

Simple Regression Analysis

Minsu Kim

10/7/2016

Abstract

In this analysis, I explore the statistical relationship between advertising budget and its effectiveness. In particular, I focus on the linear relationship between the increase in the budgets spending on TV advertisement and the increase in the number of items sold by reproducing the results from the book An Introduction to Statistical Learning.

Introduction

Nowadays, data analytics is often utilized in business sectors to effectively predict or forecast sales and earnings. In this analysis, I closely look at the relationship between advertising and sales, and implement a simple regression model to predict sales from budget spending on TV advertisement. Over the course of analysis, I first reproduce the results from the book An Introduction to Statistical Learning.

Data

The advertising data set contains sales (in thousands of units) data as well as advertising budgets (in thousands of dollars) of a certain product in 200 different markets. There are three predictors namely TV, Radio and Newspaper as well as one response variable, Sales. All of the predictors are numerical values.

Table 1: Advertising data set				
	TV	Radio	Newspaper	Sales
1	Min. : 0.70	Min. : 0.000	Min. : 0.30	Min. : 1.60
2	1st Qu.: 74.38	1st Qu.: 9.975	1st Qu.: 12.75	1st Qu.:10.38
3	Median :149.75	Median :22.900	Median : 25.75	Median :12.90
4	Mean :147.04	Mean :23.264	Mean : 30.55	Mean :14.02
5	3rd Qu.:218.82	3rd Qu.:36.525	3rd Qu.: 45.10	3rd Qu.:17.40
6	Max. :296.40	Max. :49.600	Max. :114.00	Max. :27.00

Table 1 shows the summary of the data set. The budget spending on TV is the highest on average. This intuitively makes sense because TV advertisement is often very expensive. Also, Newspaper seems to have a highly variance. So, it might be worth investigating what budget spending on newspaper has the high variance. However, in this analysis, I will mainly focus on TV and Sales columns.

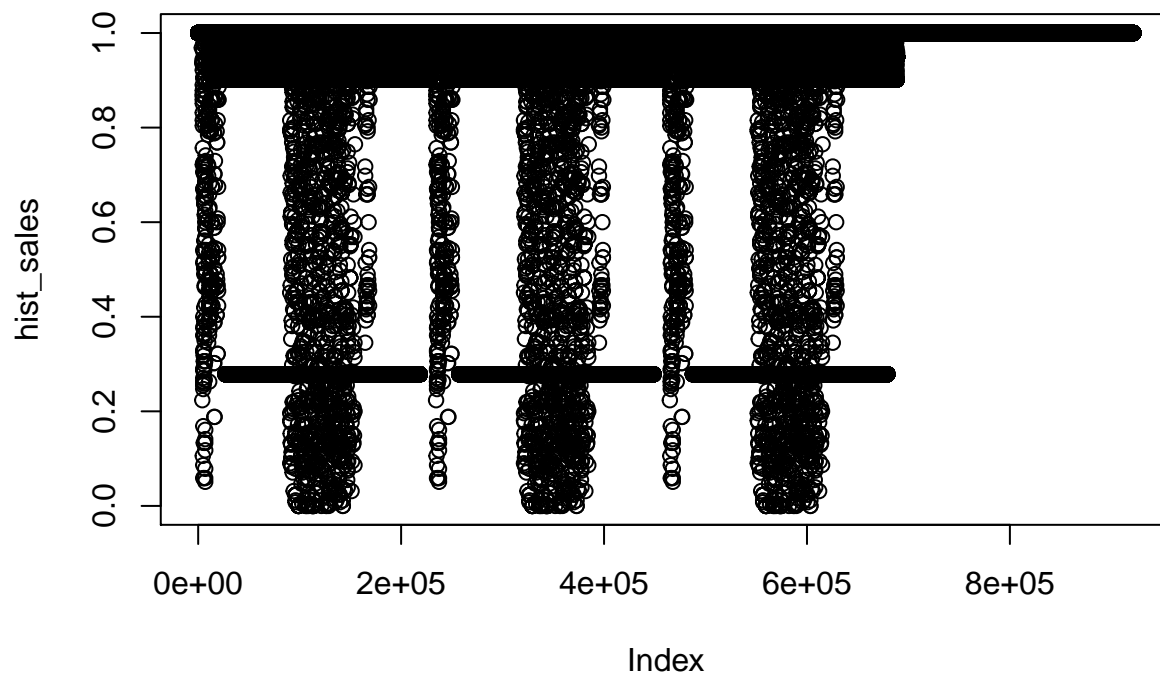
Methodology

In order to investigate the linear relationship between TV and Sales, I use a simple linear regression. It is a very straightforward approach and commonly used model for predicting a quantitative response Y on the basis of a single predictor variable X. It assumes that there is approximately a linear relationship between predictor and target variable. Mathematically, we can write this linear relationship as

In the context of this analysis, X represents TV advertising and Y represents sales. By fitting the model, I can regress sales onto TV,

Before fitting the model, I first look at the general distributions of both TV and Sales.

```
hist_sales = readPNG("../images/histogram-sales.png")  
plot(hist_sales)
```



Results

Conclusions