GCSE Mathematics

Statistics Coursework

SUMMER 2002

SECOND HAND CAR PRICES

Hypothesis

THE VALUE OF A USED CAR DECREASES EXPONENTIALLY WITH ITS AGE.

By *exponentially* it means that a graph of a certain make of car, its price on one axis and its age on the other, would show a curved decrease in price over a period of time as shown in the example on the following page.

The following experiment is being conducted to help and assist people who are buying and selling cars in magazines such as "Auto Trader".

Approach

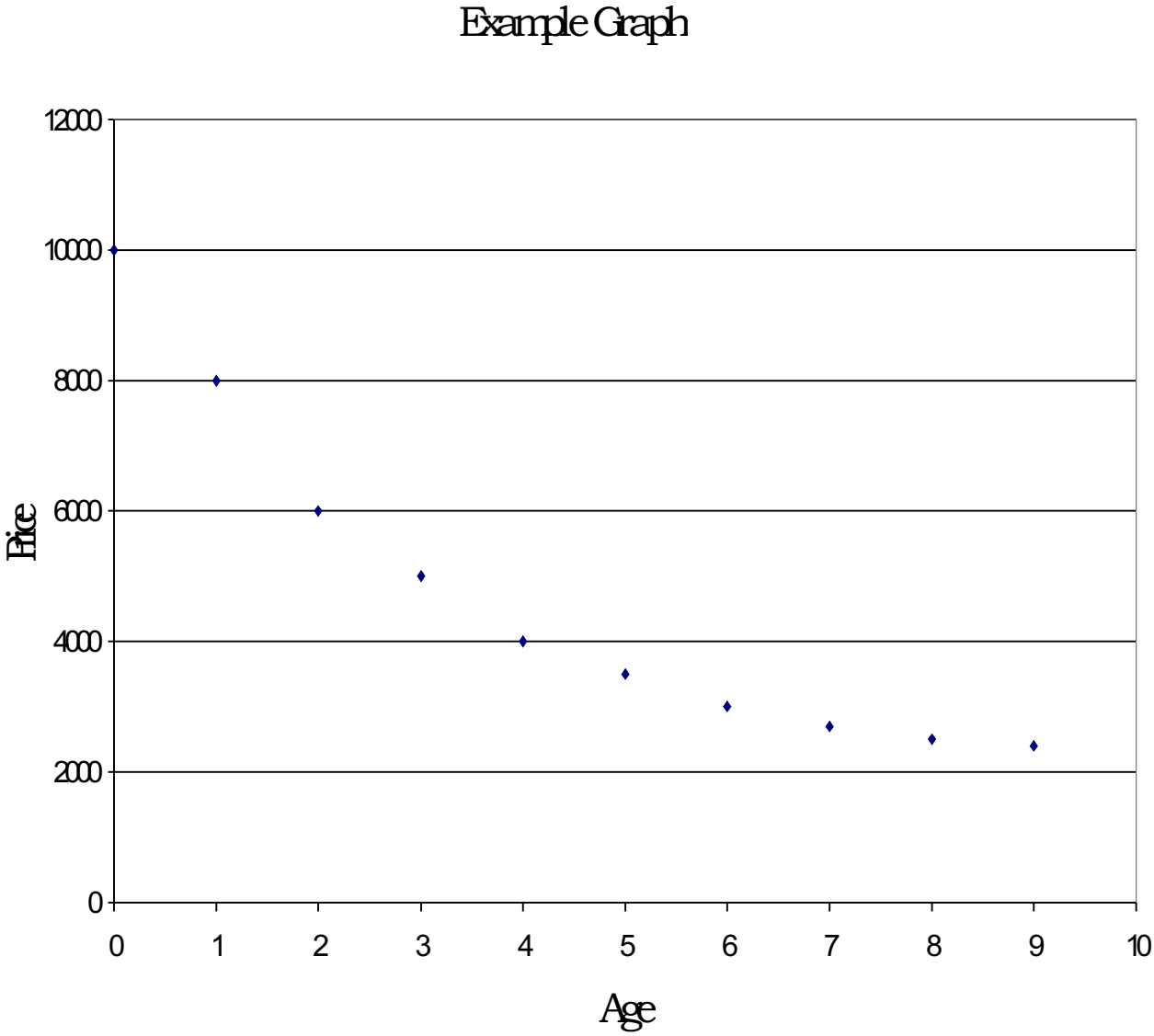
Throughout the project I have referred to the second hand car sales pages in the magazines "Auto Trader" and "Exchange & Mart" (in the local paper) for data. In these, I found a wide range of data over three weeks from which to select the type of data required. I selected relevant pieces of data in each age range for the Mercedes C-Class. If there was not enough information in the source I would have referred to another car magazine. Due to the fact that sampling concentrates on gaining information about selected sampling units, the quality of information gained is better than if the information was of units that had not been specified. Ideally, it would have been best to use random sampling but as the data available was already limited, it was decided to take all the data available over a three-week period.

