Relationship between Student Workload and Anxiety

Team 63

Section C

Roshni Balasubramanian, Steven He, Ansh Nemali, Jupiter Seong, Sathish Prasad VT

Motivation

- The relationship between student workload and anxiety is one that is universally understood. However, a layman's understanding of this is largely formed by personal or anecdotal experiences.
- Our goal is to uncover if there is a causal relationship that exists between the two. If one exists, what does it look like? Are there other factors that influence this relationship? If so, how?
- This will allow education professionals and psychologists to tackle these negative feelings early, so students are more mentally resilient as they move forward in life.
- How does academic performance relate to workload and anxiety?
  - Academic performance can be a confounder as it obfuscates students' natural ability. Gifted students can effectively manage the workload without feeling anxiety compared to their peers.
- Our own, collective experiences we developed the following hypotheses:
  - H_0: Student workload and student anxiety are orthogonal and insignificant
  - H_1: Student workload and student anxiety are related and significant

Literature Review – I

- **Rahim et al. (2016):**
  - This cross-sectional study found a positive correlation between credit hours, study hours, assignments, and stress level, with second-year biomedical students (n = 141) showing a positive and significant relationship ($p < 0.05$) between stress level and academic workload (credit hours). Statistical tools used include ANOVA (Analysis of Variance) and Spearman's correlation. However, there were no significant differences in stress levels across different years of study, suggesting that other factors, such as "personal or financial problems or college activities" may contribute to stress levels.
- **Kausar (2010):**
  - This cross-sectional study tested for coping mechanisms to manage stress among students while accounting for student workload. The primary statistical tool used for the study is regression analysis. The study had Master's student respondents (n = 150). The study found a positive relationship between academic workload and perceived stress among students. The positive correlation was 0.31 with a p-value < 0.01.

Literature Review – II

- **Ragusa et al. (2023):**
  - This relational study, conducted with high school students aged between 16 and 19 years (n = 991), examined the effect of academic self-regulation on procrastination, academic anxiety, stress, and resilience. It found that procrastination, which can be influenced by workload, positively predicted academic stress and anxiety.
- **Wunsch et al. (2017):**
  - A total of 64 university students completed five surveys, with the first at the end of the semester break (baseline) and the remaining four during the last four weeks of the semester (examination period). The participants reported their activity levels, sleep quality, and well-being. The study found that exercise during high-stress periods can increase sleep quality, and through this, decrease anxiety among students.

Literature Review – III

- **Coakley et al. (2022):**
  - This cross-sectional study aimed to examine the associations between food insecurity (FI), housing insecurity (HI), and a novel basic needs insecurity score with mental and physical health among university students. Tools used included regression analysis and Chi Squared tests. Nearly 26% of students were food insecure, and 44% were housing insecure. Almost half of the students experienced either food or housing insecurity, with 19.6% experiencing both. The anxiety score, a measure of anxiety symptom severity, was the strongest predictor of food, housing, and overall basic needs insecurity scores.
- **Li (2022):**
  - The study aimed to explore the influence of teacher-student relationships on anxiety and depression in children (n = 340). Techniques used included one-way ANOVA and regression analysis. The results indicated that teacher-student relationships significantly predicted anxiety and depression levels in higher-grade primary school students. Additionally, children's self-perception and optimism also significantly influenced their anxiety and depression levels.

Literature Review – Summary

- Plenty of literature exists on the relationship between a variety of student categorical factors included in our dataset and their effect on student anxiety and/or stress.
- The findings of our literature review highlight the following relationships among our variables:
  - Study 1: Academic (student) workload increases stress level
  - Study 2: Academic (student) workload increases stress level
  - Study 3: Academic (student) workload increases stress level and anxiety level
  - Study 4: Strong association between higher academic workload and lower sleep quality
  - Study 5: Anxiety is a strong predictor of basic needs insecurity
  - Study 6: Teacher-student relationships strongly predicted anxiety and depression levels

Novelty in Analysis

- Advanced statistical tools such as matching and regression adjustment will be employed to enhance analytical robustness.
- We plan to include an extensive set of influencing factors as confounders such as future career concerns to better capture the complexity of the causal relationship.
- Our research is pioneering as it addresses a specific causal inference model, incorporating previously unexplored confounders not considered in prior studies.
- Through our innovative approach, we seek to reveal nuanced insights into the relationship between predictors and anxiety, while addressing potential biases overlooked in the literature. Our use of instruments aims to overcome unobserved variables, particularly the personality traits of students.

