BRSM Project Report

TEAM CAR

Chayan Kochar(me) 2019114008
Debashish Roy
Aashwin Vaish

Original paper: Anxiety increases information-seeking in response to large changes, Charpentier et al, May 2022
Link: https://www.nature.com/articles/s41598-022-10813-9
Link to their github and data/source code provided:
https://github.com/affective-brain-lab/Anxiety-increases-information-seeking-in-response-to-large-changes-
Link to my code:
https://drive.google.com/file/d/1gfVSVY0lXOsdxKAeQNOOIGKIARIgwYLl/view?usp=sharing

Introduction

An adaptive reaction to anxiety can be gathering information. This is due to the fact that information can boost one's sense of control, lessen an unsettling sensation of ambiguity, and assist in directing actions[1][2]. However, if the information is unfavourable or unclear, it may exacerbate anxiety[3][4]. As a result, the knowledge people seek out when they are nervous may be for their benefit or harm. Here, we investigate if and to what extent anxiety affects the characteristics of information people seek.

Three ideas on how anxiety affects information seeking were put to the test.
First, we investigated whether a rise in the general frequency of information seeking is associated with anxiety. Regardless of the precise reason of the anxiety, increased information-seeking may be a coping tactic because knowledge can increase a person's sense of control, which is diminished in anxiety[5].
The second test was divided into two, to get an idea about how and if Trait and Induced anxiety play a role with greater information-seeking in response to large magnitude and/or valence of changes.

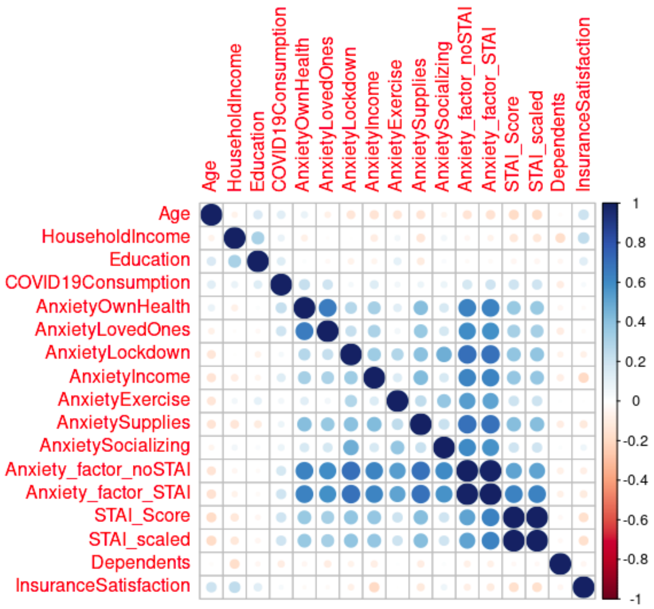
Methods

Analysis done by me

The work done by me in this project was not done by the original authors, hence it was something which we decided to implement on our own, to get a better understanding of the data.

I contributed to this team project by focusing on the first study and also helping in understanding Linear Mixed Models as well.
I was mainly undertaking the area of analysing how the different categorical variables behave and their effects on some
Though later, since the methods were the same, I extended the work to the data of study 2 as well.

First I did a low cut analysis on how the info seeking varies with respect to information seeking as well as other variables. This was done by visualising correlation heatmap as shown:



Here, I removed the categorical variables to remove any discrepancies.

Study 1:

To compare the categorical variables and get an understanding how they behave and affect anxiety scores.

Main methodology used: **ANOVA(Multiple One way ANOVAs and Factorial ANOVA)**

Testing of Assumptions:

- Normality Testing

    - QQ plots
    - Shapiro test
- Homogeneity
    - Boxplots
    - Levene Test
- Independence: From the nature of the data as was mentioned according to the paper, the data points were independent.

Study 2:

Same assumptions as above were tested. We will see that since these assumptions were not met, I proceeded with **Wilcoxon-Mann-Whitney Test** instead of regular ANOVA.

Analysis done by my Teammate:

He was mainly doing the same analysis as me as described above, but for numerical variables, Thus, the work mainly revolved around Regression and bettering the models using the VIF and heteroscedasticity concepts. The authors did do Regression and mixed models, but the concepts of VIF and heteroscedasticity was something that we thought would be better.