When choosing data, it is very important to select data to be included in testing on an unbiased selection process. This ensures that the results are reliable. If the data was biased in anyway this would mean that, although the hypothesis may seem to be true, it cannot be applied to other situations and may be of limited value.

In this statistical work I only used one type/make of car, Mercedes, and within this I used only the "C-class" range and within that, the saloon style. By doing this it restricted the type of sample chosen and I also limited the price variation to a very small number of factors – i.e. number

of doors, age. Other factors such as condition, mileage, modifications, etc. are still relevant but finding data and producing appropriate graphs would be too difficult. To validate the results further I conducted the same experiment with the Volkswagen Golf GTI Hatchback, limiting the range to five-door cars with a diesel engine sized between 1.8l and 2.2l, as well as the other factors mentioned before.

Example of an Exponential Graph Underneath

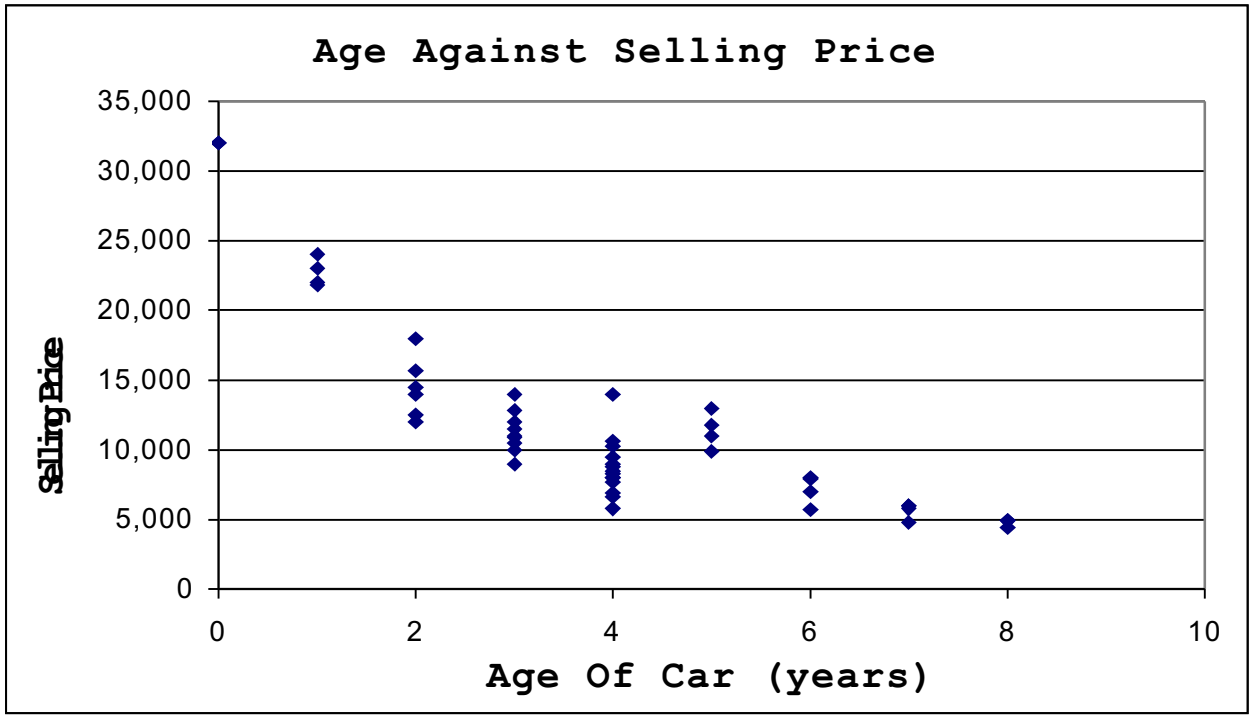


The graph above shows how the exponential line of best fit should look in the following graphs.

DATA OBTAINED

AGE	SELLING PRICE
0	31,995
0	31,995
1	22,995
1	23,995
1	21,795
1	21,999
2	15,699
2	17,999
2	13,999
2	12,499
2	11,999
2	14,499
3	13,999
3	12,799
3	11,999
3	11,499
3	10,999
3	10,499
3	9,999
3	10,899
3	8,999
4	13,995
4	10,599
4	10,299
4	9,499

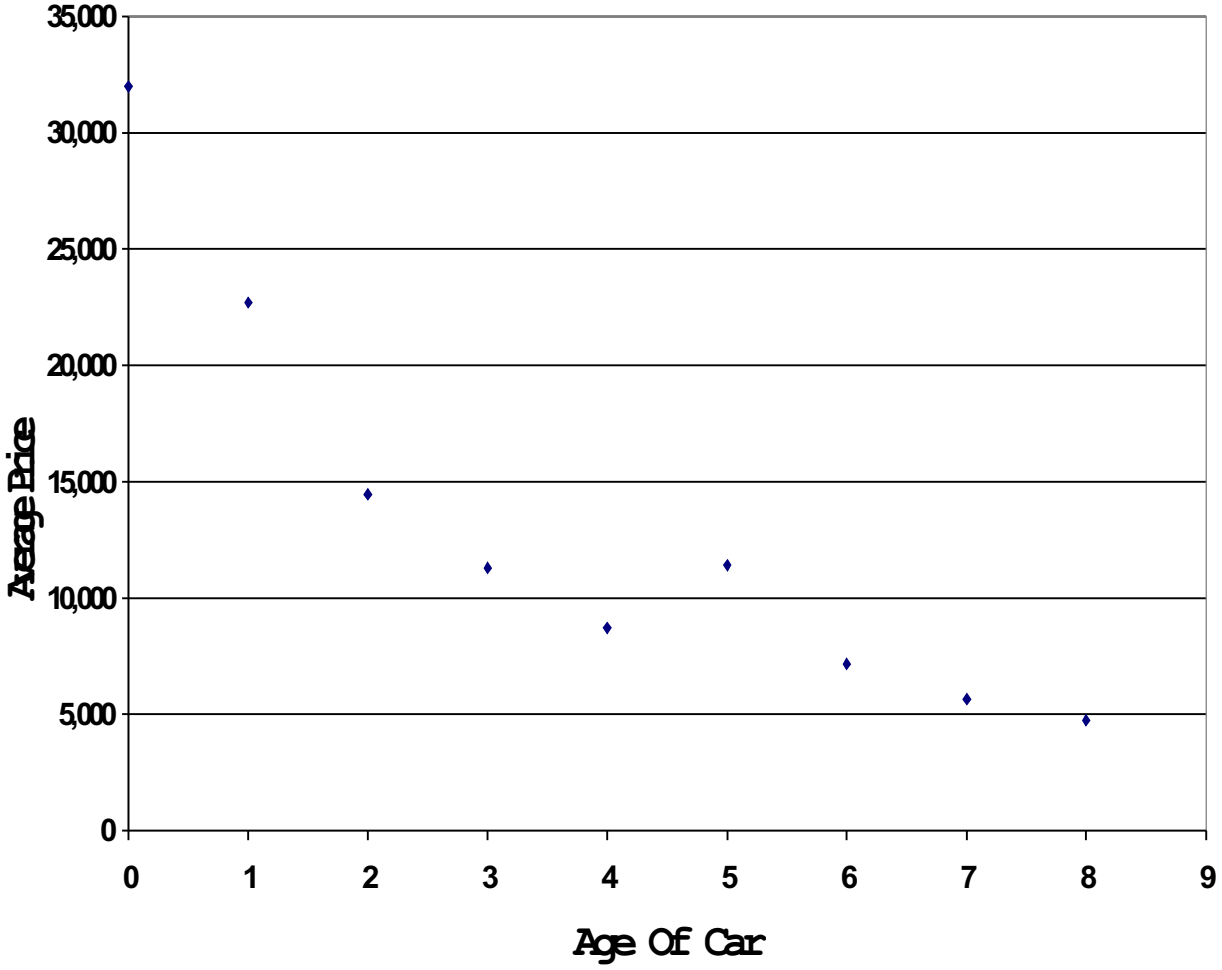
AGE	SELLING PRICE
4	8,999
4	8,799
4	8,499
4	5,799
4	8,299
4	7,999
4	6,599
4	6,899
4	7,999
4	7,699
5	12,995
5	11,795
5	10,999
5	9,895
6	7,899
6	7,995
6	6,999
6	5,695
7	5,999
7	5,995
7	4,795
7	5,795
8	4,395
8	4,899
8	4,895