Methodology

- Randomized Control Trial
- Nearest Neighbor Matching and Regression Adjustment
- Mitigating Bias with Instrument Variable

Laying Out the Problem

After constructing the DAG, we established that:

Outcome: Anxiety Levels (0 to 21)

Treatment: Having High Study Load

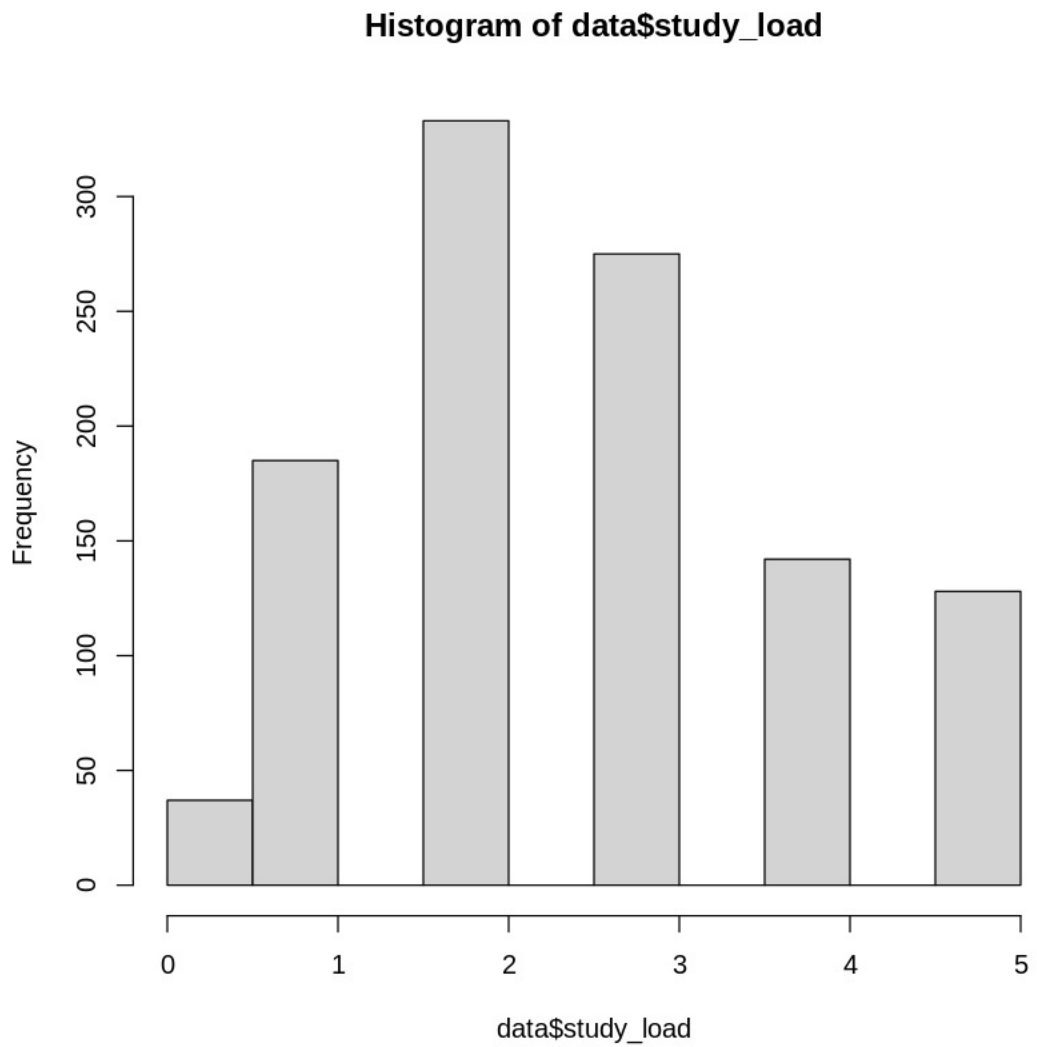
Treatment Effect: Effect of Study Load on Anxiety Levels

Treatment Group: Students having high Study Load (Study Load = 3, 4, 5)

Control Group: Students having lower study load (Study Load = 0, 1, 2)

Data

- School students in a nationwide survey.
  - Source: Kaggle dataset
- **1100 observations** across **21 features** that can be broadly broken down into **5 categories**:
  - **Psychological** - 'anxiety_level', 'self_esteem', 'mental_health_history', 'depression'
  - **Physiological** - 'headache', 'blood_pressure', 'sleep_quality', 'breathing_problem', 'stress_level'
  - **Environmental** - 'noise_level', 'living_conditions', 'safety', 'basic_needs'
  - **Academic** - 'academic_performance', 'study_load', 'teacher_student_relationship', 'future_career_concerns'
  - **Social** - 'social_support', 'peer_pressure', 'extracurricular_activities', 'bullying'

