MACSS 2018-2020 Perspectives on Computational Analysis

Assignment 4

a) Excel 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, Colorado). 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- would most likely 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, appears to be the tiny sample size of 3 (for this study) with respect to America Fact Finder's considerably larger sample. Purely by the laws of probability, a 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. The sample from this study could thus hardly be called representative for Colorado.
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 higher 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 Xbox 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) Education

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 Xbox 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 the fledgling years of the video game industry in the 1970s, a number of successful arcade and computer games such as Pong, 3D Tic-Tac-Toe, Checkers, Avalanche, Breakout and Centipede were in fact gender neutral. With exceptions like Sierra Entertainment (which was run by a female founder), 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 and marketed at toy stores.

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. Jayath (2013) asserts that under the broader umbrella of smart phone games, female participation in gaming in fact outweighs men (52%). However, our analysis limits us to the Xbox and games specifically developed for this device.

The second variable of age may relate to the extent of being 'tech savvy' in general. Smith (2014) highlights how in 2012, technology adoption by senior citizens was correlated with income, with 87% of individuals with annual incomes over $75,000 using the internet. Pinchefsky (2013) describes initiatives in the last five years to connect Xbox 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.

Thirdly, the non-representative nature of the education variable stands out most prominently for the final category- that of college graduation. As per the US Census Bureau (2014), the percentage of citizens with 'some' college (64%) falls dramatically for those with a completed Bachelor's Degree (34%). The College Atlas (2014) further points out a 56% dropout rate within 6 years of starting a 4-year degree, with 20% higher dropout rates for male students (with respect to females) and 55% lower for those enrolled full-time. My hypothesis is that a large percentage of Xbox users who answered the poll may have belonged to this category of male college dropouts. However, more research will be needed to corroborate these findings.

As a closing comment, it must be noted that Xbox's 2018 internal market research (Corden, 2018) states that the percentage of female users stands at 42% and that the majority of gamers lie in the 25-34 age bracket. Thus, if a similar study were to be conducted using data from this updated demographic, less post-stratification would be needed for creating a representative sample than was needed in 2012.

b) The authors use exit poll data from the 2008 presidential election along with data from an Xbox 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):

- Xbox (uncorrected): Victory for Mitt Romney- with between 45-50% of two-party Obama vote share
- Pollster: Victory for Obama- but very tightly contested- two-party Obama vote share at or within a small range of 50%. If taken in isolation, the second of these 3 weeks would have indicated victory for Romney, with the Obama vote share staying below 50%.
- Xbox (Post-Stratified): Victory for Obama- About 52% two-party Obama vote share - which turned out to be the approximate share in actual election results.

BIBLIOGRAPHY

'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)

Corden, Jez, '*Internal Microsoft research provides insight on Xbox One owners (exclusive)*', windowscentral.com

https://www.windowscentral.com/heres-some-interesting-stats-about-xbox-one-owners-microsoft-shared-partners (Accessed on October 30, 2018)

Jayanth, Meg, '*52% of gamers are women – but the industry doesn't know it'*. Guardian.com

https://www.theguardian.com/commentisfree/2014/sep/18/52-percent-people-playing-games-women-industry-doesnt-know (Accessed on October 31, 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)

Adults' Media Use and Attitudes: Report 2016. Ofcom. https://www.ofcom.org.uk/_data/assets/pdf_file/0026/80828/2016-adults-media-use-and-attitudes.pdf (Accessed on October 31, 2018)

Smith, A. Older Adults and Technology Use. *Report from the Pew Research Center Internet & Technology, Apr. 3, 2014;* http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ (Accessed on October 31, 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)

The College Atlas. 'U.S. College Dropout Rate and Dropout Statistics'. Collegeatlas.org

https://www.collegeatlas.org/college-dropout.html (Accessed on Oct 31, 2018)

US Census Bureau. 'Educational Attainment in the United States: 2014', census.gov

https://www.census.gov/data/tables/2014/demo/educational-attainment/cps-detailed-tables.html (Accessed on Oct 31, 2018)

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