International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1407-1412 www.ijhssm.org

Analysis of Car Sales Data

AADITYA SHAH – I01 AASHNA BHATIA – I02 ABDUL KABEER – I03 ADHIRAJ JAIN – I04 AKSHAT TOURANI – I05

Anil Surendra Modi School of Commerce Nmims Deemed To Be University
Mumbai Campus

Date of Submission: 04-04-2024 Date of Acceptance: 18-04-2024

I. INTRODUCTION

In the world of quantitative techniques, the automotive industry stands out as a dynamic arena shaped by economic conditions, consumer preferences, technological advancements, and market competition. For stakeholders such as manufacturers, dealers, investors, and policymakers, understanding the trends and patterns in car sales is paramount for informed decision-making and strategy formulation. This project delves into the application of quantitative techniques through analyses of historical data using tools like moving average analysis, coefficient of variation, index numbers, and regression analysis, we aim to uncover the underlying trends and drivers of car sales dynamics.

II. OBJECTIVES

The objective of this project is to analyse and interpret the trends in car sales over a specific period, using various statistical and analytical tools. By examining historical data on selling prices and employing techniques such as moving average analysis, coefficient of variation, index numbers, and regression analysis.

1) Conceptual Clarity of Statistics: By comprehensively applying statistical techniques to interpret trends and patterns in car sales, students gain conceptual clarity of statistics and its relevance in the global business environment.
2) Analysis of Decision-Making Situations: Working with statistical tools equipped us with the ability to evaluate data, identify trends, and draw conclusions, enhancing their analytical skills in addressing complex business scenarios.

MOVING AVERAGE ANALYSIS

A moving average is a technical indicator that market analysts and investors may use to determine the direction of a trend. It sums up the data points of a financial security over a specific time period and divides the total by the number of data points to arrive at an average. It is called a "moving" average because it is continually recalculated based on the latest price data. A moving average is a series of averages, calculated from historic data. Moving averages can be calculated for any number of time periods, for example a three-month moving average, a seven-day moving average, or a four-quarter moving average. The basic calculations are the same.

We have used 3-yearly, 4-yearly and 5 yearly averages to analyse the trend that it depicts.

3-yearly average

Years 1975-1984 are the initial years and have no prior history to compare with, hence there is no moving average calculated (N/A)

The first noticeable increase in trend of selling prices comes in the year 2000 where the moving average is 155616.667. this could be because of economic conditions- The United States, along with many other countries, had economic prosperity throughout the year 2000. Consumers may have greater disposable income during economic booms, which encourages them to spend more on expensive goods like vehicles. Prolonged economic expansion may also boost consumer confidence, which in turn may motivate consumers to make major purchases.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1407-1412 www.ijhssm.org

3 Year Moving Average		
Year	Selling Price	Moving Average
1975	7600	#N/A
1984	53500	#N/A
1987	17300	26133.33333
1989	20000	30266.66667
1990	31250	22850
1991	20000	23750
1993	11500	20916.66667
1994	27000	19500
1996	50300	29600
1997	20250	32516.66667
1998	43550	38033.33333
1999	144650	69483.33333
2000	278650	155616.6667
2001	189550	204283.3333
2002	153350	207183.3333
2003	73300	138733.3333
2004	141050	122566.6667
2005	635100	283150
2006	501200	425783.3333
2007	1569650	901983.3333
2008	1808400	1293083.333
2009	2208550	1862200
2010	3525000	2513983.333
2011	3393050	3042200
2012	4350100	3756050
2013	1710000	3151050

The next major trend increase in prices comes in the year 2005 where the moving average becomes 283150. From the year 2005 till 2012 this is a significant increase yearly. The moving average reaches 3756050 at its peak. One of the major reasons causing this, is Inflation. As the cost of goods and services increases, car manufacturers and dealers may adjust their prices to maintain profit margins. This in turn increases the selling prices and cars become costlier.

There is a significant decrease in the year 2013 where the moving average reduces to 3151050. This could be due to Market Correction. In 2013, the market might have undergone a correction following years of steady price growth. Following a period of expansion, market corrections happen when prices decline. These adjustments are frequently brought about by causes like excess supply, decreased demand, or shifts in customer preferences. Hence we see a decrease in price.

