NMC5344: Final Project

Lim Jun Yi Keith, Peng Yixian, Rajendram Rishi. Singh Roy Sourish, Vidya Meenakshi Mohan

Raja Renuga Sri Parameswari

Faculty of Arts and Social Sciences, Department of Communications and New Media

NMC5344: Coding for Communicators

Subhayan Mukerjee

November 14, 2025

Word count: 1,099

Dataset Description & Identification

The dataset comprises 27,901 observations and 18 variables aimed at analysing and predicting students' depression levels. Designed for psychology, data science, and education research, it includes demographic (age, gender, city), academic (CGPA, academic pressure, study satisfaction), and lifestyle factors (sleep duration, dietary habits, work pressure, job satisfaction, work/study hours). Additional variables capture financial stress, family history, and suicidal thoughts. The target variable, depression status, is a binary indicator denoting whether a student is experiencing depression.

Variables this study focuses on are:

1. Depression — Measures if the respondent has feelings of depression; recorded as binary, stored as integer (0 = No, 1= Yes).

2. Suicidal thoughts — Measures if the respondent has experienced suicidal thoughts; recorded as categorical, stored as character ("Yes"/ "No").

3. Academic pressure — Measures level of perceived academic pressure for respondent; recorded on a scale, stored as integer (1-5).

4. Financial stress — Measures level of perceived financial stress for each respondent; recorded on a scale and stored as character (1-5).

5. Sleep duration — Measures self-reported sleep duration of each respondent; recorded as categorical variable, stored as character (e.g. "Less than 5 hours").

6. Work/Study hours — Measures number of hours each respondent studies; recorded as integer (1-12).

7. Study satisfaction level — Measures self-reported study satisfaction level of each respondent on a scale, recorded as integer (1-5).

Dataset Summary

Using sapply() and is.numeric(), numerical variables with valid observations were kept, as seen from the Appendix (Table 1). 'Job satisfaction' and 'Work Pressure' contained only '0' values, and were hence excluded. 'Depression' was not included due to it being a binary variable.

Using the group_by function, summaries for the dataset's subgroups were found: 'Gender' (Table 2), 'City' (Table 3), 'Degree' (Table 4), 'Sleep Duration' (Table 5), 'Dietary Habits' (Table 6), 'Have you ever had suicidal thoughts?' (Table 7), and 'Family History of Mental Illness' (Table 8). Only the descriptive statistics (*M* and *SD*) were reported, as these are the core descriptive statistics for normally distributed data that measure central tendency and variability. For succinctness, only the four largest categories within each subgroup are shown in the tables, especially for variables like 'City' and 'Degree' that possess numerous categories.

Visualisations

Figure 1: Younger students (Class 12) report higher depression rates (70%) but fewer suicidal thoughts (69%); whereas university students have lower depression rates (56.6%-58.9%) but more suicidal thoughts (61.1%-65.3%). Hence, psychological interventions should prioritise emotional coping for younger students since they are more depressed; and crisis support for older students since they are more suicidal.
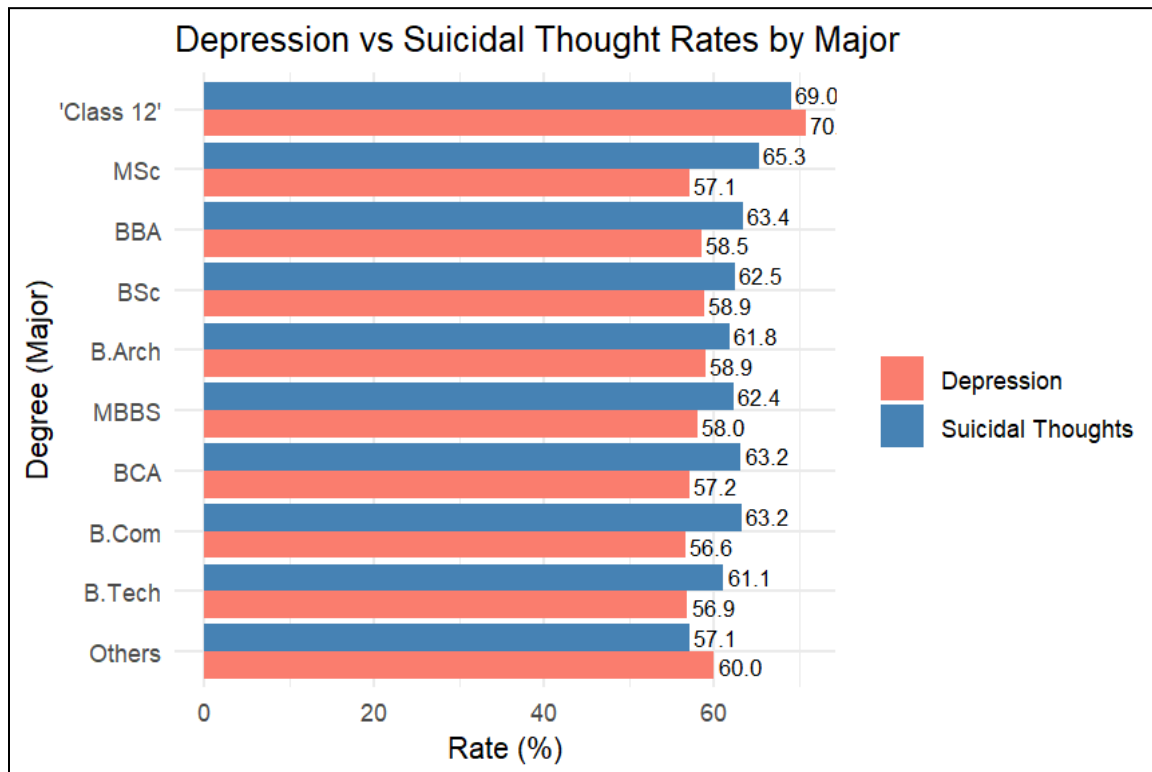
Figure 1

