CANDIDATE NUMBER : 593324

DESCRIPTIVE ANALYTICS REPORT: ANALYSIS OF SECOND-HAND CAR MARKET

**INTRODUCTION:**

This article looks at the second-hand car market in the United Kingdom with special focus on the somewhat reasonably priced and essentially useful Ford Fiesta for many drivers. The major objectives of this work are to develop a statistical model able to estimate the market worth of second-hand Ford Fiestas depending on mileage, engine size, vehicle age, additional equipment, and identify which aspects most significantly influence their pricing.

This study contrasts local market price with more general national trends using a properly chosen sample of approximately 100 used Ford Fiesta cars. Knowing these pricing policies influences consumers seeking appropriate discounts, auto dealers aiming at competitive pricing, and market analysts evaluating consumer preferences..

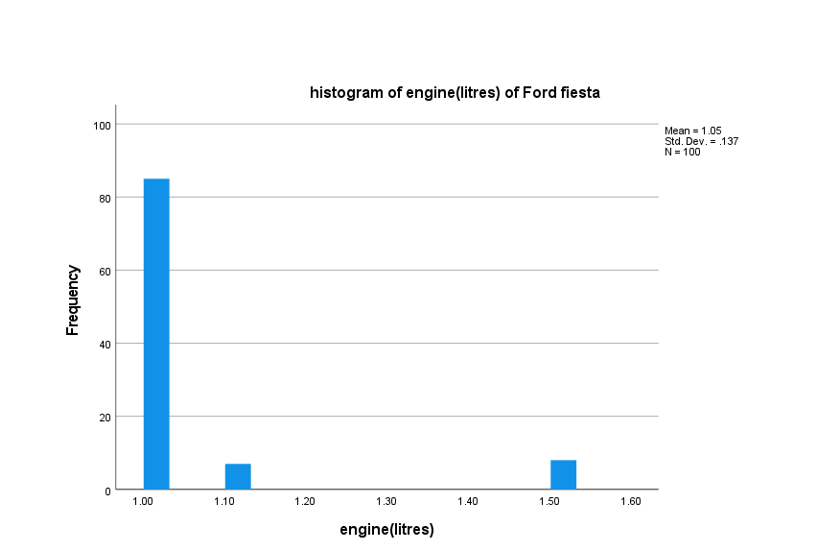
**PROBLEM DESCRIPTION AND DATA SOURCE**

Study data was gathered from Auto Trader, the UK's most prominent and widely used online car selling and buying website. The data was accessed through web scraping methods that provided an extensive and effective gathering of the pertinent car details. Cars considered for analysis were restricted to newer models registered from 2020 to 2023 to provide timeliness and relevance. Although incomplete data in the Auto Trader entries was a collection limitation, its impact is occasional and limited in scope. To be specific, some entries failed to provide adequate information on extra features or background history, i.e., services or if mileage was guaranteed. This occasional data loss minimizes the level of analysis slightly but does not adversely affect the representative nature of collected data.

Variables included in this study were chosen based on their proven effect on automobile prices, daily relevance to consumers, and availability in typical listings. Mileage, engine size, registration year, transmission, and fuel were specifically included because industry convention and contemporary market research routinely cite these as the primary determinants of used car prices. The sample of around 100 Ford Fiesta vehicles was chosen to reflect representative local market transactions carefully, so that results could be reliably generalized to comparable second-hand car markets. Omitting accident-damaged cars and demonstration models also guaranteed the accuracy and representativeness of the dataset by excluding vehicles whose prices would differ considerably from general market tendencies.