Results

Work done by me:

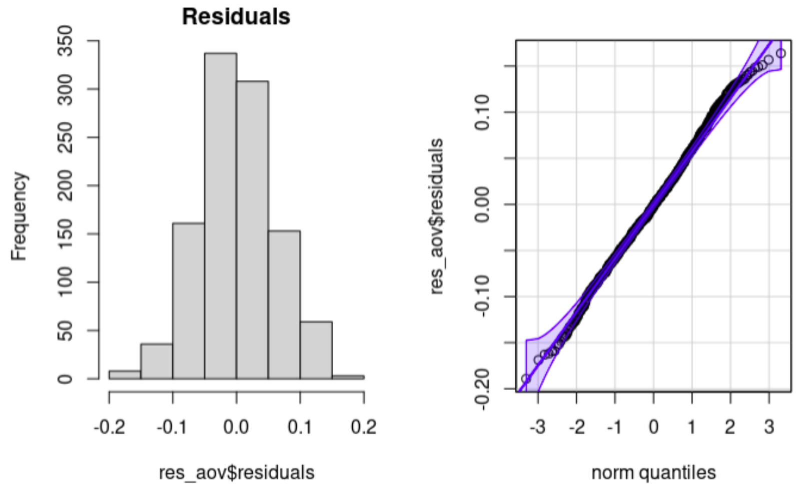
Testing for assumptions before applying ANOVA:

- **Normality testing**

    - **Data points**
      The data already had >1000 participants, hence consequently, there was no much need to check for normality. Still to check further, I went ahead with visualising QQ plots along with the Shapiro-Wilk Test.
    - **QQ plots and histogram, Shapiro Wilk Test :**
      When checking for normality, we want to check for normality for each group. I you test the normality assumption on the raw data, it must be tested for each group separately as the ANOVA requires normality in each group. This is why to make work easier and to understand better, I went ahead by checking for normality on the *residuals.*

Testing normality on all residuals or on the observations per group is equivalent, and will give similar results. Indeed, saying “The distribution of Y within each group is normally distributed” is the same as saying “The residuals are normally distributed”.

From the graphs, data appeared to be slightly skewed for normalization. We will discuss how this is not of much importance, but to be sure, I did some transformation such that after that transformation, it was satisfied via both methods.



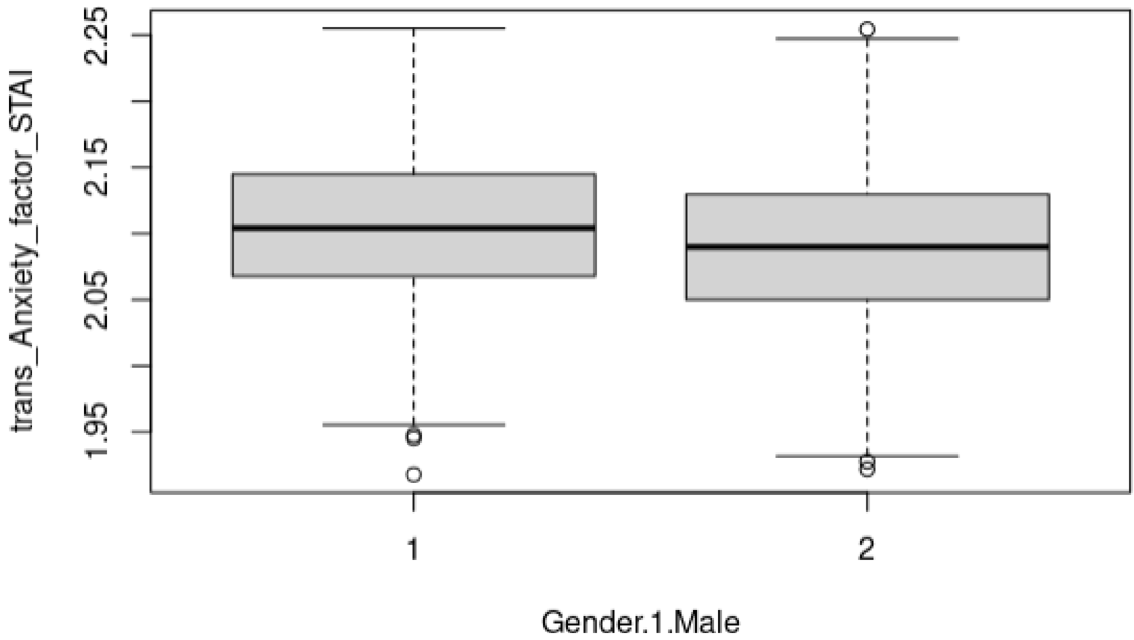
The Shapiro Wilk Test returned a p-value = 0.09291.
Since this is >0.05, we can say that the null hypothesis for Shapiro Wilk is satisfied, and the residuals are normal.

