Politics Are Afoot! - Part 2

Victor Ramirez

Intro

To discover evaluate the relationship of the total disbursement (spend) and the general votes relationship an understanding and analysis of the data was required. The analysis was conducted using the 2016 election cycle dataset provided by the fec16 package. The data set included the following data, candidate attributes, like their name, a unique id of the candidate, the election year under consideration, the office they're running for election. The results_house data contained race attributes, name of the candidates running in the election, a unique id of the candidate, the number of general_votes garnered by each candidate. The campaigns data included financial information for each house & senate campaign. This includes a unique candidate id, the total receipts, and total disbursements.

Data

The analysis of the data included the following steps, data understanding and reporting, data wrangling, data transformation, and data modeling. The following features were used to evaluate the data and build the model. The total disbursement and general votes. The data required was in two different data frames, so a join based on candid_id was required. As well as party representation needed to be explicitly cleaned. The following three options were generated during the join, Republican, Democrat and Other Party. The data required transformation so a log base 10 transformed the data and reviled a density formation in the plot. With the newly generated political party grouping we can also see the political party groupings and clustering.

Analysis

To begin the analysis, we must first confirm that the large-sample assumptions are satisfied. All assumptions satisfied: 1. Sample size is >= 100 2. IID- The distribution for the total population is true. All the data gathered and is represented independent for all observations. 3. BLP - With the data distribution there is a unique BLP that exists. All co variances are finite. There is no perfect col linearity, and the expectation of x is invertible. With the data transformation no heavy tails and no col linearity is present. Having transformed and explored the data the task of creating an applicable model was required. In the initial model building 4 different models where created: 1. The raw votes and spend data used: model_1 <- lm(general_votes ~ ttl_disb, data = df_inner_join) 2. The log of both votes and spend used: model_2 <- lm(log(general_votes) ~ log(ttl_disb), data = df_inner_join) 3. The log of votes and raw spend used: model_3 <- lm(log(general_votes) ~ ttl_disb, data = df_inner_join) 4. The log of votes and spend with the additional party coefficient used: model_4 <- lm(log(general_votes) ~ log(ttl_disb) + factor(party_new), data = df_inner_join)

Results

The Coefficient Results:

Standard Error:

The coefficient Standard Error measures the average amount that the coefficient estimates vary from the actual average value of our response variable. The tested result was a lower number relative to the coefficients.

t-value:

The coefficient t-value is a measure of how many standard deviations our coefficient estimate is far away from 0. The test result was very far away from zero, this indicates we could reject the null hypothesis - that is, we could declare a relationship between spend and votes exist.

Pr(>t):

Variable	Determination
`log(ttl_disb)`	Significant
`incumbentTRUE`	Significant
`factor(party_new)Other Party`	Significant
`factor(party_new)Republican`	Not Significant

The Pr(>t) in the model output relates to the probability of observing any value equal or larger than t. The small p-value indicates that it is unlikely we will observe a relationship between the predictor (spend) and response (votes) variables due to chance.

Conclusion

I believe that there is a relationship between the total spend and votes received. There is one noticeable piece of information regarding the vote received. Using the model with the log transformation of the data and including the political party and if the candidates was an incumbent.