**Writing Assignment 1**

**TOTAL POINTS** 100

Instructions

1.

Question 1

**Briefly summarize the main findings of the article. In other words, what’s the main take away to remember from this article? (200 words max).** [20 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

20 points

The purpose of this study was to determine the reasons why individuals flagged as binge-watchers of tv shows employ this level of media consumption. Using a principal component analysis, it was found that factors such as enjoyment, efficiency, recommendation of others, perceived control, and fandom were major drivers, accounting for almost 70% of the variance. In addition to identifying binge-watching drivers, this study also explored the relation of two populations to this type of behavior, individuals that are considered high sensation seekers as well as individuals considered high in need for cognition. The results of this study made clear that both high sensation seekers and individuals high in need for cognition are more likely to engage in this binge-watching practice than those lower on their respective scales. As a result, this study was very effective in not only empirically evaluating factors that influence a certain type of behavior, but was also able to assess that individuals with certain characteristics and drivers are more likely to not just participate in a certain practice, but have the behavior be even more pronounced.

2.

Question 2

**What is the purpose of the article? Is the article written to prove something (empirical research), persuade the reader, or describe a phenomenon? How do you know?** [10 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10 points

The main purpose of this article is to empirically prove/discern the major drivers of binge-watching and whether certain populations of individuals (high sensation seekers or high in need of cognition) are more likely to engage in this behavior. However, in doing so, the authors, in their discussion section, also attempt to persuade the reader, as well. Using the empirical research, the goal of the authors is to utilize this information to craft an argument and “convince” the reader that the methods and approach were valid, and that the behavior can truly be explained by the empirical research performed. I believe most scientific articles will always be a combination of empirical research and persuasion. Additionally, to a much lesser extent, this article did begin briefly with a description of the binge-watching phenomenon. The emergence of streaming services such as Netflix and Hulu were detailed as well as the accompanying emergence of the binge-watching practice. However, this was mainly done to set the stage for the empirical breakdown behind binge-watching. In conclusion, the major goal of this article, while persuasive and descriptive, was empirical and utilized PCA and regression in order to mathematically discern the motivations behind binge-watching.

3.

Question 3

**What concepts from this course were used in the article? Describe how they were used (200 words max per concept)** [20 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

20 points

The first concept from the course that was utilized in this article was principal component analysis. This was performed in order to understand the underlying reasons why individuals binge-watch tv shows. Using 19 statements to gauge an individual’s reason for this type of behavior, PCA was conducted to discover the most motivating factors, which were found to be enjoyment, efficiency, recommendation of others, perceived control, and fandom. By taking the eigenvalues of these principal components, it was also found that the proportion of variance explained by the five was 69.61%. PCA, and the proportion of variance explained by the top 5 principal components, was therefore the main mechanism the authors used to empirical determine the motivations behind binge-watching tv shows.

In addition to PCA, the other concept that was discussed in this article that has been covered in the course so far was regression analysis. Regression analysis was performed to determine whether any relationships existed between various individual qualities and binge-watching behavior. Mainly, one of the other major findings of the article was discerned through this analysis, as the authors, through regression analysis, were able to empirically determine that high sensation seeking and high need for cognition individuals were more likely to engage in tv binge-watching. Additionally, these types of analyses also made clear that these high need and high sensation seeking individuals were influenced more greatly by the motivations (principal components) noted above. Therefore, PCA and regression analysis were the two major topics from the course discussed and were major factors in the generation of this articles’ results.

4.

Question 4

**What were the strengths of the methods? (Why were methods appropriate or the results convincing?)** [15 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

15 points

The strength of the methods for this article is the detail the authors provided in terms of how the populations were identified. When attempting to empirically determine the motivations behind binge-watching, the authors relied on a focus group that provided 19 statements that would eventually be used to generate the principal components from the PCA. Utilizing a focus group likely allowed for statements to be generated that more encompassed the population as a whole as opposed to relying on the author’s own views, which may have been a limited view of why an individual may engage in this type of behavior. Additionally, the authors also made clear that the determinations used to identify the sensation-seeking and high cognition populations were based off of prior articles. This type of approach is very valuable in that these determinations are in alignment and allow for comparisons to be made with prior work. Additionally, the authors provide p-values and standard errors alongside all of their results, which I believe is a strength and very valuable in proving that the results are “true” and should be convincing. Lastly, the analyses used (PCA and regression analysis) are both powerful and very effective tools in discerning relationships between various variables/indicators and would lead to convincing results.

5.