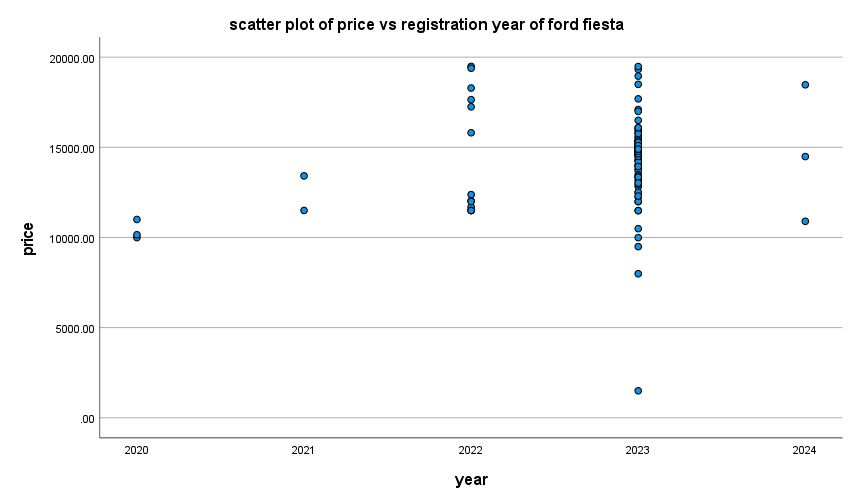
**DATA VISUALISATION:**

* **Histogram of engine (litres) of ford fiesta**

****

The majority of Ford Fiestas, as the histogram indicates, have a 1.0-litre engine, clearly the most popular among consumers because of its price and fuel efficiency. There are also some with marginally larger 1.1-litre engines, maybe special editions with marginal performance enhancements. There is also a very small group with much larger 1.5-litre engines, maybe for consumers who need greater power. Average, but with engine size around 1.05 litres, the Ford Fiestas appear to be thrifty cars for every daily need, but there are some options available for drivers who would want to have a bit of power.

* **Scatter Plot Of Price Vs Registration Year Of Ford Fiesta**

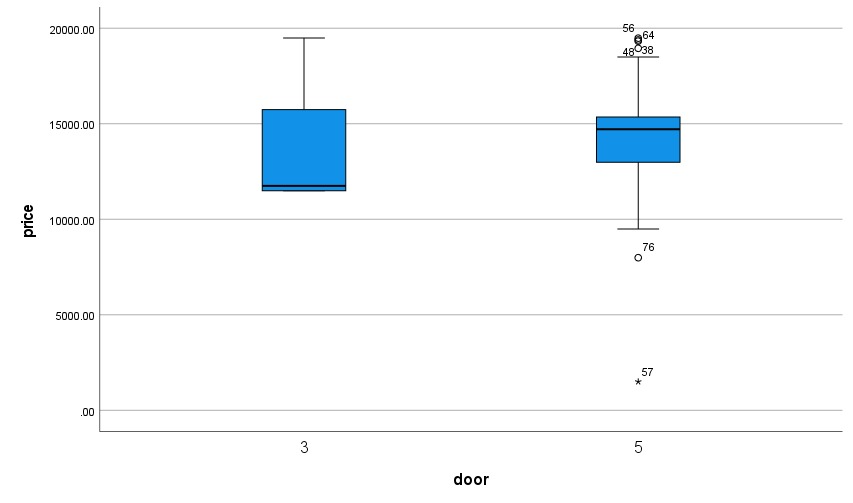


The scatter plot also shows clearly that the newer-registered Ford Fiesta vehicles, particularly the 2023-registered vehicles, are more expensive. This shows there is high demand for new-registered vehicles because they are newer, lower mileage, and newer features are introduced. The reverse is the case for older versions of 2020 and 2021 at lower prices and lower points, showing the likely depreciation as the vehicles get older and are no longer appealing to potential customers.

Surprisingly, the 2022 vehicles have less data point than the 2023 vehicles yet have high price volatility. Volatility can be attributed to volatility in the cars' state, mileage, upgrades done, or even the promotion strategy used. This creates the effect of such that, although the cars are very new, customers can notice plenty of value volatility in these cars.

Generally, the scatter plot indicates new cars to regulate market prices and supply while the prices for old cars decrease, happen less frequently, and are less attractive to purchasers

* **Boxplot Of Price By Door**

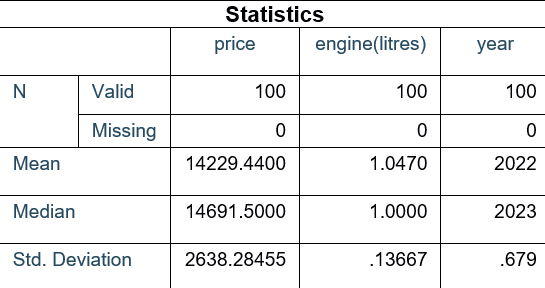


The box plot is contrasting the prices of Ford Fiestas by the doors. Three doors are less expensive and more dispersed, ranging from around $10,000 to $16,000. Five doors are more expensive and less dispersed, ranging from around $14,000 to $16,500, since they are both practical and in demand.