Average Car Price for each year

Age	Average Car Price
0	31,995
1	22,696
2	14,449
3	11,299
4	8,713
5	11,421
6	7,147
7	5,646
8	4,730

Average Price of a Mercedes C-Class Against Its Age



I used standard deviation to explain the average spread of the values from the mean for each year.

Example of SD

	KEY
∑x	SUM OF ALL VALUES
n	NUMBER OF VALUES
∑x2	SUM OF ALL SQUARED VALUES

x	$(x-x)$	$(x-x)^2$
5	-0.2	0.04
7	1.8	3.24
3	-2.2	4.84
10	4.8	23.04
1	-4.2	17.64
$\Sigma x=26$	$\Sigma(x-x)=0$	$\Sigma(x-x)2=48.8$
$n=5$		
$x=\Sigma x/n$		
$x=5.2$		

$$\frac{\Sigma(x-x)^2}{n} = \frac{48.8}{5} = 9.76$$

$$\sqrt{\frac{\Sigma(x-x)^2}{n}} = 3.124$$

ACTUAL MERCEDES DATA ANALYSED IN THE FORM OF STANDARD DEVIATION

Year	N	∑x	x	$\sum x^2$	SD
0	2	63990	31995	2047360050	0
1	4	90784	22696	2063508076	876.7
2	6	86694	14449	1276816606	2007.3
3	9	101691	11299	1166946609	1411.9
4	14	121982	8713	1116024038	1949.3
5	4	45684	11421	526881076	1131.8
6	4	28588	7147	207733252	924 .0
7	4	22584	5646	128502076	498.2
8	3	14190	4730	67277251	236.7

The results gathered from the standard deviation information show that the spread of all the plotted figures is fairly small. This is also the case for the fifth year in the Mercedes car data, which means that even though the results are out of line, the plotted data is close together within a small price range.

