**Instructions**

Close

**Overview**

**OVERVIEW**

You will read a peer-reviewed journal article and reflect on the content as it relates to math and data science. The goal of this assignment is for you to relate concepts learned in this course to real-life applications in the field of data science.

You will read and review **ONE** of the following peer-reviewed articles:

[An exploration of the motivations for binge-watching and the role of individual differences](https://search.lib.umich.edu/articles/record/cdi_crossref_primary_10_1016_j_chb_2017_12_032)

[One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections](https://search.lib.umich.edu/articles/record/cdi_crossref_primary_10_1017_S000305541900087X?query=One+Person%2C+One+Vote%3A+Estimating+the+Prevalence+of+Double+Voting+in+U.S.+Presidential+Elections&utm_source=lib-home)

You will be asked to respond to the following questions.

**Summarize the main empirical results of the article. In other words, what are the main takeaways to remember from this article? (200 words max).** [20 points]

Building up on and adding to the work of other papers published on this subject before, the article explores the motivations behind binge-watching drama TV series and the moderating effects of sensation seeking and need for cognition on binge-watching behavior in a study utilizing online survey data collected from 785 South Korean binge-watchers about their habits and preferences relating to binge-watching and other potential motivators for it.

Principal Component Analysis (PCA) of this data demonstrated that their main motivations for binge-watching stem from enjoyment, efficiency, recommendation of others, perceived control, and fandom. Moreover, regression analyses revealed that psychological traits such as sensation seeking and need for cognition have a positive effect on binge-watching behavior. The presence of these traits in individuals also boosted the effects of certain binge-watching motivations such as enjoyment, recommendation of others, and fandom.

Future studies would likely focus on incorporating multidimensional measures to more precisely account for the role played by the psychological traits under study, and also to expand both the genres of TV series as well as the pool of survey respondents to include more diversity, so that the results can be more generalizable.

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

The article is primarily written to present empirical research.

I surmised this to be the case as the study puts forth certain hypotheses about the reasons and motivations of individuals that are likely to engage in binging through media consumption in the Theoretical Framework section. The article then goes on to thoroughly address these hypotheses by means of empirical evidence, such as the findings obtained from an online survey of 785 South Korean binge-watchers of drama series about their binge-watching habits, experiences and other traits as put forth by the hypotheses, the experiential aspect of which is the essence of an empirical study.

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

Principal Component Analysis (PCA): PCA (analogous to the concepts of Eigenvalues and Eigenvectors from the course) was used on 19 focus-group interview statements in order to identify common underlying structures during the assessment of motivations of binge-watching among 23 binge-watchers. Five motivational factors (enjoyment, efficiency, recommendation of others, perceived control, fandom) were identified as the principal components for identifying motivations for binge-watching. The principal components so chosen were retained based on the proviso that Eigenvalues are greater than 1.00 and factor loadings are greater than 0.60 (i.e. conforming to Kaiser’s criteria) using an un-orthogonal Promax rotation.

Ordinary Least Squares Regression: Regression analyses (analogous to the concept of Linear Regression from the course) were instrumental in measuring relationships among the measured variables used in the study, which included demographic data, media consumption habits, motivations for binge-watching, psychological traits such as sensation seeking and need for cognition, and interaction effects between motivations for binge-watching and the psychological traits. These were organized under 4 different blocks for the purpose of the analyses. Regression analyses was helpful in supporting the hypothesis that there is a positive relationship between psychological traits and binge-watching motivators. Other hypotheses – such as the effects of binge-watching being more prominent in high sensation seekers than low sensation seekers, the establishment of need for cognition being a significant predictor of binge-watching, and the effects of binge-watching motivators on binge-watching behavior being more pronounced in those with higher need of cognition than those with lower need of cognition – were also supported by these analyses.

**What were the strengths of the statistical methods? Why were methods appropriate or the results convincing? What were the strengths of the approach to collecting the data?** [15 points]

For the purpose of this study, PCA seems to be fairly appropriate as it was instrumental to identify the common underlying structures of the motivators of binge-watching. Without PCA, we would just have 19 statements with no way to make sense of them from a statistical standpoint and would’ve been unable to proceed further in our quest to gain valuable insights about binge-watchers. The checking for the presence of linear correlations using Bartlett’s test and sampling adequacy using Kaiser’s criteria before picking the principal components was also the right move.

The choice of variables picked for the study as well as the questions (statements) for the respondents were also framed in such a manner so as to keep the findings of past studies on the subject in mind, so that this study was essentially contributing to the findings of those studies.

Moreover, the regression analyses were an obvious and perfect choice for determining the potential relationships among the chosen variables as the results so obtained support all the hypotheses put forth before the analyses.

The choice of South Korea to obtain the survey data required for the study from is a good choice since South Korea has a booming Korean TV Drama (K-Drama) scene with significant media consumption and engagement, ensuring a healthy sample size of binge-watchers for the analysis. Furthermore, the meticulousness involved in excluding incomplete and/or abnormal responses from the survey enables accurate conclusions being drawn about binge-watchers from this region.

**What were the weaknesses of the statistical methods? Was there anything missing in the article? What would you have liked to see but did not see? What were the weaknesses of the approach to collecting the data?** [15 points]

Since PCA is a linear combination of the variables in the study, there is a chance of information loss if the principal components are not chosen with caution. Outliers also tend to mess with obtaining the right principal components for the analysis, as well as with least squares regression.

