Simple Regression Analysis

Elly Wang 10/7/2016

Abstract

In section 3.1 - Simple Linear Regression of the book An Introduction to Statistical Learning, we used the data set taken from advertising and performed simple linear regression of Sales on TV. In this report we will reproduce the graphs, regression, and analysis in a reproducible manner.

Introduction

The purpose of advertising for a certain good is to hopefully increase its sales. With the technologies available today, there are many possible channels for advertisements to reach their audience. In this report, we will focus on the relationship between the budget alloted for TV advertisement and Sales and see if we can determine whether there is a relationship between the two variables.

Data

The advertising data used in the book consists of **Sales** (in thousands of units) of a particular product in 200 different markets and advertising budgets (in thousands of dollars) for the product in each of those markets. In particular, the advertising budgets were for **TV**, **Radio**, and **Newspaper**.

Methodology

In this paper, we will be focusing in particular on the relationship between the Sales (in thousands of units) and TV budget (in thousands of dollars). In particular, our model for the simple linear regression will be:

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

To esimate the coefficients for β_0 and β_1 , we will use the ordinary least squares regression in R.

Results

Using OLS, we get the estimates of the coefficients as shown in Table 1.

	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

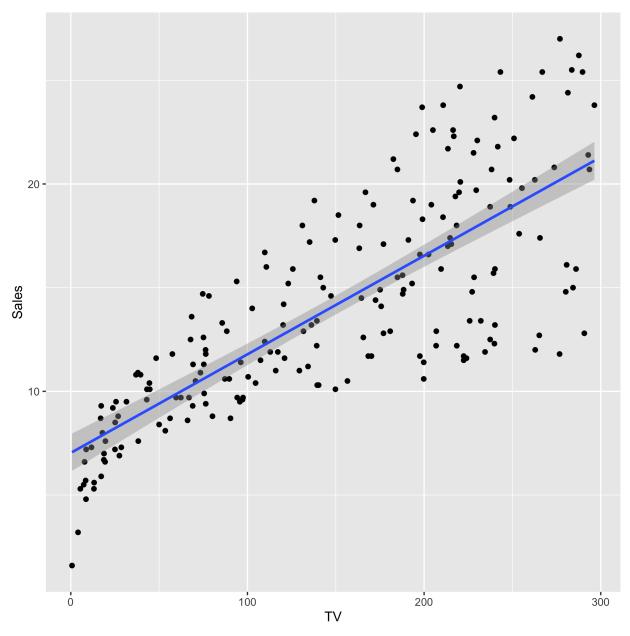
Table 1: Information about Regression Coefficients

Regarding the regression quality and the fitness of our regression, we reference the following table (Table 2).

The scatterplot of Sales on TV with the regression lines fitted is shown below:

Quantity	Value
Residual standard error	3.26
R Squared	0.61
F-Statistics	312.14

Table 2: Regression Quality Indicies



Conclusions