Simple Regression Analysis - Report

Lingjie Qiao

10/7/2016

Abstract

This report aims to reproduce the main results displayed in **section 3.1: Simple Linear Regression** of the book ***An Introduction to Statistical Learning*** and perform simple linear regression analysis on the data set `Advertising`.

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 data set `Advertising` is provided by the author of the book. This data set has four variables. It 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.

Methology

In this paper, we mainly consider the relatinoship between Sales and one media from the data set, **TV**. In order to explore this relationship, we use a simple linear model and regress `sales` onto `TV` by fitting the model:

Sales = β_0 +β_1 TV

Mathematically, β_0 represents the intercept and β_1 represents the slope terms in the linear model. With this linear model, we estimate the coefficients by minimizing the least squares criterion, which is minimizing the sum of squared errors.

Results

With the least square estimators, we compute the regression coefficients:

Table 1: Information about Regression Coefficients

Coefficients	Estimate	Std. Error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	<0.0001 (2.2e-16)
TV	0.0475	0.0027	17.67	<0.0001 (2.2e-16)

Conclusions