Since concerns of normality are not much alarming when samples >50, we avoid the hassle of transformations as it deteriorates the result and analysis of further tests. For analyses like the F or t family of tests (i.e., independent and dependent sample t-tests, ANOVAs, MANOVAs, and regressions), violations of normality are not usually a death sentence for validity. As long as the sample size exceeds 30 (even better if it is greater than 50), there is not usually too much of an impact to validity from non-normal data; something that Stevens stressed in his 2016 publication of Applied Multivariate Statistics for the Social Sciences.

- **Homogeneity Testing**
  The assumption of homogeneity is important for ANOVA testing and in regression models. In ANOVA, when homogeneity of variance is violated

there is a greater probability of falsely rejecting the null hypothesis.
Thus to do this, I used boxplots visualization and Leven Test.\

- **Boxplots:**
  From the boxplots, we can observe that there is not much difference in the variance between the levels of a group, hence satisfying homogeneity. Attached below is the boxplot for the two levels of gender. The corresponding Levene Test is given below in the Levene Test section.



- **Levene Test**

```
Levene's Test for Homogeneity of Variance (center = median)
        Df F value Pr(>F)
group    1  1.1027 0.2939
      1063
```

  Here, since the group(Gender) has only two levels, the DF here is k-1=1. The 1063 represents the df for all other residuals (N-k = 1065-2)

  Since the p-value is >0.05, here being 0.2939, the null hypothesis for Levene's Test is satisfied, and thus we conclude that the data is indeed homogeneous.

➔ **Multiple One-Way ANOVAs**

Since all the assumptions were satisfied, I then proceeded with applying One-way ANOVAs, to check how the anxiety scores vary with Gender, Ethnicity,Political Orientation and Whether they have dependents or not.

Here,
NULL Hypothesis: No difference in group means
Alternate: At least one group differs significantly from overall mean of dependent variable(Anxiety score)
IVs:Gender, Political Orientation, Whether they have dependents, Ethnicity

On doing such ANOVAs, I got some scores which can be seen in the table below.

Dependent Variables	**p-value**	**Meaning**
Gender	8 e-6	REJECT. Thus there is difference in behaviour among the groups of gender
Ethnicity	0.393	ACCEPT. Hence no significant difference in the different group means
Political Orientation	4.5e-5	REJECT
Dependents	0.01	REJECT

From this, we can infer that the anxiety scores does behave differently for different gender, people having different political orientation and people having dependents.

- **Post-Hoc Testing:**
  Post hoc tests are an integral part of ANOVA. When you use ANOVA to test the equality of at least three group means, statistically significant results indicate that not all of the group means are equal. However, ANOVA results do not identify which particular differences between pairs of means are significant. Use post hoc tests to explore differences between multiple group means while controlling the experiment-wise error rate.

  But since all those groups which came out as statistically significant had only two levels(Gender: 2, Do you have dependents: 2, Political Orientation: 2), there was no need for such testing.

Though, during the final presentation, as suggested by Vishnu Sir, Factorial ANOVA was applied. We know that a factorial ANOVA is used instead of a one-way ANOVA when using research questions looking at the influence of two or more variables. And one way ANOVA helps only if we have very strong/sound reasons not to expect an interaction between the 2+ factors. Since we don’t know about that, factorial ANOVA seems a more credible method to go for.

- **Factorial ANOVA**

  For the factorial anova, the formula basically became such that the anxiety_score is now dependent on all the categorical variables. Applying that, I got:

  fact_model = aov(trans_Anxiety_factor_STAI ~ Gender.1.Male + PoliticalOrientation + Ethnicity + Dependents, data = data_s1_t1)

```
                       Df Sum Sq Mean Sq F value   Pr(>F)    
Gender.1.Male           1  0.072 0.07161  20.526 6.55e-06 ***
PoliticalOrientation    1  0.053 0.05251  15.052 0.000111 ***
Ethnicity               5  0.022 0.00446   1.277 0.271269    
Dependents              1  0.021 0.02147   6.155 0.013256 *  
Residuals            1056  3.684 0.00349                     
---
```

  From this output we can see that both variables except ethnicity explain a significant amount of variation in average anxiety score (*p* values < 0.05). Just like above, since these variables have only two levels, we dont have to apply the post hoc testing like the TukeyHSD test.