Figure 2: Depression rate increases with both academic and financial stress across all ages, but the effect is strongest among 18-20-year-olds. Among students aged 21-29, academic pressure appears to have a stronger influence on depression, while the 30-40 group shows reduced sensitivity to both stressors. Therefore, targeted financial support, stress-management and academic counselling should be prioritised for students aged 18-24, with more academic-related support for older age groups.

Figure 2

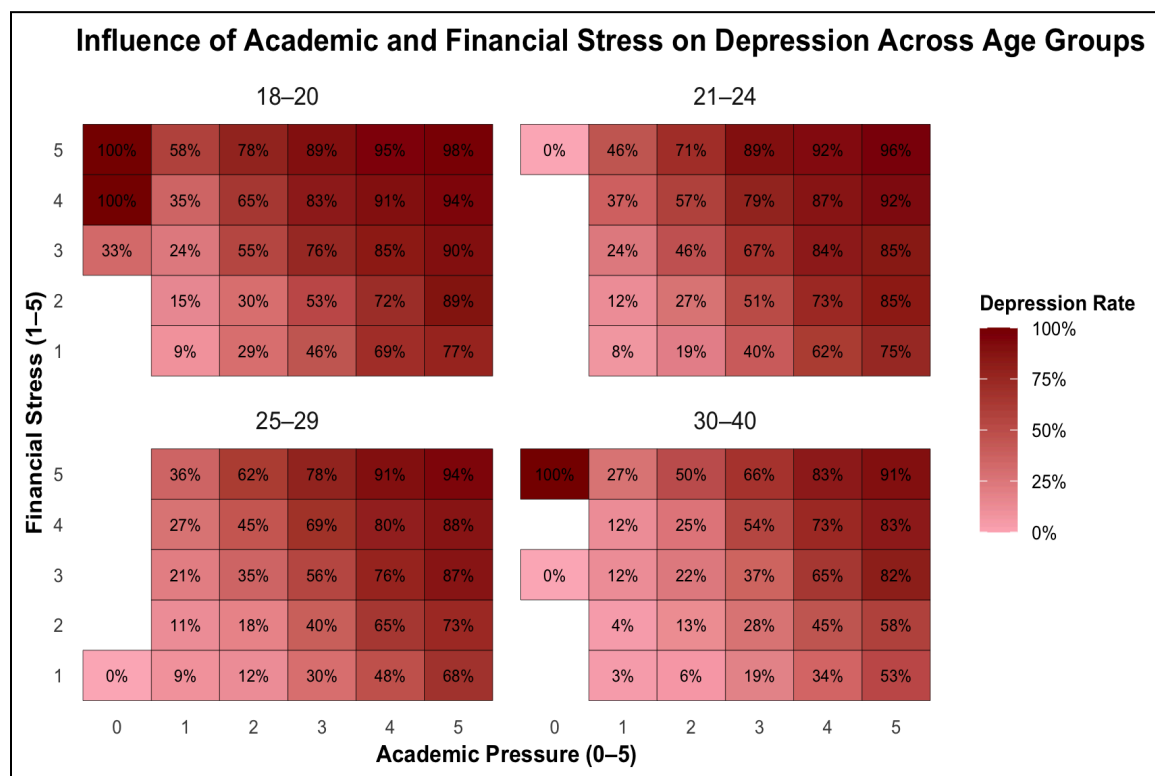Influence of Academic and Financial Stress on Depression Across Age Groups

