STATS 201 Lab Class 4

LiYuzhuo-2019210146

# Code and output

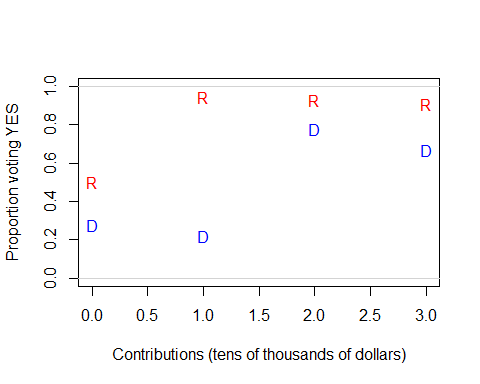
## Loading in the data.  
require(s20x)

## Loading required package: s20x

vote.df = read.table("vote-grouped.txt", header = TRUE)  
## Creating the proportion of senators who voted YES.  
vote.df$p = with(vote.df, yes/(yes+no))   
vote.df

## party contributions yes no p  
## 1 D 0 8 21 0.2758621  
## 2 D 1 2 7 0.2222222  
## 3 D 2 7 2 0.7777778  
## 4 D 3 2 1 0.6666667  
## 5 R 0 3 3 0.5000000  
## 6 R 1 17 1 0.9444444  
## 7 R 2 13 1 0.9285714  
## 8 R 3 10 1 0.9090909

## Add variable p to the dataframe here.  
## Plot the data.  
## Making a plotting area.  
plot(p ~ contributions, type = "n", data = vote.df, ylim = c(0, 1),  
 xlab = "Contributions (tens of thousands of dollars)",  
 ylab = "Proportion voting YES")  
## Making some horizontal lines at the limits of the proportion.  
abline(h = c(0, 1), col = "lightgrey")  
## Plotting the data with each point representing the senator's party.  
text(p ~ contributions, labels = party, col = ifelse(party == "D", "blue", "red"),   
 data = vote.df)



## Analyse the data.  
vote.fit = glm(p ~ party \* contributions, weight = yes + no, family = binomial, data = vote.df)  
anova(vote.fit, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: p  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 7 43.862   
## party 1 28.0207 6 15.841 1.2e-07 \*\*\*  
## contributions 1 9.2777 5 6.563 0.00232 \*\*   
## party:contributions 1 0.0042 4 6.559 0.94846   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

vote.fit2 = glm(p ~ party + contributions, weight = yes + no, family = binomial, data = vote.df)  
anova(vote.fit2, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: p  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 7 43.862   
## party 1 28.0207 6 15.841 1.2e-07 \*\*\*  
## contributions 1 9.2777 5 6.563 0.00232 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(vote.fit2)

##   
## Call:  
## glm(formula = p ~ party + contributions, family = binomial, data = vote.df,   
## weights = yes + no)  
##   
## Deviance Residuals:   
## 1 2 3 4 5 6 7 8   
## 0.33333 -1.27839 0.99946 -0.47821 -1.07372 1.31917 0.06733 -0.83229   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.1051 0.3808 -2.902 0.003704 \*\*   
## partyR 1.9959 0.5545 3.599 0.000319 \*\*\*  
## contributions 0.8025 0.2804 2.862 0.004207 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 43.8617 on 7 degrees of freedom  
## Residual deviance: 6.5633 on 5 degrees of freedom  
## AIC: 30.641  
##   
## Number of Fisher Scoring iterations: 4

1- pchisq(6.56, 5)

## [1] 0.2554729

plot(vote.fit2, which = 1, lty = 2)  
  
exp(confint(vote.fit2))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.1497489 0.6759938  
## partyR 2.5856936 23.3669682  
## contributions 1.3199865 4.0066384

(exp(confint(vote.fit2))-1)\*100

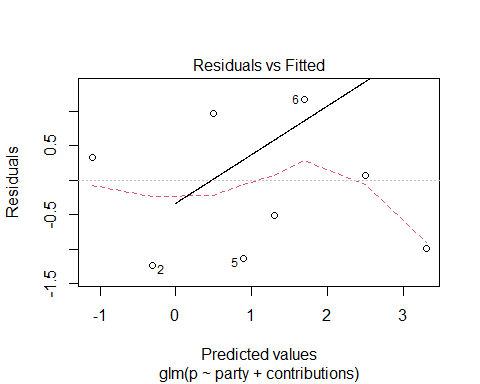
## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -85.02511 -32.40062  
## partyR 158.56936 2236.69682  
## contributions 31.99865 300.66384

vote.fit3 = glm(p ~ contributions, family = binomial, data = vote.df)

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

vote.pred = predict(vote.fit3, newdata = data.frame(contributions = seq(0, 3, length.out = 1000)))  
lines(seq(0, 3, length.out = 1000), vote.pred)



#vote.pred  
## INSERT CODE HERE.

# Methods and Assumption Checks

As the response variable is the proportion of senators who voted YES on the bill, and two explanatory variable party(factor) and contributions(numeric). We have therefore fitted a generalized linear model with a binomial response distribution.

The plot of the proportion of senators who voted YES shows an obvious association with party and contributions, and without interaction between party and contributions. We fitted a logistic regression model to these data.

And the check of residual deviance had p-value = 0.26, therefore we can trust the results from this binomial model.

Our model is:

Party is 1 when party is R, otherwise 0.

# Executive Summary

We are interested in determining if the amount of contributions from the automotive industry is related to the probability a senator votes YES on the bill. We have strong evidence that the amount of contributions from the automotive industry is related to the probability a senator votes YES on the bill.

We are also interested to determine if senators from one party are more likely to vote YES on the bill than the other. We also have strong evidence to confirm the relation between the party of the senator and the willing to vote YES on the bill .

But the effect of contributions doesn’t depend on the political party, as our first fit including the interaction item of party and contributions, the summary shows p-value is 0.95.

We estimated the mean of p will increase between 31.1% and 300.7% by the original with $10000 contributions increased.

Therefore, we can conclude that the amount of contributions from the automotive industry is related to the probability a senator votes YES on the bill and senators from Republican are more likely to vote YES than Democrat, and there is no evidence shows that the effect of contributions depend on the political party.