Your Name

Instructor Name

Course Number

Date

Regression Analysis

Introduction

Applying linear regression analysis could prove useful for answering the business question - "What is the possible impact of advertising expenditure on the sales revenue?" Such a query derives inspiration from the scenario exposed in the Linear Regression Analysis Resources. To facilitate fitting a linear regression model, abundant data is presented in the Linear Regression Analysis Resources. The data's crucial characteristics to consider for such an analysis are as follows:

Variables

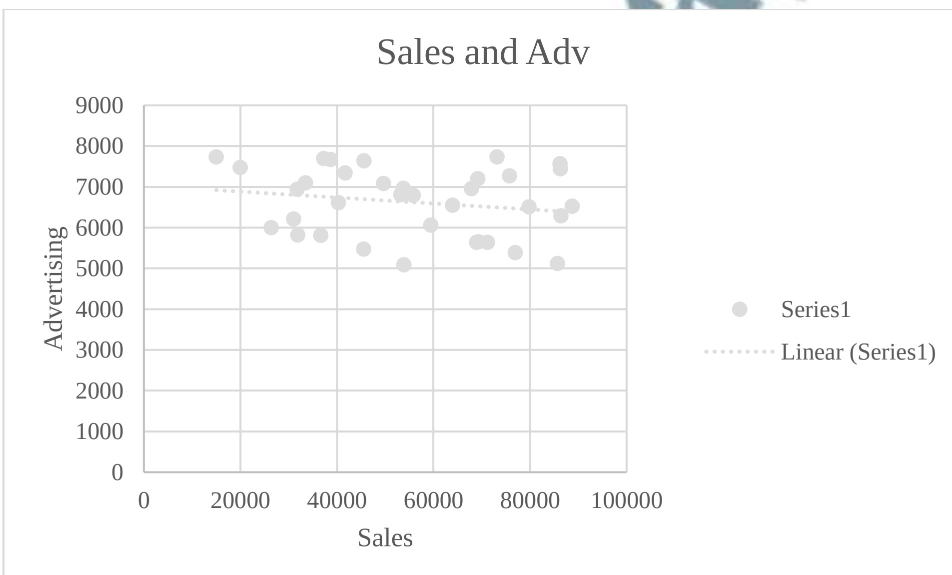
The independent variable under consideration is Advertising Expenditure. It is a key aspect of the marketing strategies implemented by firms and is often hypothesized to influence sales performance directly.

Sales Revenue functions as the dependent variable in this scenario. This factor is aimed to be forecast or explained with the help of the independent variables.

The type of data presented in the resources can be classified as numerical, more precisely, they are continuous. Both the Advertising Expenditure and Sales Revenue are expressed in dollar value, which can take any numeric value within a defined range(Salkind and Frey).

The quantity of data available would be significantly large enough to make secure assertions. Regression analysis requires a substantial volume of data for the results to be robust and reliable. In the current analysis, 35 values of each variable have been taken to assess the relationship. In summary, the strategic use of linear regression analysis using this data could provide critical insights into the relationship between advertising expenses and sales revenue. This could then inform effective decision-making around advertising strategies and budget allocation (Bennett et al.).

Figure 1: Scatter Plot between Advertisement and sales



The above diagram can be used to visualize the relationship between the two variables represented by the numbers in each line. Each observation is represented by a point in the graph, with its horizontal position determined by the first number and its vertical position determined by the second number. Analyzing given pairs of numbers, we can assume that these are coordinates (X, Y) where the X-values vary from approximately 10,000 to 87,000 and the Y-values from approximately 5,000 to 8,000. The scatter plot reveals whether there is a linear relationship between these two variables, or the pattern might suggest some other type of relationship. A positive relationship between the variables would be marked by a general trend that as the X variable increases, so does the Y variable. Judging by the numbers, it's hard to ascertain a clear pattern or trend, however.

