



Course Project Report

STA 2101: Statistics & Probability

Student Name: Sajib Chowdhury

Student ID: 242014141

University of Liberal Arts Bangladesh (ULAB)

Date: October 8, 2025

Abstract

This project investigates the intricate relationship between academic pressure, peer influence, and coping strategies among undergraduate students. Utilizing the publicly available Kaggle dataset “*Academic Stress Level Maintenance Dataset*”, the study applies a range of statistical techniques—including descriptive statistics, probability distributions, hypothesis testing, and regression analysis—to examine patterns in students’ academic stress levels. The goal is to identify the most influential stress factors and understand how different coping mechanisms contribute to maintaining psychological well-being and academic performance.

Contents

1	Milestone 1: Dataset Selection	3
2	Milestone 2: Descriptive Statistics	3
3	Milestone 3: Data Visualization	3
4	Milestone 4: Probability Distributions	3
5	Milestone 5: Hypothesis Testing	3
6	Milestone 6: Regression Analysis	4
7	Milestone 7–12: Further Analysis	4
8	Final Conclusion	4

1 Milestone 1: Dataset Selection

- **Dataset Name:** Academic Stress Level Maintenance Dataset
- **Dataset URL:** <https://www.kaggle.com/datasets/ayeshaimran123/academic-stress-level-maintenance-dataset>
- **Description:** The *Academic Stress Level Maintenance Dataset* contains responses gathered from undergraduate students regarding various dimensions of academic stress. The variables in the dataset capture multiple aspects, including peer influence, academic expectations from family, study environment, and the coping mechanisms students adopt to manage stress. Additionally, it provides data on students' self-assessed competition levels, motivation, and overall stress index.

This dataset was selected because it offers valuable insight into how social and environmental factors influence students' academic well-being. By analyzing this data, the project aims to uncover significant trends and correlations between stress triggers and coping behaviors. The findings from this analysis may contribute to a better understanding of how universities and educators can design effective support systems to promote mental health and reduce stress in academic settings.

2 Milestone 2: Descriptive Statistics

Describe summary statistics of your dataset. Include tables, mean, median, mode, standard deviation, etc.

Example of a table:

3 Milestone 3: Data Visualization

Add graphs and figures using LaTeX. Example:

4 Milestone 4: Probability Distributions

Identify probability distributions in your dataset. Perform fitting, plots, and discuss results.

5 Milestone 5: Hypothesis Testing

State hypotheses, perform tests, and report conclusions.

6 Milestone 6: Regression Analysis

Fit regression models, explain coefficients, and evaluate model fit.

7 Milestone 7–12: Further Analysis

Continue documenting each milestone here as instructed in class.

8 Final Conclusion

Summarize the overall findings of your project. Mention challenges, learning outcomes, and possible future work.

```
import pandas as pd import numpy as np import matplotlib.pyplot as plt from random
import randint import warnings warnings.filterwarnings('ignore')
— 1. Load a Synthetic Dataset (Replace with your df =
pd.read_csv('your_dataset.csv')) — — — np.random.seed(42)forreproducibilityN =
1000Populationsizenumeric_col = 'Academicstress_index'
data = 'Student_ID' : range(1, N + 1), 'Academicstage' :
np.random.choice(['Undergraduate', 'Graduate', 'PHD'], N, p =
[0.6, 0.3, 0.1]), 'StudyHoursPerDay' : np.round(np.random.normal(loc = 5, scale =
2, size = N), 1).clip(min = 0.5), 'Peerpressure_index' :
np.random.randint(1, 6, N), 'Academicstress_index' :
np.round(np.random.normal(loc = 3.5, scale = 1.0, size = N), 2).clip(min =
1.0), 'Copingstrategy' :
np.random.choice(['Intellect', 'EmotionalBreakdown', 'Other'], N, p =
[0.55, 0.3, 0.15])df = pd.DataFrame(data)
— 2. Display first 5 rows — print("1. First 5 rows of the dataset:") print(df.head())
— 3. Report dataset size and population mean —
population_size = len(df)population_mean = df[numeric_col].mean()
print("2. Dataset size (Population N):") print(f"Number of rows (N):
population_size")print(f"Number of columns :
df.shape[1]")print(f"Mean(numeric_col) : population_mean : .2f")
```