

Linear Regression

Krishna Nadiminti

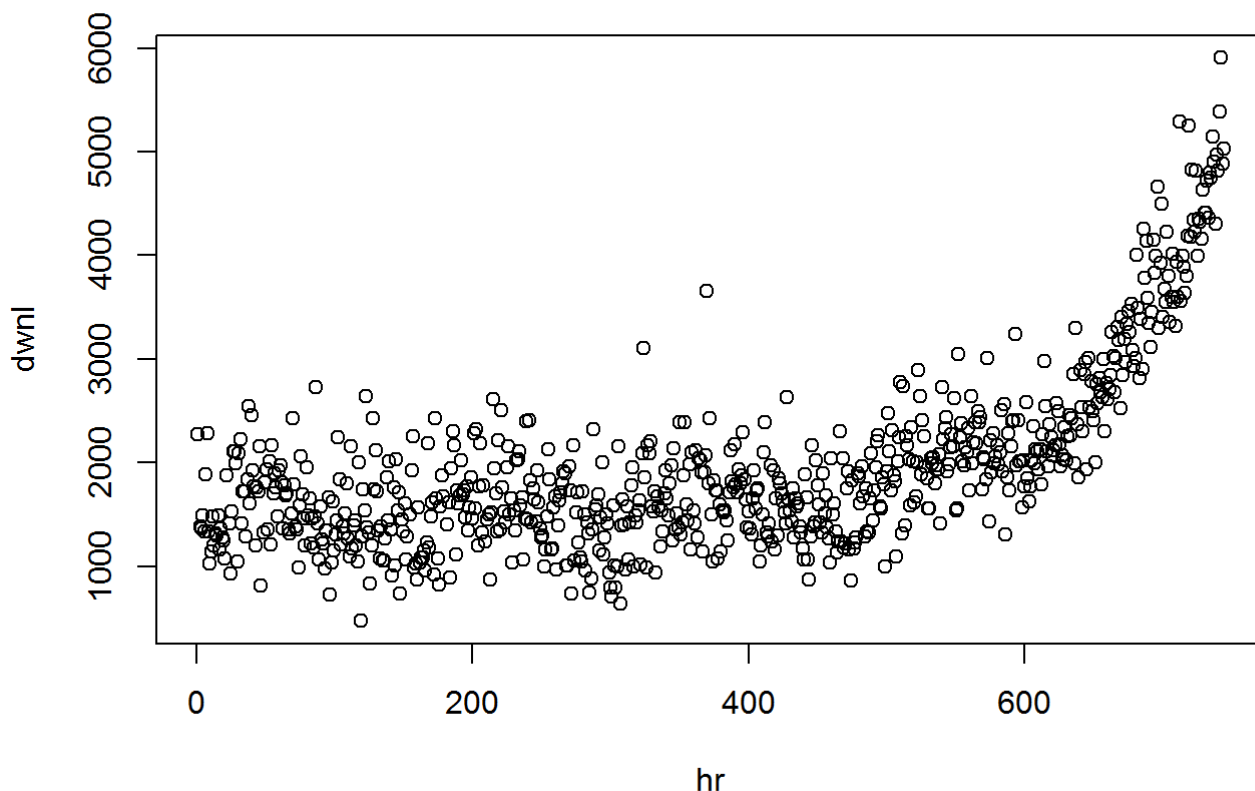
January 19, 2016

Reading and Cleaning the Data

```
docML1 <- read.csv("docML1.csv")
docML1<- subset(docML1,! (is.na(docML1$dwnlds)))
hr<-docML1$hour
dwnl<-docML1$dwnlds
### Replacing Nan with 0
#dwnl<-replace(dwnl, is.na(dwnl), 0)
```

Plotting the Elements as Scatter Plot

```
plot(dwnl~hr)
```



Computing Least Square Regression Equation

$$Y = mX + C$$

$$m = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}_i)}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

```
Sxx=0;
Sxy=0;
for(i in 1:length(hr)) {

  Sxx=Sxx+((hr[i]-mean(hr))*(dwnl[i]-mean(dwnl)))
  Sxy= Sxy+((hr[i]-mean(hr))^2)

}
```

Slope for the Linear regression is

```
(slope<- (Sxx/Sxy))
```

```
## [1] 2.619285
```

Intercept for the Linear Regression Equation is

```
(intercept<- mean(dwnl)-(slope*mean(hr)))
```

```
## [1] 983.2275
```

