Advertising Data Report

Lily Li

October 7, 2016

Abstract

This reports attempts to reproduce the results of advertising and sales found in section 3.1 Simple Linear Regression of *An Introduction to Statistical Learning*.

Introduction

The Advertising dataset contains data on sales (in thousands of units) for a particular product as a function of advertising budgets (in thousands of dollars) for TV, radio, and newspaper media. The goal is to suggest, on the basis of this data, a marketing plan for next year that will results in high product sales. In this report, we will focus specifcally on the TV budget and its relationship with sales. Some questions we would like to explore include:

1. What are the average TV ads budget and average sales in this dataset?

2. Is there a relationship between TV ads budget and sales?

3. How strong is the relationship?

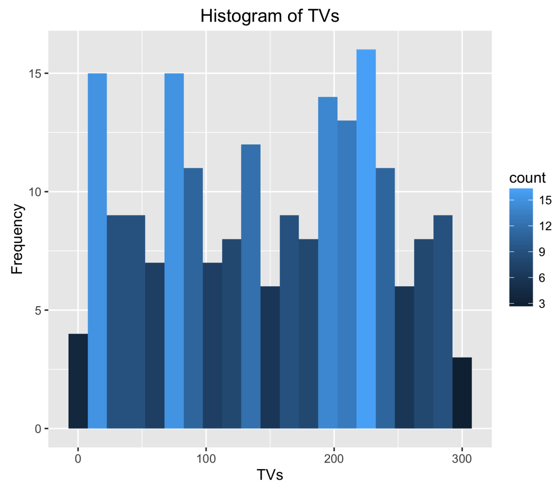
4. Is the relationship linear?

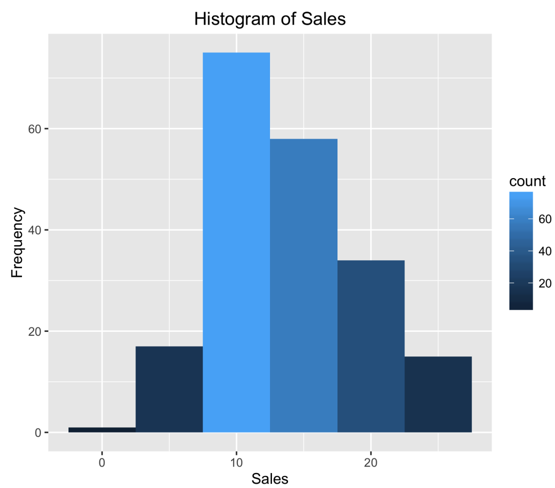
Data

This dataset has information on TV, Radio, and Newspaper budgets. On this paper, we will focus on the relationship between *TV budget* and *product sales*.
Some preliminary analysis of the dataset include:

-there are 200 observations of each TV budget and sales
-histograms below show the distribution of TV budget and sales data

Figure 1: Distributions of TV Budget and Sales Data





Methodology

We want to observe if there is a linear relationship between TV budget and Sales. Let's consider the regression model:

$$Sales = \beta_0 + \beta_1 TV$$

To estimate the coefficients β_0 and β_1, we use the least squares minimization method.

Results

After computing the regression, we found the following results:

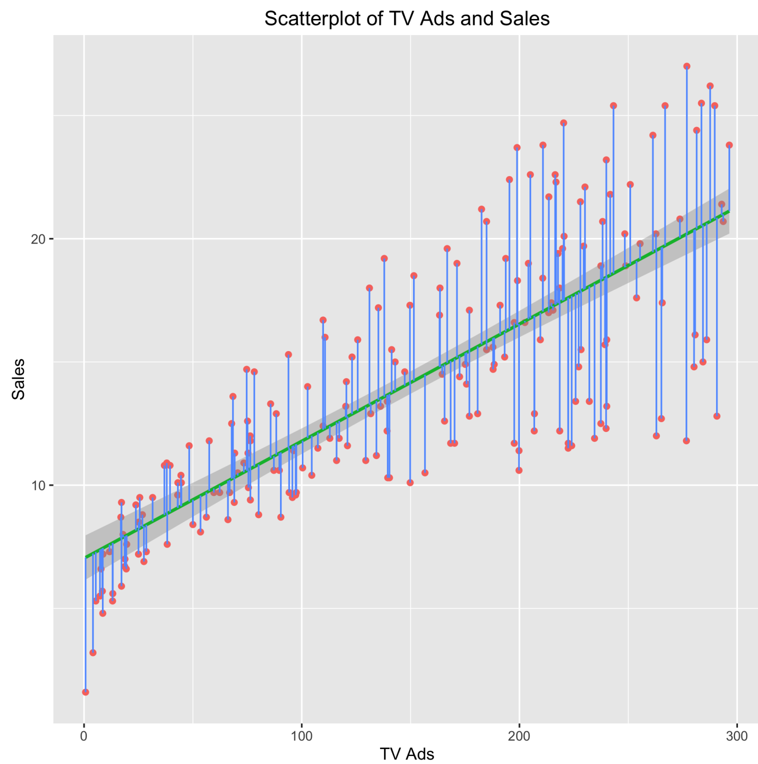
Table 1: Regression Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.0326	0.4578	15.36	0.0000
TV	0.0475	0.0027	17.67	0.0000

Table 2: Regression Quality Statistics

	Value
r_squared	0.61
f_stat	312.14
rse	3.26

Figure 2: Scatterplot of Regression



Conclusions

Looking at the *TV* regression coefficent in **Table 1** shows a positive relationship with *sales*. In fact, 1% increase in *TV* budget is associated with 4.75% increase in *sales*. The t statistic of this coefficient is so large that the p-value is less than 4 decimal places of 0. This coefficient is statistically significant.

The R squared in **Table 2** shows that 61% of the variation in sales is explained by the variation in TV budget.

However, looking at the scatterplot (**Figure 2**), sales seems to vary more with a higher TV budget and when budget is near 0, sales seem to grow expoentially with TV budget. The data is not homoscedastic and other types of regression might be worth exploring.