

26.4

By Adam Chu



# Introduction

Jurassic Software Solutions (JSS) is a small company specialising in bespoke software for local businesses. You have just completed your probationary year with them during which you shadowed your Team Leader in the project system testing phase held on-site at clients’ premises. The next stage in your development is to perform the same role without your Team Leader. For this you will need your own car to visit clients as JSS does not provide company cars.

You have a budget of £10000 and have decided that the car will be up to 10 years old. You have an idea of engine size and have narrowed your choice down to two makes (e.g. Ford). You would like to know which make represents better value for money, with regards to age and mileage for that engine size.

# Initial Research

For me I would like to choose an electric car seeing as the places I would have to drive through might have congestion charges, this would affect me even further seeing as the company will not give me a company car. This combined with the fact that I would like to get a car that is fairly recent (not prior in manufacturing that 2009) I have decided to get an electric car, this would protect me against any new laws or restrictions in the consumption of Petrol and Diesel in the use of motor vehicles.

I have set a minimum in battery size to 20Kwh as unlike with a petrol or diesel car the engine isn’t the main contributor to the consumption of fuel. In electric cars it’s the Battery Solution as the size of the battery is directly proportional to the Range of the car. Anything less than a 20Kwh battery would have less than 60 Miles of range whereas the Zoe and the Leaf have >90 Miles of range.

I have kept the same:

* Door Count of 5
* Fuel Type >> Electric
* A minimum of the battery size to keep range usable

I will gather data for 60 cars half of it will be 1 make of electric car that fits the specifications I have mentioned above with the other being another relevant model of electric car.

For the website I have chosen AutoTrader as it’s a website that is consistently as safe place to buy and sell used cars, its buyer protection is quite renowned and its known for its good and relevant information. I chose 30 cars as it allows me to get the standard deviation and variance without being too small or large of a data set for me to comfortably sort through. This allowed me to check for fully electric cars in the <£10,000 range which left me with 3 cars to choose from the:

1. Renault Twizy
2. Renault ZOE
3. Nissan Leaf

Using the requirements set up from prior to my initial search, that meant that the Twizy is disqualified seeing as it has 2 doors. Also it only has two seats which means that it would not work as a family car, as well as having the problem of a maximum speed of 50MPH. I had decided that I wanted the car to be functional as well as electric so I chose the other two cars.

I am using Excel as it would allow me to calculate and store standard deviation without using lots of formulae, also it would allow me to plot Histograms and Scatter Charts with the Trend lines without approximation.

# Averages

* A Mode of £7995
* A Median of £8329
* A Mean of £8329

The most accurate would be the mean as it is a measure of all of the values meaning it generates an average cost based on all of the information. Unlike Median or Mode which are directly related to the number of numbers inside the data set, instead of what data inside the set this leads to misinformation to be spread for example the mode for the Nissan Leaf is £7995 which is lower than the Median and the Mean. This would make people assume that a good price for this car is lower than £7995 which is not true.

# Standard Deviation

Standard Deviation and the Variance tell you the ‘spread’ of the data, what this means is how closely packed the data is. A wider spread of data tells you that you will have more extreme values (values at 0-5% & 95% - 100%) of your data set, also having a larger spread makes the Mean that I calculated above less relevant as its being affected by anomalous results. On the contrary to this if you have a smaller standard deviation then the Mean is more effective as it not being affected by Anomalous results.

* The Standard Deviation for The Nissan Leaf is 30.189
* The Standard Deviation for The Renault ZOE is 27.635

What this tells you is that the Renault ZOE has a much smaller range in its data set, this means that when looking at the Mileage Vs Cost Scatter graphs you can see that the data is closer packed and that the trend line is more accurate because of it. Unlike the Nissan Leaf which has much a larger range and has a less accurate trend line.

