Assignment 4

Ellen Hsieh

1. Non-probability sampling phone survey

(a) My area code for the survey is 214 (Dallas, Texas) and the spreadsheet (PhoneSurvey_EH.xlsx) is already pushed onto my repository.

(b) I have called 200 numbers and only two people responded to the survey questions. Beside, around 10 people picked up the phone but didn't answer the questions. Therefore, my response rate is 0.01.

(c) Both of my respondents answered the voting question and the age question.

(d) It was Saturday afternoon at Dallas, Texas, when I made those phone calls. Thus, no one picked up the phone if the numbers belong to companies. Also, for those who answered the phone, they seemed pretty occupied, for example, a lady who picked up the phone was at a funeral. Therefore, making phone calls on Saturday afternoon might be harder to get responses from people since they might want to enjoy their time instead of picking up some random calls.

(e) Since I only got two respondents, 56 and 88 years old, respectively, the median age of my respondents is 72 years old, which is the same as the average age. The reason why my sample median age might differ from the State data, which is 33.2 years old for Dallas, Texas, is that for those who are willing to answer those voting questions are people in middle age or older. Therefore, this survey has age bias, which collected more opinions from older people.

(f) 50 percent of my respondents voted Republican (Trump) in the 2016 U.S. Presidential election. None of the respondents voted Democrat (Clinton). The actual voting percentages in Texas is 52.6 percent for Trump and 43.4 percent for Clinton. Even though the sample size I have is really small, there are still 50 percent of respondents voted for trump, which matches the percentage of the vote for Republican. If the order would influence the results, then I would probably start with asking people if they have voted for the 2016 presidential election to make people feel more comfortable to answer the following questions such as which party or candidates they voted for, and then the question for the age.

2. Predicting elections survey, Wang, Rothschild, Goel, and Gelman (2015)

According to the Figure 1 in the article (Wei Wang, David Rothschild, Sharad Goel, Andrew Gelman, 2015, p982), we can tell that the most representative variables are race, state, and 2008 vote. On the contrary, the least representative variables from the Xbox sample are sex, age, and education. As for the reason why those variables are so different from the broader voting population is that the people who play Xbox are mostly male and younger such as students in high schools or college. However, from the exit poll, people voted tend to be in their middle age. Also, for people who spend more time on playing video games like Xbox might have lower education level, for example, those who addicted to video games have higher possibility to quit the school or choose to work right after graduating from high school or college.

In order to show the approach of using multilevel regression and poststratification (MRP) to predict the election, the authors use the raw Xbox data and 2008 exit polls to perform a poststratification reweighting process to transform the raw Xbox data by 2008 exit polls. Therefore, the outcome of non-presentative polls from Xbox can more accurately predict the outcome of 2012 U.S. presidential election.

During the last three weeks, the raw Xbox data have predicted that Romney would win as we can see from Figure 2 (Wei Wang, David Rothschild, Sharad Goel, Andrew Gelman, 2015, p982), that there is a decline for Obama support from around 50 percent to 45 percent. On the other hand, the prediction of Pollster.com shows that the intention of the voters seems uncertain because the percentage of two-party Obama support remains around 50. As for the Xbox poststratified data, in Figure 3 (Wei Wang, David Rothschild, Sharad Goel, Andrew Gelman, 2015, p984) we can see that after adjusting the raw data by MRP, the prediction implies the victory of Obama, which is also really close to the actual outcome, 53 percentage of support for Obama.

References

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