Behavioral Research: Statistical Methods
Guided By: Prof. Vishnu Sreekumar and Prof. Vinoo Alluri
Paper: Change in information-seeking behaviors due to Anxiety (Link)
Submitted By: Debashish Roy (2021201034)
Team: CAR

Introduction and Problem Statement:

In this project, we tried to analyze the relationship between information seeking and anxiety based on the paper "Change in information-seeking behaviors due to Anxiety". In a laboratory setup, the authors of the paper tried to figure out the relation between trait anxiety and induced anxiety with information seeking. The whole paper talks about three experiments which are as follows:

1. **COVID-related information-seeking effect on Anxiety factor and demographics across two different time periods.**

   Usually, anxiety will lead to more information seeking. During the COVID period, there are a lot of uncertainties regarding what is going to happen next. It seems like getting more and more COVID-related information can help users to mitigate these uncertainties, which makes them calm. But it can turn out to be another way also like users can be more anxious after going through many negative news stories.

   Here we will try to formulate a relationship between the COVID-related information-seeking vs anxiety factor and demographics. The key variables used in these experiments apart from demographics are COVID-anxiety scores and the SSAI scores, these are obtained using a questionnaire provided to every participant. A correlation among variables collected is on the left side.
   There are various types of anxiety scores but we will only deal with Anxiety_factor_noSTAI, Anxiety_factor_STAI, STAI_Score, and AnxietyLockdown variables.
   The data is collected in two periods with participants 1065 from 26-29 March '20 and participants 700 from 23-25 April '20.

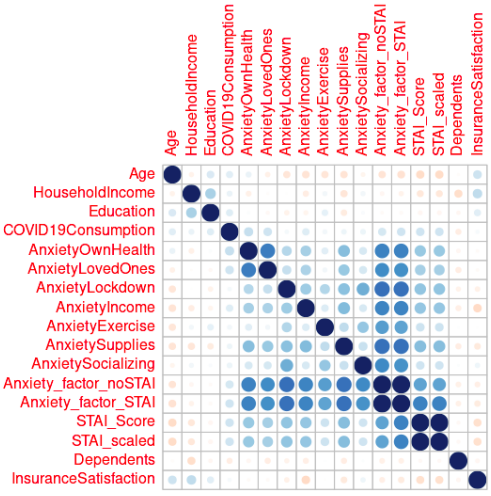


Fig1: Correlation among different variables for study 1

2. **Effect of trait anxiety and demographics to information seeking**

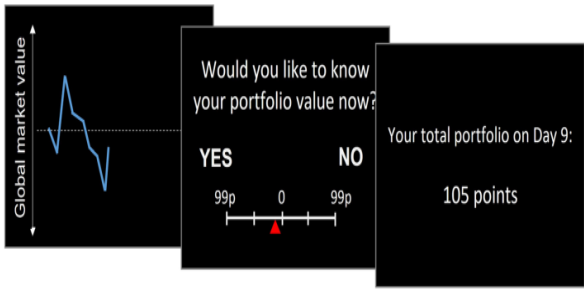
   In order to measure this the author has defined a proxy task. Instead of directly measuring this they made the participant play a stock game. The component of the stock game are as follows:

   - Every participant starts with an initial of 1000 Pounds and they need to invest it in any 2 stocks in total of 5 available stocks. These are imaginary stocks.

- For every trial the user is only able to see the change in cumulative effect of 5 available stocks, not individual stocks.
- After analyzing the market change the user can bid some value to see the price of his stock. This is treated as the willingness to pay.
- At the end of the 50 trail participant value = price of two stocks - willingness to pay for all trials.

The features used here are Subject, Gender, Age, Trait Anxiety, WTP absolute market change, WTP signed market change, and vWTP. For the study, the number of participants is 42 (15 males, 27 females, mean age = 28.02 years ± 11.62 (SD), Age: 18-66 years)
There is two datasets provided in the paper. One where the data of each trial is given with respect to each user. And second where for each user they have estimated all values in one entry.

Fig2: Stock price Game Simulation Example



We will do separate analyses with these two datasets. But the main idea is to capture the behavior of the participants using the willingness to pay variable during trials.