Analysis

As seen in the data table above, all the Mercedes C-Class cars that are five years old seem to be out of line compared to the expected pattern (they are highlighted in the table). The graph above shows these anomalous results. They could be due to one of the following:

1. Features which have been added to the car by previous owners.
2. There could have been a limited edition/selected batch of this type of car made that year.
3. Due to the rate of inflation, as all the cars that are five years old are priced higher than would be expected according to the trend. *This is by far the most likely reason.*

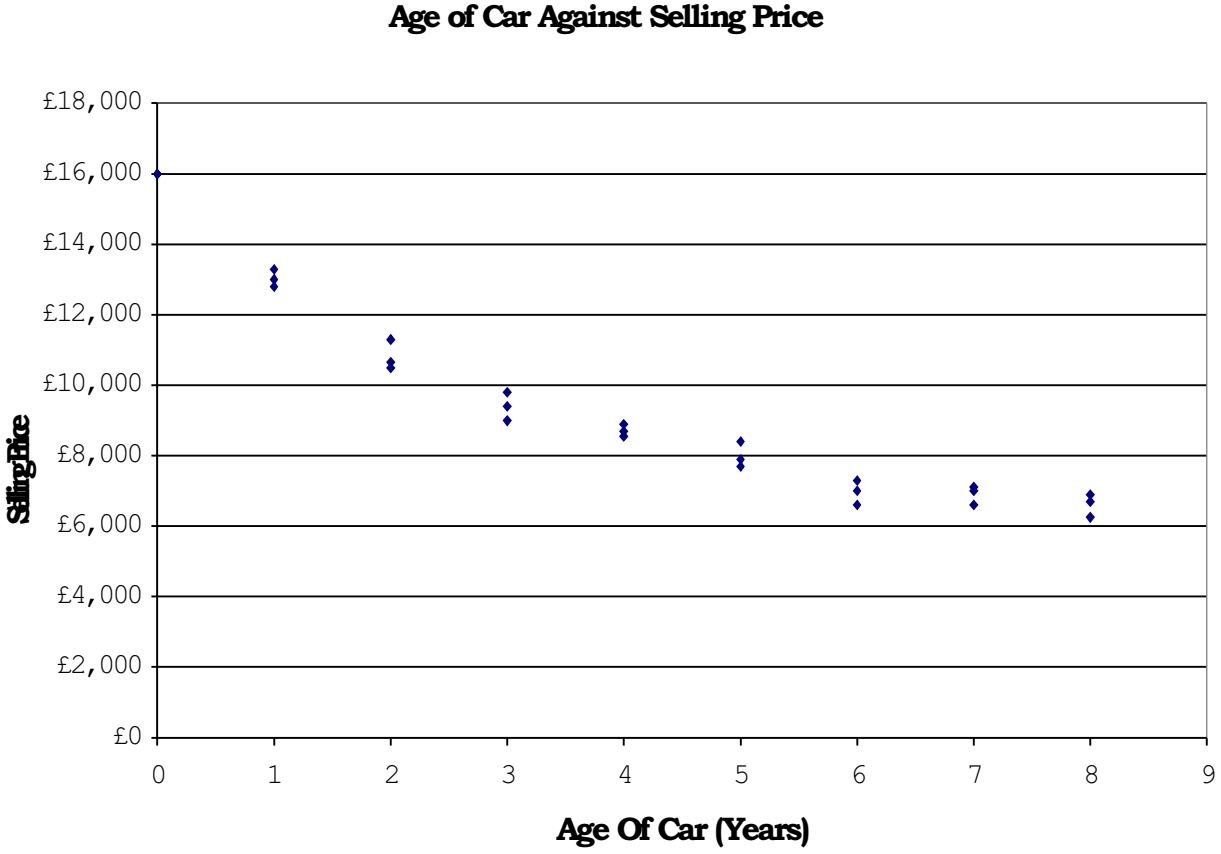
The exponential decrease of the Mercedes C-Class was discovered by examining this data. When analysing the data obtained, the average price drop after the initial year of sales was found to be from £32k to £22.5k, a drop of £9,500 (30%) compared to the original price. In the second year, the car price dropped by another £8,000 (a 35% decrease in value). Then, the rate of decrease, in terms of value, slows down, levelling out. This fits the theory of exponential devaluation indicated by the "line of best fit".

