Stat 159 HW 2

Leanne Lee

October 5, 2016

Abstract

This homework is to reproduce the analysis from Section 3.1 of Ch 3 Linear Regression from the book "An Introduction to Statistical Learning" (by James et al). The homework will analyze the advertising dataset with linear regression and summary statistics.

Introduction

From the dataset advertising, I can reproduce the simple linear regression with TV and sales. The simple linear regression can predict a quantitative response Y based on predictor variable X. With the modeling, we can tell if there is a relationship exists between TV advertising and Sales. If there exists a positive relationship, the marketing team can decide to increase Tv advertising budget to promote their sales.

Data

The dataset *Advertising.csv* comes from _"http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv_. It consists for TV, Radio, Newspaper and Sales columns. The structure of the columns are stored in numeric vectors.

Methodology

By using the simple linear regression, we can predict the future sales based on the amount of TV advertising.
The simple linear regression equation is the following:

```
Y = A + Bx + e
```

Y = Sales A = Intercept Bx = TV advertising e = error

In R command, we can fit a linear regression model by using the *lm()* command. The null hypothesis is there does not exist a relationship between TV ad and sales. The alternative hypothesis is that there exists a relationship between TV ad and sales. ##Results

```
load('./data/regression.RData')
library(xtable)
reg_table <- xtable(reg_summary)
reg_table
```

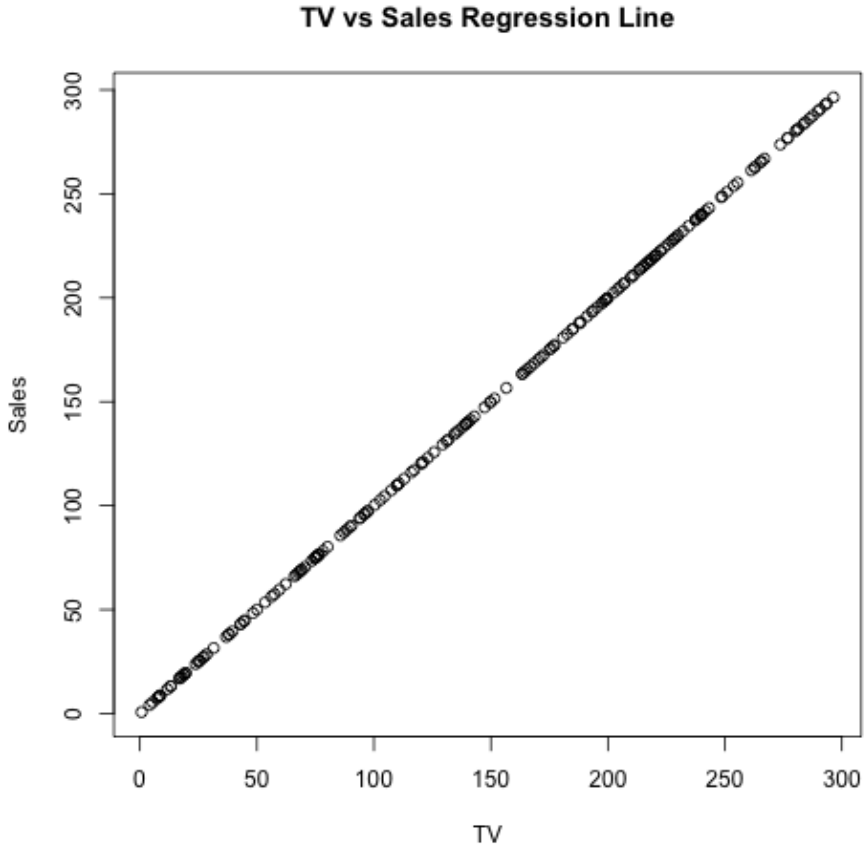

Figure 1: TV ad vs sales linear regression scatter plot

Conclusions

This homework shows the relationship between TV advertising and sales. There exists a positive correlation between the TV advertising and sales. In the future, we can apply the same model to radio and newspaper.