Multiple Linear Regression of Advertising Data

Alexander Lee

October 14, 2016

Abstract

This report seeks to replicate the research found in *Simple Linear Regression*, chapter 3 of the book **An Introduction to Statistical Learning** by Gareth James, et al., on a set of advertising data. I run a multiple linear regression on the data and create various functions to calculate several statistics, including the residual square error and the F-statistic.

Introduction

The goal of this research is to use existing data to formulate a marketing plan that will result in higher product sales. To that end, we will run a multiple linear regression to model the increase of sales against the amount of money spent on marketing the product through various media - TV, radio, and newspapers. We will analyze this to determine how advertising affects sales, and how the company should budget for advertising in order to increase sales of the product.

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`.

```
##   X    TV Radio Newspaper Sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

Methodology

We perform a simple linear regression on `Sales` for each of the three other factors, under the model:

```
Sales = a + b * (FACTOR)
```

Where `(FACTOR)` is either `TV`, `Radio`, or `Newspaper`. We predict there to be a linear relationship between `Sales` and the amount of budget placed in advertising for each of these three media. The summaries of these can be found in Tables 1-3.

Next, we look at the relationship between `Sales` and the budgets of the various forms of media using the model:

```
`Sales = a + b * TV + c * Radio + d * Newspaper`
```

That is, we predict there to be a relationship between the amount of money spent on TV, Radio, and Newspaper advertising and the number of sales of units. We approximate the values of a, b, c, and d using a multiple linear regression under the least squares criterion.

Results

We computed the correlation coefficients using the lm() function, with TV as a function of Sales.

The correlation between the factors looks like this:

The estimates of these coefficients, a, b, c, and d, are 2.9389, 0.0458, 0.1885, and −0.001, respectively. For every $1,000 increase in spending on TV advertising, sales are projected to increase by 46 units; this means that units have to cost at least $ 21.85 in order to be profitable. For every $1,000 increase in spending on Radio advertising, sales are projected to increase by approximately 189 units; this means that units have to cost at least $ 5.3 in order to be profitable. For every $1,000 increase in spending on Newspaper advertising, sales are projected to increase by −1 units.

On average, sales data will deviate from the true regression model by 1.6855 units. An R^2 of 0.8972 means 89.72% of the variability is explained by the model. An F-statistic of 570.2707 means that the coefficients found by our model are very likely to be close to the true regression values.

Conclusions

Individually, each of the factors TV, Radio, and Newspaper have a significant effect on the number of Sales. Together, only TV and Radio have significant effects; an increase in Newspaper budget actually decreases the number of Sales.

There is a positive correlation between the budget for TV advertising and Sales; however, this relationship is very minimal. For every $1,000 spent on advertising only another 48 units are sold, so each unit would have to cost at least $20.84 in order to break even or profit off the units sold.

Figures

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

	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: Summary of Simple Linear Regresssion on TV

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

	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: Summary of Simple Linear Regression on Radio

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

	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: Summary of Simple Linear Regression on 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
Newspaper	-0.00	0.01	-0.18	0.86

Table 4: Summary of Coefficients for All Independent Factors

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Fri Oct 14 16:42:34 2016

	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 of the Multiple Linear Regression

	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 6: Summary of Simple Linear Regresssion on TV