# Scatter Plots

The Scatter Plots that I have generated show that buying a Nissan Leaf you will have cars that are older mixed into the pile of newer cars probably due to the Nissan Leaf being older than the ZOE. This creates a scatter plot which is more spread out. To prove this I have plotted the variance and mean as x = equations into the graph. According to an [Online Textbook](http://my.ilstu.edu/~gjin/hsc204-hed/Module-5-Summary-Measure-2/Module-5-Summary-Measure-28.html), when plotted onto a graph with the points listed onto it 68.3% of the data is between the Mean and +1 and -1 \* σ. With the Standard deviation smaller then there will be fewer extreme values (cars that should be considered not viable).

The reason why I have used the standard deviation instead of the variance even-though the standard deviation could be negative (Variance is σ2) (this does not affect creating upper and lower bounds with it as x -- σ == x + σ & x +- σ == x - σ). Also the use of the variance removes the centralisation to the mean. With Variance it is more useful when you are analysing much larger data sets (100-200) & if the data has bias to it where data is not evenly distributed each side of the mean having a proper non-centralised measure of spread is helpful.

Both graphs have a slight negative correlation where the mileage of a car decreases as the cost increases, this is because it adds value to a car and the same could be said when comparing mileage to year of manufacturing for the car.

# Histograms

The histograms that I have created are showing the price groupings using classes. The Renault ZOE histogram shows that the largest selection of cars is in the group £7170 - £8040 and that this number is 15. Whereas in the Nissan Leaf Histogram shows that the largest selection of the cars is in the group £7995-£8995 at a group number of 11 but the lower down class of £6995- £7995 has 10 cars in this band as well.

This tells us that the data is more spread out between the classes inside of the Nissan Leafs’ Histogram, also we can infer that the Histograms are not as helpful as they could have been seeing as the class widths are quite vague if there were more classes then we might have seen a better trend. The histogram for the ZOE has followed the normal trend of a bell curve around the mean which shows a much closer knit spread which seems to evenly split over the mean.

# Conclusion:

In conclusion I would choose the Renault ZOE over the Nissan Leaf, the data that I have collected has shown that the Nissan Leaf is worse than the ZOE in the Key Features that I wanted in my car: I wanted my car to be less than 10 years old (The newer the better), I had £10,000 to spend on a car (the less I spend is better as I could use it to install an electric car outlet at home) & I had to worry about the mileage as the larger the Mileage the older the car and it might need servicing sooner as its older components.

In the data collected for the ZOE it shows that the average cost is £7830 which is significantly lower than the Nissan Leaf which has an average £8239. What this tells us is that there more of the data the Nissan Leaf’s cost is higher than that of the ZOE. Combined with the use of the Standard Deviation we know that the Spread is larger meaning that there will be much cheaper cars in the same data set as much more expensive ones.

When looking at my data on the Age values it shows that the Nissan Leaf has a larger degree of spread for this is that its first iteration (that I could find on AutoTrader) was sold with a ‘Age’ of 2011 this means that the majority of cheaper Nissan Leafs’ (Ones in my price range) are going to have larger mileages and the age of the cars in this set will be lower.

When comparing the average of the age of the models it shows that the Nissan Leaf’s ‘age’ is approximately 2012 whereas the ZOE is 2015. This tells you that whilst there are some cars in the Nissans Data set that are >= 2014 they are smaller percentile than the <2013 group, this includes the Mode of 2013. (With the ZOE it has a mode of 2015)

I calculated the Standard deviation of the ‘Age’ because this would tell you where 68.3% of the cars would reside and on the ZOE having a σ of approximately 0.9 meant most of the cars were in the 2014 - 2015 range, compared to the σ of approximately -1.01 meant that the 68.3% of the cars were from 2011 to 2013. Also the older Nissan Leaf costs more on average meaning you are spending more money on older and more used cars.

To decide which Renault ZOE that I would purchase the best graph would be a Cost V Age graph as the mileage is mainly dependent on the other two anyway and is the least important specification whereas the Age (what features it has, what speed, what battery size) as these factors will change based on the Age. And the cost which is arguably the most important part that I am checking as a car could have low mileage and be brand new but charge me all of the £10,000.   
  