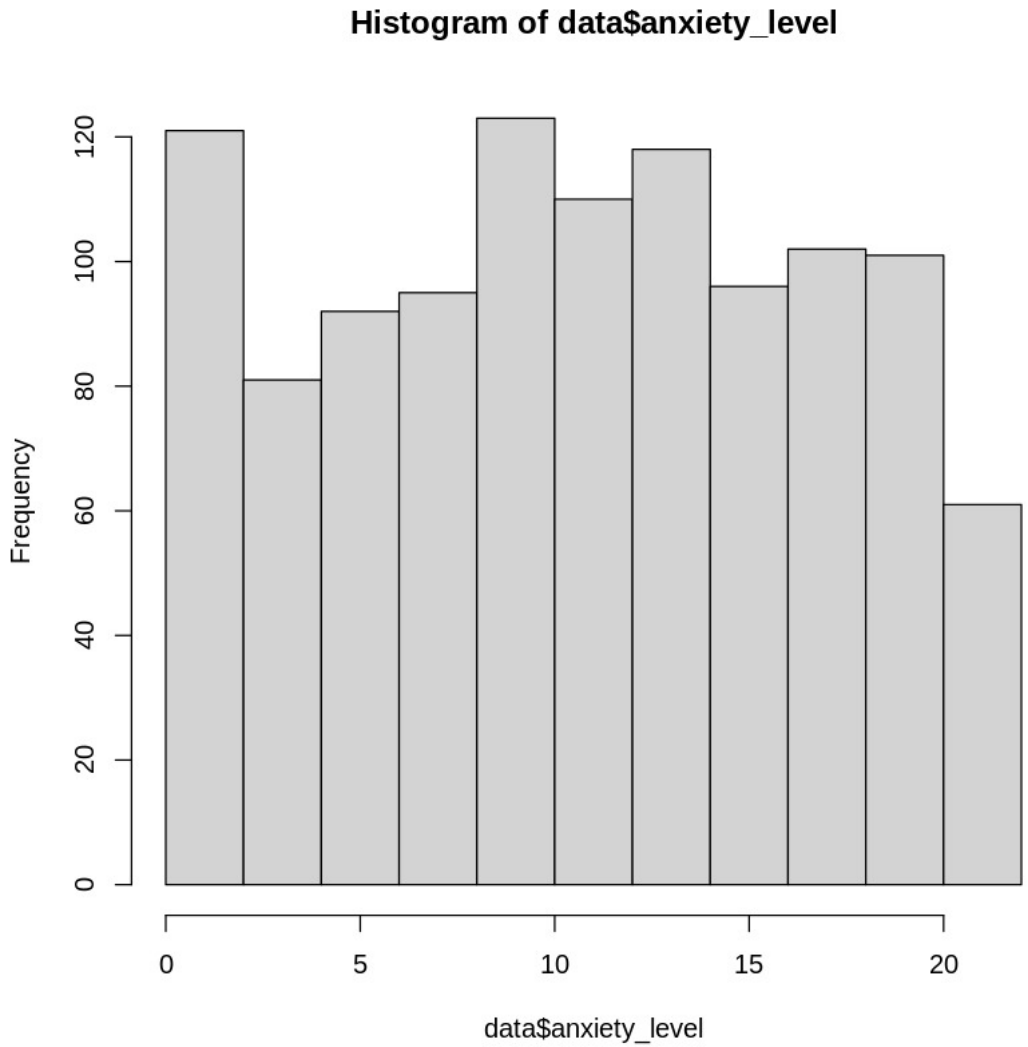


Exploratory Data Analysis I

- Histogram showing frequency of student workload (*study_load*)
- Student workload measured on a scale from 0 to 5
- 0 = lightest student workload, 5 = heaviest student workload
- Distribution of students seems to resemble lognormal

Exploratory Data Analysis II

- Histogram showing frequency of student anxiety level (*anxiety_level*)
- Student anxiety level measured on a scale from 0 to 21
- 0 = lowest student anxiety, 21 = highest student anxiety
- Relatively even distribution of students across the anxiety spectrum



Randomized Control Trials

Implementation: RCT

Why?

- RCT ensures that each participant has an equal chance of being assigned to either the treatment or control group. This randomness minimizes selection bias and ensures that any potential confounding variables are equally distributed between students with high or low study loads.

How?

