Judge Voting

J. Bravo

1/13/2022

The desire to apply logistical regression to real-world data sets served as the inspiration for this project. It is advantageous to understand how to implement classification when dealing with qualitative variables. Logistical regression seemed like a good starting point since it deals with binary response variables such 0/1 or true/false. For the data source I went to FiveThirtyEight where I found an article claiming now Supreme Court Justice Neil Gorsuch actually has a moderate record when voting on immigration from 2006 to 2016 on the Tenth Circuit Court of Appeals. Per the article “For A Trump Nominee, Neil Gorsuch’s Record Is Surprisingly Moderate On Immigration”the cases are settled by three-judge panels which implies three votes per case. This specific article also takes into account employer discrimination cases to determine voting trends for the judges.

For my analysis I will only take into account immigratin cases and Vote1, that is the first judge’s vote. For the aforementioned immigration cases, 1 is a liberal vote which is described as voting in favor of the immigration petitioner by granting help such as readjustment of immigration status or asylum. Otherwise, the vote is 0 or a conservative vote

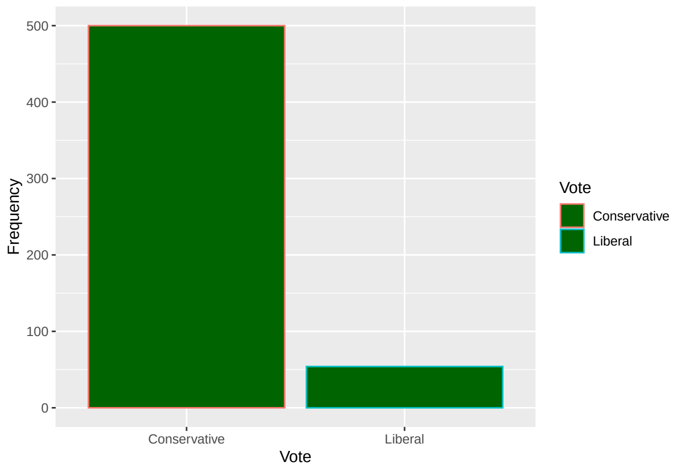
First, the `raw_data` variable will store the data set from FiveThirtyEight. I will then render a data frame which filters for the first vote (Vote1) and immigration cases. In this analysis the vote serves as the binary response variable and to continue with the analysis will make sure to convert 0 to Conservative and 1 to Liberal. This helps maintain the binary nature of the variable for the analysis and being more specific than the ambiguous 0 or 1.

```r
# df1 <- first data frame
df1 <- data.frame(raw_data$Judge1, raw_data$Vote1, raw_data$Category)
colnames(df1) <- c("Judge", "Vote", "Category")
df1 <- df1 %>%
    filter(Category == "Immigration")
df1$Vote <- factor(df1$Vote, labels = c("Conservative", "Liberal"))
```

The first comparison will be the count of conservative votes to the number of liberal votes in the initial vote. The analysis invoked the use of R pipes to clearly track the various operations.

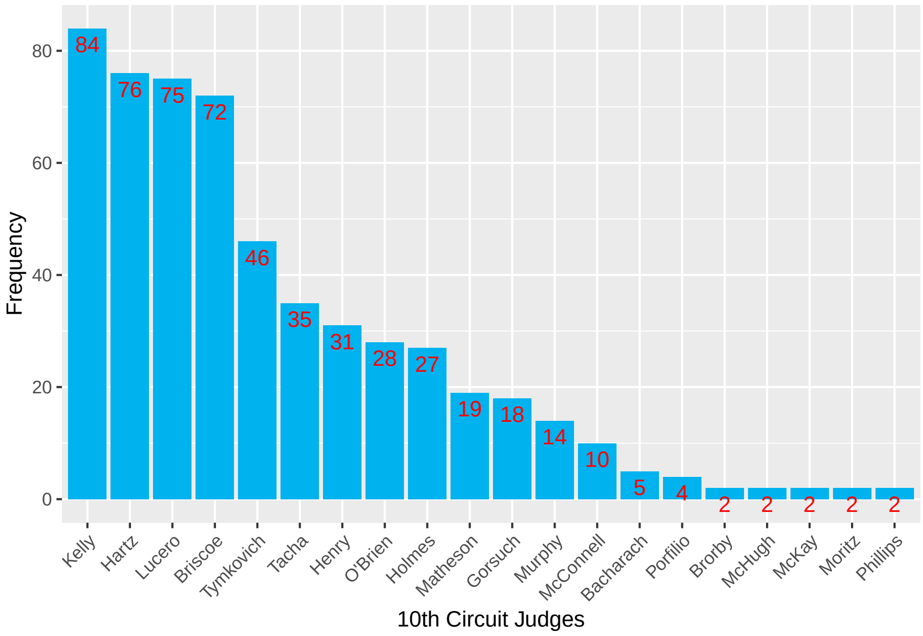
```r
by_vote <- df1 %>% group_by(Vote) %>% count()
colnames(by_vote) <- c("Vote", "Frequency")
by_vote
```

```
## # A tibble: 2 x 2
## # Groups:   Vote [2]
##   Vote         Frequency
##   <fct>            <int>
## 1 Conservative       500
## 2 Liberal             54
```