5 yearly average

The 5 yearly moving average data, years 1975-1984 are not calculated because these are the initial years which don't have prior history.

Just like in the 3 yearly moving average, the first noticeable increase in the selling price trend comes in 2005 due to factors like technological advancements, economic conditions and inflation.

In the 4 yearly average, the prices become somewhat constant again in the year 2013 thus marking an to the trend of year on year increase. This could be due to market correction.

Limitations

Sensitivity to Window Size: The selection of the window size (in this example, 3,4,5 years) can have a big impact on the outcomes. While bigger window sizes may smooth out trends but lag behind real changes, smaller window sizes may result in more responsive but noisy estimates. Absence of Predictive Power: The main uses of moving averages are historical and descriptive. They don't foresee abrupt changes in the data or have the ability to predict future patterns.

Susceptibility to Outliers: If the window size is small, moving averages may be impacted by extreme values or outliers in the data. The moving average may be distorted by outliers, which could result in incorrect readings of the underlying trend.

DESCRIPTIVE STATISTICS AND COEFFICIENT OF VARIATION

Descriptive Statistics summarizes the statistical data using the available information in the data set. The final summary shows various parameters such as mean, median, mode, skewness, etc.

In this analysis, using Descriptive Statistics played an important role as Standard Deviation and Mean were used from the table in finding the coefficient of variation. The coefficient of variation (CV) is a statistical measure used to assess the relative variability of a dataset. It is calculated by dividing the standard deviation of the dataset by the mean and expressing the result as a percentage.

sellingprice	
Mean	21571.61
Standard Error	486.8657
Median	18450
Mode	27500
Standard Deviation	12954.65
Sample Variance	1.68E+08
Kurtosis	15.67444
Skewness	2.346517
Range	152700
Minimum	1300
Maximum	154000
Sum	15272700
Count	708

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1407-1412 www.ijhssm.org

The table shows the Descriptive Statistics of the data selected by us. Now, using the values of Standard Deviation and Mean, we found the value of the Coefficient of Variation.
Coefficient of Variation:

$$= \frac{Standard\ Deviation}{Mean} * 100$$

$$= 60.05418$$

In this case, the coefficient of variation for the selling prices is calculated to be approximately 60.05%. This suggests that the selling prices exhibit moderate to high relative variability compared to their mean.

Reasons for the moderate to high variability:
Market Dynamics: The market for cars represented in the dataset is highly volatile and influenced by various factors such as supply and demand, economic conditions, and changing consumer preferences, it has led to greater variability in selling prices and has resulted in a consistent coefficient of variation.
Seller Practices: The sellers represented in the dataset have employed consistent pricing strategies and certain trends or patterns can be identified in their pricing behaviour, this has contributed to the consistency of the coefficient of variation. **Outliers**: The presence of outliers, such as extremely high or low selling prices, has impacted the consistency of the coefficient of variation.

INDEX NUMBERS
Index numbers are statistical measures that represent the relative change in a variable over time or across different categories. They are used to compare changes in the value of a group of goods, services, prices, or other measurable quantities relative to a base period or base value.
Index numbers are particularly useful for:
Tracking Changes: They help in understanding how a particular variable changes over time or across different categories.
Comparison: They facilitate comparison between different groups, periods, or categories.
Analysis: They provide insights into trends and patterns in data.

Laspeyre's Index Number:
Laspeyre's index number is a type of index number commonly used to compare the total value of a group of goods or services in the current period to the total value of the same group of goods or services in a base period, with the base period values serving as weights.

$$Pl = \frac{p1q0}{p0q0} \times 100$$

$$= 183.805671$$

The Laspeyre's index number of 183.805671 suggests that the total selling price has increased by approximately 83.81% from the base period (2008) to the current period (2012), likely indicating economic growth or changes in market conditions favouring higher prices.

