Multiple Regression Analysis

Young Hoon Kim

10/14/2016

```
## Warning: package 'testthat' was built under R version 3.2.5
```

Abstract

In this report, I am going to conduct *Multiple Regression Analysis* that is explained in the Section 3.2 *Multiple Linear Regression* (Chapter 3) 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 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 Advertising data set 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**.

Methodology

We consider one media from the data set, TV, and study its relationship with Sales. For this purpose, we use a simple linear model:

$$Sales = \beta_0 + \beta_1 TV + \beta_2 Radio + \beta_3 Newspaper \tag{1}$$

To estimate the coefficients β_0, β_1, β_2, and β_3, we first conduct simple linear regression of Sales over each predictor, and we fit a multiple regression model employing all the predictors 'TV, Radio, Newspaper' via the least squares criterion.

Results

First, we compute the simple regression coefficients for each predictors.

Table 1: Simple regression of sales on tv

	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

In order to understand that full regression model, we now fit multiple regression model to the data using all three predictors.

The correlation matrix that displays the correlation among the variables is the following.

Table 2: Simple regression of sales on radio

	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 regression of sales on newspaper

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

In order to understand the fit of the model, we need to understand how the error measures display the performance of the model.

Scatterplots of simple regressions, and the plots of multiple regression are the following:

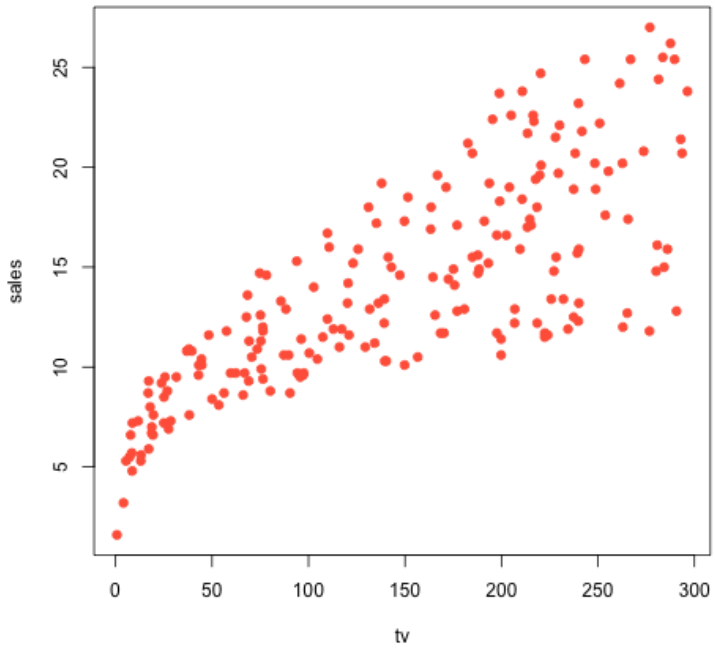

Figure 1: Scatterplot of sales and tv

Conclusions

1. Is at least one of the predictors useful in predicting the response? From the coeffiecients table of multiple regression, we could determine that tv and radio are very significant in predicting the response. We can see this from the significantly low p-values of those predictors.
2. Do all predictors help to explain the response, or is only a subset of the predictors useful? Only tv and radio are the useful predictors to explain the response.
3. How well does the model fit the data? According to the R-squared value, we can determine that the model fits the data quite well.
4. How accurate is the prediction?

Table 4: Multiple regression of sales on tv, radio, and newspaper

	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
news	-0.00	0.01	-0.18	0.86

Table 5: Correlation matrix for TV, radio, newspaper, and sales for the Advertising data.

	TV	Radio	Newspaper	Sales
TV	1.00	0.05	0.06	0.78
Radio		1.00	0.35	0.58
Newspaper			1.00	0.23
Sales				1.00