Five-door models will be more likely to be purchased by consumers, as one can see through their lower varying prices, while three-door models offer affordability with more scope for price variation.

**ADVANCED STATISTICAL ANALYSIS:**

* **Discriptive Statistics:**

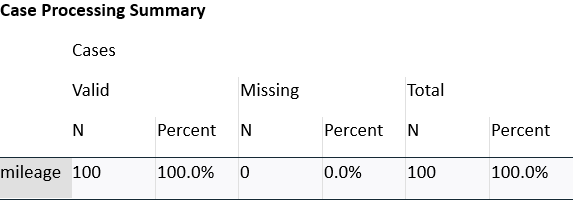
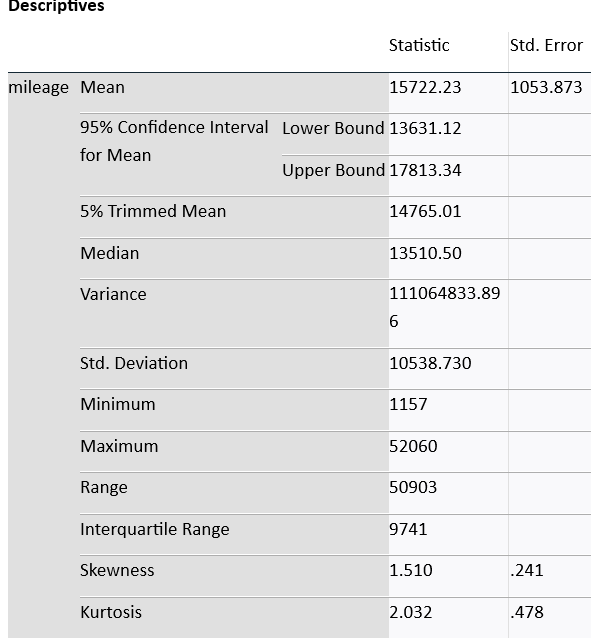


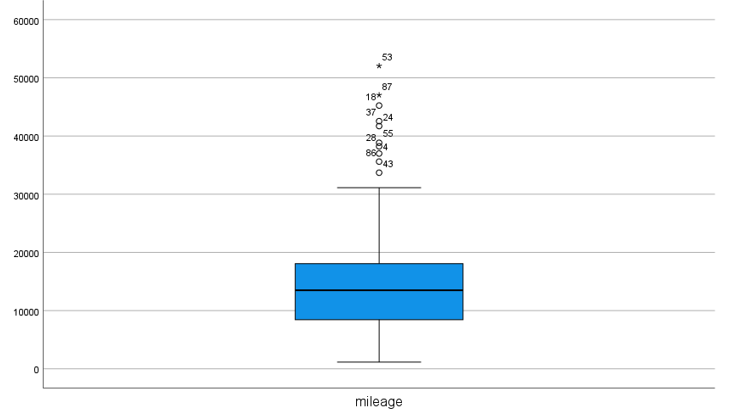
The following is the primary description of 100 Ford Fiestas. The median of approximately $14,692 is greater than the mean of around $14,229, suggesting the inclusion of some lower-priced cars that lower the mean. The range (standard deviation of $2,638) suggests difference in equipment, years, or condition.

Engine sizes also cluster at about 1.05 litres, just under the mode of 1 litre, reflecting the large number of cars being powered by these small, efficient engines.

The registration age is around 2023 on average, with the sample consisting predominantly of newer cars, thus justifying the perceived demand for newer vehicles.

* **Confidence Interval For Mileage:**





The descriptive statistics and the box plot show that the Fords Fiestas, on the whole, run approximately 9,000 up to 19,000 miles, with the median being approximately 13,500 miles. Some of the cars will run much higher, most likely from a lot of usage.

The average is approximately 15,722 miles, but this is artificially inflated by the high-mileage outliers, as indicated by the apparent skew. Most cars are fairly low-to-moderate mileage, indicating the usual usage patterns, with the higher-mileage vehicles making up a minority forming the bulk of the pulling the average up.