Figure 3 illustrates sleep duration and study intensity's interaction in predicting depression risk. Depression rates rise with longer study hours. The effect is strongest among students who sleep less than 5 hours. Those studying 11–12 hours with under 5 hours of sleep show the highest depression rate (73%). Contrastingly, students sleeping more than 8 hours maintain the lowest depression rates, even with long study hours — hence, adequate sleep strongly protects against psychological academic burden, preserving students' mental wellbeing.

Figure 3

Depression Rate by Sleep Duration and Study Hours

Sleep Duration	0-4	5-6	7-8	9-10	11-12
5-6 hours	0.42	0.53	0.58	0.66	0.66
7-8 hours	0.43	0.57	0.61	0.68	0.69
Less than 5 hours	0.50	0.63	0.68	0.70	0.73
More than 8 hours	0.37	0.50	0.53	0.61	0.61
Others					

Depression Rate
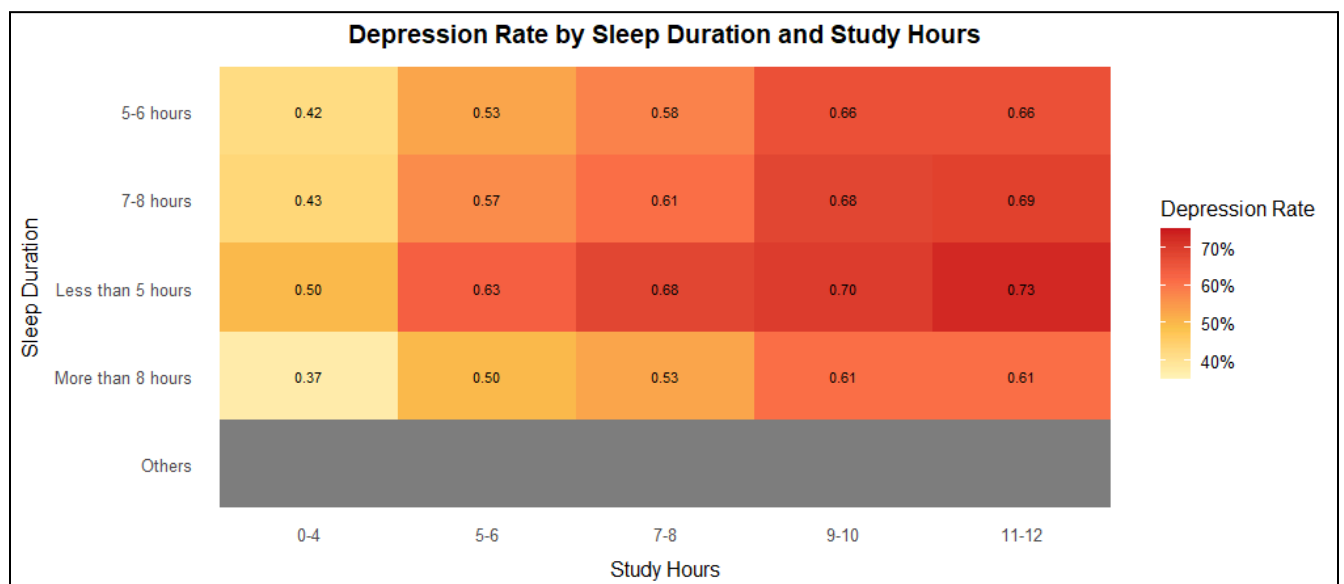- 70%
- 60%
- 50%
- 40%

Study Hours

Figure 4: high academic pressure positively associates with financial stress, while higher study satisfaction seems to be a potential moderator that helps mitigate financial stress's impact on academic pressure. Hence, students facing heavy academic workloads may experience reduced academic stress when financial burdens are eased. This highlights financial aid's importance, not just to enable equal opportunity, but also to support student mental health.

