MACS Assignment 4
Neta Grossfeld
October 31, 2018

Question 1

a) See PhoneSurvey spreadsheet.
b) I called all 200 numbers. 3 people responded according to my response variable, and 197 did not. My response rate is 1.5%.
c) For those who did respond, 2 out of 3 answered the voting question and 2 out of 3 answered the specific age question. However, 3 out of 3 answered that they were older than 18 years.
d) It was late morning to early afternoon, or 11 AM to 1 PM. During this time of day, most respondents should have been awake to answer their phone, and perhaps starting or were close to their lunch breaks. However, this is also a time when individuals are typically working, and so may not have answered due to work protocol.
e) Since only 2 out of 3 respondents responded with their ages, they are 27 and 60. The median age in Texas is 34.5. I don't have a median age based on my respondents to compare to the state's median age.
f) 50% of my respondents voted Republican, and 50% of my respondents voted Democrat. In the 2016 election, 52% of Texas voted Republican and 43% of Texas voted Democrat. One way to test if the order of the candidates/categories influence results would be to ask the question in different ways to various groups of respondents, such as "Who did you vote for? Which party did you vote for?" or simply changing the order of the question we have currently for the assignment.

Question 2

In the article, *Forecasting Elections with Non-Representative Polls*, Wang et al. used the Xbox gaming system to collect data from Xbox players in the 45 days before the 2012 U.S. presidential election. In the first poll that the Xbox respondents participated in, additional data or variables were collected, including sex, race, age, education, state, party identification, political ideology, and their 2008 vote. Out of these eight variables, the least representative were age, sex, and education. For these variables, the Xbox sample differs so greatly from the broader voting population because of the fact that Xbox is a gaming platform that is geared towards and thus most popular with the younger, male crowd. Since being younger can be correlated with the level of education a respondent has, that would also explain why the education levels of Xbox players are not representative of the broader voting population. The most representative variables were race, state, and 2008 vote.

The authors use two data sources, the raw Xbox data and the 2008 exit polls from Pollster.com, to perform post-stratification in an attempt to accurately estimate voter intent for the 2012 election by weighing observations based on the variables mentioned above to create a

more representative dataset. We can see the post-stratification results when we compare the raw Xbox data to the 2008 exit poll data to the post-stratified Xbox data. In the last 3 weeks of the election, raw Xbox data would have predicted that Romney would win. In the same timeframe, the 2008 exit polls data would have predicted that Obama would win by a small margin and the post-stratified Xbox data would have also predicted that Obama would win, but by a greater margin that is closer to the actual outcomes than the pollster data forecasted. By comparing the trends provided by these three datasets, we can see that the proper statistical adjustments such as post-stratification of the raw data can lead to greater forecasting accuracy despite the raw data being non-representative of the broader voting population, which could lead to significant cost reduction for future forecasting purposes.