I would’ve liked to see some more visualizations in the article, such as box plots for the various data points instead of just the tables. This would’ve helped me get a better sense of the spread of the data, especially with regards to outliers in order to determine whether the statistical methods chosen were appropriate for the data. Moreover, the psychological traits under study showed moderating effects only for 3 out of 5 principal components, suggesting that for this case, the measures to capture the multidimensionality of these traits were lacking.

Most of the weaknesses I found in data collection were already acknowledged by the authors of the article, like the fact that the survey data was collected only from South Korea, which would indicate that the findings of this study are not generalizable to binge-watchers from other countries. Moreover, the data in this study included binge-watchers of drama series, whereas there is evidence from another study to suggest that nearly 25% of binge-watching is done with other genres like comedy and reality TV. Therefore, the motivations for binge-watching used in this study may not hold true for binge-watchers of other genres.

**What was something you did not understand clearly from the article? Research it (for 10 minutes) and explain what you understand about it and what you still don't understand.** [10 points]

By just reading the article, I initially did not understand what Kaiser-Meyer-Olkin (KMO) statistics and Bartlett's test of sphericity were or why they were used.

After a bit of research, I understand that the KMO test is used to measure sample adequacy for factor analysis based on the proportion of variance among variables. A value of KMO statistic between 0.8 and 1 (this study has a value of 0.89) indicates that the sample is adequate. Bartlett’s test of sphericity compares the observed correlation matrix to an identity matrix. It checks for whether the variables in the study are unrelated (which would be the case if these matrices are equal). This particular study passed Bartlett’s test of sphericity, which found that the correlations among variables were suitable for the PCA.

What I still can’t seem to wrap my head around is the term of “reliability” with respect to Cronbach’s alpha (mentioned in section 4.1 of the article where it talks about the five binge-watching motivators exhibiting strong reliability).

**What future research do the authors consider to be 'the way forward'? What future research do you consider to be 'the way forward'? Are there mathematical methods that would help?** [10 points]

The authors intend to have future studies on the subject focus on adopting multidimensional measures to account for a more precise investigation of the roles played by the psychological traits sensation seeking and need for cognition in binge-watching. They are also of the opinion of incorporating the potential moderating role of other series genres (such as comedy and reality TV) and to have more diversity and international representation in their survey data to ensure that any findings gleaned from it would be generalizable.

As the “way forward”, I agree with everything the authors said above. In addition to that, I would consider adopting t-distributed Stochastic Neighborhood Embedding (t-SNE) in lieu of PCA for this analysis as it is widely considered to be one of the best dimensionality reduction techniques, and it is not as sensitive to outliers as PCA is.

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

**Review Criteria**

The response will be graded using the following rubric:

**Summary**:

20: Excellent and concise summary that shows great understanding of the take-a-ways and important points of the article.

18: Findings are well summarized and take-aways are apparent.

14: Findings are summarized and take-aways are apparent, but some minor parts are missing or not completely accurate.

10: Findings and take aways are incomplete or inaccurate.

6: Student’s findings and take aways are missing critical pieces of summarization.

0: Little to no effort was made

**Purpose:**

10: Clearly states purpose of the article AND persuasively describes how they know what the purpose is.

9: States purpose of the article AND describes how they know what the purpose is.

7: States purpose of the article, but it may be unclear or missing how they know it is the purpose.

5: Missing import parts of the article’s purpose, or poorly justified

3: Article’s purpose is incorrectly identified.

0: Little to no effort was made

**Concepts:**

20: Expertly relates the readings to material in the course. Describes how each concept was used in the context of the paper.

18: Identifies course concepts used in article AND describes how they were used.

14: Identifies course concepts used in article AND describes how they were used. Some minor errors or lack of clarity is present.

10: Identifies some concepts from the course in the article.

6: Identifies a minority of concepts from the course in the article.

0: Little to no effort was made

**Strengths:**

15: Expertly describes why methods were appropriate and convincingly support the conclusions.

13.5: Describes why methods were appropriate and convincing.

10.5: Describes why some methods were appropriate but lacks clarity or completeness.

7.5: Describes why some methods were appropriate but lacks is missing important details or contains significant errors.

4.5: Largely fails to describe why the employed methods were useful.

0: Little to no effort was made

**Limitations:**

15: Expertly describes limitations of methods employed, holes in what analysis might have been better, or misalignments between the methods and the conclusions drawn.

13.5: Describes limitations of methods employed, holes in what analysis might have been better, or misalignments between the methods and the conclusions drawn.

10.5: Lacks clarity or completeness.

7.5: Is missing important details or contains significant errors.

4.5: Largely fails to identify any weaknesses of the article.

0: Little to no effort was made

**Confusion:**

10: Articulates a concept that was not fully explained (at the level of the reader) and what knowledge or details are needed to make sense of the concept in the context of the article.

9: Describes a confusing concept and why it seems important to understanding the article.

7: Answer lacks clarity or does not explain why confusion is important to understanding the article.

5: Hard to pin point what the confusion is.

3: Largely fails to identify any confusion of the article.

0: Little to no effort was made

**Future Work**:

10: Innovative ideas for future work are described and their potential importance is identified.

9: Thoroughly describes the suggestions about future research or describes ‘the way forward’ and argues why it would be interesting.

7: Lacks clarity or contains minor misunderstandings.

5: Lists a few ways forward, but they are not very interesting and it is not argued why they would be important steps.

3: Fails to identify an interesting way forward.

0: Little to no effort was made