report

Lydia Maher

October 14, 2016

Abstract

A report detailing the performance of a simple regression analysis, based on the examples given in chapter 3 of the book **An Introduction to Statistical Learning**. In this report, we reproduce some of the graphics and statistics included in that chapter.

Introduction

The overarching goal of the project is to see the how the amount of sales is affected by the variables TV, Radio, Newspaper. In order to do this, we undertake analysis to see whether there is a concrete relationship between sales and amount of money spent on advertising. This advertising is split into three different media formats: TV, Radio and Newspaper. By developing a model which predicts sales based on funds allotted to each media domain, we can determine the optimal allocation of funds and achieve the greatest amount of sales.

Data

We are using the dataset given in the book entitled "Advertising.csv". It contains the data for 200 different markets, with the amount spent on TV, Radio and Newspaper advertising (in thousands of dollars) and the amount of sales this spending produced (in thousands of units).

Methodology

To carry out this analysis, we are using the multiple linear regression model: Sales = (Beta0) + (Beta1)TV + (Beta2)Radio + (Beta3)Newspaper + error. These coefficients are estimated with regression and least-squares in R.

Results

Here is a table with the resulting co-efficient values:

Further analysing our least squares model, it is useful to look at some quality indices:

Here are scatterplots showing the data compared to the regression (error for each point also drawn onto the graph):

