Assign4_ANSWER

by Shuyan Huang

Non-probability sampling phone survey

(b)

I called 125 numbers. 9 people over 18 years old responded, and 116 numbers were not responded or the responder was below 18. The respond rate is 7.2%.

(c)

Among the 9 responders, 3/9 of them answered the age question and 3/9 of them answered the voting question.

(d)

I called during 6pm and 9pm, when most people were off work and not sleeping. This timing might have increased my respond rate, otherwise fewer people would respond if they were working or sleeping.

(e)

The median age of my respondents is 41, which is close to the median age in Pennsylvania, 40.4. However, a sample of 3 is too small to tell anything about the population.

(f)

1(33.3%) of my respondents voted Republican (Trump) in the 2016 U.S. Presidential election. And 1(33.3%) of my respondents voted Democrat (Clinton). 1(33.3%) of them did not vote. Comparing to the actual voting percentages in Pennsylvania from the 2016 election (48.18% Republican and 47.46% Democrat), Republican's two-party voting percentage is relatively lower in my sample, while Democrat's two-party voting percentage is relatively higher.

To test if the order in which I say the candidates or categories in the survey question influences the results, I need to randomly call another 3 people and ask the survey question in a different order. Then compare the voting fractions of the two samples and see if there's a difference. Repeating this experiment several times (or having larger samples) can give a more confident result.

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

(a)

Of the eight variables reported from the respondents, sex, age and education from the Xbox sample are the least representative of the data. While race, state and 2008 vote are the most representative. Sex and age are the least representative because "young men dominate the Xbox population: men make up 93% of the Xbox sample but only 47% of the electorate; 18- to 29-year-olds comprise 65% of the Xbox dataset, compared to 19% in the exit poll."(Wang W, Rothschild D, Goel S, et al. Forecasting elections with non-representative polls[J]. International Journal of Forecasting, 2015, 31(3): 981.) That there are fewer college graduates in the Xbox population suggests that Xbox users generally have a lower education level than the voting population. Possibly because many Xbox users are too young to graduate from college.

(b)

When the authors performed a post-stratification re-weighting of the respondents, they need to divide the population into cells according to all possible combinations of the variables in the Xbox survey, and calculate the weights of each cell by the proportion of the electorate in each cell. They used exit poll data from the 2008 presidential election to obtain the weights, and then applied the weights to the Xbox survey data.

(c)

According to Fig. 2, the Xbox raw data would have predicted that Romney would win during the last three weeks before 2012 U.S. Presidential election, since the two-party Obama support was significantly below 50% in Xbox raw data. Pollster.com forecast data would have predicted that the election outcome as uncertain during the last three weeks before election, since the two-party Obama support was very close to 50% in that period. According to Fig. 3, the Xbox post-stratified data would have predicted that Obama would win during the last three weeks before 2012 election, since the two-party Obama support was significantly above 50% in Xbox post-stratified data.