There may also be outliers on the scatter plot - points that do not seem to fit the general trend of the data. For example, the pair (10307, 5291) seems to be an outlier as most of the X-values in the data hover around the 60,000 range. It is important to note that a scatter plot can only show correlation, not causation. Even if we see a pattern in the data, we can't necessarily say that changes in the X variable cause changes in the Y variable. The points on the scatter plot may also be dense in areas with frequent values, showing clusters, and sparse in areas with less frequent values. By observing the data set, there seems to be a concentration of values around (60,000 – 70,000) for the X variable and for the Y variable, the values seem to be concentrated around (6,000 – 7,000).

Regression

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.044584							
R Square	0.001988							

Adjusted R Square	-0.02826							
Standard Error	26149.43							
Observations	35							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	44941843	44941843	0.065724	0.799259			
Residual	33	2.26E+10	6.84E+08					
Total	34	2.26E+10						

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	39126.19	32756.68	1.194449	0.240817	-27517.8	105770.2	-27517.8	105770.2
X Variable 1	1.313998	5.125445	0.256368	0.799259	-9.1138	11.7418	-9.1138	11.7418

The regression analysis above is conducted to determine the relationship between advertisement (independent variable) and sales (dependent variable). A brief overview of the given statistics provides insights into this relationship. The multiple R-value (coefficient of correlation) is 0.044584. This indicates a very weak positive relationship between the dependent variable sales and the independent variable advertisement, as it is very close to zero(Brase and Brase). Essentially, it signifies that there is a negligible linear correlation between our variables.

The R-Square (coefficient of determination) value stands at 0.001988, which is very low. This indicates that only approximately 0.2% of the variation in sales can be explained by the advertising expenditure. The remaining variance will be attributed to other factors not considered in this model. This could be factors like pricing, economic conditions, competition, and so on.

The adjusted R square, which is a more accurate reflection of the strength of the model, is at -0.02826. Ordinarily, the adjusted R Square is supposed to be less than or equal to R Square but not negative(Sirkin). However, a negative adjusted R-squared implies that the model is inappropriate and does not fit the data.

The standard error statistic measures the average amount that the observed values deviate from the regression line as a result of error (i.e., an amount of variation unexplained by the predictor). Here, it is quite high (26149.43) implying substantial errors and the model's poor explanatory power. The ANOVA table reveals a very high p-value (0.799259) of the F statistic. Usually, a p-value lesser than 0.05 is considered statistically significant. The high p-value in this case implies that we fail to reject the null hypothesis that there is no relationship(Wolfe and Schneider).

Furthermore, examining the coefficients, one unit increase in advertisement results in approximately 1.31 units increase in sales, however with a very large standard error (5.125445) and again, a high p-value (0.799259).

In essence, this model is ineffective with insignificance and inappropriateness in explaining the relationship between advertisement and sales due to a very weak positive relationship, low R square (indicating that advertisement explains a very small proportion of the variations in sales), large standard errors and high p-values. Hence, other factors affecting sales should be considered to build a stronger model. Businesses should spend more on advertisement to retrieve more sales revenues.

Works Cited

Bennett, Jeff, et al. *Statistical Reasoning for Everyday Life*. Pearson Education, 2017.

Brase, Charles Henry, and Corrinne Pellillo Brase. *Understanding Basic Statistics*. Brooks/Cole, 2009.

Salkind, Neil J., and Bruce B. Frey. *Statistics for People Who (Think They) Hate Statistics*. SAGE Publications, 2019.

Sirkin, R. Mark. *Statistics for the Social Sciences*. SAGE, 2006.

Wolfe, Douglas A., and Grant Schneider. *Intuitive Introductory Statistics*. Springer, 2017.

Works Cited

AuthorLastName, FirstName. *Title of the Book Being Referenced*. City Name: Name of Publisher, Year. Type of Medium (e.g., Print).

LastName, First, Middle. “Article Title.” *Journal Title* (Year): Pages From - To. Print.