Regression

Logistic Regression

Assignment 3

BP: Suppose we are interested in the factors that influence whether a political candidate wins an election.

The outcome (response) variable is binary (0/1); win or lose.

The predictor variables of interest are the amount of money spent on the campaign,

the amount of time spent campaigning negatively and whether or not the candidate is an incumbent.

1. Now we have to create a Logistic Regression model for this , by the formula :

logreg <- glm(Result ~ factor (amount spent)+ factor (popularity rank) )

on using summary ( reg ) we get,

Null deviance: 2.4000e+00 on 9 degrees of freedom

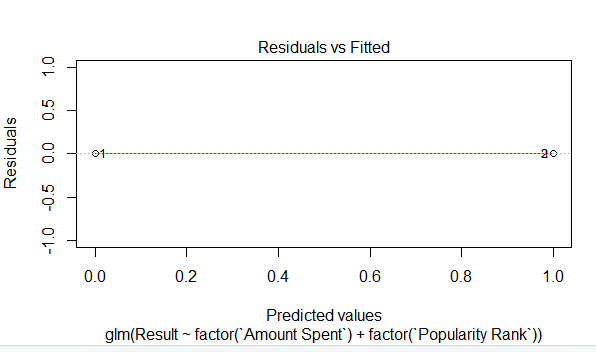
Residual deviance: 9.3711e-31 on 0 degrees of freedom

(1 observation deleted due to missingness)

AIC: -664.07

The model with least AIC is the best model .

EDA of the data is as follows:



1. Now we have to try and build model by finding the probability we can get an idea of the model i.e, the probability of candidate whether he wins or not  
   prob = predict(logreg, type=c("response"),election\_data)

On entering prob we get

prob

1 2 3 4 5 6

NA 5.726371e-16 1.000000e+00 1.000000e+00 4.037458e-16 1.000000e+00

7 8 9 10 11

-2.192009e-17 1.000000e+00 1.000000e+00 1.000000e+00 2.755483e-16

Now from the values we can assume that for the 1st value the probability of candidate winning is 0% , so from the data we can see that it may be correct.

To get the Confusion matrix ,

confusion <- table(prob>0.5,affairs$naffairs)

the output is as follows:

0 1

FALSE 4 0

TRUE 0 6

seems like there is 100% true prediction which may seem impossible in real time scenarios . So the ROC curve is :

