

Simple Regression Analysis

Yuyu Zhang

Abstract

Linear regression is an important approach in modeling. In this report we reproduce the main results presented in section 3.1 *Simple Linear Regression* of the book **An Introduction to Statistical Learning**

Introduction

The overall goal is to provide advice on how to improve sales of the particular product. More specifically, the idea is to determine whether there is an association between advertising and sales, and if so, develop an accurate model that can be used to predict sales on the basis of the three media budgets.

Data

The Advertising data set consists of the **Sales** (in thousands of units) of a particular product in 200 different markets, along with advertising budgets (in thousands of dollars) for the product in each of those markets for three different media: **TV**, **Radio**, and **Newspaper**.

Methodology

We consider one media from the data set, **TV**, and study its relationship with **Sales**. For this purpose, we use a simple linear model: $\text{Sales} = \beta_0 + \beta_1 \cdot \text{TV}$. To estimate the coefficients β_0 and β_1 we fit a regression model via the least squares criterion

Results

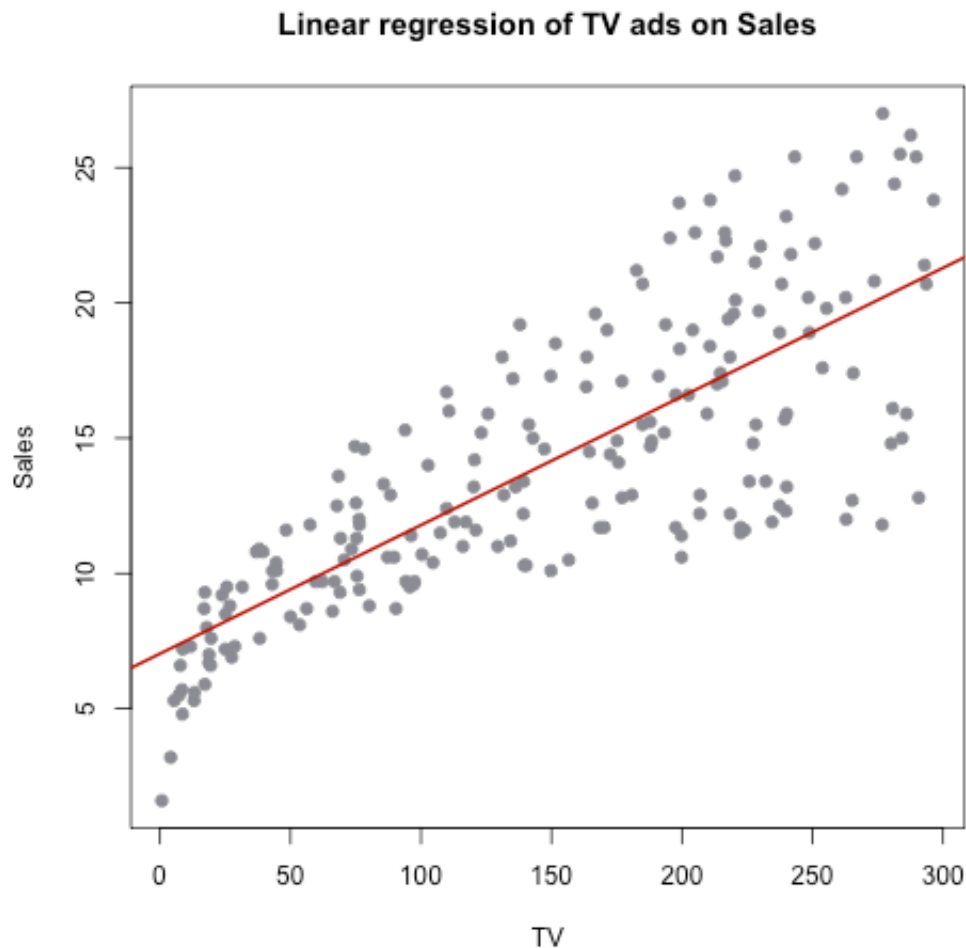
We compute the regression coefficients. The regression coefficients is summarized in Table 1 below:

Table 1: Information about regression coefficients

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.03	0.46	15.36	0.00
TV	0.05	0.00	17.67	0.00

More information about the least squares model is given in the table below:

The scatterplot of the fitted regression line is shown in Figure below.



Analysis

Visual results from Figure 1 indicates that a linear fit captures the essence of the relationship, although it is somewhat deficient in the left of the plot. From Table 1 we know that β_0 is 7.03 and β_1 is 0.05, which means that an additional \$1000 spent on TV advertising is associated with selling approximately 47.5 additional units of the product. Meanwhile, the *p-value* is close to 0, which means there **is** an association between TV ads and sales. From Table 2, R square is 0.61, which indicates that just under two-thirds of the variability in **Sales** is explained by a linear regression on **TV**. Therefore, combining with all the above analysis, we can conclude reasonably that TV advertising does have some effects on Sales of the targeted product.

Conclusion

In conclusion, by taking data and using linear regression model, computing all relevant coefficients, we see a relatively linear relationship between TV ads and sales, and hence a non-trivial effects of TV advertising on product sales.

Acknowledgement

I would like to thank the author of the book **An Introduction to Statistical Learning** Gareth James, who provides the data and such a valuable practice lesson. I would also like to thanks the TA and professor for providing help to finis this assignment.

