Multiple Linear Regression

Youngshin Kim

December 8, 2016

Abstract

The mission for this project is to reproduce results from Section 3.2 (pages 71 to 82), from the book "An Introduction to Statistical Learning" (by James et al).

Introduction

The main analysis of this project is multiple linear regression. Multiple linear regression is different from simple linear regression in that it involves more than one predictor, as the name suggests.
We will use the Advertising data that contains TV, Radio, Newspaper budget data and these will be the predictors.

Data

The Advertising data consists of values on Sales (in thousands of unit) of some product in 200 different markets and values on advertising budget (in thousands of dollars) for three different kind of marketing strategies: TV, Radio, Newspaper. For the purpose of this project, we are just interested in the advertising budget for 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

Table 1: Simple regression of sales on TV

	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 2: Simple regression of sales on radio

	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 3: Simple regression of sales on newspaper

Methodology

We use multiple regression model to model the relationship between Sales and the three predictors. The response variable here is Sales and the predictor variables are TV, Radio, and Newspaper budgets.

The regression equation is Sales = Intercept + Coefficient * TV + Coefficient * Radio + Coefficient * Newspaper. With this equation in mind, we use the lm function in R to find the relevant information for our regression. The syntax for this code is lm(Sales ~ TV + Radio + Newspaper, data = advertising) where object 'advertising' is the main data from Advertising.csv. The basic idea behind this function is that we want to find the intercept and the regression coefficients that minize the residual sum of squares.

Results

	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

Table 4: Regression table

	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 5: Correlation Matrix

	Value
RSE	1.69
R^2	0.90
F-statistic	570.27

Table 6: Various Statistic

Is at least one of the predictors useful in predicting the response?

We can be pretty confident that at least one of the predictors is useful in predicting the response, because in Table 4, p-value for TV and Radio are zero, which mean they are statistically significant.

Also, F-statistics is above 1 and in fact it is extremely high. This also shows that there is a relationship between predictors and response.

Do all predictors help to explain the response, or is only a subset of the predictors useful?

In Table 4, p-value for Newspaper is extremely high so Newspaper budget is probably not an important indicator of Sales growth. In fact, the coefficient for Newspaper is negative, so an increase in Newspaper budget probably lowers Sales growth.

How well does the model fit the data?

The model seems to fit the data pretty well, because RSE is low and R^2 is high. It also helps to notice that TV and Radio are correlated. The model that includes both variables has a lower RSE than the model that

only contains TV.

How accurate is the prediction?

If we were to do a prediction using this model, the result should be pretty satisfactory because the regression summary tells us that our coefficients for intercept, TV, and Radio are statistically signifcant. The predition would be better if we use the model without including Newspaper.

Conclusions

All the predictors have effects on our response variable Sales. This was made clear with simple linear regression. Working together, however, Newspaper proved to be harmful for Sales. TV and Radio, on the other hand, proved to have a synergy effect in Sales growth.