3. **Effect of induced trait anxiety and demographics on information seeking**
   In this study the author of the paper tries to induce anxiety based. They try to measure the anxiety before and after telling them they have to give a performance on the stage. All the participants are divided into two groups: the control and the test group. The key parameters used in this study are the same as in experiment 2. The distribution of participants is

# Participants	#Male: #Female	Age (years)	Age Range
24	10:14	23.46 ±4.79	19-34
24	10:14	21.96 ±3.51	19-32

Literature Review

We tried to use the base paper (J., 2022, #). But apart from that we have added our experiments to get more inference. (Ryan, 2019, #) gives us an idea, but more or less the paper focuses on a medical condition named Cyberchondria. This is relevant but our main goal is to focus on formulating a relationship between anxiety and information seeking. In this direction, we have briefly gone through (Christopher, 2021, #) how information-seeking changes in different domains and the formulation idea comes from this paper.

Methodology

Analysis Done By my fellow Team Mates

ANOVA (Multiple one way and factorial/mixed ANOVA):

- Normality testing
    - >1000 data points. Hence apparently no need for normality.
    - Slight deviation from normal distribution, was fitted by adding a constant
    - Finally confirmed by Shapiro test.(p-value = 0.09)
- Homogeneity
    - Boxplot
    - Levene Test
- Independence

Analysis Done By Me

Hypothesis tested by Me:

- Information seeking has a linear relation with anxiety. And both are directly proportional to each other.
- Anxiety leads to an increase in information-seeking, when large there is a change of large magnitude.

Mainly we have used Linear regression techniques:

- Linear Regression
- Generalized Linear Regression
- Mixed Linear Regression

Study1

We formulated that there is a linear relationship between anxiety scores and covid related information seeking and hence we formulated them using this equation.

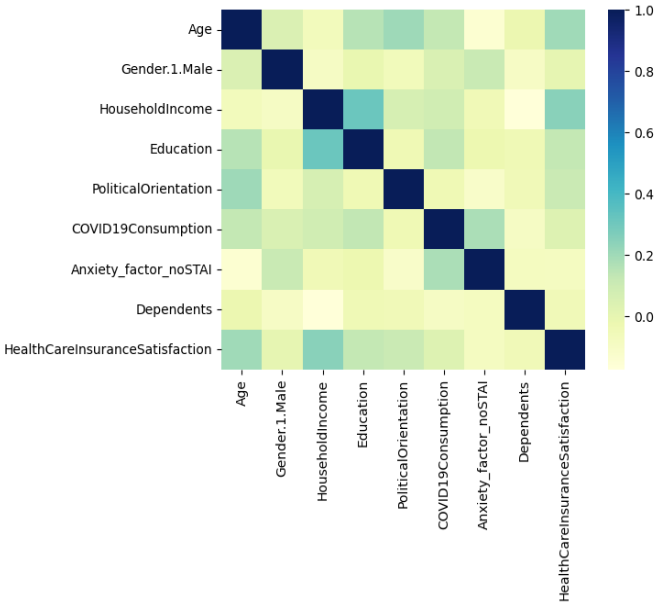
COVID-related information-seeking = β0 + β1* Anxiety Index Score + β2*Age + β3*Gender + β4*Educational level + β5*Income + β6*Political orientation + β7*Ethnicity + β8* Whether they had dependents + β9* satisfaction with their health insurance

Instead of β1's Anxiety Index we have used:

- Anxiety_factor_noSTAI,
- Anxiety_factor_STAI,
- STAI_Score,
- AnxietyLockdown

Due to the constraint of space we have only selected the factor which gives the best AIC score. But the code where available for all four analyses. This is done for both Time1 and Time2 periods.

We have used the **variance inflation factor** to test for the linear relationships among independent variables and removed the variables which have greater than 10 scores. Also for every regression model we have only considered them if they satisfy the assumptions of linear regression. For that, we have also used **White's Lagrange Multiplier Test for Heteroscedasticity**. Heteroscedasticity refers to the unequal scatter of residuals at different levels of a response variable, which violates the assumption that the residuals are equally scattered at each level of the response variable.



