Li Liu

Assignment 4

1. **Non-probability sampling phone survey**

A. See PhoneSurvey(Response).xlsx

B. I made139 phone calls and 21 of them responded. 118 numbers were not responded or I could not speak with people over 18. The response rate is around 15%.

C. 10 of the 21 responders answered the voting question and age question.

D. I called them on Sunday evening from 5 p.m to 7 p.m. Since Philadephia is one hour ahead of Chicago, most people should be at home and not busy with work. This period ensures people have enough leisure time to spend one minute on the phone with a stranger. I would expect the rate to be lower if I call during the day.

E. The median age of my respondents is 37. U.S Census Bureau (2016) estimate the median age in Pennsylvania of 2016 to be 40. The sample median doesn't quite match the State data for two main reasons. First, the sample size is too small and the variance is large. Second, Philadephia (my assigned area)'s population is younger than the whole state. The Metropolitan city is not representative of the state's demographics.

F. In my sample of 10, 50% of respondents voted Republican, 40 % respondents voted Democrats. However, the actual voting percentages for Trump and Clinton are 48.8% and 47.6% (Politico, 2016). The percentage for Republican is close, but the percentage for Democrats is lower. Since this is just a small sample, we cannot draw any significant inference from the data.

One potential issue is the order of saying the candidates or categories might influence people's willingness to collaborate with the callers. To test whether this factor matters, we could design an experiment on a large sample of respondents. We randomly divide them into two groups (treatment and control groups). When calling the people in the treatment group, we would say the order of candidates or categories differently. Then we could compare the metrics (response rate, voting percentages, etc) of two groups to determine whether the order matters.

Reference:

2012-2016 American Community Survey 5-Year Estimates, U.S. Census Bureau (2016). Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF

2016 Presidential Election Results, Politico (2016). Retrieved from https://www.politico.com/mapdata-2016/2016-election/results/map/president/

2. Response to Wang et al. (2015)

Wang, Rothschild, Goel, & Gelman (2015) demonstrate that non-representative polls could be useful for predicting elections by appropriate statistical adjustment. The eight variables reported from the Xbox respondents are “Sex (2 categories), Race (4 categories), Age (4 categories), Education (4 categories), State (51 categories), Party ID (3 categories), Ideology (3 categories) and 2008 vote (3 categories)” (Wang et al., 2015, p. 982). Among these variables, “Sex”, “Age”, and “Education” from the Xbox sample are the least representative of the 2012 Exit Poll data. The majority of the Xbox users seems to be young (18-29 years old) males with some college degrees. Forecasting using this atypical sample from the voting population would wrongly estimate the support rates. “Race”, “State”, and “2008 Vote” are the most representative of the 2012 Exit Poll data.

Wang et al., (2015) combine two data sources to perform a post-stratification re-weighting of the respondents: one is the “exit poll data from the 2008 presidential election” to calculate weights, another is the re-weighted “Xbox data on the day before the election” (p. 985-986).

According to Figure 2 (p. 982) and Figure 3 (p. 984) reported in Wang et al., (2015), in the last three weeks of the election, Xbox raw data would have predicted Romney to win as the two-party support rate for Obama is below 50%. Pollster.com’s prediction was

uncertain as the support rates for both parties are around 50%. Xbox post-stratified would have predicted Obama to win, which turns out to be true.

Reference:

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting, 31:3*, 980-991.