



**UTM**  
**UNIVERSITI TEKNOLOGI MALAYSIA**

---

**FACULTY OF COMPUTING**

**SEMESTER 2 2023/2024**

---

**SECI 1143: PROBABILITY & STATISTICAL DATA ANALYSIS**

**SECTION 01**

**GROUP 05 : BEAUTIES**

**LECTURER: DR. SHARIN HAZLIN HUSPI**

**PROJECT TITLE: DATA ANALYSIS OF STUDENTS PERFORMANCE**

**INDIVIDUAL REFLECTION**

**NAME: SEOW YEN ZHI**

**MATRICS NUMBER: A23CS0177**

Based on this data analysis conducted on the students' performance, I truly benefited a lot. Firstly, it involves selecting a suitable dataset to start this data analysis. This requires a careful consideration to ensure that the dataset chosen is relevant and comprehensive enough to provide meaningful insights. Through this process, I learned about the importance of the data selection and its impact towards the overall analysis. Next, I proceed with identifying the variables that exist in this dataset and determine each of the data type. Initially, it is a challenging part for our team to start this data analysis as we do struggle a lot. However, by having frequent discussion, we managed to come out with a better solution after discussing among ourselves and seeking valuable advices from our lecturer. Although it is challenging but this phase taught me the value of perseverance and teamwork. This collaborative effort among the team is not only the key for our team to overcome the obstacles but also managed to enhanced our problem-solving skills and strengthened our team dynamics. I find it really interesting and fascinating to explore and play around with the variables in the given dataset to investigate the relationship about them. This hands-on experience has given me a new insight towards data analysis and I'm able to feel the spark on applying the concepts we learned in our daily lives. This data analysis project underscored the importance of the concept we learned in our lectures as it is the foundation of the data analysis. Additionally, it has deepened my understanding about the topics covered in the class, this project somehow is a revision class as it enhances my comprehension towards the different tests conducted.

Based on the data analysis on the students' performance, I gain a lot and do learn how to use R Programming in a better way as R Studio is a good tool to assist our calculation in statistic. To have a deeper look into our findings in this data analysis, from the hypothesis testing for two samples conducted, we can conclude that the mean GPA for both genders do not varies a lot. For the correlation test, there is linear relationship exist between students' study time weekly and their GPA. This relationship is then undergoes further testing which is the regression test to learn more about how the students' GPA being affected by their study time weekly. Through the regression test, we are able to have a clearer picture on the relationship between the students' study time weekly and their GPA. This leads us to be able to predict the students' GPA based on their study time weekly. However, it is restricted with certain conditions as the GPA only range from 0.0 to 4.0. Lastly, the ANOVA test is a one-way ANOVA test with unequal sample size. Although this is not covered much in our lecture, our team do put extra effort in studying the way to conduct this test. We are able to come out with the conclusion that the mean GPA for different age group is the same. Based on the data analysis, I am able to identify the main key element to archive better performance is by increasing the study time weekly. Besides, based on the tests, the gender and the age do not affect much on their study performance.