ETC5513
Collaborative and Reproducible Practices

Assignment 3
Exploring Behavioral Correlates of Academic Performance

Chief Examiner: Michael Lydeamore

Author: Pranav Sajeev, Melisa, Ibtesam Ifaz

May 31, 2025

Table of contents

Executive Summary

In this report, we will be looking at how the change from in-person learning to online learning during the period of COVID-19 has affected student's academic performance. The main findings are that the students are able to maintain their academic progress regardless of the change. However, considering the dataset it based on students in India, the groups that were mainly affected by the change were students in Class 10 and Class 12, who had public exams to prepare for in those years.

Introduction

The transition to online learning during the COVID-19 pandemic significantly altered many students' academic environments and daily routines. Widespread increases in screen exposure, reductions in physical activity, and irregular sleep patterns were observed among students. These lifestyle changes have prompted growing concern regarding their potential impact on mental health, particularly with respect to stress, anxiety, and academic performance. Understanding how such factors influence student well-being is essential, especially as digital learning continues to play a central role in education.

This report analyzes survey data collected from students between March and May 2023, with the intent of identifying relationships between lifestyle habits and mental health outcomes. The data set consists of 1,000 observations of students enrolled in both schools and universities. It includes variables such as sleep duration, screen time, exercise frequency, stress levels, and exam-related anxiety. In this report, we seek to determine which behaviors are most closely associated with mental health challenges. The insights gained may support the development of targeted interventions to promote student wellness in remote or hybrid learning settings.

Methodology

This study applied multinomial logistic regression to evaluate the association between multiple independent variables and a categorical outcome with three levels: "Same", "Increase", and "Decrease". The independent variables used in the model were a combination of numerical (e.g., age, screen time) and categorical predictors (e.g., gender, education level).

Figure 1: Multinomial logistic regression model

This method is appropriate when the response variable is nominal with more than two categories and does not follow a natural order. The "Same" category was treated as the reference level in this model.

The regression model estimates the log-odds of each non-reference category relative to the reference category. Separate sets of coefficients are produced for predicting "Improved" and "Declined" outcomes compared to "Same". The model was estimated using maximum likelihood estimation.

The output of the model included regression coefficients, standard errors, and p-values for each predictor variable across both comparisons. The significance of each variable was evaluated using Wald tests, with a threshold of $p < 0.05$ considered statistically significant. This approach allowed for identifying which variables had a meaningful association with changes in the response category.

This methodology provided a clear framework for determining the statistical significance of each predictor and contributed to understanding the factors associated with outcome variations in the study population. For the variables information, please see Table 1

Table 1: Data Set Description

Name	
Gender	Male, Female, Other
Age	15 – 26
Education Level	Class 8, MSc, BTech, BA, Class 11, MTech, Class 9, Class 10, MA, BSc, Class 12
Screen Time (hrs/day)	2 – 12

Table 1: Data Set Description

Sleep Duration (hrs)	4 – 9
Physical Activity (hrs/week)	0 – 10
Stress Level	Medium, High, Low
Anxious Before Exams	No, Yes
Academic Performance Change	Same, Improved, Declined

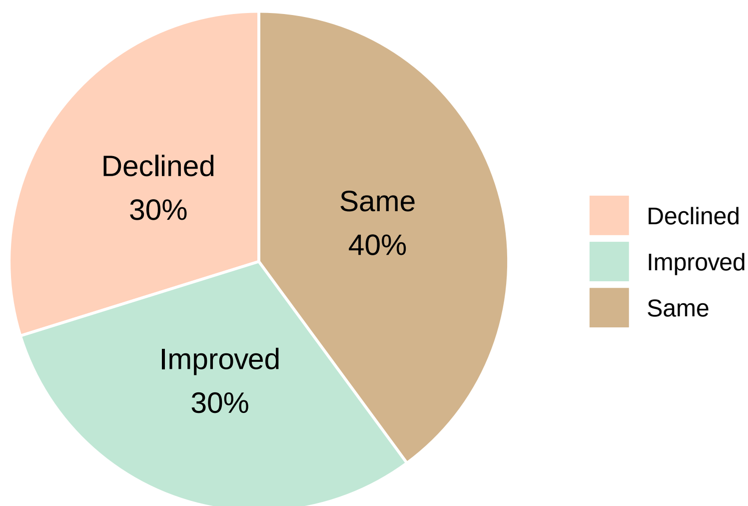

Figure 2: Proportion of Student based on Academic Performance

Based on Figure 2, most students either maintained or improved their academic performance, with only 30% reporting a decline. This reflects a mixed impact of the learning environment, where some students thrived while others struggled.