According to the first plot, it is clear that an overwhelming number of votes were conservative. Now, I will arbitrarily perform further analysis on the conservative votes because the sample size is so much larger than the liberal votes.

```
cons_vote <- df1 %>% group_by(df1$Judge) %>% count("Conservative") %>% arrange(desc(n))
colnames(cons_vote) <- c("Judge", "Vote", "Frequency")
cons_vote <- cons_vote %>% arrange(desc(Frequency))
```



The logic for this plot includes sorting the votes in descending order. Such order allows to visualize the progression from the judge with the most conservative votes to the least number of conservative votes.

Now conisder the following code.

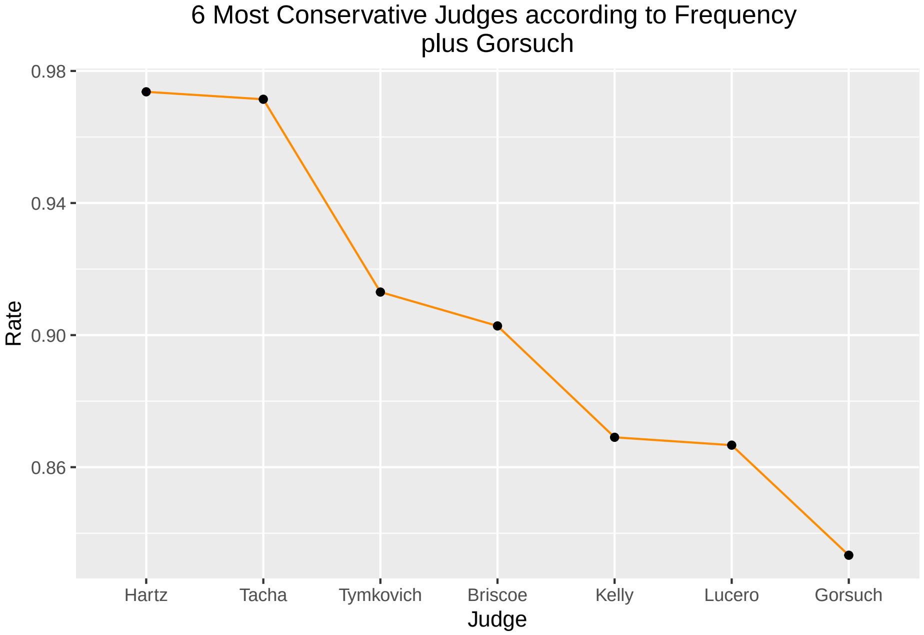
```
median(cons_vote$Frequency)
```

```
## [1] 18.5
```

It is not surprising to see that Judge Gorsuch is close to the median of 18.5 conservative votes. Visually, Judge Gorsuch is near the bisect of the plot which implies its proximity to the median. Indeed Judge Gorsuch had 18 conservative votes in this first round.

Having determined the six most conservative judges according to number of votes, want to compute the conservative voting rate. The rate is defined as the ratio of conservative votes by each judge to the total number of votes cast by the judge. I will then incorporate Judge Gorsuch's own voting rate into the next plot to compare his voting record.

```
conservative_prop <- df1 %>% filter(Vote == 'Conservative') %>% count(Judge)
table1 <- df1 %>% group_by(Judge) %>% count()
total_vote <- table1 %>% pull(n)
# Conservative proportions
conservative_prop <- cbind(conservative_prop, total_vote)
colnames(conservative_prop) <- c("Judge", "n", "TotalVotes")
conservative_prop <- conservative_prop %>% arrange(desc(n))
# Initiate empty column
conservative_prop["Rate"] <- NA
# For loop to fill empty column and calculate rate.
for(j in 1:length(conservative_prop$n)) {
  conservative_prop$Rate[j] <- conservative_prop$n[j]/conservative_prop$TotalVotes[j]
}
# New data frame to store just the top 6 conservative judges and their voting rate.
df2 <- data.frame(head(conservative_prop$Judge), head(conservative_prop$Rate))
colnames(df2) <- c('Judge', 'Rate')
gorsuch3 <- conservative_prop[11, ]
gorsuch4 <- subset(gorsuch3, select = -c(n, TotalVotes))
# t6cj <- Top 6 conservative judges + gorsuch
t6cj <- rbind(df2, gorsuch4)
```



The last plot seems to suggest that Judge Gorsuch is not as conservative as six of his more conservative colleagues because he holds a smaller conservative voting rate. This helpes support the argument of the original article.