What I created was a graph using inequalities so it highlights the areas inside of it so that I will choose cars that are below the upper bounds from the cost’s AND the lower bound from the Age column. This would restrict the cars that I would buy to being newer than 2014 and less than £8539 to narrow my field down. Looking at this data the car I would select being a 2015 model that has done 8200 miles, even though it is not the cheapest one in that set it has much lower mileage on the car itself.

# What Influences Information Validity?

## Validity Vs Reliability

Validity is a characteristic of a test it measures what it was intended to in its creation for example. If you do a test which says you weigh 70 Kg and you actually weigh 75Kg then the test itself is invalid as the information produced is not correct.

Reliability is a synonym for ‘Consistency’ this means that if a test is consistently wrong then it can be classed as reliable. This is because the test is wrong ‘but it has to be in the same way and magnitude wrong’, if a test is unreliable it means that it is incorrect and changes it is not reliable.

Reliability and Validity are independent from each other meaning that a test can be Valid but not Reliable, this means that the data generated is accurate but not consistent. On the other hand the use of Reliable but not Valid data means that the user is collecting incorrect data but the test results are closely knit.

This was true with both cars with the reliability of the data being good seeing as whilst there were some anomalous results. The Nissan Leaf with a cost of £9995 and multiple ZOEs which was £9489 which are treated as such to not affect the ‘cost-effective’ purpose of my research, and would affect the cars that I would choose to heavily.

The age values as well were reliable and valid because the majority of the values are reliable they are in the range of 2015 - 2017, and valid seeing as they are in the range of 2010 - 2019. This means that the ages of the car are fit for purpose and relevant to the study I am doing.

The mileages for the Nissan Leaf were more spread due to the variance in age between each model being larger, but they still are valid seeing as I did not put restrictions on the actual size of the mileage itself. This is understandable as I am more interested in the cost of the car and the generation of it due to it affecting features and functions for the car, compared to how long its been on the road. Because these issues making the car undriveable should be fixed prior to selling the car to us the users.

# Reflect on the results of the study for data collection

## AutoTrader

AutoTrader is a helpful tool and is a good idea as the user has the options: to limit the total cost and make of the cars using AutoTrader’s Filters also the data was quite reliable, as when I generated the averages and standard deviation showed that the spread of ‘Costs’ were closely knit. This means that the sellers are all giving good prices and sticking to the market value of the car (with a similar mark-up). This made the information collected Accessible and I could maintain a high level of detail compared to other sellers for example: Mileage, Year, Door Count, Fuel Type, Gearbox Type, 0-60MPH times, Expected Range, Battery Capacity, Maximum Speed. On AutoTrader these all have to be shown on all car listings. This large amount of data helped me when creating my graphs and allowed me to compare mileage to cost & cost to age.

## What would I have changed?

Whilst AutoTrader is a good source of information I should have split it to get a secondary source like using the Facebook Marketplace or using AA’s Used car selling utility. This would allow me to make sure the data is reliable and cost effective, if people are selling ZOE’s at a similar price and age then the study is Valid and reliable. Also this would allow for checking if the data collected on AutoTrader is more cost effective than the ones sold on the Facebook Marketplace or Gumtree is a bonus to us a potential buyer.

Having multiple ‘middle men’ type providers could try to sell me the same car so what I would do is record the Licence REG so that I could not find multiple advertisement for the same car. This would have been helpful on this run for AutoTrader as well.

## Other Methods

I could have called up potential sellers and asked for there current pricing on an electric car that is used and under £10,000. The problem of manually calling up different sellers 60-times, is that this is time consuming: due to the fact that I would have to find a seller wait for them to answer or ring me back. This also could increase the chance for data redundancy issues as I am manually phoning people and copying the answers onto the excel spreadsheet, unlike using a static webpage.

Also using the system of ringing a dealership up and waiting means that you will end up calling more than the required dealerships as they might have the car you want but at too high of a price, you lose out on the filter and sort systems of web-based car dealers.