**HYPOTHESIS TESTING**

**INDEPENDENT SAMPLE T-TEST**

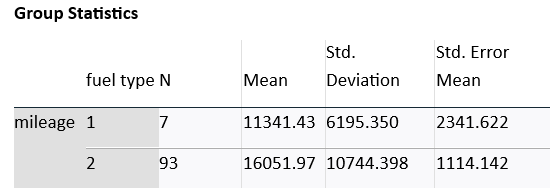
**Null Hypothesis (H₀):**

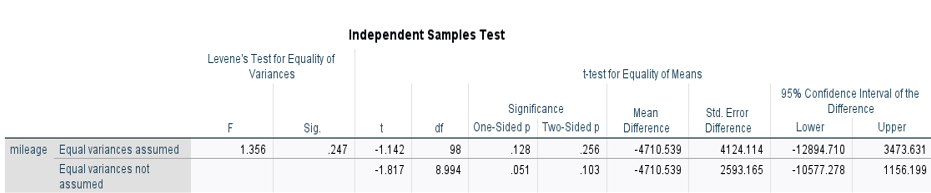
There is no significant difference in the average mileage between different fuel types.

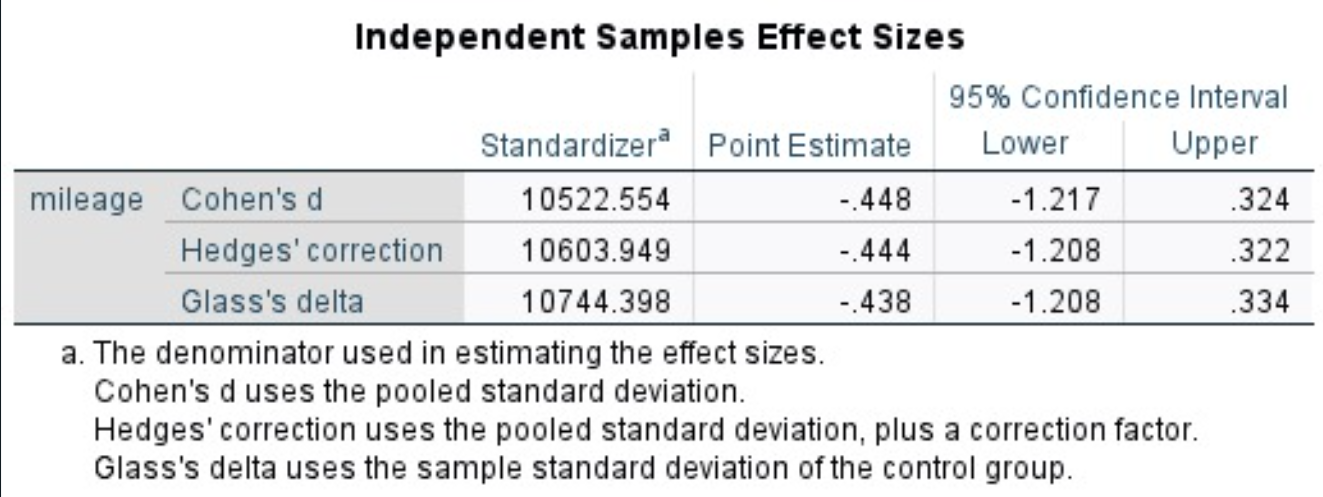
**Alternative Hypothesis (H₁):**

There is a significant difference in the average mileage between at least two fuel types.