Figure 4

Mean Academic Pressure by Financial Stress and Study Satisfaction Level
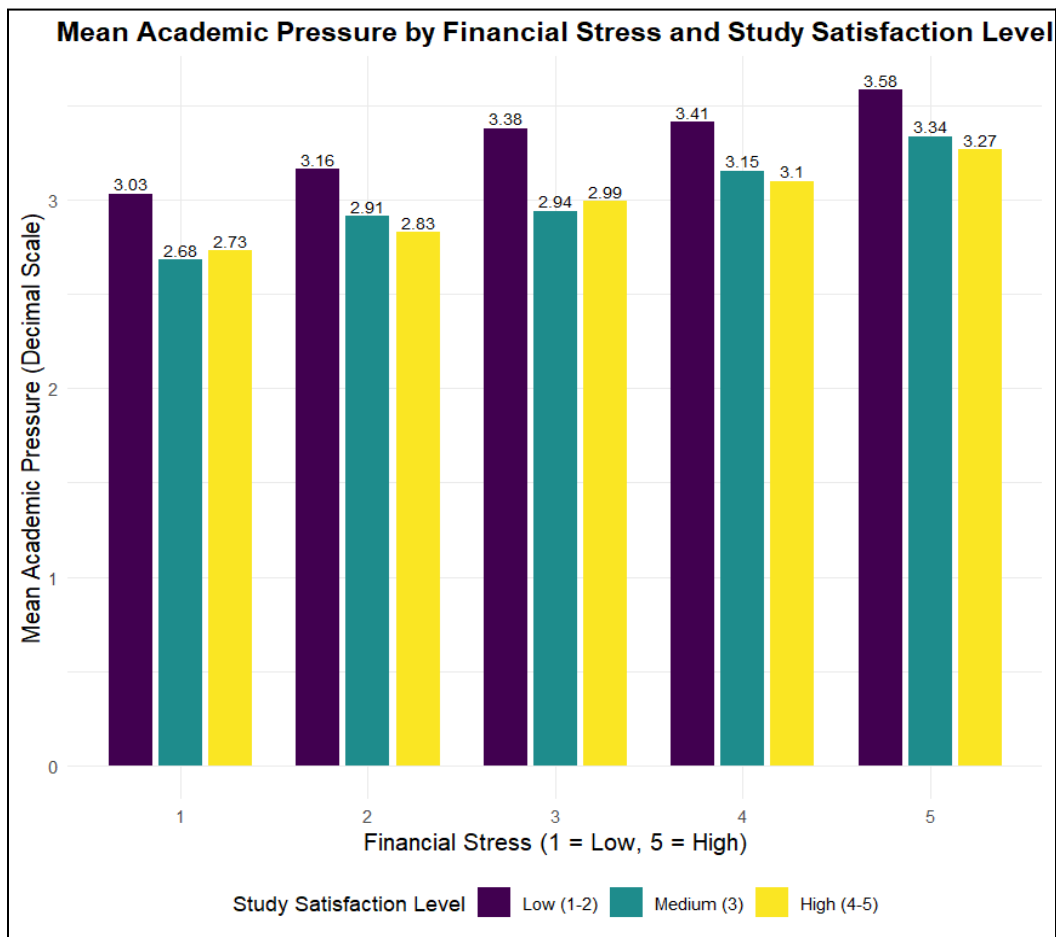
Models

Linear Regression — Financial stress and Academic pressure

Motivation

Students from low-income families often rely on education to support household stability. Higher financial stress increases academic pressure (Figure 4). However, our linear regression predicting CGPA from financial stress ($\beta = 0.006$, $p = 0.326$) was nonsignificant. This supports that financial stress elevates academic pressure and has a negative impact on student mental health, underscoring the need for stronger financial aid systems. Additionally, we examined if study satisfaction moderated the relationship between financial stress and academic pressure.

Model results

The regression examined whether study satisfaction moderates the effect of financial stress on academic pressure. Financial stress showed a significant positive effect ($\beta = 0.144$, $p < .001$), indicating that increases in financial stress correspond to higher academic pressure. Although the effect size is modest, each 1-unit rise in financial stress causes a 0.144 increase in academic pressure.

Study satisfaction demonstrated a significant negative relationship with pressure ($\beta = -0.098$, $p < .001$), suggesting that increased study satisfaction results in lower academic pressure overall. The interaction term between financial stress and study satisfaction was nonsignificant ($\beta = -0.0015$, $p = .716$), indicating that study satisfaction does not moderate the relationship between financial stress and academic pressure.

