# Chapter 12: Use Case Notebook for Instructors

Ram Gopal, Dan Philps, and Tillman Weyde

Summer 20202

## Contents



## Load packages

```
library(bbmle)
```

## Use Case: Profit Forecasting using Linear Regression

Forecasting profits, revenues, debt levels, and a whole host of other metrics and analytics is crucial in business. Linear regression provides the most widely used tool kit for forecasting.The dataset gives us a snapshot of annual data for a range of startup companies: research and development spending (R&D Spend), Administration expenses, Marketing Spend, tells us the State the startup is based in, and the Profit generated:

```
df = read.csv("../../data/50_Startups.csv")
startups = sample(1:nrow(df),3)
head(df)
```

```
##   R.D.Spend Administration Marketing.Spend      State Profit
## 1    165349         136898          471784   New York 192262
## 2    162598         151378          443899 California 191792
## 3    153442         101146          407935    Florida 191050
## 4    144372         118672          383200   New York 182902
## 5    142107          91392          366168    Florida 166188
## 6    131877          99815          362861   New York 156991
```

## New Startups

We will randomly select three new startups we want to assess from the dataframe. We will build the regression model based on 47 startups (we will term it training data) and then test how well the model predicts for the three new startups (we will term it the test data):

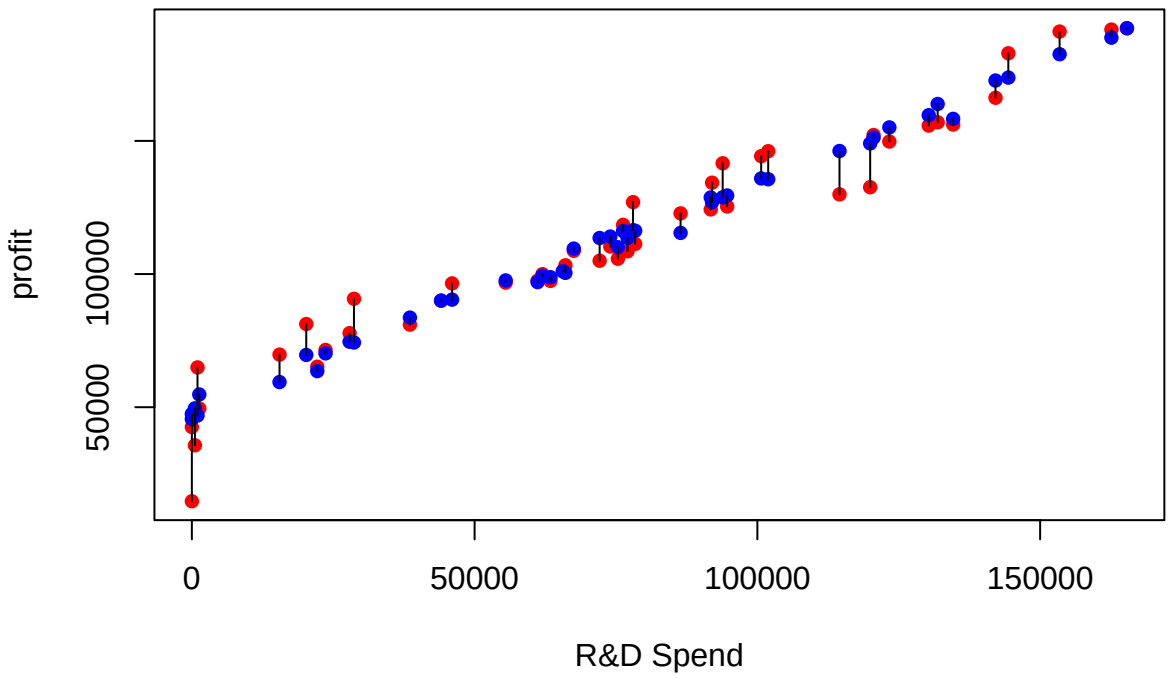
```
df_test = df[startups,]
df_train =df[-startups,]
```

## Building the Model

```
reg1 = lm(Profit ~ .-State, data=df_train)
summary(reg1)
```

```
##
## Call:
## lm(formula = Profit ~ . - State, data = df_train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32765  -4413    -76   6699  18034
##
## Coefficients:
##                   Estimate Std. Error t value             Pr(>|t|)
## (Intercept)     51352.6134  6832.4708    7.52         0.0000000023 ***
## R.D.Spend           0.8179     0.0479   17.09 < 0.0000000000000002 ***
## Administration     -0.0429     0.0541   -0.79                 0.43
## Marketing.Spend     0.0246     0.0171    1.44                 0.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9410 on 43 degrees of freedom
## Multiple R-squared:  0.951,  Adjusted R-squared:  0.948
## F-statistic:  280 on 3 and 43 DF,  p-value: <0.0000000000000002
```

```
plot(df_train$R.D.Spend,df_train$Profit,col="red",pch=16,xlab="R&D Spend",ylab="profit")
points(df_train$R.D.Spend,reg1$fitted.values,col="blue",pch=16)
segments(df_train$R.D.Spend,df_train$Profit,df_train$R.D.Spend,reg1$fitted.values)
```