**T-TEST**





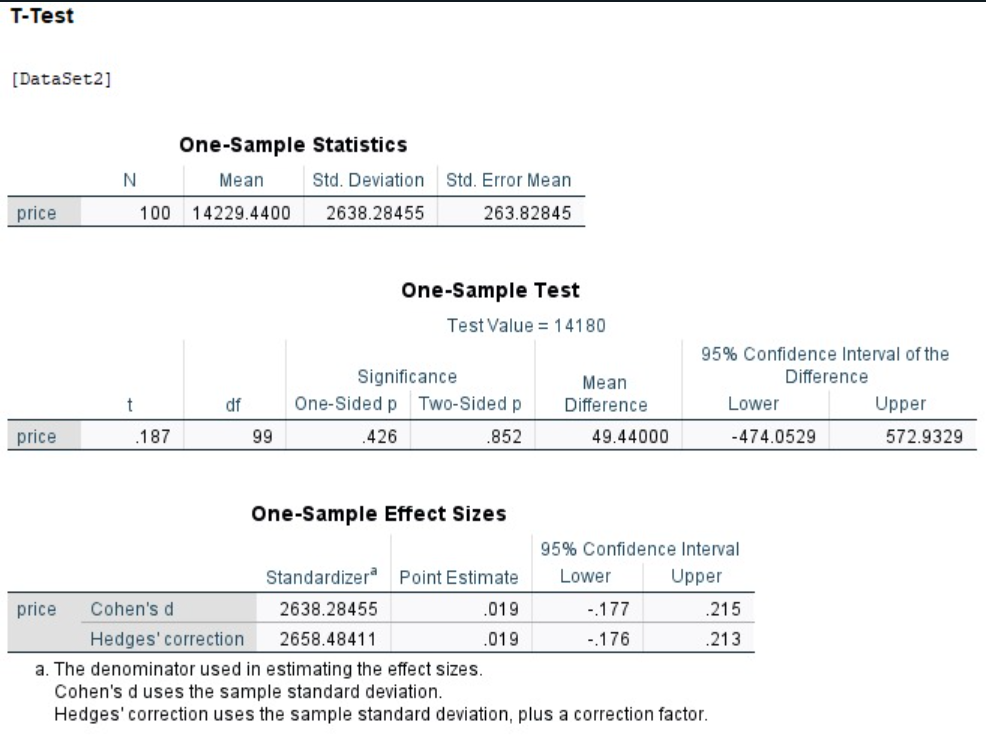


This is a test of if two fuel types of Ford Fiesta vehicles would be significantly different from each other. Cars of fuel type 1 had the lower average of about 11,341 miles, whereas the cars of fuel type 2 had the higher average of about 16,052 miles. At face value, this indicates cars of fuel type 2 would likely be driven more.

However, the independent samples t-test p-value was 0.256, demonstrating that the difference on mileage is not statistically significant. With only 7 of the cars of fuel type 1, however, we can't know whether fuel type really does any difference for the mileage. Even in the absence of statistical significance, the effect size (Cohen’s d of -0.448) shows a practical difference of small to medium size. That shows a tendency for the cars of fuel type 2 being of higher mileage, although further data would be required to know if the tendency is actually significant.

**COMPARISON WITH MARKET TRENDS:**

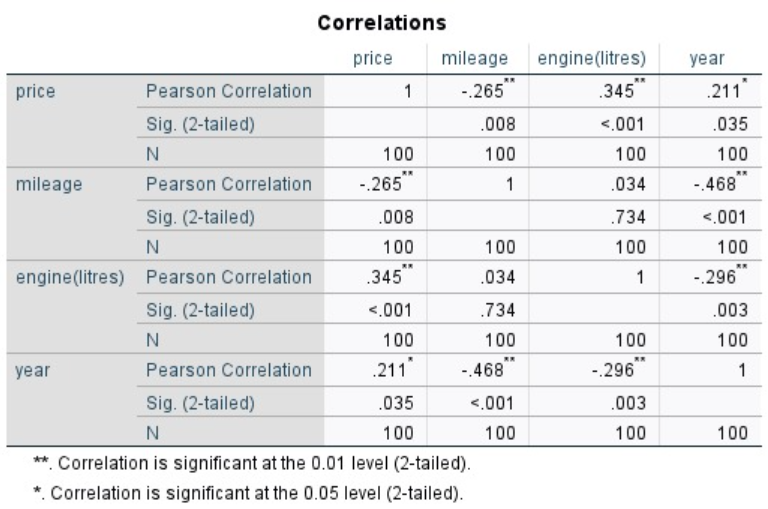
This compares local prices of Ford Fiestas for the sample with more generalised prices across the whole of the UK. It is useful for establishing if local dealership prices conform with the general market conditions or if prices deviate considerably.



The one-tailed test for a difference from a known population is used here to find if the local sample of the Ford Fiesta vehicle with a mean of $14,229 is statistically dissimilar from the national level of $14,180. The 0.852 p-value indicates we find no statistical difference here. Our local prices would be incredibly uniform with national averages if this is the case.

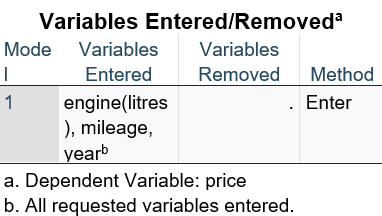
Also, the effect size is extremely small (Cohen’s d = 0.019), also supporting the fact that whatever difference exists among the prices is of no practical significance. Therefore, we can conclude with full assurance that locally the prices of the Ford Fiesta cars conform with the overall market tendency of the UK.