Paasche's Index Number:
The Paasche index number is another type of index commonly used to measure changes in the value of a group of goods or services over time. Unlike Laspeyre's index, which uses base period values as weights, the Paasche index uses current period values as weights.

$$Pp = \frac{p1q1}{p0q1} \times 100$$

$$= 263.404756$$

The Paasche's index number of approximately 263.40 indicates that the total selling price has increased by approximately 163.40% from the base period (2008) to the current period (2012). The rising Paasche's index number has suggested that there has been economic growth and expansion. It implies the fact that the total value of cars sold in 2012 is substantially higher than in 2008, reflecting economic progress or changes in market conditions favouring higher prices.

Fisher's Index Numbers:
The Fisher index is a composite index number which allows us to study the increase in the cost of living (inflation). It is the geometric mean of two index numbers: The Laspeyres index, and. The Paasche index.

Advantages
1) Corrects the upward bias and downward bias for laspeyer's and paasche's index numbers respectively

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1407-1412 www.ijhssm.org

Limitations

1) More complex construct
2) Quantities of the future years have to be forecasted

$$I_{01}^{F} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} X \frac{\sum p_1 q_1}{\sum p_0 q_1}} \; X 100$$

= 220.0347425

Dorbish and Bowley's Index Number

Dorbish and Bowley have suggested simple arithmetic mean of the two indices (Laspeyres and Paasche) so as to take into account the influence of both the periods, i.e., current as well as base periods.

Advantages

1) it is free from bias
2) This method considers values of both, the current year and the base year.

Limitations

1) It is tedious and time consuming
2) As the data of the current year and the base year is required, the data collection is costly and time-consuming

$$P_{01} = \frac{\frac{\sum p_1 q_0}{\sum p_0 q_0} + \frac{\sum p_1 q_1}{\sum p_0 q_1}}{2} \times 100$$

= 223.6052135

Marshall Edgeworth's Index Number

The Marshall Edgeworth Method for the index number, credited to Marshall (1887) and Edgeworth (1925), is a weighted relative of the current period to base period sets of prices. This index uses the arithmetic average of the current and based period quantities for weighting.

Advantages

1) simple to understand
2) easy to calculate

Limitations

1) It needs current weights every time an index is computed.
2) It needs all the data relating to price and quantities of both the base and current years for which the work becomes tedious and expensive.

$$P_{01}^{ME} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

= 241.2694705

Limitations of Index Numbers

1. **Quality of data:** Index numbers are only as reliable as the data they are based on. If the underlying data is inaccurate, incomplete, or biased, the index numbers may not accurately reflect the true changes in the variable being measured.
2. **Selection of base period:** The choice of base period can significantly affect the interpretation of index numbers. Different base periods may lead to different index values and can make it difficult to compare index numbers over time.
3. **Quality adjustments:** Index numbers may not always account for changes in the quality of goods or services over time. For example, improvements in product quality may not be reflected in the price index, leading to an overstatement of inflation.

.

Total Selling Price		Total Condition					
2008(p0)	2012(p1)	2008(q0)	2012(q1)	p1q0	p0q0	p1q1	p0q1
410050	579950	591	1072	342750450	242339550	621706400	439573600
87250	272150	96	571	26126400	8376000	155397650	49819750
182950	433950	259	678	112393050	47384050	294218100	124040100
110500	457250	156	789	71331000	17238000	360770250	87184500
218500	483000	222	885	107226000	48507000	427455000	193372500
87350	382300	83	562	31730900	7250050	214852600	49090700
93250	633250	151	1260	95620750	14080750	797895000	117495000
188850	393650	167	735	65739550	31537950	289332750	138804750
43400	351700	50	600	17585000	2170000	211020000	26040000
11750	130100	29	261	3772900	340750	33956100	3066750
194050	176400	279	356	49215600	54139950	62798400	69081800
180500	56400	194	123	10941600	35017000	6937200	22201500
				934433200	508381050	3476339450	1319770950

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1407-1412 www.ijhssm.org

REGRESSION ANALYSIS

Regression analysis is a statistical technique used to model the relationship between one or more independent variables (predictors) and a dependent variable (response). It aims to understand how changes in the independent variables are associated with changes in the dependent variable. Regression analysis is widely used in various fields, including economics, finance, psychology, sociology, and epidemiology, to analyse and predict relationships between variables.

The most common form of regression analysis is linear regression, where the relationship between variables is described using a straight line. The equation for a simple linear regression model with y on x can be represented as:

y = a + bx

Where:

→ y is the dependent variable

→ x is the independent variable

→ a is the intercept.