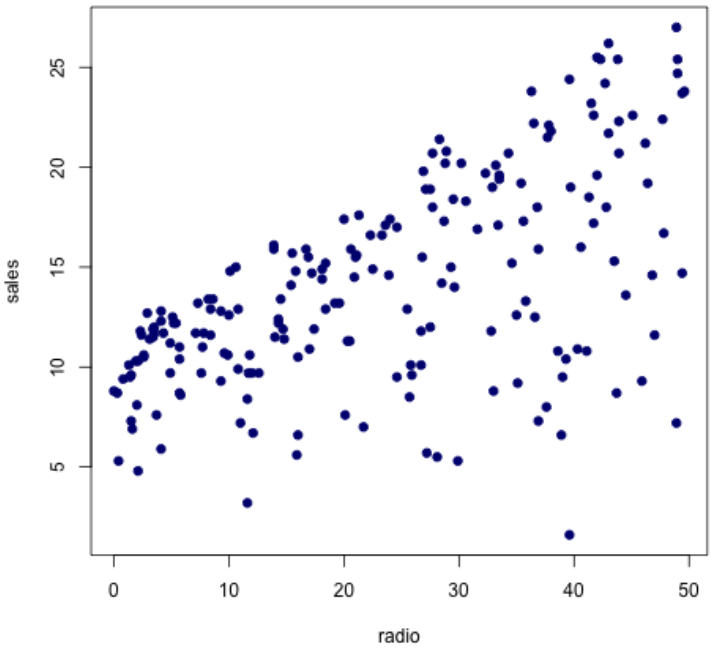

Figure 2: Scatterplot of sales and radio

Table 6: Regression Quantity Indices

	Quantity	Value
1	Residual standard error	1.69
2	R2	0.897
3	F-stat	570

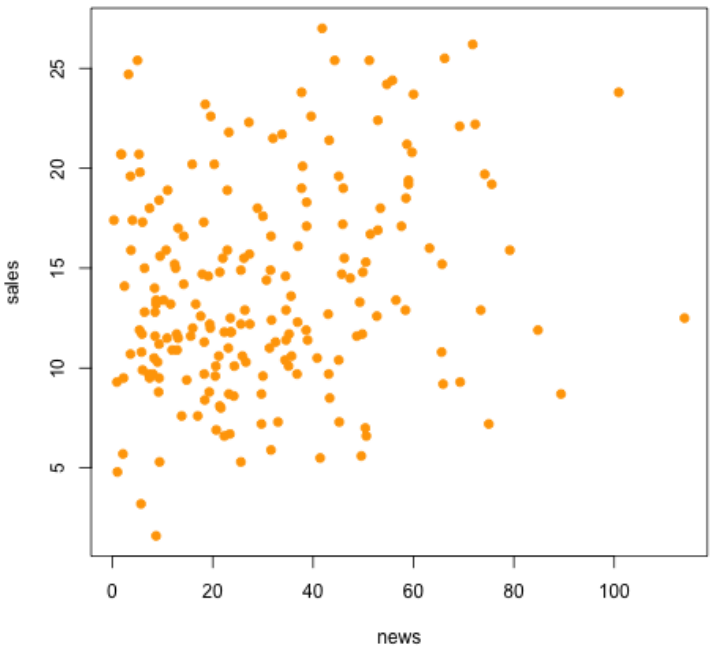

Figure 3: Scatterplot of sales and newspaper

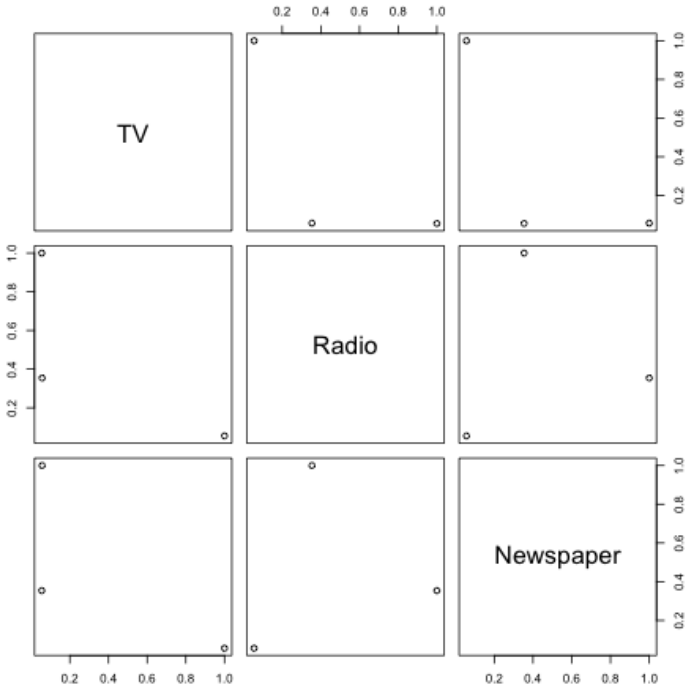