Question 5

**What were the weaknesses of the methods? (Was there anything missing in the article? What would you have liked to see, but did not see?)** [15 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

15 points

While the article did make note that various demographic variables were accounted and controlled for (age, gender, education level, etc.) I would have liked to have seen the manner in which these were controlled for and how these demographic variables may have even played a role in the propensity to binge-watch, as well. Additionally, I would have also liked to have seen how the data was prepared for the PCA, as well, in terms of dimensionality reduction and standardization. The standardization and preparation of the data for the most part took a back seat in this article (understandably but as someone newly becoming oriented to this type of work, would have been very variable to see these types of approaches and tools put into action).

6.

Question 6

**What was something you did not understand clearly from the article?** [10 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10 points

With this being my first introduction into PCA I had difficulty wrapping my head around Table 2, which shows the five principal components and their relationships to one of the 19 binge-watching statements. I am still struggling to wrap my head around how this table would be generated from PCA and mainly just around PCA in general, as well, in terms of establishing covariance relationships, etc., but this article was valuable in identifying this deficiency and I am hoping to pull in more external resources moving forward to fill in some of the gaps I currently have in stats and linear algebra. I do understand that this table is displaying covariance and is a valuable visualization in one these principal components were selected and that each have different levels of covariance with each of the binge-watching prompts.

7.

Question 7

**What suggestions are being made about future research into this topic/phenomenon or what does the researcher consider to be 'the way forward'? Are there mathematical methods that would help?** [10 points]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10 points

The authors noted a couple of paths forward for this research. For this article, South Korean binge-watching of tv dramas was what was investigated. However, the authors did pose questions around what may be seen if different countries are evaluated and compared to each other and whether different genres, such as comedies and reality tv shows, are evaluated and compared to each other. This was one of the paths forward posed by the authors and relationships between countries and various dramas could easily be evaluated through the regression analysis methods seen in this article, as well. Additionally, the authors did not that another path forward would be to perhaps create some granularity in the “high sensation seeking” and “high cognition” populations, to perhaps see if there are any other breakouts within these buckets that may also have interesting relationships with binge-watching motivations. Again, PCA and regression analysis could be employed to handle this future research.

**Writing Assignment 2**

**TOTAL POINTS** 100

**Briefly summarize the main findings of the article. In other words, what’s the main takeaway to remember from this article? (200 words max).** [20 points]

The purpose of this article was to estimate the number of individuals who casted a double vote in the 2012 presidential election. The authors pursued this study as there often is discussion around the trade offs that may exist between improving voter accessibility and avoiding or minimizing election fraud. Public perception would play a large role in policy related to these two issues, so the authors of the article aimed to generate a model that would create an estimate of this type of fraud to better frame this discussion empirically and determine whether this is a widespread issue. The authors, through their approach and model, found that, in the 2012 election, 1 in 4,000 voters may have double-cast a ballot. However, the authors note further, that despite this number, clerical errors, for example, could be a major, if not the total, driver for this finding and that actions that seek to stifle this behavior may, in fact, lead to more valid votes being thrown away in the process. As a result, the authors’ model made clear that this type of fraud is not prevalent and not a threat to United States elections.

**What is the purpose of the article? Is the article written to prove something (empirical research), persuade the reader, or describe a phenomenon? How do you know?**

**The purpose of this article was to prove something – the authors generated a model from multiple sources of demographic and voting data, and, through the model they created, empirically provided/estimated that only 1 in 4,000 voters cast a double ballot during the 2012 presidential election. Therefore, through their model, the authors empirically proved that election fraud, specifically double vote casting, is not prevalent in United States elections and is not a current threat to the integrity of elections. Interestingly, in this article, while the major goal was to empirical prove this major finding, the authors framed a lot of their introduction and discussion around what they described as the “current ‘voting wars’”. As a result, I also believe this article and generated model were a means of persuading the reader, as well, to the fact that this type of behavior is not prevalent and should not hinder steps being taken toward voting accessibility. Therefore, I believe the article both proved something as shown above but also aimed to persuade the reader to favor a policy argument centered around accessibility as opposed to combatting a threat proven by the authors to be minimal/non-existent.**

**What concepts from this course were used in the article? Describe how they were used by giving a summary of the method. (200 words max per concept)**

**One concept that was discussed in this course was confidence intervals, and the authors utilized these to make determinations with varying levels of certainty (95%, for example) that the “true” value lies within the interval range. In regards to the paper itself, the authors put confidence intervals on the doubling voting estimate, providing a range the authors could confidently say contained the true number of voters casting two ballots. This is a common practice and is a way to easily capture the variance tied to the estimate as well, with smaller confidence intervals corresponding to lower levels of variance.**