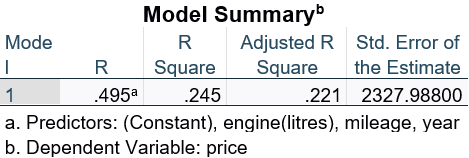
**CORRELATION:**

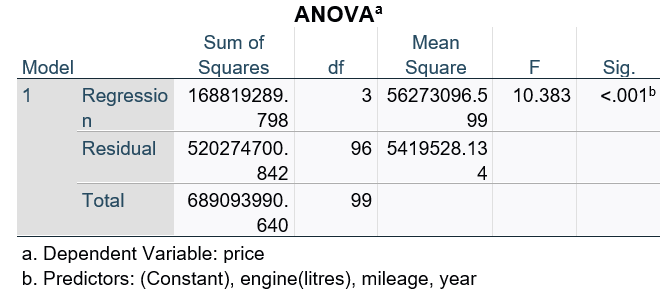


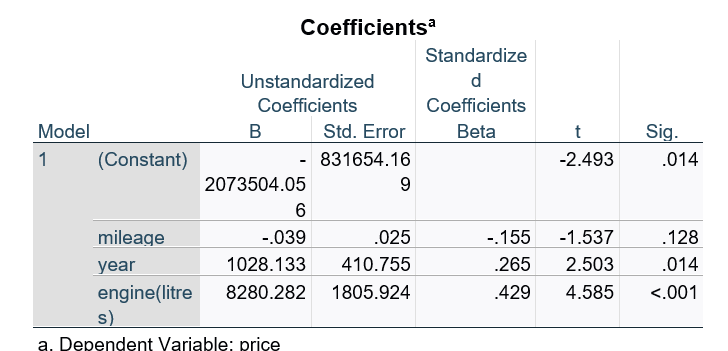
Some of the most interesting links affecting Ford Fiestas' pricing are briefly shown by the correlation study. The most noteworthy positive association with price is that for the engine size (0.345), meaning that more expensive cars are those with larger engines. The miles travelled (-0.265) negatively affect price, meaning that greater distance driven autos are less expensive. The car's age indicates that newer models are more expensive since it is favorably correlated with the price (0.211). These findings show that the market pricing of Ford Fiesta vehicles are much influenced by the engine size, kilometers driven by the car, and vehicle age.

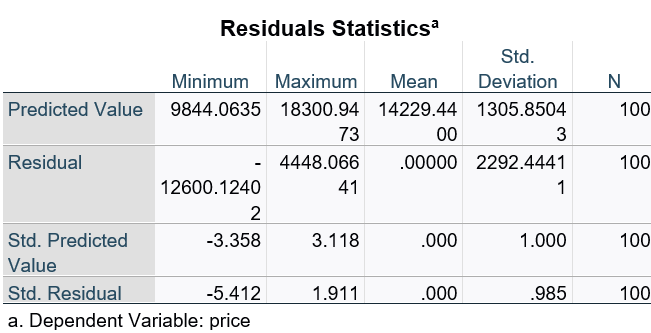
**REGRESSION ANALYSIS****:**











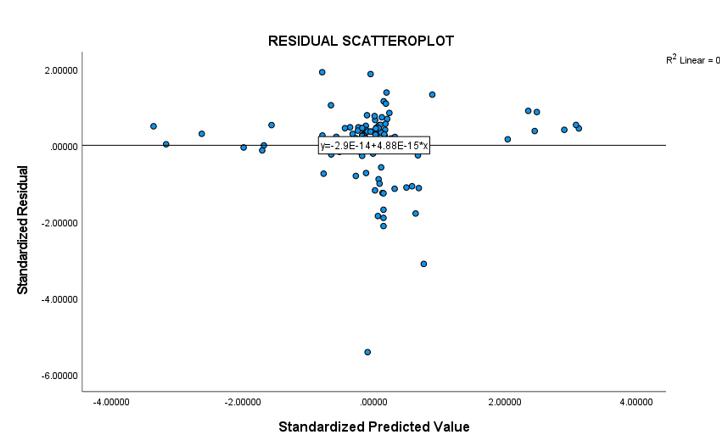
The analysis of regression examines the effect of mileage, engine size, and year on Ford Fiesta prices. Collectively, these factors explain around 24.5% of the variation of prices, indicating that they are significant but slight influencers.