One of the drawbacks in analysing the data was that the averages were taken for each year. This type of average, i.e. the mean, suffers from distortion based on single anomalous or outlying values. Under the circumstances a different type of average e.g. the median may have been more appropriate. One point to make here is that the mean price for the 5-year-old vehicles was obtained from four sample values. If this year's average had been discounted from the plotted data, a much truer analysis would have been possible. Therefore, when I marked in the exponential curve "line of best fit", I discarded the anomalous result and drew the expected trend of devaluation for the car.

Also, the average for the 4th year's data is likely to be more accurate than the 8^{th} year's, because it is an average of 14 pieces of information whereas, for the 8^{th} year, it is an average of only three pieces of data. In general, the greater the range of data, the more accurate the constructed picture can be.

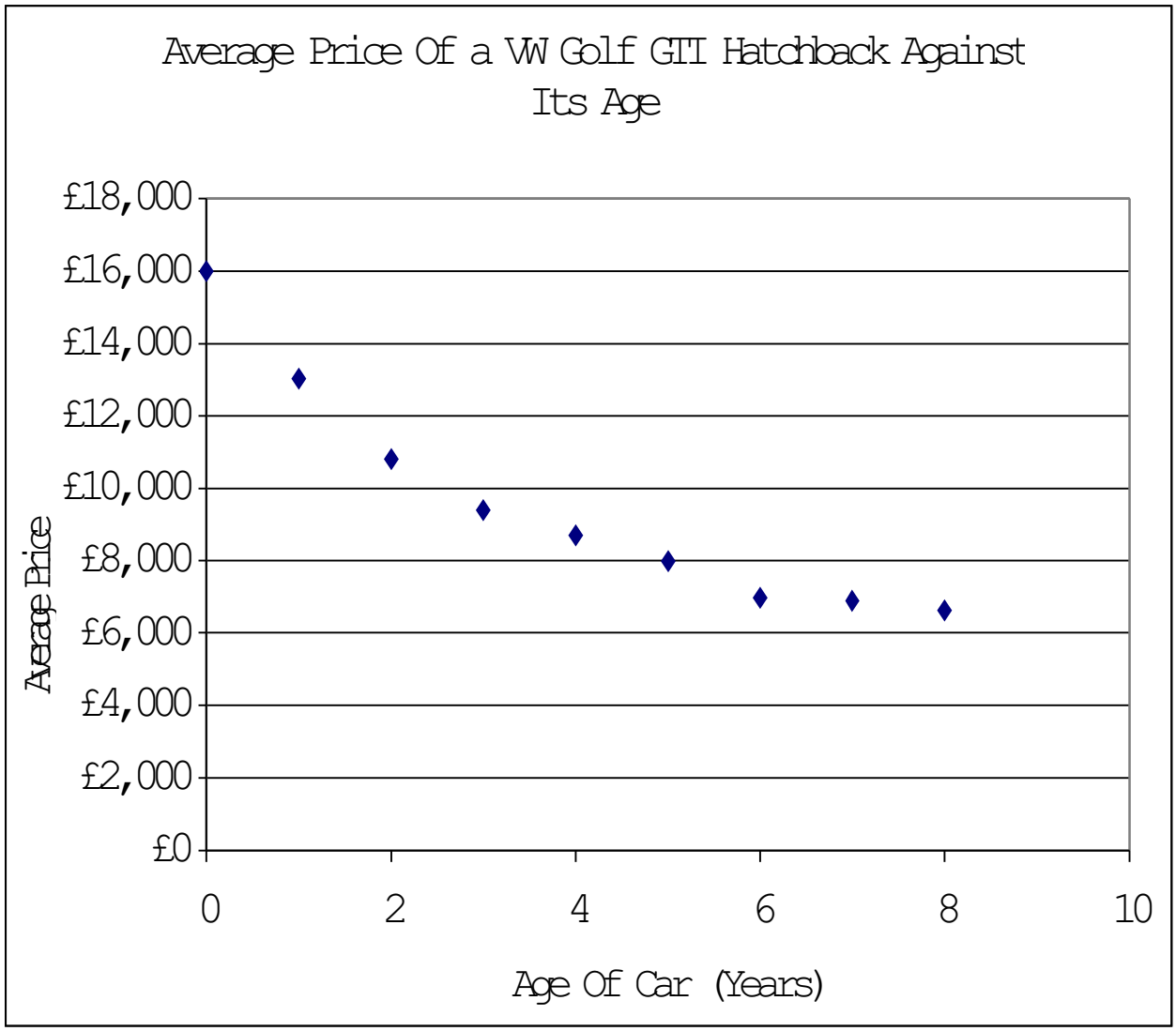
The same experiments were carried out on the Volkswagen Golf GTI Hatchback to confirm the results of the hypothesis further:

