MACSS 2018-2020 Perspectives on Computational Analysis

Assignment 4

a) Sheet provided in the A4 Folder in the Repository

b) I first ran an API-based script from www.numverify.com to check which of the randomly generated numbers were indeed valid (in this repository). I got 60 numbers from there, which can still be seen coloured green in the Excel. I automatically coded the response variable for the remaining 140 numbers to 0. I then called those 60 numbers. Among those 60 numbers, many turned out to be no longer in service. This status does not seem to get caught by numverify.com's algorithm

   Only 3 people responded according to the Response Variable. Accordingly, 197 did not respond. This sets the response rate to 1.5%.

   As an additional comment, if we were to reduce the denominator to the valid numbers only, the proportion rises to 3 of 60, or 5%. If we focus further on only those who picked up the call and were private individuals (not companies), the figure rises further to 3 of 5, or 60%.

c) 100%. All those who agreed to respond then provided answers to both the age and voting questions

d) My time of calling was only about 5:30-7:00 PM as per the local time (in Denver). I don't think this is a very late hour to call on a weekday. However, a large number of calls were either not picked up or went to answering machines (marked as 'Not Available' in the Excel Sheet). This figure stood at 32 of 60 (more than 50% of numbers). So calling at a different time- perhaps on the weekend or during the day- may have resulted in a much higher response rate.

e) The median age of my 3 respondents is 50 years. As per America Fact Finder, the median age for the state of Colorado is in fact 36.4 years. The most salient reason for the disparity of course, is the tiny sample size of 3 (for this study) vis-a-vis America Fact Finder's considerably larger sample. Purely by the laws of probability, aA random sample of 3 could have a median highly different from that of the entire population (or the relatively huge sample used for the state-level estimate).

   Apart from this, the data may not be entirely random. Those who answered the calls were older individuals- 48 years and above. They may have been retired, or with fewer time commitments than college youth, young professionals or young couples with families. This may have made them more willing to devote time to answering the call. In that sense, there exists a selection bias among the respondents. Other (younger) sub-populations I just mentioned would be under-represented. It should not come as a surprise then that the median age for the state lies considerably lower than that found by this study.

f) The figures in my data for Trump and Clinton were 0% and 33% respectively. The remaining 66% (2 respondents) said they voted for the Libertarian Party. This is an

intriguing anomaly. The actual voting percentages for these 3 parties in the 2016 Presidential Elections stood at 44.2%, 47.2% and 5% respectively.
To test if the order of listing the questions would make any difference to the answers provided, I would need a larger random sample (of valid numbers). The precise split of this sample would depend on how many different permutations we wanted to test. Let's assume there were 2. In the first one, we place 'Democrats' as the first option. In the second one, we do the same with 'Republicans'. We would assign half the sample to each group and make calls accordingly. We would then calculate the proportions in the two samples for one of the 2 parties (say Democrats), and check if the difference in proportions is statistically different from zero. This could be achieved through a two-sided t-Test.
The process would have to control for non-response rates in the two groups (which has clearly been a challenge in this assignment). As long as the rate is roughly the same in both groups, we can proceed to the next steps. We could also consider alternative means of collecting data with high response rates, and then stratify for under-represented sub-populations. This approach will be investigated in the next answer in the context of the use of X-Box data by Wang et al (2015).

2.
a) The statements in this section are based on the Fig 1 on Page 3 of Wang et al (2015). Of the eight variables, the three that are least representative of the larger population are:
i) sex
ii) age
iii) party ID.
The most representative three would be
i) State
ii) Race and
iii) 2008 Vote

The Xbox population, at a very intuitive level, would be drawn mostly from a younger, male population. However, for the sake of methodological rigour, it is important to lay down more precise arguments as to why this may be the case.

Differences in the representation of the sexes among X-Box users may be broadly explored from the individual effects and interplay of two key factors- sociological gender stereotypes and marketing efforts of game development companies. In either case, applying a historical lens to the industry provides key insights.

Lien (2015) describes how in their fledgling years of the industry in the 1970s, a number of gender-neutral arcade and computer games such as Pong, 3D Tic-Tac-Toe,

Checkers, Avalanche, Breakout and Centipede. With exceptions like Sierra Entertainment (which gained fame for its adventure game series), the composition of the industry was indeed male-dominated. However, games were being designed by men, and not necessarily for men.

The nascent industry boomed and then floundered in 1983-1985 with an almost 97% crash in revenues, largely due to saturation of the market with low quality games and limited regulation. Nintendo Entertainment System stepped in to restore consumer confidence through stricter regulations and much more concerted niche marketing efforts. Video games were strategically incorporated into 'toy culture' and made available through traditional toy retail channels.

Now treading carefully on limited budgets, Nintendo decided to zero in on a specific market. Publishers of the Nintendo Power magazine organized and inspected tournaments and market research initiatives to understand the user demographic. At these specific events, boys were in larger attendance than girls. Also, under the extant social norms, boys showed a higher likelihood of adopting and engaging with new forms of technology, with higher support from school and family to choose STEM careers. Since this juncture, through the 1990s, games (such as Gran Turismo and Tomb Raider) were developed and marketed specifically for male audiences between the ages of 10 and 15. By the time Xbox arrived under Microsoft in 2001, games began to be developed for these young adults to ensure that the gaming habit continued among its loyal male adherents.

The variable of age may relate to tech savviness in general. Pinchefsky (2013) describes initiatives in the last five years to connect X-Box with other use cases in the household and thus universalize their usage in the family. However, as highlighted in the explanation of the differences in sex, the target group has largely remained younger and narrower than the general population age range.

The party affiliation may itself be closely correlated with age, which Wang et al (2015) themselves concede in this paper hold a strong correlation with voting behaviour- through a reference to Kaufman and Petrocik (1999).

b) The authors use exit poll data from the 2008 presidential election along with data from an X-Box opt-in poll available in the 45 days run-up to the 2012 Presidential Election.

c) Predictions of each of the three data sources for the last 3 Weeks (Oct 15-Nov 05):

X-Box (uncorrected)- Victory for Mitt Romney with between 45-50% of two-party Obama vote share
Pollster- Victory for Obama with between (with slight fall in confidence after the first Presidential Debate)
X Box (Post-Stratified)- Victory for Obama- About 52%- which turned out to be the actual two-party Obama vote share.

BIBLIOGRAPHY

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

Kaufmann, K. M., & Petrocik, J. R. (1999). The changing politics of American men: understanding the sources of the gender gap. American Journal of Political Science, 43, 864–887.

'2016 Colorado Presidential Election Results'. Politico.com.

https://www.politico.com/2016-election/results/map/president/colorado/ (Accessed on October 29, 2018)

'Community Facts'. factfinder.census.gov

https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml (Accessed on October 29, 2018)

Lien, Tracy, 'No Girls Allowed'. Polygon.com

https://www.polygon.com/features/2013/12/2/5143856/no-girls-allowed (Accessed on October 30, 2018)

Pinchefsky, Sara 'The Xbox One Is Targeting The Largest Market Of All: Television Watchers', Forbes.com

https://www.forbes.com/sites/carolpinchefsky/2013/05/22/the-xbox-one-is-targeting-the-largest-market-of-all-television-watchers/#6b41836e6e69 (Accessed on October 30, 2018)