Stat 159 Hw3 Report

Ellen Hwang

10/7/2016

Abstract

In this report, I am performing a simple linear regression analysis. I am using the Advertising dataset, taken from the webpage for the text, *An Introduction to Statistical Learning*, to perform this simple linear regression. This report will include a description of the data, methodology, and results for the linear regression.

Introduction

One of the basic models data scientists should understand is simple linear regression. Regression, in its most simple terms, is a statistical process that estimates the relationship between a dependent variable and one or more independent variables. In this report, we will only be examing the relationship between one dependent variable and one independent variable to understand regression. This report will specifically examine the relationship between TV advertising budget and number of sales using linear regression.

Data

We will be working with the Advertising dataset. This dataset holds information for 200 different markes for 4 different variables: *Sales*, *TV*, *Newspaper*, and *Radio*. *Sales* represents the amount of units sold (in thousands). *TV*, *Newspaper*, and *Radio* each represent the advertising budget spent on those platforms.

Methodology

For running a regression, we using the `lm()` function. More specially, we regress *Sales* on *TV* using this code: `lm(Sales ~ TV, data = advertising)`. The immediate output of this code is the estimated coefficient and the constant value. Second, I use the `summary`function on the regression object to see further coefficient information, residuals, and other statistics.

Results

Based on the regression of Sales on TV, we found that TV is a statistically significant variable because of a p-value far below zero. The R squared value is at .6099 meaning about 61% of the variance from the population regression line is explained by TV.

Regression Object Output

```
load('../data/regression.RData')
simp_reg
```

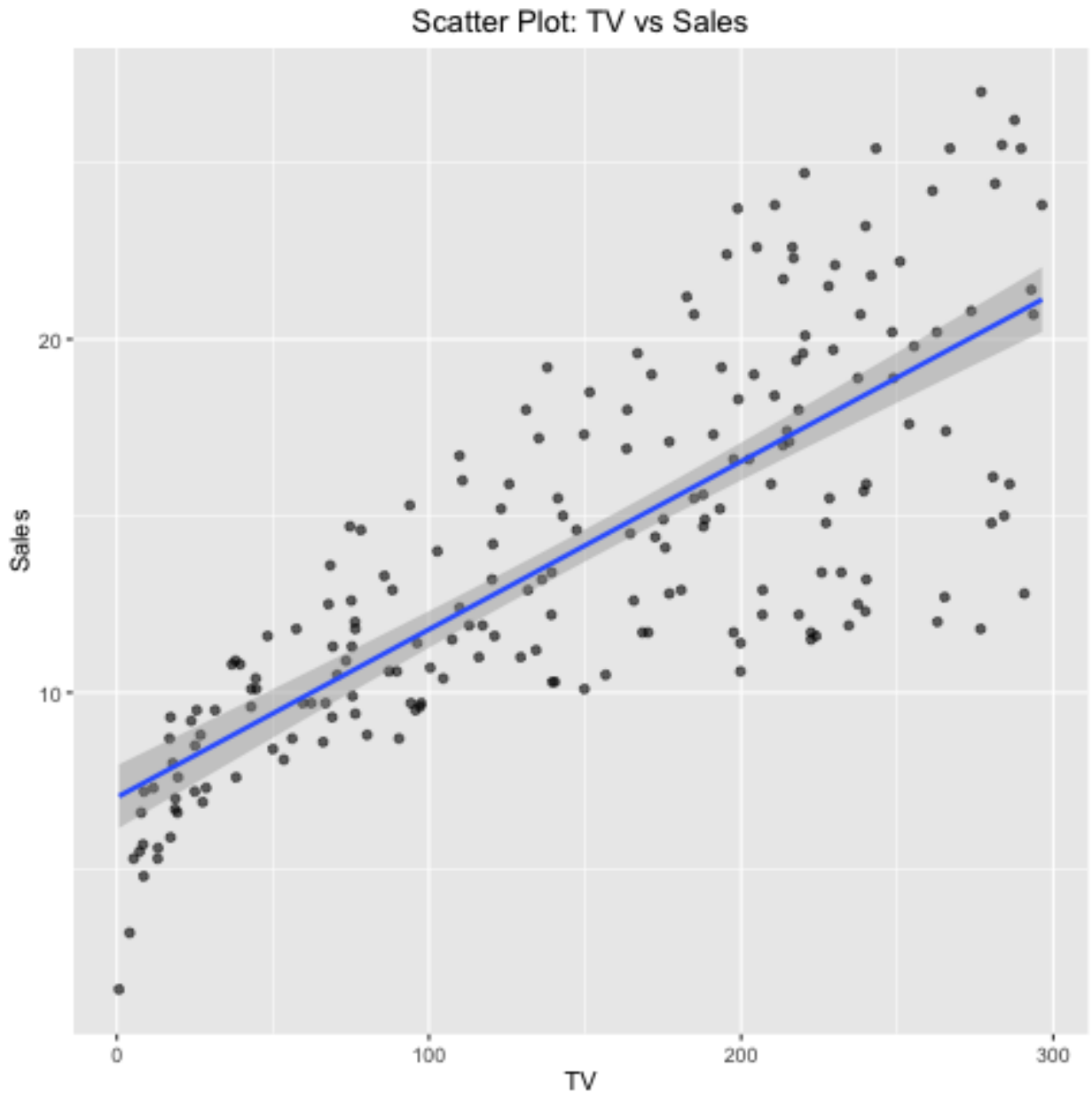
```
##
## Call:
## lm(formula = Sales ~ TV, data = advertising)
##
## Coefficients:
## (Intercept)           TV
##     7.03259      0.04754
```

Summary on Regression Object

```
sum_simp_reg
```

```
##
## Call:
## lm(formula = Sales ~ TV, data = advertising)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## TV          0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

Scatter Plot with Regression of Sales on TV



Conclusions

Ultimately, we can see