

Report - Multiple Linear Regression

Cheng Peng

10/14/2016

Abstract

This report reproduced the main results displayed in section 3.2: Multiple 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 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 downloaded from the website. 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

To explore this multiple variable relationship, we use a multiple linear model and regress **sales** onto TV, Radio, Newspaper by fitting the model:

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

Results

Table 1: Simple Linear Regression on TV and Sales

Table 1: Simple Linear Regression - TV and Sales				
	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

Table 2: Simple Linear Regression on Radio and Sales

Table 2: Simple Linear Regression - Radio and Sales				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.31	0.56	16.54	0.00
Radio	0.20	0.02	9.92	0.00

Table 3: Simple Linear Regression on Newspaper and Sales

Table 3: Simple Linear Regression - Newspaper and Sales

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.35	0.62	19.88	0.00
Newspaper	0.05	0.02	3.30	0.00

Table 4: Regression Coefficients

Table 4: Multiple Linear Regression Coefficients

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.94	0.31	9.42	0.00
TV	0.05	0.00	32.81	0.00
Radio	0.19	0.01	21.89	0.00
Newspaper	-0.00	0.01	-0.18	0.86

For the Advertising data, least squares coefficient estimates of the multiple linear regression of number of units sold on radio, TV, and newspaper advertising budgets.

We see that **TV**, **Radio** and **Newspaper** all have very small p-value, which rejects the null hypothesis and shows that they have an impact on the sales data.

I think the linear model fits the data well. However, even though we fit correctly the linear model, the predictions will not be as accurate as we imagine it to be.

Table 5: Correlation Matrix

	Table 5: Correlations Matrix			
	TV	Radio	Newspaper	Sales
TV	1.00	0.05	0.06	0.78
Radio	0.05	1.00	0.35	0.58
Newspaper	0.06	0.35	1.00	0.23
Sales	0.78	0.58	0.23	1.00

Table 6: Regression Quality Indices

Conclusions

I explored the linear relationship between TV, Radio and Newspapers versus Sales with a multiple linear regression model.

Table 6: Regression Quality Indices

	Quantity	Value
1	Residual Standard Error	1.69
2	R-squared	0.89
3	F-statistic	504.94