Data Exploration

For Figure 3, the charts show that 51% of students reported feeling anxious before exams. Most participants were enrolled in master's programs (41%). The gender distribution was balanced between male and female (both 48%), with 5% identifying as "Other." Regarding stress levels, 49% experienced medium stress, 33% low stress, and 18% high stress.

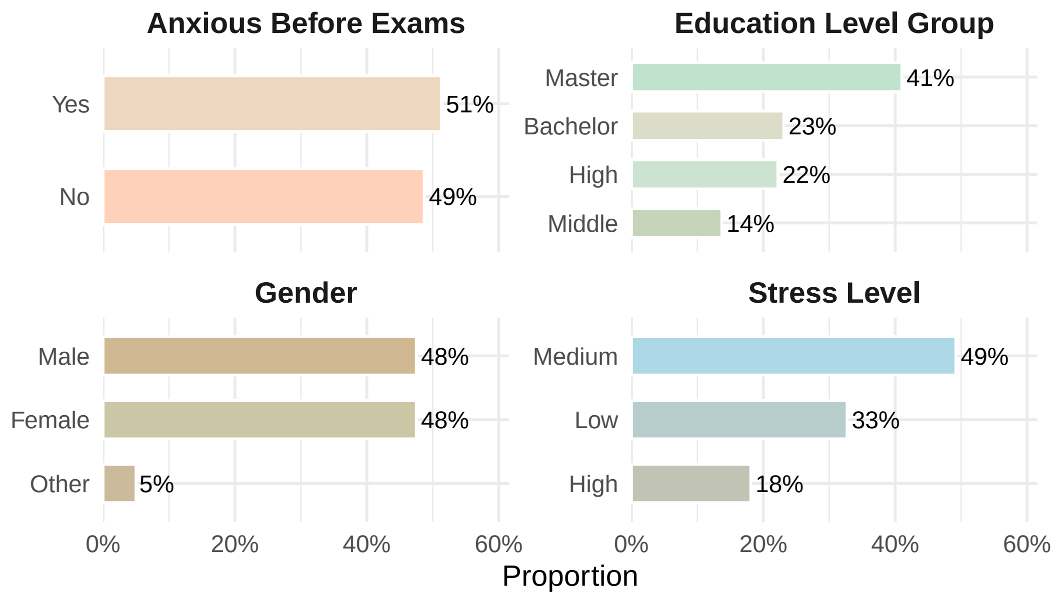

Figure 3: Summary of Students Proportion based on Categorical Variables

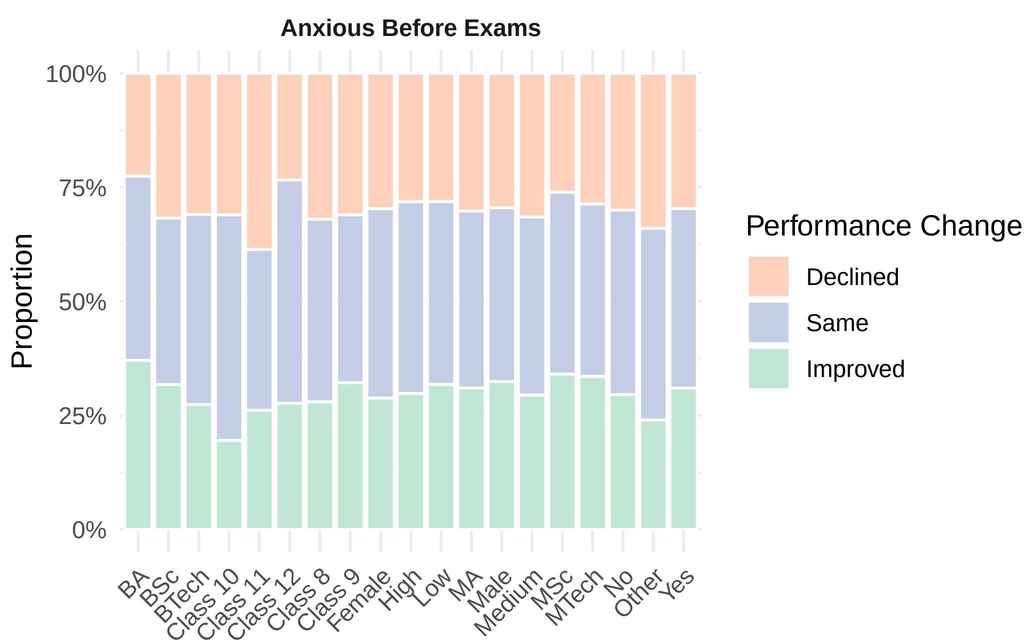

Figure 4: Performance Outcome Breakdown by Categorical Variables

Results

The key variable in this report is academic performance. We compared the observations on the basis of the students' academic performance. Meanwhile Figure 4 shows there is no notable difference in academic performance change across groups based on exam anxiety, gender, or stress level, as the proportions of declined, improved, and unchanged performance remain relatively similar.

