Assignment #4

Cosette L. Hampton

MACS 30000, Dr. Evans

Due Wednesday, Oct. 31 at 11:30am

(a) Submit your filled out version of the PhoneSurvey.xlsx spreadsheet.

See spreadsheet.

(b) How many numbers did you call? How many people responded according to your Response variable? How many people did not respond according to your Response variable? What is your response rate?

I called 137 numbers over the course of 3 hours.

(c) What fraction of those for whom Response = 1 answered the voting question? What fraction of those for whom Response = 1 answered the age question?

I only had two people response = 1 and both of them answered the voting question (100%) . Only one person answered the age question (50%)

(d) What time of day was it in the area codes you called when you called them? What role did the time of day play in your response rate?

I split calls between two ours, one 3-4pm CST, one 5-6pm CST, and the other 10-11am CST. I only got answers during 6-7pm, which in New York is 6-7pm CST. People did pick up the phone 3-4 and 10-11 but as soon as I said I was from the University of Chicago they would hang up on me. I thought through the best times to call though before I started calling, thinking about lunch hours and hours outside of work or classtime.

(e) What is the median age of your respondents? How does that compare to the average age in the state of the phone numbers you called? What are some reasons why your sample median does or does not match the State data?

The only age I got was 68. According to Fact Finder[[1]](#footnote-1) based on the 2016 American Community Survey, the average age in NY is 38. I think it does not match because maybe older people are more likely to answer calls from numbers they do not know. Younger people may be more wary of answering unknown callers, or even responding to survey responses. Older people also may not be working as much as younger adults who may be busier.

(f) What percent of your respondents voted Republican (Trump) in the 2016 U.S. Presidential election? What percent of your respondents voted Democrat (Clinton)? How do those percentages compare to the actual voting percentages from the 2016 election? How might you test if the order in which you say the candidates or categories in the survey question in

the results?

0% of my respondents voted for trump. 50% voted Democrat and 50% did not vote at all or if we’re only counting those who did vote, 100% voted for Clinton. According to Politico 58.8% voted for Clinton of all people who voted. If not based on the Electoral College, both my outcome and Politico’s outcome would have had HRC winning. If I were testing if the order I’m using is biased, I’d increase my caller size to 400 and ask 200 if they voted Republican first and 200 if they voted Democrat first, and after each the opposite party, then “Other.” This would provide some comparison in answers given the order of the question so there’s no priming-bias.

**2. Predicting elections survey, Wang, Rothschild, Goel, and Gelman (2015) (3 points). Read the paper Wang et al. (2015), and write a one-to-two-page responding to the following questions.**

(a) Of the eight variables reported from the respondents, which three from the Xbox sample are the least representative of the data and which three are the most representative? For the three least representative variables, why do you think the Xbox sample would be so different from the broader voting population?

(b) What two data sources do the authors use to perform a post-stratification re-weighting of the respondents? Figure 5 shows the results of degree of representativeness of the re-weighted data.

(c) Consider Xbox raw (unweighted) data, Pollster.com forecast data, and Xbox post-stratified data. And consider a prediction of a 2012 U.S. Presidential election outcome as Obama wins, Romney wins, or uncertain. What would Xbox raw have predicted in the last three weeks of the election? What would Pollster.com have predicted during the last three weeks of the election? What would Xbox post-stratified have predicted in the last three weeks of the election?

Wang et al. (2015) use Xbox and Pollster.com to collect data about respondents voting intentions for the 2012 U.S. presidential election. The goal of the paper was to look at data samples that were and were not demographically and characteristically representative of the total population, and use different types of analyses, like mainly “multilevel regression and poststratification” (MRP) (Wang et al., 981) to improve the accuracy of responses.

The eight variables that were reported from the Xbox respondents were: 1) sex, 2) race, 3) age, 4) education, 5) State, 6) party ID, 7) political ideology and 7) who they voted for during the 2008 presidential election (Wang et al., 981). According to the chart on page 982, the variables least representative of the data are sex, age and maybe party ID. I think the Xbox sample is different in these 3 categories from the broader voting population because the Xbox was introduced in 2001[[2]](#footnote-2) or only 12 years before the survey was taken, and young people tend to be more familiar with newer technology and as a gaming system, young people have more leisure time to game than older adults. Studies show that young men are more likely to play videogames than young women for various intrinsic and extrinsic reasons (Romrell, 1), so it explains why for Xbox respondents men may be oversampled. Finally, according to Wang et al., “…age and sex are strongly correlated with voting preferences,” (Wang et al., 981) so with young people and males being least representative, it may follow that party ID may have the same difference. While the composition of Republicans for the general population and Xbox players isn’t too far off from one another, there are more “Other” in the Xbox population and more Democrats in the general survey population than the Xbox population. The three most representative variables are the race, state, and the 2008 vote.

The purpose of poststratification, according to Wang et al., is to split data into groups of cells and to weigh them according to the proportion of the electorate in each cell, made appropriate for actual levels at a geographic subpopulation level (Wang, et al., 983). The data they used to re-weigh respondents for accuracy was U.S. Current Population Survey (CPS) and exit poll data from the 2008 presidential election. According to the question, Figure 5 in the referenced paper displays the representativeness of the Xbox data based on exit polls.

All of the data sources eventually were used by Wang et al. to determine the composition of respondents voting for whom in the 2012 presidential election. The question wants us to make the prediction 3 weeks out, which is 21 days off from November 5. This would be October 15. According to figure 2, the raw Xbox data predict that Obama would have had ~45% of two-party share. Pollster data and post-stratified Xbox data from Figure 3 predict that Obama would have had a little over 50% (maybe ~50.5%) and a little over ~51% share respectively. Thus, raw Xbox data would have predicted Mitt Romney would win, post-stratified Xbox data would have predicted Obama would win, and Pollster data would have predicted either uncertain, or that Obama would win but just by an incredibly narrow margin if decimal points were used. The limitation here is that estimates on one day are not a really good judge of the actual outcome. Wang et al. state that it doesn’t account for, “[1.] voter preferences on the particular day…. [2.] fading early lead bias… [the fact that] [3.] preferences expressed weeks or months before election day are at best a noisy indicator of individuals’ eventual votes… [and 4.] estimates of vote shares require a model of likely voters” (Wang et al., 987). Accounting for these differences in a new model may produce different results.

References

**Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman,** “Forecasting Elections with Non-Representative Polls,” *International Journal of Forecasting*, 2015, *31* (3), 980-991.

**Romrell, Danae**, “Gender and Gaming: A Literature Review,” *Idaho State University*, 2013.

1. <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF> [↑](#footnote-ref-1)
2. Wikipedia contributors. "Xbox." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 25 Oct. 2018. Web. 31 Oct. 2018. [↑](#footnote-ref-2)