Assignment 4

Donghyun Kang

1. Non-probability sampling phone survey

(a) My *PhoneSurvey.xlsx* spreadsheet is now on my github.

(b) I made two hundred calls and found that all the numbers assigned to me were invalid numbers. Thus I did not have anyone responded and my response rate is either, I think, zero or not applicable in this case.

(c) I was not able to compute the fraction.

(d) My area was Ohio where the EST is used. I called from 4:30 pm to 6:00 pm in the central time, which corresponds to the range from 5:30 pm to 7:00 pm. However, as like I mentioned, all the phone numbers I had were invalid so it is hard for me to think of the role of the time of day in my response rate. My general sense is that relatively old people would participate the phone survey because people in other age groups might be working or returning to home.

(e) It is impossible to compute the median age of my respondents.

(f) The actual voting percentages for Trump and Hillary were 52.1% and 43.5%, respectively. However, it is again impossible to answer the question.

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

(a) The authors collected eight variables from the Xbox users: Sex, Race, Age, Education, State, Party ID, Ideology, and who they voted in 2008. Among them, Race, State, and Ideology appeared to be the most representative while Sex, Age, and Education seemed to be the least representative. As the authors reported, the majority of the respondents (93%) were young males who were less educated than general voting population, comparing to the 2012 exit poll (Wang et. al, 2015, p. 981). This might not be striking if we consider some level of general knowledge about the typical users of these gaming machines - again, young males whose ages are between 18 to 29. Then, we can raise question of why young males are gravitated to games when other ages and gender groups are not. Here I suggest some possible explanations: young males could satisfy their desire for competition (which might be attributable to the high level of testosterone) by playing those games; they maybe do not have to work hard to support their families whereas other people in ages groups; or perhaps older and educated people appreciate high-brow cultural products such as performing arts in their spare time while young less educated males did not have an opportunity to develop those kinds of tastes.

(b) The *"Current Population Survey" (CPS)* is often used perform poststratification but authors the exit polls from the 2008 presidential election (ibid., p. 984) because 1) the CPS data does not contain crucial variables such as party identification and 2) the 2008 exit pool can be used to predict the result of the 2012 presidential election. Then, the authors use *historical data from the previous presidential elections (2000, 2004, 2008)* to estimate the voters' intents because the election outcome is eventually determined by those who cast a ballot. (ibid., p. 987)

(c) According to Figure 2, the unweighted Xbox raw data would have predicted a victory for Republican candidate, Mitt Romney, during the last three weeks of the 2012 presidential election. And Pollster.com would have predicted that the outcome would be uncertain given that the proportion of the Obama supporters stayed around the 50% percent, even though there would have been slight increase of the supporting rates for Obama (ibid., p. 982) as time passed. In contrast, Figure 3 demonstrates that the adjusted Xbox data after post-stratification would have predicted the Obama's win even if the 95% confidence intervals are taken into consideration (ibid., p. 984).