In addition we looked into the correlation between screen time and academic performance because research shows that problematic smartphone use is associated with poor mental health outcomes among youth [1]. Therefore, in this analysis we assume that screen time is negatively correlated academic performance, meaning high screen time would result in poor academic performance.

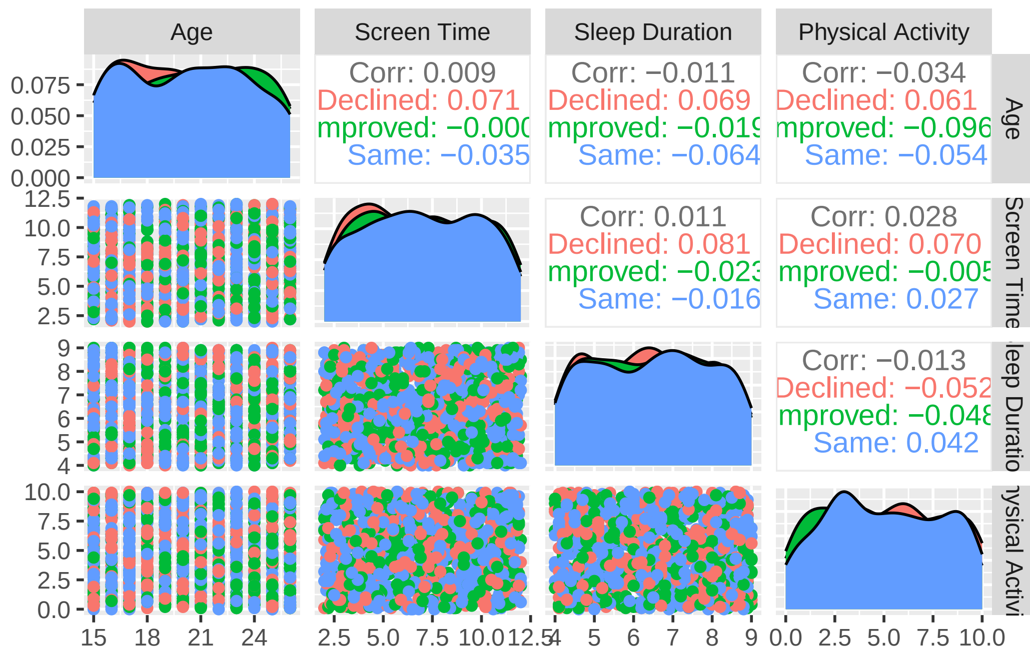

Figure 5: Performance Outcome Breakdown by Categorical Variables

Figure 5 illustrates there is no distinct pattern that classifies academic performance change based on the observed continuous variables. For example, students with less screen time are not consistently associated with improved performance. Correlation values across age, screen time, sleep duration, and physical activity remain weak, suggesting no strong relationship with performance outcomes.

Key finding through data modelling

From Table 2, it tells about students whose performance remained the same, Class 10 students experienced a significantly higher rate of academic decline (p = 0.027). This suggests that the decline in performance was most prominent among Class 10 students.

Table 2: P_value Summary

Outcome	Education_Level_Class 10
Declined	0.5392
Improved	0.0270

Conclusion

From the analysis of the data we can see that the p-values show the only significant variable to the academic performance is the level of education. From this we can infer that the groups mainly affected are students with public exams. This may be due to the fact that the students in those classes require more personal attention that online classes may not be able to provide.

Recommendations

- For data collection, we need more numeric variables as change in performance needs to be recorded more accurately for data analysis.
- For the purpose of comparison, data from before COVID-19 would be highly beneficial to enable more efficient comparisons.
- Use different models to reach solutions. More complex models, like ones that add variable interaction, may be able to find more connections between the variables.
- Additional variables that measure quality of the interaction in the lecture, internet connection, and metrics on motivation to study would greatly increase the utility of the data.

References

[1] Xiaoyu Li et al. "Problematic Smartphone Use and Mental Health in Children and Youth: A Meta-Analysis". In: *Frontiers in Public Health* 11 (2023), p. 1252371. DOI: 10.3389/fpubh.2023.1252371. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1252371/full.