- RCT was implemented by randomly allocating treatment.
  - The categorization of the covariate into 'high' and 'low' groups was determined by employing a mean-based split. Values below the mean of the covariate were designated as 0, indicating the 'low' category, while values equal to or above the mean were assigned a value of 1, representing the 'high' category.
- Covariate balance can be checked via KS Test and QQ Plots
  - For attributes such as extracurricular activities, teacher student relationship, future career concerns, academic performance, anxiety levels, the covariates were balanced in the relative frequencies of occurrence
- Using Simple Treatment assignment, treatment status can be applied using the 'simple_ra' function

Results of RCT

Covariate splits were relatively balanced with 49.09% Treated and 50.9% control (lower study loads)

The Kolmogorov-Smirnov (KS) test provided confidence in the matching results.

- The large p-Value implies that we fail to reject the null hypothesis, implying that density of academic performance among the treated group is the same as those in the untreated group. That's a good sign that our RCT will work.

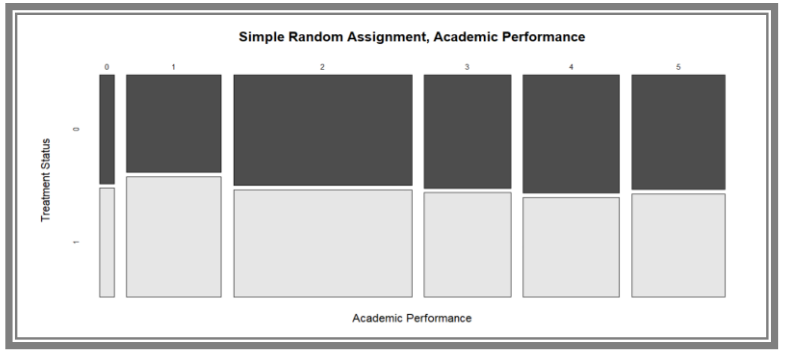


Figure: Mosaic Plot to Visually Represent Split between Treatment and Control Assignment in simple_ra

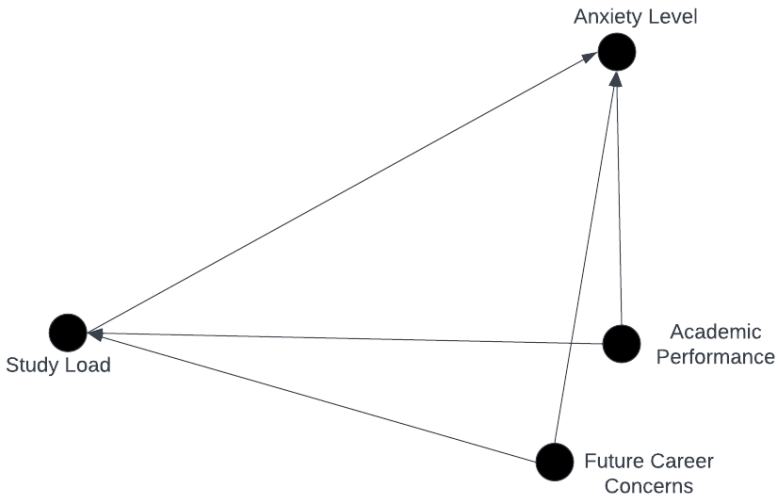
Results of RCT

- 46.8% of the treated sample have poor academic performance and 51.07% have good academic performance. The null hypothesis is that these are equal, given the large p-value, we fail to reject the null and this implies that the proportions of poor/good academic performance is similar in treated and untreated groups. This suggests that our co-variate of academic performance, is balanced and initially gave confidence in the RCT.
- The Difference in Means (DiM) showed an average anxiety level difference of 0.096 for students with a higher study load compared to those with a lower study load.

Results of RCT

- The Welch Two Sample t-test was conducted to assess the significance of the observed DiM.
  - The test concluded that the difference in anxiety levels between the groups was **statistically insignificant** (p-value > 0.05).
  - The 95% Confidence Interval included 0, further showing the insignificance of results
  - This insignificant was confirmed using regression OLS as well since results had a large p value and very low R squared value.
- While there is a numerical difference, it is not practically significant, as the statistical test did not support the observed difference as being beyond what could occur due to random chance.

