

Simple Linear Regression

Zhongling Jiang

November 2, 2016

Abstract

In this report, we will look at the Advertising.csv dataset and reproduce the regression result covered in book session 3.1 Simple Linear Regression of the book An introduction to the Statistical Learning.

1 Introduction

The goal of our analysis is to provide advice on improving sales of a particular product. More specifically, we are looking at the relationship between sales and advertising. If so, we want to quantify the relationship between them.

2 Data

The Advertising.csv dataset contains Sales (in thousands of units), advertising budget (in thousands of dollars) for three different media: TV, Radio and Newspaper.

3 Methodology

We focus on one of the independent variables: **TV**. Assume the relationship between the response and independent variable forms a simple linear relationship. Then we have the model:

$$Sales = \beta_0 + \beta_1 * TV \quad (1)$$

4 Result

We run the regression analysis in R by using `lm()` function. The resulting table shows regression coefficients, their standard error, and their p-value.

	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

More information above least-square model is below:

5 conclusion

From the regression statistics, we find that p-value for slope is $1.4673897 \times 10^{-42}$ which shows there is a significant correlation between two variables.

However, this is not enough to conclude that the simple linear model is a good fit to the data. We still have to look at residual plots to determine whether residuals follow a random pattern.

Quality	Value
RSS	0.6118751
R2	3.2586564
F-stat	312.1449944