES

New & Used Cars for Sale - Auto Trader UK. (N.D.). New & Used Cars for Sale -Auto Trader UK. https://www.autotrader.co.uk/car-search?journey=FEATURED_LISTING_JOURNEY&sort=relevance&advertisinglocation=at_cars&make=Land%20Rover&model=Range%20Rover%20Evoque&postcode=b4%207dr&radius=35&year,from=2018&flrfc=1#202307039232909.

Used Land Rover Range Rover Evoque Price Guide, Average Prices, Average Mileage and Prices by Year. (N.D.). Used Land Rover Range Rover Evoque Price Guide, Average Prices, Average Mileage and Prices by Year. *https://www.carsite.co.uk/used-car-price-guide/land-rover/range-rover-evoque*.

Evans, J.R.,*Business Analytics. Methods, Models and Decisions*. (2016). Part 3, Chapter 8, pp. 275-289