# STATS 201 Experimental Class 4

Runze liao 222020321102007

Code and output

```
## Loading in the data.
vote.df = read.table("vote-grouped.txt", header = TRUE)
## Creating the proportion of senators who voted YES.
vote.df$p = vote.df$yes/(vote.df$yes+vote.df$no)## Add variable p to th
e 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", "blu
e", "red"),
     data = vote.df)
```
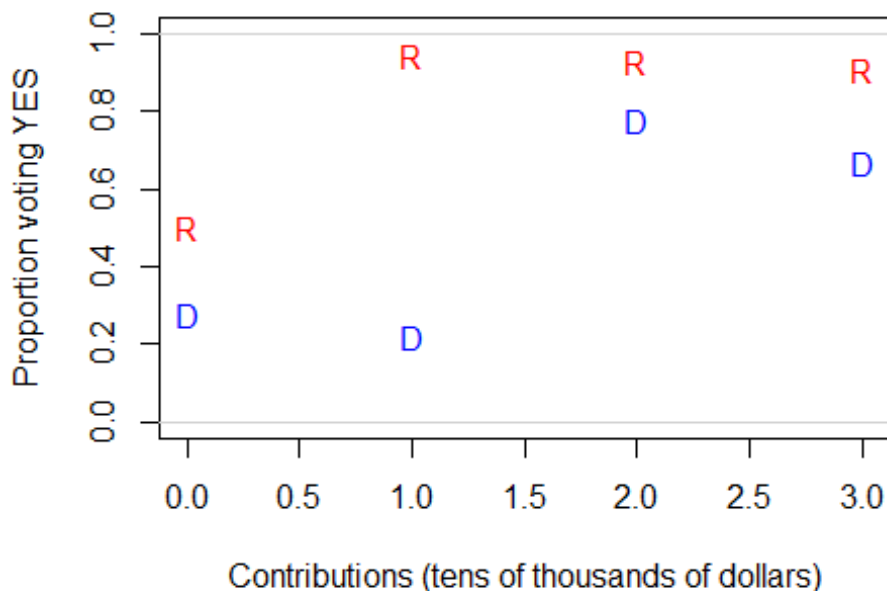
From this plot, we can see that R of the proportion of yes is larger than D. And for each party, there's no clear linear relation between the contributions and Proportion for voting yes. The proportions retained seem to follow an "S" shape.

We will fit a logistic regression model to fit our data.

```
vote.fit = glm(cbind(yes,no)~party*contributions, family = binomial, data = vote.df)

plot(vote.fit,which =1)
```
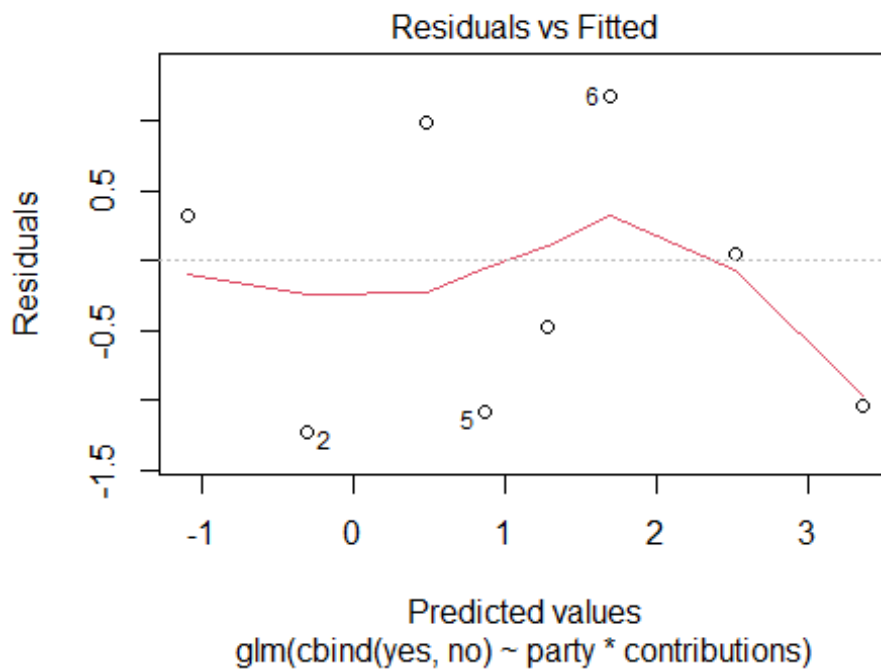

