

# Election Prediction

*Daniel Spakowicz*

*11/7/2016*

Created as part of STAT 625: Statistical Case Studies

This script was created after discussions with Dennis Walsh, though no code was shared.

This script reads in a table of poll results, reduces the set to those after the national conventions, and computes a margin of victory based on the votes for Clinton and Trump only. The distribution of the margins of victory are used to calculate a weighted mean, where the weights are adjustments for the grade assigned to each poll and its age. A normal distribution using the mean and variance of the margin of victory distribution is used to calculate a probability of Clinton winning the state.

	State	MOV	MOV.lwb	MOV.upb	ProbClinton
1	Florida	0.01	-0.05	0.07	0.57
2	Michigan	0.07	0.02	0.12	0.92
3	North Carolina	0.01	-0.03	0.06	0.62
4	Pennsylvania	0.06	0.02	0.10	0.92
5	Colorado	0.07	-0.01	0.16	0.80
6	Ohio	-0.01	-0.06	0.04	0.41
7	Wisconsin	0.09	0.00	0.17	0.85
8	Minnesota	0.11	0.06	0.16	0.99
9	Arizona	-0.00	-0.06	0.06	0.50
10	New Hampshire	0.09	-0.03	0.21	0.78
11	Nevada	0.01	-0.04	0.06	0.60
12	Georgia	-0.06	-0.13	0.01	0.21
13	Iowa	-0.01	-0.08	0.06	0.44
14	Texas	-0.10	-0.17	-0.03	0.07

## Appendix

```
# Used for display of the result
library(xtable)

# Read in table of recent polls.
x <- read.csv("president_polls_Nov1.csv", as.is = TRUE)

# States to study
states <- c("Florida", "Michigan", "North Carolina", "Pennsylvania", "Colorado",
            "Ohio", "Wisconsin", "Minnesota", "Arizona", "New Hampshire",
            "Nevada", "Georgia", "Iowa", "Texas")

# Get only the states of interest
x <- x[x$state %in% states,]

# Format date cols
x$startdate <- as.Date(x$startdate, format = "%m/%d/%y")
x$enddate <- as.Date(x$enddate, format = "%m/%d/%y")

# # Order by state, then date
# x <- x[order(x$state, -x$enddate),]
```

```

# Subset to only those that occurred after the last convention
x <- x[x$enddate >= "2016-07-21",]

#####
### Weighted mean of all post-convention polls, adjusted for grade and age

# Create a data frame of grade correction scores
grade.cor <- data.frame(frac = seq(1, 0.5, length.out = 10),
                        grade = c("A+", "A", "A-", "B+", "B", "B-", "C+",
                                "C", "C-", "D"))

# Correct the poll sample size by grade
for (i in 1:nrow(x)) {
  if (x$grade[i] == "") {
    x$samplesize.grade[i] <- x$samplesize[i] * 0.75
  } else {
    x$samplesize.grade[i] <- x$samplesize[i] *
      grade.cor$frac[which(grade.cor$grade == x$grade[i])]
  }
}

# Correct for the age of the poll
dates <- seq.Date(from = min(x$startdate), to = Sys.Date(), by = 1)
age.cor <- data.frame(date = dates,
                      correction = seq(from = 0.2, to = 1,
                                       length = length(dates)))

# Correct the sample size by age of poll
for (i in 1:nrow(x)) {
  x$samplesize.grade.age[i] <- x$samplesize.grade[i] *
    age.cor$correction[which(age.cor$date == x$enddate[i])]
}

# Calculate the proportion of Clinton votes (b/n Clinton and Trump only)
x$pc <- (x$rawpoll_clinton * x$samplesize.grade.age) /
  ((x$rawpoll_clinton * x$samplesize.grade.age) +
   (x$rawpoll_trump * x$samplesize.grade.age))

# Calculate the proportion of Trump votes (b/n Clinton and Trump only)
x$pt <- 1 - x$pc

# Find the Margin of Victory (>0 Clinton, <0 Trump)
x$MOV <- x$pc - x$pt

# Preallocate memory to data frame
df <- data.frame(State = states,
                  MOV = NA,
                  MOV.lwb = NA,
                  MOV.upb = NA,
                  ProbClinton = NA)

# Calculate the weighted mean MOV and standard deviations for each state

```

```

for (i in 1:length(states)) {
  df$MOV[i] <- round(weighted.mean(x$MOV[x$state == states[i]],
                                   w = x$samplesize.grade.age[x$state ==
                                                                    states[i]]), 3)
  df$MOV.lwb[i] <- round(df$MOV[i] - sd(x$MOV[x$state == states[i]]), 3)
  df$MOV.upb[i] <- round(df$MOV[i] + sd(x$MOV[x$state == states[i]]), 3)
}

# Assign probability based on the density > 0 in the MOV distribution
df$ProbClinton <- round(1 - pnorm(0, df$MOV, sd = (df$MOV - df$MOV.lwb)), 3)

# Write results to table
write.csv(df, "Election2016_Nov8_djs88.csv", row.names = FALSE)

print(xtable(df), caption = "Margin of victories (>0 Clinton, <0 Trump) with one standard deviation bound",
      caption.placement = "top", comment = FALSE)

```