

IDS 701 - Team 1 Project Proposal

Chen Elisa, Liu Xiaoquan, Qu Genesis

2023-03-27

Topic

College students today appear to be more stressed and anxious than ever before. This is a major problem affecting the student's mental and physical well-being, and greatly impacts the student's academic performance. According to a survey conducted by the National Institutes of Health, 71% of the respondents indicated increased stress and anxiety due to the COVID-19 outbreak. There were several stressors that were identified, but one of the major indicators listed was decreased social interactions due to physical distancing (86% of respondents)[1].

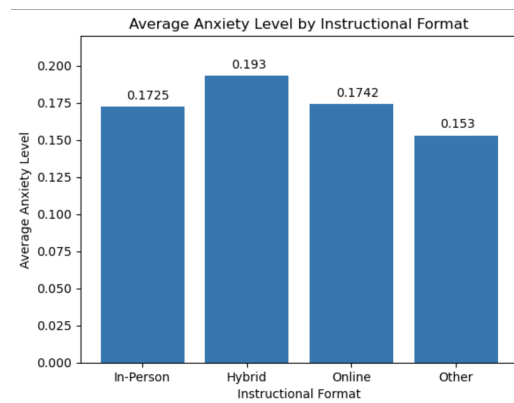
During the pandemic, many students and academic staff realized how online / hybrid learning could make education more engaging and accessible for many students. While there could be many benefits to this, it is important to assess whether this comes at a cost of the well-being of students[2].

Exploratory Data Analysis

Given our topic and the problem we want to address (higher anxiety levels of college students in the US), we want to explore the following questions in depth:

1. What is the portion of students with high anxiety level by instructional format? Is the difference statistically significant?

Answering this question would validate the problem that we're observing. The students in the survey were given multiple prompts to assess their anxiety level. The researchers who conducted the survey created a binary field to indicate whether the student experienced high anxiety level based on the responses given by students.

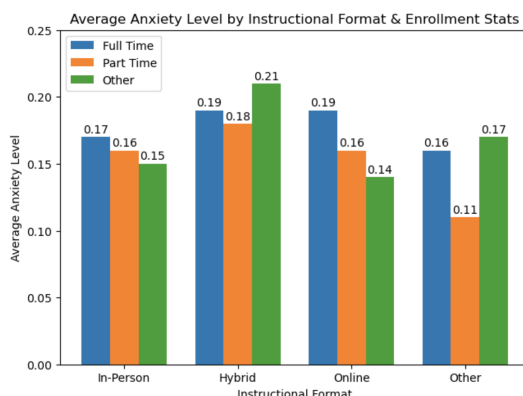


It appears that the portion of students who experienced high anxiety levels were about the same for online and in-person students. The difference between online and in-person anxiety levels were statistically insignificant.

However, when we measured the difference between hybrid and in-person students' anxiety levels, we observed a statistically significant difference. At a first glance it would appear that the anxiety level is fairly consistent across all instructional formats. However, it is good to keep in mind that there are various factors that we have not accounted for yet including the student's access to mental health resources and enrollment status.

2. Does the anxiety level differ by the enrollment status and instructional format?

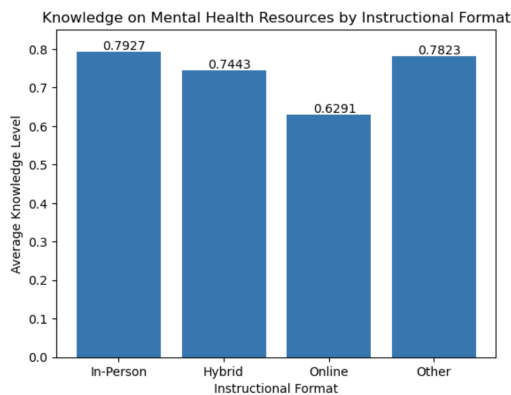
Answering this question would help us understand whether the trend we observe vary by enrollment status perhaps. This could help us potentially narrow down our scope of study.



Motivated by the first question, we wanted to explore whether the anxiety level would vary if we also accounted for the enrollment status. Interestingly, we observe now that the portion of students with high anxiety level increased for hybrid and online students if they were enrolled in school full-time. This could potentially signal that we want to control for the student's enrollment status as well when we measure their anxiety levels.

3. Do students know how to access mental health resources offered by their college (across different instructional forms)?

Answering this question would help us understand whether there's a potential relationship between anxiety level and whether the student knows where to access mental health resources in their college. In the survey, the students were asked to evaluate on a binary scale if they were "to seek professional help for [their] mental or emotional health, [they] would know where to access [their] school's resources."



It appears from the above graph that students doing school entirely remote are significantly less (~20% less) aware of their school's mental health resource offerings compared to the students studying in-person or in a hybrid model. The lack of knowledge about mental health resources could be one of the explanations to why students doing school online experience a higher anxiety level on average.

Project Question

In this project, we aim to explore the effects of social distancing measures on the anxiety levels of college students in the US during the years 2021-2022. As supported by our results from the EDA, the anxiety level is ~1-2% higher for students who did online / hybrid schooling and is even more pronounced among full-time students. Given our findings, the question that we're interested in answering is:

Did remote / hybrid schooling cause higher anxiety levels among college students in the US during the years 2021 - 2022?