Correlation matrix across the Independent variables

VIF	Column
10.895716	HealthCareInsuranceSatisfaction
10.617572	Dependents
10.175761	Age
9.488231	Gender.1.Male
8.765398	STAI_Score
8.619276	PoliticalOrientation
6.558334	Education
6.235379	HouseholdIncome

VIF values for all the Independent variables

Null (H0): Homoscedasticity is present (residuals are equally scattered)
Alternative (HA): Heteroscedasticity is not present (residuals are not equally scattered)

P-value is less than 0.05 (a critical value that we have fixed), which means residuals are equally scattered. We have also tried **Q-Q Plots**. It is an easy way to check for normality distribution.

Study 2

We have formulated that there is a linear relationship between Traint Anxiety and willingness to pay in different ways. So we have formulated the equation in the form:

Willingness to pay for abs market change = β0 + β1 Trait Anxiety score + β2*Age + β3*Gender*

We have also experimented with valence market change and signed market change as a dependent variable instead of a willingness to pay for absolute market change. In this experiment, since we have the transaction data, for every person, we have tried the **Linear Mixed model**. We have taken different people as a random effect

and Anixety, age, and gender as the fixed effect. The reason is that the willingness to pay varies from person to person and each person has a different personality.

We also tried to explore the **generalized linear model** but we changed our hypothesis. We saw that age and trait anxiety might have a correlation. Maybe with age, people get more anxious, (conditioned over this game only), so we try to explore that. We have formulated our GLM equations as

Willingness to Pay = TraitAnxiety + age + gender + TraitAnxiety * age

Study 3

We have repeated Study 2 methodology with induced trait anxiety this time. Here also we have performed Linear Regression, Linear Mixed model, and Generalized linear models.

Results

Results from my Team Mates:

ANOVAs:

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

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

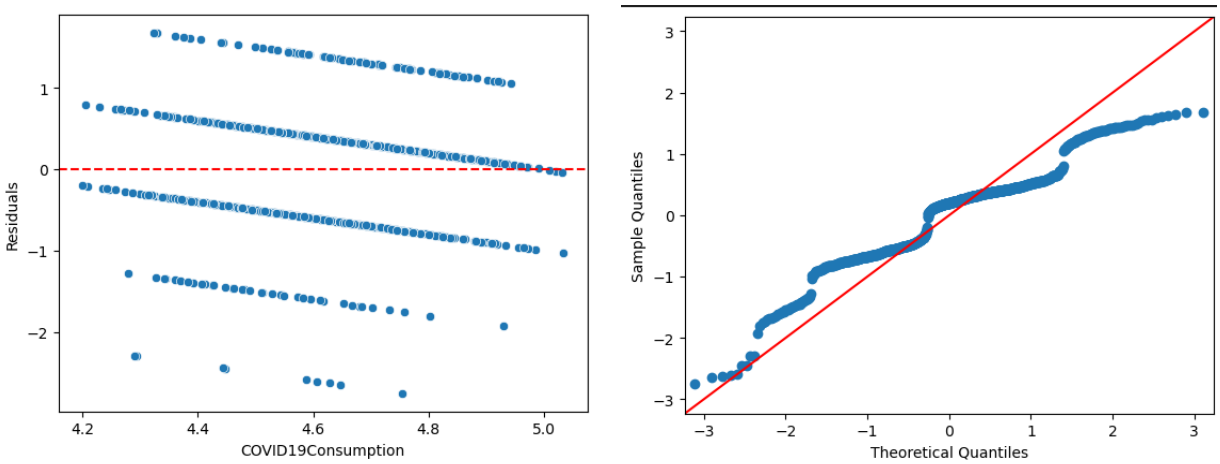
Results from Me

I have done total of 4*2=8 Linear Regression for study 1, 3 Linear Regression, 1 GLM and 1 Linear Mixed Model for study 2 and study 3.
All the code and results are present in https://github.com/debashish05/tgcn, these are in Python.

Study 1

Independent variables	AIC Values
Anxiety_factor_noSTAI	2251.
Anxiety_factor_STAI	2242.
AnxietyLockdown	2287.
STAI_Score	2236.

Based on these scores we will show an analysis of Anxiety Lockdown only due to space constraints.