The greatest contribution comes from the size of the engine (p < 0.001). Age of the vehicle (year of registration) is also significant, with newer vehicles being most expensive (p = 0.014). Neither of these was significant individually (p = 0.128), which indicates that, for this group, car age and size of the engine are less determinant factors of cost than their combinations.

**RESIDUAL ANALYSIS**

* **Residual scatter plot:**

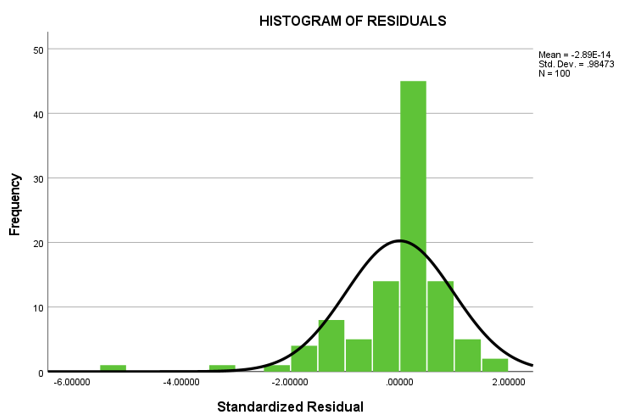
.

****

The residual scatterplot checks if the model fits. If the assumptions of the model, especially the homoscedasticity assumption, hold, the residuals will be randomly spread around zero with no patterning.

But we notice some of the data points (outliers) have bigger negative residuals, indicating the model is occasionally off for some of the cars. These may be instances we would need closer examination on to determine why their prices are this far off from predictions.

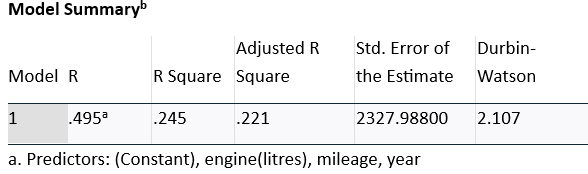
* **Histogram of residuals:**



The histogram of the residuals is one graphical check of whether the model for the prices of Ford Fiesta cars will make reasonable predictions. If the actual car prices, less the predicted prices, would cluster around zero, then this histogram is essentially showing that the most of the residuals will be near zero, and thus the predictions will be extremely near the actual prices of the cars from the market.

Nonetheless, the distribution is not even remotely normally distributed; we see a spike for zero, and some of the residues creep into negative extremes. The negative outliers indicate that for some of the cars, the model is under-pricing below their present market worth. Even though the model periodically makes such mistakes, the shape of the residues still remains fairly normally.

* **Durbin Watson Statistics:**

****

The Durbin-Watson statistic for the model is 2.107, which is roughly the ideal of 2. That means the residuals of the model—i.e., the differences of predicted prices from actual prices—don't show large autocorrelation. That is, no observation is independent of the preceding one. That kind of independence is important because it means the model is realistic, and it satisfies one of the primary assumptions required for accurate and legitimate results.

**CONCLUSION:**

Consequently, the research was undertaken for the determination of prices of used Ford Fiestas on the British market using a dataset of approximately 100 vehicles. The principal conclusions of the research are that size of the engine, registration age, and level of mileage are the determinants of the price with bigger engines, newer vehicles, and higher levels of the distance traveled decreasing car prices, respectively. The research deduced that prices of cars from the local dealerships closely conform with the common national market prices, showing competitive local pricing strategies.

The model is correctly predicting the car prices from the most salient car features, with some of the predictions at the lower end of actual car prices at some point. The robustness of the model is also supported with the help of residual analysis, with the homoscedasticity and normality assumptions being met mainly.

The derived statistical regression model for predicting the market price of a Ford Fiesta is:

Price=−20735.04+(8280.28×Engine size in litres)+(1028.13×Registration year)−(0.039×Mileage)

Overall, dealers can confidently use this model to strategically price Ford Fiestas with close proximity to the expectations of the market and enhance their competitiveness. Nevertheless, further research using larger data sets can be employed to enhance accuracy and eliminate sporadic prediction deviations.

**REFERRENCE:**

1. <https://www.autotrader.co.uk>