**The other concept in this article that has been discussed in our course was the few cases where the authors calculated probabilities for various scenarios within their data. For instance, the authors calculated the probability that two voting records with the same name and date of birth belong to the same voter. This particular scenario also ties in well with the concept of a p-value. The authors found that this particular probability was less than 0.05, which is often used as a threshold when rejecting a null hypothesis as the finding had a <5% probability of being due to chance. The authors calculated probabilities around combinations of first and last names having the same birthdays based around the seasonality and historical precedent for various names being chosen for newborns. In regards to voting specifically, the authors also calculated probabilities around the probability of registration being duplicated and the duplicated record being removed, as well as the probability of false positives and negatives when comparing registrations to polling records. Probability play a major role in this study and the author’s eventual development of their model.**

Question 4

**What were the strengths of the methods? (Why were methods appropriate or the results convincing?)**

The strength I believe of this article and study is how thorough the authors were in ensuring a clean dataset. This issue appears frequently in my own day to day as data cleansing is such an important but often overlooked portion of any study or job and can greatly influence a finding. Removing votes where there was not a clean first name, last name, and date of birth associated with it, removing votes with birthdays at the start of the month as these often serve as placeholders and may have skewed results, and removing votes from states where poor data collection is apparent were all extremely valuable steps to take to ensure accurate findings. The authors still, in their study, then took these removals into consideration to ensure that their end model would be indicative of the larger national dataset. Additionally, the authors, building off of a prior study, and accounting for greater frequencies of certain names during particular parts of the year, utilizing confidence intervals when reporting their estimate for double voting in the 2012 presidential election, accounting for human clerical error in the registration and polling data, and looking nationally as opposed to one single state all contribute I believe to a very comprehensive study.

**What were the weaknesses of the methods? (Was there anything missing in the article? What would you have liked to see, but did not see? What is a case you can imagine affects the results that they did not account for?)**

**As noted above in the strengths section, I believe that the authors were extremely thorough and comprehensive in both cleaning their datasets as well as accounting for various errors or phenomenon (newborn naming trends, for example) that could be driving a perhaps even higher than actual estimate of the number of voters who cast more than one ballot during the 2012 election. However, while the authors focused heavily on clerical errors perhaps accounting for most if not all of the double votes, the authors did not focus much at all on the possibility of legal double voting contributing to the estimate, and I believe that this is a weakness of the article and its discussion of the results. As seen most notably in the most recent 2020 election cycle, there is the possibility for individuals to vote by absentee but also in-person (perhaps out of fear of the absentee not being received in time, an issue being found on the absentee, etc.). As a result, in this article I would have been curious to see how these types of situations would have been handled by the authors and how much of a role it may play in driving the estimate produced by the authors.**

**What was something you did not understand clearly from the article? If you perfectly understood everything, explain Equation 2.**

**While I understood the need and purpose for this step of the analysis, I had a hard time following the process around the birthday distribution. The authors were expanding on a prior study and article and wanted to take into account the likelihood that a higher percent of the population will be named Autumn, for example, between September and November, but I mainly am having trouble following how this work will impact equation 2. I believe this probability would behave essentially as a sliding scale and will have taken into account seasonal and historical naming data, but its impact is not immediately tangible to me after the reading. While this was a difficult portion of the paper to follow, I believe it was another great example in ensuring that your data is as normalized or uniformly distributed as possible. Through this approach, I believe the authors were attempting to uniformly distribute an event (naming) that is not uniform and can change depending on the year and even the time of year itself. I believe my follow up for this week is the same as the one from last week, to better familiarize myself with how samples like the birthday distribution can be “converted” from something perhaps random to something more uniform and more easily pulled in for analyses. \**

Question 7

**What suggestions are being made about future research into this topic/phenomenon or what does the researcher consider to be 'the way forward'? (Are there mathematical methods that would help?**

I believe, in this particular topic of election fraud, the authors were extremely thorough and built upon already existing work within the field, and I had some difficulty identifying potential future areas of focus or methods to improve upon or enhance. However, the authors in their discussion did note that work and improvements can really be made policy-wise and in the court of public perception. The authors determined that efforts to try to prevent election fraud would likely lead to more valid votes being removed per every invalid vote removed, which is a clear call for future policy around voter accessibility as opposed to trying to avoid a fraud that does not appear prominently in the data itself. Further, the authors noted that significant progress could be made in convincing the public that fraud is not prominent by highlighting that “fraud”, in actuality, are cases where clerical errors were made, for example, or can easily be explained through registration and polling record logging discrepancies. Utilizing this type of estimate (that 1 in 4,000 voters cast more than one ballot) that clearly displays double voting as a rare event and presenting the cases where there were double votes as human error driven, public perception could be altered significantly and beneficially.