

Simple Regression Analysis

Abstract

In this document, we will reproduce the main results of section 3.1 Simple Linear Regression of the book An Introduction to Statistical Learning. Simple linear regression is a very straightforward and useful method for predicting a quantitative response Y on the basis of a single predictor variable X . Linear regression would assume a linear relationship between X and Y , and has two parameters: slope and intercept. In this case, we will predict a response variable Sales from the predictor variable TV and visualize the result in a scatter plot with regression line.

Introduction

The main goal of this analysis is to find a way to improve sales of the particular product. In detail, we are trying to find if there is a linear relationship between sales and the other three variables (TV, Radio, Newspaper) and use the linear model to predict sales based on TV, Radio, Newspaper. From the model, we can analyze which one of the three factors contribute most to sales, and hence improve our sale records accordingly.

Data

In this analysis, we use the dataset of Advertising data in the book An Introduction to Statistical Learning. The dataset consists of four variables including three predictors (TV, Radio, Newspaper) and one response variable (Sales). The dimension of the dataset is 200×4 , which means it consists the sales of a particular product in 200 different markets.

Methodology

In this case, we only use one predictor variable namely TV, and study its relationship with the response variable Sales. With this purpose, we can just use a simple linear regression with $\text{Sales} = \text{Beta-0} + \text{Beta-1} \times \text{TV}$ where Beta-0 is the intercept and Beta-1 is the slope of the regression line. Hence we can formulate a linear regression via the least square criterion.

Results

We first compute the coefficients of the regression line to visualize how one unit of TV contribute to the sale and also the standard error and t value of the regression line.

Table 1: Information about regression coefficients

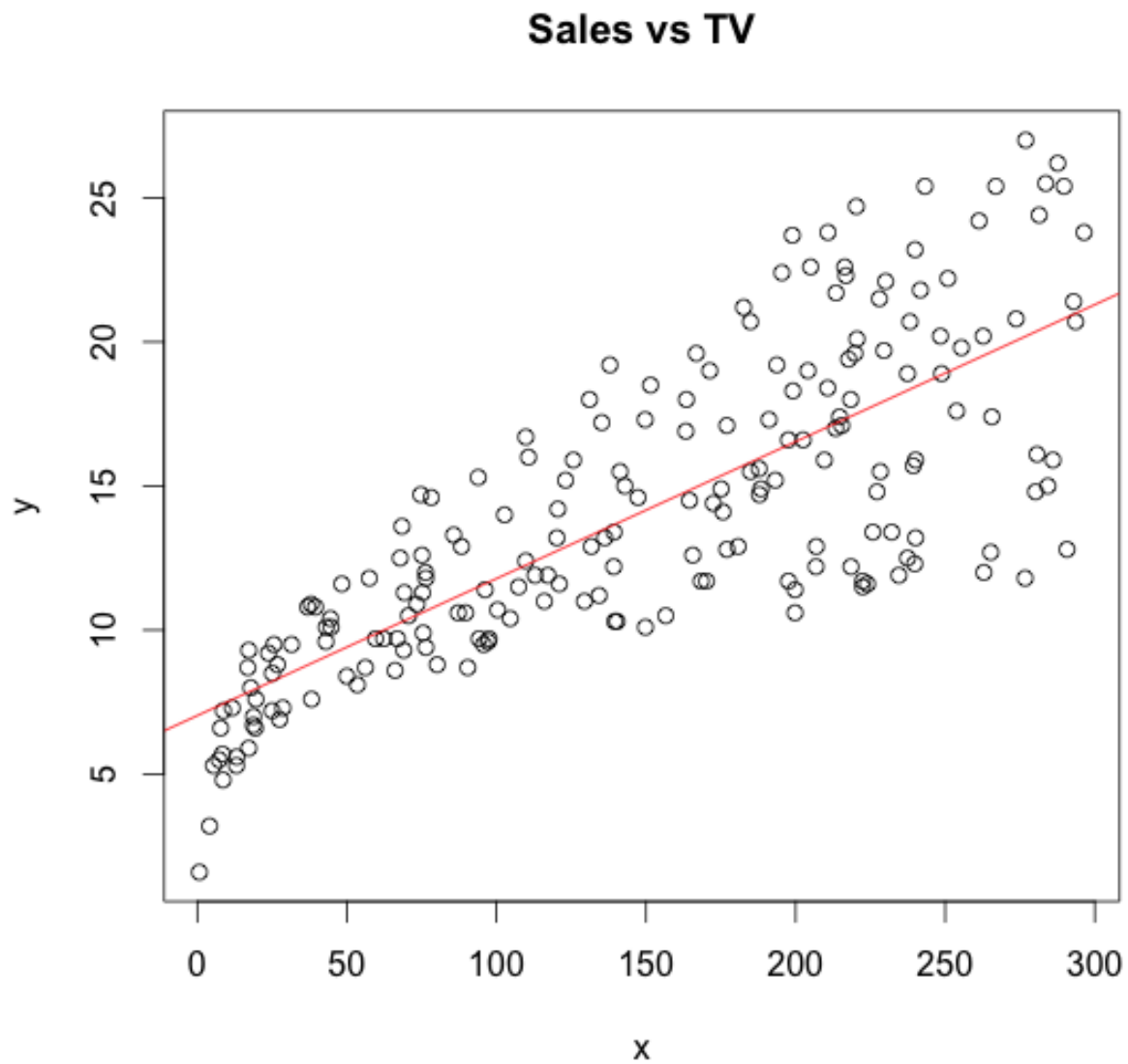
##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	7.03259355	0.457842940	15.36028	1.40630e-35
## x	0.04753664	0.002690607	17.66763	1.46739e-42

Furthermore, we can also get more details of regression model by computing RSS, R square, and F statistics. The results of these statistics are displayed in the table

Table 2: Regression Quality Indices

##	Quantity	Value
## A	RSS	3.25865636865046
## B	R2	0.611875050850071
## C	F-stat	312.144994372713

Finally we will visualize the scatter plot of Sales vs TV and the fit the regression line in the plot



Conclusion

In this analysis, we analyzed the relationship between the predictor variable TV and the response variable Sales and fit a linear regression to predict the response variable Sales from TV. Then we compute statistics to study the regression line such as RSS, R square and F statistics. Finally we visualize the regression line with the scatter plot to provide a visual understanding of the dataset.