Answering the proposed causal question would shed more light on whether online schooling will increase the anxiety levels of college students in the US and by how much. In the case that online school models will become more prevalent, it will become increasingly important to be aware of how this impacts the well-being of students and ensure the availability of mental health resources.

Ideal Experiment

To investigate the causal relationship between instructional format (in-person schooling and remote/hybrid schooling) and anxiety levels among college students in the US during 2021-2022, an ideal experiment would involve the following steps:

1. Randomly select a sample of college students in the US who were enrolling in in-person schooling before 2021 and divide them into two groups (treatment group and control group). The sample size should be large enough to detect meaningful differences between the treatment and control groups.
2. Administer a baseline anxiety level test to both groups to ensure that the anxiety levels are the same (no baseline difference) at the beginning of the study.
3. Implement the treatment: Remote/hybrid schooling for the treatment group and continue in-person schooling for the control group.
4. After one academic year (2021-2022), administer a follow-up anxiety level test to both groups.
5. Analyze the results and compare the change in anxiety levels between the treatment and control groups.

Study Context

In order to obtain data that includes variation in the treatment variable and measures the outcome variable, we could get administrative data from colleges and universities that tracks students' enrollment status (remote/hybrid vs. in-person) and academic performance, as well as demographic information such as age, gender, and socioeconomic status. Those administrative data can help us in analyzing the relationship between enrollment status and anxiety levels, controlling for other factors that may affect anxiety. We could also obtain clinical data from mental health clinics or counseling centers that serve college students, which would provide a more objective measure of anxiety levels. This data could include diagnoses of anxiety disorders, self-reported anxiety levels, and other clinical measures of anxiety.

Project Design

We are given observational data of students from the National Institute of Health, which means that there will almost certainly be baseline differences between students who respond with high anxiety levels and student who respond with low anxiety levels. To accommodate for the baseline differences and ensure that we get as close an approximation to the Average Treatment Effect as possible, we will first check for balance in the other demographic and clinical variables in the data.

We will first determine whether matching is needed to ensure that our treatment (hybrid/online schooling) and our control (in-person schooling) observations share a common support. If we identify variables where

treatment and control lack common support, we will apply the FLAME algorithm and determine whether it is feasible to prune our data.

We will run a regression model to estimate the effect of hybrid/online schooling on student anxiety levels. To account for baseline differences, we will include demographic and clinical variables as fixed-effects.

Model Results

Below are two hand-drawn regression tables showing model results if our hypothesis is true and if our hypothesis is false. Since our null hypothesis is that hybrid/online schooling has no effects on college students' anxiety levels in the US during year 2021-2022, these are our hypothetical model results:

Hypothesis == True

OLS Regression Results						
Dep. Variable:	Anxiety				R-Squared	
Model:	OLS				Adj. R-Squared	
Method:	Least Squares				Prob	
No. Observations:	95861				Log-Likelihood	
Df. Residual:	95810				AIC	
Df. Model:	50				BIC	
Covariance Type:	Robust					
	coef	std. err.	t	P > t	[0.025	0.975]
Intercept	0.1	0.02	-1	0	0.12	0.08
Hybrid/Online	0.01	0.1	0.1	0.95	-0.15	0.2
Race: Hispanic	0.005	0.004	0.2	0.8	0.0048	0.0052
...						

* Statistically Insignificant

Hypothesis == False

OLS Regression Results							
Dep. Variable:	Anxiety	R-Squared					
Model:	OLS	Adj. R-Squared					
Method:	Least Squares	Prob					
No. Observations:	95861	Log-Likelihood					
Df. Residual:	95810	AIC					
Df. Model:	50	BIC					
Covariance Type:	Robust						
	coef	std. err.	t	P > t	[0.025	0.975]	
Intercept	0.1	0.02	-1	0	0.12	0.08	* Statistically Significant
Hybrid/Online	0.02	0.0005	4	0	0.018	0.022	
Race:Hispanic	0.005	0.004	0.2	0.8	0.0048	0.0052	

Final Variables Required

We need instructional format data (remote/hybrid vs. in-person) of college students in the US during 2021-2022 academic year, demographic information such as age, gender, and socioeconomic status of college students in the US, as well as data in terms of measurement of their anxiety levels and other potential clinical information(depression, eating and body image, etc.).

Data Sources

Considering the characteristics we want for our study, we gathered information from The Healthy Minds Network (HMN), one of the nation's leading research organizations dedicated to adolescent and young adult mental health. HMN's major investigators are located at the University of California-Los Angeles, the University of Michigan, Wayne State University, and Boston University. The dataset that we are going to use is collected by the University of Michigan where they sampled 100k college (undergrad, graduate and phd) students in the United States and asked questions related to both demographic and clinical information (e.g. anxiety levels, depression, eating and body image, etc.) during the 2021-2022 academic year.

References

- [1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7473764/>
- [2] <https://www.weforum.org/agenda/2022/06/online-learning-higher-education-covid-19/>