Predicted Downloads on noon Fifth day of next month

No of Hours on of Fifth day is 852

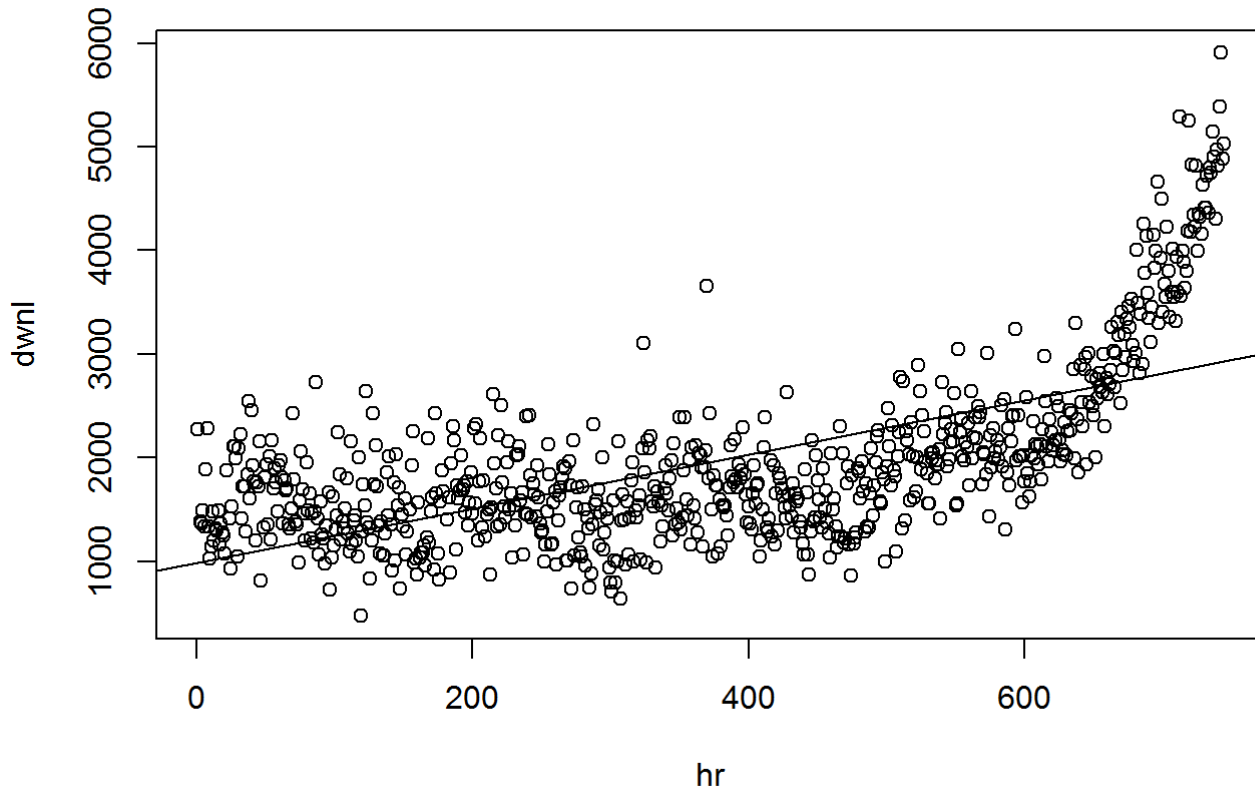
```
(Response<- slope*852+intercept)
```

```
## [1] 3214.858
```

Total number of Downloads at 852 hours is 3236.685

Visualizing the Regression Equation on the scatterplot

```
plot(dwnl~hr)
abline(intercept,slope)
```



Upon Plotting the Scatterplot we can consider that the popularity of the book has been steadily increasing over the period of time and the variables are nearly linear fashioned.

Comparing the results with Existing LinearModel

```
modellinear<- lm(dwnlds~hour,data = docML1)
summary(modellinear)
```

```
##
## Call:
## lm(formula = dwnlds ~ hour, data = docML1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1364.8  -461.4   -92.3   335.3  2979.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  983.227     49.268   19.96  <2e-16 ***
## hour         2.619       0.114   22.97  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 662.2 on 734 degrees of freedom
## Multiple R-squared:  0.4182, Adjusted R-squared:  0.4174
## F-statistic: 527.7 on 1 and 734 DF,  p-value: < 2.2e-16
```

Seems both the Functions have yielded the similar Results