Data Collected for the VW Golf Hatchback GTI			
Age	**Price**	4	£8,695
0	£15,996	4	£8,550
0	£15,996	5	£8,400
0	£15,996	5	£7,895
1	£12,999	5	£7,689
1	£13,295	6	£7,000
1	£12,795	6	£7,295
2	£11,295	6	£6,595
2	£10,655	7	£7,099
2	£10,499	7	£6,999
3	£8,999	7	£6,599
3	£9,795	8	£6,899
3	£9,399	8	£6,695
4	£8,895	8	£6,250



Average car Prices for each year and Standard Deviation

Age	Average Car Price	SD
0	£15,996	0
1	£13,030	205.272719
2	£10,816	344.408027
3	£9,398	324.967007
4	£8,713	141.440997
5	£7,995	298.697543
6	£6,963	286.947537
7	£6,899	216.02469
8	£6,615	270.973964



Analysis

The use of standard deviation for the results from the VW Golf GTI was to show the average spread of the values in each year range. The smaller the value, the closer the results and, as seen in both the second table and graph, the annual values are quite close to their respective means. All results for the standard deviation are below £345.

The points on both the graphs indicate that an exponential relationship exists between these two factors because they would not correlate well enough to fit a straight "line of best fit". The VW Golf GTI holds its price well compared to the Mercedes as, in the first year its price dropped from £16,000 to £13,000 (a 19% decrease). In the second year, the car price dropped by a further £2,200 (a 17% decrease in value). In the third year the Golf GTI loses £1420, a vast difference compared to the first two annual decreases. This then seems to level out and the decrease between the 7^{th} and 8^{th} year falls to £284. As you can see in the table of averages, the Golf GTI holds its price better in percentage terms than the Mercedes C-Class. This may be due to the size of the car, the target group of the car, speed of the car etc.

Although both the Mercedes C-Class and the VW Golf's values decrease exponentially, there is a large difference in the uses of these cars. The price of the Volkswagen Golf GTI starts at around £16,000 because this would be suited to the needs of a small family, young singles, etc. and would be very useful in and around city centres. On the other hand, the Mercedes C-Class Saloon range would be aimed at high-powered business executives, as they need to give a good impression when meeting with clients and its price would be within their substantial salaries.

Conclusion

The first graph shows us that there is a definite correlation between the age of a Mercedes C-Class and the price of the car. The second graph, which shows the average price of a Mercedes against its age, comfirms that there is a strong exponential correlation. Therefore, this proves that the hypothesis made initially is correct.

The results shown above would have been more accurate if the same number of data points were available in each 'year group', and this would give a more accurate mean. This was noticed and the data for the VW Golf GTI was adjusted accordingly to see if there was any significant data. There may well have been, but more data points would have to be available for each year range.

The results for the VW Golf GTI show that, with fewer data points for each range, the standard deviation tends to be lower. This implies greater accuracy, but a greater range of points may show that this is not the case. On the other hand, the limitations of the data would improve the level of accuracy.

Also, all the data used is from two particular types of newspaper/magazine. Not everyone who has a Mercedes C-Class to sell decides to advertise in these particular publications. Therefore, if I were to repeat the experiment or develop it further, I would have to take into consideration data from other newspapers/magazines.

Using spearman's rank may produce a more accurate line of best fit. If there was more time to conduct this experiment, I would have investigated this area.