Figure 4: Scatterplot of the correlation matrix

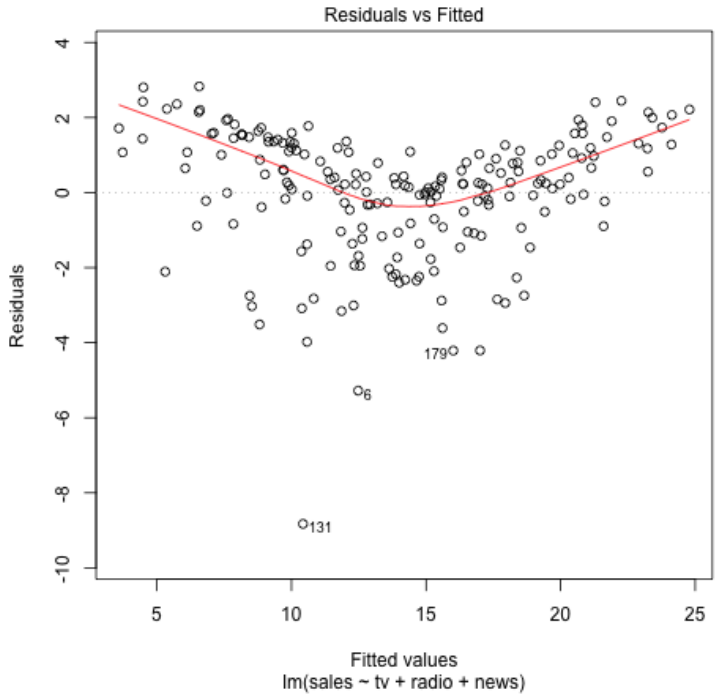

Figure 5: residual plot of multiple regression

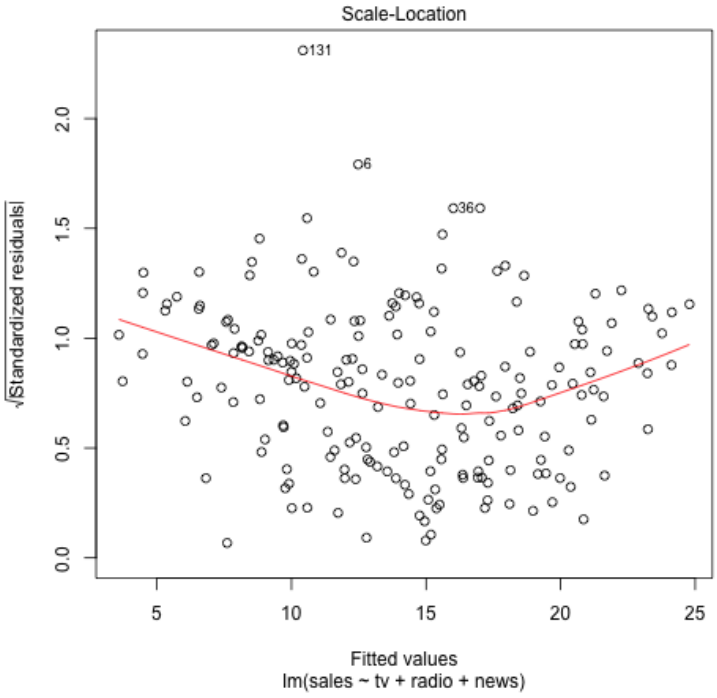

Figure 6: scale-location plot

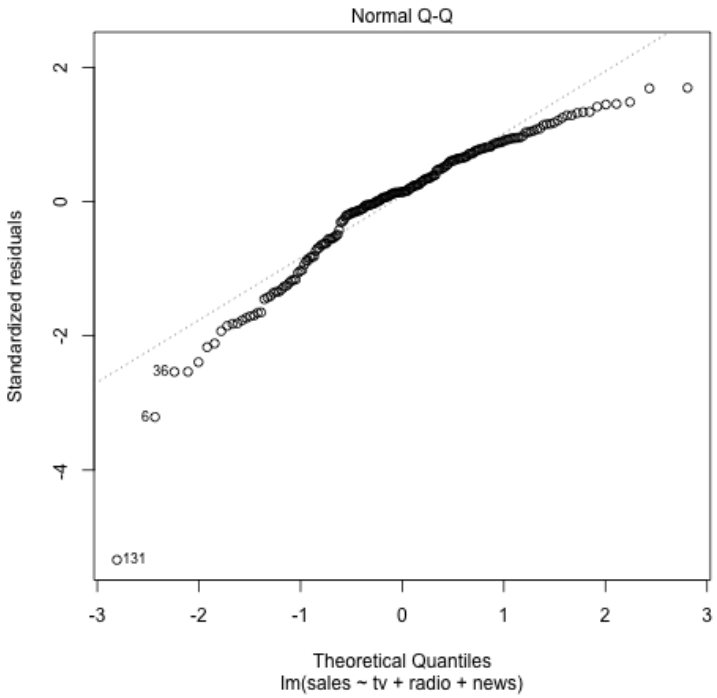

Figure 7: Normal Q-Q plot of multiple regression