Q-Q Plots

The plot on left hand side on last page, shows a random scatter of points with no discernible pattern, then the residuals are homoscedastic and have constant variance across the range of fitted values. This is a desirable property of a regression model, as it suggests that the model's assumptions are met and the results are reliable.

With the Het white test, Test p values come out to be 0.02, which is less than 0.05. Which means residuals are equally scattered.

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.6319	0.022	214.374	0.000	4.590	4.674
Age	0.1015	0.023	4.353	0.000	0.056	0.147
Gender.1.Male	0.0279	0.022	1.269	0.205	-0.015	0.071
HouseholdIncome	0.0465	0.024	1.917	0.056	-0.001	0.094
Education	0.0645	0.023	2.762	0.006	0.019	0.110
PoliticalOrientation	-0.0483	0.022	-2.159	0.031	-0.092	-0.004
AnxietyLockdown	0.0420	0.022	1.914	0.056	-0.001	0.085
Dependents	-0.0507	0.022	-2.288	0.022	-0.094	-0.007
HealthCareInsuranceSatisfaction	-0.0096	0.023	-0.416	0.677	-0.055	0.035

Summary of Valence Anxiety Lockdown Linear Regression Model

Let's interpret the final results:

Coefficient Suggest the predicted Value will increase by a coef unit (for example for Age it is 0.1015) when Dependent Variable (Eg. Age) increases by 1 unit, and rest all remain constant (assuming all values are normalized).

By looking at this the most important variable is Age. In other Linear Regression, the most important variable was anxiety. So we can conclude that age and anxiety play a key role in information seeking. P and t values are also tabulated. Range [0.025 and 0.975] are both measurements of values of our coefficients within 95% of our data, or within two standard deviations. P tells us how likely the value of the coefficient we get is by chance. All the values are low except health care insurance.

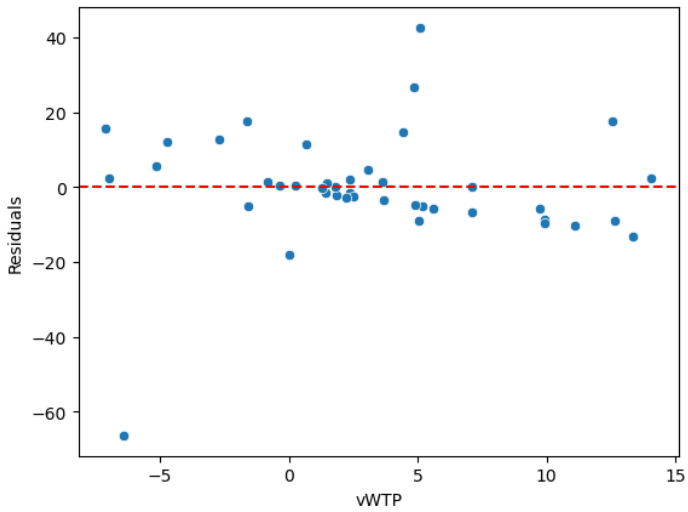
Study 2

Willingness to pay for abs market change = β0 + β1* Trait Anxiety score + β2*Age + β3*Gender

We have also experimented with valence market change and signed market change as a dependent variable instead of a willingness to pay for absolute market change. The signed version gives an AIC score of 233, the valence version gives a score of 357.7 and the absolute version gives AIC score of 211.8. So we will consider the valence version. One thing to note here is that the signed version failed the Hat Whitey test.

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.1857	2.495	1.277	0.210	-1.870	8.241
Subject	-5.2097	2.895	-1.799	0.080	-11.076	0.657
Gender	3.8968	2.620	1.487	0.145	-1.412	9.206
Age	1.7867	2.851	0.627	0.535	-3.991	7.564
Trait Anxiety	1.2568	2.597	0.484	0.631	-4.006	6.520

Summary of Valence WTP Linear Regression Model



The plot of valence willingness to Pay v/s the

VIF	Column
6.268014	Trait Anxiety
6.230957	Age
5.444430	Subject
2.866686	Gender

VIF scores of independent variables

Residual.