The residual plot is strange, seems that not satisfy the EOV check, but we will still keep it by having glm.

```
summary(vote.fit)

##
## Call:
## glm(formula = cbind(yes, no) ~ party * contributions, family = binomial,
ial,
##      data = vote.df)
##
## Deviance Residuals:
##          1          2          3          4          5          6          7
          8
##    0.31227   -1.27471    1.01919   -0.45896   -1.03712    1.32283    0.04196
   -0.87183
##
```

```
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.09613    0.40441  -2.710  0.00672 **
## partyR               1.95564    0.83203   2.350  0.01875 *
## contributions        0.79097    0.33152   2.386  0.01704 *
## partyR:contributions 0.04011    0.62157   0.065  0.94855
## ---
## 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:  6.5591  on 4  degrees of freedom
## AIC: 32.637
##
## Number of Fisher Scoring iterations: 5
```

We can see the interaction part of coefficient is out of the 95% confidence interval, so we will fit a model without interaction.

```
vote.fit2 = glm(cbind(yes,no)~party+contributions, family = binomial, d
ata = vote.df)
plot(vote.fit,which =1)
```


Residuals vs Fitted
Predicted values
glm(cbind(yes, no) ~ party * contributions)
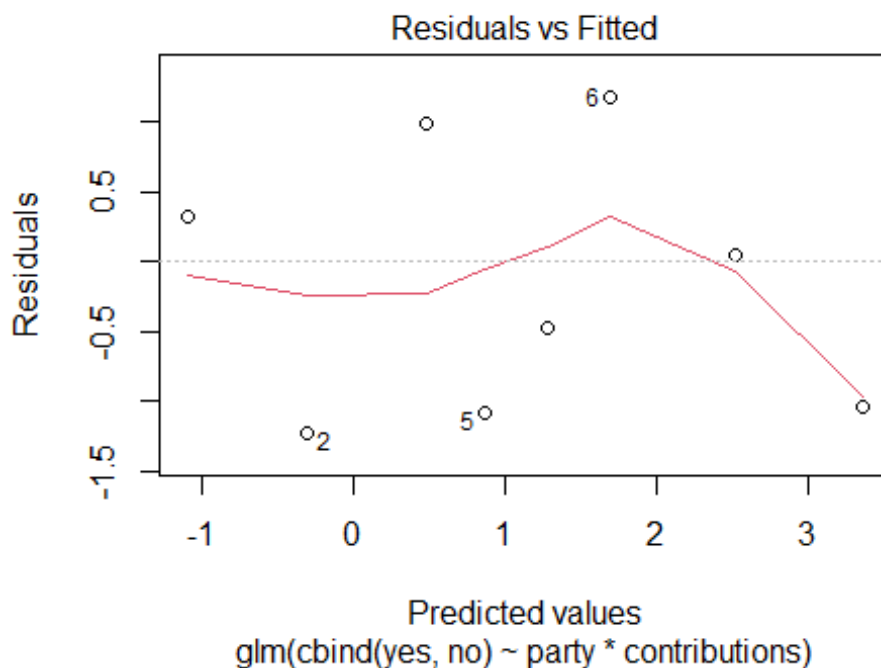
```
summary(vote.fit2)
```

```
## 
## Call:
## glm(formula = cbind(yes, no) ~ party + contributions, family = binom
ial,
##     data = vote.df)
## 
## 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(6.5633,5)

## [1] 0.2551956
```

It is a good model we can trust as its' all coefficient are in 95% CI(p-value < 0.5). and the check of our residual deviance p-value 0.255 is larger than the 0.05.

