Stat159 HW03 Multiple Regression Analysis

Siyu Chen

10/14/2016

Abstract

The purpose of this assignment is to extend the scope of the previous HW. Not only we will look into relationships among variables, but also we will write funtions to compute some important statistics. In addition, we need to write unit tests to test our written functions.

Introduction

The overall goal is to examine the effect of TV, Newspaper and Radio advertising budgets on Sales by calculations and plotting graphs. If we do find a relationship, then we want to build a good linear model that can be used for Sales prediction based on TV, Newspaper or Radio advertising budgets.

Data

The advertising dataset consists of 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, Newspaper and Radio).

Methodology

We consider Sales vs TV, Newspaper and Radio advertising budgets in our dataset and try to fit them in a multiple linear regression model:

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

In order to estimate four coefficients β_0, β_1, β_2 and β_3 we fit the linear regression model via the least square criterion.

Results

Before analysis about multiple linear regression, it's good to examine the relationships between Sales and each of TV, Newspaper and Radio advertising budgets.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.0326	0.4578	15.36	0.0000
advertising$TV	0.0475	0.0027	17.67	0.0000

Table 1: Regressing Sales on TV advertising budgets

As we can see, the p-value of predictor is pretty small, which is 0. So we can infer that there is an association between the predictor(TV advertising budgets) and the response(Sales).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.3514	0.6214	19.88	0.0000
advertising$Newspaper	0.0547	0.0166	3.30	0.0011

Table 2: Regressing Sales on Newspaper advertising budgets

As we can see, the p-value of predictor is pretty small, which is 0.0011. So we can infer that there is an association between the predictor(Newspaper advertising budgets) and the response(Sales).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.3116	0.5629	16.54	0.0000
advertising$Radio	0.2025	0.0204	9.92	0.0000

Table 3: Regressing Sales on Radio advertising budgets

As we can see, the p-value of predictor is pretty small, which is 0. So we can infer that there is an association between the predictor(Radio advertising budgets) and the response(Sales). ## Is at least one of the predictors useful in predicting the response?