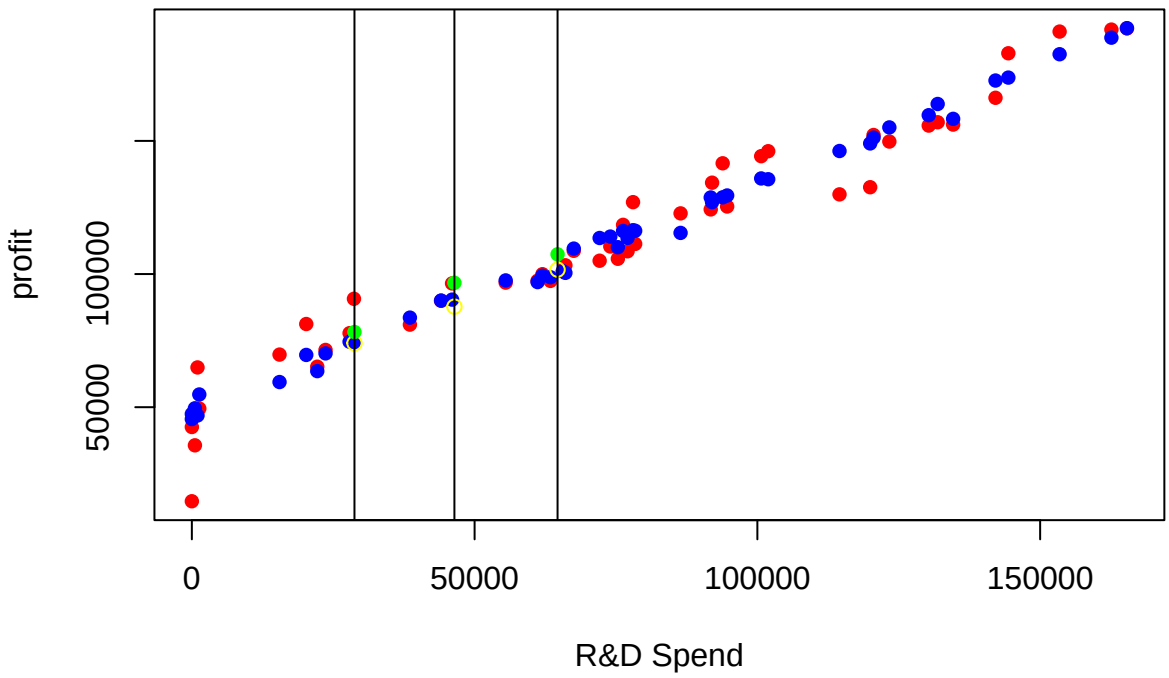



## Forecast Startup Profits

```
df_test$prediction = predict.lm(reg1,df_test)
df_test
```

```
##    R.D.Spend Administration Marketing.Spend      State Profit prediction
## 35     46426          157694          210798 California  96713      87746
## 26     64665          139553          137963 California 107404     101648
## 41     28754          118546          172796 California  78240      74037
```

```
plot(df_train$R.D.Spend,df_train$Profit,col="red",pch=16,xlab="R&D Spend",ylab="profit")
points(df_train$R.D.Spend,reg1$fitted.values,col="blue",pch=16)
#segments(df_train$R.D.Spend,df_train$Profit,df_train$R.D.Spend,reg1$fitted.values)
points(df_test$R.D.Spend,df_test$prediction,col="yellow",pch=)
points(df_test$R.D.Spend,df_test$Profit,col="green",pch=16)
#segments(df_test$R.D.Spend,df_test$Profit,df_test$R.D.Spend,df_test$prediction,lwd = 5)
abline(v = df_test$R.D.Spend)
```




## Dealing with Categorical Data

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
dummy = dummyVars(" ~ .", data=df)
final_df = data.frame(predict(dummy, newdata=df))
head(final_df)
```

```
##   R.D.Spend Administration Marketing.Spend StateCalifornia StateFlorida
## 1    165349         136898          471784                0            0
## 2    162598         151378          443899                1            0
## 3    153442         101146          407935                0            1
## 4    144372         118672          383200                0            0
## 5    142107          91392          366168                0            1
## 6    131877          99815          362861                0            0
##   StateNew.York Profit
## 1             1 192262
## 2             0 191792
## 3             0 191050
## 4             1 182902
## 5             0 166188
## 6             1 156991
```