



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

SECI 2143-01

PROBABILITY STATISTICAL & DATA ANALYSIS

21/22 - 2

GROUP 4

PROJECT 2

Part 1:

Description on selected data set & proposed analysis

Submitted to:

DR. Azurah A Samah

NAME	MATRIC NUMBER
AI SYAH BINTI MOHD NADZRI	A21EC0011
NUR IMMAL HAYATI BINTI HASMI ANUAR	A21EC0111
AHMAD MIRZA ARMAND BIN SHAZRIL FARIZA	A21EC0006
MUHAMMAD IZAT BIN MD KAMIL	A21EC0082

1.0 DATASET

Understanding students' behavior is an essential part in order to improve students' effectiveness. An ideal student is said to have a balanced life socially and academically. In example, the student's performance as well as behavior is heavily influenced by their hobby, study habits, general behavior etc,. Therefore, the main purpose of this project is to specify and show the key datas for students' behavior as well as to apply the usage of statistical analysis tests in order to prove whether there is a relationship between the datas. To achieve this objective, a few potential variables were selected and a series of test analysis must be carried out.

The dataset that we have chosen regarding students' behavior is a secondary data we have retrieved from [www.Kaggle.com](https://www.kaggle.com) website. The details of the variables, type of variable and its level of measurement is stated in the table below:

Dataset Description

Data Set URL: [Student Behavior | Kaggle](https://www.kaggle.com)

Population: Students from various Universities.

Sample: 235 university students.

Variable (Description)	Variable Type	Level of Measurement
Daily studying time	Quantitative	Interval
College mark	Quantitative	Ratio
Stress level	Qualitative	Ordinal
Time spent on social media & video games	Quantitative	Interval
Gender	Qualitative	Nominal
Degree certificate of completion	Qualitative	Nominal
Department	Qualitative	Nominal
Height (cm)	Quantitative	Ratio

Weight (kg)	Quantitative	Ratio
10th Mark	Quantitative	Ratio
12th Mark	Quantitative	Ratio
Hobby	Qualitative	Nominal
Study time preference	Qualitative	Nominal
Salary expectation	Quantitative	Ratio
Degree of students' likeness towards their degree	Qualitative	Nominal
Possibility of choosing their career based on their degree	Quantitative	Ratio
Time of travel	Quantitative	Interval
Degree of financial status	Qualitative	Ordinal
Degree of doing part-time job	Qualitative	Nominal

2.0 PROPOSED ANALYSIS

Selected Variables	Objectives	Test Analysis & Expected outcome
1. Daily studying time 2. College mark	To test whether the mean of the college mark obtained by the students is equal to the mean of the daily studying time spent by the students.	Analysis: 2-sample of Hypothesis Testing Expected outcome: The mean of the college mark obtained by the students is equal to the mean of the daily studying time spent by the students.
1. Daily studying time 2. Stress Level	To test whether linear relationship between daily studying time and stress level using Pearson's Product Moment Correlation Coefficient, at 95% confidence level.	Analysis: Correlation Analysis Expected outcome: There is a strong linear relationship between the daily studying time and stress level at confidence level 95%. The larger the daily studying time, the higher stress level.
1. College mark 2. Time spent on social media & video games.	To test whether the value of time spent on social media & video games depends on the value of college marks, using college marks as the independent variable(x) and college marks as the dependent variable(y).	Analysis: Regression Analysis Expected outcome: The value of college marks depends on the value of time spent on social media & video games. The higher the time spent on social media & video games the lower the college marks.
1. Gender 2. Degree certificate of completion	To conduct a test between degree certificate of completion by the students and gender. This test is utilized to observe if there is any relationship or relation between the two variables.	Analysis: Chi-Squared Test Expected outcome: The degree certificate of completion and gender variables are not related and independent at 0.05 significance level.