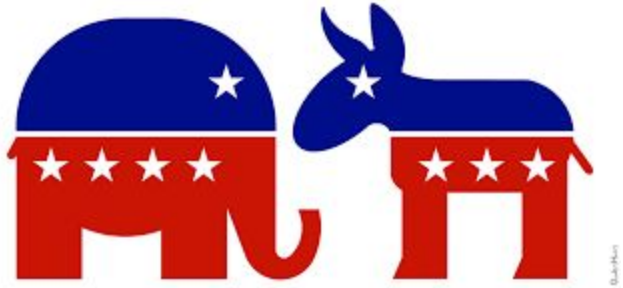
Analyzing Campaign Finance Data in California using SAS & R

By Robby Born

Background Info

- ➔ **In the US in every even year, Americans cast their vote in all 50 states in national, state, and local elections**
- ➔ **This data was collected on kaggle from the Federal Election Commission (FEC) for campaign finance data from 2019-2020 leading up to the 2020 election for all major US house of representatives candidates**
- ➔ **I chose to further analysis campaign finance data for my home state of California to better understand where each party was collecting their political campaign contributions from**

Voting!



- Live in California
- Difference in campaign finance contributions by political party?
- Pertinent as the 2022 midterm elections occurred this month

The Data

The LOGISTIC Procedure

Model Information	
Data Set	WORK.CALI
Response Variable	CAND_PTY_AFFILIATION
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	330
Number of Observations Used	330

Response Profile		
Ordered Value	CAND_PTY_AFFILIATION	Total Frequency
1	DEM	180
2	IND	5
3	REP	145

FEC.gov

Variables ~

- **CAND_PTY_AFFILIATION** = political party
- **CAND_LOANS** = loans from the candidate
- **Receipts** = total $ receipts for the candidate
- **TTL_INDIV_CONTRIB** = contributions made from the candidate themselves

Data Cleaning


```
proc import out=cand
  datafile="/home/u60876478/410Project/All candidates.csv"
   dbms=csv replace;
run;

proc sql;
    create table cali as
    select *
    from cand
    where CAND_OFFICE_ST = 'CA';

proc sql;
   delete
      from cali
      where CAND_PTY_AFFILIATION = 'UNK' or CAND_PTY_AFFILIATION = 'OTH'
      or CAND_PTY_AFFILIATION = 'NON' or CAND_PTY_AFFILIATION = 'GRE' or
      CAND_PTY_AFFILIATION = 'NPA' or CAND_PTY_AFFILIATION = 'NNE' or
      CAND_PTY_AFFILIATION = 'LIB';
```


```
library(dplyr)
library(tidyverse)

Cali = All_candidates %>%
  rename(state = CAND_OFFICE_ST)
Cali = Cali %>%
  filter(state == "CA") %>%
  rename(party = CAND_PTY_AFFILIATION) %>%
  rename(receipts = TTL_RECEIPTS)

Cali = Cali %>%
  filter(party != "GRE")
Cali = Cali %>%
  filter(party != "NPA")
Cali = Cali %>%
  filter(party != "NON")
Cali = Cali %>%
  filter(party != "LIB")
Cali = Cali %>%
  filter(party != "OTH")
Cali = Cali %>%
  filter(party != "UNK")
Cali = Cali %>%
  filter(party != "NNE")
```

SAS and R Codes

SAS:

```
proc logistic;
 model CAND_PTY_AFFILIATION(ref="DEM") = TTL_RECEIPTS TTL_DISB CAND_LOANS TTL_INDIV_CONTRIB / link=glogit;
 run;
```

R:

```
library(nnet)

fit = multinom(party ~ receipts + CAND_LOANS + TTL_INDIV_CONTRIB,
               data = Cali)
summary(fit)
null = multinom(party ~ 1, data = Cali)

print(deviance <- deviance(null)-deviance(fit))

print(p.value<- pchisq(deviance, df=12, lower.tail = FALSE))
```

