1. Non-probability sampling phone survey

(a) Please see the PhoneSurvey.xlsx

(b) I have called all of the 200 numbers. However, none of them responds. About 80 % numbers are invalid and for the rest 20%, they didn’t pick up my phone or just go to the mail box. Therefore, unfortunately, my respond rate is 0.

(c) Considering that no one answered my phone, I can’t answer this question.

(d) The area code Professor Evans designated to me is 305. It is the area code for all of Miami, Florida, Miami-Dade County, and the part of Monroe County in the Florida Keys in the United States. There is a one-hour time difference between Chicago and Miami. I called three times on Friday’s 11AM, 2PM and 5PM, respectively. Considering that I made my call on Friday’s working hours, maybe they are too busy to pick up my call. Therefore, the time I chose might contribute to my extremely low respond rate.

(e) I don’t have data to answer this question. However, just like what we have discussed during the class, the median age of respondents might be higher than the average age in that state. The reason for this phenomenon is that people who tend to pick up unknown phone numbers and chat with them are those who have plenty of free times, that is, old people. Therefore, the average age of respondents might be higher than the overall average

age.

(f) Still, I don’t have data to answer this question. Nevertheless, just like what I have mentioned before, there might be an age biased in the survey data. Therefore, the result might be different from the actual voting percentages.

As for the order of the candidates, considering that those interviewees might think that I have a preference of Clinton if I say her name before Trump, I would like to say Clinton’s name before Trump’s for half of the interviewees and Trump’s name before Clinton’s for the other half. Then, I will calculate the voting percentage of each half and compare those percentages to find if there are some differences between them to decide whether the order of the candidates indeed matters.

I would do the same thing to find how the order of different categories of survey questions might influences the result, like ask interviewee’s age first for half of the interviewees and ask interviewee’s choice of president first for the other half. However, considering that age is a relatively sensitive question, interviewee might think that you are offensive to reach their personal privacy; it might be a better choice to ask their choice of president first and then ask whether they are willing to tell you their age.

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

To begin with, I must admit that this paper is really interesting and inspiring. As a senior player, I have never thought about that we could conduct a serious survey such like the presidential election on a gaming platform. What's more, their result that 'with proper statistical adjustment, non-representative polls can be used to generate accurate election forecasts, and that this can often be achieved faster and at a lesser expense than traditional survey methods'[1] is also very exciting.

However, when the authors tried to 'compare the demographic composition of the Xbox participants to that of the general electorate'[2], we could see that there are still some differences in several variables. The three least representative variables are sex, age and education. For the sex variable, 'men make up 93% of the Xbox sample but only 47% of the electorate'[3]. As for the age variable, Xbox sample has more young people. For the education variable, Xbox sample has more people with low-education and less people with higher education. The main reason for these differences is that 'young men with relatively low education dominate the Xbox population'[4]. However, there are still some similarities between Xbox participants and 2012 Exit Poll, and the three most representative variables are race, state and 2008 Vote.

Considering that Xbox data is not representative, the authors choose to use the 'multilevel regression and post-stratification' method to 'transform the raw Xbox data into accurate estimates of voter intent in the general electorate.'[5] In this process, 'the core idea is to partition the population into cells based on combinations of various demographic and political attributes, use the sample to estimate the response variable

[1] Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman, Forecasting Elections with Non Representative Polls," International Journal of Forecasting, 2015, 31 (3), Abstract in Page 980.

[2] Same paper, 2. Xbox data in Page 981 and Fig.1 in Page 982

[3] Same paper, 2. Xbox data in Page 981

[4] Same paper, 2. Xbox data in Page 981

[5] Same paper, 3. Estimating voter intent with multilevel regression and poststratification in Page 981

within each cell, and finally aggregate the cell-level estimates up to a population-level estimate by weighting each cell by its relative proportion in the population.'[6] Finally, they decided to 'use exit poll data from the 2008 presidential election' to compute cell weights. Therefore, the authors used the Xbox raw (unweighted) data and the exit poll data from the 2008 presidential election in the post-stratification re-weighting process.

As for the prediction in the last three weeks of the election, we could see clearly that from Figure 2[7], the Xbox raw data predicated 'a landslide victory for Mitt Romney' because the two-party Obama support rate was below 50% in the last three weeks. As for the Pollster.com forecast data, the prediction was rather uncertain. The two-party Obama support rate was very close to 50% in the last three weeks. From Figure 3[8], the Xbox post-stratified data predicted that Obama would win because the two-party Obama support rate was higher than 50% in the last three weeks. From these different predictions, we could see that the post-stratified data's prediction 'is much more reasonable, and the voter intent in the last few days is close to the actual outcome.' Therefore, we could reach the aforementioned conclusion that 'the adjustment of non-representative polls is success.'

Reference

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

All the above answers are cited from this paper.

[6] Same paper, 3. Estimating voter intent with multilevel regression and poststratification in Page 982

[7] Same paper, Page 982

[8] Same paper, Page 984