MULTIVARIABLE REGRESSION

The following dataset is a record of sales of a company and its investments on advertisements through TV and Newspapers, with this dataset we will try to predict a regression model best suited to the data set and check its relevance to the dataset.

|  |  |  |
| --- | --- | --- |
| Sales (in Cr.) | Newspaper (in lakhs) | TV (in lakhs) |
| 4.8 | 1 | 8.6 |
| 7.2 | 75 | 8.7 |
| 9.2 | 65.9 | 23.8 |
| 10.4 | 45.1 | 44.5 |
| 11.3 | 18.3 | 69.2 |
| 11.8 | 23.5 | 57.5 |
| 12 | 69.3 | 17.2 |
| 12.5 | 114 | 67.8 |
| 12.6 | 24.2 | 66.1 |
| 13.2 | 11.6 | 120.2 |
| 13.7 | 7.2 | 97.5 |
| 14.6 | 19.1 | 147.3 |
| 15.6 | 21.2 | 199.8 |
| 16.5 | 58.5 | 151.5 |
| 17.4 | 4 | 214.7 |
| 17.9 | 58.4 | 180.8 |
| 19 | 46 | 204.1 |
| 22.1 | 69.2 | 230.1 |
| 22.4 | 52.9 | 195.4 |
| 24.4 | 55.8 | 281.4 |

**CORRELATION: -**

****

We can see that the relation between Sales and TV variables is having a very high degree of correlation.

**DESCRIPTIVE STATISTICS: -**



**SCATTER PLOT: -**



******

******

**REGRESSION OUTPUT: -**

Here Sales is taken as dependent variable and Newspaper, TV are taken as independent variables because in general the sales of a product depend on the number of investments done in the advertisements or mode of advertisements (as in this case).



**Estimated model:**  Sales = 6.269 + 0.033\*(Newspaper) + 0.057\*(TV)

**CONCLUSION: -**

As we look at the regression output, we can derive that the R-squared value is greater than 0.9 i.e., 0.906 which means very less data deviates from the model that we estimated and the probability values are also less than 0.05 apart from this the prob(F-statistic) value is 0.00 which is a good sign. From the correlation we can say that the dependence of sales revenue is biased towards the TV investments than the newspaper investments, which is quite justifiable. Overall, we can conclude that the model we estimated is a good fit for the given dataset.

NAME: S.KRISHNAN

REG.NO: 200968068

ROLL.NO: 19

SEC: DSE-A