The final part of this analysis will include configuring a training data set to train the logistical (logit) model.

```
# Want a subset for logistical regression model.
training_data <- subset(df1, select = c(Judge, Vote))
# Will fit a logit model in order to predict Vote using Judge.
```

```
logit.model <- glm(Vote ~ Judge, data = training_data, family = binomial)
summary(logit.model)
```

```
## 
## Call:
## glm(formula = Vote ~ Judge, family = binomial, data = training_data)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -1.1774  -0.5298  -0.4523  -0.2408   2.6972  
## 
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)  
## (Intercept)     -1.386e+00  1.118e+00  -1.240   0.2150  
## JudgeBriscoe    -8.422e-01  1.187e+00  -0.710   0.4779  
## JudgeBrorby      1.386e+00  1.803e+00   0.769   0.4419  
## JudgeGorsuch    -2.231e-01  1.285e+00  -0.174   0.8621  
## JudgeHartz      -2.225e+00  1.328e+00  -1.675   0.0939 .
## JudgeHenry      -5.232e-01  1.240e+00  -0.422   0.6730  
## JudgeHolmes     -6.931e-01  1.275e+00  -0.544   0.5866  
## JudgeKelly      -5.063e-01  1.164e+00  -0.435   0.6636  
## JudgeLucero     -4.855e-01  1.168e+00  -0.415   0.6778  
## JudgeMatheson   -1.504e+00  1.518e+00  -0.991   0.3219  
## JudgeMcConnell   2.437e-14  1.369e+00   0.000   1.0000  
## JudgeMcHugh     -1.518e+01  1.697e+03  -0.009   0.9929  
## JudgeMcKay      -1.518e+01  1.697e+03  -0.009   0.9929  
## JudgeMoritz     -1.518e+01  1.697e+03  -0.009   0.9929  
## JudgeMurphy     -1.179e+00  1.525e+00  -0.773   0.4397  
## JudgeO'Brien    -1.179e+00  1.337e+00  -0.881   0.3781  
## JudgePhillips    1.386e+00  1.803e+00   0.769   0.4419  
## JudgePorfilio   -1.518e+01  1.200e+03  -0.013   0.9899  
## JudgeTacha      -2.140e+00  1.510e+00  -1.417   0.1563  
## JudgeTymkovich -9.651e-01  1.234e+00  -0.782   0.4343  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 354.00  on 553  degrees of freedom
## Residual deviance: 333.71  on 534  degrees of freedom
## AIC: 373.71
## 
## Number of Fisher Scoring iterations: 15
```

From the summary of the logistical model above consider the coefficient for Judge Hartz at -2.225.In this multiple logisitic model, the negative coefficient implies that when votes of the other judges are fixed, Judge Hartz is less likely to vote Liberal than their collegues. Judge Tacha has a similar coeffieint at -2.14. Although the p-value of 0.0939 for Judge Harts and 0.1563 for Judge Tacha are the smallest values, they are still relatively large implying insufficient evidence of a real association between Judge Hartz or Judge Tacha and Vote.It is also noteworthy that both Judge Hartz and Judge Tacha have the highest conservative voting rate as demonstrated in the earlier plot.

The small absolute value for z-statistic for all the judges implies evidence for the null hypothesis which in this case is that the probability of Vote does not depend on Judge.

Now I want to predict the probability that a Judge votes Liberal given the predictor values.

```
# Invoke predict function in R.
logit.prob <- predict(logit.model, type = "response")
```

In the following logic, will make sure that R assigned dummy variable 1 for liberal. In order to test the model will create a vector of 554 conservative votes and transform all elements whose predicted probability exceeds 0.50 to Liberal vote.

```
contrasts(training_data$Vote)
```

```
##              Liberal
## Conservative       0
## Liberal            1
```

```
logit.pred <- rep("Conservative", 554)
logit.pred[logit.prob > 0.5] = "Liberal"
```

The final steps for this analysis project is to use the table function to get the confusion matrix where the diagonal elements are the correct predictions.

```
table(logit.pred, training_data$Vote)
```

```
## 
## logit.pred     Conservative Liberal
##   Conservative          499      53
##   Liberal                 1       1
```

```
print((499+1)/554)
```

```
## [1] 0.9025271
```

```
# Alternatively
mean(logit.pred == training_data$Vote)
```

```
## [1] 0.9025271
```

So logistic regression predicted correctly 90.2% of time. This implies that the training error rate is 9.8%.

This model has three distinct limitations. First, I only took into account Vote1. A more realistic model incorporates Vote2 and Vote3 especially because the original article mentions that "judges' votes are influenced by their panel colleagues". If judges' opinions are distorted by their colleagues' actions this is something to take into account. Second, the original article also incorporated how the judges in the 10th circuit voted with respect to employment discrimination cases. Including voting trends with respect to another case type would give deeper insight into the voting trends of the judges. It is convenient that the qualitative binary aspect is maintained. Last, the model was trained and tested on the same data set. To obtain a more accurate model it may be beneficial to split the data set into two and use one for training the model and then other half for testing.

So, this initial and quick analysis demonstrated the logistical regression as a viable classification method when dealing with binary response variables. This is an example of developing statistical learning models from qualitative data.