

# Simple Regression Analysis

Thomas Sun

October 7, 2016

## Abstract

This report attempts to reproduce the results found in Chapter 3 of the book **An Introduction to Statistical Learning**. In this chapter, a regression analysis is run on the *Advertising* dataset, containing data on sales and advertising budget for a particular product. Using a simple linear regression model, I find the same estimates of the coefficients, obtain the same quality index results, and similar looking plots as the ones contained in the book.

## Introduction

The goal is to determine whether there is an effect of advertising on sales, ideally to increase product sales. Specifically, if increasing spending on certain mediums of advertising has a relationship with the amount of sales on a product. The chapter mainly considers one medium, **TV**, and fits a regression model to it with **Sales**. It finds that there is a strong positive relationship between **TV** and **Sales** and the data points fit the regression line closely.

## Data

Data was obtained by downloading the *Advertising* dataset available on the textbook's website. It contains data of the size of advertising budget for **TV**, **Radio**, and **Newspaper** (in thousands of dollars) for a product in 200 different markets, in addition to the number of sales (in thousands of units) for the product in each market. We specifically are interested in the data for **TV** and **Sales**.

## Methodology

In order to estimate the relationship between **TV** and **Sales**, a simple linear model is used.

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

Where  $\beta_0$  and  $\beta_1$  are the regression coefficients. To find an estimate for these coefficients, we use the ordinary least squares method to fit the model. OLS regression was run through RStudio, where the regression coefficients and quality indices (mean squared error, R-squared, F-statistic) are calculated. We also plot **TV** against **Sales** to replicate the scatterplot in the chapter.

## Results

Table 1: Regression Coefficients

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.033	0.458	15.360	0
TV	0.048	0.003	17.668	0

More information in table below

Table 2: Regression Quality Indices

Quantity	Value
MSE	3.25865636865046
R-squared	0.611875050850071
F-Stat	312.144994372713

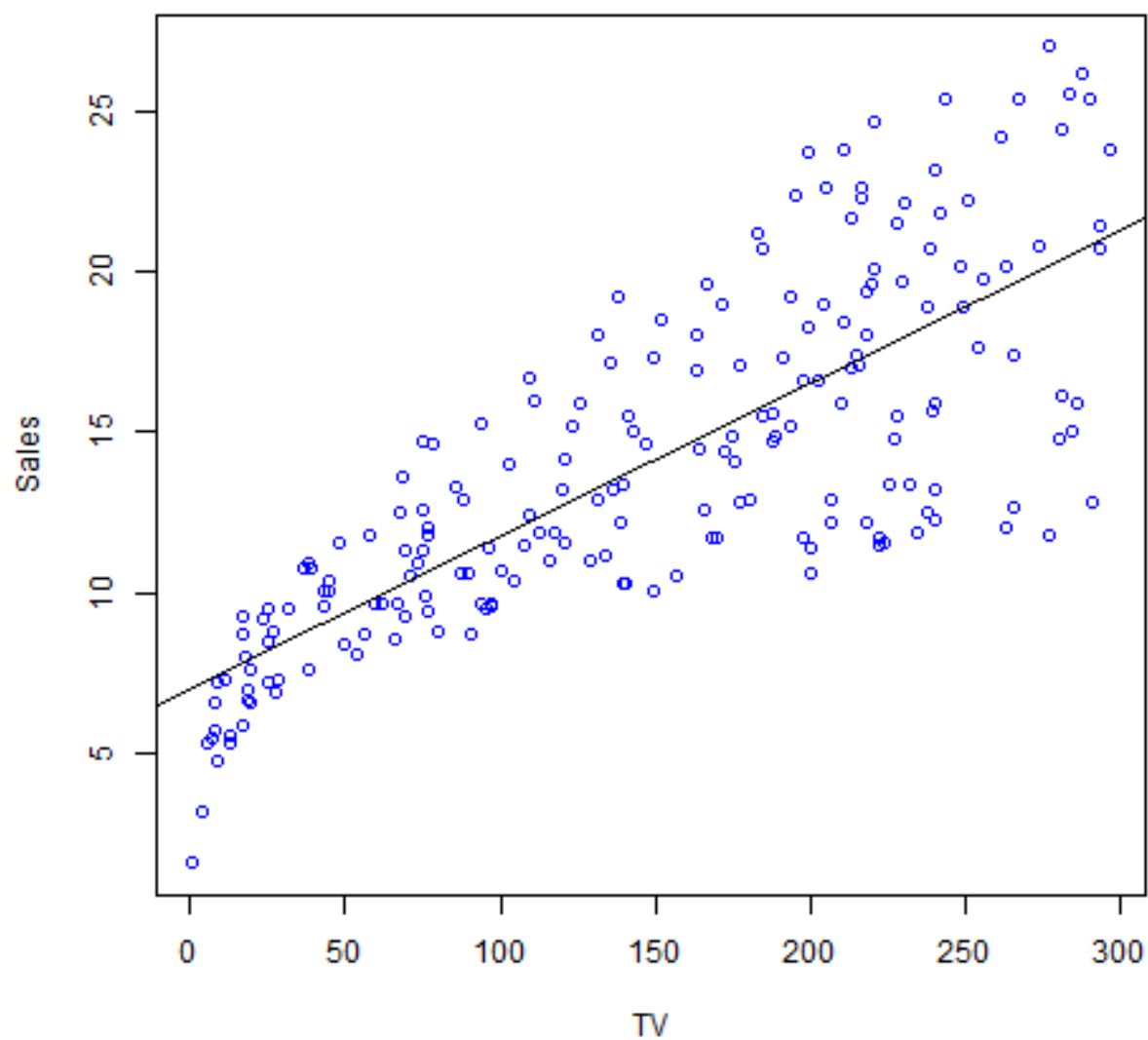


Figure 1: Scatterplot of TV and Sales with fitted regression line

## Conclusions

I successfully managed to reproduce the regression analysis and results found in the textbook.