  To know the exact interaction of these variables while doing the factorial ANOVA, I replaced the ‘+’ sign with ‘*’ sign.

  *inter_model = aov(trans_Anxiety_factor_STAI ~ Gender.1.Male * PoliticalOrientation * Ethnicity * Dependents, data = data_s1_t1)*

  By observing the boxplots as well, we can add to this conclusion and infer which level of the group affected the anxiety score in what way. For example, clearly, the Gender termed as ‘1’ has higher anxiety scores on avergae when compared to Gender ‘2’.

Note: If the group had more than 2 levels and was statistically significant, Post-hoc Testing would be necessary to solve the confusion.

➔ **Wilcoxon-Mann-Whitney Test**

All the above tests could be done for Study 2 dataset, since the dependent variable followed normality. The dependent variable in the study 2 data, which is the WTP scores(vWTP, etc - denoting the info seeking behaviour) is not normalized and heavily skewed.
Hence I proceeded with the Mann-Whitney Test.

Null Hypothesis: vWTP scores(info seeking) is similar for both genders
Alternate hypothesis: vWTP scores(info seeking) is different for both genders.
The p-value comes out to be 0.67. Hence we accept the NULL Hypothesis. Therefore vWTP scores are not much different between male/female.

Work Done by my teammate:

My teammate worked on Linear Regression and also tried hands with LInear Mixed models. This work was also done originally by the authors of the paper.
Results from Study 1 suggest that individuals who reported higher anxiety were more likely to seek information related to the pandemic.
Using those methods in Study 2, we showed how induced anxiety as well trait anxiety led to greater information-seeking in response to the large magnitude (but not valence) of changes.

Conclusion and Discussion

On observing and analyzing the data provided, we see how information seeking behaviour changes with anxiety scores in general. And also how trait/induced anxiety leads to greater information-seeking in response to the large magnitude (but not valence) of changes.

On the other hand, we also separately analyzed the effect of the categorical variables on the anxiety scores itself using various ANOVAs. Such analysis was not done by the authors in the original paper. It showed how Gender, Political Orientation and whether a person has Dependents do affect the anxiety scores. Ethnicity on the other hand didn't have much effect on the anxiety scores.

The findings of our ecological study also support previous research[6][7][8] in that participants who expressed more worry during the pandemic sought out more information regarding COVID-19. The pandemic caused significant environmental changes and a negative valence, which may be the cause of the rise in information seeking.

Our research shows a connection between anxiety and increased information seeking in response to significant change, which sheds fresh light on the complex interplay between anxiety and information seeking and also how categories like gender and dependents, etc might affect the anxiety scores. Anxiety might cause people to look for more knowledge when they are preparing for significant changes in their environment, such changing jobs or moving locations. On the one hand, this might make things more adaptable and less unclear, but in some situations, it might also result in indecision and information overload.

References

1. Kelly, C. & Sharot, T. Individual differences in information-seeking. *Nat. Commun.* **12**(1), 1–13 (2021).
2. Kobayashi, K. & Hsu, M. Common neural code for reward and information value. *Proc. Natl. Acad. Sci.* **116**(26), 13061–13066 (2019).
3. Dörnemann, A., Boenisch, N., Schommer, L., Winkelhorst, L., & Wingen, T. (2021). How do good and bad news impact mood during the Covid-19 pandemic? The role of similarity.
4. Johnston, W. M. & Davey, G. C. L. The psychological impact of negative TV news bulletins: The catastrophizing of personal worries. *Br. J. Psychol.* **88**, 85–91 (1997).
5. Gallagher, M. W., Naragon-Gainey, K. & Brown, T. A. Perceived control is a transdiagnostic predictor of cognitive–behavior therapy outcome for anxiety disorders. *Cogn. Ther. Res.* **38**(1), 10–22 (2014).
6. Drouin, M., McDaniel, B. T., Pater, J. & Toscos, T. How parents and their children used social media and technology at the beginning of the COVID-19 pandemic and associations with anxiety. *Cyberpsychol. Behav. Soc. Netw.* **23**(11), 727–736 (2020).
7. Ebrahim, A. H. *et al.* COVID-19 information-seeking behavior and anxiety symptoms among parents. *OSP J. Health Care Med.* **1**(1), 1–9 (2020).
8. Loosen, A. M., Skvortsova, V. & Hauser, T. U. Obsessive–compulsive symptoms and information seeking during the Covid-19 pandemic. *Transl. Psychiatry* **11**(1), 1–10 (2021).