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

## 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

plot(p~contributions,type="n",data=vote.df,ylim=c(0,1),xlab="Contributi
ons (tens of thousands of dollars)",ylab="Proportion voting YES")
abline(h=c(0,1))
text(p~contributions,labels=party,col=ifelse(party=="D","blue","red"),d
ata=vote.df)
x2=seq(-1,4,length.out=1000)
pred.r=predict(vote.fit2,newdata=data.frame(party=rep("R",1000),contrib
```
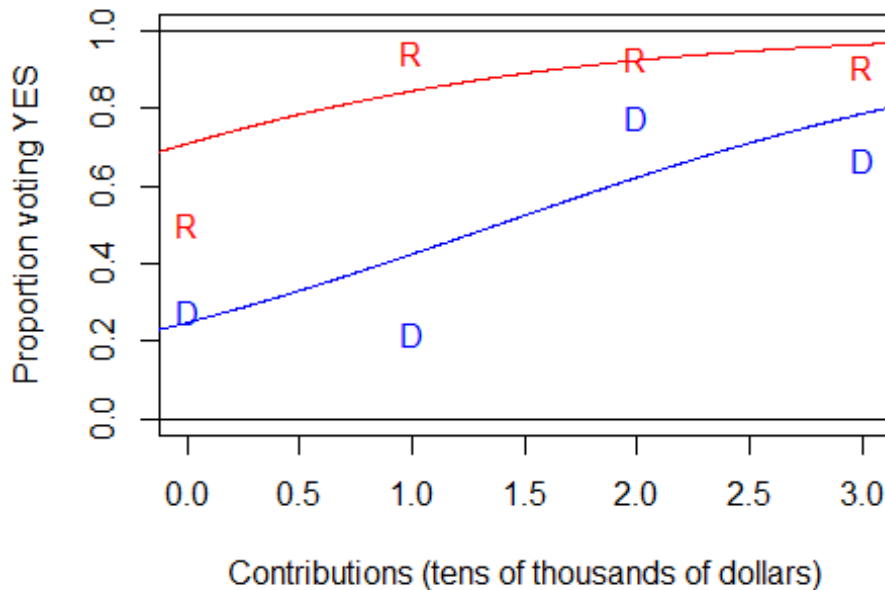
```
utions=x2),type="response")
pred.d=predict(vote.fit2,newdata=data.frame(party=rep("D",1000),contrib
utions=x2),type="response")
lines(x2,pred.r,col="red")
lines(x2,pred.d,col="blue")
```


We plot a predict line above with the specific functions in R. We can see that the increases of contributions come with the increases of proportion of voting yes.

Methods and Assumption Checks

As the response variable stands for the number of senators who voted YES and No on the bill for a particular combination of party, we have therefore generalized linear model with a binomial family, we have two explanatory variables, party(factor) and contributions(numeric) respectively. The plot of the voting proportions show a S association with contribution, so we fit a logitic regression model to these data.

The check of the residual deviance had p-value of 0.255 > 0.05, we can trust our binomial model.

Our final model is:

$$log(Odds_i) = \beta_0 + \beta_1 \times party_i + \beta_2 \times contributions$$

Where $Odds_i$ is the odds of the retention of voting yes, and party is indicating variable, taking 1 if it is Republican otherwise Democrat.

Executive Summary

We estimate that:

- Every 1 contribution(tens of thousands of dollars) increase comes with the increase of the odds between 32.0% and 300.1%.

- With the party Republican, the Odds will be between 2.59 and 23.37 times than the party Democrat.

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.

From the estimation above, the answer is yes. We can see that the coefficient of the contributions is in the 95% CI(p-value < 0.05)

We are also interested to determine if senators from one party are more likely to vote YES on the bill than the other.
The Republic is larger than the Democrat.

Also, does the effect of contributions depend on the political party? Nope, we see that there's no interaction between the contributions and the party, so the effect of contributions do not depend on the political party.

Nope, the contribution and political party have no interaction, so the effect of contributions do depend on the political party.