SAS and R Outputs

R

```
Call:
multinom(formula = party ~ receipts + CAND_LOANS + TTL_INDIV_CONTRIB,
    data = Cali)

Coefficients:
    (Intercept)      receipts   CAND_LOANS TTL_INDIV_CONTRIB
IND  -2.8902295 -3.840985e-04 3.049641e-04      3.853073e-04
REP  -0.2283919 -9.474754e-07 1.277966e-05      1.040096e-06

Std. Errors:
     (Intercept)     receipts   CAND_LOANS TTL_INDIV_CONTRIB
IND 1.033757e-08 5.809849e-05 1.063057e-04      5.825450e-05
REP 4.579551e-09 4.865105e-07 4.817382e-06      5.806167e-07

Residual Deviance: 474.0105
AIC: 490.0105
```

```
print(deviance
L] 24.57901
```

```
print(p.value
1] 0.01694874
```

SAS and R Outputs

SAS

Parameter	CAND_PTY_AFFILIATION	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	IND	1	-2.9924	0.4967	36.2889	<.0001
Intercept	REP	1	-0.2284	0.1242	3.3808	0.0660
TTL_RECEIPTS	IND	1	-0.00012	0.000180	0.4178	0.5181
TTL_RECEIPTS	REP	1	-9.47E-7	4.977E-7	3.6248	0.0569
CAND_LOANS	IND	1	0.000043	0.000244	0.0306	0.8610
CAND_LOANS	REP	1	0.000013	4.857E-6	6.9215	0.0085
TTL_INDIV_CONTRIB	IND	1	0.000117	0.000181	0.4196	0.5171
TTL_INDIV_CONTRIB	REP	1	1.04E-6	5.875E-7	3.1338	0.0767

Likelihood Ratio	23.9767	6	0.0005

AIC	502.590	490.613

Fitted Model, Significance, & Interpretation

Fitted model:

$$\frac{\widehat{P}(IND)}{\widehat{P}(DEM)} = exp(-2.9924 - .00012(receipts) + .000244(candidate\ loans) + .000117(individual\ contributions)$$

$$\frac{\widehat{P}(REP)}{\widehat{P}(DEM)} = exp(-0.2284 - 9.47E-7(receipts) + .000013(candidate\ loans) + 1.04E-6(individual\ contributions)$$

Significant Predictor - CAND_LOANS for the odds of REP vs. DEM

As candidate loans increase by $1, the estimated odds in favor of republicans against democrats increase by (exp(.000013)-1)*100% = 0.0013%

Prediction SAS & R

SAS

```
 data pred;
input TTL_RECEIPTS CAND_LOANS TTL_INDIV_CONTRIB;
cards;
570000 123450 300
;

data Cali;
set cali pred;
run;

proc logistic;
 model CAND_PTY_AFFILIATION(ref="DEM") = TTL_RECEIPTS CAND_LOANS TTL_INDIV_CONTRIB / link=glogit;
 output out=outdata p=poutcome;
run;
proc print data=outdata noobs;
 var _level_ poutcome;
run;
```

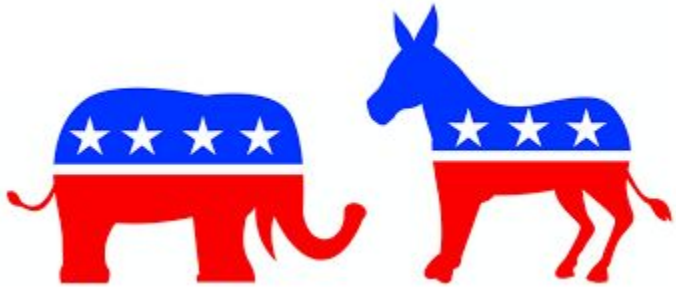
Prediction Output

DEM	0.30802
IND	0.00000
REP	0.69198

Prediction SAS & R

R

```
> pred = predict(fit, data.frame(receipts=570000, CAND_LOANS = 123450,
+                  TTL_INDIV_CONTRIB = 300), type = "prob")
> print(pred)
          DEM           IND           REP
3.079954e-01 3.556468e-81 6.920046e-01
```



Thank you!

I would like to thank

- **My professor, Dr Olga**
- **All of my STAT 410 class**
- **The United States Federal Election Commission (FEC) for providing data on campaign finance contributions for all parties**