

Statistics 406 Lab 10
GSI: Adam Rothman
Tuesday, November 21

1. Load the data frames:

```
house_polls.dat  
house_results.dat
```

into R and display both tables.

Solution:

```
P <- read.table('house_polls.dat')  
R <- read.table('house_results.dat')  
P  
R
```

2. Convert the race results into percentages.

Solution:

```
T <- R[,2] + R[,3]  
R[,2] <- 100 * R[,2] / T  
R[,3] <- 100 * R[,3] / T
```

3. Add an additional column to the result table corresponding to Democrat percentage minus Republican percentage. Display the resulting table with the extra column.

Solution:

```
R <- cbind(R, R[,2]-R[,3])
```

4. Create a list of race names for which there is polling data, call this list D. Then figure out how many races have polling data.

Solution:

```
D <- unique(as.vector(P[,2]))  
D  
length(D)
```

5. Find the row in the result table that corresponds to race NC – 08. Then display this row.

Solution:

```
i2 <- which(R[,1] == "NC-08")
R[i2,]
```

6. List the races for which the Republican won.

Solution:

```
as.vector(R[R[,4] < 0, 1])
```

7. Calculate the standard deviations of the poll averages for polls taken after day 290. Assume a margin of error of 3.5 percent.

Solution: First note that the margin of error of k polls is $1/\sqrt{k}$ times the margin of error of any one poll. Thus the overall margin of error for k polls is $2 * 3.5/\sqrt{k}$. The poll standard deviation is $\hat{\sigma}/\sqrt{n}$.

$$1.96 * \hat{\sigma}/\sqrt{n} = 2 * 3.5/\sqrt{k}$$

$$SD_{\text{pollAve}} = \hat{\sigma}/\sqrt{n} = 7/(1.96\sqrt{k})$$

```
SD <- NULL
## Loop through the races.
for (d in D)
{
  ## Get the polls for race d.
  i1 <- which(P[,2] == d)

  ## Get the recent polls (roughly three weeks back).
  i1 <- i1[P[i1,1] > 290]
  if (length(i1) == 0) { next }

  ## The standard deviation for the poll average difference.
  sdev <- 7/(1.96*sqrt(length(i1)))

  SD <- c(SD, sdev)
}
```

8. Repeat 7 using only a single poll for each race.

Solution:

```
SD <- NULL
## Loop through the races.
for (d in D)
{
```

```
## Get the polls for race d.
i1 <- which(P[,2] == d)

## Get the recent polls (roughly three weeks back).
i1 <- i1[P[i1,1] > 290]
if (length(i1) == 0) { next }

## Take the first poll only
i1 <- i1[1]
## The standard deviation for the poll average difference.
sdev <- 7/(1.96*sqrt(length(i1)))

SD <- c(SD, sdev)
}
```

9. Say an individual is two attributes X and Y . Both are standard normal random variables. Suppose we wish to sample individuals for a poll, but that there is a small bias in the polling. Assume that anyone who is sampled for inclusion in the poll but that has $X > 1$ is randomly excluded with probability θ . If a person is excluded, then generate another person as a replacement (if the replacement has $X > 1$ then repeat the random selection, possibly leading to a third candidate being considered, and so on). Generate ten individuals that will be included in the poll, display each individual's X and Y value.

Solution:

```
k <- 1
n <- 10
theta <- 0.5
xytable <- array( 0, c(n,2) )

while(1)
{
  includeMe <- TRUE
  X <- rnorm(1)
  Y <- rnorm(1)

  if ( X > 1 && 1*(runif(1) < theta) )
  {
    includeMe <- FALSE
  }

  if ( includeMe )
  {
    xytable[k,1] <- X
    xytable[k,2] <- Y
    k <- k+1
  }

  if( k > n)
  {
    break
  }
}
xytable
```