Let's interpret the final results from the summary in the last bottom page:

Coefficient Suggest the predicted Value will increase by a coef unit (for example for Trait anxiety it is 1.2568) when Dependent Variable (Eg. Trait Anxiety) increases by 1 unit, and the rest all remain constant (assuming all values are normalized).

By looking at the summary the most important variable is Gender. In other Linear Regression, the most important variables were anxiety and age. So we can conclude that gender and anxiety play a key role in information seeking. Rest all the values are the same as described in Study 1.

Linear Mixed Model

Here we have taken different people as a random effect and Anixety, age, and gender as the fixed effect. The reason is that the willingness to pay varies from person to person and each person has a different personality. So all transactions made by each person will fall into one group. The results of the linear mixed model are as follows:

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-0.001	0.094	-0.014	0.989	-0.186	0.183
TraitAnxiety	0.017	0.096	0.174	0.862	-0.171	0.205
age	-0.019	0.094	-0.201	0.841	-0.203	0.165
gender	0.105	0.096	1.091	0.275	-0.083	0.292
Group Var	0.370	0.107				

Here the unit change in a variable, creates an exponential of coefficient of that variable increase. The results are not that good for this model, but we wanted to experiment. The P > |Z| values are very high, which signifies the outcomes are majorly by chance. This highly suggest that keeping person as a random variable is not a good measure here. But there are no other alternate available for the data.

Generalise Mixed Linear Regression

vWTP = B1*Gender + B2*Age + B3*Trait_Anxiety + B4*Age * Trait_Anxiety
Despite of the large normalization the factor of Age * Trait_Anxiety caused an overflow in the library. And the results are as follows

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-8.338e+15	1.05e+07	-7.96e+08	0.000	-8.34e+15	-8.34e+15
Gender	7.728e+15	1.07e+07	7.19e+08	0.000	7.73e+15	7.73e+15
Age	3.316e+14	1.06e+07	3.14e+07	0.000	3.32e+14	3.32e+14
Trait_Anxiety	-2.253e+13	1.1e+07	-2.04e+06	0.000	-2.25e+13	-2.25e+13
Age:Trait_Anxiety	-2.569e+14	1.72e+07	-1.49e+07	0.000	-2.57e+14	-2.57e+14

Summary of GLM Model

Here for gender simple exponentiation of the coefficient will give the change in vwtp at different gender. But for other Age we need to add to fix trait anxiety as constant and it will be expoentiation of B2+B4. Simillarly for Trait_Anxiety it will be exp(B3+B4). Here the values are not accurate so I kept it in form of Bi's.

Study 3

For study 3 we have replicated study 2 with induced anxiety score. Just to save space the details are not provided. But all the code and experiments are present in the Github Repository.

Conclusion and Discussion

Conclusion:

We have proved our hypothesis

- Information seeking has a linear relation with anxiety. And both are directly proportional to each other.
- Anxiety leads to an increase in information-seeking, when large there is a change of large magnitude (valence).
- Age and Gender plays a crucial role in Information Seeking
- Dependents plays a very important role in information seeking.

Discussion:

- The output of the Linear mixed model was not accurate as the parameter were suggesting that there is high probability that it relation comes via chance. We should have tried different libraries.
- There is an overflow issue with the global mixed model which we won't able to resolve.
- Other than this rest of the model performed well. Not only we performed linear regression we have also check the assumptions it needs to satisfy, in some cases it is not satisfying that. So we neglected the results.
- I learnt a new way of doing behavioral research via proxy task. Here we used stock game as proxy task. Similarly for some complex task we can map it to a smaller proxy task.

References

Christopher, K. (2021). Individual differences in information-seeking. *Nature*, *7062*(12), -. https://doi.org/10.1038/s41467-021-27046-5

J., C. (2022). Anxiety increases information-seeking in response to large changes. *Scientific Reports*, *7385*(12), -. https://doi.org/10.1038/s41598-022-10813-9

Ryan, M. (2019). The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. *Journal of Affective Disorders*, *245*(2019), 270-278. https://www.sciencedirect.com/science/article/abs/pii/S0165032718315775#preview-section-abstract

Paper by Code (in R): https://github.com/affective-brain-lab/Anxiety-increases-information-seeking-in-response-to-large-changes-

Code by Me: https://github.com/debashish05/tgcn