Matching & Regression Adjustment



Choosing Controls

The controls chosen were based on testing theoretical soundness as well as checking significance on estimate

Good Controls:

- Future Career Concerns
- Academic Performance

Bad Controls that were identified and thus removed from next steps include teacher student relationship, peer pressure and social support

Implementation: Matching

01

Matching was done using academic performance on study load to find the smallest distance between subjects

02

The ‘ratio’ was set to 1 and this matched a single treated to a single control

03

Nearest Neighbor method was implemented using the Mahalanobis distance measure since Mahalanobis accounts for correlation of the covariates

Results: Matching

- A total of 545 matches were made using nearest neighbor matching; 10 subjects from the control were left unmatched. However, using this technique without a caliper, the quality of matches were poor.
- Setting a caliper of 0.3 for academic performance, led to better quality matches based on meeting all the criteria that follow:
  - Std Mean Diff Absolute Values < 0.1
  - Var Ratio Close to 1
  - eCDF Mean and eCDF Max close to 0
- However, following this criteria, the number of matches reduce to 252. We continue to RA with good quality matches.

Implementation and Results: Regression Adjustment

- Regression analysis on matched data offers a more nuanced estimation of treatment effects than the Difference of Means approach
- Conducting an OLS regression analysis on the variables of study load, future career concerns, and academic performance reveals a treatment effect of 0.9986 on anxiety and this result is statistically significant
- Hence, with every unit increase in study load, anxiety increases by ~1 unit (0.9986). This supports our initial hypothesis that high study load increases/has an impact on anxiety.
- Since we expected the effect of study load to be greater than 0.99 units, we suspect there may be some bias.

Mitigating Bias Using Instruments

Do We Have Omitted Variables Causing Bias?

- After conducting Matching and Regression Adjusted, we recognized the potential for unobserved confounding in this causal relationship due to the subtle impacts that could exist between a myriad of factors such as ambition and drive.
- We suspected that personality traits, ambition, familiar pressures could play a huge role in anxiety. An instrument would be perfect in this case to mitigate this bias. The instrument chosen for this was extracurricular activities.
- This follows all the properties of being a valid instrument theoretically.
  - Relevance: It is relevant to the treatment of high study load since those who are likely to participate in many extra curriculars could be high achievers and might perceive high study load as well.
  - Exclusive: Extra curriculars do not directly affect anxiety levels.
  - Independent: Extra curriculars is independent from the bias.

Requirement for IV

- We compared the estimates of with IV and OLS.
- With OLS: Treatment Effect = 0.92
- With IV: Treatment Effect = 5.48
- Due to the significant difference in treatment effects discovered after inclusion of IV, we may conclude that IV is necessary.
- This implies that with the increase in a unit of future career concern, anxiety increases by 5.48 units. There is likely OVB.
- Correlation between the treatment variable and instrument is 0.54 which is large enough to suggest it is a relevant instrument.

Requirement for IV

Weak Instrument Test:

- The small p-value of the weak instrument test suggest that we reject the null. We have confidence that the instrument is valid and thus is related sufficiently to study load.

Wu Hausman Test:

- On interpreting the Wu Hausman test, we reject the null due to the small p-value and that we do have an endogeneity problem caused by OVB. This supports our use in IV.

```
Diagnostic tests:
                 df1  df2 statistic p-value
Weak instruments   1 1098     460.4  <2e-16 ***
Wu-Hausman         1 1097     318.1  <2e-16 ***
```

Results

- We arrive at the estimate of 5.48. **With an increase in study load, we see a statistically significant effect on anxiety levels, i.e., with a unit increase in study load, anxiety increases by 5.48 units**.
- Checking for marginal percentage effects to remove any abstract meaning from 'units' we convert the variables into log units. We thus arrive at the following conclusion: With a 1% increase in study load, anxiety increases by nearly 2%.

Conclusion

- Our findings represent a departure from previous literature based on:
  - Literature focuses on older students (high school and above) → different generalizability profile
  - Research typically isolated the effect of one specific variable on anxiety or stress. Our model seeks to go above and beyond this by combining factors to form a model that explains student anxiety at a younger age to tackle the issue earlier.
- We statistically prove that increase in study load does indeed have an impact on anxiety levels for students using techniques such as Matching, Regression Adjustment and regression analysis after inclusion of an Instrument Variable to omit possible biases.
- **Limitation**: Some limitations of our study include that the data is primarily self reported and thus includes the possibility of participants underreporting or overreporting their characteristics such as academic performance, study load, anxiety and future career concerns.

References

- https://scialert.net/abstract/?doi=jas.2016.108.112
- https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4d35c0aba8529bfd66482284d79dddfaa8d924ec
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9936312/#ref27
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5414656/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8818275/
- https://drpress.org/ojs/index.php/EHSS/article/view/1691/1619
- https://waterprogramming.wordpress.com/2018/07/23/multivariate-distances-mahalanobis-vs-euclidean