Overall, financial stress reliably increases academic pressure, while study satisfaction independently lowers it, with no interaction between the two variables.
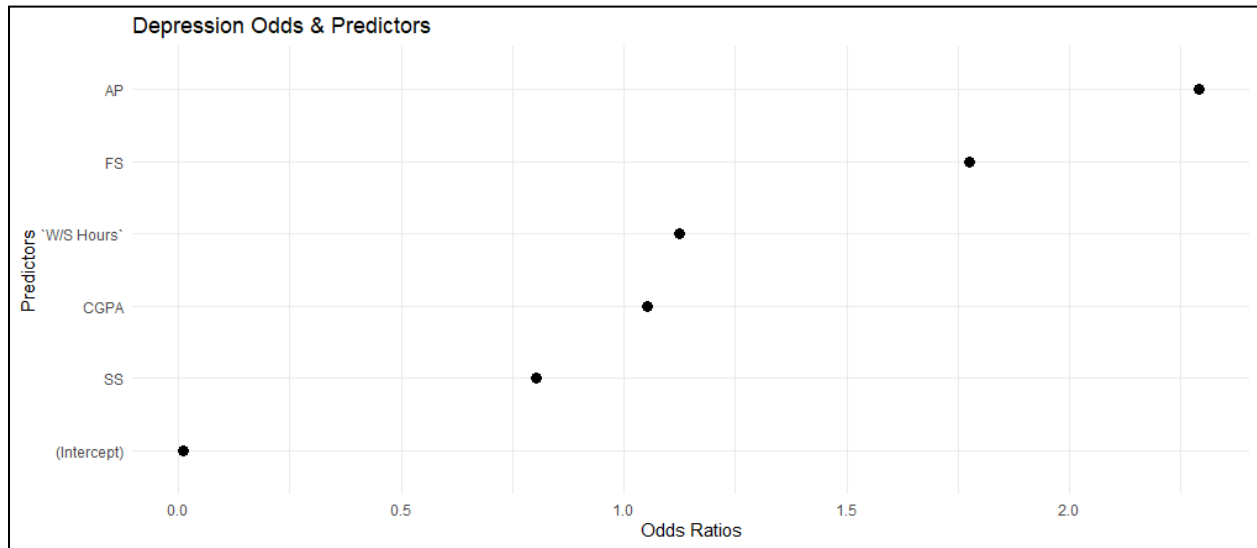
Logistic Regression — Depression Odds & Predictors

Motivation

We sought to find which variable had the greatest impact on predicting depression. Knowing which variables are the biggest (and significant) predictors could help prioritise the right solutions and mental health approaches for students. As we aimed to predict a binary outcome (i.e. Depression is a binary variable), we thus used logistic regression.

Model results

Figure 5: Academic pressure (AP) was the strongest predictor of depression ($\beta = 0.813$, $p < .05$), more than doubling the odds per unit increase. Financial stress (FS) also raised depression risk ($\beta = 0.573$, $p < .05$), followed by Work/Study hours (W/S hours) ($\beta = 0.117$, $p < .05$). CGPA showed a small but significant positive effect ($\beta = 0.0513$, $p < .05$). However, study satisfaction (SS) significantly reduced depression likelihood ($\beta = -0.225$, $p < .05$). Overall, interventions ought to prioritise reducing AP due to its criticality, and SS should be promoted simultaneously as a protective factor.

Figure 5

Comparison of Models

Our two models are incompatible for comparison because:

1. Our logistic regression model predicts a binary outcome; a linear regression model predicts a continuous outcome.

2. Logistic regression's fit is measured with AIC or pseudo-R^2; linear regression's fit is assessed using adjusted R^2.

3. Both models are asking fundamentally different questions — "What predicts academic pressure variation?" versus "What factors increase the odds of depression?"

4. Logistic regression assumes a non-linear S-shaped relationship between variables; linear regression assumes linearity, making their model structures different.

GenAI Use & Acknowledgement

We asked ChatGPT (GPT-4 and GPT-5) for suggestions on improving and refining our code. We also used GenAI to help cut down on our words and ensure clarity. We are responsible for the content and quality of the submitted work.

