Report

Atul Lanka

October 14th, 2016

Abstract

This report will replicate the main results displayed in section 3.2 **Multiple Linear Regression** (chapter 3) of the book *An Introduction to Statistical Learning*.

Introduction

The primary goal of this analysis is to give advice on how to boost sales of the product given the current information on advetising budgets. More specifically, the idea is to determine whether there exists an correalation between advertising and sales, and if so, formulate an accurate model that can be used to predict sales from media budget. For this analysis in particular, a combination of simple linear regression and multiple linear regression.

Data

The Advertising data set comprises of the Sales (in thousands of units) in 200 different markets, along with the advertising budgets (in thousands of dollars) in each market for three different forms of media: TV, Radio, and Newspaper. In this report, the relation between each of them and Sales, and the possible relation between Sales and the three of them combined are observed and studied/analyzed.

We may first look at the table of summary statistics below:

	TV	Radio	Newspaper	Sales
1	Min. : 0.70	Min. : 0.000	Min. : 0.30	Min. : 1.60
2	1st Qu.: 74.38	1st Qu.: 9.975	1st Qu.: 12.75	1st Qu.:10.38
3	Median :149.75	Median :22.900	Median : 25.75	Median :12.90
4	Mean :147.04	Mean :23.264	Mean : 30.55	Mean :14.02
5	3rd Qu.:218.82	3rd Qu.:36.525	3rd Qu.: 45.10	3rd Qu.:17.40
6	Max. :296.40	Max. :49.600	Max. :114.00	Max. :27.00

Table 1: Summary Statistics

Histograms for each variable:

Methodology

Single Linear Regression

We may first consider only using one media fromthe data set, *TV*, *Radio* and *Newspaper* respectively, and study its relationship with the dependent variable *Sales*. The null hypothesis here would be that each of the independent variables would not have an effect on *Sales*, and the alternative hypothesis is that they do have an effect on *Sales*. For this purpose, we use a simple linear model:

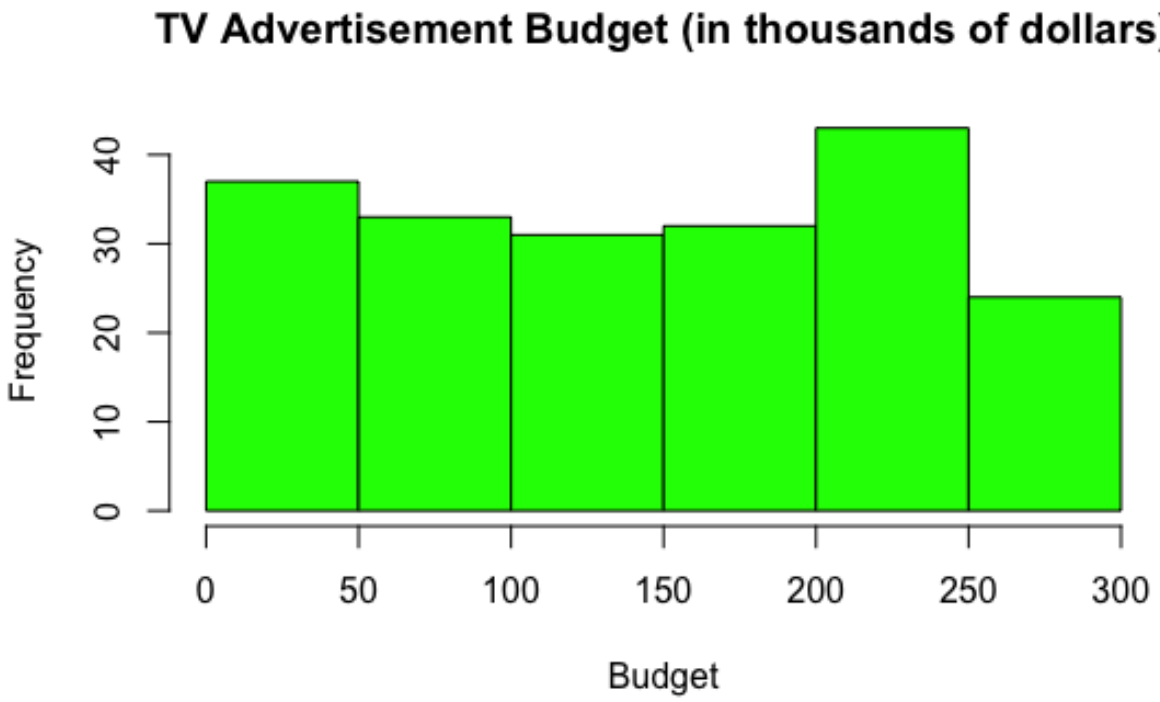

Figure 1: Figure 1: Histogram for TV

$$Sales = \beta_0 + \beta_1(TV|Radio|Newspaper)$$

Multiple Linear Regression

Instead of fitting a separate simple linear regression model for each predictor, a better approach might be to extend the simple linear regression model so that it can directly accommodate multiple predictors. We can do this by giving each predictor a separate slope coefficient in a single model. Given that we have three distinct predictors here, the multiple linear regression model takes the form:

$$Sales = \beta_0 + \beta_1 TV + \beta_2 Radio + \beta_3 Newspaper$$

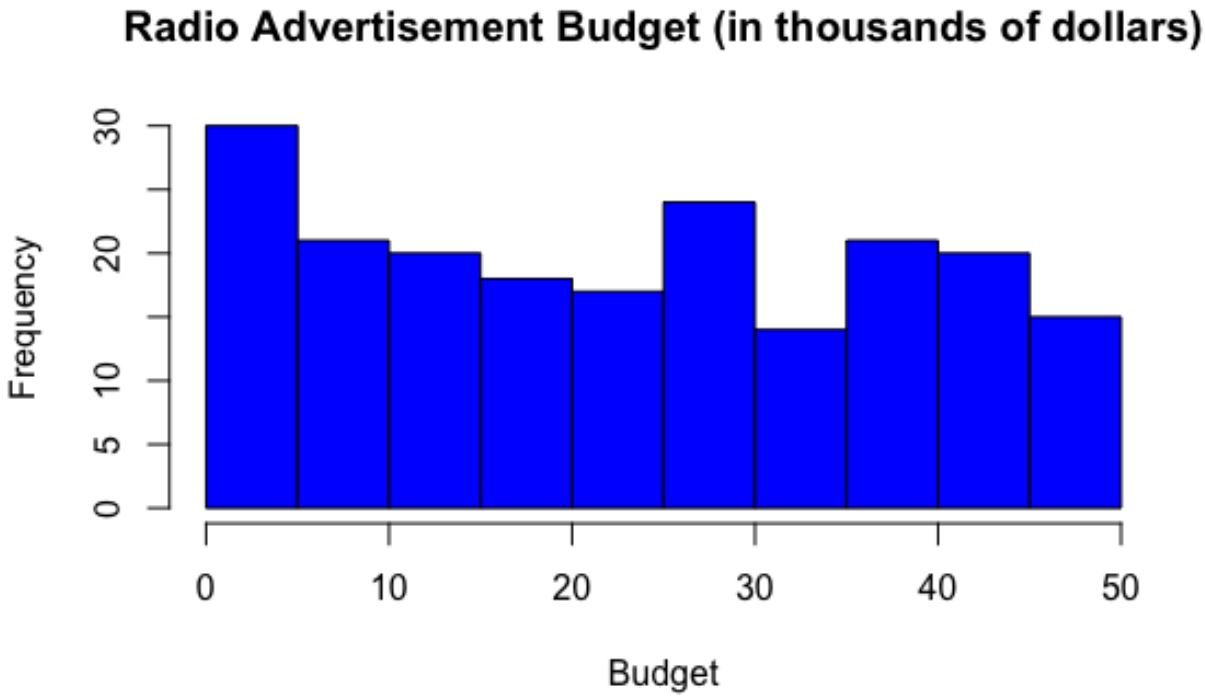

Figure 2: Figure 2: Histogram for Radio

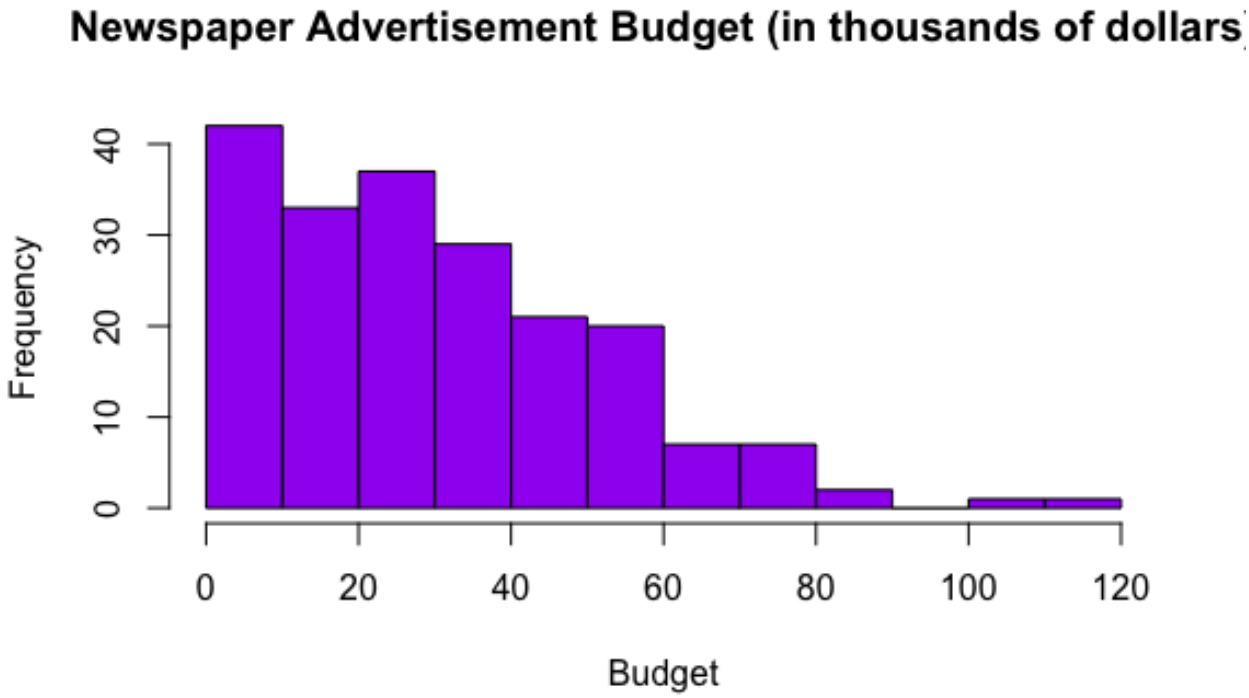

Figure 3: Figure 3: Histogram for Newspaper

Figure 4: Figure 4: Histogram for Sales