→ b is the slope coefficient

Year	Sum of selling price
1996	50300
1997	20250
1998	43550
1999	135550
2000	178800
2001	155300
2002	111350
2003	73300
2004	91800
2005	510050
2006	353450
2007	1407450
2008	1640150
2009	2143750
2010	3291000
2011	3105650
2012	1579850
2013	193000

In our project, we've taken the two variables as years (independent variable) and selling price (dependent variable). As talked about above, through our analysis, we can predict, explain, and control the movements of our sum of selling price in relation to the independent variable, years.

Through our data, we could find the summation of selling price for various years as shown above. In order to carry out our regression analysis, we are considering years to be "x" and selling price to be "y", hence we can use the equation of "y on x" which is "**y=a + bx"** where, x (i.e. years) is independent variable and y (i.e. selling price) is dependent variable.

We then find our 2 constants a & b by finding the intercept (a) and slope (b) of equation y on x.

Here we can see our a = 838030.5556 and our b = 70255.34056

	Coefficients
Intercept	838030.5556
X= Year- Middle year	70255.34056

Now we use these constants in our equation of y on x to derive the trend of the selling price throughout the given time period, this trend can further be used to predict the selling price of future years.

Year	Sum of selling price	X= Year-Middle year	Y estimated
1996	50300	-17	-356310.2339
1997	20250	-15	-215799.5528
1998	43550	-13	-75288.87169
1999	135550	-11	65221.80943
2000	178800	-9	205732.4905
2001	155300	-7	346243.1717
2002	111350	-5	486753.8528
2003	73300	-3	627264.5339
2004	91800	-1	767775.215
2005	510050	1	908285.8961
2006	353450	3	1048796.577
2007	1407450	5	1189307.258
2008	1640150	7	1329817.939
2009	2143750	9	1470328.621
2010	3291000	11	1610839.302
2011	3105650	13	1751349.983
2012	1579850	15	1891860.664
2013	193000	17	2032371.345
2014		19	2172882.026
2015		21	2313392.707
2016		23	2453903.388
2017		25	2594414.069
2018		27	2734924.751
2019		29	2875435.432
2020		31	3015946.113

Then these values were plotted on a graph with the green line indicating the trend and the orange line indicating the actual values. The trendline generated from regression analysis visually represents the overall direction and pattern of the relationship between variables in a dataset. There is an upward trend in the above graph. An upward trend in selling prices over time when using the least squares method could be attributed to a combination of factors such as inflation, changes in demand and supply dynamics, production costs, technological advancements, market competition and market trends.

Limitations of Least Square Method

The least squares method is a powerful tool for fitting a mathematical model to a set of data points by minimizing the sum of the squares of the differences between observed and predicted values. However, like any statistical technique, it has its limitations. Here are some of the main limitations of the least squares method:

1. **Sensitive to outliers:** The least squares method can be highly sensitive to outliers in the data. Outliers, or data points that deviate significantly from the rest of the data, can disproportionately influence the fit of the model, leading to biased estimates of the model parameters.

2. **Assumption of linearity:** The least squares method assumes that the relationship between the independent and dependent variables is linear. If the relationship is non-linear, the least squares estimates may be biased or inefficient, and alternative approaches may be more appropriate.

3. **Assumption of independence:** The least squares method assumes that the observations are independent of each other. If the observations are correlated or exhibit serial dependence, the least squares estimates may be biased or inefficient, and alternative modelling approaches may be needed.

III. Conclusion

In conclusion, our exploration of the automotive industry through the lens of quantitative techniques has provided valuable insights into the intricacies of car sales dynamics. By analysing historical data with the help of moving average analysis, coefficient of variation, index numbers, and regression analysis, we have uncovered trends and patterns. Armed with knowledge, people are better equipped to make informed decisions and formulate strategies that adapt to the ever-evolving landscape of the automotive market.

As technology advances and new data sources become available, opportunities for deeper insights and more accurate predictions will emerge, further enhancing our ability to navigate the challenges and capitalize on the opportunities within this dynamic sector.

References

Data Set-

[1]. https://www.kaggle.com/