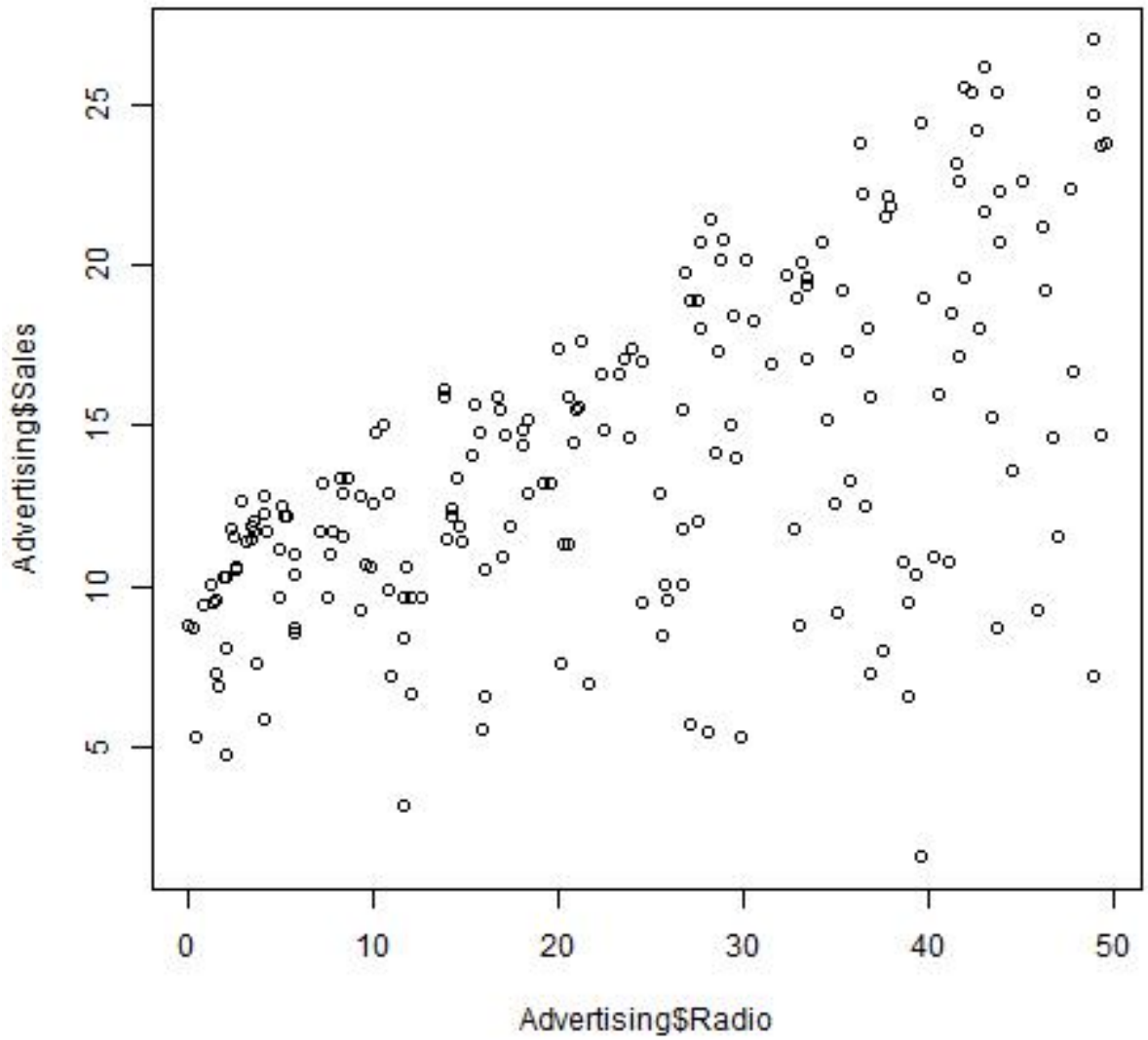

Figure 1: Radio Scatterplot

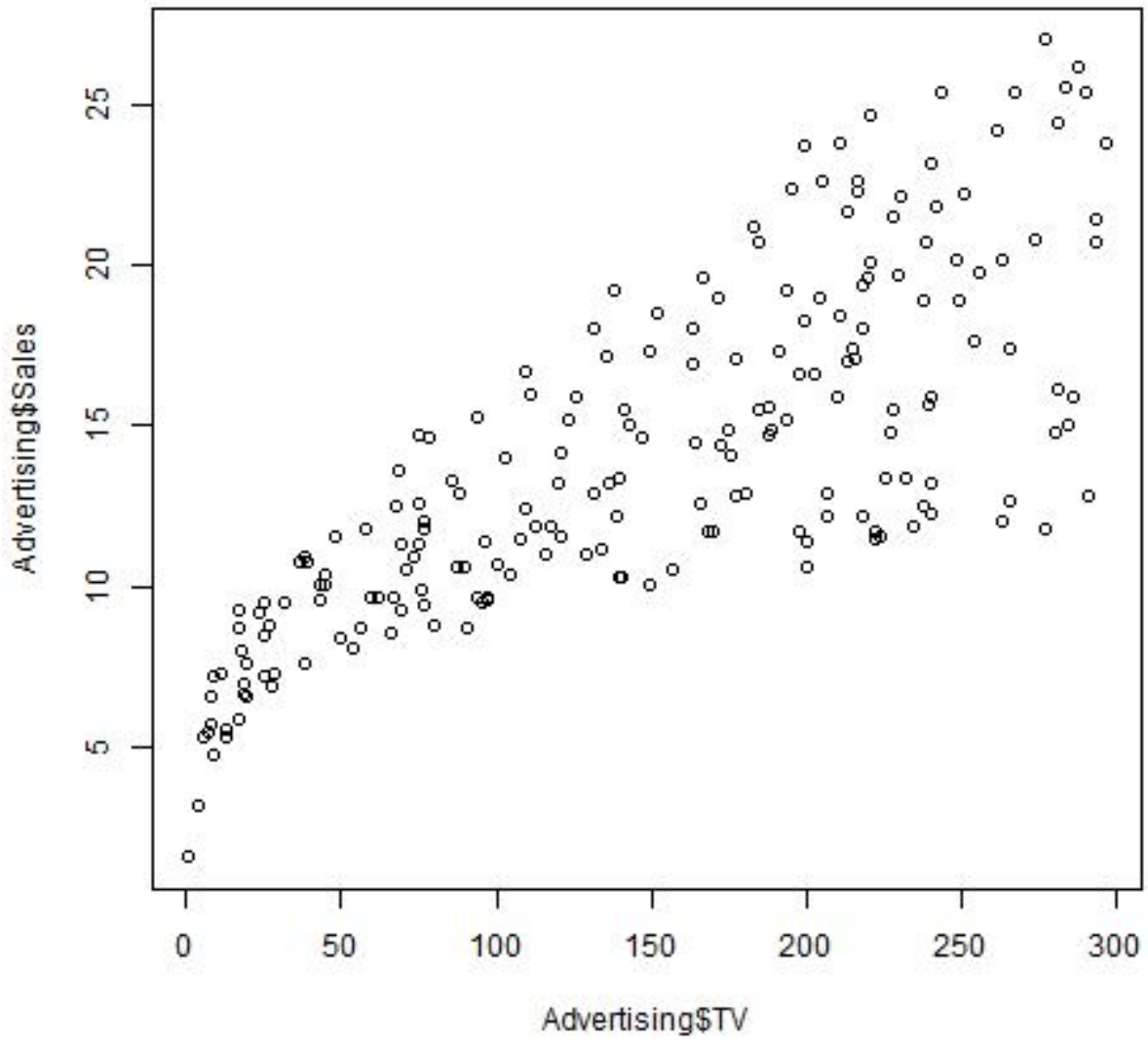

Figure 2: TV Scatterplot

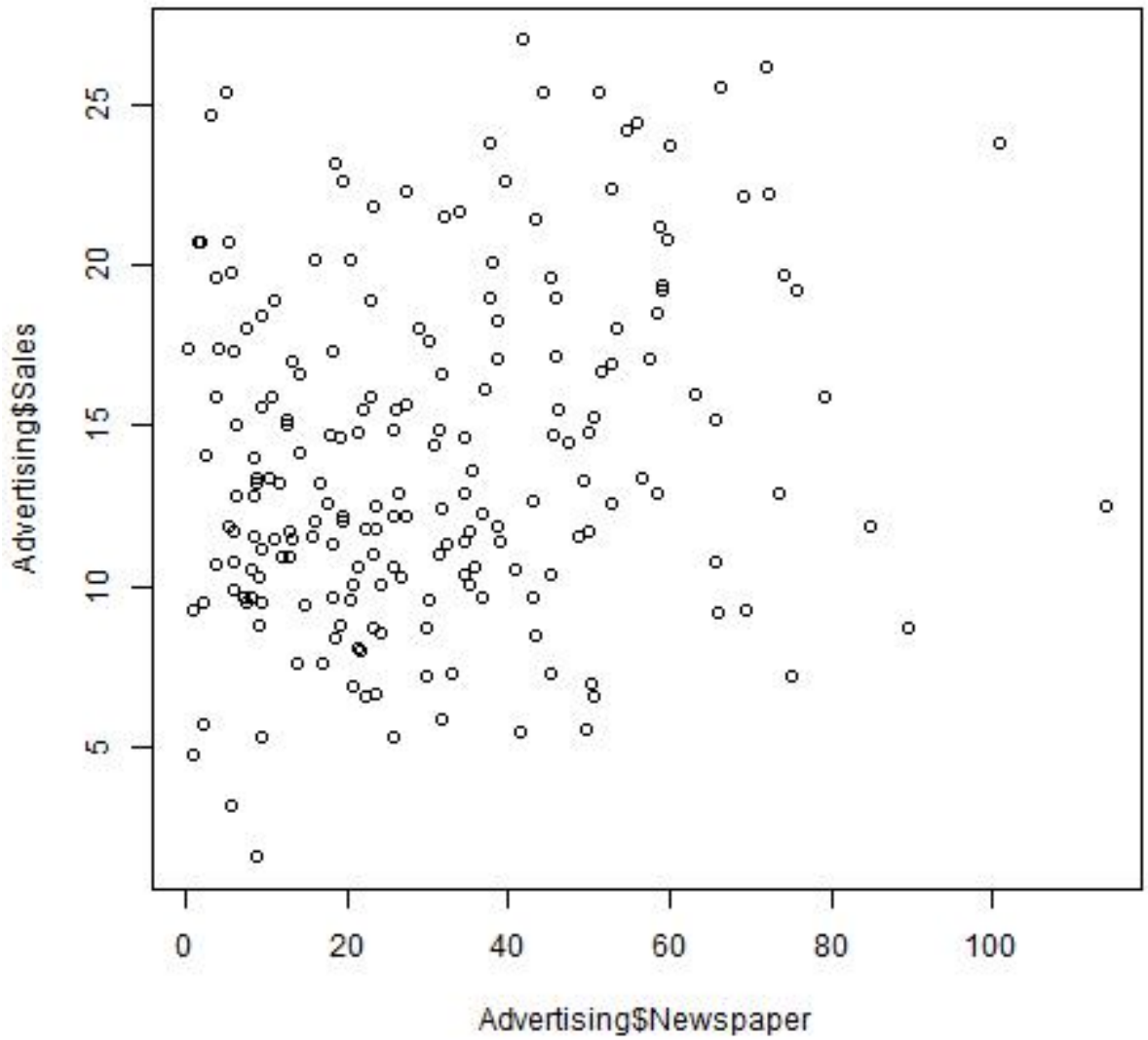

Figure 3: TV Scatterplot

This shows graphically that there is quite a lot of error, but also a positive correlation.

Conclusions

According to our data, there is a positive correlation between TV and Sales of about 0.6. However, our model is by no means perfect and the simple regression has a lot of error as shown by the RSS statistic of 3.26. It would be useful if we had more data to be able to refine our model. Based on this simple analysis though, it would seem that more spending on TV advertising will result in more sales.