Appendix

Tables

Table 1

Dataset Summary

	Mean	Median	Standard Deviation	Maximum	Minimum
Age	25.82	25	4.91	59	18
Academic Pressure	3.14	3	1.38	5	0
CGPA	7.66	7.77	1.47	10	0
Study Satisfaction	2.94	3	1.36	5	0
Work/Study Hours	7.16	8	3.71	12	0
Financial Stress	3.14	3	1.44	5	1

Table 2

Gender Subgroup

Gender	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
Female	25.77 (4.90)	3.18 (1.38)	7.60 (1.49)	2.97 (1.37)	7.10 (3.74)	3.15 (1.44)
Male	25.86 (4.91)	3.11 (1.38)	7.70 (1.46)	2.92 (1.35)	7.20 (3.68)	3.13 (1.43)

Table 3

City Subgroup

City	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
Delhi	25.65 (4.92)	3.42 (1.34)	7.48 (1.48)	3.05 (1.30)	7.12 (3.74)	3.20 (1.45)
Agra	25.48 (4.85)	2.98 (1.42)	7.76 (1.45)	2.95 (1.34)	6.95 (3.74)	3.06 (1.41)
Ahmedabad	24.83 (5.04)	3.25 (1.28)	7.82 (1.47)	2.99 (1.34)	7.24 (3.69)	3.25 (1.43)
Bangalore	24.78 (5.12)	3.19 (1.37)	7.82 (1.39)	2.94 (1.38)	7.17 (3.64)	3.17 (1.48)

Table 4

Degree Subgroup

Degree	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
Class 12	20.13 (3.24)	3.36 (1.30)	7.59 (1.45)	3.08 (1.38)	7.28 (3.61)	3.28 (1.41)
B.Arch	25.64 (3.81)	3.06 (1.41)	7.62 (1.44)	2.84 (1.35)	7.27 (3.73)	3.17 (1.45)
B.Com	26.19 (3.92)	3.08 (1.41)	7.67 (1.47)	2.94 (1.31)	7.23 (3.65)	3.12 (1.46)
B.Ed	27.99 (3.90)	3.12 (1.40)	7.85 (1.49)	2.90 (1.32)	7.16 (3.71)	3.08 (1.44)

Table 5

Sleep Duration

Sleep Duration	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
5–6 hours	25.95 (4.93)	3.12 (1.37)	7.69 (1.45)	2.95 (1.35)	7.28 (3.73)	3.11 (1.41)
7–8 hours	25.79 (4.90)	3.14 (1.36)	7.69 (1.48)	2.94 (1.35)	7.27 (3.70)	3.17 (1.42)
<5 hours	25.79 (4.86)	3.23 (1.38)	7.64 (1.46)	2.92 (1.37)	7.19 (3.68)	3.15 (1.47)
>8 hours	25.78 (4.95)	3.05 (1.41)	7.61 (1.49)	2.98 (1.38)	6.85 (3.71)	3.11 (1.44)
Others	27.33 (4.65)	3.17 (0.92)	7.60 (1.59)	3.11 (1.49)	6.78 (4.31)	2.67 (1.46)

Table 6

Dietary Habits Subgroup

Dietary Habits	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
Healthy	26.1 (4.96)	2.99 (1.38)	7.65 (1.48)	3.0 (1.36)	7.02 (3.73)	3.0 (1.43)
Moderate	26.0 (4.94)	3.09 (1.39)	7.66 (1.48)	2.92 (1.38)	7.13 (3.71)	3.08 (1.44)
Others	26.4 (5.14)	3.08 (1.68)	7.53 (1.52)	3.42 (1.44)	4.58 (3.12)	3.17 (1.59)
Unhealthy	25.4 (4.81)	3.30 (1.36)	7.66 (1.45)	2.93 (1.34)	7.29 (3.69)	3.31 (1.43)

Table 7

Suicidal Thoughts Subgroup

Suicidal Thoughts	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
No	26.6 (4.99)	2.67 (1.38)	7.64 (1.49)	3.09 (1.36)	6.57 (3.84)	2.75 (1.42)
Yes	25.4 (4.80)	3.42 (1.31)	7.67 (1.46)	2.86 (1.36)	7.50 (3.59)	3.37 (1.40)

Table 8

Family History of Mental Illness Subgroup

Family History of Mental Illness	Age (M, SD)	Academic Pressure (M, SD)	CGPA (M, SD)	Study Satisfaction (M, SD)	Work/Study Hours (M, SD)	Financial Stress (M, SD)
No	25.8 (4.88)	3.10 (1.39)	7.66 (1.47)	2.95 (1.36)	7.09 (3.73)	3 (1.43)
Yes	25.8 (4.94)	3.18 (1.37)	7.65 (1.47)	2.94 (1.38)	7.22 (3.69)	3 (1.44)